Journal of Computational Physics 228 (2009) 2906–2917

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Trim-to-Coherence Fourier Transform

M. Böhm [a], M. Tasche [b], B. Seifert [a,c], F. Mitschke [a,*]

[a] Institut für Physik, Universität Rostock, 18051 Rostock, Germany
[b] Institut für Mathematik, Universität Rostock, 18051 Rostock, Germany
[c] Faculdad de Fisica, Pontificia Universidad Catòlica de Chile, Santiago 22, Chile

ARTICLE INFO

Article history:
Received 22 May 2008
Received in revised form 3 December 2008
Accepted 27 December 2008
Available online 12 January 2009

Keywords:
Fourier transform
Numerical methods
Scalloping error

ABSTRACT

We introduce a discrete Fourier transform technique which extracts more spectral information from a given time series data set than conventional discrete Fourier transform (DFT). Valid information is obtained between the spectral bins of conventional DFT, scalloping error is greatly reduced, and amplitude and phase of Fourier components are more true to the process under study as with conventional DFT. We call the general idea Trim-to-Coherence Fourier Transform, and its particular embodiment 'Phase-Rotation Fourier Transform'. Treatment of the raw data is minimally invasive; e.g. there is no zero padding.

1. Introduction

Fourier transform of the temporal variation of physical quantities provides information on the spectral content which in many cases gives valuable insight into the mechanism creating the temporal variation. Examples are abundant and run the gamut from acoustics to civil engineering to wireless communications to physiology.

Consider a time-varying function $s(t)$ (signal), sampled at the time instants $t = 0, \Delta t, 2\Delta t, \ldots, (N-1)\Delta t$ where Δt is the constant sampling time chosen so that Nyquist's sampling condition is fulfilled. The samples are denoted by $s_n = s(n\Delta t)$ $(n = 0, \ldots, N-1)$. Each sample represents a time slot of duration Δt; a record of N successive data points $s = \{s_0, s_1, \ldots, s_{N-1}\}$ (time series) is the discrete representation of a partial history of duration $T = N\Delta t$ (time window) of the signal. This partial, discretized representation reflects the signal's properties only within certain bounds of accuracy. The question typically is to which precision the frequency, amplitude and phase of some oscillation in the signal can be retrieved from the time series. We present a variation on the established procedure which can, in some circumstances, provide improved accuracy on this count.

1.1. Continuation and edge effects

A Fourier transform on a finite time series results in a discrete spectrum. Since the phase of the resulting complex spectral values is 2π-periodic and thus carries ambiguity, the Fourier transform can therefore be considered as a Fourier transform of a periodic continuation of the finite time series.

Consider as a first example that the time series represents a single harmonic oscillation of fixed frequency, amplitude and phase. Further assume for now that an integer number of oscillation periods fits precisely into the window length; this

* Corresponding author. Tel.: +49 381 498 6820; fax: +49 381 498 6822.
E-mail addresses: michael.boehm@uni-rostock.de (M. Böhm), fedor.mitschke@uni-rostock.de (F. Mitschke).
URL: http://www.physik.uni-rostock.de/optik (M. Böhm).

0021-9991/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2008.12.032

coincidence is called *coherent sampling* or *synchronous sampling* [1]. Then the periodic continuation of the window transforms the finite-length, rectangularly truncated oscillation into a continuous, unperturbed oscillation. In the spectral domain, one particular bin will represent that frequency exactly. This is the most favorable case in which both amplitude and phase at the underlying frequency are obtained without error. Since N samples generate $N/2$ different frequency bins, the frequency is certainly located to better than the separation to the neighboring bins. Specifically, if the window has N sample points and k such points constitute one period of oscillation (k divides N in the coherent case), then the signal is found in the N/kth bin. The distance of neighboring bins is equal to k/N. Note that this may be a small number; it constitutes an upper bound for the relative spectral half-width, but any finer subdivisions of the frequency domain are fictitious at this point.

Now consider the *non-coherent case*: the period of the oscillations is such that a non-integer number of periods is represented within the length of the time series. Periodic continuation now produces an oscillation in which phase jumps occur periodically (every N data points), as sketched in Fig. 1. The result is a broadened spectral line. Also, the frequency falls between two bins in the spectral domain, so that both neighboring bins will contain nonzero amplitude values; this gives rise to the *scalloping error*. It attains its maximum when the frequency falls halfway between two bins, where the frequency is off by one half the bin separation, and the amplitude by -3.92 dB [2].

The usual strategy to reduce the frequency and amplitude inaccuracies is to subject the time series to an apodization [2], also referred to as a windowing function or tapering. Windowing eliminates the artifacts from convolution of the spectrum with that of the rectangular window (false spectral peaks) by smearing out the phase jumps from the continuation in the non-coherent case, and by way of broadening the spectral line it also reduces the amount of scalloping amplitude error.

To determine values between the bins one can use improved interpolation methods. Assuming that there is no additional information outside the measured range, one can apply the Whittaker–Shannon interpolation formula [3] for the spectral domain. This method works fine in the temporal domain, where the data set is always real, but in the frequency domain this interpolation formula is usually inapplicable. The results of the methods presented here are similar to the Whittaker–Shannon interpolation formula but are universally applicable.

2. A novel approach

To explain the core idea about the *Trim-to-Coherence Fourier Transform*, consider again the case of a simple harmonic oscillation of which only a very few periods are within the recorded time series. In our proposal, a synchronization of the signal and the measured window is used to evoke the error free coherent case. In particular, we introduce what we call *Phase-Rotation Fourier Transform* (PFT); the Appendix presents a full mathematical formulation.

Start with the available time series s with arbitrary but fixed sample length N. By DFT of length N, one obtains

$$\tilde{s}_k = \frac{1}{N}\sum_{n=0}^{N-1} s_n \exp\left(-\frac{2\pi i}{N} nk\right) \quad (k = 0, \ldots, N-1). \tag{1}$$

We now show how one can approach the coherence condition by, in effect, rescaling the frequency axis. To this end we shift the signal through a rotation of its phase. The data points s_n $(n = 0, \ldots, N-1)$ are treated as complex values even if they are originally real. Each data point is rotated in phase by multiplication with a term of absolute value one:

$$s_n \to s_n^{(\delta)} = s_n e^{-i\varphi} = s_n \cdot \exp\left(-i\delta\pi\left(\frac{2n}{N} - 1\right)\right) \tag{2}$$

with the spectral resolution constant δ. In principle, one could choose any δ to access a certain frequency. To achieve equally spaced spectral information, one should choose $\delta = \frac{1}{D}$ with a positive integer number of intermediate points D. The rotation angle starts with $\varphi = -\delta\pi$ at $n = 0$, increases with n, and reaches its maximum of $\varphi = \delta\pi\frac{N-2}{N}$ at $n = N-1$. In other words, the phase is linearly *ramped* across the time window. When this modified data set is Fourier-transformed, it has acquired a shift of the frequency scale according to the displacement law of Fourier transform

$$f_k = \frac{k}{N\Delta t} \to f_k^{(\delta)} = f_k + \delta\frac{1}{N\Delta t}. \tag{3}$$

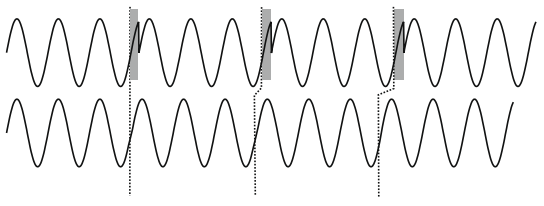

Fig. 1. Top: Sine wave with phase disruptions, as it occurs from periodic continuation in the non-coherent case. Bottom: By truncation of the time series (discarding the part indicated by shading) such that the coherence condition is met, a smooth continuation is achieved. This is the core idea of our proposal, but coherence can also be attained without sacrificing data points.

The variable δ defines the amount of spectral shift. At $\delta = 0$ the spectrum is not shifted at all, and $\delta = 1$ the spectrum is shifted by a full bin. Therefore, it is fully sufficient to use $0 < \delta < 1$ to get information between bins. Specifically, if one wishes to achieve D intermediate values, one can choose $\delta = 0/D, 1/D, \ldots, (D-1)/D$, and perform a Fourier transform at each of these D steps.

The additional phase-rotation in effect rescales the frequency. As the frequency is increased, the underlying frequency and the given time window will eventually reach the coherence point within the precision given by the D discrete steps. Here the spectrum will reach a maximum amplitude. Plotting all spectral data $\tilde{s}_k^{(\delta)}$ at the new frequencies $f_k^{(\delta)}$ yields a high resolution spectrum.

2.0.1. Procedure for PFT

The procedure for PFT in compact form is this:

Input: $N \geqslant 8$, $\Delta t > 0$, $D \in \mathbb{N}$ (one may conveniently choose N as a power of 2 and $D \in \{4, 8, 16, \ldots\}$). $s_n = s(n\Delta t)$ $(n = 0, \ldots, N-1)$ sampled values of a sufficiently smooth, quasiperiodic function $s(t)$:

- For $\delta = 0, \frac{1}{D}, \ldots, \frac{D-1}{D}$ compute by DFT (or if possible by FFT) of length N for $(k = 0, \ldots, N-1)$

$$\tilde{s}_k^{(\delta)} := \frac{1}{N} \sum_{n=0}^{N-1} s_n \exp\left(-i\delta\pi\left(\frac{2n}{N} - 1\right)\right) \cdot \exp\left(-\frac{2\pi i}{N} nk\right). \tag{4}$$

- Determine from the PFT spectrum $\left(\frac{k+\delta}{N\Delta t}, \tilde{s}_k^{(\delta)}\right)$ for $\delta = 0, \frac{1}{D}, \ldots, \frac{D-1}{D}$ and $k = 0, \ldots, N-1$ the local maxima of the linear splines through the points $\left(\frac{k+\delta}{N\Delta t}, |\tilde{s}_k^{(\delta)}|^2\right)$.
- If the linear spline has at $\frac{k'+\delta'}{N\Delta t}$ $(\delta' \in \{0, \frac{1}{D}, \ldots, \frac{D-1}{D}\};\ k' \in \{0, \ldots, N-1\})$ a local maximum with significant value $|\tilde{s}_k^{(\delta)}|^2$, then $\frac{k'+\delta'}{N\Delta t}$ is the wanted frequency and $\tilde{s}_{k'}^{(\delta')}$ is the corresponding complex amplitude.

Output: $\frac{k'+\delta'}{N\Delta t}$ frequency of $s(t)$, $\tilde{s}_{k'}^{(\delta')}$ complex amplitude.

3. Demonstrations

We will now demonstrate that this technique is beneficial: amplitude, frequency and phase of the spectral maximum are located with improved precision. We go through several examples to illustrate the difference.

3.1. Simple harmonic oscillation

3.1.1. Fixed frequency

Consider as the simplest case a sinusoidal wave

$$s_n = \cos(2\pi f_1 n) \quad (n = 0, \ldots, N-1) \tag{5}$$

with the normalized frequency $f_1 = f\Delta t$ sampled with $N = 2^8$ and $1/k = 0.06 = f_1$ (the latter implies that $k \approx 16.7$ points constitute one cycle). The time series then represents $N/k \approx 15.4$ periods of this oscillation. Fig. 2 compares conventional DFT (with no apodization used) and PFT. Frequency is here scaled to the Nyquist frequency which is half the sampling frequency. Only the low frequency part is shown for clarity. The lowest nonzero frequency is at $1/N$. In comparison, DFT produces sparse

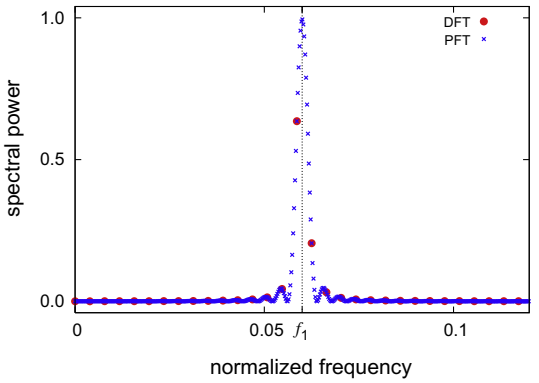

Fig. 2. The sinusoidal signal equation (5) with a normalized frequency of $1/k = 0.06$ (see vertical dotted line), analyzed with DFT (large symbols) and PFT (small symbols). Frequency is scaled to the Nyquist frequency; the lowest frequency bin is at 1/256.

data points from which it is hard to assess the spectral peak. It is evident that the signal is identified with much better precision when PFT is employed.

On close inspection, there are small secondary maxima in the power spectrum. These occur because off the position of the maximum, the coherence condition is not precisely met, and the continuation gives rise to phase jumps; therefore the square window is felt and convoluted into the spectrum.

To illustrate this last point, we present Fig. 3. It is based on the same data as in Fig. 2 but has a logarithmic vertical axis. In addition, it contains the result from a modified procedure in which a four-term Blackman–Harris apodization window was employed [2]. The apodization suppresses the background, which consists of ripples, by about eight orders of magnitude. This highly desirable modification comes at the price that the width of the spectral peak is increased by a factor of roughly 2, and the amplitude is also altered. The apodization affects both DFT and PFT in the same way, but PFT gives information between the original bins and therefore provides the peak position to a much better precision.

3.1.2. Variation of frequency

To obtain more precise statements about the achieved precision in the determination of frequency and amplitude, we consider simple harmonic oscillations at various frequencies f_1. Whenever the coherence condition is fulfilled by coincidence, even conventional DFT will return a precise frequency and amplitude value. For all other cases, however, the best

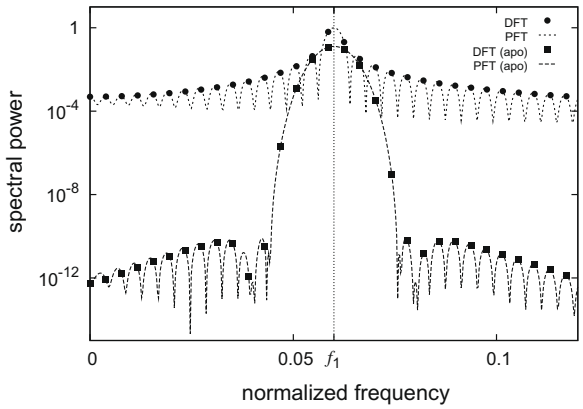

Fig. 3. Same as Fig. 2 but on a logarithmic scale. The spectral power retrieved by DFT and PFT is shown with and without apodization.

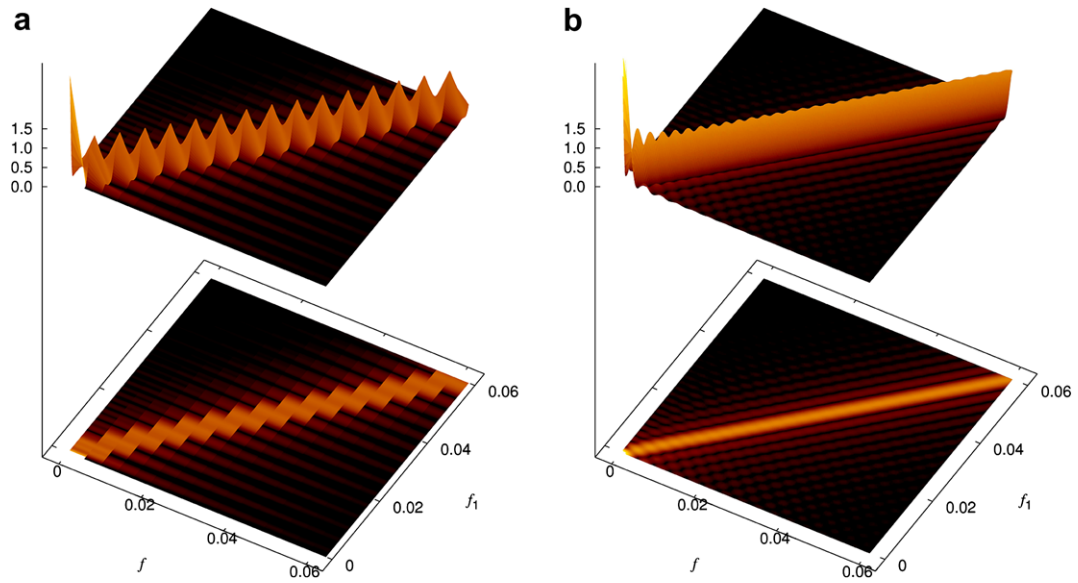

Fig. 4. Power spectra of a sinusoidal signal equation (5) calculated with DFT (a), and PFT (b). Shown are 3D representations (top) and contours in the $f - f_1$ plane. The lowest nonzero frequency is at 1/256. The underlying frequency f_1 varies form 0 to 0.06. The scalloping error in frequency and amplitude can be seen.

value for frequency and amplitude is given by that data point that has the highest amplitude of all power spectra (compare Fig. 2). Due to the coarse steps in DFT, both frequency and amplitude have scalloping errors.

Fig. 4 displays the obtained power spectrum (modulus squared of the amplitude vs. frequency f) when the true frequency f_1 is varied (i.e. k is varied). Due to the coarse resolution of DFT, the spectral maximum moves in a stair-like fashion; a mild improvement can be had by apodization. For PFT, the result is much smoother immediately.

For a more immediate comparison of frequency and amplitude errors, see Figs. 5 and 6. The frequency error and the amplitude error are plotted for the total meaningful range of f_1 for the considered example in Fig. 4. The frequency error alternates between plus/minus half a bin width for conventional DFT. PFT consistently keeps the error to one order of magnitude less than DFT. It shows relative weaknesses only at the lowest and highest few bins but still beats DFT everywhere. Part of the residual error stems from a phase dependent error described in Section 3.2. The remainder can be reduced by increased sampling rate.

The amplitude error (scalloping error) shows up in the DFT case as false dips by 60%, corresponding to −3.92 dB. This can be reduced to −0.86 dB with the best windowing functions [2]. PFT, in contrast, has much less deviation over most of the frequency range considered. For comparison we indicate the −0.86 dB as a dotted line in Fig. 6.

3.2. Bichromatic oscillation

While for a monochromatic signal the coherence condition can always be met, for two noncommensurate frequencies it cannot be met for both components at the same time. However, Trim-to-Coherence Fourier Transform still provides an advantage.

As data sets we prepared the sum of two monochromatic signals with same amplitudes

$$s_n = \cos(2\pi f_1 n) + \cos(2\pi f_2 n) \quad (n = 0, \ldots, N-1), \tag{6}$$

and the frequencies vary as in Fig. 4. As an example we use $f_2 = 0.06 - f_1$. Fig. 7(a) shows that resolution is restricted in the case of DFT due to the coarse steps in frequency. In contrast, PFT (b) allows to follow both frequencies through their crossing. Obviously, at very close separation just adjacent to the crossing point PFT has an enormous advantage.

The secondary maxima explained above appear here, too. Note, however, that these artifacts may now interfere with the main peaks and create amplitude errors. These errors are not specifically created by PFT: they are always there, but in DFT they go unnoticed because they are swamped by coarse resolution. Just as above, the artifacts and the errors they create can be removed by apodization, but only at the price of reduced resolution.

For the case of a closely spaced doublet of oscillations, Fig. 8 shows how the two retrieved spectral lines depend on the relative phase of the oscillations. With DFT (a), the doublet is not quite resolved. Variation of the relative phase affects the

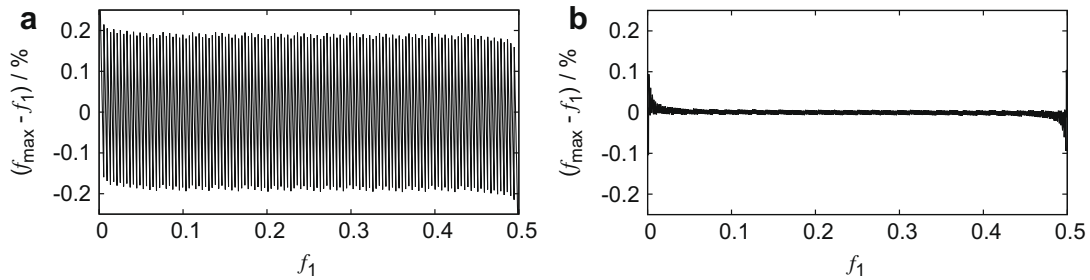

Fig. 5. Frequency error as function of frequency for conventional DFT (a) and PFT (b).

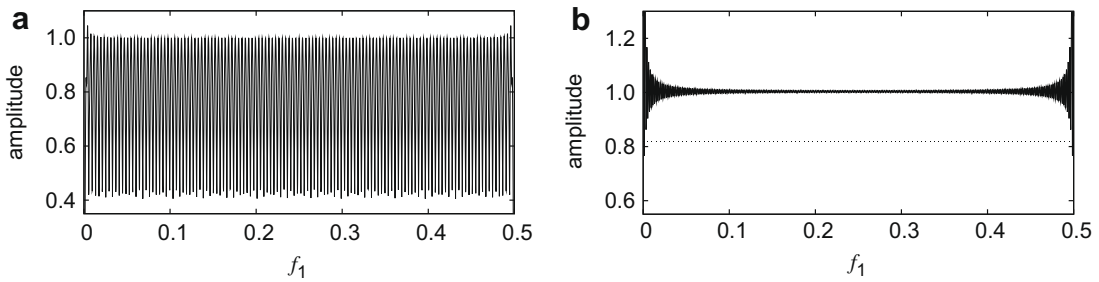

Fig. 6. Amplitude error (scalloping error) as a function of frequency for conventional DFT (a) and PFT (b). The dotted line indicates the most favorable case of windowed DFT, see text.

reconstructed amplitudes. With PFT (b), the doublet is fully resolved, and we see the same amplitude error arising again. In addition, an error becomes apparent in the assessment of the precise frequency difference. This error was swamped in the coarse resolution of DFT and went unnoticed; with the improved resolution of PFT it becomes visible.

3.3. Nontrivial chirp reconstruction: phase information

We turn to data representing a chirped Gaussian wave packet. Chirp refers to a non-constant value of the frequency of the wave over the duration of the envelope. Chirped pulses play a major role in radar technology, laser science, data transmission, etc. We here choose a linear spectral chirp, which means that the spectral phase is described by a parabola. The signal is given by [4]

$$s(T) = \exp\left(-\frac{1+Ci}{2}\left(\frac{t}{t_0}\right)^2\right)\cdot\cos\left(2\pi f_0\left(\frac{t}{t_0}\right)\right), \tag{7}$$

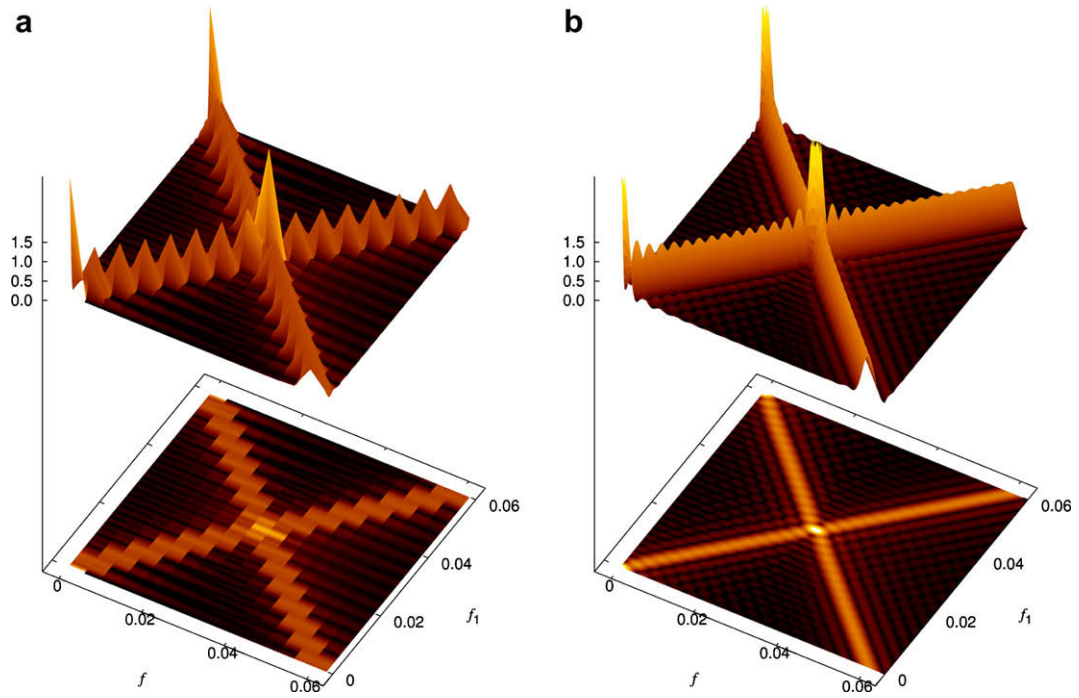

Fig. 7. Power spectra calculated with DFT (a) and PFT (b) of a doublet consisting of the sum equation (6) of two sinusoidal signals with the frequencies f_1 and $f_2 = (0.06 - f_1)$, where f_1 varies form 0 to 0.06. The lowest nonzero frequency is at 1/256. The scalloping error in frequency and amplitude can be seen.

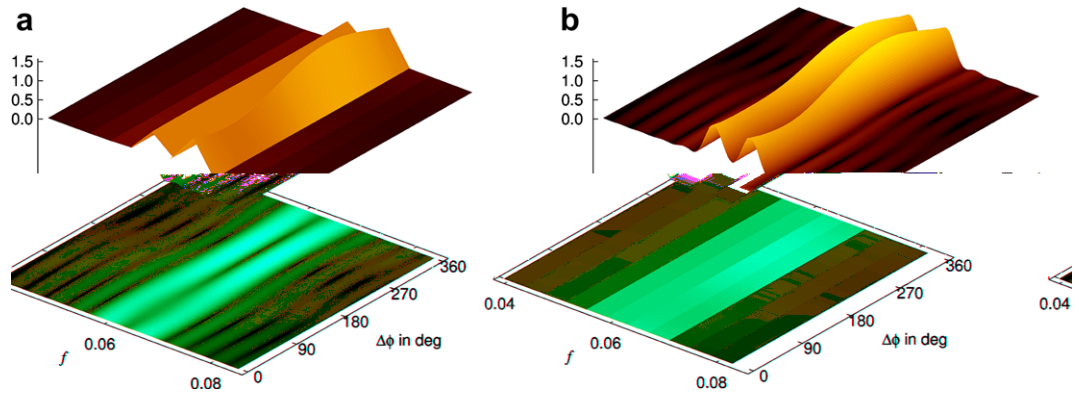

Fig. 8. Scalloping error of a doublet consisting of the sum Eq. (6) as a function of their phase difference $\Delta\phi$; the lowest nonzero frequency is at 1/256. The underlying frequencies are 0.0600 and 0.0655. (a) DFT; (b) PFT.

which has the spectral function

$$\tilde{s}(f) = \sqrt{\frac{2\pi t_0}{1+Ci}} \exp\left(-\frac{(2\pi(f-f_0))^2 t_0^2}{2(1+Ci)}\right). \tag{8}$$

We restricted the example to $N = 256$ sampling points t_n ($t_n = 0, \ldots, 255$). For convenience we use a centered wave packet which is shifted by 128. In the example we use a normalized duration of 80, which leads to the normalized time $\tau_n = (t/t_0 - 128)/80$ instead of $t_n = t/t_0$. The chirp C is set to 0.5 and the normalized carrier frequency f_0 to 6.2. The real part of this chirped wave packet is shown in Fig. 9.

By DFT and PFT we calculate power (Fig. 10) and phase (Fig. 11). The power spectrum shows the behavior described before. The advantage of DFT should speak for itself.

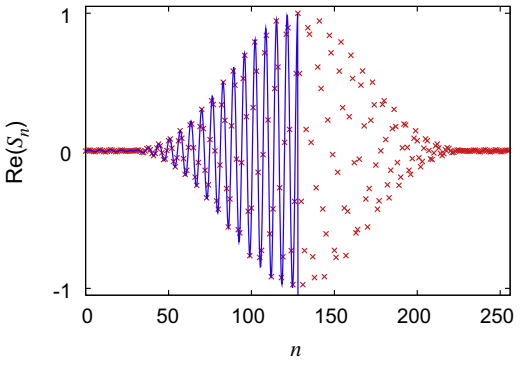

Fig. 9. Sampled chirped wave packet together with the function Eq. (7) which is shown for clarity only between 0 and 128.

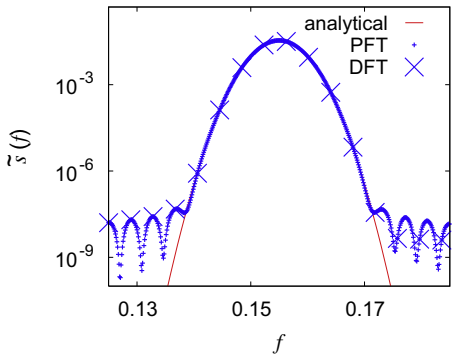

Fig. 10. PFT power spectrum of the wave packet shown in Fig. 9 together with the analytical expression Eq. (8) and a standard DFT.

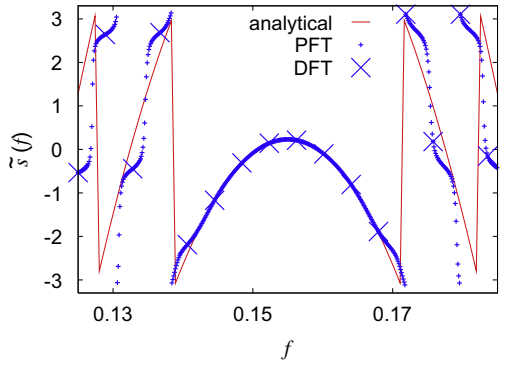

Fig. 11. PFT spectral phase of the wave packet shown in Fig. 9 together with the analytical expression Eq. (8) and a standard DFT.

4. Specifications

4.1. Resolution

With PFT the precision with which the exact frequency will be determined is no more given by the inverse of the time window length T. There is an improvement reflecting the intermediate frequencies between DFT bins. The constant δ can be chosen for a compromise between computation time and resolution increase.

4.2. Computational cost and efficiency

DFT is particularly efficient when the number of data points fulfils certain conditions. Most frequently, one uses powers-of-two, and applies *fast Fourier transform* (FFT) [6] or a refinement thereof [7,8]. When the available data happen not to conform to this criterion, a recommended strategy is zero padding [11] to reach the next power-of-two. If done judiciously, this strategy has its merits, but ultimately it adds something to the data that was not there in the original signal. It does not remedy the phase jumps from the continuous continuation, and it adds information about the signal which is false. Zero padding is driven by an interest in computational efficiency, not in accuracy in the representation of the underlying signal! For another comment on zero padding see below.

Computation time for DFT of length N scales as N^2, for FFT as $N\log(N)$. Trim-to-Coherence Fourier Transform has a disadvantage here. For PFT we need D transforms in which FFT may or may not be applicable. In the worst case, therefore, the computation time for PFT scales as DN^2; in more favorable circumstances, this is reduced to $DN\log(N)$.

In the example of 256 data points, the comparison would be $256\log_2(256) = 2^{11}$ for FFT, $256^2 = 2^{16}$ for DFT, and $D \cdot 256^2 = 16 \cdot 2^{16} = 2^{20}$ for PFT. These are considerable increases: 16-fold for PFT compared to DFT. However, even a standard desktop PC performs the PFT in well under one second. We therefore do not anticipate that PFT will be used where data need to be processed in real time. Rather, we see its domain of usefulness in those problems where one deals with unique data, i.e. data which stem either from a singular, non-repeatable incident, or are extremely difficult or costly to obtain. One then wants to extract as much information as possible, regardless of the size of the effort.

5. Example: sunspot cycle

Data sets which are hard or even impossible to be measured in more detail seem to be natural opportunities for PFT. Consider the historic data set of the number of sunspots, which is known to vary with a certain periodicity. In order to obtain the period with better precision, data acquisition for longer time (for more centuries) is not an option. Therefore methods like the one presented here are useful.

The number of sunspots has been counted and recorded on a daily basis for almost 260 years. It is well known that there is a variation with a period of about 11 years, known as the Schwabe cycle. More recent measurements, taking into account magnetic field data, indicate that magnetic polarity between northern and southern solar hemisphere alternates every other Schwabe cycle so that the true period really is twice as long (the Hale cycle). Longer periods such as ≈80 years (Gleissberg cycle) or ≈200 years have been reported by some researchers [12].

We used the monthly sunspot averages published in [13], considering $N = 3111$ data points (from January 1749 to March 2008) (Fig. 12). The spectra obtained from DFT and PFT are shown in Fig. 13. In the DFT spectrum there is some indication of a maximum near 3 nHz, but data points are so sparse that statements about any detail are not possible. From PFT, one immediately sees the maximum at 2.88 nHz. This frequency corresponds to a period of 11.00 years.

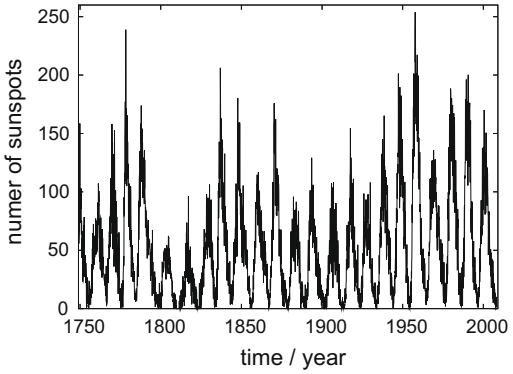

Fig. 12. Variation of the number of sunspots over the last centuries since beginning of systematic daily recordings.

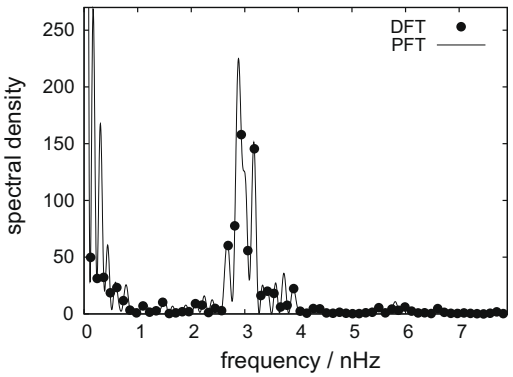

Fig. 13. Fourier transform of the time series shown in Fig. 12. Dots: DFT; line: PFT.

Other maxima can also be determined. The most prominent is at 3.17 nHz, which corresponds to 10.00 years. A closer examination of the involved frequencies, i.e. a PFT over a shorter time window, reveals that the frequency varies in time. Therefore, the fixed frequencies of 2.88 and 3.17 nHz are only an illusion, and the 'integer' number of years is purely coincidental. There is only one prominent frequency varying in the described range.

Then there is a weak overtone at twice the fundamental frequency, indicating that the sunspot variability is not a harmonic oscillation. Furthermore, there seems to be an indication of a subharmonic at one half the fundamental frequency which belongs to a period of 22 years, or the Hale cycle. Note that in the case of perfect symmetry of the two half-cycles in a Hale cycle the amplitude of this subharmonic would be zero. We conjecture that the nonzero value is more likely attributable to a statistical asymmetry, rather than an actual astrophysical asymmetry of the two hemispheres of the sun.

For extremely low frequencies there are some maxima, but at these frequencies the systematic errors are comparably large (compare with Fig. 6). Therefore, a prediction of sub-nHz frequencies corresponding to periods of more than 100 years is not meaningful.

6. Discussion

Phase unwrapping. In some cases conventional DFT is unable to reconstruct the spectral phase because phase variations from bin to neighboring bin reach or exceed the order of 2π. Since the transform provides the phase only modulo 2π, correct phase continuation is a severe problem which has already attracted quite some consideration [14]. Trim-to-Coherence Fourier Transform resolves many frequencies between the original DFT frequency bins; therefore the phase differences are reduced by a corresponding factor so that the continuation will, in most cases, cease to be a problem.

Iterative frequency search. With the additional knowledge that there is only one dominant frequency component in the signal, one can modify PFT to find this frequency in an iterative way. Instead of calculating D steps between each bin, δ is used as a variable which bisects the interval. Now the interval can be reduced by choosing that interval, which contains the higher spectral power. This can be repeated until the desired precision is reached. This procedure considerably decrease the computational time.

Another way to trim to coherence. The central idea about our technique is to avoid the phase jumps in the continuous continuation; PFT achieves that by rescaling the frequency. Another way would be to simply chop data points of the time series until an integer number of oscillations is left (to within the step width). This was schematically shown in Fig. 1. We have systematically explored this variant and find that it has advantages over DFT similar to those of PFT. However, we feel that PFT should be preferred because removing data points discards valuable information.

Zero padding. In the same spirit, zero padding also adjusts the length of the data set: conventionally, one appends zeroes until the next power-of-two is reached. As commented above, this is driven by an argument about computational efficiency, not about signal structure. It is also unfortunate because zeroes are almost always false information about the underlying signal. And in contrast to our technique, zero padding does nothing about the phase jumps which are the root cause of most problems.

A variant would be to add many more zeroes so that at some point one again obtains the coherent case to within the grid resolution. We have even explored this variant of trimming to coherence, and find the following: a given signal frequency can be trimmed to coherence exactly if it is commensurate with the data set length. PFT accommodates both commensurate and incommensurate cases directly by appropriate choice of δ. Whenever both happen to be commensurate, both approaches yield identical results. It is true that massive zero padding can approximate an irrational to any arbitrary precision when even more zeroes are appended, but then the computation time for zero padding diverges, whereas PFT can treat any incommensurate frequency in one go.

7. Conclusion

The method presented here provides spectral data points in between the frequency bins of standard DFT. According to conventional wisdom, the inverse of the bin separation equals the duration of the time series. It goes without saying that the increased number of data points in the spectrum from Trim-to-Coherence Fourier Transform in no way implies that after inverse transformation a longer piece of the temporal history can be reconstructed than was originally measured. If the data are truly stationary, such conclusion would be trivially true; if they are not, it would be wrong.

Several more tenets of conventional wisdom are disputed here:

- One should always strive to use FFT, and hence $N = 2^n$. If necessary, zeros can be padded.We have shown here that zero padding is not the only, and often not the best, way to reveal information about the signal.
- Apodization windowing is always beneficial.We have shown here that when the frequency of the spectral peak needs to be determined, apodization is not helpful. However, it can reduce spurious side peaks, if at the expense of reduced spectral resolution; this carries over to Trim-to-Coherence Fourier Transform.
- Spectral resolution is always limited by the total length of the time series.

We have shown (see Section 3.1.2) that the effective resolution can be improved by increasing the sampling rate beyond the minimum requirement of Nyquist's sampling condition. This is a major reinterpretation of the sampling condition which states that the full information is present as long as the sampling frequency is at least twice that of the maximum frequency in the signal. In Trim-to-Coherence Fourier Transform a significant oversampling (many more than two points per period) plays out its benefit in terms of enhanced precision in frequency, amplitude and phase.

We envision that Trim-to-Coherence Fourier Transform can be applied to a multitude of situations. It will prove most beneficial where data sets contain only few oscillations, as may be the case whenever (a) one prefers to look only at short time series because the stationarity of the process is in question, e.g. through drift of parameters or when (b) a unique data set from a singular, non-repeatable incident is all one has to work with. Both conditions are met in the sunspot example above; the latter can include historic data, or data taken from rare or catastrophic events which cannot be repeated. Consider analysis of the audio record of an anonymous telephone call in crime investigation, etc.

Appendix

Finally, we present a mathematical description of our method. For simplicity, we consider first a harmonic oscillation of fixed frequency. Then the exponential function $s(t) = \mathrm{e}^{2\pi i f_1 t}$ has the real frequency f_1. Instead of the function $s(t)$, only $N+1$ sampled values $s(n\Delta t)$ $(n = 0, \ldots, N)$ are given. Here $\Delta t > 0$ is the grid spacing. Let $T = N\Delta t$. Assume that the Nyquist sampling condition $|f_1| < 1/(2\Delta t)$ is fulfilled (see [15, p. 183]). It is our aim to determine the frequency f_1 approximately.

First we compute the Fourier coefficients $c_k^{(T)}(s)$ of $s(t)$ with respect to the interval $[0, T]$. As usually, $\mathbb{Z}$ denotes the set of all integers. If $f_1 T \in \mathbb{Z}$, then we obtain

$$c_k^{(T)}(s) = \frac{1}{T}\int_0^T s(t)\mathrm{e}^{-2\pi i t k/T}\,\mathrm{d}t = \delta(f_1 T - k) \quad (k \in \mathbb{Z}),$$

where $\delta(k)$ denotes the Kronecker symbol, i.e., $\delta(0) = 1$ and $\delta(k) = 0$ for $k \neq 0$. But if $f_1 T$ is not an integer, then

$$\boxed{c_k^{(T)}(s) = \mathrm{e}^{\mathrm{i}\pi(f_1 T - k)}\,\mathrm{sinc}(f_1 T - k) \quad (k \in \mathbb{Z}).}$$

Here we use $\mathrm{sinc}\, t = \sin(\pi t)/(\pi t)$ for $t \neq 0$ and $\mathrm{sinc}\, 0 = 1$. The Fourier coefficient $c_k^{(T)}(s)$ with index k closest to $f_1 T$ will have the largest magnitude. Fourier coefficients with neighboring indices k (i.e. $|f_1 T - k|$ is small) typically will have nonzero values. This is the phenomenon of *sidelobes*. The appearance of these sidelobes is a symptom of *leakage*. This leakage effect produces frequency components that are *not* present in the original function $s(t)$. Further, the Fourier coefficients oscillate and converge to zero like $|f_1 T - k|^{-1}$ for $|k| \to \infty$.

The DFT has a similar property of leakage. By the trapezoidal rule, we can compute $c_k^{(T)}(s)$ $(k = -N/2 + 1, \ldots, N/2)$ via DFT of length N, where we set $h(t) = \mathrm{e}^{2\pi\mathrm{i}(f_1 T - k)t/T}$:

$$c_k^{(T)}(s) \approx \frac{1}{N}\left[\frac{1}{2}h(0) + \sum_{n=1}^{N-1} h(n\Delta t) + \frac{1}{2}h(T)\right] = \tilde{s}_k^{(N)} = \frac{1}{N}\sum_{n=0}^{N-1} s_n w_N^{nk} \quad (w_N = \mathrm{e}^{-2\pi\mathrm{i}/N}).$$

Using average value at endpoints, the input values s_n of DFT of length N are given by

$$s_n = \begin{cases} s(n\Delta t) & \text{for } n = 1, \ldots, N-1, \\ \frac{1}{2}(s(0) + s(T)) & \text{for } n = 0. \end{cases}$$

Now we calculate the values $\tilde{s}_k^{(N)}$ $(k=-N/2+1,\dots,N/2)$ of the DFT of length N explicitly by the known summation formula

$$\sum_{n=1}^{N-1} \mathrm{e}^{\mathrm{i}nx} = \mathrm{e}^{\mathrm{i}Nx/2}\,\frac{\sin((N-1)x/2)}{\sin(x/2)}$$

with $x=2\pi(f_1T-k)/N$. Note that from $|f_1|<1/(2\Delta t)$ and $|k|\leqslant N/2$ it follows $|x|<2\pi$. By simple calculation, we get for $k=-N/2+1,\dots,N/2$

$$\boxed{c_k^{(T)}(s)\approx\tilde{s}_k^{(N)}=\mathrm{e}^{\mathrm{i}\pi(f_1T-k)}\,\frac{\cos(\pi(f_1T-k)/N)}{\operatorname{sinc}((f_1T-k)/N)}\cdot\operatorname{sinc}(f_1T-k).}$$

For the *Phase-Rotation Fourier Transform*, we multiply $s(t)$ by $\mathrm{e}^{-2\pi\mathrm{i}\delta t/T}$ $(\delta\in[0,1[)$ and consider

$$s^{(\delta)}(t)=\mathrm{e}^{2\pi\mathrm{i}(f_1T-\delta)t/T}.$$

Then the Fourier coefficients of $s^{(\delta)}(t)$ with respect to the interval $[0,T]$ read as follows:

$$c_k^{(T)}(s^{(\delta)})=\mathrm{e}^{\mathrm{i}\pi(f_1T-\delta-k)}\operatorname{sinc}(f_1T-k-\delta)\quad(k\in\mathbb{Z}).$$

By the trapezoidal rule, we can compute $c_k^{(T)}(s^{(\delta)})$ approximately via DFT of length N. Let $\tilde{s}_k^{(\delta)}$ $(k=-N/2+1,\dots,N/2)$ be the values of this DFT. Similarly as above, we obtain for $k=-N/2+1,\dots,N/2$

$$c_k^{(T)}(s^{(\delta)})\approx\tilde{s}_k^{(\delta)}=\mathrm{e}^{\mathrm{i}\pi(f_1T-\delta-k)}\,\frac{\cos(\pi(f_1T-k-\delta)/N)}{\operatorname{sinc}((f_1T-k-\delta)/N)}\cdot\operatorname{sinc}(f_1T-k-\delta).$$

Thus, $\tilde{s}_k^{(\delta)}$ is an approximate value of the Fourier coefficient $c_k^{(T)}(s^{(\delta)})$. By

$$c_k^{(T)}(s^{(\delta)})=\frac{1}{T}\int_0^T s(t)\,\mathrm{e}^{-2\pi\mathrm{i}t(k+\delta)/T}\,\mathrm{d}t,$$

$c_k^{(T)}(s^{(\delta)})$ coincides with the kth Fourier coefficient of $s(t)$ with respect to the rotated harmonic wave $\mathrm{e}^{2\pi\mathrm{i}t(k+\delta)/T}$ with frequency $f=(k+\delta)/T$. Now we substitute f into $\tilde{s}_k^{(\delta)}$ and obtain for $k=-N/2+1,\dots,N/2$

$$\tilde{s}_k^{(\delta)}=\mathrm{e}^{\mathrm{i}\pi(f_1-f)}\cdot\frac{\cos(\pi(f_1-f)\Delta t)}{\operatorname{sinc}((f_1-f)\Delta t)}\cdot\operatorname{sinc}T(f_1-f),$$

such that the power spectrum reads as follows:

$$\boxed{|\tilde{s}_k^{(\delta)}|^2=\left(\frac{\cos(\pi(f_1-f)\Delta t)}{\operatorname{sinc}((f_1-f)\Delta t)}\right)^2\cdot(\operatorname{sinc}T(f_1-f))^2\leqslant 1.}$$

Note that the estimate $x\leqslant\tan x$ for all $x\in[0,\pi)$ provides

$$\left(\frac{\cos(\pi t)}{\operatorname{sinc}t}\right)^2\leqslant 1\quad(t\in(-1,1)).$$

The maximum value of the power spectrum is one at $f=f_1$. For $T\gg 1$, the power spectrum has a sharp peak at $f=f_1$. We choose $k'\in\{-N/2+1,\dots,N/2\}$ and $\delta'\in[0,1[$ such that $|\tilde{s}_{k'}^{(\delta')}|^2\approx 1$. Consequently, the wanted frequency f_1 is approximately given by $f=(k'+\delta')/T$.

These results can be immediately extended to the general case of an *exponential sum*

$$s(t)=\sum_{j=1}^{J} a_j\,\mathrm{e}^{2\pi\mathrm{i}f_jt}$$

with complex coefficients $a_j\neq 0$ and real frequencies f_j with a small integer $J>1$. We know only the sampled values $s(n\Delta t)$ (

$$a_j \mathrm{e}^{\mathrm{i}\pi T(f_j - f)} \frac{\cos(\pi(f_j - f)\Delta t)}{\operatorname{sinc}((f_j - f)\Delta t)} \cdot \operatorname{sinc}(T(f_j - f))$$

dominates all the other terms of the sum $\tilde{s}_k^{(\delta)}$, since the magnitude of each other term is smaller than ε near f_j. By $T \gg 1$, in the power spectrum we can observe sharp peaks at $f = f_j$ $(j = 1, \ldots, J)$ of height $|a_j|^2$. In the power spectrum, we determine all maximum values at $(f_j, |a_j|^2)$ $(j = 1, \ldots, J)$ with $|a_j| \geqslant J\varepsilon$ approximately.

References

[1] A. Ferrero, R. Ottoboni, High-accuracy Fourier analysis based on synchronous sampling techniques, IEEE Trans. Instrum. Meas. 41 (1992) 780–785.
[2] F.J. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, Proc. IEEE 66 (1978) 51–83.
[3] C.E. Shannon, Communication in the presence of noise, Proc. IRE 37 (1949) 10–21.
[4] G.P. Agrawal, Nonlinear Fiber Optics, Academic Press, San Diego, 1995.
[6] J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, Math. Comput. 19 (1965) 297–301.
[7] M. Frigo, S.G. Johnson, FFTW: an adaptive software architecture for the FFT, in: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 3, 1998, pp. 1381–1384.
[8] M. Frigo, S.G. Johnson, The design and implementation of FFTW3, Proc. IEEE 93 (2005) 216–231.
[11] S.M. Kay, S.L. Marple Jr., Spectrum analysis: a modern perspective, Proc. IEEE 69 (1981) 1380–1419.
[12] W. Gleissberg, A long-periodic fluctuation of the sun-spot numbers, Observatory 62 (1939) 158–159.
[13] National Oceanic & Atmospheric Administration (NOAA), Washington, DC, USA. Data are taken from the website: <ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SUNSPOT_NUMBERS/MONTHLY.plt>.
[14] J.M. Tribolet, A new phase unwrapping algorithm, IEEE Trans. Acoust. Speech Sig. Process. 25 (1977) 170–177.
[15] W.L. Briggs, V.E. Henson, The DFT: An Owner's Manual for the Discrete Fourier Transform, SIAM, Philadelphia, 1995.