APPROXIMATE SPECTRAL ANALYSIS BY LEAST-SQUARES FIT

Successive Spectral Analysis

PETR VANÍČEK

University of Liverpool Tidal Institute, The Observatory, Birkenhead, Ches., England, and Czech Technical University, Prague, Czechoslovakia

(Received 28 January, 1969)

Abstract. An approximate method of spectral analysis called 'successive spectral analysis' based upon the mean-quadratic approximation of an empirical function by generalised trigonometric polynomial with both unknown frequencies and coefficients is developed. A few quotations describing some properties of the method as well as one of the possible methods for numerical solution are given.

1. Introduction

In astronomy, geophysics, oceanography and other experimental sciences we often have to express a given empirical function in the form of a generalised trigonometric polynomial the frequencies of which are not known beforehand. To solve the problem the classical approximate methods of spectral analysis based either on Fourier transformation or on statistical auto-correlation are used.

In this paper we are going to discuss another approach based upon the minimisation process, using mean-quadratic distance as a criterion. We shall prove that such an approach has certain useful properties. The method, built on this approach, which we shall call 'successive spectral analysis' has given encouraging results when applied to several geophysical problems.

2. Description of Successive Spectral Analysis

Let the given function F be defined on

$$\mathscr{M}_n \equiv \left\{ -\pi + \frac{2\pi}{n-1} i \right\}, \; i = 0, 1, \ldots, n-1,$$

where n is any odd integer greater than 2. We require a general trigonometric polynomial

$$\overset{m}{T}(t) = a_0 + \sum_{j=1}^{m-1} r_j \cos(\omega_j t - \varphi_j), \tag{1}$$

($t \in \mathscr{M}_n$, $\frac{1}{2}(n+1) \geqslant m \geqslant 2$ given), which has the least mean-quadratic distance from F in the space $\mathscr{G}_{\mathscr{M}_n}$ of all the functions defined on $\mathscr{M}_n$. We assume that neither amplitudes a_0, r_j nor frequencies ω_j and phases φ_j are known beforehand. The solution, should it exist, is obviously very difficult. It involves solving a system of non-linear algebraic equations. This was the reason why we have chosen to resolve the problem approximately only, step by step.

Let us, for this purpose write (1) in the following form

$$\overset{m}{T}(t) = \sum_{j=1}^{m-1} (a_{0j} + r_j \cos(\omega_j t - \varphi_j)) = \sum_{j=1}^{m-1} T_j(t), \sum_{j=1}^{m-1} a_{0j} = a_0 . \tag{1'}$$

Introducing $a_j = r_j \cos\varphi_j$, $b_j = r_j \sin\varphi_j$ we can write

$$T_j(t) = a_{0j} + a_j \cos\omega_j t + b_j \sin\omega_j t . \tag{2}$$

Obviously, if the frequencies ω_j of $F \in \mathscr{G}_{\mathscr{M}_n}$ are integers then we can determine each T_j separately instead of seeking $\overset{m}{T}$ as a whole in one go. This is because the matrix of the system of normal equations (see for instance Berezin and Zidkov, 1962; Karatajev, 1963; Laurent, 1965) for such frequencies is diagonal and the system degenerates into $2m-1$ independent linear equations for individual coefficients. Generally though, the polynomial $\overset{m}{T}{}^*$ with components T_j^* found separately, will differ from $\overset{m}{T}$.‡

To find T_1^*, giving the largest 'contribution' to F, we shall exploit the fact that for any rational $\omega_1 \in (0, \frac{1}{4}(n-1)) \equiv \mathscr{E}_{\max}$ the optimal coefficients a_{01}, a_1, b_1 are given by the set of normal equations

$$\left.\begin{array}{r}(1, 1)\, a_{01} + (1, \cos\omega_1 t)\, a_1 + (1, \sin\omega_1 t)\, b_1 = (1, F) \\ (\cos\omega_1 t, 1)\, a_{01} + (\cos\omega_1 t, \cos\omega_1 t)\, a_1 + (\cos\omega_1 t, \sin\omega_1 t)\, b_1 = (\cos\omega_1 t, F) \\ (\sin\omega_1 t, 1)\, a_{01} + (\sin\omega_1 t, \cos\omega_1 t)\, a_1 + (\sin\omega_1 t, \sin\omega_1 t)\, b_1 = (\sin\omega_1 t, F),\end{array}\right\} \tag{3}$$

where by (f, g) we denote the scalar product of $f, g \in \mathscr{G}_{\mathscr{M}_n}$.

In (3) $(1, \sin\omega_1 t)$, $(\cos\omega_1 t, \sin\omega_1 t)$ are identically zero because $\sin\omega_1 t$ is odd,1, $\cos\omega_1 t$ are even function and $\mathscr{M}_n$ is symmetrical.

We get thus

$$\begin{aligned} a_{01} &= \frac{1}{n}\Big(\sum_{t\in\mathscr{M}_n} F(t) - a_1 \sum_{t\in\mathscr{M}_n} \cos\omega_1 t\Big) = \frac{1}{n}\Big(\sum F - a_1 Q\Big), \\ a_1 &= \frac{n \sum_{t\in\mathscr{M}_n} (F(t)\cos\omega_1 t) - \sum_{t\in\mathscr{M}_n} F(t) \cdot \sum_{t\in\mathscr{M}_n} \cos\omega_1 t}{n \sum_{t\in\mathscr{M}_n} \cos^2\omega_1 t - \big(\sum_{t\in\mathscr{M}_n} \cos\omega_1 t\big)^2} \\ &= \frac{n \sum F \cos - Q \sum F}{nQ_1 - Q^2}, \\ b_1 &= \sum_{t\in\mathscr{M}_n} (F(t)\sin\omega_1 t) / \sum_{t\in\mathscr{M}_n} \sin^2\omega_1 t = \sum F \sin / Q_2 , \end{aligned} \tag{4}$$

where $nQ_1 - Q^2 > 0$ follows from Schwarz inequality and Q_2 is obviously also positive.

‡ To what extent does the difference depend on distribution of frequencies, amplitudes and phases remains to be established. From this view the studies of the method have not yet been completed. Although, from a number of experiments carried out for hypothetical as well as observed functions it seems clear that the deviations should be reasonably small.

The square of mean-quadratic distance $\varrho(F, T_1^*)$ is given by

$$\varrho^2 = \sum_{t \in \mathscr{M}_n} (F(t) - T_1^*(t))^2 \in [0, \sum F^2]. \tag{5}$$

This can be rewritten as

$$\varrho^2 = \sum F^2 - a_{01} \sum F - a_1 \sum F \cos - b_1 \sum F \sin \tag{5'}$$

and for the given F is a function of ω_1 only.

The quantity ϱ^2 can be thus regarded as a transformation of F into frequency space. The minima show the approximate values of frequencies 'present' in F in a similar way as the peaks do in either power spectrum or Fourier transform periodogramme. If we were interested in locating unknown frequencies only then this transform or a 'least-square periodogramme' $\sigma = \sum F^2 - \varrho^2 \in [0, \sum F^2]$, showing the dominant frequencies also in form of peaks, could be used.

The argument ω_1^* of the absolute minimum of ϱ^2 in $\mathscr{E} \equiv [\nu_1, \nu_2] \subset \mathscr{E}_{\max}$ is the first frequency we seek. Together with the coefficients a_{01}, a_1, b_1 given by (4) for $\omega_1 = \omega_1^*$ it determines T_1^*. ϱ^2 is analytic on $\mathscr{E}$ so that the minima are sharp; case $\varrho^2 = \text{const.}$ is trivial. If there were several absolute minima in $\mathscr{E}$ we would take the one corresponding to the smallest frequency.

Having found T_1^*, we can calculate the first residue $\Delta^1 F = F - T_1^*$. The procedure of finding the following most distinctive component T_2^* will be exactly the same if we only replace everywhere F by $\Delta^1 F$. Similarly, for any T_j^*, $\Delta^{j-1} F$ shall replace F.

3. Some Properties of the Method

(a) The solution is unique. It is not difficult to see that there exists one and only one polynomial $\overset{m}{T}{}^*$ for any given m.

(b) The fit improves with the increase of m. As this is not obvious let us prove it. To do it, we shall use the identity $\varrho(T, F) \equiv \varrho(F - T, \mathcal{O})$, $\mathcal{O}$ is the zero element of $\mathscr{G}_{\mathscr{M}_n}$.

Is $\Delta^{m-1} F = \Delta^{m-2} F - T_{m-1}^* = F - \overset{m}{T}{}^*$. Hence $\varrho(\overset{m}{T}{}^*, F) = \varrho(\Delta^{m-1} F, \mathcal{O})$ and $\varrho(\overset{m+1}{T}{}^*, F) = \varrho(\Delta^m F, \mathcal{O})$. Therefore $\varrho(\Delta^m F, \mathcal{O}) = \varrho(\Delta^{m-1} F - T_m^*, \mathcal{O}) = \varrho(\Delta^{m-1} F, T_m^*)$. From the fundaments of the method follows $\varrho(\Delta^{m-1} F, T_m^*) \leqslant \varrho(\Delta^{m-1} F, \mathcal{O})$ and we get: $\varrho(\Delta^m F, \mathcal{O}) \leqslant \varrho(\Delta^{m-1} F, \mathcal{O})$ or $\varrho(\overset{m+1}{T}{}^*, F) \leqslant \varrho(\overset{m}{T}{}^*, F)$. The equality thus takes place if and only if $T_m^* \equiv 0$.

(c) If F is represented by a simple sinusoidal curve then the fit is precise. This is obvious and requires no proof.

(d) The fit is invariable in the transformation $F \to F + \text{const}$. It means that $\overset{m}{T}{}^*|_{(F+K)} = \overset{m}{T}{}^*|_F + K$ if $K = \text{const}$.

To prove it let us take (5′) and substitute for a_{01}, a_1, b_1 from (4). We get:

$$\varrho^2|_F = \sum F^2 - \frac{(\sum F)^2}{n} - \frac{(\sum F \sin)^2}{Q_2} - \frac{(n \sum F \cos - Q \sum F)^2}{n^2 Q_1 - n Q^2}. \tag{5''}$$

Similarly, for $F+K$,

$$\begin{aligned}\varrho^2\big|_{(F+K)} &= \sum(F+K)^2 - \frac{(\sum F + \sum K)^2}{n} - \frac{(\sum F\sin - \sum K\sin)^2}{Q_2} - \\ &- \frac{(n\sum F\cos + n\sum K\cos - Q\sum F - Q\sum K)^2}{n^2Q_1 - nQ^2} = \\ &= \varrho^2\big|_F + \varrho^2\big|_K + \Delta\varrho^2\big|_{(F+K)},\end{aligned}$$

where by $\Delta\varrho^2\big|_{(F+K)}$ we denote the term

$$2\left(\sum FK - (1/n)\sum F\sum K - Q_2 b_1\big|_K b_1\big|_F - \frac{nQ_1 - Q^2}{n}\, a_1\big|_K a_1\big|_F\right.$$

Evidently we obtain: $a_1\big|_K=0$, $b_1\big|_K=0$, $a_{01}\big|_K=K$ and finally $\Delta\varrho^2\big|_{(F+K)}=0$, $\varrho^2\big|_K=0$. Therefore $\varrho^2\big|_{(F+K)}=\varrho^2\big|_F$. Equally easily we can prove: $a_{01}\big|_{(F+K)}=a_{01}\big|_F+K$, $a_1\big|_{(F+K)}= =a_1\big|_F$, $b_1\big|_{(F+K)}=b_1\big|_F$. We can conclude that $\Delta^1F=\Delta^1(F+K)$ and the deduction $\overset{m}{T}\big|^*_{(F+K)}=\overset{m}{T}\big|^*_F+K$ is thus evident.

4. Remarks

Now, it is obvious that mean $\bar{F}=(1/n)\sum_{t\in\mathcal{M}_n}F(t)$ can be removed beforehand from all the values $F(t)$ without any influence on results of analysis (with the exception of a_0 that becomes $a_0-\bar{F}$). Removal of the mean will simplify formulae (4), (5′), (5″).

To find the numerical values of minima of ϱ^2 in $\mathscr{E}$ and their arguments we can use any iterative method. In the program, used for analyses carried out so far, the following method of separation of minima was adopted.

A real positive h – the fundamental step – is chosen so that ϱ^2 has no more than one sharp local minimum in any interval $[\omega-h, \omega+h]\subset\mathscr{E}$. We now calculate the values $\varrho^2(\omega)$ for $\omega=v_1, v_1+h, \ldots, v_1+ih, \ldots, v_1+qh$ $(qh\leqslant v_2-v_1<(q+1)h)$ from (5′). If for any $i\in\{1, 2, \ldots, q-1\}$ the conditions:

$$\varrho^2(v_1+(i-1)h) \geqslant \varrho^2(v_1+ih) \leqslant \varrho^2(v_1+(i+1)h) \tag{6}$$

are satisfied then one sharp minimum is separated in the interval $[v_1+(i-1)h, v_1+(i+1)]$. Experiments have shown that the values of $h=\frac{1}{16}$ for $\omega\leqslant 1$ and $\frac{1}{3}$ for $\omega>1$ are adequate.

To determine the numerical value of one minimum as well as the value of its argument, the simple iterative method of dividing the interval into 4 divisions has been used (see for instance Demidovic and Maron, 1966). The precision of iteration $(1/\varepsilon)$ had to be fixed in advance. Transform σ could be used in the same manner as ϱ^2 with the only exception that absolute maxima instead of minima would be sought.

In order to shorten the computation time the values of Q, Q_1, Q_2 can be calculated from the formulae

$$\begin{aligned}Q &= \sin(n\pi\omega/(n-1))/\sin(\pi\omega/(n-1)),\\ Q_1 &= \tfrac{1}{2}\left(n + \sin(2n\pi\omega/(n-1))/\sin(2\pi\omega/(n-1))\right),\\ Q_2 &= \tfrac{1}{2}\left(n - \sin(2n\pi\omega/(n-1))/\sin(2\pi\omega/(n-1))\right),\end{aligned} \tag{7}$$

which are equivalent to the initial ones (see Vancl and Čuřík, 1921).

If $\sum F = 0$ then σ can be written as

$$\sigma(\omega) = \frac{1}{Q_2}\left(\sum F \sin\right)^2 + \frac{n}{nQ_1 - Q^2}\left(\sum F \cos\right)^2 . \tag{8}$$

It can be noticed hence that σ has a similar form to Fourier transform periodogramme given (according to Lanczos, 1957; Serjebrjenikov and Pjervozvanskij, 1965) by

$$R^2(\omega) = (2/n)\left(\left(\sum F \sin\right)^2 + \left(\sum F \cos\right)^2\right). \tag{9}$$

Since σ is a linear transformation of ϱ^2 it retains the properties described in Section 3(c), 3(d). These properties distinguish the least-square periodogramme clearly from Fourier transform periodogramme. The latter has none of the described properties as can be easily proved. The advantage of the former periodogramme is particularly noticeable in lower frequency range. In comparison with power spectrum, the advantage of the least square periodogramme is in higher power of distinction of lower frequencies too. These qualities were also confirmed by several experimental analyses.

The computation of $\sigma(\omega)$ is almost as fast as the computation of $R^2(\omega)$ and much faster than that of power spectrum. Computation of the whole polynomial $\overset{m}{T}{}^*$ (consists of determination of both frequencies and coefficients) is, indeed, much slower. This is not surprising if we consider that the process corresponds to two separate procedures with classical methods – spectral analysis for determining the frequencies and solution of normal equations for the found frequencies.

Evidently, the presented method can be also used for analysing a function defined on a set $\mathcal{N}_n$ of unequidistant arguments. In that case, expressions (4), (5′), (5″), (8) become more complicated and slow the computation down.

References

Berezin, I.S. and Zidkov, N. P.: 1962, *Metody vycislennij* **I**, G.I.F.M.L., Moscow, 5, § 3/1.

Demidovic, B. P. and Maron, I. A.: 1966, *Osnovy vycislitelnoj matematiki* (Czech translation, S.N.T.L., Prague), **IV**, 3.

Karatajev, G. I.: 1963, *Nekotoryje svedenija iz funkcionalnovo analyza*, Primenenije E.C.V.M. pri resenii zadac geofiziki, I.A.N. S.S.S.R., Novosibirsk, § 4.

Lanczos, C.: 1957, *Applied Analysis*, Pitman, London.

Laurent, P.-J.: 1965, *Cours de Théorie de l'Approximation*, Faculty of Science, Grenoble, § 1, 2.

Serjebrjenikov, M. G. and Pjervozvanskij, A. A.: 1965, *Vyjavljenije skrytych perijodicznostjej*, Nauka, Moscow.

Vancl, J. and Čuřík, F.: 1921, *Technický průvodce*, Česká matice technická, Prague, p. 146.