Inconsistencies in the harmonic analysis of the time series of pulsating stars observed by space missions

J. Pascual-Granado,* R. Garrido,† and J. C. Suárez‡
Instituto de Astrofísica de Andalucía (CSIC), 18008, Granada, Spain
(Dated: July 24, 2013)

Purpose: The current paradigm of asteroseismology establishes the harmonic analysis of the time series of a pulsating star as the method to find unambiguously its frequency content. On the light of the challenges raised by the ultra-precise photometry supplied by space missions this assumption is here reconsidered. In the present work we test the analyticity of the underlying functions describing the observations.

Method: The reliability of the paradigm is studied using a method based on the Stone-Weierstrass theorem. More specifically we check the correspondence of some properties derived from the spectral components found on the stellar light curves with classical Fourier algorithms. This is done by examining a fine-scale property of the time series, the connectivity, which is intrinsically related with the differentiability of the underlying functions. We performed a connectivity analysis to a multiperiodic pulsating δ Scuti star HD 174936 observed by *CoRoT*. The connectivities of the *CoRoT* data were compared to those obtained from an analytic model typically representative of the frequency content of the star. In order to discard any possible contribution of the instrumentation to the results the method was applied also to the hybrid δ Scuti-γ Dor star KIC 006187665 observed by the *Kepler* space mission.

Results: The comparison of the connectivities of *CoRoT* data and the analytic model yielded very different results demonstrating that the light curve cannot be represented by the analytic model. In addition, the correlated residuals resulting from the connectivity analysis for *CoRoT* and *Kepler* data demonstrate that the functions underlying the light curves are not analytic. Instrumental effects are discarded, since the method was applied to stars observed by two different satellites of different characteristics (precision, sampling, etc.).

Conclusions: The dramatic consequence is that the periodogram is not a consistent estimator of the spectral density of the time series. Therefore, the light curves of these pulsating stars cannot be expanded in Fourier series.

Keywords: methods: data analysis — methods: statistical — stars: oscillations — *Kepler*

INTRODUCTION

Since *SoHO* started the era of asteroseismology from space with the Sun observations [1], outstanding advances in the knowledge of the solar structure have been made. However, some of the objectives of this space mission remains unsolved: the non-detection of solar g-modes [2] and the physical origin of the near surface frequency effects [3] are good examples. More recently Hanasoge et al. [4] using data from NASA satellite SDO (Solar Dynamics Observatory) found a velocity field on the solar atmosphere two orders of magnitude lower than theoretically predicted, which lead them to prompt the following question: "(sic) What mechanism transports the heat flux of a solar luminosity?".

Thanks to *CoRoT* [5] and later *Kepler* [6] missions, ultra-high quality asteroseismic data from space were also available for other stars than the Sun [see the impact of these missions into the physics of stellar structure models in 7], eliminating ground-based observational windows and removing the noise produced by the atmospheric transparency variations. Analysis of such precise time series of multi periodic pulsating stars [e.g. 8–10] has revealed a new puzzle not yet understood: the detection of a huge number of frequencies together with non-gaussian residuals of the light curve fitting.

The unprecedented quality of the data obtained from the above mentioned space missions allows the improvement of the theoretical pulsation models and therein to infer a better description of the stellar structures. But, for a correct understanding of the nature of the observed physical processes, the mathematical description used to interpret the data has to be able to reproduce the complete data set while only leaving as residuals an independent random series. This rarely happens in practice where the residuals, small as they are, keep some structure, which makes them far different from an independent random sequence. This is especially the case in some δ Scuti stars [10, 11] and γ Dor stars [12] observed by *CoRoT* .

These challenges lead us to reconsider some of the basic foundations of the harmonic analysis of data in asteroseismology, where Fourier series (i.e. Lomb-Scargle periodogram, [13]) have been used widely as mathematical description.

In this paper we check the correspondence of some properties derived from the spectral components found on the stellar light curve of HD 174936 using classical Fourier algorithms with properties derived from the original data. With this aim, a fine-scale property called connectivity, which is intrinsically related with the differentiability of the function describing the observations, has been examined in detail.

This allows us to check the identity between the obtained spectral representation and the data, and at the same time to test the validity of the assumption of the square integrability of the signal, which is a necessary condition for the Fourier expansion to be convergent.

The paper begins in Sec. 2 with some esential considerations concerning the reliability of the expansion of a function in Fourier series depending on its analyticity. In Sec. 3 the definition of connectivity is given, which is, basically, a function describing how smoothly connected are the data points of a given sampled function. Thus defined, connectivity gives an adequate approximation of the differentiability for discrete time series, provided that the function is band-limited and sufficiently sampled. In Sec. 4, two approximations are used to calculate the connectivities: one is based on a closed-form mathematical model, the cubic spline polynomials, and the other one is based on a non-closed-form model given by autoregressive techniques (ARMA). The results of the tests on real data obtained from the asteroseismic space mission *CoRoT* are presented in Sec. 5, where we try to reproduce the results obtained by fitting a model based on the assumption that the time series consisted on a discrete set of individual frequencies with white noise added. We compared the results obtained for both real and fitted light curves showing large differences. In order to discard eventual instrumental effects, the same test is performed on a time series obtained by another space instrument, *Kepler* (Sec. 6). Finally, in Sec. 7, a discussion of the main results of the paper is presented together with the relevant consequences derived from our findings and summarized in the Conclusions.

ANALYTICITY

Fourier analysis is a useful technique for frequency detection but it has a variety of difficulties that are not always considered in the literature. If the sequence analyzed belongs to a nonperiodic or quasiperiodic data set, the number of spectral components needed to represent the signal can be so large that the physical information provided by its Fourier representation could be no longer useful. Constraining the number of spectral components used for the harmonic analysis only increases the probability of misleading information as a result. In any case, and more importantly, in order to apply Fourier techniques to the function, Parseval's theorem must hold. This theorem states that the integral of the squared modulus of a function is equal to the integral of the squared modulus of its Fourier transform. For this to be true the function must be square integrable

$$\int_{-\infty}^{+\infty} |F(x)^2| dt < \infty \qquad (1)$$

The function is square integrable if the Fourier series of the function converge uniformly pointwise. Unlike the Fourier series, which is an infinite sum, the Discrete Fourier Transform (DFT) and its inverse, the IDFT, which is the analogous to the Fourier series, always converges if the sequence is finite, because it is a finite sum. But when the Fourier series of the function is divergent, or not convergent to the value of the function at this point, the DFT no longer provides a physical description of the function based on its frequency components. That is, a frequency is only a well-defined physical observable when the Fourier series converge.

Therefore, before a Fourier analysis can be applied to a time series, the convergence of the Fourier series expansion of the function must be demonstrated.

On the other hand, the analyticity of a function is a sufficient condition for the square integrability. A function is said to be analytic if it is infinitely differentiable so that its Taylor expansion is convergent. If this condition is not met the convergence of the Fourier series is not guaranteed. That is the reason why it is so important to evaluate the analyticity of the function. In this paper the term analyticity refer to the property that a given function can be expanded in Fourier series. To study the analyticity of a function we give an approximation of the differentiability for discrete time series, the connectivity.

CONNECTIVITY

Time dependent astronomical data are given as a discrete sequence of samples from a function, which is the representation of an observable satisfying the Lebesgue measure definition[25]. In the case of photometric data, such a discrete sequence is called light curve.

Let us consider here a discrete series as a finitely close sequence of data points and the function being sampled, a set of infinitely close datapoints. The first question arising is: can a discrete sequence of values fully and uniquely determine the properties of the function? This is the well-known problem of sampling. The Nyquist-Shannon theorem states that the function can be reconstructed perfectly well by means of a regular sampling of the function if, and only if, it is band-limited to less than half the sampling rate, i.e. the Nyquist frequency.

When the conditions above mentioned are satisfied, then all the properties of the function can be studied from a discrete series, even those concerning the pointwise limits of the function. For our purposes this means that the differentiability of the function can be evaluated adequately.

Concept

Here we are interested on how smoothly the discrete sequence can be connected in order to reproduce exactly the function. We call this property: the connectivity of the function at a given point. This property is closely related to the differentiability.

In general it can be said that a discrete sequence of data points generated from a function is composed of a stochastic sequence plus a deterministic contribution that can have or not a wave-like structure. In fact, the Wold decomposition theorem [14] states that every stationary process can be represented always as an stochastic process with a purely deterministic component. The random component of the stochastic process appears to complicate the study of a time series when the connectivity of the data is being analysed but a very robust mathematical result, the Kolmogorov continuity theorem, which applies for stochastic time series, allows us to fully characterise the properties of the function [15]. This theorem permits to extend properties of stochastic processes, like differentiability, which would require infinitely close data points, to a function, given a finite number of samples. This is a consequence of the topological separability of the function, and it is applicable if certain constraints on the moments describing the data differences are satisfied. Basically, the condition states that the moments of the data differences change at most with a power law. Therefore, it is not expected that the random components of the data difficult seriously the characterization of the function by means of its connectivities.

Definition

It is usual to study the continuity of a function at a given point by calculating both one-sided limits in order to check whether they converge to the same value and coincide with the value of the function at that point. In that case the function can be said to be continuous at that point.

In a similar way, given a datapoint from a discrete sequence, a forward and backward extrapolation from a subset bracketing the selected datapoint can be computed to check whether they both converge to the same value and coincide with the value of the selected datapoint of the sequence.

The above theoretical framework permits the following self-consistent definition of the connectivity $\mathcal{C}_n$ of a datapoint x_n of a discrete sequence as sampled from the function $F(x)$ as

$$C_n = \epsilon_n^f - \epsilon_n^b \,, \tag{2}$$

where ϵ_n^f, ϵ_n^b, are the deviations of the forward and backward extrapolations from the datapoint x_n.

$$\begin{aligned} \epsilon_n^f &= x_n^f - x_n \\ \epsilon_n^b &= x_n^b - x_n \,. \end{aligned} \tag{3}$$

are the deviations of the forward and backward extrapolations from the datapoint x_n, represented by x_n^f and x_n^b, respectively.

A discrete approximation for the pointwise derivability condition for $F(x_n)$ can be:

$$\frac{x_n^b - x_{n+1}}{\Delta t} = \frac{x_n^f - x_{n-1}}{\Delta t} \tag{4}$$

where Δt is the sampling rate of the sequence. It is evident from these equations that connectivity is closely related to derivability. The numerical approximation of the point derivative $\mathcal{D}_n$ at x_n can be expressed as:

$$\mathcal{D}_n = \frac{\mathcal{C}_n + x_{n+1} - x_{n-1}}{2\Delta t} \tag{5}$$

which reduces to the typical point derivative for discrete data when $\mathcal{C}_n = 0$. When connectivities are not zero but independent randomly distributed values, the derivative is still well-defined. In that case, the connectivities can be considered simply as deviations from the true derivative at this point. On the contrary, when the connectivities are correlated with the data the derivability condition is not fulfilled.

METHODS

The forward and backward extrapolations necessary to obtain the connectivities have to be calculated numerically. Two numerical approaches, conceptually very different, have been considered. The first one is based on a representation given by a closed-form function, the cubic spline polynomials, the second one is based on an autoregressive model, which is not described by any closed-form function. The detailed description of these calculations and the properties of these mathematical tools are given in the next sections.

Cubic splines

The Stone-Weierstrass theorem [16] states that if a function is uniformly continuous in a closed interval it can be approximated arbitrarily well, i.e. as closely as desired, by a polynomial of degree n, with n a natural number. Then, the entire function can be expressed as a piecewise polynomials function of finite degree. But piecewise polynomials functions have a drawback, the Runge's phenomenon, resulting in oscillations at the edges of the approximated function, which is the equivalent to the Gibbs phenomenon in the Fourier description.

Another crucial point here is that piecewise polynomials are continuous functions but not necessarily analytic because they can show instabilities at high order derivatives. Then, instead of choosing any polynomials to model the function, piecewise smooth polynomials were chosen. Adding smoothness as an additional constraint guarantees analyticity and then the convergence of Fourier series is assured, also giving a natural solution to the Runge's phenomenon. That is the reason why cubic splines are selected as the mathematical technique to calculate forward and backward extrapolations using an analytic approach.

Formulated in this very general way any particular property can be extended to any analytic function. That is, if a continuous function can be fitted arbitrarily well with a given cubic spline parametric model then the analyticity condition is automatically satisfied, otherwise the convergence of the Fourier series is not guaranteed for that particular function.

The approximation using cubic splines gives us key information about local features of the function which generates the observed discrete sequence. Other non-smooth approaches, such as the Hermite polynomials, are able to capture more features from the function when it is non-analytic. This is the case for the time series analyzed here as will be discussed later in Sec. VII. However, we are more interested now in another approximation using autoregressive methods, because they can fit a non-analytic function also and no closed-form formula is needed for that.

ARMA

Fourier decomposition is useful for describing periodic processes with small non-periodic contributions. That is because the expansion in Fourier series of a periodic function has a finite number of terms. In contrast, when every feature of a function requires independent modelling and there are non-periodic features whose Fourier series expansion might be infinite, the autoregressive methods can be more appropiate for the description.

In the present work, we used the well-known autoregressive method ARMA (Auto Regressive Moving Average) [17] for the connectivity analysis, which is particularly interesting for our purposes because it can fit fairly well the data even when the function is non-analytic and no closed-form formula is necessary (see below Eq. 6)

ARMA models have been used to calculate a forward and backward extrapolation from a subset bracketing every datapoint of the dataset. The ARMA model is composed of two statistical processes: the autoregressive (AR) and a moving average (MA) contribution:

$$x_n + \sum_{i=1}^{p} a_i x_{n-i} = \epsilon_n + \sum_{j=1}^{q} b_j \epsilon_{n-j} \qquad (6)$$

where the left hand side represents the autoregressive process being x_n the data points and the right hand side corresponds to the moving average process being ϵ_n the terms of a white noise process. The a_i, b_j parameters are numerical coefficients containing the correlation between one term of a given sequence and the previous one.

Just like cubic splines modelling, the accuracy of the predictions with ARMA models depends on the number of data points of the segment to be modelled in order to calculate forward and backward extrapolations as defined in Sec. III. In addition to this, ARMA extrapolations depend on two parameters: the orders p and q, which are the number of autoregressive coefficients and the number of moving average coefficients, respectively. Hereafter, ARMA descriptions of time series will be denoted by its coefficients as ARMA(p, q).

In general, an ARMA model is able to reproduce a sinus wave with only three terms. Accordingly, the number of terms increases with the frequency content of the signal.

The ARMA modelling involves three steps:

1. Identification of the order of the starting model. This affects principally the velocity of convergence but does not modify the result.

2. Calculation of the parameters a_i, b_j using a Steiglitz-McBride algorithm [18] given the orders for the AR and MA contributions.

3. Evaluation of the validity of the model based on the goodness of the fit in backward and forward extrapolations. The algorithm starts again with increased orders until a minimum in the residuals is reached.

In the following section both techniques will be applied, cubic splines and ARMA, to our case of study: the time series corresponding to light variations of a star given as the integrated flux as measured by the photometric instrument onboard a space satellite, where the satellites are *CoRoT* and *Kepler*, and the stars are the δ Scuti HD 174936 and the hybrid pulsating star KIC 006187665.

COROT TARGET: HD 174936

The light curve of the δ Scuti star HD 174936 has been analyzed using classical Fourier analysis in García-Hernández et al. [11]. They found 422 significant frequencies in the power spectrum of this star. The authors follow the conservative limit given by Breger et al. [19] to consider a given peak in the power spectrum as a statistically significant frequency as the Fisher's test for periodicity establishes.

However, the 422 frequencies are distributed in such a large range of periods that none of the stellar theoretical models studied in the paper to represent the internal structure of this star is able to reproduce.

The physical origin of the excitation of the peaks interpreted as frequencies in the power spectra of this kind