SPECTRUM ANALYSIS WITH RECURSIVE LEAST SQUARES
ESTIMATES ON AMPLIFIED G-HARMONICS

Benoît TRUONG-VAN
Département de Mathématiques - CNRS - UA 1204 - Université de Pau
Avenue de l'Université, 64000 PAU, FRANCE.

RESUME

Nous proposons une nouvelle méthode efficace pour obtenir un estimateur convergent du spectre mixte de signaux pouvant être représentés par un modèle non-harmonique général. Cette méthode est fondée sur une régression récursive sur des G-harmoniques amplifiés. Nous en donnons quelques résultats d'application.

MOTS CLES : estimateur convergent, modèle non-harmonique, ARMA.

SUMMARY

We present a computationally efficient new method to obtain consistent estimate of the mixed spectrum of signals that are represented by a general non-harmonic model. This technique is based upon recursive least squares on amplified G-harmonics. A description of the method and some numerical results are given.

KEYWORDS : Consistent estimate, non-harmonic model, ARMA.

1 - Introduction

The problem considered herein is the estimation of the spectrum of a signal \( x_t \) that can be additively decomposed into an oscillatory component \( H_t \) and an "irregular" component \( \eta_t \) from part of a realization of the signal, say \( x_t, \ t = 1, \ldots, N \).

We assume that the \( H_t \) component can be written as
\[
H_t = \sum_{j=1}^{m} \sum_{j}^{} c_j \cos (\omega_j t + \phi_j) (1)
\]
where the amplitudes \( c_j \), the phases \( \phi_j \), the angular frequencies or pulsations \( \omega_j \) and the number \( m \) are unknown, with the \( \omega_j \)'s being arbitrary in \( (0, \pi) \). The "irregular" part is assumed to be a stationary purely non-deterministic ARMA process, that we shall briefly call regular ARMA. (The case where \( \eta_t \) is some non-stationary ARMA has been studied in [4]). In other words, the signals considered herein are assumed to be represented by the following general non-harmonic model
\[
x_t = H_t + \eta_t (2)
\]
where \( H_t \) satisfies (1) and \( \eta_t \) is a discrete regular ARMA \((p,q)\) process. By extension, each sinusoid in (1) will be termed G-harmonic.

The non-harmonic model that is commonly adopted in most of recent works is a special case of (2), in which \( \eta_t \) is a white noise. For this model, two techniques for estimating line spectrum are well-known, that are the Pisarenko harmonic decomposition and the extended Prony method (see (2)). The first one requires complex calculations and is computationally expensive. The second method consists in estimating the model parameters by least squares (LS) criterion and in solving this non-linear LS problem in several steps. The deficiency of this technique is that the LS estimates of the autoregressive (AR) parameters in its AR filtering step are inconsistent. Another technique for estimating mixed spectrum is the Whittle method (see (1)) that gives consistent estimates. However, when the number of G-harmonics is greater than two, it leads to intractable calculations. We present here a new efficient method to estimate the mixed spectrum of signals of the form (2). The theoretical aspect and an algorithm of the method are first described, then some numerical results and illustrations are given.

2 - Least squares estimates on amplified G-harmonics

Since several methods are now available (see e.g. [3]) to get consistent estimates for the parameters of regular ARMA processes, the problem of estimating the mixed spectrum of a signal \( x_t \) of the form (2) amounts to that of identifying the G-harmonics in (2). In this aim, we introduced in [5] the notion of G-harmonic amplifiers, that is stated below.

Proposition. For each \( j = 1, \ldots, m \), let \( a_j = \cos (\omega_j) \) and let \( \beta_j \) be some real. Define \( \xi_j(t) \) as the unique solution of the following equation
\[
\xi_j(t) = 2 \sum_j \xi_j(t-1) + \xi_j(t-2) = x_t (3)
\]
with the initial conditions \( \xi_j(t_0) = \xi_j(t_0+0) = 0 \) (4)

If \( \beta_j = a_j \), then \( \xi_j(t+t_0) = \sum_j \xi_j(t+t_0) + \sum_j \xi_j(t+t_0) \),
\( t \geq 1 \) where
\[ u_j(t + \omega_j) = (2 \sin \omega_j)^{-1} C_j(t \sin \omega_j(t+1) = \phi_j) - (\sin \omega_j)^{-1}(\sin \phi_j) \sin(\omega_j(t)) \quad (5) \]

\[ z_j(t + \omega_j) = (\sin \omega_j)^{-1} \sum_{k=0}^{t-1} \left[ (\sin(k+1)\omega_j) \times \frac{\partial}{\partial \omega_j} \cos(\omega_j(t-1-k) + \phi_j + \phi_j) + \phi_j \right] - \frac{\partial}{\partial \omega_j} \phi_j \quad (6) \]

Since the first term in (5) is an oscillation with pulsation \( \omega_j \), the amplitude of which quasi-linearly increases with \( t \) whereas the amplitude of the \( G \)-harmonics other than the \( j \)th remains constant in time (see (6)), hence the \( \xi_j(t) \) process will be called harmonic amplifier relative to \( \beta_j \) when \( \beta_j \) is near \( a_j \).

The notion of Recursive Least Squares on Amplified Harmonics (RLSOAH) estimates is now defined.

**Definition.** For each \( j = 1, \ldots, m \) let \( \{ \delta_{j,n} \}_{n \in \mathbb{N}} \) be the sequence defined over the set \( \mathbb{N} \) of non-negative integers as follows:

1. \( \delta_{j,0} = a_j \) where \( a_j \) is some initial estimate for \( a_j \).
2. For any \( n \in \mathbb{N} \), \( \delta_{j,n+1} \) is the LS estimate for \( a_j \) defined by the regression equation

\[ \xi_{j,n}(t) + \xi_{j,n}(t-2) = 2a_j \xi_{j,n}(t-1) + e_{j,n}(t), \quad t \geq 0 \]

where \( e_{j,n}(t) \) is the error term and the \( \xi_{j,n}(t) \)’s are the unique solution of (3), in which \( \beta_j \) is replaced by \( \delta_{j,n} \), with the initial conditions (4). Note that the \( \xi_{j,n}(t) \)’s depend on \( \delta_{j,n} \). The \( \{ \delta_{j,n} \} \) is called the sequence of RLSOH estimates for \( a_j \).

The following basic result (see (5)) establishes for each \( G \)-harmonic \( j \), the conversion of the RLSOH estimates \( \{ \delta_{j,n} \} \) towards \( a_j \) up to \( O_p(N^{-1}) \).

**Theorem.** For any \( j \in \{ 1, \ldots, m \} \) and for any given \( n \in \mathbb{N} \), let \( \tilde{\delta}_{j,n} = \delta_{j,n} \) and \( \delta_{j,n+1} = \tilde{\delta}_{j,n} \).

Define \( \tilde{a}_j = \tilde{\delta}_{j,n} \) and \( a_j = \delta_{j,n+1} \).

If \( \delta_{j} = O(N^{-1}) \) and if \( N \) is sufficiently large, then either \( \delta_{j} - a_{j} = O(N^{-1}) \) or \( |\delta_{j} - a_{j}| \leq (1-\varepsilon_j) |\tilde{a}_{j} - a_{j}| \) with \( \varepsilon_j = r_j + o_p(N^{-1}) \) and \( r_j \) being a real satisfying \( 0 < r_j < 1 \).

The theorem yields a very simple procedure to perform consistent estimates for the pulsations \( \omega_j \), since for each \( G \)-harmonic, e.g. the \( j \)th, it consists of iterations of the following scheme, called one RLSOH operation:

1. Perform the \( \xi_{j,n}(t) \) amplifier relative to \( \delta_{j,n} \).
2. Compute the \( \delta_{j,n+1} \) by the formula

\[ \delta_{j,n+1} = \frac{[(\xi_{j,n}(t) + \xi_{j,n}(t-2)) \xi_{j,n}(t-1)]}{2(\xi_{j,n}(t-1))^2} \]

The convergence criterion is \( |\delta_{j,n+1} - \tilde{\delta}_{j,n}| \leq \varepsilon \).

Generally, \( \varepsilon \) can be chosen equal to \( 10^{-8} \). For each \( G \)-harmonic, the number of RLSOH operations necessary to obtain convergence with \( \varepsilon = 10^{-8} \) is about 25 when starting with improper initial estimates (see (6)). For series \( x_t \) that contain some \( G \)-harmonics with great amplitudes and the others with most weaker amplitudes, we associate with the preceding RLSOH schema, an additional step that reduces the effect of the greater \( G \)-harmonics and thus derive an efficient algorithm called recursive regression on amplified harmonics (RROAH) for estimating the \( G \)-harmonics in (2). This algorithm can be outlined as follows. (For a detailed description, see (6)).

**Step 1.** Starting with initial estimates \( \tilde{\omega}_1, \ldots, \tilde{\omega}_m \), we perform RLSOH operations on the series \( x_t \), then identify the number \( s \) of the dominating \( G \)-harmonics (called of the first level) and obtain primary estimates \( \tilde{a}_1, \ldots, \tilde{a}_s \) for the \( a_j \)’s.

**Step 2.** An improvement of the primary estimates \( \tilde{\delta}_j \) and a reduction of the effect of the \( G \)-harmonics identified in step 1 are here achieved (if \( s > 2 \)) by means of a LS technique (the details of which is given in (6)) that yield final RLSOH estimates \( \delta_{1,n}, \ldots, \delta_{s,n} \) and a series, say \( x_t \), which corresponds to the residuals of the \( x_t \)’s after removing the \( G \)-harmonics identified in step 1. The estimates \( \delta_{j} \) of the \( \omega_j \)’s are then computed by \( \delta_j = c(j) \).

**Step 3.** Since all the remaining \( G \)-harmonics are in the residuals \( \tilde{x}_t \), we replace the \( x_t \)’s by the \( \tilde{x}_t \)’s then iterate steps 1 and 2 with initial estimates \( \tilde{\omega}_{s+1}, \ldots, \tilde{\omega}_m \) and identify new \( G \)-harmonics, called of the second level. The procedure is thus continued until all the \( G \)-harmonics in the \( x_t \)’s are identified.

3. Some numerical results and illustrations

We now illustrate some properties of the RROAH procedure on 4 following simulated series with two \( G \)-harmonics and with the same length \( N = 98 \). They are of the form

\[ x_t = C_1 \cos(2\pi/P_1) + C_2 \cos(2\pi/P_2) + \varepsilon_t, \]

where \( \varepsilon_t \) is an MA(1) process with the parameter \( \theta = 0.7 \) and the noise standard deviation \( \sigma = 0.2 \).

**Series A:** \( C_1 = 15, C_2 = 3, P_1 = 144, P_2 = 6.5 \)
**Series B:** \( C_1 = 15, C_2 = 14, P_1 = 144, P_2 = 6.5 \)
**Series C:** \( C_1 = 15, C_2 = 4, P_1 = 144, P_2 = 6.5 \)
**Series D:** \( C_1 = 8, C_2 = 7, P_1 = 6.5, P_2 = 6.87 \).
1) The accuracy of the RLSOAH estimates is very satisfying. For example, for series B2, the estimates are $\hat{\alpha}_1 = 0.90617$ ($P_1 = 14.389$) and $\hat{\omega} = 0.56823$ ($P_2 = 5.501$).

2) Even when the initial estimates are far from the true values, the RROAH procedure converges to the same limits. For instance, for series Bl, with the following initial estimate $\hat{\alpha}_1 = 0.955$, 0.148 or 0.891, we obtain the same value for $\hat{\alpha}_1$, namely 0.90619 ($P_1 = 14.39$). Further results are given in (6) as well as methods to obtain the initial estimates $\hat{\alpha}_j$.

3) For G-harmonics with very weak amplitudes, the accuracy of the RLSOAH estimates remains very satisfying. For example, with series B3, the second G-harmonic of which has an amplitude nearly equal to the noise deviation $\sigma$, we obtain $\hat{\alpha}_1 = 0.90618$ ($P_1 = 14.39$) at the first level and $\hat{\alpha}_2 = 0.56625$ ($P_2 = 6.48$) at the second level.

4) According to the theorem, the RROAH procedure has a good frequency resolution in a certain range of frequencies: Two frequency peaks, say at $\omega_1$ and $\omega_2$, such that $|\omega_1 - \omega_2| > 1/N$, can be easily distinguished by the RLSOAH estimates. Thus, with series B4 that contains two periods $P_1 = 6.5$ and $P_2 = 6.87$ near each other, we obtain $\hat{\alpha}_1 = 0.56424$ ($P_1 = 6.47$) and $\hat{\alpha}_2 = 0.61425$ ($P_2 = 6.91$). The RROAH estimation procedure was also successfully applied to analyse spectrum of actual data, since with this method, we have shown that the logarithmically transformed series of the well-known Canadian lynx trapping series has a mixed spectrum constituted by two G-harmonics at the periods 9.62 and 5.13 and by the continuous spectrum of an AR(2) process. For the detailed results of this application and further complements, we refer the reader to (6).

Fig. 1. Series Bl

Fig. 2. $\hat{\alpha}_1$-amplifier of series Bl with $\hat{\alpha}_1 = 0.9061965$

Fig. 3. $\hat{\alpha}_2$-amplifier of series B4 with $\hat{\alpha}_2 = 0.618801$

References


