Extension of Whittle likelihood to transient signal parameters estimation. Application to stellar speckle interferometry

A. Ferrari and Y. Cavallin
Université de Nice Sophia-Antipolis
Laboratoire d'Astrophysique, UMR 6525
Parc Valrose, 06108 Nice Cedex 2, France
ferrari@unice.fr, cavallin@unice.fr

ABSTRACT

This paper addresses the problem of estimating the parameters of a transient random process. In many applications, the absence of model for the samples leads to the impossibility to use the classical estimation theory tools. This paper develops a cost function for the case where a parametric model for the second order spectral characteristics exists. This cost function is build from the statistical properties of the signal discrete Fourier transform (DFT) when the number of samples on the signal support tends to infinity. When the DFT bins are decorrelated, this criterion has exactly the same form as the Whittle likelihood for the stationary signals case, the power spectrum being replaced by the energy spectrum. Application of this estimation framework to visibility estimation in stellar speckle interferometry is proposed.

1 INTRODUCTION

The estimation of the parameters of a transient signal is a problem arising in many physical situations. If a model of the measurements is available, it can be treated using the classical estimation theory tools, [4]. However, contrarily to the case of a stationary signal, the number of samples is not relevant to study the asymptotic properties of the estimators of transient signal parameters. Thus, as for a deterministic signal plus noise, this study is generally carried out considering that a properly defined signal-to-noise ratio tends to infinity, [4].

However in many situations a temporal model of the signal is not available. In stellar interferometry for example, it is only possible to obtain solely a parametric model of the energy spectrum of the interferometric image, [6]. In the stationary context, situations where only a parametric model of the power spectrum is available, or where a spectral model is more tractable than the temporal model, also often occur, like for time series with long range dependence, [8]. A very valuable tool to analyze these type of processes is the Whittle's likelihood, [9, 2], that gives an approximate expression of the signal likelihood as a function of the periodogram computed from the measured data and the theoretical power spectrum. The use of this cost function has recently received an increased interest [8, 3].

The objective of this communication is to derive a cost function for the estimation of transient signals parameters that possesses the same structure as the Whittle likelihood, i.e. built on the discrete Fourier transform (DFT) of the data. The Whittle likelihood deeply relies on the statistical properties of the DFT when the number of samples is large, [3]. As mentioned above this approach is irrelevant for a transient signal. Moreover the definition of a signal-to-noise ratio is not always possible.

After a presentation of the notations used in this communication, the third section will study the statistical properties of the DFT of a transient signal when the number of samples on the support of the signals tends to infinity, i.e. when the sampling frequency tends to infinity. It demonstrates that the DFT bins are asymptotically Gaussian distributed with a covariance function that depends on the two dimensional Fourier transform of the signal autocorrelation. Departing from this result, the fourth section derives a new cost function for the parameter estimation problem. This cost function possesses the same structure as the Whittle likelihood in the case where the DFT bins are asymptotically decorrelated. Finally, the last section is devoted to the application of this estimation scheme to the estimation of the visibility in stellar speckle interferometry from simulated data. These computer experiments prove the superiority of the proposed method with respect to the technique usually used.

2 PROBLEM STATEMENT AND GENERAL FORMALISM

Let $x(l)$ be a real valued continuous time transient random signal defined without loss of generality on the interval $[0, T]$. For simplicity, the development is restricted herein to one-dimensional functions. It is assumed that $x(l)$ is a finite energy signal:

$$\int_0^T \mathsf{E}[x(l)^2]dl < \infty. \tag{1}$$

The "energy spectrum" of the nonstationary signal $x(l)$ can thus be defined as:

$$\tilde{S}(f) = \mathsf{E}\left[\left|\int_0^T x(l)e^{-j2\pi fl}dl\right|^2\right] = \int_{-T}^{T} c(\tau)e^{-j2\pi f\tau}d\tau \tag{2}$$

where the "covariance" function $c(\tau)$ is:

$$c(\tau) = \int_0^{T-|\tau|} \mathsf{E}[x(l)x(l+|\tau|)]dl. \quad (3)$$

Note that under condition (1) the existence of $c(\tau)$ is verified. It is assumed that the "energy spectrum", denoted $\tilde{S}(f;\boldsymbol{\theta})$ in the sequel, can be parametrized by a set of unknown parameters θ_i to be estimated.

N samples of this continuous signal are acquired in the interval $[0, T]$ with a sampling interval $T_e = T/N$:

$$x_n = x(nT_e),\ n = \{0, \ldots, N-1\}. \quad (4)$$

Using the usual formalism [7], the DFT of x_n is defined as:

$$X_k = \sqrt{T_e} \sum_{n=0}^{N-1} x_n \exp(-j2k\pi n T_e),\ 0 \le k < N. \quad (5)$$

3 ASYMPTOTIC DISTRIBUTION OF THE DFT

The purpose of this section is to derive the joint asymptotic distribution (for N large) of the X_k, $k = \{0, \ldots, N-1\}$. The analysis is split in two parts: first the x_n are assumed independent and then statistical dependence is introduced through a linear filtering of the previous independent sequence.

Let $\kappa_{x,m}(l)$ be the m-th order cumulant of x_n. It is assumed in the sequel that:

$$\forall m,\ \int_0^T |\kappa_{x,m}(l)| dl < \infty \quad (6)$$

The asymptotic joint Gaussian distribution of the X_k is derived following the proof given in [1] for the traditional stationary case. The cumulant $\mathsf{cum}[X_{k_1}, \ldots, X_{k_m}]$ can be written:

$$\mathsf{cum}[X_{k_1}, \ldots, X_{k_m}] = (T_e)^{\frac{m}{2}} \sum_{n=0}^{N-1} \kappa_{x,m}(n) e^{-j2\pi(\sum_{q=1}^m k_q) n T_e}$$

In this expression the k_l can be negative in order to take into account the case where the conjugated DFT $X^*_{-k_l}$ is considered. The derivation of an equivalent of this cumulant for N large, as in [1] for the stationary case, seems impossible to obtain in our case without further information on the $\kappa_{x,m}(n)$. For this reason the proof will reduce to show the convergence toward 0 of this cumulant in the case $m > 2$. This can be achieved using the rough majoration:

$$|\mathsf{cum}[X_{k_1}, \ldots, X_{k_m}]| \le (T_e)^{\frac{m}{2}-1} T_e \sum_{n=0}^{N-1} |\kappa_{x,m}(n)|.$$

When N tends to infinity, $T_e^{\frac{m}{2}-1}$ will tend to zero for $m > 2$ whereas the second term will converge to $\int_0^T |\kappa_{x,m}(l)| dl$ recognizing in the expression above the associated Riemann sum. Consequently, for N sufficiently large the X_k will be approximately jointly Gaussian distributed.

The case where the measurements are dependent is considered now. This derivation is obtained following [7] where the x_n are modeled as the output of a linear filter driven by an independent sequence e_n that is of course assumed nonstationary herein. The coefficients of the filter are assumed to be dependent of the sampling period T_e in order to prevent that as T_e tends to zero, the transfer function $H(w)$ of the filter tends to $H(0)$. The impulse response of the filter is denoted h_{n,T_e}. Taking into account the finite support of the considered signals, x_n is:

$$x_n = \sum_{v=0}^{n} h_{v,T_e} e_{n-v} \quad (8)$$

Denoting for simplicity $W_N = \exp(-2j\pi k T_e)$ and $H_k = \sum_{v=0}^{N-1} h_{v,T_e} W_N^v$, the DFT of x_n becomes:

$$\begin{aligned} X_k &= \sqrt{T_e} \sum_{n=0}^{N-1} \sum_{v=0}^{n} h_{v,T_e} e_{n-v} W_N^{n-v} W_N^v \\ &= \sqrt{T_e} \sum_{v=0}^{N-1} h_{v,T_e} W_N^v \sum_{n=v}^{N-1} e_{n-v} W_N^{n-v} \\ &= \sqrt{T_e} \sum_{v=0}^{N-1} h_{v,T_e} W_N^v \left[E_k - \sum_{n=N-v}^{N-1} e_n W_N^n \right] \\ &= H_k E_k - \rho_k \end{aligned} \quad (9)$$

where:

$$\rho_k = \sqrt{T_e} \sum_{v=0}^{N-1} h_{v,T_e} W_N^v \sum_{n=N-v}^{N-1} e_n W_N^n \quad (10)$$

$$= \sqrt{T_e} \sum_{p=1}^{N-1} e_p W_N^p \sum_{q=N-p}^{N-1} h_{q,T_e} W_N^q \quad (11)$$

As a consequence, the mean and variance of ρ_k verify:

$$\mathsf{E}[\rho_k] = \sqrt{\frac{T_e}{N}} \sum_{p=1}^{N-1} \kappa_{e,1}(p) W_N^p \left(\sqrt{N} \sum_{q=N-p}^{N-1} h_{q,T_e} W_N^q \right) \quad (12)$$

$$\mathsf{var}[\rho_k] \le T_e \sum_{p=1}^{N-1} \kappa_{e,2}(p) \left(\sum_{q=N-p}^{N-1} h^2_{q,T_e} \right) \quad (13)$$

If the second sum in (12) and (13) tends to zero when N tends to infinity, which is true if $h_{q,T_e} W_N^q$ and h^2_{q,T_e} are abslolutely summable sequences, the dominated convergence theorem proves that $\lim_{N\to\infty} \mathsf{E}[\rho_k] = 0$ and $\lim_{N\to\infty} \mathsf{var}[\rho_k] = 0$. Consequently, equation (9) decomposes X_k in a sum of an asymptotically Gaussian term and a term that tends to 0 in the mean square sense. This implies that the DFT of the correlated signal x_n follows an asymptotically jointly Gaussian law.

A straightforward computation of $\mathsf{E}[X_k]$ and $\mathsf{E}[X_k X_q^*]$ from (5) shows that:

$$\lim_{N\to\infty} (\sqrt{T_e})\mathsf{E}[X_k] = m(k;\boldsymbol{\theta}) \quad (14)$$

$$\lim_{N\to\infty} T_e \mathsf{E}[X_k X_q^*] = S(k,q;\boldsymbol{\theta}) \quad (15)$$

where

$$m(f;\boldsymbol{\theta}) = \int_0^T \mathsf{E}[x(l)] \exp(-j2\pi f l) dl \quad (16)$$

$$S(f_1,f_2;\boldsymbol{\theta}) = \int_{-T}^{T} \left(\int_0^{T-|\tau|} \mathsf{E}[x(l)x(l+|\tau|)] \exp\{-j2\pi(f_1 l - f_2(l+\tau))\} dl \right) d\tau \quad (17)$$

Consequently the asymptotic distribution of the vector of the DFT components $\boldsymbol{X} = (\sqrt{T_e})(X_1,\ldots,X_N)^t$ is:

$$\boldsymbol{X} \sim \mathcal{N}(\mu(\boldsymbol{\theta}), \Sigma(\boldsymbol{\theta}) - \mu(\boldsymbol{\theta})\mu(\boldsymbol{\theta})^t) \quad (18)$$

where $\Sigma(\boldsymbol{\theta})_{k,q} = S(k,q;\boldsymbol{\theta})$, with $k,q = \{0,\ldots,N-1\}$, and $\mu(\boldsymbol{\theta})_k = m(k;\boldsymbol{\theta})$.

This result suggests several remarks. First, noticing that $S(f,f;\boldsymbol{\theta}) = \tilde{S}(f;\boldsymbol{\theta})$, the asymptotic variance of $(\sqrt{T_e})X_k$ equals $\tilde{S}(k;\boldsymbol{\theta})$. This result can be compared with the one obtained in the stationary case, where the variance of the DFT asymptotically equals the power spectrum, [1].

Then, if the signal is second order stationary $\mathsf{E}[x(l)x(l+|\tau|)]$ can be substituted by $c(\tau)$ in (3). If we now consider a signal observed on $[0,\infty]$, the limit of (3) can be computed when $T \to \infty$. It is of course important to take the precaution of dividing this quantity by T before the computation of the limit, the signal under scope becoming as T tends to ∞ a finite power signal. In this case a simple computation shows that:

$$\lim_{T\to\infty} \frac{1}{T} S(f_1,f_2;\boldsymbol{\theta}) = \delta(f_1-f_2) \int_{-\infty}^{+\infty} c(\tau) e^{-j2\pi f_1 \tau} d\tau.$$

This leads to the classical result: if x_n is a stationary signal, under mild assumptions the vector $\boldsymbol{X}$ is asymptotically Gaussian distributed and the X_k are independent with a variance that equals the power spectrum of the signal.

Finally, if X_k are asymptotically independent:

$$S(f_1,f_2;\boldsymbol{\theta}) = m(f_1;\boldsymbol{\theta})m(f_2;\boldsymbol{\theta}) \text{ for } f_1 \neq f_2. \quad (19)$$

Then the asymptotic distribution of $\boldsymbol{X}$ is:

$$\boldsymbol{X} \sim \mathcal{N}(\mu(\boldsymbol{\theta}), \mathrm{Diag}[\{\tilde{S}(k;\boldsymbol{\theta}) - m(k;\boldsymbol{\theta})^2\}_{k=0\ldots N-1}]).$$

This assumption presents a significant interest for the establishment of the distribution of the periodogram defined herein as $I_k = (T_e)|X_k|^2$. Indeed, since $\boldsymbol{X}$ is decorrelated, each component I_k of the periodogram follows a χ_2^2 distribution and the joint distribution of vector $\boldsymbol{I}$ is a product of χ_2^2 distributions.

4 COST FUNCTION FOR PARAMETERS ESTIMATION

The purpose of this section is to derive a cost function for the estimation of $\boldsymbol{\theta}$ from the asymptotic distribution of $\boldsymbol{X}$. To preserve a general aspect of this study, the dependence of the X_k will be considered. In this case the distribution of $\boldsymbol{I}$ is very complicated to derive and moreover the correlation between the DFT bins suggests to study a statistic that also includes their products. Let us assume that M independent realizations of the signal x_n are available and that the vector $\mu(\boldsymbol{\theta})$ is known (otherwise $\mu(\boldsymbol{\theta})$ will be substituted by the empirical mean obtained from the M DFT $\boldsymbol{X}^{(m)}$, $m = 1\ldots M$). Let the covariance matrix $\hat{\boldsymbol{C}}_M$ be defined as:

$$\hat{\boldsymbol{C}}_M = \frac{1}{M}\sum_{m=1}^{M} (\boldsymbol{X}^{(m)} - \mu(\boldsymbol{\theta}))(\boldsymbol{X}^{(m)} - \mu(\boldsymbol{\theta}))^H. \quad (20)$$

When N is large, the asymptotic distribution of the $\boldsymbol{X}^{(m)}$ implies that $\hat{\boldsymbol{C}}_M$ follows a complex Wishart distribution of dimension N and degrees of freedom M [1]:

$$\hat{\boldsymbol{C}}_M \sim \mathcal{W}_N^c\left(M, \frac{1}{M}\boldsymbol{\Sigma}(\boldsymbol{\theta})\right).$$

The log-likelihood of $\hat{\boldsymbol{C}}_M$ will be used as a cost function for the estimation of $\boldsymbol{\theta}$. If the terms that are not function of $\boldsymbol{\theta}$ are suppressed the expression for this cost function reduces to:

$$\mathcal{C}(\boldsymbol{\theta}) = -\log(|\boldsymbol{\Sigma}(\boldsymbol{\theta})|) - \mathrm{tr}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\hat{\boldsymbol{C}}_M\right) \quad (21)$$

$\mathcal{C}(\boldsymbol{\theta})$ takes a simplified form when the X_k are independent, since $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a diagonal matrix with $\boldsymbol{\Sigma}(\boldsymbol{\theta})_{k,k} = \tilde{S}(k;\boldsymbol{\theta})$, and thus relation (21) becomes:

$$\mathcal{C}(\boldsymbol{\theta}) = -\sum_{k=0}^{N-1}\left[\log\{\tilde{S}(k;\boldsymbol{\theta})\} + \frac{1}{M}\frac{\sum_{m=1}^{M} I_k^{(m)}}{\tilde{S}(k;\boldsymbol{\theta})}\right] \quad (22)$$

This expression of the cost function has exactly the same structure as the approximately Gaussian log-likelihood introduced by P. Whittle [9, 2] for the stationary processes where the power spectrum has been replaced by $\tilde{S}(k;\boldsymbol{\theta})$.

5 APPLICATION TO STELLAR SPECKLE INTERFEROMETRY

The cost function (22) has been applied to the problem of visibility estimation in stellar speckle interferometry.

When a stellar source is observed at the focus of a large telescope the phase of the wave coming from the object is disturbed by the atmospheric turbulence, so the image $x(l)$ must be considered as a random function. The technique of stellar speckle interferometry proposed by Labeyrie [5], based on the computation of the second order statistics $\tilde{S}(\boldsymbol{f})$, enables to freeze the turbulence during the measurement.

The acquired image equals the object convolved with the optical transfer function characterizing both the telescope and the atmospheric turbulence. The computation of the energy spectrum of this image gives, [6]:

$$\tilde{S}(\boldsymbol{f};\boldsymbol{\theta}) = |O(\mathbf{0})|^2 B^2(\boldsymbol{f}) + \frac{\sigma_c}{2s}\Bigg(|O(\mathbf{0})|^2 T_o(\boldsymbol{f}) + \frac{|O(\boldsymbol{f_0})|^2}{2}\left(T_o(\boldsymbol{f}+\boldsymbol{f_0}) + T_o(f-\boldsymbol{f_0})\right)\Bigg) \quad (23)$$

$O(\boldsymbol{f_0})$ is the Fourier Transform of the object at the spatial frequency $\boldsymbol{f_0}$ corresponding to the distance between the two telescopes of the interferometer. $T_0(\boldsymbol{f})$ is the optical transfer function of the telescope (normalized autocorrelation of one telescope of aperture area s). $B(\boldsymbol{f}) = B_\psi(\boldsymbol{f})T_0(\boldsymbol{f})$ where $B_\psi(\boldsymbol{f})$, the second-order moment of the wavefront, characterizes the effect of the turbulence on the incident wave. $\sigma_c = \int B_\psi^2(\boldsymbol{f})d\boldsymbol{f}$ is the coherence area of the wavefront.

The energy spectrum $\tilde{S}(\boldsymbol{f};\boldsymbol{\theta})$ is then decomposed in two low frequency contributions: $|O(\mathbf{0})|^2 B^2(\boldsymbol{f})$ called the seeing peak and $|O(\mathbf{0})|^2(\sigma_c/2s)T_o(\boldsymbol{f})$ called the speckle peak, and one high frequency contribution called the fringes peaks: $|O(\boldsymbol{f_0})|^2(\sigma_c/4s)T_o(\boldsymbol{f} \pm \boldsymbol{f_0})$.

$\boldsymbol{\theta} = [|O(\mathbf{0})|^2, |O(\boldsymbol{f_0})|^2]^t$ is the vector parameter to estimate in order to compute the squared fringes visibility $V^2(\boldsymbol{f_0}) = |O(\boldsymbol{f_0})|^2/|O(\mathbf{0})|^2$. $V(\boldsymbol{f_0})$ is generally estimated by carrying out the ratio between the estimated fringes peak energy and the speckle peak energy computed from locally averaged periodogram bins. This estimator will be denoted as "energy ratio" in the sequel. This simulation proposes to compare this estimator with the estimator obtained maximizing (22).

The interferometric images have been generated using two circular telescopes with a diameter of $5\,m$ separated by a $15\,m$ baseline. The atmospheric turbulence is modelled by $B_\psi(f) = \exp(-3.44(\lambda|\boldsymbol{f}|/r_0)^{5/3})$ with $r_0 = 2\,cm$. The simulations consider an unresolved source (delta function), i.e. $\boldsymbol{\theta} = [1,1]^t$. The values of $\boldsymbol{\theta}$ have been estimated using series of $M = 50$ images each one with size $N = 256 \times 256$. An example of the shape of the cost function $\mathcal{C}(\boldsymbol{\theta})$ is given in Fig. 1.

For each serie the visibility has been estimated as $\widehat{V^2(\boldsymbol{f_0})} = |\widehat{O(\boldsymbol{f_0})}|^2/|\widehat{O(\mathbf{0})}|^2$. 40 values of these parameters have been estimated from independent series of image realizations using the proposed method and the "energy ratio" method. The minimization of $\mathcal{C}(\boldsymbol{\theta})$ has been performed using the `Matlab` routine `fminsearch` initialized at $\{0.5, 0.5\}$. The estimated bias and the mean squared error (MSE) are given in table 1. These results clearly prove the efficiency of the parametric method proposed in this paper.

6 CONCLUSION

This paper studied the problem of parametric estimation of transient signals based on energy spectrum models. A criterion similar to the Whittle likelihood for the stationary case has been derived. The interest of the method have been validated by computer simulations. The extension of this approach to the computation of lower bounds on the estimated parameters variance is currently under investigation.

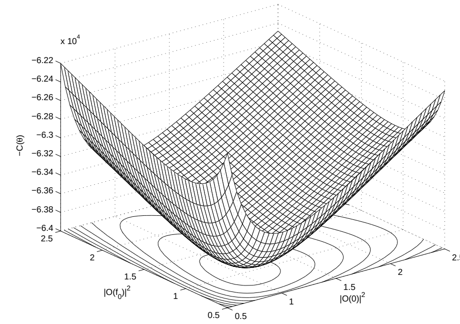

Figure 1: $-\mathcal{C}(\boldsymbol{\theta})$ as a function of θ_1 and θ_2. The thrue value is $\{1, 1\}$.

	Estimated bias		
	$\widehat{\lvert O(\mathbf{0})\rvert^2}$	$\widehat{\lvert O(\boldsymbol{f_0})\rvert^2}$	$\widehat{C(\boldsymbol{f_0})}$
Max. of $\mathcal{C}(\boldsymbol{\theta})$	34.10^{-4}	50.10^{-4}	18.10^{-4}
Energy ratio	62.10^{-3}	-63.10^{-3}	12.10^{-2}

	Estimated MSE variation		
	$\widehat{\lvert O(\mathbf{0})\rvert^2}$	$\widehat{\lvert O(\boldsymbol{f_0})\rvert^2}$	$\widehat{C(\boldsymbol{f_0})}$
Max. of $\mathcal{C}(\boldsymbol{\theta})$	13.10^{-5}	23.10^{-5}	35.10^{-5}
Energy ratio	22.10^{-2}	28.10^{-2}	14.10^{-3}

Table 1: Bias and MSE for the proposed method and "energy ratio" method.

References

[1] D. R. Brillinger. *Times Series: Data Analysis and Theory.* McGraw-Hill, 1981.

[2] K. Dzhaparidze. *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer Series in Statistics, 1985.

[3] G. Fay. *Non-linear Functionals of the Periodogram.* PhD thesis, Ecole Nationale Supérieure des Télécommunications, January 2000.

[4] S. M. KAY. *Fundamentals of Statistical Signal Processing - Estimation Theory.* Prentice Hall International Editions, 1993.

[5] A. Labeyrie. *Progress in Optics*, volume 14, pages 49–87. E Wolf, 1976.

[6] F. Roddier. *Progress in Optics*, volume XIX. E. Wolf, 1981.

[7] P. Stoica and R. Moses. *Introduction to Spectral Analysis.* Prentice Hall, 1997.

[8] M. Taqqu and V. Teverovsky. Whittle estimator for non-Gaussian long-memory time serie. *Stochastic Models*, 13:723–757, 1997.

[9] P. Whittle. Gaussian estimation in stationary time series. *Bulletin for the International Institute of Statistics*, 39:105–129, 1962.