A&A 388, 1124–1138 (2002)
DOI: 10.1051/0004-6361:20020520

Astronomy & Astrophysics

Joint Time–Frequency Analysis: A tool for exploratory analysis and filtering of non-stationary time series

R. Vio[1,2] and W. Wamsteker[2]

[1] Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S.Liberale di Marcon, 30020 Venice, Italy
[2] ESA-VILSPA, Apartado 50727, 28080 Madrid, Spain
e-mail: willem.wamsteker@esa.int

Received 24 July 2001 / Accepted 3 April 2002

Abstract. It is the purpose of the paper to describe the virtues of time-frequency methods for signal processing applications, having astronomical time series in mind. Different methods are considered and their potential usefulness respectively drawbacks are discussed and illustrated by examples. As areas where one can hope for a successful application of joint time-frequency analysis (JTFA), we describe specifically the problem of signal denoising as well as the question of signal separation which allows to separate signals (possibly overlapping in time or frequency, but) which are living on disjoint parts of the time-frequency plane. Some recipes for practical use of the algorithms are also provided.

Key words. methods: data analysis – methods: numerical – methods: statistical

1. Introduction

The analysis of time series has always been an issue of central interest to astronomers. Indeed, the most natural way to get some insights on the characteristics of a physical system consists in studying its behavior over time. Unfortunately, maybe because until few years ago the most interesting objects known to change their luminosity were periodic/semiperiodic stars and multiple star systems, time series analysis became synonymous of research of periodicities and power-spectrum became the tool for signal analysis. Only in the eighties people began to realize that various astrophysical objects, principally extragalactic sources, can show very complex and unpredictable time evolutions (see, for example, Miller & Wiita 1991; Duschl et al. 1991). In spite of that, although inadequate, power-spectrum continued to be the preferred approach by most of researchers. Such a predilection can be explained with the substantial failure of the methods proposed as alternatives (Vio et al. 1992). The main limitations of such methods are two:

– The implicit assumption of stationarity of the processes underlying the signals. In other words, an observed signal is assumed to be characterized by statistical properties that do not change with time. However, real signals in their vast majority are nonstationary.

– The implicit assumption that it is possible to obtain the equations describing the evolution of a given system solely through the analysis of an experimental signal. In reality, if one takes into account that a given time series is only the one-dimensional output of a system which very often can be described only in terms of a set of ordinary and/or partial differential equations (in other words a time series is the one-dimensional projection of a more complex dynamics), it is easy to understand that this position is not well founded.

The consequence of these points is that if not carried out in a well defined physical context, signal analysis can be expected to be useful only for explorative purposes (e.g. feature detection) and/or for filtering tasks (e.g. removal of unwanted components as, for example, noise or spurious contributions). Furtehrmore, in order to obtain meaningful results, the nonstationarity of signals must be explicitly considered in the analysis.

2. Data exploration

2.1. Limits of the classical approach

It is well known that the power-spectrum $\mathrm{PS}(\nu)$ of a continuous and integrable (possibly complex) signal $x(t)$ is defined as

$$\mathrm{PS}(\nu) = |\mathrm{FT}(\nu)|^2, \tag{1}$$

Send offprint requests to: R. Vio,
e-mail: robertovio@tin.it

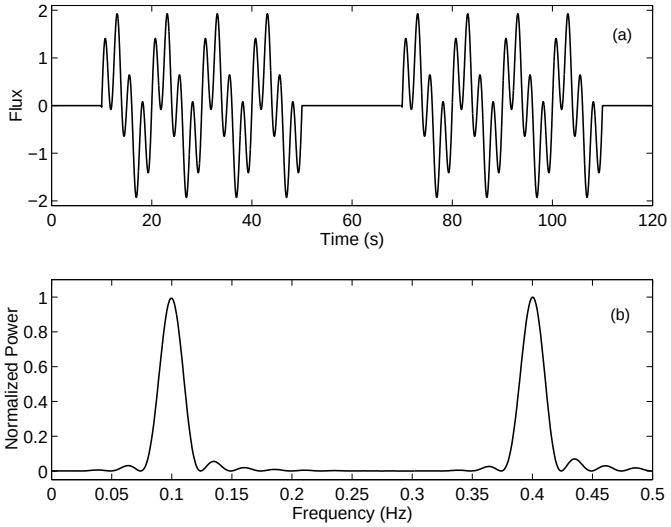

Fig. 1. a) Realization of the non-stationary process described in the text. **b)** Corresponding normalized power-spectrum.

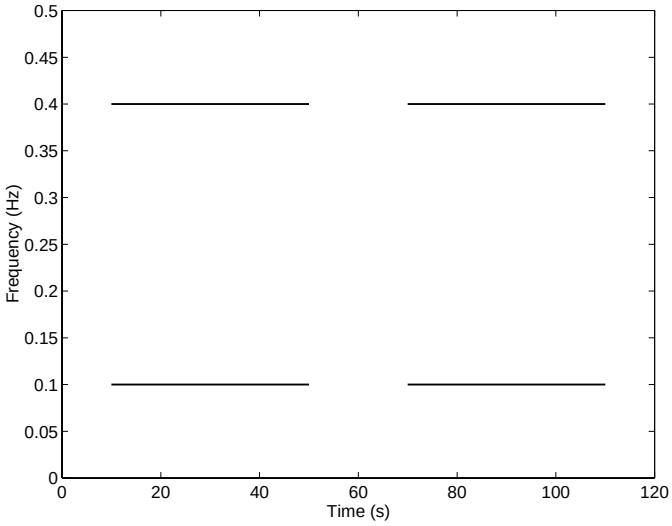

Fig. 2. Grayscale image of the ideal power-spectrum $\mathrm{IPS}(t,\nu)$ corresponding to the signal shown in Fig. 1a.

where the frequency range is the full real line, i.e. $-\infty < \nu < +\infty$, and

$$\mathrm{FT}(\nu) = \int_{-\infty}^{+\infty} x(t)\ \mathrm{e}^{-j2\pi\nu t}\ \mathrm{d}t \tag{2}$$

is the Fourier transform of $x(t)$. $\mathrm{PS}(\nu)$ can be considered as a measure of the similarity between signal $x(t)$ and the complex sinusoid $\exp(-j2\pi\nu t)$. Indeed, if $\mathrm{FT}(\nu)$ is again integrable, the Fourier inversion formula reads

$$x(t) = \int_{-\infty}^{+\infty} \mathrm{FT}(\nu)\ \mathrm{e}^{j2\pi\nu t}\ \mathrm{d}\nu, \tag{3}$$

i.e., the signal can be written (uniquely) as a kind of superposition of pure frequencies. Therefore, the larger $\mathrm{PS}(\nu)$, the more important is the contribution of the pure frequency ν within $x(t)$. This is the reason why $\mathrm{PS}(\nu)$ is so useful in the search of periodic components. However, this is also its more serious limit in the analysis of non-periodic signals (Feichtinger & Strohmer 1998). Figure 1a shows a signal constituted by two sinusoids with amplitudes different from zero only within a short time interval:

$$x(t) = a_1(t)\sin(2\pi\nu_1 t) + a_2(t)\sin(2\pi\nu_2 t) \tag{4}$$

with $\nu_1 = 0.1$ Hz, $\nu_2 = 0.4$ Hz and

$$a_1(t), a_2(t) = \begin{cases} 1 & \text{if } 10 \le t \le 50 \text{ and } 70 \le t \le 110\ ; \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

with t expressed in seconds. The two key facts are:

- this signal clearly is not periodic and
- it does not maintain its characteristics constant over time (in poor words, it is non-stationary).

From Fig. 1b it is evident that $\mathrm{PS}(\nu)$ is not able to extract the salient features of $x(t)$. In particular, the power spectrum clearly contains no information about the time intervals on which these two frequencies occur. Moreover, since $\mathrm{PS}(\nu) \neq 0$ also at frequencies $\nu \neq \nu_1, \nu_2$, one could erroneously infer that $x(t)$ contains more than two sinusoidal components. The main problem is that, if a signal changes its characteristic over time, a representation limited exclusively to the frequency domain is absolutely inadequate. Indeed, the function $\exp(-j2\pi\nu t)$ does not have any temporal localization: it has neither start nor end. This is an useful characteristic only in case of signals with constant statistical properties (i.e. periodic or stationary signals). Unfortunately, in nature this kind of signals is not so common. Therefore, for a good analysis of most "natural" signals it is necessary to use a representation that is able to consider also the temporal dimension. A possible solution is a representation, that we call *ideal power-spectrum*, $\mathrm{IPS}(t,\nu)$, defined as the squared amplitude of the sinusoids that at a given time instant t constitute the signal of interest. That is shown in Fig. 2 where it is possible to realize that with $\mathrm{IPS}(t,\nu)$, contrary to $\mathrm{PS}(\nu)$, the characteristics of the signal in Fig. 1a are very well defined; it is clear that such a signal consists of two pure sinusoids having different frequencies but amplitudes which evolve in the same way.

2.2. Time–frequency representation of continuous signals

In the previous section we have seen that in case of non-stationary signals a joint time-frequency representation (JTFR) might be a good analysis tool. However, the JTFR in Fig. 2 has been obtained directly from the model with which the time series has been simulated. Unfortunately, in an experimental context the dynamical model underlying a given time series will usually be unknown. Consequently, the problem consists in the way such a JTFR can be estimated from a set of available data.

The simplest solution is given by the so called *Fourier Spectrogram*, $\mathrm{FS}(t,\nu)$,

$$\mathrm{FS}(t,\nu) = |\mathrm{STFT}(t,\nu)|^2, \tag{6}$$

where $-\infty < t < +\infty$, $-\infty < \nu < +\infty$, and

$$\mathrm{STFT}(t,\nu) = \int_{-\infty}^{+\infty} w^*(t-\tau)\; x(\tau)\; \mathrm{e}^{-j2\pi\nu\tau}\; \mathrm{d}\tau, \tag{7}$$

(symbol "*" stands for complex conjugation), is the so called *short time fourier transform*. Here

$$\begin{array}{ll} w(t) \neq 0 & \text{for } -T/2 \leq t \leq T/2; \\ \simeq 0 & \text{otherwise} \end{array} \tag{8}$$

is a window function usually indicated with the term of *analysis function*. The rationale is that, for signals $x(t)$ which maintain their harmonic components unchanged at least approximately within a time interval of length T and centered at t_o, $\mathrm{FS}(t_\mathrm{o},\nu)$ can be considered an estimate of $\mathrm{IPS}(t_0,\nu)$. Although effective, this methods is characterized by two drawbacks:

- since $\mathrm{FS}(t,\nu)$ is estimated via the Fourier transform of only a segment of the available signal, it suffers a degradation of the frequency resolution *inversely* proportional to the length of $w(t)$. For example, if $w(t)$ is the rectangular window

  $$w(t) = \begin{cases} 1 & \text{for } -T/2 \leq t \leq T/2; \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

  it is easy to show that

  $$\mathrm{STFT}(t_\mathrm{o},\nu) = \mathrm{FT}(\nu) \otimes \mathrm{sinc}_T(t_\mathrm{o},\nu), \tag{10}$$

  where the symbol "$\otimes$" stands for convolution, and

  $$\mathrm{sinc}_T(t_\mathrm{o},\nu) = \frac{\sin(T\pi\nu)}{\pi\nu}\; \mathrm{e}^{-j2\pi\nu t_\mathrm{o}}. \tag{11}$$

  This means that, in $\mathrm{FS}(t_\mathrm{o},\nu)$, the narrowest structure along the frequency direction is charaterized by a frequency support equal to $2/T$. Such value corresponds to the *full width at zero intensity* of the main lobe in the graphical representation of $|\mathrm{sinc}_T(t_\mathrm{o},\nu)|^2$;
- since $\mathrm{FS}(t,\nu)$ is calculated by using all the points in a interval of length T centered at t, it suffers also a degradation of the resolution along the time dimension which is *proportional* to the length of $w(t)$. Indeed, in case of a signal constituted by an impulse at time t_0, and again with a rectangular $w(t)$, $\mathrm{FS}(t,\nu) \neq 0$ when $t_o - T/2 \leq t \leq t_o + T/2$. In other words, the narrowest structure along the time direction is charaterized by a time support equal to T.

From these two points it is easy to understand that the main limitation of $\mathrm{FS}(t,\nu)$ lies in the impossibility to increase simultaneously the time and the frequency resolutions. In other words, there is a trade-off: if $w(t)$ is chosen to have good time resolution (i.e., smaller T), then its frequency resolution must be deteriorated and vice versa. In principle, a possible way to mitigate this occurrence is the choice of a window $w(t)$ for which the degradation of the time-frequency resolution is reduce to a minimum. In this regard, it has been proved (e.g., Qian & Chen 1996) that, if the time-frequency resolution of a representation is measured by its energy concentration in the time-frequency domain, then the *Gaussian* function

$$w(t) = \left(\frac{\alpha}{\pi}\right)^{\frac{1}{4}} \mathrm{e}^{-\frac{\alpha}{2}t^2} \tag{12}$$

represents an "optimal" choice. Without any a priori information on the spectral characteristics of the signal, α is a free parameter with which it is possible to balance between an increased time or frequency resolution according to the current necessities. In reality, if $w(t)$ is a smooth function, well localized around zero, it will serve well as a window for the above purpose, independently from the precise analytic form. In this regard note that, because of its lack of smoothness, the rectangular window, used above only for ease of discussion, does not represent a good choice because of its very poor frequency resolution.

2.3. Time-frequency representation of discrete signals

In practical situations a signal is available only in discrete form, say $x[k]$, obtained by sampling $x(t)$ on a regular grid of L_s time instants $k = 0, 1, 2, \ldots, L_\mathrm{s}-1$. Consequently, $\mathrm{FS}(t,\nu)$ and $\mathrm{STFT}(t,\nu)$ have to be modified to their discrete variants:

$$\mathrm{DFS}[m,n] = |\mathrm{DSTFT}[m,n]|^2; \tag{13}$$

$$\mathrm{DSTFT}[m,n] = \sum_{k=0}^{L_s-1} w^*[k-m]\; x[k]\; \mathrm{e}^{-j2\pi nk/L_\mathrm{s}}, \tag{14}$$

with m, n, $0 \leq m,n \leq L_\mathrm{s}-1$, representing the discrete time and frequency indices, respectively.

Figure 3 shows the representation $\mathrm{DFS}[m,n]$ of a time series obtained by sampling the signal in Fig. 1 with a time step of one second: in spite the loss of resolution described above, the salient characteristics of the signal remain well defined.

2.4. Some practicalities

2.4.1. Digital implementation

The calculation of $\mathrm{DFS}[m,n]$ requires some technical adjustments. The first one regards the edge effects, due to the finite length of the signal, that do not permit the calculation of this representation for $m < L_w/2$ and $m > L_\mathrm{s} - 1 - L_w/2$, with L_w the length of $w[k]$. Usually this problem is solved through one of the following three methods:

1) by assuming $x[k]$ and $w[k]$ to be periodic with period L_s (NB. since usually $L_w < L_\mathrm{s}$, $w[k]$ has to be zero padded);

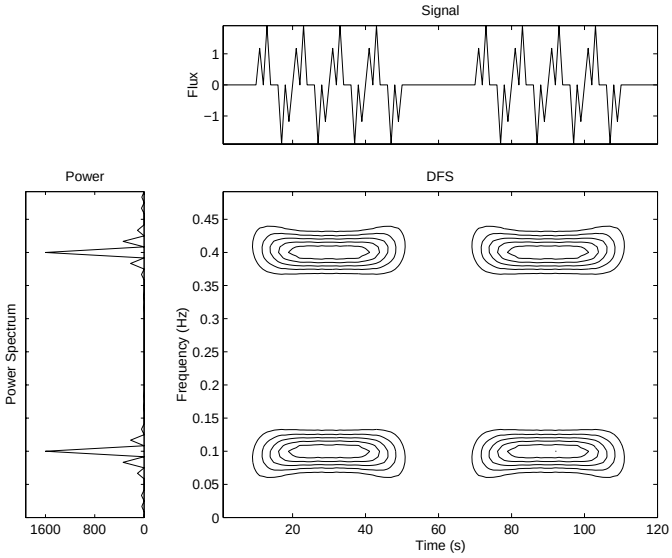

Fig. 3. Contour plot of the DFS$[m, n]$ representation for the discrete time series, shown in the uppermost panel, corresponding to the signal in Fig. 1a when sampled at a frequency rate of 1.0 Hz (the corresponding power-spectrum is shown in the leftmost panel). In this example the analysis function is an *Hamming* window with length $L_w = 31$. Levels are uniformed distributed between the 15% and 85% of the peak value.

2) by zero padding one of the ends of $x[k]$ with L_w data and assuming the new time series to be periodic with period $L_\mathrm{s} + L_w$ (for an efficient implementation of this method see Qian & Chen 1996);
3) by imposing that $0 \leq m \leq M_\mathrm{T}$, with $M_\mathrm{T} = L_\mathrm{s} - L_w + 1$ the number of possible shifts of $w[k]$ inside the signal length (for an efficient implementation of this method see Munk 1997, 2001);

If not stated otherwise, in the following the discussion will be developed according to the first approach.

A more serious drawback is represented by the fact that, by increasing the length of $x[k]$, the corresponding DFS$[m, n]$ becomes quickly an huge matrix. Indeed, already $L_\mathrm{s} = 1000$ will result in a matrix with $\simeq 10^6$ elements. A viable solution is to cut the signal in shorter segments and then to analyze them separately. An alternative comes out from the observation that DFS$[m, n]$ is an highly redundant representation; it is composed by $\simeq L_\mathrm{s}^2$ elements that, however, have been obtained from only L_s data. Since L_s is the dimension of the signal space, at most L_s elements can be linear independent. This means that, in principle, it is possible to recover arbitrary signals if samples of DFS$[m, n]$ for at least L_s coordinates are available. According to this observation, Eqs. (13), (14) could be modified in

$$\mathrm{SDFS}[m, n] = |a_{mn}|^2; \tag{15}$$

$$\begin{aligned} a_{mn} &= \mathrm{DSTFT}[m\Delta M, n\Delta N] \\ &= \sum_{k=0}^{L_s-1} w^*[k - m\Delta M]\; x[k]\; \mathrm{e}^{-j2\pi nk\Delta N/L_\mathrm{s}} \end{aligned} \tag{16}$$

where $0 \leq m < M = L_\mathrm{s}/\Delta M$, $0 \leq n < N = L_\mathrm{s}/\Delta N$, with ΔM and ΔN integer divisors of L_s such that $L_\mathrm{s}/(\Delta M \times \Delta N) \geq 1$. In such a signal representation the coefficients a_{mn} are called the *Gabor coefficients* and the Eq. (16) goes by the name *Gabor representation* (GR).

Given the above arguments, in order to avoid "unnecessary" redundancy, one could be tempted to use a SDFS$[m, n]$ containing only L_s elements. In reality, numerical stability of the expansion and easiness of the analysis (Daubechies 1992, and see below), make it necessary to work with redundant representations. The degree of redundancy is expressed through the so called *oversampling rate*, say α,

$$\alpha = \frac{\text{number of elements of SDFS}[m, n]}{\text{length of } x[k]}, \tag{17}$$

that will take values in the range $[1, L_\mathrm{s}]$. When $\alpha = 1$, SDFS$[m, n]$ is said to be sampled at a *critical rate*, whereas when $\alpha > 1$ it is said *oversampled*. Using a full short-time Fourier transform corresponds to the case $\alpha = L_\mathrm{s}$.

2.4.2. Which L_w for $w[k]$?

One important step in the construction of JTFR's, such as SDFS$[m, n]$, is the choice of the analysis function $w[k]$. Although the exact shape of this function is not so critical, the value of L_w can severely influence the final result. In fact, in order to be meaningful, a SDFS$[m, n]$ has to reflect both the instantaneous frequency content and the (possible) non-stationarity of the signal. However, very often these two requirements conflicts one with the other. The reason is that a large L_w is preferable when the frequency content is of interest and shows little changes, whereas the time evolution of a signal can be better tracked with a small L_w. Again, we are forced to decide within a trade-off.

Despite its importance and its long-time use in signal processing, surprisingly this subject has found little attention in current literature (however, see Stankovic & Katkovnik 1999; Stankovic 2001, and references therein). A possible explanation is that this problem is not so well defined. Indeed, the concept of "optimal" L_w is strictly linked to the problem at hand. For example, if one is interested in tracking the "long term" evolution of a signal rather than possible short "bursts", then larger L_w's represent the "best" choice. The contrary is true if the short term evolution is of interest. However, in certain applications (e.g. automatic signal detection) it should be useful to have an "objective" method able to provide a "good" L_w without the intervention of an operator. Of course, that requires the adoption of an optimality criterion. At this regards, a possible choice is that L_w provides the representation with the "best" global time-frequency resolution. In this case the following procedure is useful:

- a set, say Λ, of integer values for L_w is fixed;
- the SDFS$[m, n]$'s corresponding to the $L_w \in \Lambda$ are calculated;

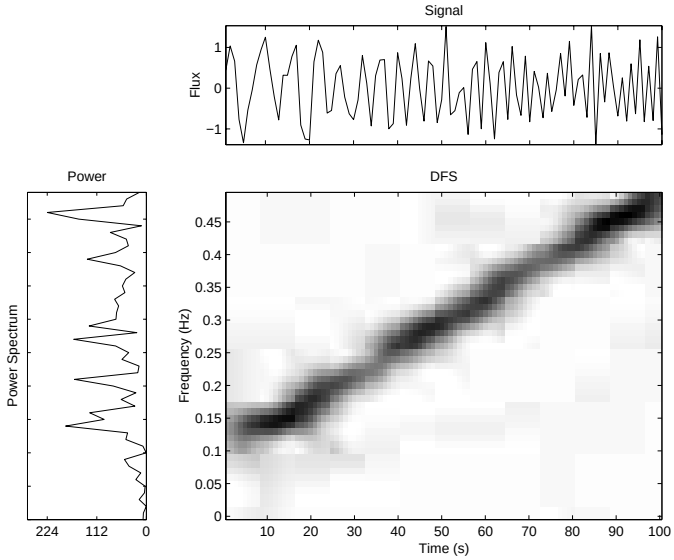

Fig. 4. Grayscale image of the DFS$[m, n]$ representation relative to the chirp signal $x(t) = \sin[2\pi t\nu(t)]$ with $\nu(t) = 4.0 \times 10^{-3} \times t + 0.10$ and $0 \leq t \leq 100$ s, when sampled with a frequency rate of 1.0 Hz and contaminated with a discrete, Gaussian white noise process with standard deviation equal to 0.3.

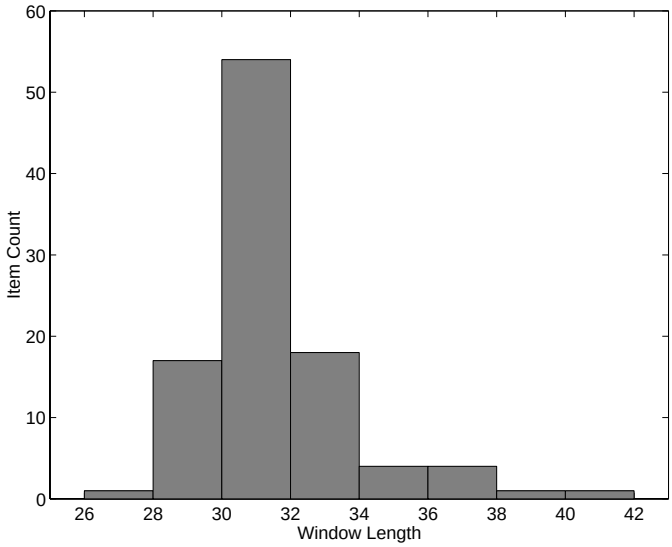

Fig. 5. Histogram concerning the lengths L_w provided in 100 simulations by the autocorrelation method described in the text. The signal used in this simulation it is shown in the uppermost panel of Fig. 4 and the target value is $L_w = 31$.

– for each of these SDFS$[m, n]$ the two-dimensional autocorrelation function ACF$[k, l]$ is computed:

$$\mathrm{ACF}[k,l] = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \mathrm{SDFS}[m,n]\ \mathrm{SDFS}[m+k,n+l] \qquad (18)$$

with $-M < k < M$ and $-N < l < N$.
– the value of L_w to which corresponds the "sharpest" ACF$[m, n]$ is selected as the searched length.

The idea behind such a procedure is that the concept of "best" global resolution implies a representation maximally concentrated in the time-frequency plane. Indeed, a value of L_w too large or too small will results in a bad resolution along the time or frequency dimension and therefore in a representation that tends to spread in the time-frequency domain. The use of ACF$[k, l]$ stems by the observation that the larger the spread of the representation, the more similar (correlated) the values are of the adjacent elements in SDFS$[m, n]$. This will reflect in ACF$[k, l]$'s with significant values also for lags k, l different from zero. The only open question in this procedure is the measure of the "sharpness" of ACF$[k, l]$. At this regards there are various possibilities. In our numerical experiments we have seen that the simple count of the elements of ACF$[k, l]$ with values larger than a given threshold, say 50% of the maximum, is able to provide already satisfactory results. This method for the estimation of L_w works very well even in the situation of very noisy data (see Figs. 4 and 5). The only concern regards the calculation of ACF$[k, l]$ that, if carried out on the basis of its definition (18), can be computationally very expensive. Indeed, for saving computing time, this operation has to be carried out in the Fourier domain via the inverse double Fourier transform of the bidimensional power-spectrum of SDFS$[m, n]$. Although very fast, with this approach it is implicitly assumed that SDFS$[m, n]$ is periodic with regards to both indices; this fact introduces edge effects. Fortunately, they can be avoided via zero padding; if interested in the correlation for lags as large as $\pm K$ along both dimensions, then both ends of SDFS$[m, n]$ must be zero padded in such a way to obtain a matrix with dimensions $(M + K) \times (N + K)$.

2.5. How to improve the resolution of the time-frequency representations?

The main limitation of SDFS$[m, n]$ is its bad resolution. In many practical situations that is not a so serious problem. However, in certain applications (e.g. tracking of quasi-periodic sources) more precise results can be required. This is a subject that has been extensively treated in current literature. Here three methods are shortly presented that are able to provide interesting results. We warn, however, that this list is by no means complete and other important approaches are available in current literature as, for instance, the reduced interference distributions (Cohen 1995).

2.5.1. Wigner-Ville distribution

The loss of resolution suffered by the FS(t, ν) is due to the fact that, for each time instant t, STFT(t, ν) is calculated by using only a segment of the signal. In principle, if such windowing were avoided, it could be possible to obtain a JTFR with a much better resolution. Indeed such a JTFR does exist, it is named Wigner-Ville Distribution,

and comes out from the problem of defining a representation, say $\mathrm{WVD}(t,\nu)$, with the property that:

$$\int_{-\infty}^{+\infty} \mathrm{WVD}(t,\nu)\ \mathrm{d}\nu\ =\ |x(t)|^2; \tag{19}$$

$$\int_{-\infty}^{+\infty} \mathrm{WVD}(t,\nu)\ \mathrm{d}t\ =\ \mathrm{PS}(\nu). \tag{20}$$

In other words, the marginals with respect to frequency and time have to provide, respectively, the squared signal and the corresponding power-spectrum. The form of this function, that however is not the only one sharing properties (19) and (20), is given by (Flandrin 1999; Groĉhenig 2001)

$$\mathrm{WVD}(t,\nu) = \int_{-\infty}^{+\infty} x(t+\tau/2)\ x^*(t-\tau/2)\ \mathrm{e}^{-j2\pi\nu\tau}\ \mathrm{d}\tau. \tag{21}$$

In reality, the use of this JTFR has been rather limited because of two drawbacks:

- $\mathrm{WVD}(t,\nu)$ can go negative. In fact, except very particular cases, e.g., $x(t)$ given by a Gaussian function, it will go negative for all signals. The non-positive property prohibits an energy interpretation of this representation;
- the quadratic nature of $\mathrm{WVD}(t,\nu)$ determines the occurrence of non-linear effects when dealing with multicomponents signals.

This last point is particularly troublesome. In fact, if $x(t) = g(t) + r(t)$, it happens that

$$\mathrm{WVD}_x(t,\nu) = \underbrace{\mathrm{WVD}_g(t,\nu) + \mathrm{WVD}_r(t,\nu)}_{\mathrm{Auto-Terms}} + 2\cdot \underbrace{Re\{\mathrm{WVD}_{g,r}(t,\nu)\}}_{\mathrm{Cross-Terms}} \tag{22}$$

where

$$\mathrm{WVD}_{g,r}(t,\nu) = \int_{-\infty}^{+\infty} g(t+\tau/2)\ r^*(t-\tau/2)\ \mathrm{e}^{-j2\pi\nu t}\ \mathrm{d}\tau. \tag{23}$$

In general, the $\mathrm{WVD}(t,\nu)$ corresponding to a signal containing N components will be characterized by N *auto-terms*, that constitutes the desired information, and $0.5\times N\times(N-1)$ real *cross-terms* which can be considered a sort of unwanted interference disturb. The worst point is that the intensity of such *cross-terms* is larger than those of the corresponding *auto-terms*. As shown in Fig. 6, that can mean unreadable results. In spite of these problems, $\mathrm{WVD}(t,\nu)$ has very interesting properties not shared by most of the other representations. For this reason, in these last years, one of the most important subject in the field of time-frequency analysis has been the research of methods able to reduce the importance of the interference terms with the minimum deterioration of such properties.

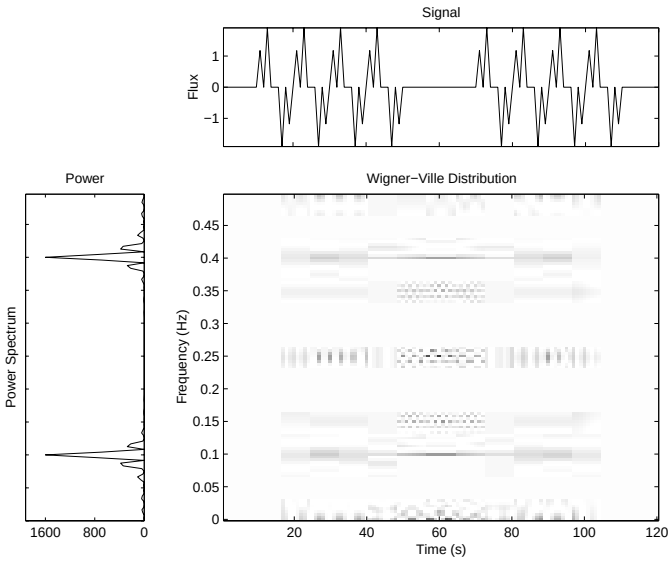

Fig. 6. Grayscale image of the $\mathrm{WVD}[m,n]$ representation for the time series shown in the uppermost panel of Fig. 3.

2.5.2. The reassignment method

It is possible to show (Flandrin 1999) that the spectrogram $\mathrm{FS}(t,\nu)$ is the 2D-convolution of the Wigner-Ville distribution of the signal $x(t)$, say $\mathrm{WVD}_x(t,\nu)$ with the Wigner-Ville distribution, say $\mathrm{WVD}_w(t,\nu)$ of the window function $w(t)$:

$$\mathrm{FS}(t,\nu) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{WVD}_x(s,\xi)\ \mathrm{WVD}_w(t-s,\nu-\xi)\ \mathrm{d}s\ \mathrm{d}\xi. \tag{24}$$

From this equation it is possible to see that the reason why $\mathrm{FS}(t,\nu)$, contrary to $\mathrm{WVD}(t,\nu)$, does not present the interference terms is due to the fact that such terms are smoothed out by the convolution. Such operation, however, deteriorates the time-frequency resolution.

A closer look at Eq. (24) shows that $W_{\mathrm{w}}(t-s,\nu-\xi)$ delimits a time-frequency domain at the vicinity of the (t,ν) point, inside which a weighted average of the WVD_x values is performed. This is the reason of the bad time-frequency resolution of $\mathrm{FS}(t,\nu)$. Indeed, the key point of the reassignment principle is that these values have no reason to be symmetrical distributed around (t,ν), which is the geometrical center of this domain; their average should not be assigned at this point, but rather at the center of gravity of this domain, which is much more representative of the local energetic distribution of the signal. In practice, with the reassignment approach one moves each value of the spectrogram computed at any point (t,ν) to

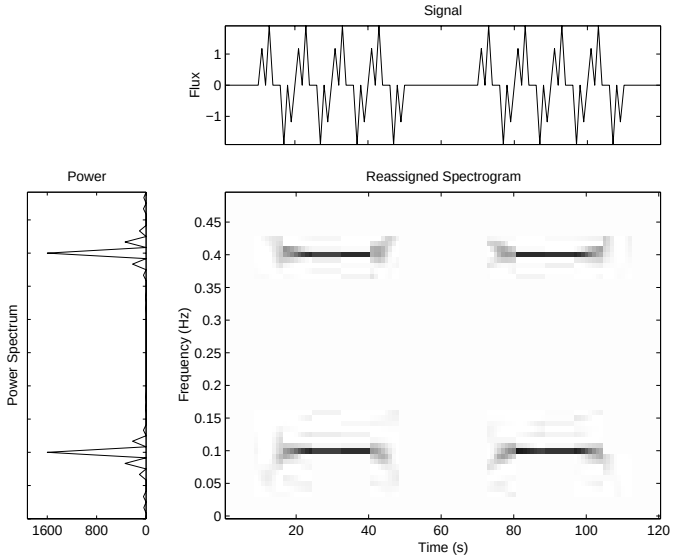

Fig. 7. Grayscale image of the $\mathrm{FS}_r[m,n]$ representation for the time series shown in the uppermost panel of Fig. 3.

another point $(\hat{t},\hat{\nu})$ which is the center of gravity of the signal energy distribution around (t,ν):

$$\hat{t}(t,\nu) = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} s\ \mathrm{WVD}_w(t-s,\nu-\xi)\ \mathrm{WVD}_x(s,\xi)\ \mathrm{d}s\mathrm{d}\xi}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{WVD}_w(t-s,\nu-\xi)\ \mathrm{WVD}_x(s,\xi)\ \mathrm{d}s\mathrm{d}\xi} \tag{25}$$

$$\hat{\nu}(t,\nu) = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \xi\ \mathrm{WVD}_w(t-s,\nu-\xi)\ \mathrm{WVD}_x(s,\xi)\ \mathrm{d}s\mathrm{d}\xi}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{WVD}_w(t-s,\nu-\xi)\ \mathrm{WVD}_x(s,\xi)\ \mathrm{d}s\mathrm{d}\xi} \tag{26}$$

and thus leads to a reassigned spectrogram, whose value at any point (t',ν') is the sum of all the spectrogram values reassigned to this point:

$$\mathrm{FS}_r(t',\nu') = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{FS}(t,\nu)\ \delta(t'-\hat{t}(t,\nu))\ \delta(\nu'-\hat{\nu}(t,\nu))\ \mathrm{d}t\ \mathrm{d}\nu. \tag{27}$$

Figure 7 show $\mathrm{FS}_r[k,l]$ for the data shown in Fig. 1. A detailed description of the algorithm and of its digital implementation can be found in Chassande-Mottin (1998).

2.5.3. Gabor spectrogram

Recently, Qian & Chen (1996) have presented an interesting technique for improving the time-frequency resolution of $\mathrm{FS}(t,\nu)$. They start from the consideration that each *cross-term* in $\mathrm{WVD}(t,\nu)$ is located midway the pair of corresponding *auto-terms* and tends to highly oscillate in both time and frequency directions. On the other hand, useful properties, such as the time marginal, frequency marginal and instantaneous frequency, are obtained by averaging $\mathrm{WVD}(t,\nu)$. For example, the mean instantaneous frequency ν_{inst}, a quantity that is commonly used for characterizing non-stationary signals, can be evaluated via:

$$\nu_{\mathrm{inst}} = \frac{\int \nu\ \mathrm{WVD}(t,\nu)\ \mathrm{d}\nu}{\int \mathrm{WVD}(t,\nu)\ \mathrm{d}\nu}. \tag{28}$$

These observations suggest that, if $\mathrm{WVD}(t,\tau)$ can be decomposed as the sum of two-dimensional (time and frequency) localized harmonic functions, then we could use the low-order harmonic terms to delineate the time-dependent spectrum with reduced interference. High-order harmonics that introduce high oscillation have relatively small averages and thereby have negligible influence on the useful properties. The signal energy and useful properties are mainly determined by a few low-order harmonic terms. On these bases, they propose the so called *Gabor spectrogram* GS_D representation

$$\mathrm{GS}_D[k,l] = \sum_{d=0}^{D} P_{\mathrm{d}}[k,l] \tag{29}$$

where

$$P_{\mathrm{d}}[k,l] = 2\sum_{A_{\mathrm{d}}} a_{mn}\ a^*_{mn} \times \exp\left\{-\frac{[k-\Delta M(m+m')/2]^2}{2\sigma_k^2} - \frac{[l-\Delta N(n+n')/2]^2}{2\sigma_l^2}\right\} \times \exp\left\{\frac{j2\pi}{L_w}\left[k(n-n')\Delta N + \left(l-\frac{n+n'}{2}\Delta N\right)\Delta M(m'-m)\right]\right\}, \tag{30}$$

where

$$A_{\mathrm{d}} = \{(m,m'),(n,n');|m-m'|+|n-n'|=d\}, \tag{31}$$

$$\sigma_{\mathrm{k}} = \frac{T_1}{2\sqrt{\pi}}, \qquad \sigma_{\mathrm{l}} = \frac{L_w}{T_1 2\sqrt{\pi}}, \tag{32}$$

with $0 \le m,m' \le M_{\mathrm{tot}}-1$, $-N/2 \le n,n' \le N/2-1$, $0 \le k < L_{\mathrm{s}}$, and $L_w/2 \le l < L_w/2-1$. Here, T_1 is a parameter whose value can be chosen $\approx\sqrt{\Delta M \cdot N}$ and a_{mn} are the Gabor coefficients calculated via Eq. (16) that here we rewrite in a slightly modified form

$$a_{mn} = \mathrm{e}^{-j2\pi mn\Delta M/N} \times \sum_{k=0}^{L_w-1} w^*[k]\ x[k+m\ \Delta M]\ \mathrm{e}^{-j2\pi kn/N}, \tag{33}$$

$0 \le m \le M_{\rm T} = [(L_{\rm s} - L_w)/\Delta M + 1]$, $0 \le n < N$. It is possible to show that the larger the value of D the better is the resolution of the representation but, at the same time, the higher is the contribution of the high-order harmonics in the decomposition of WVD(t, ν). That means that there is a trade-off between time-frequency resolution and the importance of the interference terms. Usually $D = 3$, 4 is sufficient for many practical applications. For the time series of Fig. 3, $\mathrm{GS}_4[k, l]$ provides results very similar to those of Fig. 7. A detailed description of the algorithm and of its digital implementation can be found in Munk (1997, 2001).

3. Time-variant filtering of discrete signals

Although data exploration is an interesting application of time-frequency analysis, the issue where such an approach shows all its potency is signal filtering (Matz & Hlawatsch 2002; Hlawatsch & Matz 2002). Indeed, given the capability of a JTFR in characterizing a non-stationary signal, we may expect that a filtering technique, developed within this framework, be able to adapt its action to the *instantaneous* properties of the signal itself. Since it is much easier to treat this subject in the context of the linear representations (WVD, GS_D and FS_r are bilinear representations carrying no phase information in a form easy to use), we shall concentrate in our discussion on the use of the Gabor transform.

3.1. Some preliminary notes

We have seen that the GR of a discrete signal corresponds to map a one-dimensional array onto a two dimensional matrix. Of course, in order to benefit of this kind of representation for signal filtering purposes, it is necessary to be able to do the reverse job. It can be shown that the relationship between a discrete signal $x[k]$ and the corresponding GR is given by (Groĉhenig 2001)

$$x[k] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} a_{mn}\ g[k - m\Delta M]\ \mathrm{e}^{j2\pi kn/N} \tag{34}$$

with $0 \le k < L_{\rm s}$ and

$$a_{mn} = \sum_{k=0}^{L_{\rm s}-1} w^*[k - m\Delta M]\ x[k]\ \mathrm{e}^{-j2\pi kn/N} \tag{35}$$

with $0 \le m < M$, $0 \le n < N$. Here, the *synthesis function* $g[k]$ and the *analysis function* $w[k]$ have to satisfy the so called *biorthogonality condition* (or *Wexler–Raz identity*).

$$\sum_{k=0}^{L_{\rm s}-1} w^*[k]\ g[k + pN]\ \mathrm{e}^{-j2\pi kq/\Delta M} = \frac{\Delta M}{N}\ \delta_p\ \delta_q \tag{36}$$

with $-\Delta N < q < \Delta N$, $0 \le p < \Delta M$. Pairs of functions satisfying Eq. (36) are termed *dual functions* or windows. Since this formula is symmetric with respect to $g[k]$ resp. $w[k]$, duality is a symmetric relation.

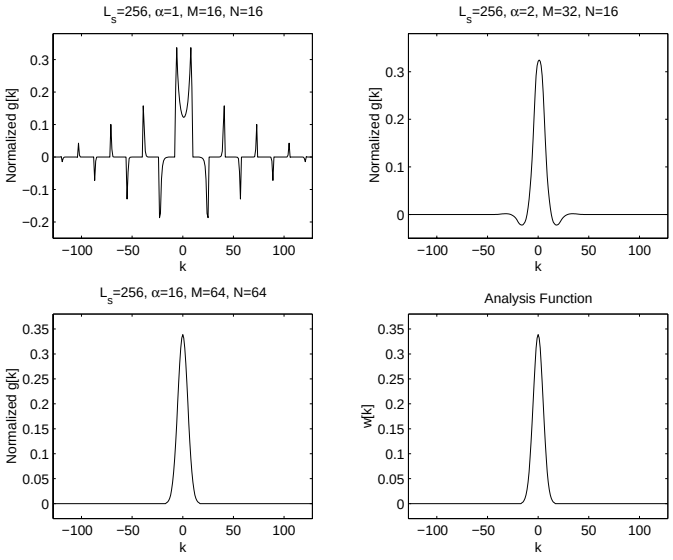

Fig. 8. Gaussian analysis window $w[k]$ and the corresponding synthesis windows $g[k]$ for different values of the oversampling rate α. For easyness of comparison, functions $g[k]$ are normalized to unit area.

For oversampled GR, with $\alpha > 2L_w/(L_w + \Delta M)$, the system of linear Eq. (36) is underdetermined. In other words, $g[k]$ and $w[k]$ do not determine each other in a unique way. Hence, for a given $g[k]$ there is a certain freedom in the choice of the corresponding dual complementary $w[k]$. The solution suggested by Qian & Chen (1996) is to solve system (36) with the constraint, say Γ, that the shapes of $g[k]$ and $w[k]$ be as close as possible (in least squares sense):

$$\Gamma = \min \sum_{k=0}^{L_s-1} \left(\frac{w[k]}{\|w\|} - \frac{g[k]}{\|g\|} \right)^2 . \tag{37}$$

This choice turns out to correspond to the solution of the system (36) via a *pseudo-inverse* method. For example, by supposing $g[k]$ fixed and rewriting system (36) in a matrix notation,

$$\boldsymbol{H}\boldsymbol{w}^* = \boldsymbol{u}, \tag{38}$$

the standard synthesis function $\boldsymbol{w}$, i.e., the window satisfying Eq. (36) with minimal norm, is given by

$$\boldsymbol{w}^* = \boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{H}^*)^{-1}\boldsymbol{u}. \tag{39}$$

It can be shown that the larger α the more similar are $g[k]$ and $w[k]$ (see Fig. 8). In particular, when $\alpha = L_{\rm s}$ (i.e., $\Delta M = \Delta N = 1$) then the two functions are identical.

In case of a critical sampled GR (i.e. $\alpha = 1$,) $w[k]$ and $g[k]$ are univocally determined. However, according to the Balian-Low theorem, both the functions cannot be simultaneously well localized in time and frequency (see, for example, Mallat 1998). In the practical application, the consequence is that $w[k]$ and $g[k]$ have very different shapes. This is the main reasons why in the filtering problems it is preferable to work with redundant representations (see below).

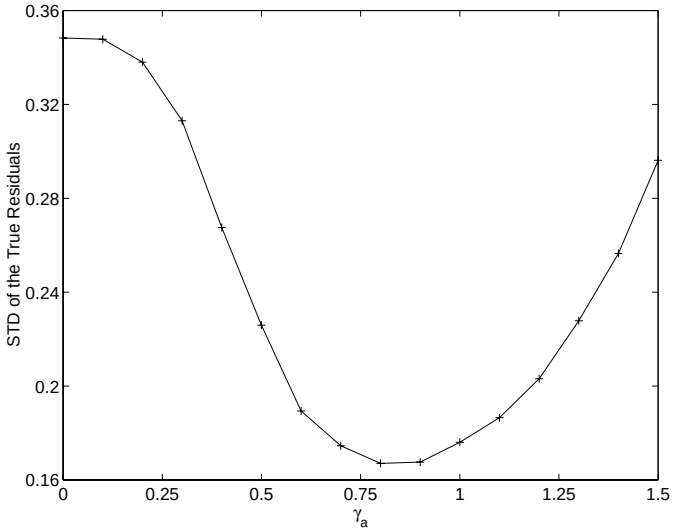

Fig. 9. Dependence of the standard deviation (STD) of the residuals between the *true* and the reconstructed signals on the value of the noise threshold γ_a. The signal used in this example is that shown in the uppermost panel of Fig. 4.

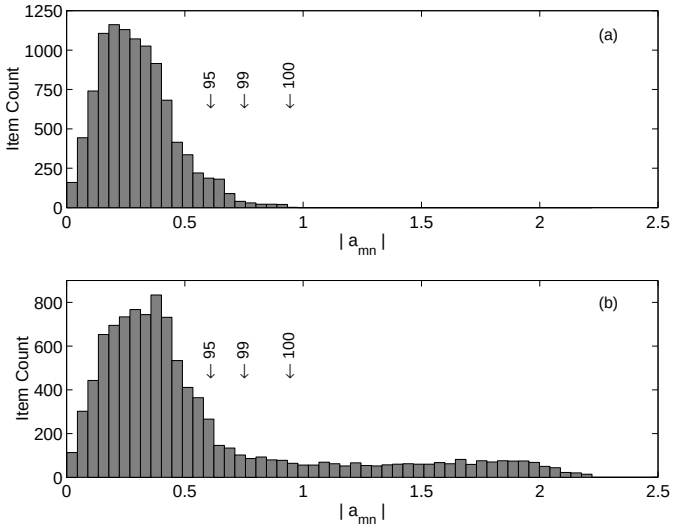

Fig. 10. a) Histogram of the $|a_{mn}|$ coefficients for a realization of a noise process with the same characteristics of that contaminating the signal of Fig. 4. **b)** Histogram of the $|a_{mn}|$ coefficients for the signal of Fig. 4. The three arrows indicate the 95%, 99%, 100% empirical percentiles relative to the histogram of the simulated noise process and provide three possible choices for γ_a.

3.2. Time–variant signal denoising

As shown in Fig. 9, the choice of γ_a is an important step in a filtering procedure. Unfortunately, without substantial a priori information (typical situation in many applications), there are no so many ways to select the "optimal" value for this threshold.

In many practical situations the signal of interest is contaminated by a broad–band random noise (white–noise). The key point is that such a noise tends to spread evenly over the entire joint time-frequency domain, whereas the signal component is usually concentrated in a relatively small region (e.g., see Fig. 4). This fact suggests a simple procedure, similar to the hard thresholding method introduced by Donoho for wavelets (see, for example, Donoho 1995), to filter out the noise contribution:

1) determination of those indices for which the corresponding coefficients a_{mn} may be considered as contributed by the noise. This step can be carried out in several ways according to the problem at hand. A typical solution consists in setting a threshold value, say γ_a^2, for the elements $|a_{mn}|^2$ of SDFS$[m, n]$. The idea is that the coefficients a_{mn} of GR, whose absolute value is smaller than γ_a, can be assumed unaffected by the signal contribution;
2) masking (zeroing) of such coefficients in order to obtain a "cleaned" version of the coefficients, denoted by $\tilde{a}_{mn}$, equal to zero in areas only occupied by noise;
3) synthesis of the filtered signal $\hat{x}[k]$ via Eq. (34).

Although this approach is fairly simple, its application requires some care in order to avoid unpleasant side effects. In particular, one point is important:

> once the analysis function $g[k]$ is fixed, it is advisable that the synthesis function $w[k]$ be calculated via Eq. (39). In other words, the shapes of $g[k]$ and $w[k]$ have to be as close as possible.

The good reason for this is that $g[k]$ is often chosen to have a good localization both in time and frequency. If $w[k]$ would not be forced to share this property it might be badly localized along one or both of these two dimensions. The consequence is that the synthesis of $\hat{x}[k]$, via the coefficient $\tilde{a}_{mn}$, could provide different results from what expected;

Here it is necessary to stress that, in general, the set of modified coefficients $\tilde{a}_{mn}$ will not constitute a valid GR representation of any signal whatever. Indeed, it happens that

$$\hat{x}[k] \neq \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \tilde{a}_{mn}\; g[k - m\Delta M]\; \mathrm{e}^{j2\pi kn/N}. \qquad (40)$$

In particular the coefficients $\hat{a}_{mn}$, obtained from $\hat{x}[k]$ via transform (35), can be different from zero even for those indices m_0, n_0, for which $\tilde{a}_{m_0 n_0} = 0$. The reason is that coefficients a_{mn} are obtained via a convolution operation and therefore no physical signal can correspond to a representation $\tilde{a}_{mn}$ characterized by the sharp cutoffs created by the masking procedure. In any case, it can be shown (see Appendix A) that, if $w[k]$ and $g[k]$ satisfy the constraint (37), $\hat{x}[k]$ is a signal whose Gabor coefficients $\hat{a}_{mn}$ are as close as possible (in least squares sense) to $\tilde{a}_{mn}$.

A last note regards the fact that, in certain situations, the zeroing of coefficients a_{mn} can constitute a too drastic operation. In fact, it can introduce some spurious structures in the reconstructed signals. In this case, a possible alternative is represented by the weighting of coefficients a_{mn}, according to the expected level of noise contamination, in a way similar to that suggested by Donoho (1995) (soft thresholding) in the context of wavelets.

3.2.1. Practical determination of an "optimal" γ_a

Apart the very common assumption of Gaussianity, the typical information available about the noise is an estimate of its standard deviation. Conversely to the case of classical Fourier analysis, the statistical characterization of a pure random Gaussian process in the time-frequency domain is much more complex. The problem is that coefficients a_{mn} are not independent. This situation suggests that the simplest (and safest) way to determine the threshold γ_a is the following:

- simulation of a noise process (not necessarily Gaussian) via a pseudo-random generator;
- calculation of the corresponding SDFS$[m, n]$ with the same modalities used for the calculation of the GR of the signal;
- estimation of a simple statistical index (e.g. empirical percentiles), describing the characteristics of such a representation, that provides the desired threshold for the SDFS$[m, n]$ of the signal.

Although very simple, in our numerical simulations this method has proved to be rather solid (see Fig. 10 to compare with Fig. 9). An alternative to the last step could be the calculation of a set of filtered time series $\hat{x}[k]$ corresponding to different choices of γ_a. After that, among these solutions it is chosen the time series that satisfies some statistical requirements as, for example, that the standard deviation of the residuals between the original and the filtered signal is equal to the nominal value of noise level. Although appealing, from our numerical experiments this approach appears to be not so stable.

In the case one lacks also the information regarding the standard deviation of the noise, the simplest solution consists in considering directly the SDFS$[m, n]$ of the signal and then in determining the threshold interactively by delimiting, via some graphical facility, the region where the signal is supposed to give its contribution. Another possibility is the use of some statistical indices of the values of SDFS$[m, n]$ (e.g. percentiles).

3.2.2. Could the SNR be further improved?

The synthesis of $\hat{x}[k]$, via the coefficients $\tilde{a}_{mn}$, permits to filter out the noise contribution corresponding to the coefficients a_{mn} that have been zeroed. However, what about the contamination of the remaining coefficients? In their book, Qian & Chen (1996) suggest a method that, according to their opinion, should further improve the SNR of the final results. Their idea is very simple and consists in an iterative application of the procedure described in the previous section. In other words, once obtained the filtered $\hat{x}[k]$, to this signal is applied the same procedure used for $x[k]$ and so on.

The most serious problem of this method consists in the lack of theoretical argumentation that guarantee its reliability. Indeed, the repeated application of the masking procedure corresponds to a *contraction*, namely to an operation for which at the n-th iteration

$$\|x_n[k]\|^2 < \|x_{n-1}[k]\|^2. \tag{41}$$

In other words, the total energy of the filtered signal is continuously reduced with the number of iterations. At first sight, one could accept this fact as useful since effectively the noise tends to increase the power of an observed time series. Unfortunately, such an increase regards the global characteristics of a signal whereas the procedure suggested by Qian & Chen is applied to the coefficient a_{mn} that reflect its *local* properties. In other words, given an *observed* Gabor coefficient, the only thing one is allowed to claim is that, because of the noise contamination, it will be different from the *true* one. For a correct filtering, however, we should be able to decide at least whether the contaminated coefficient is larger or smaller than the uncontaminated one.

The only situation where the Qian & Chen procedure can be expected to work is when γ_a is chosen too low. In situations like this one, a large fraction of the unmasked coefficient a_{mn} could be erroneously assumed linked to the signal. In general, these coefficients are smaller, in absolute value, than the coefficients effectively reflecting the signal contribution. Consequently, the contraction connected with the Qian's & Chen's technique can be able to smooth out them before the signal component be too negatively affected. However, we stress that, contrary to what claimed by the two authors, the improvement is due to the removal of Gabor coefficients still related to the noise rather that a better estimate of the coefficients representing the contribution of the signal. Better results are to be expected with a more appropriate choice of γ_a. All that is shown in the Fig. 11 to compare with Fig. 9.

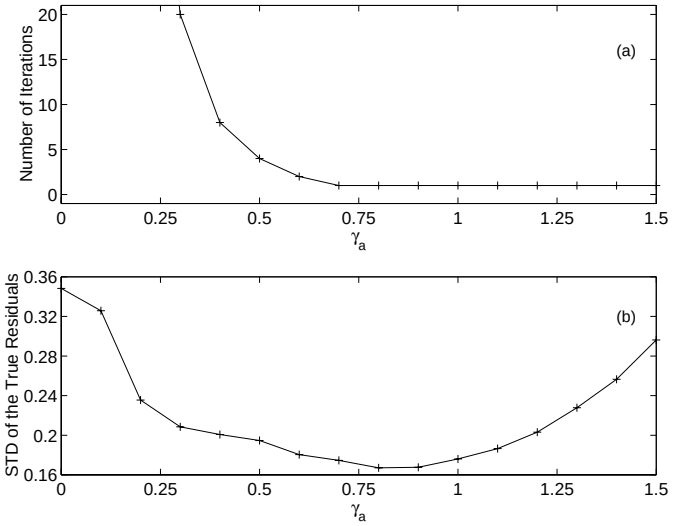

Fig. 11. a) Number of iteration for the Qian & Chen method providing the best reconstruction of the signal of Fig. 4 as function of the noise threshold γ_a. **b)** Standard deviation (STD) of the corresponding residuals between the *true* and the reconstructed signals.

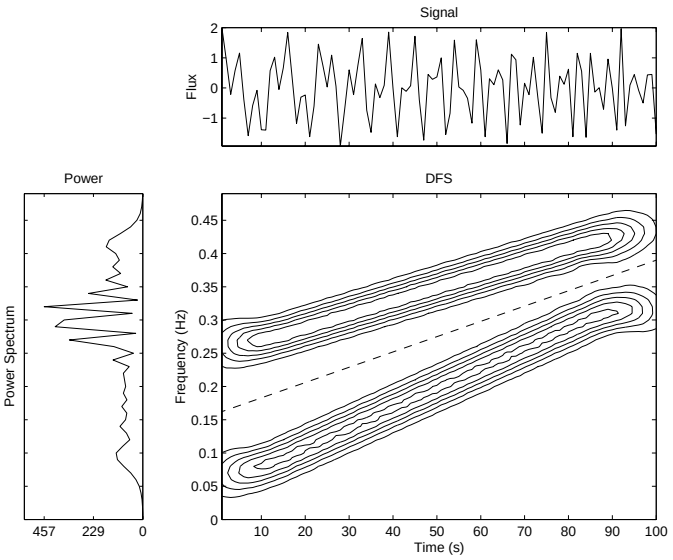

Fig. 12. Contour plot of the DFS$[m, n]$ representation for the signal mixture, given by two components disjoint in the time-frequency domain, described in the text. Levels are uniformly distributed between the 10% and 90% of the peak value. The dotted line delimits the two regions of the time-frequency domain that have been used to separate the components shown in Fig. 13.

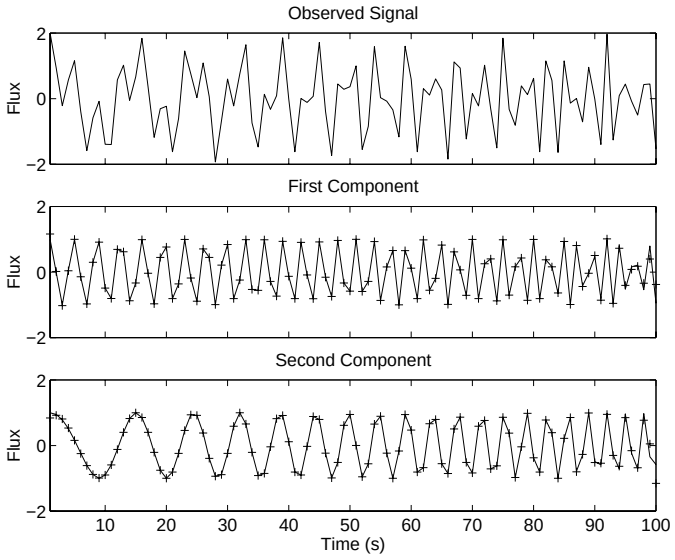

Fig. 13. Observed signal, first component, and second component corresponding to the DFS$[m, n]$ representation of Fig. 12. Continuous line indicates the original signals whereas crosses indicate the corresponding reconstructions obtained via the method described in the text.

3.3. Signal separation

In certain practical situations it can happen that a given signal is the mixture of two or more components. Often it can be of interest to separate the various contributions. In this respect, it can be shown (Hopgood 2000, and Appendix B) that, within the context of linear filtering, this operation is possible (to a given degree of accuracy) only when a representation of signal is available where the various components are disjoint. For example, in the Fourier domain, two components are "perfectly" separable if their power-Spectra do not overlap. Therefore, in the context of the joint time-frequency representations, it can be expected that separability of signal mixtures is possible if the components do not overlap in the time-frequency domain (see Figs. 12 and 13). A simple and yet effective procedure is the same as that presented in Sect. 3.2: each component is reconstructed by zeroing the Gabor coefficients corresponding to the remaining ones.

In order to understand the potential but also the limits of such an approach, it is necessary to understand what does it mean the fact that two signal are disjoint in the time-frequency domain. To discuss this problem, let suppose to have a signal $x(t)$ given by the sum of two components, say $x_1(t)$ and $x_2(t)$. If SDFS$[m, n]$ is interpreted as an energy distribution, no overlap in the time-frequency plane implies that

$$\int |x(t)|^2 \, \mathrm{d}t = \int |x_1(t)|^2 \, \mathrm{d}t + \int |x_2(t)|^2 \, \mathrm{d}t. \tag{42}$$

In other words, the energy of the observed signal is equal to the sum of the energies corresponding to the single components. This point is very important since indicates that $x_1(t)$ and $x_2(t)$ can be separated only if they do not interference and the total energy is conserved. From the statistical point of view this condition means that the two components are uncorrelated. In fact, if $x_1(t)$ and $x_2(t)$ are supposed, without loss of generality, zero mean processes, then the integrals in Eq.(42) provide the variance of the signals. However, it is well known that

$$\sigma^2_{x(t)} = \sigma^2_{x_1(t)} + \sigma^2_{x_2(t)} + 2\sigma_{x_1(t)x_2(t)} \tag{43}$$

where the last term provides the covariance between $x_1(t)$ and $x_2(t)$. Therefore, Eq. (42) implies that $\sigma_{x_1(t)x_2(t)} = 0$.

From the above discussion it is evident that, in the separation of signals overlapping in the time-frequency domain, things have to be expected much more complex. For example, Fig. 14 shows the SDFS$[m, n]$ corresponding to three signals, sampled with a frequency of one Hertz, that have been obtained through the following models:

$$x(t) = \sin[2\pi t \nu_a(t)] + \sin[2\pi t \nu_b(t) + \phi_i] \tag{44}$$

with $i = 1, 2, 3$, $\nu_a(t) = 4.0 \times 10^{-3} \times t + 0.10$, $\nu_b(t) = -3.5 \times 10^{-3} \times t + 0.45$, $0 \le t \le 100$ s, and $\phi_1 = 0$, $\phi_2 = \pi/2$, and $\phi_3 = \pi$ rad. In practice, the only difference lies in the phase ϕ of the second component. The key point to note is that, although the SDFS$[m, n]$'s are different, these signals are characterized by the same IPS(t, ν) (see Fig. 14). In other words, in this specific case the equality (42) does not hold. Of course, that is the consequence of the interference of the components present in $x(t)$. Unfortunately, the details of such a interference will depend not only on the amplitudes of the components at a given time instant, but also on the corresponding phases. Therefore, their separation should require that each coefficient a_{mn} be able to provide information about four parameters. At least of some a priori information (e.g. the value of the phases),

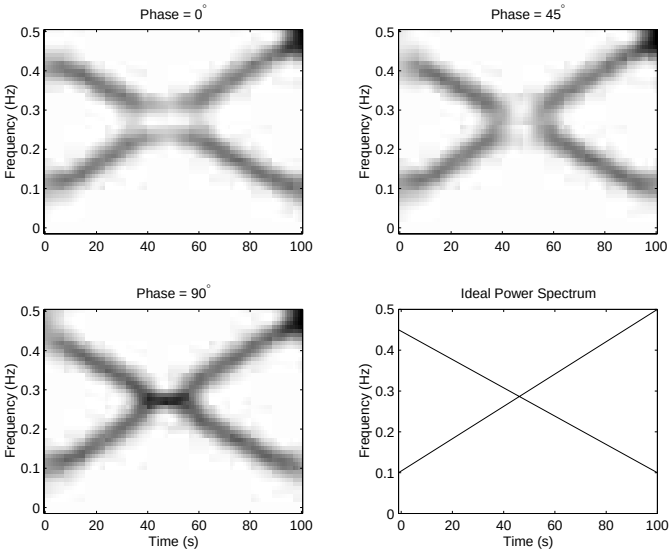

Fig. 14. Grayscale image of the DFS$[m, n]$ representation for three chirps described in the text and the corresponding ideal power-spectrum IPS(t, ν).

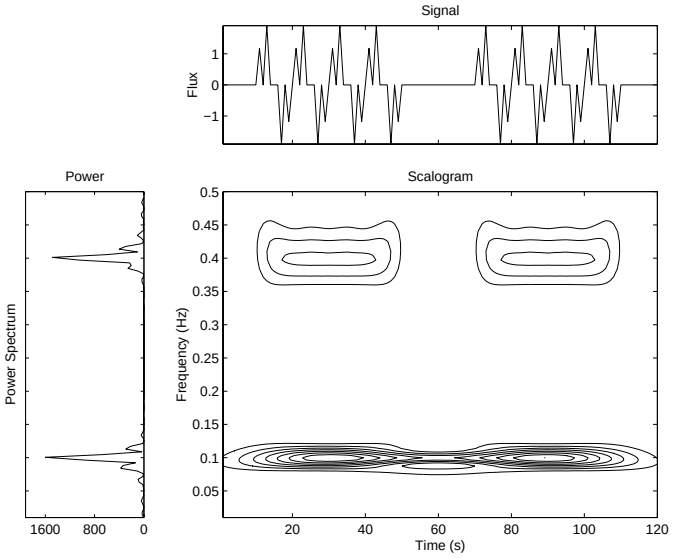

Fig. 15. Scalogram for the time series shown in the uppermost panel of Fig. 3.

this is clearly an underdetermined problem. The situation becomes worst in case the number of overlapping signals is larger than two.

One time more, these facts suggest that in order to obtain some insights on the characteristics of a complex system, in general, the analysis of a signal, if not developed within a given physical context, is absolutely insufficient.

4. JTFR better than wavelets?

Joint time-frequency analysis does not constitute the only possible tool for the analysis of non-stationary signals. Nowadays the so-called *Wavelet Analysis* arises as a new alternative to JTFA. In the continuous case the (*Scalogram*) of a signal $x(t)$ is given by:

$$\mathrm{SCL}(t, a) = |\mathrm{WT}(t, a)|^2 \tag{45}$$

with $W(t, a)$ being the continuous wavelet transform given by

$$\mathrm{WT}(t, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(s)\ \Psi^* \left(\frac{s - t}{a} \right)\ \mathrm{d}s. \tag{46}$$

Here $\Psi(t)$ is a function, named *wavelet*, with the property that $\int \Psi(t) = 0$. The idea behind such an approach is that, if $\Psi(t)$ is centered in the neighborhood of $t = 0$ and has a Fourier transform centered at a frequency $\nu_0 \neq 0$, then in the time-frequency plane $\Psi\left[(t - b)/a\right]$ will be centered at $t = b$ and $\nu = \nu_0/a$, respectively. Consequently, SCL(b, a) will reflect the signal behavior in the vicinity of such time instant and frequency. The main difference with FS(t, ν) is that the variable a corresponds to a *scale* factor, in the sense that taking $|a| > 1$ dilates Ψ and taking $|a| < 1$ compress it. Therefore, in contrast to the analysis function $w(t)$ in JTFR which maintains a fixed duration (or width) but changes its shape due to frequency modulation, in wavelet analysis when the scale factor a is varied $\Psi(t, a)$ maintains its shape but changes its duration and consequently its bandwidth. The practical consequence is that, while FS(t, ν) is characterized by a time-frequency resolution constant over the entire time-frequency domain, SCL(t, a) presents a better time resolution at high frequencies but a lower resolution at low frequencies (see Fig. 15 to compare with Fig. 3). At this point, the natural question arises concerning which choice, between the time-frequency or wavelet approach, is the most appropriate one in the context of the analysis of astrophysical signals. Of course, the answer strictly depends on the problem at hand. In particular, it is useful to have a representation characterized by a non-constant time-frequency resolution when the signal of interest presents a non-stationary evolution on very different time scales as, for example, in case of non-stationary self similar (fractal) signals. On the other hand, especially in problems of exploratory data analysis, when the time behavior of a signal is dominated by one or more non-stationary processes which evolve on similar time scales (e.g. a periodic process that changes slowly its period), then it is by far better to use a representation with a constant time-frequency resolution since it permits an easier tracking of the changing characteristics of the signal. In the astrophysical applications, very often this is the situation of interest.

It is necessary to mention that, especially in the field of *discrete* wavelets, the above indicated limitations have been overcome in the sense that some approaches have been devised that permit a large flexibility in the choice of the resolution across the time-frequency plan (e.g., wavelet packets, local cosine bases, cf. Mallat 1998). In spite of that, in the exploratory analysis of the typical astrophysical signals the joint-time frequency approach maintains its supremacy since it is more intuitive and with better

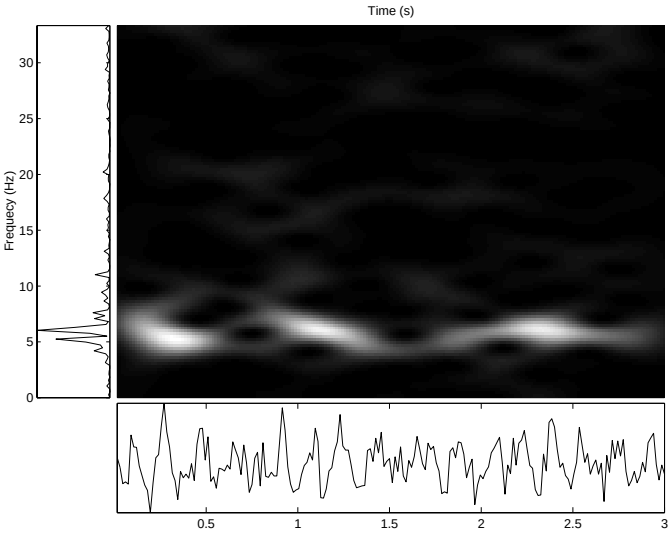

Fig. 16. Grayscale image of the DFS representation of a 3 s segment extracted from the X-ray light curve of SCO-X1 described in the text.

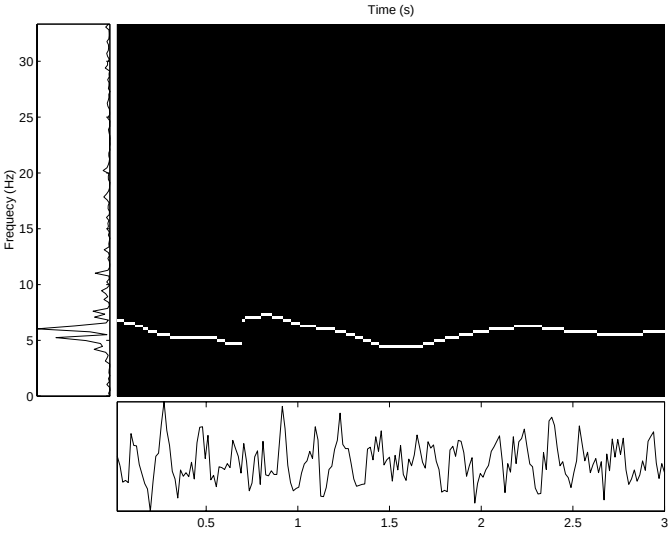

Fig. 17. Ridges of the grayscale image 16.

developed theoretical backgrounds as far as the interpretation is concerned. Indeed, while JTFA essentially makes use of *localized* sinusoids, the shape of the wavelets is much more difficult to explain from the physical point of view. Consequently, the results provided by the time-frequency approach are more amenable for a direct physical interpretation.

5. Practical applications

To illustrate the analysis strengths of JFTA in astronomical applications, we have applied some simple JFTA algorithms to a sequence of X-ray observations of low mass X-ray binary Sco X-1 (van der Klis, priv. comm.) made with the *Proportional Counter Array* (PCA) on the Rossi-XTE spacecraft (Bradt et al. 1993). The time series used, shown in Fig. 16 (lower panel), is a 3 s subset of the available sequence that is 1500 s long. Sampling is regular with a time

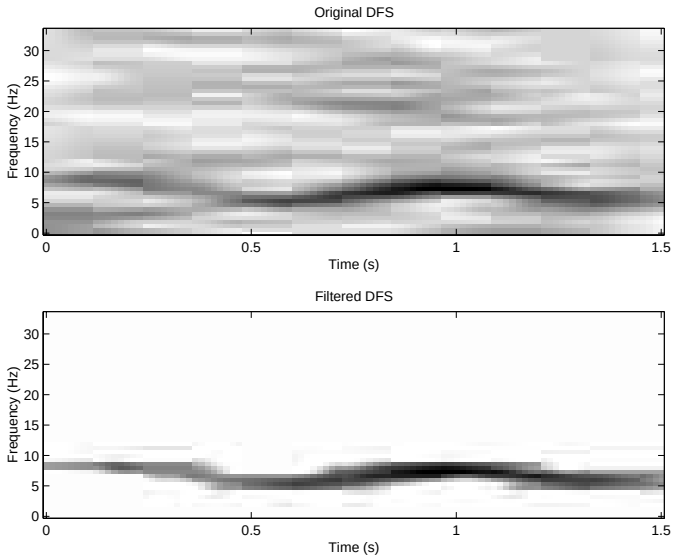

Fig. 18. Original vs. filtered DFS representation of a 1.5 s segment extracted from the X-ray light curve of SCO-X1 described in the text. Filtering has been carried out via the method described in Sect. 3.2 with γ_a corresponding to the 95% percentile of the DFS coefficients.

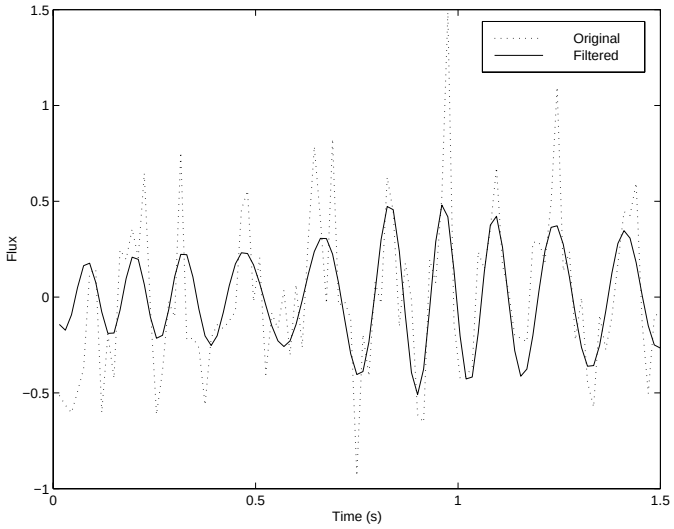

Fig. 19. QPO component vs. original time series obtained via the filtering shown in Fig. 18.

step of 125 microsecond. As Sco X-1 shows quasi-periodic oscillations (QPO) in the X-ray light emission, its light curve is characterized a periodic component that changes its period over time. Therefore, an important experimental activity consists in tracking such changes. Figs. 16, 17 show that even a very simple JTFR as DFS is able to provide very interesting results, whereas Figs. 18, 19 show how it is possible to extract the component of interest from the observed signal. Although still preliminary and partial, this example represents a good illustration of the possibilities of the application of the technique to astrophysical (and other) time series.

6. Summary and conclusions

In this paper we have considered the joint time-frequency domain as a possible framework for the analysis of the astrophysical signals. The methodologies based on such a domain appear very interesting since they allow reliable analyses and filterings of non-stationary time series that are very difficult, if not impossible, to carry out in the standard Fourier domain. These methodologies appear also useful for the separation of signal mixtures.

Acknowledgements. We thank Prof. M. van der Klis for the availability of the unpublished data on Sco X-1 and Prof. H. Feicthinger for his useful comments.

Appendix A: Masking of the Gabor Coefficients as Least Squares Filtering

Transforms (34) and (35) can be rewritten as matrix multiplications. In particular, the Gabor transform (35) described as a linear mapping from the signal space to the coefficient space $\mathcal{C}^{MN}$ takes the form

$$\boldsymbol{a} = \boldsymbol{G}_w^* \boldsymbol{x} \tag{A.1}$$

where $\boldsymbol{a} = \mathrm{vec}\{a_{mn}\}$ and $\boldsymbol{G}_g$ is an $(MN) \times L_\mathrm{s}$ matrix

$$\boldsymbol{G}_w = \left\{ \begin{array}{c} \boldsymbol{w}^* \\ T_{\Delta M}\ \boldsymbol{w}^* \\ T_{2\Delta M}\ \boldsymbol{w}^* \\ \vdots \\ T_{M\Delta M}\ \boldsymbol{w}^* \\ M_{\Delta N}\ \boldsymbol{w}^* \\ M_{\Delta N} T_{\Delta M}\ \boldsymbol{w}^* \\ M_{\Delta N} T_{2\Delta M}\ \boldsymbol{w}^* \\ \vdots \\ M_{\Delta N} T_{M\Delta M}\ \boldsymbol{w}^* \\ M_{2\Delta N}\ g \\ M_{2\Delta N} T_{\Delta M}\ \boldsymbol{w}^* \\ \vdots \\ M_{N\Delta N} T_{(M-1)\Delta M}\ \boldsymbol{w}^* \\ M_{N\Delta N} T_{M\Delta M}\ \boldsymbol{w}^* \end{array} \right\} \tag{A.2}$$

where M_q and T_p are, respectively, the *modulation* and the *translation* operators

$$\begin{array}{ll} (M_q\ w)[k] & = w[k]\ \mathrm{e}^{i2\pi qk/L_\mathrm{s}} \\ (T_p\ w)[k] & = w[k-p], \end{array} \tag{A.3}$$

and

$$(M_q T_p\ w)[k] = w[k-p]\ \mathrm{e}^{i2\pi qk/L_\mathrm{s}}. \tag{A.4}$$

Since $\boldsymbol{G}_w$ is a full rank matrix, from Eq. (A.1) it easy to show that the transform (34) can be expressed in the form

$$\boldsymbol{x} = \boldsymbol{G}_w^\dagger \boldsymbol{a}, \tag{A.5}$$

where

$$\boldsymbol{G}_w^\dagger = (\boldsymbol{G}_w^* \boldsymbol{G}_w)^{-1} \boldsymbol{G}_w^* \tag{A.6}$$

is the *pseudo-inverse* matrix of $\boldsymbol{G}_w$ with dimensions $L_\mathrm{s} \times (MN)$.

Here the key point is that, if in Eq. (A.1) we suppose $\boldsymbol{a}$ and $\boldsymbol{G}_w$ fixed and $\boldsymbol{x}$ unknown, then we are dealing with a system of $M \times N$ equations in L_s unknowns. For oversampled GR the number $M \times N$ is larger than L_s and therefore the system of Eq. (A.1) is overdetermined. In this case, it can be shown that Eq. (A.5) delivers the corresponding least squares solution (Bjorck 1990). The same still holds also when $\boldsymbol{a}$ is not in the range of $\boldsymbol{G}_w$. That happens exactly, for example, when the set of a_{mn} does not constitutes a valid GR for $x[k]$. In this case Eq. (A.1) has to be rewritten as

$$\tilde{\boldsymbol{a}} \overset{\mathrm{LS}}{=} \boldsymbol{G}_w \hat{\boldsymbol{x}} \tag{A.7}$$

in order to highlight that the equality has to be intended in least squares sense. However, $\hat{x}[k]$ will be characterized by a GR $\hat{a}_{mn} = \boldsymbol{G}_w \hat{\boldsymbol{x}}$. Consequently, Eq. (A.7) can be rewritten as

$$\tilde{\boldsymbol{a}} \overset{LS}{=} \hat{\boldsymbol{a}}. \tag{A.8}$$

The meaning of this last expression is that, although $\tilde{a}_{mn}$ is not a valid GR, $\hat{x}[k]$ synthesized via Eq. (A.5) corresponds to a signal whose GR is as close as possible to $\tilde{a}_{mn}$ in least squares sense.

Appendix B: Signal separability

In this section we briefly present the conditions according to which, given an observed signal $x(t) = \mathrm{d}(t) + n(t)$, the component of interest $\mathrm{d}(t)$ and the unwanted component $n(t)$ can be separated.

A signal $x(t)$ may be represented on an arbitrary domain λ by the following general integral transform:

$$X(\lambda) = \int_T x(t)\ K(t,\lambda)\ \mathrm{d}t \qquad \forall\ \lambda \in \Lambda \tag{B.1}$$

$$x(t) = \int_\Lambda X(\lambda)\ k(\lambda,t)\ \mathrm{d}\lambda \qquad \forall\ t \in T \tag{B.2}$$

where Λ is the region in the signal space that the representation $X(\lambda)$ lies in[1]. Here the functions $K(t,\lambda)$, called *direct transform kernel*, and $k(\lambda,t)$, called the *inverse transform kernel*, are linked by the relationships:

$$\int_\Lambda k(\lambda,t)\ K(\tau,\lambda)\ \mathrm{d}\lambda = \delta(t-\tau) \tag{B.3}$$

$$\int_T k(\lambda,t)\ K(t,\hat{\lambda})\ \mathrm{d}t = \delta(\lambda-\hat{\lambda}). \tag{B.4}$$

For example in case of the Fourier transform λ is equal to the Fourier frequency ν, $K(t,\nu) = \exp(-i2\pi\nu t)$ and $k(\nu,t) = K^*(t,\nu)$.

[1] In case of stochstic signals, integrals are to be interpreted in a mean-square limit.

If signal $\mathrm{d}(t)$ has to be reconstructed by linear filtering of the observed signal $x(t)$ (i.e., by making use of linear operators), then it has to hold (see, for example, Hopgood 2000)

$$\hat{\mathrm{d}}(t) = \int_T h(t,\tau)\ x(\tau)\ \mathrm{d}\tau, \tag{B.5}$$

where $h(t,\tau)$ is the response at time t given an impulse occurred at the filter input at time τ. A sufficient condition for the existence of $h(t,\tau)$ is that it can be written in the form

$$\begin{aligned} h(t,\tau)\ &= \int_{\Lambda_\mathrm{d}} k(\lambda,t)\ K(\tau,\lambda)\ \mathrm{d}\lambda \\ &+ \int_{\Lambda_0}\int_{\Lambda_0} H_0(\lambda,\hat{\lambda})\ k(\lambda,t)\ K(\tau,\hat{\lambda})\ \mathrm{d}\lambda\ \mathrm{d}\hat{\lambda}. \end{aligned} \tag{B.6}$$

Here Λ_d is the region of the Λ space over which the spectral component of $\mathrm{d}(t)$ does not overlap the spectral component of $n(t)$, whereas Λ_0 is the spectral region where these two components overlap. $H_0(\lambda,\hat{\lambda})$ is the solution of

$$P_{\mathrm{d}x}(\lambda,\hat{\lambda}) = \int_{\Lambda_0} H_0(\lambda,\tilde{\lambda})\ P_{xx}(\tilde{\lambda},\hat{\lambda})\ \mathrm{d}\tilde{\lambda}, \quad \forall\ \lambda,\hat{\lambda}\in\Lambda_0 \tag{B.7}$$

where

$$P_{yz}(\lambda,\hat{\lambda}) = \begin{cases} Y(\lambda)Z^*(\hat{\lambda}) & \text{if } y(t), z(t) \text{ deterministic;} \\ \mathrm{E}[Y(\lambda)Z^*(\hat{\lambda})] & \text{if } y(t), z(t) \text{ stochastic;} \end{cases} \tag{B.8}$$

is the *generalized cross power-spectrum* concerning to the processes $y(t)$ and $z(t)$. It can be shown (Hopgood 2000) that the reconstruction $\hat{\mathrm{d}}(t)$, as provided by Eqs. (B.5) and (B.6), is characterized by a variance $\sigma^2(t) = \mathrm{E}[|\hat{\mathrm{d}}(t) - \mathrm{d}(t)|^2]$ equal to

$$\sigma^2(t) = \int_{\Lambda_0}\int_{\Lambda_0} P_{\sigma\sigma}(\lambda,\hat{\lambda})\ k(\lambda,t)\ k^*(\hat{\lambda},t)\ \mathrm{d}\lambda\ \mathrm{d}\hat{\lambda} \tag{B.9}$$

with $P_{\sigma\sigma}$ the generalized power-spectrum of $\sigma^2(t)$.

Here the important points are three:

1) the first term in the rhs of Eq. (B.6) is independent of any signal. Moreover, it constitutes a so called *ideal filter* for Λ_d, namely a filter which passes without distorsion all *generalized frequencies* components falling in Λ_d and rejects all others;
2) the second term in the rhs of Eq. (B.6) depends, via Eq. (B.7), on the crosscorrelation function between $x(t)$ and $\mathrm{d}(t)$ and therefore, since $x(t) = d(t) + n(t)$, on the crosscorrelation between $\mathrm{d}(t)$ and $n(t)$;
3) the magnitude of $\sigma^2(t)$, given by Eq. (B.9), depends on the extension of the overlap between $d(t)$ and $n(t)$.

This last point is particularly interesting since it indicates that signals $d(t)$ and $n(t)$ can be perfectly separated only when they have disjoint supports in Λ. Indeed, in such a case, Λ_0 is an empty set and the three points above imply, respectively, that:

– Eq. (B.6) simplifies in

$$h(t,\tau) = \int_{\Lambda_\mathrm{d}} k(\lambda,t)\ K(\tau,\lambda)\ \mathrm{d}\lambda; \tag{B.10}$$

– $\mathrm{d}(t)$ and $n(t)$ are uncorrelated signals;
– $\sigma^2(t) = 0$.

In other words, the perfect separation of two signals, via a linear filtering, requires the following condition be satisfied:

> there exists some domain where the generalized spectral representation of the desired components are disjoint. Thus, the desired components must be uncorrelated processes.

All that holds also in the context of joint time-frequency analysis. The only different thing is that the function $h(t,\tau)$, with arguments in the time domain, has to be changed, via a Fourier transform with respect to τ, in $H(t,\nu)$ which has arguments in the mixed time-frequency domain.

References

Bradt, H. V., Rotschild, R. E., & Swank, J. H. 1993, A&AS, 97, 355

Bjorck, A. 1990, Numerical Methods for Least Squares Problems (SIAM, Philadelphia)

Chassande-Mottin, E. 1998, Ph.D. Thesis, University of Cergy-Pointoise

Cohen, L. 1995, Time-Frequency Analysis (Prentice Hall: London)

Daubechies, I. 1992, Ten Lectures on Wavelets (SIAM: Philadelphia)

Donoho, D. L. 1995, IEEE Transaction on Information Theory, 41, 613

Duschl, W. J, Wagner, S. J., & Camenzind, M. 1991, Variability of Active Galaxies (Springer-Verlag: Berlin)

Feichtinger, H. G., & Strohmer, T. 1998, Gabor Analysis and Algorithms: Theory and Applications (Birkhäuser: Boston)

Flandrin, P. 1999, Time-Frequency/Time-Scale Analysis (Academic Press, London)

Groc̆henig, K. 2001, Foundations of Time-Frequeny Analysis (Birkhäuser: Boston)

Hlawatsch, F., & Matz, G. 2002, in Time-Frequeny Signal Analysis and Processing, ed. B. Boashash (Prentice Hall: London)

Hopgood, J. R. 2000, Ph.D. Thesis, University of Cambridge

Mallat, S. 1998, a Wavelet Tour of Signal Processing (Academic Press, London)

Matz, G., & Hlawatsch, F. 2002, in Applications in Time-Frequency Signal Processing, ed. A. Papandreu-Suppappola (CRC Press: Boca Raton)

Miller, H. R., & Wiita, P. J. 1991, Variability of Active Galactic Nuclei (Cambridge University Press: Cambridge)

Munk, F. 1997, Ph.D. Thesis, Aalborg University

Munk, F. 2001, compendium for the course "Joint Time Frequency Analysis", SB8-2, Aalborg University

Qian, S., & Chen, D. 1996, Joint Time-Frequency Analysis (Prentice Hall: London)

Stankovic, L., & Katkovnik, V. 1999, IEEE Trans. of Signal Processing, 47, 1053

Stankovic, L. 2001, Signal Processing, 81, 621

Vio, R., Cristiani, S., Lessi, O., & Provenzale, A. 1992, ApJ, 391, 518