Phase retrieval for wavelet transforms

Irène Waldspurger

***Abstract*—This article describes a new algorithm that solves a particular phase retrieval problem, with important applications in audio processing: the reconstruction of a function from its scalogram, that is, from the modulus of its wavelet transform.**

It is a multiscale iterative algorithm, that reconstructs the signal from low to high frequencies. It relies on a new reformulation of the phase retrieval problem, that involves the holomorphic extension of the wavelet transform. This reformulation allows to propagate phase information from low to high frequencies.

Numerical results, on audio and non-audio signals, show that reconstruction is precise and stable to noise. The algorithm has a linear complexity in the size of the signal, up to logarithmic factors, and can thus be applied to large signals.

***Index Terms*—Phase retrieval, scalogram, iterative algorithms, multiscale method**

I. Introduction

The spectrogram is an ubiquitous tool in audio analysis and processing, eventually after being transformed into mel-frequency cepstrum coefficients (MFCC) by an averaging along frequency bands. A very similar operator, yielding the same kind of results, is the modulus of the wavelet transform, sometimes called scalogram.

The phase of the wavelet transform (or the windowed Fourier transform in the case of the spectrogram) contains information that cannot be deduced from the single modulus, like the relative phase of two notes with different frequencies, played simultaneously. However, this information does not seem to be relevant to understand the perceptual content of audio signals [1], [2], and only the modulus is used in applications. To clearly understand which information about the audio signal is kept or lost when the phase is discarded, it is natural to consider the corresponding inverse problem: to what extent is it possible to reconstruct a function from the modulus of its wavelet transform? The study of this problem mostly begun in the early 80's [3], [4].

On the applications side, solving this problem allows to resynthesize sounds after some transformation has been applied to their scalogram. Examples include blind source separation [5] or audio texture synthesis [6].

The reconstruction of a function from the modulus of its wavelet transform is an instance of the class of *phase retrieval problems*, where one aims at reconstructing an unknown signal $x \in \mathbb{C}^n$ from linear measurements $Ax \in \mathbb{C}^m$, whose phase has been lost and whose modulus only is available, $|Ax|$. These problems are known to be difficult to solve.

Two main families of algorithms exist. The first one consists of iterative algorithms, like gradient descents or alternate projections [7], [8]. In the case of the spectrogram, the oldest such algorithm is due to Griffin and Lim [3]. These methods are simple to implement but, because the phase retrieval problem is non-convex, they are not guaranteed to converge towards the correct solution, and often get stuck in local minima. For measurement matrices A that are chosen at random (according to a suitable distribution), this problem can be overcome with a careful initialization of the algorithm [9]–[11]. However, these methods do not apply when measurements are not random. In the case of the spectrogram or scalogram, the reconstructed signals tend to present auditive artifacts. For the spectrogram, the performances can be improved by applying the algorithm window by window and not to the whole signal at the same time [12]. If additional information on the nature of the audio signal is available, it can also be taken into account in the algorithm [13], [14]. Nevertheless, in the general case, the reconstruction results are still perfectible.

More recently, convexification methods have been proposed [15], [16]. For generic phase retrieval problems, these methods are guaranteed to return the true solution with high probability when the measurement matrix A is chosen at random. In the case of the spectrogram or scalogram, the matrix is not random and the proof does not hold. However, numerical experiments on small signals indicate that the reconstruction is in general correct [17], [18]. Unfortunately, these methods have a high complexity, making them difficult to use for phase retrieval problems whose size exceeds a few hundred.

In this article, we present a new algorithm for the reconstruction of a function from its scalogram. As convexification methods, it offers a reconstruction of high quality. However, it has the complexity of an iterative method (roughly proportional to the size of the signal, up to logarithmic factors) and can be applied to large signals. The memory it requires is also proportional to the size of the signal.

The algorithm is multiscale: it performs the reconstruction frequency band by frequency band, from the lowest frequencies up to the highest ones.

The main idea of this algorithm is to introduce an equivalent formulation of the phase retrieval problem (by using the analytic extension of the wavelet transform).

This reformulation gives an explicit method to propagate towards higher scales the phase information reconstructed at the low scales. Moreover, the local optimization algorithm naturally derived from this reformulation, although non-convex, seems very robust to the problem of local minima.

Additionally:

- we introduce a multigrid error correction method, to detect and correct eventual errors of the reconstruction algorithm afterwards

The author is with the Département d'informatique de l'École Normale Supérieure, Paris (e-mail: waldspur@di.ens.fr).

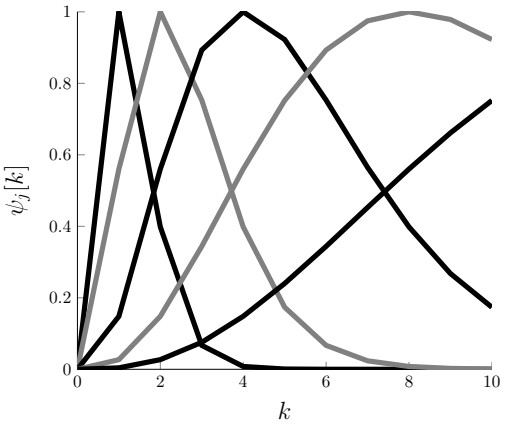


Figure 1. Example of wavelets; the figure shows ψ_j for $J-5 \leq j \leq J$ in the Fourier domain. Only the real part is displayed.

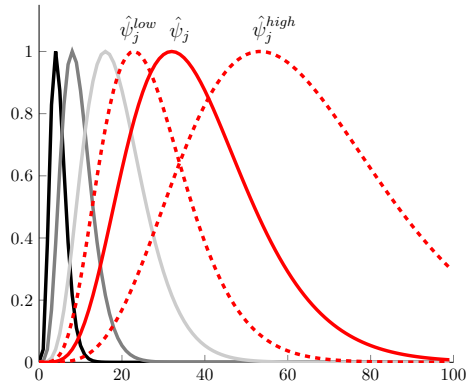


Figure 2. $\psi_J, ..., \psi_{j+1}, \psi_j$ (in the Fourier domain), along with ψ_j^{low} and ψ_j^{high} (dashed lines)

- we use our algorithm to numerically study the intrinsic stability of the phase retrieval problem, and highlight the role played by the sparsity or non-sparsity of the wavelet coefficients.

Section II is devoted to definitions and notations. We explain our reformulation of the phase retrieval problem in Section III, prove its equivalence with the classical formulation, and describe its advantages. In Section IV, we describe the resulting algorithm. In Section V, we discuss the superiority of multiscale algorithms over non-multiscale ones. Finally, in Section VI, we give numerical results, and empirically study the stability of the underlying phase retrieval problem.

The source code, and several reconstruction examples, are available at:

http://www.di.ens.fr/∼waldspurger/wavelets_phase_retrieval.html

II. Definitions and assumptions

All signals $f[n]$ are of finite length N. Their discrete Fourier transform is defined by:

$$\hat{f}[k] = \sum_{s=0}^{N-1} f[n] e^{-2\pi i \frac{kn}{N}} \qquad k = 0, ..., N-1$$

and the convolution always refers to the circular convolution.

We define a family of wavelets $(\psi_j)_{0 \leq j \leq J}$ by:

$$\hat{\psi}_j[k] = \hat{\psi}(a^j k) \qquad k = 0, ..., N-1$$

where the *dilation factor* a can be any number in $(1; +\infty)$ and $\psi : \mathbb{R} \to \mathbb{C}$ is a fixed *mother wavelet*. We assume that J is sufficiently large so that $\hat{\psi}_J$ is negligible outside a small set of points. An example is shown in Figure 1.

The *wavelet transform* is defined by:

$$\forall f \in \mathbb{R}^N, \quad Wf = \{f \star \psi_j\}_{0 \leq j \leq J}$$

The problem we consider here consists in reconstructing functions from the modulus of their wavelet transform:

$$\text{Reconstruct } f \text{ from } \{|f \star \psi_j|\}_{0 \leq j \leq J}$$

Multiplying a function by a unitary complex does not change the modulus of its wavelet transform, so we only aim at reconstructing functions up to multiplication by a unitary complex, that is *up to a global phase*.

All signals are assumed to be analytic:

$$\hat{f}[k] = 0 \quad \text{when } N/2 < k \leq N-1 \tag{1}$$

Equivalently, we could assume the signals to be real but set the $\hat{\psi}_j[k]$ to zero for $N/2 < k \leq N-1$.

III. Reformulation of the phase retrieval problem

In the first part of this section, we reformulate the phase retrieval problem for the wavelet transform, by introducing two auxiliary wavelet families.

We then describe the two main advantages of this reformulation. First, it allows to propagate the phase information from the low-frequencies to the high ones, and so enables us to perform the reconstruction scale by scale. Second, from this reformulation, we can define a natural objective function to locally optimize approximate solutions. Although non-convex, this function has few local minima; hence, the local optimization algorithm is efficient.

A. Introduction of auxiliary wavelets and reformulation

Let us fix $r \in]0; 1[$ and define:

$$\forall k = 0, ..., N-1, \qquad \hat{\psi}_j^{low}[k] = \hat{\psi}_j[k] r^k$$
$$\hat{\psi}_j^{high}[k] = \hat{\psi}_j[k] r^{-k}$$

This definition is illustrated by Figure 2. The wavelet $\hat{\psi}_j^{low}$ has a lower characteristic frequency than $\hat{\psi}_j$ and $\hat{\psi}_j^{high}$ a higher one. The following theorem explains how to rewrite a condition on the modulus of $f \star \psi_j$ as a condition on $f \star \psi_j^{low}$ and $f \star \psi_j^{high}$.

Theorem III.1. *Let $j \in \{0, ..., J\}$ and $g_j \in (\mathbb{R}^+)^N$ be fixed. Let Q_j be the function whose Fourier transform is:*

$$\hat{Q}_j[k] = r^k \widehat{g_j^2}[k] \tag{2}$$
$$\forall k = \left\lfloor \frac{N}{2} \right\rfloor - N + 1, ..., \left\lfloor \frac{N}{2} \right\rfloor$$

For any $f \in \mathbb{C}^N$ satisfying the analycity condition (1)*, the following two properties are equivalent:*

1) $|f \star \psi_j| = g_j$
2) $(f \star \psi_j^{low})(\overline{f \star \psi_j^{high}}) = Q_j$

Proof. The proof consists in showing that the second inequality is the analytic extension of the first one, in a sense that will be precisely defined.

For any function $h : \{0, ..., N-1\} \to \mathbb{C}$, let $P(h)$ be:

$$\forall z \in \mathbb{C} \qquad P(h)(z) = \sum_{k=\lfloor \frac{N}{2} \rfloor - N + 1}^{\lfloor \frac{N}{2} \rfloor} \hat{h}[k] z^k$$

Up to a change of coordinates, $P(|f \star \psi_j|^2)$ and $P(g_j^2)$ are equal to $P((f \star \psi_j^{low})\overline{(f \star \psi_j^{high})})$ and $P(Q_j)$:

Lemma III.2. *For any f satisfying the analycity condition* (1)*, and for any $z \in \mathbb{C}$:*

$$P(|f \star \psi_j|^2)(rz) = P((f \star \psi_j^{low})\overline{(f \star \psi_j^{high})})(z)$$
$$\text{and} \quad P(g_j^2)(rz) = P(Q_j)(z)$$

This lemma is proved in the appendix B. It implies the result because then:

$$\begin{aligned}
& |f \star \psi_j| = g_j \\
\iff & |f \star \psi_j|^2 = g_j^2 \\
\iff & \forall z, P(|f \star \psi_j|^2)(z) = P(g_j^2)(z) \\
\iff & \forall z, P(|f \star \psi_j|^2)(rz) = P(g_j^2)(rz) \\
\iff & \forall z, P((f \star \psi_j^{low})\overline{(f \star \psi_j^{high})})(z) = P(Q_j)(z) \\
\iff & (f \star \psi_j^{low})\overline{(f \star \psi_j^{high})} = Q_j
\end{aligned}$$

□

By applying simultaneously Theorem III.1 to all indexes j, we can reformulate the phase retrieval problem $|f \star \psi_j| = g_j, \forall j$ in terms of the $f \star \psi_j^{low}$'s and $f \star \psi_j^{high}$'s.

Corollary III.3 (Reformulation of the phase retrieval problem)**.** *Let $(g_j)_{0 \leq j \leq J}$ be a family of signals in $(\mathbb{R}^+)^N$. For each j, let Q_j be defined as in* (2)*. Then the following two problems are equivalent:*

Find f satisfying (1) *such that:*

$$\forall j, \quad |f \star \psi_j| = g_j$$

$\iff$

Find f satisfying (1) *such that:*

$$\forall j, \quad (f \star \psi_j^{low})\overline{(f \star \psi_j^{high})} = Q_j \tag{3}$$

B. Phase propagation across scales

This new formulation yields a natural multiscale reconstruction algorithm, in which one reconstructs f frequency band by frequency band, starting from the low frequencies.

Indeed, once $f \star \psi_J, ..., f \star \psi_{j+1}$ have been reconstructed, it is possible to estimate $f \star \psi_j^{low}$ by deconvolution. This deconvolution is stable to noise because, if r is sufficiently small, then the frequency band covered by ψ_j^{low} is almost included in the frequency range covered by $\psi_J, ..., \psi_{j+1}$ (see figure 2). From $f \star \psi_j^{low}$, one can reconstruct $f \star \psi_j^{high}$, using (3):

$$f \star \psi_j^{high} = \frac{\overline{Q}_j}{\overline{f \star \psi_j^{low}}} \tag{4}$$

Finally, one reconstructs $f \star \psi_j$ from $f \star \psi_j^{high}$ and $f \star \psi_j^{low}$.

The classical formulation of the phase retrieval problem does not allow the conception of such a multiscale algorithm. Indeed, from $f \star \psi_J, ..., f \star \psi_{j+1}$, it is not possible to directly estimate $f \star \psi_j$: it would require performing a highly unstable deconvolution. The introduction of the two auxiliary wavelet families is essential.

C. Local optimization of approximate solutions

From the reformulation (3), we can define a natural objective function for the local optimization of approximate solutions to the phase retrieval problem. This is also possible from the classical formulation but the objective function then has numerous local minima, which make it difficult to globally minimize. Empirically, the objective function associated to the reformulation suffers dramatically less from this drawback.

The objective function has $2J + 3$ variables: $(h_j^{low})_{0 \leq j \leq J}$, $(h_j^{high})_{0 \leq j \leq J}$ and f. The intuition is that f is the signal we aim at reconstructing and the h_j^{low}, h_j^{high} correspond to the $f \star \psi_j^{low}$'s and $f \star \psi_j^{high}$'s. The objective function is:

$$\begin{aligned}
& \text{obj}(h_J^{low}, ..., h_0^{low}, h_J^{high}, ..., h_0^{high}, f) \\
= & \sum_{j=0}^{J} ||h_j^{low} \overline{h_j^{high}} - Q_j||_2^2 \\
& + \lambda \sum_{j=0}^{J} \left(||f \star \psi_j^{low} - h_j^{low}||_2^2 + ||f \star \psi_j^{high} - h_j^{high}||_2^2 \right)
\end{aligned} \tag{5}$$

We additionally constrain the variables $(h_j^{low})_{0 \leq j \leq J}$ and $(h_j^{high})_{0 \leq j \leq J}$ to satisfy:

$$\forall j = 0, ..., J-1 \qquad h_j^{low} \star \psi_{j+1}^{high} = h_{j+1}^{high} \star \psi_j^{low} \tag{6}$$

The first term of the objective ensures that the equalities (3) are satisfied, while the second term and the additional constraint (6) enforce the fact that the h_j^{low}'s and h_j^{high}'s must be the wavelet transforms of the *same* function f.

The parameter λ is a positive real number. In our implementation, we choose a small λ, so that the first term dominates over the second one.

A similar objective function can also be derived directly from the classical formulation. However, empirically, it appears to have much more local minima than the function (5); hence, it is more difficult to efficiently minimize. A possible explanation is that the set of zeroes of the first term of (5) (which dominates the second one) has a smaller dimension when the reformulation is used, thus reducing the number of local minima it contains.

IV. Description of the algorithm

In this section, we describe our implementation of the multiscale reconstruction algorithm introduced in Section III. We explain the general organization in Paragraph IV-A. We then describe our exhaustive search method for solving phase retrieval problems of very small size (paragraph IV-B), which our algorithm uses to initialize the multiscale reconstruction. In Paragraph IV-C, we describe an additional multigrid correction step.

A. Organization of the algorithm

We start by reconstructing $f \star \psi_J$ from $|f \star \psi_J|$ and $|f \star \psi_{J-1}|$. We use an exhaustive search method, described in the next paragraph IV-B, which takes advantage of the fact that $\hat{\psi}_J$ and $\hat{\psi}_{J-1}$ have very small supports.

We then reconstruct the components of the wavelet transform scale by scale, as described in Section III.

At each scale, we reconstruct $f \star \psi_j^{low}$ by propagating the phase information coming from $f \star \psi_J, ..., f \star \psi_{j+1}$ (as explained in Paragraph III-B). This estimation can be imprecise, so we refine it by local optimization, using the objective function defined in Paragraph III-C, from which we drop all the terms with higher scales than j. The local optimization algorithm we use in the implementation is L-BFGS ([19]), a low-memory approximation of a second order method.

We then reconstruct $f \star \psi_j^{high}$ by the equation (4).

At the end of the reconstruction, we run a few steps of the classical Gerchberg-Saxton algorithm to further refine the estimation.

The pseudo-code 1 summarizes the structure of the implementation.

Algorithm 1 overview of the algorithm

Input: $\{|f \star \psi_j|\}_{0 \leq j \leq J}$

1: Initialization: reconstruct $f \star \psi_J$ by exhaustive search
2: **for all** $j = J : (-1) : 0$ **do**
3: Estimate $f \star \psi_j^{low}$ by phase propagation
4: Refine the values of $f \star \psi_J^{low}, ..., f \star \psi_j^{low}, f \star \psi_J^{high}, ..., f \star \psi_{j+1}^{high}$ by local optimization
5: Do an error correction step
6: Refine again
7: Compute $f \star \psi_j^{high}$ by $f \star \psi_h^{high} = \overline{Q}_j / \overline{f \star \psi_j^{low}}$
8: **end for**
9: Compute f
10: Refine f with Gerchberg-Saxton

Output: f

B. Reconstruction by exhaustive search for small problems

In this paragraph, we explain how to reconstruct $f \star \psi_j$ from $|f \star \psi_j|$ and $|f \star \psi_{j-1}|$ by exhaustive search, in the case where the support of $\hat{\psi}_j$ and $\hat{\psi}_{j-1}$ is small.

This is the method we use to initialize our multiscale algorithm. It is also useful for the multigrid error correction step described in the next paragraph IV-C.

Lemma IV.1. *Let $m \in \mathbb{R}^N$ and $K \in \mathbb{N}^*$ be fixed. We consider the problem:*

$$\begin{aligned} &\text{Find } g \in \mathbb{C}^N \text{ s.t. } |g| = m \\ &\qquad\qquad \text{and } Supp(\hat{g}) \subset \{1, ..., K\} \end{aligned}$$

This problem has at most 2^{K-1} solutions, up to a global phase, and there exist a simple algorithm which, from m and N, returns the list of all possible solutions.

Proof. This lemma is a consequence of classical results about the phase retrieval problem for the Fourier transform. It can for example be derived from [20]. We give a proof in the appendix A. □

We apply this lemma to $m = |f \star \psi_j|$ and $|f \star \psi_{j-1}|$, and construct the lists of all possible $f \star \psi_j$'s and of all possible $f \star \psi_{j-1}$'s. The true $f \star \psi_j$ and $f \star \psi_{j-1}$ are the only pair in these two lists which satisfy the equality:

$$(f \star \psi_j) \star \psi_{j-1} = (f \star \psi_{j-1}) \star \psi_j$$

This solves the problem.

The number of elements in the lists is exponential in the size of the supports of $\hat{\psi}_j$ and $\hat{\psi}_{j-1}$, so this algorithm has a prohibitive complexity when the supports become large. Otherwise, our numerical experiments show that it works well.

C. Error correction

When the modulus are noisy, there can be errors during the phase propagation step. The local optimization generally corrects them, if run for a sufficient amount of time, but, for the case where some errors are left, we add, at each scale, a multigrid error correction step. This step does not totally remove the errors but greatly reduces their amplitude.

1) Principle: First, we determine the values of n for which $f \star \psi_j^{low}[n]$ and $f \star \psi_{j+1}^{high}[n]$ seems to have been incorrectly reconstructed. We use the fact that $f \star \psi_j^{low}$ and $f \star \psi_{j+1}^{high}$ must satisfy:

$$(f \star \psi_j^{low}) \star \psi_{j+1}^{high} = (f \star \psi_{j+1}^{high}) \star \psi_j^{low}$$

The points where this equality does not hold provide a good estimation of the places where the values of $f \star \psi_j^{low}$ and $f \star \psi_{j+1}^{high}$ are erroneous.

We then construct a set of smooth "windows" $w_1, ..., w_S$, whose supports cover the interval on which errors have been found (see figure 3), such that each window has a small support. For each s, we reconstruct $(f \star \psi_j^{low}).w_s$ and $(f \star \psi_{j+1}^{high}).w_s$, by expressing these functions as the solutions to phase retrieval problems of small size, which we can solve by the exhaustive search method described in Paragraph IV-B.

As w_s is smooth, the multiplication by w_s approximately commutes with the convolution by ψ_j, ψ_{j+1}:

$$|(f.w_s) \star \psi_j| \approx |(f \star \psi_j).w_s| = w_s |f \star \psi_j|$$

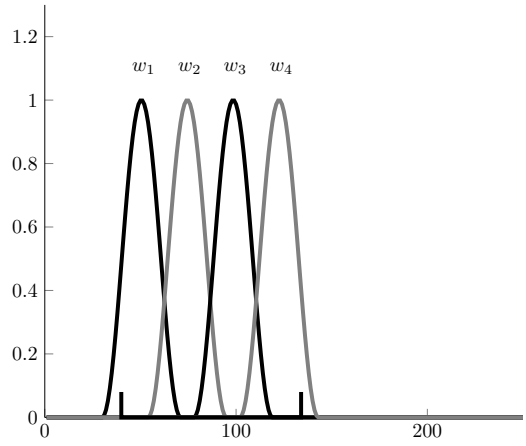


Figure 3. Four window signals, whose supports cover the interval in which errors have been detected

$$|(f.w_s) \star \psi_{j+1}| \approx |(f \star \psi_{j+1}).w_s| = w_s|f \star \psi_{j+1}|$$

The wavelets ψ_j and ψ_{j+1} have a small support in the Fourier domain, if we truncate them to the support of w_s, so we can solve this problem by exhaustive search, and reconstruct $(f.w_s) \star \psi_j$ and $(f.w_s) \star \psi_{j+1}$.

From $(f.w_s) \star \psi_j$ and $(f.w_s) \star \psi_{j+1}$, we reconstruct $(f \star \psi_j^{low}).w_s \approx (f.w_s) \star \psi_j^{low}$ and $(f \star \psi_{j+1}^{high}).w_s \approx (f.w_s) \star \psi_{j+1}^{high}$ by deconvolution.

2) Usefulness of the error correction step: The error correction step does not perfectly correct the errors, but greatly reduces the amplitude of large ones.

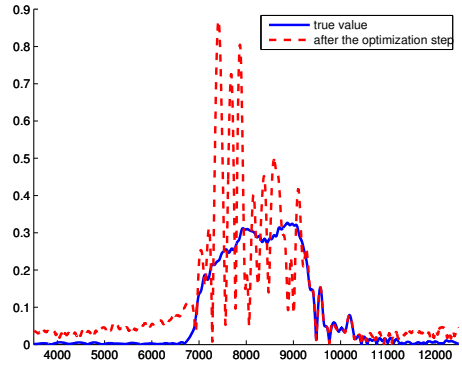
Figure 4 shows an example of this phenomenon. It deals with the reconstruction of a difficult audio signal, representing a human voice saying "I'm sorry". Figure 4a shows $f \star \psi_7^{low}$ after the multiscale reconstruction at scale 7, but before the error correction step. The reconstruction presents large errors. Figure 4b shows the value after the error correction step. It is still not perfect but much closer to the ground truth.

So the error correction step must be used when large errors are susceptible to occur, and turned off otherwise: it makes the algorithm faster without reducing its precision.

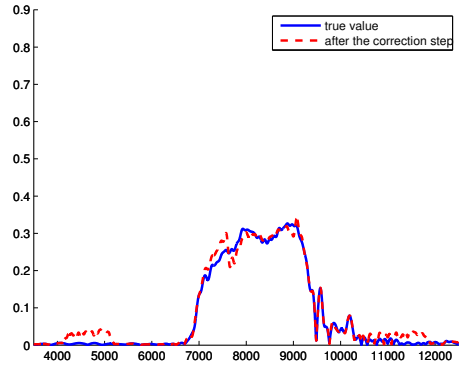
Figure 5 illustrates this affirmation by showing the mean reconstruction error for the same audio signal as previously. When 200 iterations only are allowed at each local optimization step, there are large errors in the multiscale reconstruction; the error correction step significantly reduces the reconstruction error. When 2000 iterations are allowed, all the large errors can be corrected during the local optimization steps and the error correction step is not useful.

V. Multiscale versus non-multiscale

Our reconstruction algorithm has very good reconstruction performances, mainly because it uses the reformulation of the phase retrieval problem introduced in Section III. However, the quality of its results is also due to its multiscale structure. It is indeed known that, for the reconstruction of functions from their spectrogram or scalogram, multiscale algorithms perform better than non-multiscale ones [12], [21].



(a)



(b)

Figure 4. For an audio signal, the reconstructed value of $f \star \psi_7^{low}$ at the scale 7 of the multiscale algorithm, in modulus (dashed line); the solid line represents the ground true. (a) Before the error correction step (b) After the error correction step

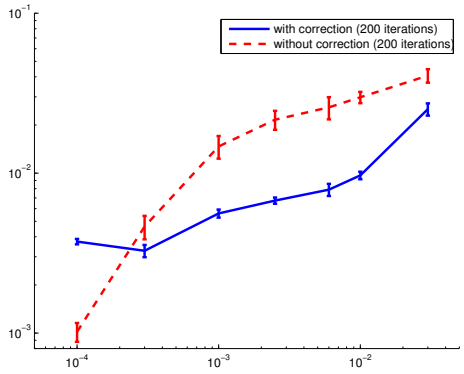
In this section, we propose two justifications for this phenomenon (paragraph V-A). We then introduce a multiscale version of the classical Gerchberg-Saxton algorithm, and numerically verify that it yields better reconstruction results than the usual non-multiscale version (paragraph V-B).

A. Advantages of the multiscale reconstruction

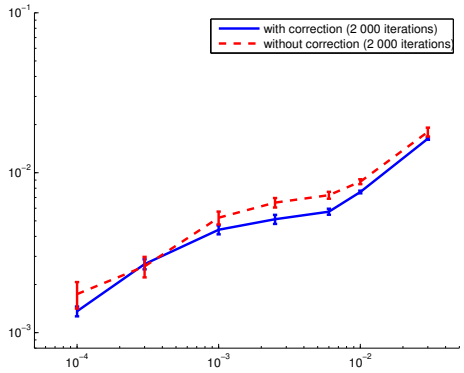
At least two factors can explain the superiority of multiscale methods, where the $f \star \psi_j$'s are reconstructed one by one, and not all at the same time.

First, they can partially remedy the possible ill-conditioning of the problem. In particular, if the $f \star \psi_j$'s have very different norms, then a non-multiscale algorithm will be more sensitive to the components with a high norm. It may neglect the information given by $|f \star \psi_j|$, for the values of j such that this function has a small norm. With an multiscale algorithm where all the $|f \star \psi_j|$'s are successively considered, this happens less frequently.

Second, iterative algorithms, like Gerchberg-Saxton, are very sensitive to the choice of their starting point (hence the care given to their initialization in the literature [9], [10]). If all the components are reconstructed at the same time and the starting point is randomly chosen, the algorithm almost never



(a)



(b)

Figure 5. Mean reconstruction error (9) as a function of the noise, for an audio signal representing a human voice. (a) Maximal number of iterations per local optimization step equal to 200 (b) Maximal number equal to 2000.

converges towards the correct solution: it gets stuck in a local minima. In a multiscale algorithm, the starting point at each scale can be chosen so as to be consistent with the values reconstructed at lower scales; it yields much better results.

B. Multiscale Gerchberg-Saxton

To justify the efficiency of the multiscale approach, we introduce a multiscale version of the classical Gerchberg-Saxton algorithm [8] (by alternate projections) and compare its performances with the non-multiscale algorithm.

The multiscale algorithm reconstructs $f \star \psi_J$ by exhaustive search (paragraph IV-B).

Then, for each j, once $f \star \psi_J, ..., f \star \psi_{j+1}$ are reconstructed, an initial guess for $f \star \psi_j$ is computed by deconvolution. The frequencies of $f \star \psi_j$ for which the deconvolution is too unstable are set to zero. The regular Gerchberg-Saxton algorithm is then simultaneously applied to $f \star \psi_J, ..., f \star \psi_j$.

We test this algorithm on realizations of Gaussian random processes (see Section VI-B for details), of various lengths. On Figure 6, we plot the mean reconstruction error obtained with the regular Gerchberg-Saxton algorithm and the error

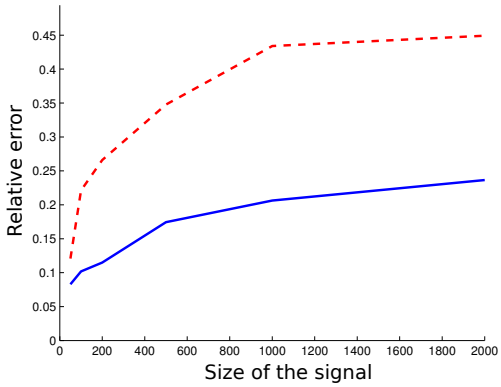


Figure 6. Mean reconstruction error, as a function of the size of the signal; the solid blue line corresponds to the multiscale algorithm and the dashed red one to the non-multiscale one.

obtained with the multiscale version (see Paragraph VI-A for the definition of the reconstruction error).

None of the algorithms is able to perfectly reconstruct the signals, in particular when their size increases. However, the multiscale algorithm clearly yields better results, with a mean error approximately twice smaller.

VI. Numerical results

In this section, we describe the behavior of our algorithm. We compare it with Gerchberg-Saxton and with *PhaseLift*. We show that it is much more precise than Gerchberg-Saxton. It is comparable with *PhaseLift* in terms of precision, but significantly faster, so it allows to reconstruct larger signals.

The performances strongly depend on the type of signals we consider. The main source of difficulty for our algorithm is the presence of small values in the wavelet transform, especially in the low frequencies.

Indeed, the reconstruction of $f \star \psi_j^{high}$ by the equation (4) involves a division by $f \star \psi_j^{low}$. When $f \star \psi_j^{low}$ has small values, this operation is unstable and induces errors.

As we will see in Section VI-C, the signals whose wavelet transform has many small values are also the signals for which the phase retrieval problem is the least stable (in the sense that two functions can have wavelet transforms almost equal in modulus without being close in l^2-norm). This suggests that this class of functions is intrinsically the most difficult to reconstruct; it is not an artifact of our algorithm.

We describe our experimental setting in Paragraph VI-A. In Paragraph VI-B, we give detailed numerical results for various types of signals. In Paragraph VI-C, we use our algorithm to investigate the stability to noise of the underlying phase retrieval problem. Finally, in Paragraph VI-D, we study the influence of various parameters on the quality of the reconstruction.

A. Experimental setting

At each reconstruction trial, we choose a signal f and compute its wavelet transform $\{|f \star \psi_j|\}_{0 \leq j \leq J}$. We corrupt

it with a random noise n_j:

$$h_j = |f \star \psi_j| + n_j \tag{7}$$

We measure the amplitude of the noise in l^2-norm, relatively to the l^2-norm of the wavelet transform:

$$\text{amount of noise} = \frac{\sqrt{\sum_j ||n_j||_2^2}}{\sqrt{\sum_j ||f \star \psi_j||_2^2}} \tag{8}$$

We run the algorithm on the noisy wavelet transform $\{h_j\}_{0 \leq j \leq J}$. It returns a reconstructed signal f_{rec}. We quantify the reconstruction error by the difference, in relative l^2-norm, between the modulus of the wavelet transform of the original signal f and the modulus of the wavelet transform of the reconstructed signal f_{rec}:

$$\text{reconstruction error} = \frac{\sqrt{\sum_j || \, |f \star \psi_j| - |f_{rec} \star \psi_j| \, ||_2^2}}{\sqrt{\sum_j ||f \star \psi_j||_2^2}} \tag{9}$$

Alternatively, we could measure the difference between f and f_{rec}:

$$\text{error on the signal} = \frac{||f - f_{rec}||_2}{||f||_2} \tag{10}$$

But we know that the reconstruction of a function from the modulus of its wavelet transform is not stable to noise [22]. So we do not hope the difference between f and f_{rec} to be small. We just want the algorithm to reconstruct a signal f_{rec} whose wavelet transform is close to the wavelet transform of f, in modulus. Thus, the reconstruction error (9) is more relevant to measure the performances of the algorithm.

In all the experiments, unless otherwise specified, we use dyadic Morlet wavelets, to which we subtract Gaussian functions of small amplitude so that they have zero mean:

$$\hat{\psi}(\omega) = \exp(-p(\omega - 1)^2) - \beta \exp(-p\omega^2)$$

where $\beta > 0$ is chosen so that $\hat{\psi}(0) = 0$ and the parameter p is arbitrary (it controls the frequency bandwidth of the wavelets). For $N = 256$, our family of wavelets contains eight elements, which are plotted on Figure 18a. The performances of the algorithm strongly depend on the choice of the wavelet family; this is discussed in Paragraph VI-D1.

The maximal number of iterations per local optimization step is set to 10000 (with an additional stopping criterion, so that the 10000-th iteration is not always reached). We study the influence of this parameter in Paragraph VI-D2.

The noises are realizations of Gaussian white noises.

The error correction step described in Paragraph IV-C is always turned on.

Gerchberg-Saxton is applied in a multiscale fashion, as described in Paragraph V-B, which yields better results than the regular implementation.

We use *PhaseLift* [23] with ten steps of reweighting, followed by 2000 iterations of the Gerchberg-Saxton algorithm. In our experiments with *PhaseLift*, we only consider signals

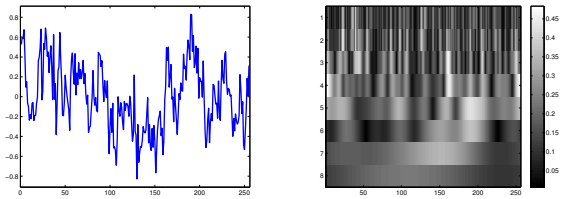


Figure 7. Realization of a Gaussian process (left) and modulus of its wavelet transform (right)

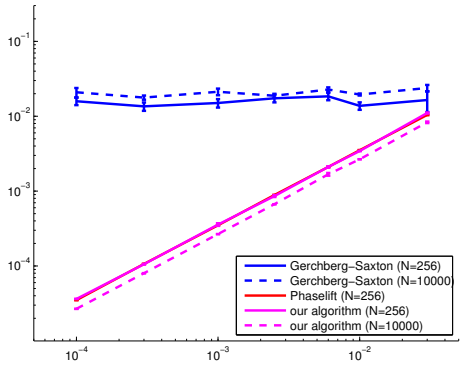


Figure 8. Mean reconstruction error as a function of the noise, for Gaussian signals of size $N = 256$ or 10000

of size $N = 256$. Handling larger signals is difficult with a straightforward Matlab implementation.

B. Results

We describe four classes of signals, whose wavelet transforms have more or less small values. For each class, we plot the reconstruction error of our algorithm, Gerchberg-Saxton and *PhaseLift* as a function of the noise error.

1) Realizations of Gaussian random processes: We first consider realizations of Gaussian random processes. A signal f in this class is defined by:

$$\begin{aligned} \hat{f}[k] &= \frac{X_k}{\sqrt{k+1}} \quad \text{if } k \in \{1, ..., N/2\} \\ &= 0 \qquad\qquad \text{if not} \end{aligned}$$

where $X_1, ..., X_{N/2}$ are independent realizations of complex Gaussian centered variables. The role of the $\sqrt{k+1}$ is to ensure that all components of the wavelet transform approximately have the same l^2-norm (in expectation). An example is displayed on Figure 7, along with the modulus of its wavelet transform.

The wavelet transforms of these signals have few small values, disposed in a seemingly random pattern. This is the most favorable class for our algorithm.

The reconstruction results are shown in Figure 8. Even for large signals ($N = 10000$), the mean reconstruction error is proportional to the input noise (generally 2 or 3 times smaller); this is the best possible result. The performances of *PhaseLift* are exactly the same, but Gerchberg-Saxton often fails.

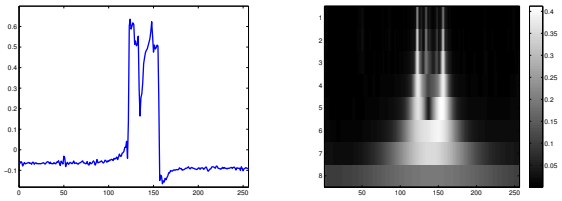


Figure 9. Line from an image (left) and modulus of its wavelet transform (right)

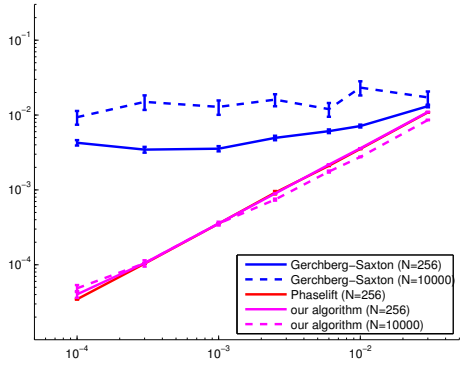


Figure 10. Mean reconstruction error as a function of the noise, for lines extracted from images, of size $N = 256$ or 10000

2) Lines from images: The second class consists in lines randomly extracted from photographs. These signals have oscillating parts (corresponding to the texture zones of the initial image) and smooth parts, with large discontinuities in between. Their wavelet transforms generally contain a lot a small values, but, as can be seen in Figure 9, the distribution of these small values is particular. They are more numerous at high frequencies and the non-small values tend to concentrate on vertical lines of the time-frequency plane.

This distribution is favorable to our algorithm: small values in the wavelet transform are mostly a problem when they are in the low frequencies and prevent the correct initialization of the reconstruction at medium or high frequencies. Small values at high frequencies are not a problem.

Indeed, as in the case of Gaussian signals, the reconstruction error is proportional to the input noise (figure 10). This is also the case for *PhaseLift* but not for Gerchberg-Saxton.

3) Sums of a few sinusoids: The next class of signals contains sums of a few numbers of sinusoids, multiplied by a window function w to avoid boundary effects. Formally, a

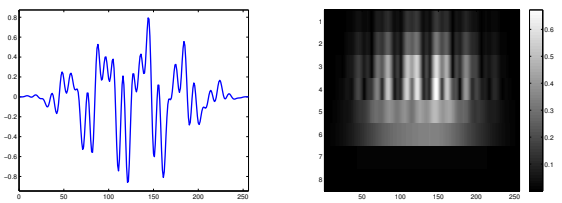


Figure 11. Random sum of sinusoids, multiplied by a window function (left) and modulus of its wavelet transform (right)

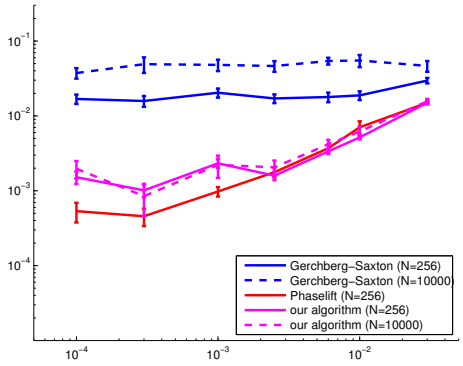


Figure 12. Mean reconstruction error as a function of the noise, for random sums of sinusoids multiplied by a window function, of size $N = 256$ or 10000

signal in this class is of the form:

$$f[n] = \left[\sum_{k=1}^{N/2} \alpha_k \exp\left(i\frac{2\pi kn}{N}\right)\right] \times w[n]$$

where the α_k are zero with high probability and realizations of complex Gaussian centered variables with small probability.

The wavelet transforms of these signals often have components of very small amplitude, which may be located at any frequential scale (figure 11). This can prevent the reconstruction.

The results are on Figure 12. Our algorithm performs much better than Gerchberg-Saxton but the results are not as good as for the two previous classes of signals.

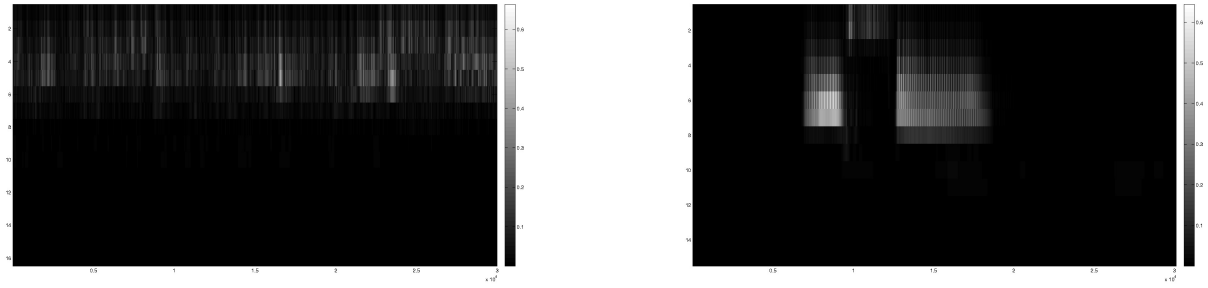
In most reconstruction trials, the signal is correctly reconstructed, up to an error proportional to the noise. But, with a small probability, the reconstruction fails. The same phenomenon occurs for *PhaseLift*.

The probability of failure seems a bit higher for *PhaseLift* than for our algorithm. For example, when the signals are of size 256 and the noise has a relative norm of 0.01%, the reconstruction error is larger than the noise error 20% of the time for *PhaseLift* and only 10% of the time for our algorithm. However, *PhaseLift* has a smaller mean reconstruction error because, in these failure cases, the result it returns, although not perfect, is more often close to the truth: the mean reconstruction error in the failure cases is 0.2% for *PhaseLift* versus 1.7% for our algorithm.

4) Audio signals: Finally, we test our algorithm on real audio signals. These signals are difficult to reconstruct because they do not contain very low frequencies (as the human ear cannot hear them, these frequencies are not included in the recordings), so the first components of their wavelet transforms are very small.

The reconstruction results may vary from one audio signal to the other. We focus here on two representative examples.

The first signal is an extract of five seconds of a musical piece played by an orchestra (the *Flight of the Bumblebee*, by Rimsky-Korsakov). Figure 13a shows the modulus of its



(a) (b)

Figure 13. Wavelet transforms of the audio signals (a) Rimsky-Korsakov (b) "I'm sorry"

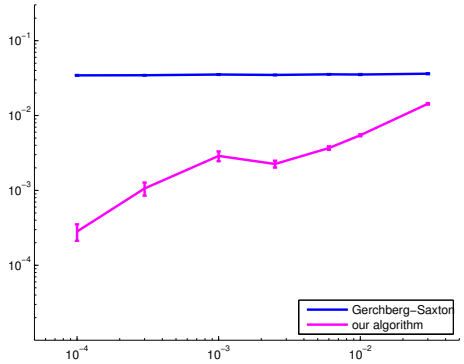


Figure 14. mean reconstruction error as a function of the noise, for the audio signal "Rimsky-Korsakov"

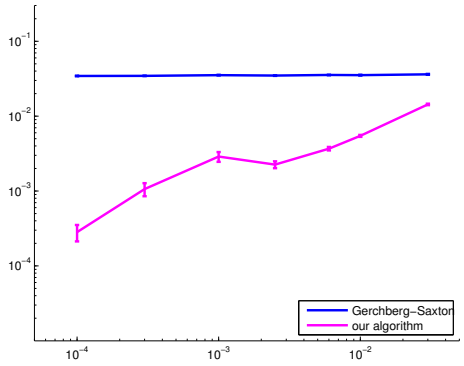


Figure 15. mean reconstruction error as a function of the noise, for the audio signal "I'm sorry"

wavelet transform. It has 16 components and 9 of them (the ones with lower characteristic frequencies) seem negligible, compared to the other ones. However, its non-negligible components have a moderate number of small values.

The second signal is a recording of a human voice saying "I'm sorry" (figure 13b). The low-frequency components of its wavelet transform are also negligible, but even the high-frequency components tend to have small values, which makes the reconstruction even more difficult.

The results are presented in Figures 14 and 15. For relatively high levels of noise (0.5% or higher), the results, in the sense of the l^2-norm, are satisfying: the reconstruction error is smaller or equal to the amount of noise.

In the high precision regime (that is, for 0.1% of noise or less), the lack of low frequencies does not allow a perfect reconstruction. Nevertheless, the results are still good: the reconstruction error is of the order of 0.1% or 0.2% when the noise error is below 0.1%. More iterations in the optimization steps can further reduce this error. By comparison, the reconstruction error with Gerchberg-Saxton is always several percent, even when the noise is small.

C. Stability of the reconstruction

In this section, we use our reconstruction algorithm to investigate the stability of the reconstruction. From [22], we know that the reconstruction is not globally stable to noise: the reconstruction error (9) can be small (the modulus of the wavelet transform is almost exactly reconstructed), even if the error on the signal (10) is not small (the difference between the initial signal and its reconstruction is large).

We show that this phenomenon can occur for all classes of signals, but is all the more frequent when the wavelet transform has a lot of small values, especially in the low frequency components.

We also experimentally show that, when this phenomenon happens, the original and reconstructed signals have their wavelet transforms $\{f \star \psi_j(t)\}_{j\in\mathbb{Z},t\in\mathbb{R}}$ equal up to multiplication by a phase $\{e^{i\phi_j(t)}\}_{j\in\mathbb{Z},t\in\mathbb{R}}$, which varies slowly in both j and t, except maybe at the points where $f \star \psi_j(t)$ is close to zero. This has been conjectured in [22].

We perform a large number of reconstruction trials, with various reconstruction parameters. This gives us a large number of pairs (f, f_{rec}), such that $\forall j,t, |f\star\psi_j(t)| \approx |f_{rec}\star\psi_j(t)|$. For each one of these pairs, we compute:

$$\text{error on the modulus} = \frac{\sqrt{\sum_j || \, |f \star \psi_j| - |f_{rec} \star \psi_j| \, ||_2^2}}{\sqrt{\sum_j ||f \star \psi_j||_2^2}} \quad (9)$$

$$\text{error on the signal} = \frac{||f - f_{rec}||_2}{||f||_2} \quad (10)$$

The results are plotted on Figure 16, where each point corresponds to one reconstruction trial. The x-coordinate represents the error on the modulus and the y-coordinate the error on the signal.

We always have:

$$\text{error on the modulus} \leq C \times (\text{error on the function})$$

with C a constant of the order of 1. This is not surprising because the modulus of the wavelet transform is a Lipschitz operator, with a constant close to 1.

As expected, the converse inequality is not true: the error on the function can be significantly larger than the error on the modulus. For each class, an important number of reconstruction trials yield errors such that:

$$\text{error on the signal} \approx 30 \times \text{error on the modulus}$$

For realizations of Gaussian random processes or for lines extracted from images (figures 16a and 16b), the ratio between the two errors never exceeds 30 (except for one outlier). But for sums of a few sinusoids (16c) or audio signals (16d), we may even have:

$$\text{error on the signal} \geq 100 \times \text{error on the modulus}$$

So instabilities appear in the reconstruction of all kinds of signals, but are stronger for sums of sinusoids and audio signals, that is for the signals whose wavelet transforms have a lot of small values, especially in the low frequencies.

These results have a theoretical justification. [22] explain how, from any signal f, it is possible to construct g such that $|f \star \psi_j| \approx |g \star \psi_j|$ for all j but $f \not\approx g$ in the l^2-norm sense.

The principle of the construction is to multiply each $f \star \psi_j(t)$ by a phase $e^{i\phi_j(t)}$. The function $(j,t) \rightarrow e^{i\phi_j(t)}$ must be chosen so that it varies slowly in both j and t, except maybe at the points (j,t) where $f \star \psi_j(t)$ is small. Then there exist a signal g such that $(f \star \psi_j(t))e^{i\phi_j(t)} \approx g \star \psi_j(t)$ for any j,t. Taking the modulus of this approximate equality yields:

$$\forall j,t \qquad |f \star \psi_j(t)| \approx |g \star \psi_j(t)|$$

However, we may not have $f \approx g$.

This construction works for any signal f (unless the wavelet transform is very localized in the time frequency domain), but the number of possible $\{e^{i\phi_j(t)}\}_{j,t}$ is larger when the wavelet transform of f has a lot of small values, because the constraint of slow variation is relaxed at the points where the wavelet transform is small (especially when the small values are in the low frequencies). This is probably why instabilities occur for all kinds of signals, but more frequently when the wavelet transforms have a lot of zeroes.

From our experiments, it seems that the previous construction describes all the instabilities: when the wavelet transforms of f and f_{rec} have almost the same modulus and f is not close to f_{rec}, then the wavelet transforms of f and f_{rec} are equal up to slow-varying phases $\{e^{i\phi_j(t)}\}_{j,t}$.

Figure 17 shows an example. The signal is a sum of sinusoids. The relative difference between the modulus is 0.3%, but the difference between the initial and reconstructed signals is more than a hundred times larger; it is 46%. The right subfigure shows the difference between the phases of the two wavelet transforms. It indeed varies slowly, in both time and frequency (actually, it is almost constant along the frequency axis), and a bit faster at the extremities, where the wavelet transform is closer to zero.

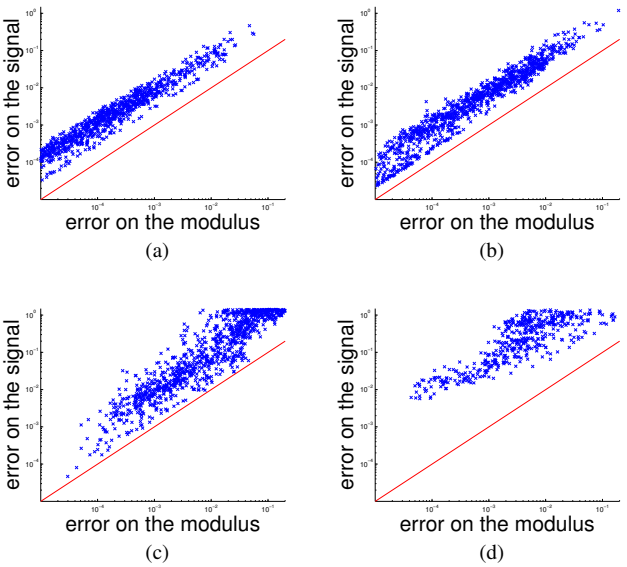


Figure 16. error on the signal (10) as a function of the error on the modulus of the wavelet transform (9), for several reconstruction trials; the red line $y = x$ is here to serve as a reference (a) Gaussian signals (b) lines from images (c) sums of sinusoids (d) audio signal "I'm sorry"

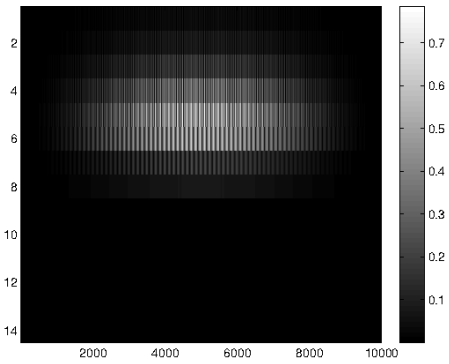
D. Influence of the parameters

In this paragraph, we analyze the importance of the two main parameters of the algorithm: the choice of the wavelets (paragraph VI-D1) and the number of iterations allowed per local optimization step (paragraph VI-D2).

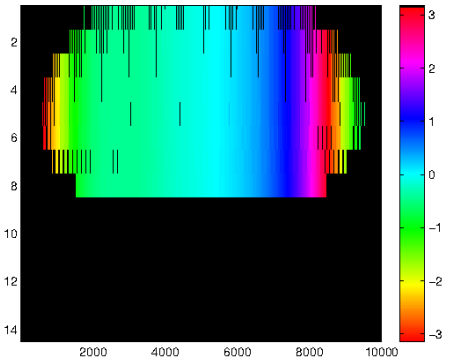
1) Choice of the wavelets: Two properties of the wavelets are especially important: the exponential decay of the wavelets in the Fourier domain (so that the Q_j's (2) are correctly computed) and the amount of overlap between two neighboring wavelets (if the overlap is too small, then $f \star \psi_J, ..., f \star \psi_{j+1}$ contain not much information about $f \star \psi_j$ and the multiscale approach is less efficient).

We compare the reconstruction results for four families of wavelets.

The first family (figure 18a) is the one we used in all the previous experiments. It contains dyadic Morlet wavelets. The second family (figure 18b) also contains Morlet wavelets, with a smaller bandwidth (Q-factor ≈ 8) and a dilation factor of $2^{1/8}$ instead of 2. This is the kind of wavelets used in audio processing. The third family (figure 18c) consists in dyadic Laplacian wavelets $\hat{\psi}(\omega) = \omega^2 e^{1-\omega^2}$. Finally, the wavelets of the fourth family (figure 18d) are (derivatives of) Gammatone wavelets.

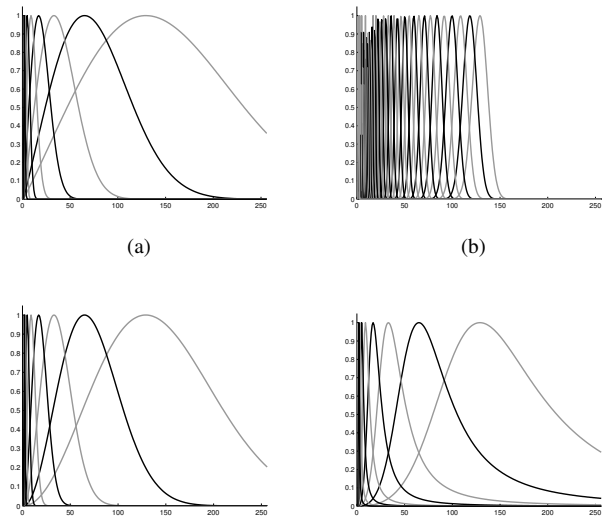


(a)



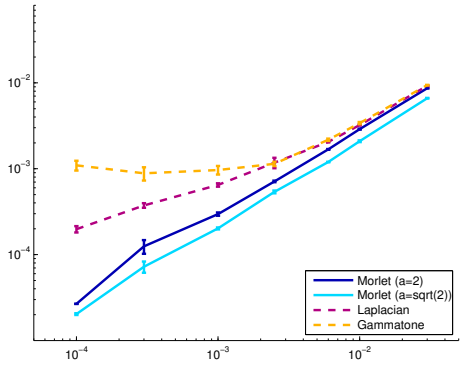
(b)

Figure 17. (a) modulus of the wavelet transform of a signal (b) phase difference between the wavelet transform of this signal and of its reconstruction (black points correspond to places where the modulus is too small for the phase to be meaningful)

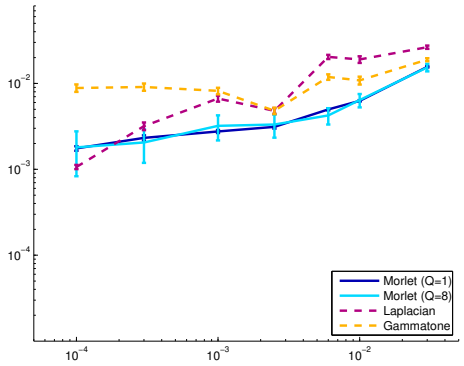


(a) (b) (c) (d)

Figure 18. Four wavelet families. (a) Morlet (b) Morlet with dilation factor $2^{1/8}$ (c) Laplacian (d) Gammatone



(a)



(b)

Figure 19. Mean reconstruction error as a function of the noise for the four wavelet families displayed in 18. (a) Lines from images (b) Audio signal "I'm sorry"

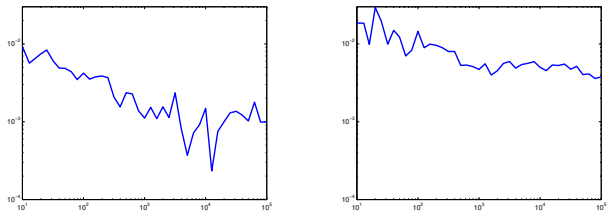
Figure 19 displays the mean reconstruction error as a function of the noise, for two classes of signals: lines randomly extracted from natural images and audio signals.

Morlet wavelets have a fast decay and consecutive wavelets overlap well. This does not depend upon the dilation factor so the reconstruction performances are similar for the two Morlet families (figures 19a and 19b).

Laplacian wavelets are similar, but the overlap between consecutive wavelets is not as good. So Laplacian wavelets globally have the same behavior as Morlet wavelets but require significantly more computational effort to reach the same precision. Figures 19a and 19b have been generated with a maximal number of iterations per optimization step equal to 30000 (instead of 10000) and the reconstruction error is still larger.

The decay of Gammatone wavelets is polynomial instead of exponential. The products Q_j cannot be efficiently estimated and our method performs significantly worse. In the case of lines extracted from images (19a), the reconstruction error stagnates at 0.1%, even when the noise is of the order of 0.01%. For audio signals (19b), it is around 1% for any amount of noise.

2) Number of iterations in the optimization step: The maximal number of iterations allowed per local optimization



(a) (b)

Figure 20. for the audio signal "I'm sorry", reconstruction error as a function of the maximal number of iterations (a) with 0.01% of noise (b) with 0.6% of noise

step (paragraph III-C) can have a huge impact on the quality of the reconstruction.

Figure 20 represents, for an audio signal, the reconstruction error as a function of this number of iterations. As the objective functions are not convex, there are no guarantees on the speed of the decay when the number of iterations increases. It can be slow and even non-monotonic. Nevertheless, it clearly globally decays.

The execution time is roughly proportional to the number of iterations. It is thus important to adapt this number to the desired application, so as to reach the necessary precision level without making the algorithm exaggeratedly slow.

VII. Conclusion

We have presented an new iterative algorithm that reconstructs a signal from its scalogram. This algorithm is based on a new reformulation of the reconstruction problem, using the analytic extension of the wavelet transform. It is precise and stable to noise, and has a sufficiently low complexity to be applied to audio signals.

In future works, we plan to investigate further ways to speed up the reconstruction, including parallelization, and to test our algorithm on concrete applications, like source separation.

Appendix A
Proof of lemma IV.1

Lemma. *(IV.1) Let $m \in \mathbb{R}^N$ and $K \in \mathbb{N}^*$ be fixed. We consider the problem:*

$$\begin{aligned} &\text{Find } g \in \mathbb{C}^N \text{ s.t. } |g| = m \\ &\qquad \text{and } Supp(\hat{g}) \subset \{1, ..., K\} \end{aligned}$$

This problem has at most 2^{K-1} solutions, up to a global phase, and there exist a simple algorithm which, from m and N, returns the list of all possible solutions.

Proof. We define:

$$P(g)(X) = \hat{g}[1] + \hat{g}[2]X + ... + \hat{g}[K]X^{K-1}$$

We show that the constraint $|g| = m$ amounts to knowing $P(g)(X)\overline{P(g)}(1/X)$. This is in turn equivalent to knowing the roots of $P(g)$ (and thus knowing g) up to inversion with respect to the unit circle. There are in general $K-1$ roots, and each one can be inverted. This gives 2^{K-1} solutions.

We set:

$$\begin{aligned} Q(g)(X) &= \overline{P(g)}(1/X) \\ &= \overline{\hat{g}[K]}X^{-(K-1)} + \overline{\hat{g}[K-1]}X^{-(k-2)} + ... + \overline{\hat{g}[1]} \end{aligned}$$

The equation $|g|^2 = m^2$ is equivalent to $\widehat{|g|^2} = \widehat{m^2}$, that is $\frac{1}{N}\hat{g} \star \hat{\bar{g}} = \widehat{m^2}$. For each $k \in \{-(K-1), ..., K-1\}$:

$$\hat{g} \star \hat{\bar{g}}[k] = \sum_s \hat{g}[k-s]\overline{\hat{g}[-s]}$$

This number is the coefficient of order k of $P(g)(X)Q(g)(X)$, so $|g| = m$ if and only if:

$$P(g)(X)Q(g)(X) = N \sum_{k=-(K-1)}^{K-1} \widehat{m^2}[k]X^k \qquad (11)$$

Let us denote by $r_1, ..., r_{K-1}$ the roots of $P(g)(X)$, so that:

$$\begin{aligned} P(g)(X) &= \hat{g}[K](X - r_1)...(X - r_{K-1}) \\ Q(g)(X) &= \overline{\hat{g}[K]}(1/X - \overline{r}_1)...(1/X - \overline{r}_{K-1}) \end{aligned}$$

From (11), the equality $|g| = m$ holds if and only if $\hat{g}[K], r_1, ..., r_{K-1}$ satisfy:

$$\begin{aligned} &|\hat{g}[K]|^2 \prod_{j=1}^{K-1} (X - r_j)(1/X - \overline{r}_j) \\ &\qquad = N \sum_{k=-(K-1)}^{K-1} \widehat{m^2}[k]X^k \qquad (12) \end{aligned}$$

If we denote by $s_1, 1/\overline{s}_1, ..., s_{K-1}, 1/\overline{s}_{K-1}$ the roots of the polynomial function $\sum_{k=-(K-1)}^{K-1} \widehat{m^2}[k]X^k$, then the only possible choices for $r_1, ..., r_{K-1}$ are, up to permutation:

$$r_1 = s_1 \text{ or } 1/\overline{s}_1 \qquad r_2 = s_2 \text{ or } 1/\overline{s}_2 \qquad \dots$$

So there are 2^{K-1} possibilities. Once the r_j have been chosen, $\hat{g}[K]$ is uniquely determined by (12), up to multiplication by a unitary complex.

From $r_1, ..., r_{K-1}, \hat{g}[K]$, $P(g)$ is uniquely determined and so is g. The algorithm is summarized in 2. □

Algorithm 2 reconstruction by exhaustive search for a small problem

Input: K,m

1: Compute the roots of $\sum_{k=-(K-1)}^{K-1} \widehat{m^2}[k]X^k$
2: Group them by pairs $(s_1, 1/\overline{s}_1), ..., (s_{K-1}, 1/\overline{s}_{K-1})$
3: List the 2^{K-1} elements $(r_1, ..., r_{K-1})$ of $\{s_1, 1/\overline{s}_1\} \times ... \times \{s_{K-1}, 1/\overline{s}_{K-1}\}$
4: **for all** the elements **do**
5: Compute the corresponding $\hat{g}[K]$ by (12)
6: Compute the coefficients of $P(g)(X) = \hat{g}[K](X - r_1)...(X - r_{K-1})$
7: Apply an IFFT to the coefficients to obtain g
8: **end for**

Output: the list of 2^{K-1} possible values for g

APPENDIX B
PROOF OF LEMMA III.2

Lemma (III.2)**.** *For any f satisfying the analycity condition* (1)*:*

$$\forall z \in \mathbb{C} \qquad P(|f \star \psi_j|^2)(rz) = P((f \star \psi_j^{low})\overline{(f \star \psi_j^{high})})(z)$$
$$\textit{and} \quad P(g_j^2)(rz) = P(Q_j)(z)$$

Proof. Recall that, by definition, for any $h \in \mathbb{C}^N$:

$$\forall z \in \mathbb{C} \qquad P(h)(z) = \sum_{k=\left\lfloor \frac{N}{2} \right\rfloor - N+1}^{\left\lfloor \frac{N}{2} \right\rfloor} \hat{h}[k] z^k$$

So for any two signals h, H, the condition $P(h)(rz) = P(H)(z), \forall z \in \mathbb{C}$ is equivalent to:

$$\forall k = \left\lfloor \frac{N}{2} \right\rfloor - N + 1, ..., \left\lfloor \frac{N}{2} \right\rfloor \qquad \hat{h}[k] r^k = \hat{H}[k] \tag{13}$$

Applied to g_j^2 and Q_j, this property yields the equality $P(g_j^2)(rz) = P(Q_j)(z), \forall z \in \mathbb{C}$: by the definition of Q_j in (2), the equation (13) is clearly satisfied.

Let us now show that:

$$P(|f \star \psi_j|^2)(rz) = P((f \star \psi_j^{low})\overline{(f \star \psi_j^{high})})(z), \forall z \in \mathbb{C}$$

It suffices to prove that (13) holds, that is:

$$\forall k \in \left\lfloor \frac{N}{2} \right\rfloor - N + 1, ..., \left\lfloor \frac{N}{2} \right\rfloor,$$
$$\widehat{|f \star \psi_j|^2}[k] r^k = \widehat{(f \star \psi_j^{low})\overline{(f \star \psi_j^{high})}}[k]$$

Indeed, because the analycity condition (1) holds, we have for all k:

$$\begin{aligned}
\widehat{|f \star \psi_j|^2}[k] &= \frac{1}{N} \left(\widehat{f \star \psi_j} \right) \star \left(\overline{\widehat{f \star \psi_j}} \right)[k] \\
&= \frac{1}{N} \sum_{l=1}^{\lfloor N/2 \rfloor} \hat{f}[l] \hat{\psi}_j[l] \overline{\hat{f}[l-k] \hat{\psi}_j[l-k]} \\
&= \frac{r^{-k}}{N} \sum_{l=1}^{\lfloor N/2 \rfloor} \hat{f}[l] \hat{\psi}_j^{low}[l] \overline{\hat{f}[l-k] \hat{\psi}_j^{high}[l-k]} \\
&= \frac{r^{-k}}{N} \left(\widehat{f \star \psi_j^{low}} \right) \star \left(\overline{\widehat{f \star \psi_j^{high}}} \right)[k] \\
&= r^{-k} \left(\widehat{(f \star \psi_j^{low})\overline{(f \star \psi_j^{high})}} \right)[k]
\end{aligned}$$

□

REFERENCES

[1] R. Balan, P. Casazza, and D. Edidin, "On signal reconstruction without noisy phase," *Applied and Computational Harmonic Analysis*, vol. 20, pp. 345–356, 2006.

[2] J.-C. Risset and D. L. Wessel, "Exploration of timbre by analysis and synthesis," in *The psychology of music*, D. Deutsch, Ed. Academic Press, 1999, pp. 113–169.

[3] D. Griffin and J. S. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, pp. 236–243, 1984.

[4] S. H. Nawab, T. F. Quatieri, and J. S. Lim, "Signal reconstruction from short-time fourier transform magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pp. 986–998, 1983.

[5] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.

[6] J. Bruna and S. Mallat, "Audio texture synthesis with scattering moments," *Preprint*, 2013, http://arxiv.org/abs/1311.0407.

[7] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[8] R. Gerchberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[9] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Advances in Neural Information Processing Systems 26*, 2013, pp. 1796–2804.

[10] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: theory and algorithms," *IEEE Transactions of Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[11] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Preprint*, 2015.

[12] J. Bouvrie and T. Ezzat, "An incremental algorithm for signal reconstruction from stft magnitude," in *International conference on spoken language processing*, 2006.

[13] K. Achan, S. T. Roweis, and B. J. Frey, "Probabilistic inference of speech signals from phaseless spectrograms," in *Advances in Neural Information Processing Systems 16*, 2004, pp. 1393–1400.

[14] Y. C. Eldar, P. Sidorenko, D. G. Mixon, S. Barel, and O. Cohen, "Sparse phase retrieval from short-time fourier measurements," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 638–642, 2015.

[15] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inverse Problems*, vol. 27, no. 1, 2011.

[16] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: exact and stable signal recovery from magnitude measurements via convex programming," *Communications in Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[17] D. L. Sun and J. O. Smith, "Estimating a signal from a magnitude spectrogram via convex optimization," in *Audio Engineering Society 133rd Convention*, 2012.

[18] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.

[19] J. Nocedal, "Updating quasi-newton matrices with limited storage," *Mathematics of Computation*, pp. 773–782, 1980.

[20] M. H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 2, pp. 140–154, 1982.

[21] J. Bruna, "Scattering representations for recognition," Ph.D. dissertation, École Polytechnique, Palaiseau, 2013.

[22] S. Mallat and I. Waldspurger, "Phase retrieval for the cauchy wavelet transform," *to appear in the Journal of Fourier Analysis and Applications*, 2014.

[23] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 199–225, 2011.