Synchrosqueezing Transforms: from low to high frequency modulations and perspectives

Sylvain Meignen\textsuperscript{a}, Thomas Oberlin\textsuperscript{b} Duong-Hung Pham\textsuperscript{c}

\textsuperscript{a} Laboratoire LJK - Bâtiment IMAG, Université Grenoble Alpes
700 Avenue Centrale, Campus de Saint Martin d'Hères
38401 Domaine Universitaire de Saint-Martin-d'Hères, France

\textsuperscript{b} IRIT, Toulouse INP - ENSEEIHT, 2, rue Charles Camichel, B.P. 7122, 31071 Toulouse Cedex 7, France

\textsuperscript{c} ICube UMR 7357 Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie
300 bd Sébastien Brant - CS 10413 67412 Illkirch Cedex, France

Abstract

The general aim of this paper is to introduce the concept of synchrosqueezing transforms (SSTs) that was developed to sharpen linear time-frequency representations (TFRs), like the short-time Fourier or the continuous wavelet transforms, in such a way that the sharpened transforms remain invertible. This property is of paramount importance when one seeks to recover the modes of a multicomponent signal (MCS), corresponding to the superimposition of AM/FM modes, a model often used in many practical situations. After having recalled the basic principles of SST and explained why, when applied to an MCS, it works well only when the modes making up the signal are slightly modulated, we focus on how to circumvent this limitation. We then give illustrations in practical situations either associated with gravitational wave signals or modes with fast oscillating frequencies and discuss how SST can be used in conjunction with a demodulation operator, extending existing results in that matter. Finally, we list a series of different perspectives showing the interest of SST for the signal processing community.

Résumé

La transformée synchrosqueezée, adaptation aux signaux fortement modulés et perspectives Dans cet article, nous présentons le principe de la transformée synchrosqueezée (SST) développée pour améliorer, en utilisant des techniques de réallocation, la qualité des représentations linéaires temps-fréquence, comme les transformées de Fourier à court terme ou en ondelettes. Les transformées réaffectées demeurent inversibles ce qui est fondamental pour l'étude des signaux multicomposantes (MCS), i.e. la superposition de modes modulés à la fois en amplitude et en fréquence, utilisés comme modèle dans de nombreuses applications. Cependant, la SST, dans sa formulation initiale, n'est pas adaptée aux signaux composés de modes fortement modulés, et des améliorations, que nous présentons ici, ont récemment été proposées pour pallier ce défaut. Pour illustrer ces nouveaux développements, nous étudierons le cas des ondes gravitationnelles et des modes à phase oscillante. Dans un second temps, nous montrerons comment utiliser conjointement la SST et la démodulation pour améliorer la reconstruction des modes d’un MCS, et terminerons en évoquant différentes perspectives actuelles autour de la SST.

Key words: Time-frequency analysis; multicomponent signals; reassignment techniques

Mots-clés : Analyse temps-fréquence; signaux multicomposantes; techniques de réallocations

Preprint submitted to Elsevier Science

July 1, 2019
1. Introduction

Since its foundation, the field of signal and image processing has always made intensive use of frequency representations, and today, Fourier analysis, or more generally speaking harmonic analysis, is the fundamental tool to analyze or process signals or images [1]. Many reasons can explain why the use of Fourier theory is so widespread inside the signal processing community. First, the notion of frequency is essential to describe oscillatory signals, as for instance sounds. Second, the Fourier transform (FT) is intrinsically related to linear time-invariant filters often used to model the response of physical systems or sensors. Last, when considering random measurements, FT allows to easily model or simulate stationary Gaussian random fields. In this paper, we will particularly focus on the first point, that is to design a frequency analysis framework but adapted to non-stationary signals.

Indeed, the main limitation of FT is that it does not locate in time the frequency information computed on a given signal, while most signals from the physical world are intrinsically non-stationary, in the sense that their frequencies vary along time. Analyzing or processing these signals efficiently thus requires the use of local transformations, called time-frequency representations (TFRs), among which the most popular is probably the short-time Fourier transform (STFT), which basically windows the signal around the time of interest before applying FT. Note that the window cannot be of arbitrary length since, according to Heisenberg Gabor uncertainty principle [1], a small temporal window (associated with a good time localization) leads to a bad frequency resolution, and vice-versa. Another popular TFR is the continuous wavelet transform (CWT) [2], which shares many properties with STFT, but is based on a different frequency resolution. While both transforms are invertible and allow to process non-stationary signals, they suffer from a strong limitation that the time-frequency (TF) resolution is constrained by the choice of window or wavelet, limiting both the adaptivity of the proposed analysis and the readability of the TFR.

Many works in the past decades have tackled this limitation, by using, for instance, quadratic TFRs, e.g. Wigner-Ville distributions [3], which are not constrained by the uncertainty principle, but exhibit strong interference hampering the representation and are not invertible. Another attempt, called the reassignment method (RM), dating back to the work by Kodera et al. [4] in the 1970s and then further developed in [5], essentially proposed a means of improving the TFR readability. Unfortunately, the reassigned representation was also no longer invertible. This in particular means that, when applied to the TFR of a multicomponent signal (MCS) made of AM/FM modes, RM does not allow for an easy retrieval of the components of the MCS. Other works, like the Empirical Mode Decomposition (EMD) [6], precisely focused on this latter aspect and consists of a simple algorithm to adaptively decompose an MCS into modulated AM/FM waves. While it proved to be interesting in many practical applications, it lacks mathematical foundations and behaves like a filter bank resulting in mode-mixing [7]. To partially overcome these drawbacks, recent works tried to mimic EMD but using a more stable framework, either based on wavelet transforms [8] or convex optimization [9,10].

Synchrosqueezing transform (SST) is yet another approach whose initial goal was to improve the readability of the TFR given by CWT [11] using a reassignment procedure, in such a way that the reassigned transform was still invertible. This has the nice consequence that the retrieval of the modes of an MCS is then possible from the reassigned transform [12], which is of great practical interest. Indeed, real-life signals are often modeled by means of MCS, found in a wide range of applications such as audio recordings, structural stability [13,14], or physiological signals [15,16]. Because, in these instances, the modes are essentially AM/FM, TFRs are well adapted to represent such non-stationary signals and SST proved to be an efficient technique to obtain improved TFRs while allowing for mode retrieval.

SST has recently been extended in many ways: it was adapted to STFT, known as STFT-based SST (FSST) in [17], for which mathematical analyses are available in [18,19,20], a bidimensional extension was proposed by means of the monogenic wavelet transform [21,22], and a multivariate extension adapted to quasi-circular modulated signals was also proposed in the context of brain electrical recordings [23]. In another direction, a multi-taper approach was proposed to increase the resolution and the robustness of
SST, as in the ConceFT method [24], and the case of non-harmonic waves investigated in [25]. In spite of all these advances, one major problem associated with SST in its original formulation is that it cannot deal with MCSs containing modes with strong frequency modulation, which are very common in many fields of practical interest, as for instance radar [26], speech processing [27], gravitational waves [28,29] or in the analysis of otoacoustic emissions [30]. In this regard, an adaption of FSST to better handle that type of signals, known as the second order synchrosqueezing transform (FSST2), was introduced in [20], and its theoretical foundations settled in [31]. Such a new transform was also used in a demodulation algorithm [32] and extended to better deal with modes with fast oscillating phase [33].

In this paper, we will briefly recall some useful definitions in Section 2, and then those of SST in the STFT and CWT settings in Section 3, before putting the emphasis on the different behaviors of these two transforms. Then, we will present an overview of recent developments on SST, including those on higher-order SST [31,33] and on the design of the demodulation algorithm in the SST context [32]. Then, we will give details on practical implementation, an aspect often left apart in the literature.

2. Background

Consider a signal $f \in L^1(\mathbb{R})$, its Fourier transform corresponds to:

$$\hat{f}(\xi) = \mathcal{F}\{f\}(\xi) = \int_{\mathbb{R}} f(t) e^{-i2\pi \xi t} dt,$$

(1)

and its short-time Fourier transform (STFT) is defined using any sliding window $g \in L^\infty(\mathbb{R})$ by:

$$V^g_f(t, \xi) = \int_{\mathbb{R}} f(\tau) g(\tau - t) e^{-i2\pi \xi (\tau - t)} d\tau = \int_{\mathbb{R}} f(t + \tau) g(\tau) e^{-i2\pi \xi \tau} d\tau.$$

(2)

If $f, \hat{f}, g$ and $\hat{g}$ are all in $L^1(\mathbb{R})$, $f$ can be reconstructed from its STFT as soon as $g$ is non-zero at 0:

$$f(t) = \frac{1}{g(0)} \int_{\mathbb{R}} V^g_f(t, \xi) d\xi,$$

(3)

where $\overline{Z}$ denotes the complex conjugate of $Z$.

In this paper, we will intensively study multicomponent signals (MCSs) defined as a superimposition of AM-FM components or modes:

$$f(t) = K \sum_{k=1} f_k(t) \quad \text{with} \quad f_k(t) = A_k(t) e^{i2\pi \phi_k(t)},$$

(4)

for some finite $K \in \mathbb{N}$, $A_k(t)$ and $\phi_k(t)$ being respectively the instantaneous amplitude (IA) and frequency (IF) $f_k$ satisfying: $A_k(t) > 0, \phi_k(t) > 0$ and $\phi_{k+1}(t) > \phi_k(t)$ for all $t$. Such a signal admits an ideal TF (ITF) representation defined as:

$$\text{TI}_f(t, \omega) = \sum_{k=1}^{K} A_k(t) \delta(\omega - \phi_k(t)),$$

(5)

where $\delta$ denotes the Dirac distribution. In practice, the IF of the modes cannot be recovered by estimating the IF of $f$ as is done in the theory of analytical signals, and to locate the modes in the TF plane is essential before computing the IF of each modes: this is one of the goal of SST whose construction is recalled hereafter.

3. The synchrosqueezing transform

SST pursues two main objectives: first to sharpen a given linear TFR, and then, when applied to an MCS, to separate and retrieve the modes. The transform is based on an estimation, from the TFR, of the IF of each mode, then used to sharpen the energy of the representation around so-called ridges, approximating the curves $(t, \phi_k(t))$. SST operating on a linear TFR, e.g. STFT or CWT, also allows for mode retrieval.
exploiting the synthesis formula (3), as will be recalled later. SST was first introduced in [11,12] in the CWT setting, and then adapted to STFT in [18,19,20]. In this paper, we mainly focus on STFT-based SST, and will therefore only mention, when necessary, the one based on CWT.

3.1. STFT-based synchrosqueezing transform

STFT based synchrosqueezing (denoted by FSST in the sequel) is based on the definition of the local instantaneous frequency estimate \( \hat{\omega}_f \) [5]:

\[
\hat{\omega}_f(t, \xi) = \frac{\partial \arg V_f^\gamma(t, \xi)}{\partial t} = \Re \left\{ \frac{1}{2\pi} \frac{\partial_t V_f^\gamma(t, \xi)}{V_f^\gamma(t, \xi)} \right\}, \text{ wherever } V_f^\gamma(t, \xi) \neq 0.
\] (6)

The FSST of \( f \) with threshold \( \gamma \) is then obtained by moving any coefficient \( V_f^\gamma(t, \xi) \) with magnitude larger than \( \gamma \) to location \( (t, \hat{\omega}_f(t, \xi)) \). This can be formally written as:

\[
T_f^\gamma(t, \omega) = \int_{|V_f^\gamma(t, \xi)| > \gamma} V_f^\gamma(t, \xi) \delta(\omega - \hat{\omega}_f(t, \xi)) d\xi.
\] (7)

Any mode \( f_k \) can then be reconstructed by summing FSST coefficients around the \( k \)th ridge, which amounts to modifying the synthesis formula (3) to select only the coefficients related to the \( k \)th mode, namely:

\[
f_k(t) \approx \frac{1}{C'} \int_{|\omega - \varphi_k(t)| < d} T_f^\gamma(t, \omega) d\omega.
\] (8)

3.2. Wavelet-based synchrosqueezing transform

Let \( \psi \) be a wavelet, we define at time \( t \) and scale \( a > 0 \), the continuous wavelet transform (CWT) of \( f \) by:

\[
W_\psi^\psi f(t, a) = \frac{1}{a} \int_{\mathbb{R}} f(\tau) \overline{\psi(\tau - t/a)} d\tau.
\] (9)

If \( f \) and \( \hat{f} \) are in \( L^1(\mathbb{R}) \) and \( f \) or \( \psi \) is analytic (meaning not containing any negative frequency), and assuming \( C'_{\psi} = \int_0^{+\infty} \overline{\psi(\xi)} d\xi < \infty \), the wavelet-based SST (WSST) local instantaneous frequency estimate is defined by:

\[
\hat{\omega}_f(t, a) = \Re \left\{ \frac{1}{i2\pi} \frac{\partial_t W_\psi^\psi f(t, a)}{W_\psi^\psi f(t, a)} \right\}.
\] (10)

WSST then consists of vertically moving the coefficients according to the map \( (t, a) \mapsto (t, \hat{\omega}_f(t, a)) \):

\[
S_f^\gamma(t, \omega) = \int_{|W_\psi^\gamma f(t, a)| > \gamma} W_\psi^\gamma f(t, a) \delta(\omega - \hat{\omega}_f(t, a)) \frac{da}{a}.
\] (11)

The \( k \)th mode can then be approximately reconstructed by integrating \( S_f^\gamma(t, \omega) \) along the frequency axis as follows:

\[
f_k(t) \approx \frac{1}{C'} \int_{|\omega - \varphi_k(t)| < d} S_f^\gamma(t, \omega) d\omega.
\] (12)

3.3. Approximation results for FSST and WSST

As pointed out in the introduction, the success of SST is partly due to the nice theoretical results pioneered in [12], of which we here give a flavor (referring the reader to [12] and [20] for details on WSST and FSST respectively). These results consider MCSs as in equation (4), assuming that:
(i) Each \( f_k \) is a perturbation of a pure wave, i.e. its IA \( A_k \) and IF \( \phi'_k(t) \) are slow-varying. 
(ii) The different modes are separated in the TF plane, i.e. for each time \( t \), \( \phi'_k(t) \) for \( k = 1 \cdots K \) are significantly different.

Even though FSST and WSST are very similar in nature, they behave well on different types of MCSs. Indeed, while both techniques enable a perfect representation and reconstruction of purely harmonic modes \([12,20]\), this is no longer the case when the modes are modulated.

The main difference between the two transforms is related to the frequency resolution: STFT has a fixed frequency resolution whereas the resolution of CWT depends on the scale (or frequency). Thus the slow variation for \( f'_k \), i.e. \( |A'_k(t)|, |\phi''_k(t)| \ll \phi'_k(t) \) and \( \frac{\phi'_{k+1}(t) - \phi'_k(t)}{\phi'_{k+1}(t) + \phi'_k(t)} \geq \Delta \) (\( \Delta \) being the frequency bandwidth of the wavelet), respectively, whereas in the STFT setting these are constant all over the TF plane, i.e. \( |A'_k(t)|, |\phi''_k(t)| \ll 1 \) and \( \phi'_{k+1}(t) - \phi'_k(t) \geq 2\Delta \) (\( \Delta \) being the frequency bandwidth of the window \( g \)), respectively. As a first simple example to illustrate these differences, one may consider a signal made of two parallel linear chirps (i.e. with linear IFs), for which the separation assumption is more stringent at high than low frequencies. As just explained, the applicability of FSST is restricted to a class of MCSs composed of slightly perturbed pure harmonic modes \([34,31]\). More precisely, one first defines a complex reassignment operators \( \tilde{\omega}_f(t, \xi) \) and \( \tilde{l}_f(t, \xi) = t - \frac{\partial \tilde{\omega}_f(t, \xi)}{2\pi \nu_f(t, \xi)} \), and then defines a complex frequency modulation operator as \([34]\):

\[
\tilde{q}_f(t, \xi) = \frac{\partial \tilde{\omega}_f(t, \xi)}{\partial t \tilde{l}_f(t, \xi)} = \frac{\partial_t \left( \frac{\partial \nu_f(t, \xi)}{2\pi \nu_f(t, \xi)} \right)}{2\pi - \partial_t \left( \frac{\partial \nu_f(t, \xi)}{2\pi \nu_f(t, \xi)} \right)},
\]

The second-order local modulation operator then corresponds to \( \Re \{ \tilde{q}_f(t, \xi) \} \), and the second order complex IF estimate of \( f \) is defined by:

\[
\tilde{\omega}_f^{[2]}(t, \xi) = \begin{cases} 
\tilde{\omega}_f(t, \xi) + \tilde{q}_f(t, \xi)(t - \tilde{l}_f(t, \xi)) & \text{if } \partial_t \tilde{l}_f(t, \xi) \neq 0 \\
\tilde{\omega}_f(t, \xi) & \text{otherwise},
\end{cases}
\]

4. Higher order SST

4.1. Second order synchrosqueezing transforms

As just explained, the applicability of FSST is restricted to a class of MCSs composed of slightly perturbed pure harmonic modes. To overcome this limitation, an extension of FSST was introduced based on a more accurate IF estimate, then used to define an improved synchrosqueezing operator, called second-order STFT-based synchrosqueezing transform (FSST2) \([34,31]\). More precisely, one first defines a second-order local modulation operator, which is then used to compute the new IF estimate. To do so, one introduces complex reassignment operators \( \tilde{\omega}_f(t, \xi) = \frac{\partial \nu_f(t, \xi)}{2\pi \nu_f(t, \xi)} \) and \( \tilde{l}_f(t, \xi) = t - \frac{\partial \nu_f(t, \xi)}{2\pi \nu_f(t, \xi)} \), and then defines a complex frequency modulation operator as \([34]\):

\[
\tilde{q}_f(t, \xi) = \frac{\partial \tilde{\omega}_f(t, \xi)}{\partial t \tilde{l}_f(t, \xi)} = \frac{\partial_t \left( \frac{\partial \nu_f(t, \xi)}{2\pi \nu_f(t, \xi)} \right)}{2\pi - \partial_t \left( \frac{\partial \nu_f(t, \xi)}{2\pi \nu_f(t, \xi)} \right)},
\]

The second-order local modulation operator then corresponds to \( \Re \{ \tilde{q}_f(t, \xi) \} \), and the second order complex IF estimate of \( f \) is defined by:

\[
\tilde{\omega}_f^{[2]}(t, \xi) = \begin{cases} 
\tilde{\omega}_f(t, \xi) + \tilde{q}_f(t, \xi)(t - \tilde{l}_f(t, \xi)) & \text{if } \partial_t \tilde{l}_f(t, \xi) \neq 0 \\
\tilde{\omega}_f(t, \xi) & \text{otherwise},
\end{cases}
\]
and we then put $\hat{\omega}^{[2]}_f(t, \xi) = \Re \{ \hat{\omega}^{[2]}_f(t, \xi) \}$. It was proven in [31] that $\hat{\omega}^{[2]}_f(t, \xi) = \phi'(t)$, when $f$ is a Gaussian modulated linear chirp. It is also worth mentioning here that $\hat{\omega}_f(t, \xi)$ can be computed by means of five different STFTs. Finally, FSST2 is defined by replacing $\hat{\omega}_f(t, \xi)$ by $\hat{\omega}^{[2]}_f(t, \xi)$ in (7), to obtain the so called $T^{[2]}_f$, and mode reconstruction is then performed by replacing $T^{[2]}_f$ by $\hat{T}^{[2]}_f$ in (8).

Similarly, complex reassignment operators $\hat{\omega}_f(t, a)$ and $\hat{\tau}_f(t, a)$ in the CWT context are respectively defined, for any $(t, a)$ s.t. $W^{\psi}_f(t, a) \neq 0$, by:

$$\hat{\omega}_f(t, a) = \frac{1}{i2\pi} \frac{\partial_t W^{\psi}_f(t, a)}{W^{\psi}_f(t, a)} \quad \text{and} \quad \hat{\tau}_f(t, a) = \int_{\mathbb{R}} \tau f(\tau) \frac{1}{a} W^{\psi}_f(\tau, a) d\tau = t + a \frac{W^{\psi}_t(t, a)}{W^{\psi}_f(t, a)}.$$ (15)

The second-order local complex modulation operator corresponds to $\tilde{\varphi}_f(t, a) = \frac{\partial_t \hat{\omega}_f(t, a)}{\partial_t \hat{\tau}_f(t, a)}$ and the definition of the improved complex IF estimate associated with CWT is derived as:

$$\hat{\omega}^{[2]}_f(t, a) = \begin{cases} \hat{\omega}_f(t, a) + \tilde{\varphi}_f(t, a)(t - \hat{\tau}_f(t, a)) \text{ if } \partial_t \hat{\tau}_f(t, a) \neq 0, \\ \hat{\omega}_f(t, a) \end{cases} \quad \text{if } \partial_t \hat{\tau}_f(t, a) = 0.$$ (16)

whose real part $\hat{\omega}^{[2]}_f(t, a) = \Re \{ \hat{\omega}^{[2]}_f(t, a) \}$ is the desired IF estimate. It was shown in [35] that $\Re \{ \tilde{\varphi}_f(t, a) \} = \phi''(t)$ when $f$ is a Gaussian modulated linear chirp, i.e. $f(t) = A(t)e^{i2\pi\phi(t)}$ where both $\log(A(t))$ and $\phi(t)$ are quadratic and that $\Re \{ \hat{\omega}^{[2]}_f(t, a) \}$ is an exact estimate of $\phi''(t)$ for that kind of signals. For a more general mode with Gaussian amplitude, its IF can be estimated by $\Re \{ \hat{\omega}^{[2]}_f(t, a) \}$. As for FSST2, $\hat{\omega}_f(t, a)$ and $\tilde{\varphi}_f(t, a)$ can be computed by means of only five CWTs. The second-order WSST (WSST2) is then defined by simply replacing $\hat{\omega}_f(t, a)$ by $\hat{\omega}^{[2]}_f(t, a)$ in (11):

Figure 1. Illustration of STFT, CWT, FSST and WSST on a synthetic signal.
and phase \(\phi(t)\). To introduce the technique, we restrict ourselves to the STFT context but the technique is finally retrieved by replacing \(S^*_f(t,\omega)\) with \(S^*_f(t,\omega)\) in (12).

### 4.2. Higher order synchrosqueezing transforms

Despite FSST2 definitely sharpens the TFR it is based on, it is proved to provide a truly sharp TFR only for perturbations of linear chirps with Gaussian modulated amplitudes. To handle signals containing more general types of AM-FM modes having non-negligible \(\phi^{(n)}(t)\) for \(n \geq 3\), especially those with fast oscillating phase, one defines new SST operators based on third- or higher-order approximations of both the amplitude and phase [33]. To introduce the technique, we restrict ourselves to the STFT context but the technique presented hereafter could easily be extended to the CWT setting. Let \(f(t) = A(\tau)e^{i2\pi\phi(\tau)}\) with \(A(\tau)\) (resp. \(\phi(\tau)\)) being equal to its \(L^{th}\)-order (resp. \(N^{th}\)-order) Taylor expansion for \(\tau\) close to \(t\), namely:

\[
\log(A(\tau)) = \sum_{k=0}^{L} \frac{[\log(A)](k)(t)}{k!} (\tau - t)^k \quad \text{and} \quad \phi(\tau) = \sum_{k=0}^{N} \frac{\phi^{(k)}(t)}{k!} (\tau - t)^k,
\]

where \(Z^{(k)}(t)\) denotes the \(k^{th}\) derivative of \(Z\) evaluated at \(t\). Such a mode, with \(L \leq N\), can be written as:

\[
f(\tau) = \exp\left(\sum_{k=0}^{N} \frac{1}{k!} \left[\log(A)\right]^{(k)}(t) + i2\pi\phi^{(k)}(t)\right) (\tau - t)^k,
\]

since \([\log(A)]^{(k)}(t) = 0\) if \(L + 1 \leq k \leq N\). Its corresponding STFT writes:

\[
V^j_f(t,\xi) = \int_{\mathbb{R}} f(\tau + t)g(\tau)e^{-i2\pi\xi\tau}d\tau = \int_{\mathbb{R}} \exp\left(\sum_{k=0}^{N} \frac{1}{k!} \left[\log(A)\right]^{(k)}(t) + i2\pi\phi^{(k)}(t)\right) \tau^k g(\tau)e^{-i2\pi\xi\tau}d\tau.
\]

By taking the partial derivative of \(V^j_f(t,\xi)\) with respect to \(t\) and then dividing by \(i2\pi V^j_f(t,\xi)\), the local complex reassignment operator \(\tilde{\omega}_f(t,\xi)\) defined in Section 4.1 can be written, when \(V^j_f(t,\xi) \neq 0\), as:

\[
\tilde{\omega}_f(t,\xi) = \sum_{k=1}^{N} r_k(t) \frac{V^{k-1}_f(t,\xi)}{V^j_f(t,\xi)} = \frac{1}{i2\pi} \left[\log(A)\right]'(t) + \phi'(t) + \sum_{k=2}^{N} r_k(t) \frac{V^{k-1}_f(t,\xi)}{V^j_f(t,\xi)},
\]

where \(r_k(t) = \frac{1}{i2\pi} \left(\frac{1}{(k-1)!} \left[\log(A)\right]^{(k)}(t) + \phi^{(k)}(t)\right)\). It is clear that to get an exact IF estimate for the studied signal, one needs to subtract \(\Re\left(\sum_{k=2}^{N} r_k(t) \frac{V^{k-1}_f(t,\xi)}{V^j_f(t,\xi)}\right)\) to \(\Re\{\tilde{\omega}_f(t,\xi)\}\), which requires the calculation of \(r_k(t)\) for \(k = 2,\ldots,N\). For that purpose, one derives a frequency modulation operator \(\tilde{q}^{[k,N]}_f(t,\xi)\), equal to \(r_k(t)\) for the type of modes just introduced, and obtained by differentiating different STFTs with respect to \(\xi\) (to differentiate with respect to \(\xi\) rather than \(t\) leads to much simpler expressions). \(\tilde{q}^{[k,N]}_f(t,\xi)\) for \(2 \leq k \leq N\) can be derived recursively, following [33]:

\[
\tilde{q}_f^{[N,N]}(t,\xi) = y_N(t,\xi) \quad \text{and} \quad \tilde{q}_f^{[j,N]}(t,\xi) = y_j(t,\xi) - \sum_{k=j+1}^{N} x_{k,j}(t,\xi) \tilde{q}_f^{[k,N]}(t,\xi), \quad N - 1 \geq j \geq 2,
\]

where \(y_j(t,\xi)\) and \(x_{k,j}(t,\xi)\) are defined as follows:

\[
y_1(t,\xi) = \tilde{\omega}_f(t,\xi) \quad \text{and} \quad x_{1,1}(t,\xi) = \frac{V^{k-1}_f(t,\xi)}{V^j_f(t,\xi)}, \quad 1 \leq k \leq N,
\]

\[
y_j(t,\xi) = \frac{\partial y_{j-1}(t,\xi)}{\partial \xi x_{j-1}(t,\xi)} \quad \text{and} \quad x_{k,j}(t,\xi) = \frac{\partial x_{k,j-1}(t,\xi)}{\partial \xi x_{j-1}(t,\xi)}, \quad 2 \leq j \leq N \text{ and } j \leq k \leq N.
\]
The definition of the $N$th-order IF estimate then follows [33]:

\[
\tilde{\omega}_f^{[N]}(t, \xi) = \begin{cases} 
\tilde{\omega}_f(t, \xi) + \sum_{k=2}^{N} q_f^{[k,N]}(t, \xi) (-x_{k,1}(t, \xi)) , & \text{if } V_g^f(t, \xi) \neq 0, \text{ and } \partial_{\xi} x_{j,j-1}(t, \xi) \neq 0, 2 \leq j \leq N \\
\tilde{\omega}_f(t, \xi) , & \text{otherwise.}
\end{cases}
\]

The real part $\hat{\omega}_f^{[N]}(t, \xi) = \Re\{\tilde{\omega}_f^{[N]}(t, \xi)\}$ is the desired IF estimate which is, by construction, exact for $f$ satisfying (19). As for FSST2, the $N$th-order FSST (FSST$^N$) is defined by replacing $\hat{\omega}_f(t, \xi)$ by $\hat{\omega}_f^{[N]}(t, \xi)$ in (7) to obtain $T_{N,f}^g(t, \omega)$ and the modes of the MCS can be reconstructed by replacing $T_f^g(t, \omega)$ by $T_{N,f}^g(t, \omega)$ in (8).

![Simulated signal](image1.png)

![STFT](image2.png)

![STFT zoom](image3.png)

![FSST zoom](image4.png)

![FSST2 zoom](image5.png)

![FSST3 zoom](image6.png)

![FSST4 zoom](image7.png)

Figure 2. Illustration of the difference between second, third and fourth order FSST: (a): real part of $f$ with $A$ superimposed; (b): modulus of the STFT of $f$; (c): STFT zoom of a small TF patch (delimited by a red rectangle) extracted from (b); (d) FSST zoom carried out on the STFT shown in (c); from (d) to (g), same as (d) but for FSST2, FSST3 and FSST4 respectively.

4.3. Illustration

To illustrate the benefits brought by higher order FSSTs, we consider a synthetic mono-component signal defined as: $f(t) = A(t)e^{i2\pi \phi(t)}$ with $A(t) = 1 + 3t^2 + 4(1-t)^7$ and $\phi(t) = 240t - 2 \exp(-2t) \sin(14\pi t)$ for $t \in [0, 1]$, which corresponds to a damped-sine function signal containing very strong nonlinear sinusoidal frequency modulations and high-order polynomial amplitude modulations. Such a function is sampled at a rate $N = 1024$ Hz on $[0, 1]$. The STFT of $f$ is then computed with the $L^1$–normalized Gaussian window $g(t) = \sigma^{-1}e^{-\pi \sigma^2 t^2}$, where $\sigma$ is the optimal value determined by the Rényi entropy technique introduced in [36]. To use this optimal value for the window length parameter is crucial since it allows for a good tradeoff between time and frequency resolutions and also good mode retrieval performance [33]. In Figure 2 (a) and (b), the real part of the signal $f$ superimposed with $A$ and its STFT modulus are respectively displayed. Then, in the second row of Figure 2, we show close-ups of STFT and of reassigned representations given by the first four order FSSTs. It is clear that the higher the order of FSST the sharper the TFR, especially when the IF of the mode has a non negligible curvature $\phi''(t)$.

A second illustration considers the gravitational wave signal GW151226 [37], produced by the collision of two black holes and recorded by the LIGO and Virgo Scientific Collaborations. It mainly consists of a
chirp whose IF is increasing until the merging phase. This signal has been pre-processed as specified in the Gravitational Wave Open Science Center. However such a pre-processing does not remove all the noise which makes the chirp hardly distinguishable (Figure 3 (b) (bottom)). We applied WSST2, which seems to effectively sharpen the TFR given by CWT (Figure 3 (a)). This enabled the retrieval of the main mode, displayed in Figure 3 (b) (top) along with the theoretical predicted gravitational signal. We can observe a good correspondence until $t = 0$, after what the signal is no longer an AM/FM mode. It is worth remarking here that the reassignment operation is directly performed on the noisy CWT and that no extra denoising procedure is needed, which demonstrates that SST is somewhat robust, even though little has been done so far to quantify this robustness.

5. Practical implementation

5.1. Settings parameters

Most of the theoretical works on SST studied the continuous time and frequency settings, while in practice both time and frequency are discretized and this requires special attention. In particular and as already shown in Figure 1, the frequency resolution directly impacts the sharpness of the reassigned transform. Furthermore, SSTs depend on two main parameters: the choice of window or wavelet defining the TFR, and the threshold $\gamma$ used in the definition of the SSTs. Although there is no ideal way to determine the window length, a common practice is to consider the window (or wavelet) length that minimizes the Rényi entropy of the considered TFR [38,32]. Regarding the determination of the threshold $\gamma$, a small $\gamma$ is clearly favored to avoid some data associated with the signal to be removed.

As far as mode retrieval is concerned, a critical issue is the computation of the estimation $\varphi_k$ of the IF of mode $k$, used in (8) and carried out by ridge extraction, performed by fitting smooth curves with the locations of the highest energy coefficients of the considered STFT. This is usually done by considering non convex variational problem, whose optimal solution can be approximated using stochastic sampling or

heuristic techniques [39,12,19]. This difficult problem somehow reduces the scope of the theoretical results, since the latter assumes the ridge information is known.

5.2. Influence of frequency resolution

We here discuss the influence of the frequency resolution on FSST. Indeed, assume $f$ is with finite length, typically defined on the interval $[0,T]$, discretized into $f(\frac{kT}{N})_{n=0,\ldots,N-1}$, and $g$ supported on $[-\frac{LT}{N},\frac{LT}{N}]$, with $L < N/2$, the STFT of $f$ is then computed as follows:

$$V_f^g(t,\xi) = \int_{-LT/N}^{LT/N} f(t+\tau)g(\tau)e^{-2\pi i \tau \xi}d\tau \approx \frac{T}{N} \sum_{n=-L}^{L} f(t + \frac{nT}{N})g(\frac{nT}{N})e^{-i2\pi \frac{n}{N}\xi},$$  (22)

from which we infer that, for $0 \leq p \leq N - 1$:

$$V_f^g(\frac{pT}{N},\frac{pN}{MT}) \approx \frac{T}{N} \sum_{n=-L}^{L} f(\frac{(q+n)T}{N})g(\frac{nT}{N})e^{-i2\pi \frac{p}{N}\xi},$$

for some $M \geq 2L + 1$. The last sum is computed by means of a discrete FT, and the frequency resolution equals $\frac{N}{MT}$, which has the consequence that the TFR associated with FSST is all the more compact that the frequency resolution is low.

6. Mode reconstruction for strong modulations

In this section, we are interested in assessing the quality of mode reconstruction given by formula of type (8), when $d = 0$ and for different frequency resolutions, to clearly state the importance of the latter in mode reconstruction based on FSST2 evaluated on the ridge. Furthermore, we explain how FSSTs can be used in a demodulation algorithm for the purpose of mode reconstruction, the principle of which is recalled hereafter and whose performance are then compared to FSST2.

6.1. Demodulation algorithm based on FSST for mode reconstruction

We briefly recall how FSST (or FSST2) was combined in [32] with a demodulation algorithm for the purpose of mode reconstruction. Assume that for $f(t) = \sum_{k=1}^{K} A_k(t)e^{2\pi i \phi_k(t)}$, one has computed, from FSST (or FSST2), $K$ estimates $(\varphi_k(t))_{k=1,\ldots,K}$ of the IFs of the modes. Then, associated with each of these estimates, one defines $K$ demodulation operators $e^{-i2\pi \int_{0}^{t} \varphi_k(x)dx - \varphi_0 t}$, in which $\varphi_0$ is some positive constant frequency. One then remarks that, if $\varphi_k(t)$ is a good IF estimate for mode $k$, the signal $f_{D,k}(t) = f(t)e^{-i2\pi \int_{0}^{t} \varphi_k(x)dx - \varphi_0 t}$ should contain an almost purely harmonic component at frequency $\varphi_0$, so that computing FSST for $f_{D,k}$ and then extracting the information contained in the vicinity of frequency $\varphi_0$, one can recover $f_k$ through

$$f_k(t) \approx \left(\int_{|\omega - \varphi_0(t)| < \epsilon} T_{f_{D,k}}(t,\omega)d\omega\right)e^{i2\pi \int_{0}^{t} \varphi_k(x)dx - \varphi_0 t}.$$  (23)

In [32], $(\varphi_k(t))_{k=1,\ldots,K}$ were estimated as the piecewise constant curves associated with the ridges extracted from FSST (or FSST2), and were therefore greatly dependent on the frequency resolution $M$. We propose here a much more relevant technique to compute IFs estimates than the one based on crude ridge extraction and which will prove to be only slightly dependent on the frequency resolution. Indeed, while performing FSST (or FSST2), one computes $\hat{\varphi}_f(t,\xi)$ (or $\hat{\varphi}_f^{[2]}(t,\xi)$) which, evaluated on the ridge associated with the $k^{th}$ mode, i.e. $\hat{\varphi}_f(t,\varphi_k(t))$ (or $\hat{\varphi}_f^{[2]}(t,\varphi_k(t))$), leads to a much smoother IF estimate than the one proposed in [32]. This is due to the fact that the former is not constrained by the frequency resolution. In
what follows, we denote for the sake of simplicity \( \hat{\omega}_f(t, \varphi_k(t)) \) (or \( \hat{\omega}_f^{[2]}(t, \varphi_k(t)) \)) by \( \varphi_k(t) \) depending on the studied cases.

6.2. Performance of the demodulation algorithm based on FSST, comparison with FSST2

We here consider two types of signal to illustrate the benefits of the demodulation procedure based on FSST (or FSST2) introduced in the previous section: the first tested signal is \( f(t) = f_1(t) + f_2(t) \) made of the linear chirp \( f_1(t) = 2e^{i2\pi(130t+100t^2)} \) and the quadratic chirp \( f_2(t) = 2e^{i2\pi(230t+50t^3)} \), for \( t \in [0, 1] \) sampled at \( N = 1024 \) Hz. The STFT is computed with the same \( L^1 \)-normalized Gaussian window as previously, for which the optimal \( \sigma \) leads to a filter length \( 2L + 1 = 123 \). The representations of the FSST2 of these signals when \( M = N \) and \( M = N/8 = 128 \) are displayed in Figures 4 (a) and (b), when the input SNR equals 10 dB. Then in Figures 4 (c) and (d), we display the reconstruction results associated with modes \( f_1 \) and \( f_2 \) respectively, when \( d = 0 \) (meaning we only consider the information on the ridge). We either consider the direct reconstruction based on FSST2 displayed in (a) and (b) to show that a more concentrated representation results in a better reconstruction from the information on the ridge. Furthermore, we remark that using FSST or FSST2 in the demodulation process (with \( M = N/8 \) in the definition of STFT) does not make any difference for that type of modes (these cases are denoted by ‘FSST demod’ and ‘FSST2 demod’ in the considered figures) and that the results obtained with these latter techniques are significantly better than those obtained using FSST2 only. As a second illustration, we consider the signal \( f(t) = f_1(t) + f_2(t) \) in which the modes have cosine phases, namely \( f_1(t) = 2e^{i2\pi(190t+9\cos(3\pi t))} \) and \( f_2(t) = 2e^{i2\pi(330t+16\cos(3\pi t))} \). As previously, \( t \) belongs to \([0, 1]\), \( N = 1024 \), and the window \( g \) remains the same (the optimal \( \sigma \) given by the Rényi entropy being very similar in both cases). The FSST2 of this signal computed with \( M = N \) or \( M = N/8 \) are displayed in Figures 4 (e) and (f), respectively. Then, we again investigate the quality of the reconstruction process, assuming \( d = 0 \) and when either reconstruction from FSST2 is considered or the demodulation process based on either FSST or FSST2. As in the previous case, the demodulation algorithm based on FSST2 still behaves better than FSST2 alone as reported in Figures 4 (g) and (h). Furthermore, we remark that while to use FSST or FSST2 to compute the demodulation operator does not make any difference for mode \( f_1 \) (Figure 4 (g)), for the more modulated mode \( f_2 \), taking into account the modulation in the demodulation operator by means of FSST2 leads to much better results (Figure 4 (h)). To conclude, this pleads in favor of using demodulation based on FSST2 to reconstruct the modes rather than FSST2 only.

7. Conclusion and perspectives

The synchrosqueezing transform enters now its mature age, and has proven to be useful when analyzing a wide range of real-life signals. Many variants have been proposed to extend its domain of application, but open questions still remain:

(i) It is still not clear whether SST is the perfect tool for mode decomposition. The main reason is that the latter highly depends on the ridge estimation step, which is a difficult task: extracting the ridges associated with the modes from the reassigned transform seems to be a better option than doing the same thing on STFT but some theoretical developments still need to be carried out on that matter [40]. Furthermore, SST when used for mode reconstruction cannot operate on a STFT downsampling in time, which is a strong constraint when one has to deal with long, high-rate signals. To circumvent this last limitation, further works are therefore needed, and, in this regard, the comparison of the results obtained with SST and alternative representations in the context of audio source separation is a first step [41].

(ii) SST suffers from the intrinsic limitation that it operates on a linear TFR, associated with a fixed TF resolution given by a global window or wavelets. This can be mitigated by using multi-taper methods [24], though there is still a lack of flexibility. More critically, any technique based on linear TFR
Figure 4. (a): FSST2 of the two-chirps signal when $M = N$ ($N$ being the signal length); (b): same as (a) when $M = N/8$; (c): reconstruction results for mode $f_1$ using FSST2 or the demodulation algorithm based on FSST2 or FSST (with $d = 0$, for different input SNR, and when $f(t)$ is the two-chirp signal); (d): same as (c) but for mode $f_2$; (e): FSST2 of the two modes with cosine phase when $M = N$; (f): same as (e) but when $M = N/8$; (g): reconstruction results for mode $f_1$ using FSST2 or the demodulation algorithm based on FSST2 or FSST (with $d = 0$, for different input SNR, and $f(t)$ is the signal made with two modes with cosine phase); (h): Same as (g) but for mode $f_2$.

requires TF separation for the components, which is not the case in many applications. Some recent works considered interfering or even crossing components [42], but with limited generality.

These limitations of SST make some researchers question the usefulness of SST [43], but in that paper the analysis was restricted to the original SST and the numerical investigations were carried out on simple slightly modulated modes. We definitely think that recent extensions of SST brought about two strong and clear contributions:

— The kind of results shown in [12,20], by using local slow-variation and separation assumptions, can be easily extended to other adaptive decompositions based on STFT or wavelets.

— The local instantaneous frequency and its extensions based on higher-order phase derivatives are perfectly adapted for defining a meaningful instantaneous frequency for real-life signals. Furthermore, this quantity also enables fast and accurate post-processing operations, such as ridge detection, reassignment, and mode retrieval.

References


