Spectral Modes Estimation using Whittle-type Criterion. Application to Stellar Speckle Interferometry

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

ABSTRACT

This communication adresses the parameter estimation problem using the minimization of a Whittle-type criterion when the spectrum parametrization is linear versus the unknown parameters. In this case, the global convergence of a particular iterative minimization algorithm is proved. This estimation method applies to stellar speckle interferometry for the visibility estimation problem. Simulations show the efficiency of the proposed method.

1 Introduction

The use of Whittle-type [12, 5] criterion for parameters estimation has recently received an increased interest. This approach allows to simply cope with the situations where only a parametric model of the power spectrum is available, or where a spectral model is more tractable than the temporal model. It is the case for time series with long range dependence, [11]. Extension of the criterion to non stationary signals has been proposed in [4]. In [6] this criterion has been applied to the transient signals. For this, the power spectrum is simply replaced by the continuous signal energy spectrum and the convergence is expected for an increased number of samples on the signal support.

In many situations the parametrization of the spectrum is linear versus the unknown parameters, as when the spectrum is compounded by modes with a known shape but unknown amplitudes. The estimation problem then reduces to the minimization of a particular multivariate function under positivity constraints. This communication adresses the minimization problem. It proposes to perform the minimization using an iterative algorithm recently devised in [3] in the image restoration field. The main result consists in the proof of the global convergence of this algorithm.

Application of the method to the visibility estimation in stellar interferometry is presented. In this case, the phase of the wave coming from the object is disturbed by the atmospheric turbulence and the use of Kolmogorov's law allows to obtain solely a parametric model of the energy spectrum of the interferometric image, [10]. The spectrum model shows that the Whittle estimator extended to finite support signal together with the iterative algorithm is particularly adapted to this problem.

2 Estimation technique

2.1 Problem statement and general formalism

Let x_n, $n = \{0 \dots N-1\}$ be a zero mean time serie which depends on an unknown vector parameter $\boldsymbol{\theta}$. The Whittle type estimator of $\boldsymbol{\theta}$ is the value $\widehat{\boldsymbol{\theta}}$ which minimizes the cost function:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} \log(S(k;\boldsymbol{\theta})) + \frac{I_k}{S(k;\boldsymbol{\theta})}. \tag{1}$$

The structure of this cost function encloses a wide range of estimators.

- if x_n is stationary, $S(k;\boldsymbol{\theta})$ denotes its power spectrum at frequency k/N and I_k the periodogram of the signal at the same frequency, $\widehat{\boldsymbol{\theta}}$ is the standard Whittle estimator, [12]. The same criterion structure arises for the aggregated [11] or tapered Whittle estimator.

- if $x_n = x(nT/N)$ where $x(t)$ is a continuous time transient random signal defined on the interval $[0, T]$, $S(k;\boldsymbol{\theta})$ is the energy spectrum of $x(t)$ at frequency k/T and I_k the periodogram of x_n, $\mathcal{L}(\boldsymbol{\theta})$ is the extension of the Whittle likelihood to transient signal parameters estimation recently proposed in [6]. Application of this criterion to visibility estimation in stellar speckle interferometry is presented in section 3.

In the sequel, the problem will be reduced to the case where the parametrization of $S(k;\boldsymbol{\theta})$ is linear versus $\boldsymbol{\theta}$, i.e., the spectrum $S(k;\boldsymbol{\theta})$ is composed by M modes $S_q(k)$ having a known shape but with an unknown amplitude θ_q. In order to ensure the positivity of the estimated spectrum the constraint $\forall q\ \theta_q \geq 0$ must be added to the minimization of $\mathcal{L}(\boldsymbol{\theta})$:

$$\widehat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}\in\mathbb{R}^{+M}} \sum_{k=0}^{N-1} \log\left(\sum_{q=1}^{M} \theta_q S_q(k)\right) + \frac{I_k}{\sum_{q=1}^{M} \theta_q S_q(k)}. \tag{2}$$

2.2 Optimization algorithm

An iterative algorithm for the minimization of a criterion analogue to (2) has been recently proposed in [3] in the image reconstruction context. The criterion aroused from the minimization of the cross Burg Entropy (or Itakura-Saito distance) between the received image and the convolution between a known point spread function and the restored image. Denoting S the $N \times M$ regression matrix associated to the $S_q(k)$, the iteration is:

$$\theta_j^{[k+1]} = \theta_j^{[k]} \left[\mathsf{S}^t \boldsymbol{p}^{[k]}\right]_j / \left[\mathsf{S}^t \boldsymbol{q}^{[k]}\right]_j \tag{3}$$

$$\text{with}: \quad \boldsymbol{p}_\ell^{[k]} = I_\ell / (\left[\mathsf{S}\boldsymbol{\theta}^{[k]}\right]_\ell)^2, \; \boldsymbol{q}_\ell^{[k]} = 1/\left[\mathsf{S}\boldsymbol{\theta}^{[k]}\right]_\ell \tag{4}$$

It was obtained setting to zero the lower bound of the difference $\mathcal{L}(\boldsymbol{\theta}^{[k]}) - \mathcal{L}(\boldsymbol{\theta}^{[k+1]})$ and consequently ensures that the cost will never increase. Moreover, it can be easily check from (3,4) that the iteration maps $\mathbb{R}^{+M}$ inside $\mathbb{R}^{+M}$ and then verifies the non-negativity constraint.

A major problem is the study of the global convergence of (3, 4). It was investigated in [3] only by computer simulations where global convergence was always observed. This section is devoted to the proof of the global convergence for the application under scope.

First, let us establish that $\hat{\boldsymbol{\theta}}$ is in the interior of $\mathbb{R}^{+M}$. The second order partial derivatives of $\mathcal{L}(\boldsymbol{\theta})$ are:

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial\theta_p \partial\theta_q} = \sum_{k=0}^{N-1} S_p(k) S_q(k) \frac{2I_k - S(k;\boldsymbol{\theta})}{S(k;\boldsymbol{\theta})^3}, \tag{5}$$

and the Hessian of $\mathcal{L}(\boldsymbol{\theta})$ is:

$$H_{\boldsymbol{\theta}} = \mathsf{S}^t \mathrm{Diag}\left\{\frac{2I_k - S(k;\boldsymbol{\theta})}{S(k;\boldsymbol{\theta})^3}\right\} \mathsf{S} = H_{ij}\big|_{i,j=1,\ldots,M} \tag{6}$$

$$\text{with} \quad H_{ij} = \sum_{k=0}^{N-1} \frac{[2I_k - S(k;\boldsymbol{\theta})] S_i(k) S_j(k)}{S(k;\boldsymbol{\theta})^3}. \tag{7}$$

We denote by $\mathcal{C}$ the set where the criterion $\mathcal{L}(\boldsymbol{\theta})$ is convex. The Kuhn-Tucker conditions, [9], with $\boldsymbol{\theta} \geq \mathbf{0}$ are:

$$\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{\theta}^*) - \boldsymbol{\mu}^t = \mathbf{0} \tag{8}$$

$$\boldsymbol{\mu}^t \boldsymbol{\theta}^* = 0 \tag{9}$$

where $\boldsymbol{\theta}^*$ is a relative minimum point and $\boldsymbol{\mu} \in \mathbb{R}^{+M}$. These conditions mean that if the global minimum of $\mathcal{L}(\boldsymbol{\theta})$ is not in the interior of $\mathbb{R}^{+M}$ each components of the solution of (2) can be in the interior of $\mathcal{C}$ or on the boundaries of $\mathbb{R}^{+M}$. This last case, i.e. $\exists q$, $\hat{\theta}_q = 0$, clearly suggest that the number of modes as been overestimated and the estimation should be derived again substituting M by $M-1$. Consequently, we will only consider in the sequel the case where $\widehat{\boldsymbol{\theta}}$ is inside $\mathcal{C}$. This assumption is of course valid as long as $\mathcal{C} \cap \mathbb{R}^{+M}$ is not empty and does not reduce to the boundaries of $\mathbb{R}^{+M}$.

Notice that, as long as $N > M$, $\mathcal{C}$ strictly includes the polyhedral set $\mathcal{S} = \{\boldsymbol{\theta} \,|\, \forall k \quad S(k;\boldsymbol{\theta}) \leq 2I_k\}$. Under mild assumptions on the signal, the I_k are asymptotically (N large) independent and exponentially distributed with a variance that equals twice the spectrum at the same frequency. See for example [2] for the stationary case. Consequently for large N, $\forall k$, $I_k \neq 0$. It is then possible to find a vector $\boldsymbol{\theta}$ having identical components that belongs to $\mathcal{S} \cap \mathbb{R}^{+M}$ and that is not on the boundaries of $\mathbb{R}^{+M}$:

$$\forall q = 1 \ldots M, \quad \theta_q = \min_{k,k=0\ldots N-1} \frac{2I_k}{\sum_{q=1}^{M} S_q(k)}. \tag{10}$$

However, given the true parameter $\boldsymbol{\theta}^\circ$, the probability that the I_k's are included in $\mathcal{S}$ for N large is:

$$p(\boldsymbol{\theta} \in \mathcal{S}) = \exp\left(-\frac{1}{2}\sum_{k=0}^{N-1} \frac{S(k;\boldsymbol{\theta}^\circ)}{S(k;\boldsymbol{\theta})}\right). \tag{11}$$

This proves that, as long as $S(k;\boldsymbol{\theta}^\circ)/S(k;\boldsymbol{\theta})$ is bounded from below, $\hat{\boldsymbol{\theta}}$ will certainly be outside $\mathcal{S}$ when N increases.

The last point is the global convergence of (3,4). In [8] the authors propose a general framework for the derivation of minimization algorithms with non-negativity constraints. It relies on the use of a steepest descent method to solve the Kuhn-Tucker first-order optimality conditions, [9]. The application of this method to (2) directly leads to the iterations (3,4) when the step size is set equal 1. It is worthy to note that [3] proves that the choice of this step size value will guaranty that the cost will not increase. Consequently if $\boldsymbol{\theta}^{[k]}$ converges to a value $\boldsymbol{\theta}^*$ and that $H_{\boldsymbol{\theta}^*}$ is positive defined, $\boldsymbol{\theta}^*$ can be a local minimum.

The proof that $\boldsymbol{\theta}^*$ is the global minimum relies on the quasiconvex property of $\mathcal{L}(\boldsymbol{\theta})$, [1], i.e. for $\alpha \geq 0$, $\beta \geq 0$:

$$\mathcal{L}(\alpha\boldsymbol{\theta}^1 + \beta\boldsymbol{\theta}^2) \leq \max[\mathcal{L}(\boldsymbol{\theta}^1), \mathcal{L}(\boldsymbol{\theta}^2)], \quad \alpha + \beta = 1. \tag{12}$$

The demonstration of this result is straightforward writing $\mathcal{L}(\boldsymbol{\theta}) = \phi(\boldsymbol{f}(\boldsymbol{\theta}))$ where:

$$\phi(\boldsymbol{\mu}) = \sum_{k=0}^{N-1} \log(\mu_k) + \frac{I_k}{\mu_k} \tag{13}$$

is a real-valued function on $\mathbb{R}^{+M}$ and $\boldsymbol{f}(\boldsymbol{\theta})$ are N linear functions of $\boldsymbol{\theta}$. The quasiconvexity of $\phi(\boldsymbol{\mu})$ is a direct consequence of the quasiconvexity of $\log(x) + 1/x$. Since $\boldsymbol{f}(\boldsymbol{\theta})$ are linear functions defined on the convex set $\mathbb{R}^{+N}$, the quasiconvexity of $\phi(\boldsymbol{\theta})$ implies the quasiconvexity of $\mathcal{L}(\boldsymbol{\theta})$, [1, theorem 6.9]. In addition, the level sets of a quasiconvex real-valued function defined on a convex set $X \subset \mathbb{R}^{+M}$ are convex for every $\gamma \in \mathbb{R}^+$, [1, theorem 6.1]. This shows that $\boldsymbol{\theta}^*$ is also the global minimum of $\mathcal{L}(\boldsymbol{\theta})$ on $\mathbb{R}^{+M}$, under the assumption that $H_{\boldsymbol{\theta}^*}$ is positive defined.

3 Stellar speckle interferometry

3.1 Spectrum model

The cost function (1) has been applied to the problem of visibility estimation in stellar speckle interferometry. An interferometer (composed of a pupil containing 2 telescopes in our case) samples the Fourier transform of the brightness distribution of a source through measurements of fringes visibility. With a sufficiently large number of spatial frequencies $\boldsymbol{f}$, several informations about the observed source, like its angular diameter, can be reconstructed.

However, 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. To improve the measurement, Labeyrie proposed the technique of stellar speckle interferometry [7], based on the computation of a sequence of short exposure images. The short exposure enables to freeze the turbulence during the measurement, but presents the main drawback to reduce considerably the signal-to-noise ratio. It is thus necessary to acquire a large number of interferometric images.

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 $S(\boldsymbol{f};\boldsymbol{\theta})$ of this image, defined as

$$S(\boldsymbol{f}) = \mathsf{E}\left[\left|\int x(\boldsymbol{\ell})e^{-j2\pi\boldsymbol{f}\boldsymbol{\ell}}d\boldsymbol{\ell}\right|^2\right] \qquad (14)$$

requires an expression of the fourth-order moment of the complex amplitude $\psi(\boldsymbol{u})$ of the randomly distorted incoming wave front. Assuming that $\psi(\boldsymbol{u})$ is a complex Gaussian process, $S(\boldsymbol{f})$ is given by [10]:

$$S(\boldsymbol{f}) = |O(\mathbf{0})|^2 B^2(\boldsymbol{f}) + \frac{\sigma_c}{2s}\Bigg(|O(\mathbf{0})|^2 T_o(\boldsymbol{f}) + \frac{|O(\boldsymbol{f_0})|^2}{2}(T_o(\boldsymbol{f}+\boldsymbol{f_0}) + T_o(f-\boldsymbol{f_0}))\Bigg) \qquad (15)$$

where:

- $O(\boldsymbol{f_0})$ is the Fourier Transform of the object at the spatial frequency $\boldsymbol{f_0}$ corresponding to the distance between the two telescopes. The vector parameter to estimate is $\boldsymbol{\theta} = [|O(\mathbf{0})|^2, |O(\boldsymbol{f_0})|^2]^t$.

- $T_0(\boldsymbol{f})$, the optical transfer function of the telescope, equals to the normalized autocorrelation of one telescope. Considering a circular telescope of radius r and neglecting the central obstruction:

$$T_0(\boldsymbol{f}) = \frac{2}{\pi}\left[\arccos\left(\frac{\lambda|\boldsymbol{f}|}{2r}\right) - \frac{\lambda|\boldsymbol{f}|}{2r}\left(1-\frac{\lambda^2|\boldsymbol{f}|^2}{4r^2}\right)^{1/2}\right]$$

- $B(\boldsymbol{f}) = B_\psi(\boldsymbol{f})T_0(\boldsymbol{f})$ where $B_\psi(\boldsymbol{f})$ is the second-order moment of $\psi(\boldsymbol{u})$. According to the Kolmogorov turbulence model:

$$B_\psi(\boldsymbol{f}) = \exp(-3.44(\lambda|\boldsymbol{f}|/r_0)^{5/3}) \qquad (16)$$

  where r_0 is the seeing parameter. r_0 will be assumed known herein.

- $\sigma_c = \int B_\psi^2(\boldsymbol{f})d\boldsymbol{f} = 0.342r_o^2$ and s is the aperture area of a single telescope.

The energy spectrum $S(\boldsymbol{f};\boldsymbol{\theta})$ is then decomposed in:

- two low frequency contributions: $|O(\mathbf{0})|^2B^2(\boldsymbol{f})$ called the seeing peak and $|O(\mathbf{0})|^2(\sigma_c/2s)T_o(\boldsymbol{f})$ called the speckle peak,

- one high frequency contribution called the fringes peaks: $|O(\boldsymbol{f_0})|^2(\sigma_c/4s)T_o(\boldsymbol{f}\pm\boldsymbol{f_0})$.

The aim of this estimation problem is to compute the fringes visibility $V(\boldsymbol{f_0})$ (or the contrast $C(\boldsymbol{f_0})$):

$$V(\boldsymbol{f_0}) = \sqrt{C(\boldsymbol{f_0})} = \frac{|O(\boldsymbol{f_0})|}{|O(\mathbf{0})|}.$$

$V(\boldsymbol{f_0})$ is generally estimated by carrying out the ratio between the estimated fringes peaks energy computed from locally averaged periodogram bins and the speckle peak energy. This last is estimated fitting by linear least squares $T_0(\boldsymbol{f})$ on periodogram bins at $|\boldsymbol{f}| > 2r/\lambda$, where the contribution of the seeing peak is negligible. This estimator will be denoted as "energy ratio" in the sequel.

In [6], it has been proved that the cost function (1) can be used to estimate the parameters of a continuous time signal with finite support $[0, T]$ and energy spectrum $S(\boldsymbol{f})$ sampled at $T_s = T/N$ when N is large. For this purpose, $S(k;\theta)$ in (1) denotes $S(\boldsymbol{k}/NT_s)$ and I_k the averaged periodogram of short exposure interferometric images. This simulation proposes to compare the estimation of $V(\boldsymbol{f}_0)$ obtained maximizing (1) using the iterations (3,4) to the energy ratio estimator.

The interferometric images have been generated using two circular telescopes (without central obstruction) with a diameter of 5 m separated by a 20 m baseline. The phase $\phi(\boldsymbol{u})$ of the complex amplitude of the disturbed incident wave front is generated filtering a normally distributed independent phase screen by the filter:

$$H_\phi(\boldsymbol{f}) = 0.151r_0^{-5/6}|\boldsymbol{f}|^{-11/6}. \qquad (17)$$

The interferometric image $x(\boldsymbol{\ell})$ is then the squared Fourier transform of $\psi(\boldsymbol{u})$ masked by the pupil. The seeing parameter r_0 will take values in the interval $[5\,cm, 20\,cm]$ for a parametric study of its effects.

Simulations consider an unresolved source (delta function), i.e. $|O(\mathbf{0})|^2 = |O(\boldsymbol{f_0})|^2 = 1$. Values of $|O(\mathbf{0})|^2$, $|O(\boldsymbol{f_0})|^2$ have been estimated using series of 20 short exposure images each one with size $N = 800 \times 800$.

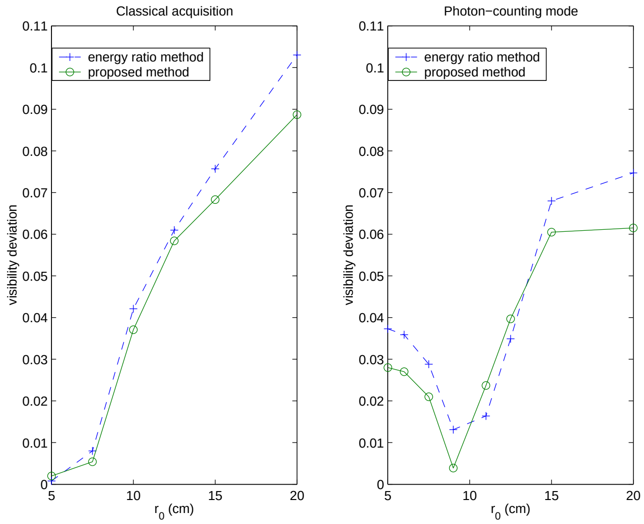

Figure 1: Estimated bias for the classical image's acquisition and in the photon-counting mode.

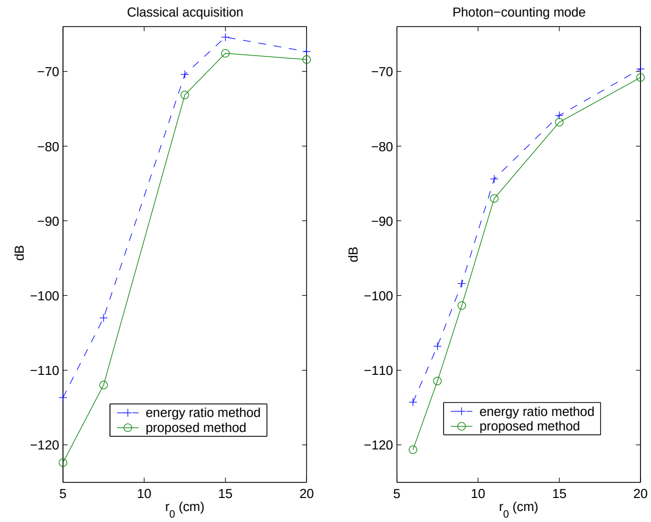

Figure 2: Estimated variance for the classical image's acquisition and in the photon-counting mode.

For each serie the visibility has been estimated as $|\widehat{O(\boldsymbol{f_0})}|/|\widehat{O(\mathbf{0})}|$. 50 values of these parameters have been estimated from independent series using the proposed method and the "energy ratio" method. Two images acquisition procedure are distinguished in the simulations: the classical acquisition with a CCD camera, where a large photons flux is available for the measurements but with a poor resolution; the photon-counting mode where each pixel follows a Poisson distribution conditionally to the classical mode. In this case, the average number of photons by image is fixed to 10^5.

4 Simulation results and conclusion

The algorithm (3,4) has been initialized at {10,10}. The estimated bias (Fig.1) and variance (Fig.2) for the two methods in the two modes of images acquisition are given as a function of the seeing parameter r_0. The stopping criterion of the iterations is $|\mathcal{L}(\boldsymbol{\theta}^{[k+1]}) - \mathcal{L}(\boldsymbol{\theta})^{[k]}| < 10^{-5}$. Simulation have shown that optimization algorithm (3,4) is about four times faster than the `Matlab` routine `fminsearch` used for the simulations in [6], with the same termination tolerance for $\mathcal{L}(\boldsymbol{\theta})$.

These results clearly prove the efficiency of a parametric method using the spectrum model as an *a priori* information. In both situations, large or weak photons flux, and for typical values of r_0, bias and variance estimation shows a higher precision compared to the traditionnal method. The advantages of this technique can be applied to a wide range of poblems as long as the parametrization of the spectrum is linear.

References

[1] M. Avriel. *Non-linear Programming - Analysis and Methods*. Prentice Hall, 1976.

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

[3] Y. Cao, P. Eggermont, and S. Terebey. Cross Burg entropy maximization and its application to ringing suppression in image reconstruction. *IEEE Transactions on Image Processing*, 8(2):286–292, February 1999.

[4] R. Dahlhaus. Fitting time series models to nonstationary processes. *Annals of Statistics*, 25:1–37, 1997.

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

[6] A. Ferrari and Y. Cavallin. Extension of Whittle likelihood to transient signal parameters estimation. Application to stellar speckle interferometry. In *EUSIPCO*, 2002.

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

[8] H. lantéri, M. Roche, O. Cuevas, and C. Aime. A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints. *Signal Processing*, 81:945–974, 2001.

[9] D. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 1989.

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

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

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