JOURNAL OF COMPUTATIONAL PHYSICS 36, 347–365 (1980)

A New Algorithm to Determine the Parameters of a Sinusoidal Signal

J. LAVERGNAT, J. PORTES, AND M. SYLVAIN

Laboratoire de Géophysique Externe, 4, Avenue de Neptune, 94100 St-Maur, France

Received January 3, 1979; revised October 25, 1979

The main poblem in least-squares estimation of a nonlinear function is to find starting values for the mathematical iterative process. In this paper, we describe an algorithm which, in the case of a sinusoidal signal, is free from this restriction.

1. OUTLINE OF THE PROBLEM

In most space experiments, the spacecraft (rocket, satellite, or balloon) is spin stabilized. As a consequence, measurements made on board are modulated at the spin frequency and appear in the form of sinusoidal signals (generally with noise) the characteristics of which are slowly varying with time. The physical quantity one tries to measure has a magnitude given by the amplitude of the sinusoid and a direction, in the case of a vectorial quantity such as an electric or magnetic field, given by the phase of the signal. Thus, we are faced with the problem of estimating parameters of a sinusoidal signal. Let

$$s(t) = \mathscr{A} \cos(\omega t + \phi) + C. \tag{1}$$

Many works have been devoted to this "hidden periodicity" model. It was recognized a long time ago that the number of nonlinear parameters can be decreased by expanding the harmonic component in the form

$$s(t) = A \cos \omega t + B \sin \omega t + C,$$

where ω is the only nonlinear parameter.

If we have a priori knowledge of the frequency $f = \omega/2\pi$, estimation of A, B, and C by the least-squares method is a linear problem presenting no difficulty. Unfortunately, we are generally not in such favourable circumstances. Two methods of attack seem possible:

— To determine first the frequency by one of the existing methods, spectral analysis [1, 2] or the maximum entropy method [3, 4], and then deduce A, B, and C by a linear regression. The drawback of spectral analysis is the need for a large amount of data and a rather poor resolution of the frequency. The maximum entropy

0021-9991/80/090347-19$02.00/0

method has a good resolution, even with a limited amount of data, but gives a biased estimate of the frequency [5]. Its results are compared with those of the proposed method in Section 5.

— To proceed directly to a nonlinear least-square estimation [6, 7]. Problems in that case result from the possibility for the mathematical equations to have several solutions. The result then depends crucially on the initial guess, which must be close to the true value.

In this paper, we describe an algorithm which, in the case of sinusoidal signals, allows one to be free from the limitation due to the initial guess and to find satisfactory estimates of the parameters with a minimum of a priori information. All qualities of the estimators are not fully demonstrated, but the method is a posteriori justified by the success of its application. We begin with a summary of least-squares estimation in the case of a nonlinear problem (Section 2); then we describe the principles of the proposed algorithm (Sections 3 and 4); the method is tested on simulated data and applied to true data (Section 5).

2. Estimation of Parameters by the Method of Least Squares

Let the interpretative model be

$$s(t) = A \cos \omega t + B \sin \omega t + C. \tag{2}$$

The problem is to deduce, from observations

$$y(t) = s(t) + n(t), \tag{3}$$

where $n(t)$ is noise, an estimation $\tilde{s}$ of s, defined by estimations $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, and $\tilde{\omega}$ of the parameters of the model. In what follows, we suppose that

— the model is true;
— time (the independent variable) is measured without error;
— observations are made on the time interval $[0, T]$.

Moreover, as it is physically reasonable, and to make computations simpler, we assume that the noise is white.

Under these assumptions, estimation according to the least-squares method consists in looking for values $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, and $\tilde{\omega}$ of the parameters such that the sum of the residues

$$S = \int_0^T (s(t) - y(t))^2 \, dt = \int_0^T \varepsilon^2(t) \, dt \tag{4}$$

is minimized.

Mathematically, this is equivalent to solving the system of equations

$$\partial S/\partial \tilde{A} = 0, \tag{5a}$$

$$\partial S/\partial \tilde{B} = 0, \tag{5b}$$

$$\partial S/\partial \tilde{C} = 0, \tag{5c}$$

$$\partial S/\partial \tilde{\omega} = 0. \tag{5d}$$

Theoretical properties of the estimators were studied by Walker [8], who showed that, for stationary noise with zero mean, finite variance σ^2 and no temporal correlation:

— $\tilde{A}$, $\tilde{B}$, and $\tilde{\omega}$ are consistent;

— and their joint distribution is asymptotically normal and unbiased, with mean (A, B, ω) and covariance matrix

$$\frac{2\sigma^2}{A^2+B^2} \cdot \begin{bmatrix} \frac{A^2+4B^2}{m} & \frac{-3AB}{m} & \frac{-6B}{m^2} \\ \frac{-3AB}{m} & \frac{4A^2+B^2}{m} & \frac{6A}{m^2} \\ \frac{-6B}{m^2} & \frac{6A}{m^2} & \frac{12}{m^3} \end{bmatrix}, \tag{6}$$

where m is the number of observations. With these results, practically no theoretical problem remains in the application of the least-squares method to the harmonic model. The last difficulty is to obtain an estimate of $\tilde{\omega}$.

The system of equations (5) is not linear and generally has several solutions. In fact, what we want to find is the absolute minimum of S and not only a local minimum. Classical methods of solving such a system consist in an iterative process starting from an initial guess $(A_1, B_1, C_1, \omega_1)$ of the parameters' vector, differences between them coming from the iterative process adopted (steepest descent, Newton's method, Gauss's method, etc.) [9].

The iterative process converges to the true solution only if the guess vector is not too far from it. As an illustration, we have used the program BSOLVE from Marquardt [10] on a signal without noise:

$$y(t) = A\cos(\omega t + \phi) = A\cos(2\pi f t + \phi)$$

$$\text{with } A = 1 \qquad \omega = 2\pi \quad \text{or} \quad f = 1$$

$$\phi = 0.2$$

sampled at 100 points per period. In order that the program converge to the true solution, we need to take the guess value f_1 of f between

0.7 and 1.5 when $T = 1$ sec (1 period),

0.9 and 1.1 when $T = 5$ sec,

0.995 and 1.055 when $T = 20$ sec.

Thus, it is clear we need a rather good approximation of the frequency to get the method to work correctly, all the more as we intend to use a large number of data. But on the other hand, we need a large interval of data when the noise increases and when, as a consequence, our a priori knowledge of the parameters decreases.

The main object of the algorithm we describe now is principally to overcome this difficulty.

3. Derivation of the Method

3.1. *Elimination of the Linear Parameters*

For a given value of the frequency $\tilde{\omega}$, the solution of (5a) to (5c) is a linear regression problem which can be solved algebraically. Solving formally the partial system,

$$\frac{\partial S}{\partial \tilde{A}} = \frac{\partial S}{\partial \tilde{B}} = \frac{\partial S}{\partial \tilde{C}} = 0 \tag{7}$$

and substituting its solutions $\tilde{A}(\tilde{\omega})$, $\tilde{B}(\tilde{\omega})$, and $\tilde{C}(\tilde{\omega})$ in $S(\tilde{\omega}, \tilde{A}, \tilde{B}, \tilde{C})$, the sum of squares is now expressed as a function of only one variable:

$$\mathscr{S}(\tilde{\omega}) = S(\tilde{\omega}, \tilde{A}(\tilde{\omega}), \tilde{B}(\tilde{\omega}), \tilde{C}(\tilde{\omega})). \tag{8}$$

The method proposed by Lawton and Sylvestre [11] to eliminate linear parameters in nonlinear regression is based on similar considerations.

Now, we can find the zero of $d\mathscr{S}/d\tilde{\omega}$. It is given by

$$\frac{d\mathscr{S}}{d\tilde{\omega}} = \frac{\partial S}{\partial \tilde{\omega}} + \frac{\partial S}{\partial \tilde{A}} \cdot \frac{\partial \tilde{A}}{\partial \tilde{\omega}} + \frac{\partial S}{\partial \tilde{B}} \cdot \frac{\partial \tilde{B}}{\partial \tilde{\omega}} + \frac{\partial S}{\partial \tilde{C}} \cdot \frac{\partial \tilde{C}}{\partial \tilde{\omega}} = 0, \tag{9}$$

which, taking account of (7), implies $\partial S/\partial \tilde{\omega} = 0$.

Thus it is demonstrated, in the simple case of a sinusoidal signal, that the solution of (9) is mathematically equivalent to that of the initial system (5). A more general demonstration is given in Golub and Pereyra [12].

However, if the initial problem has been reduced to the determination of the single parameter $\tilde{\omega}$, the difficulty caused by the need of a good initial guess remains.

3.2. *Delimitation of the Frequency Range*

Let us now show that in the case of a noisy sinusoidal signal, it is possible to determine a frequency range $[\omega_A, \omega_B]$ in which the sum of squares $\mathscr{S}(\tilde{\omega})$ exhibits its absolute minimum and no secondary extremum, which solves the initial guess problem.

The proposed method is based on the fact that, for sufficiently large values of ωT, finding the minimum of

$$S(\tilde{\omega}) = \int_0^T [y(t) - \tilde{s}(t)]^2 \, dt$$

is equivalent to finding the maximum of

$$F(\tilde{\omega}) = \int_0^T y(t) \, \tilde{s}(t) \, dt,$$

$F(\tilde{\omega})$ being evaluated for fixed, but arbitrary, values of $\tilde{\mathscr{A}}$, $\tilde{\phi}$, and $\tilde{C}$ (see Appendix A). The practical interest of this result is due to the fact that the study of $F(\tilde{\omega})$ is simpler than that of $\mathscr{S}(\tilde{\omega})$.

Substituting $y(t)$ (Eq. (3)) and $\tilde{s}(t)$ (Eq. (1)) in the expression of $F(\tilde{\omega})$, we find

$$F(\tilde{\omega}) = D(\tilde{\omega}) + N(\tilde{\omega}),$$

where $D(\tilde{\omega}) = \int_0^T s(t) \, \tilde{s}(t) \, dt$ is the deterministic part of the function and $N(\tilde{\omega}) = \int_0^T n(t) \, \tilde{s}(t) \, dt$ is a random term describing the effect of noise.

It can be shown (see Appendix B) that $D(\tilde{\omega})$ is pseudo periodic with period $2\pi/T$ and has its maximum near the true value $\tilde{\omega} = \omega$ provided that the two following conditions were satisfied:

$$(1°) \quad \omega T \, |\cos(\tilde{\phi} - \phi)| \gg 1, \tag{10}$$

$$(2°) \quad \frac{\mathscr{A}}{2} |\cos(\tilde{\phi} - \phi)| \gg C. \tag{11}$$

These conditions imply that $|\tilde{\phi} - \phi|$ differs from $\pi/2$ and that C, the true value of the zero shifting, is small compared to the amplitude of the signal. Under these conditions, and due to its pseudo periodicity, $D(\tilde{\omega})$ attains no other maximum in a $4\pi/T$ wide interval of values of $\tilde{\omega}$ surrounding ω. Thus it is possible to determine a frequency interval $[\omega_A, \omega_B]$ on which the only extremum of $D(\tilde{\omega})$ is the maximum near ω, and this without any assumption concerning the values of $\tilde{\mathscr{A}}$, $\tilde{\phi}$, and $\tilde{C}$. This interval also contains the main minimum of $\mathscr{S}(\tilde{\omega})$ and in principle may be kept small enough to let it be the only extremum of $\mathscr{S}(\tilde{\omega})$ on it. Experience shows that an interval of width $2\pi/T$ is generally suitable.

The condition (11) concerning the shifting of the zero is not so restrictive as it may seem. In fact, this quantity generally has no physical interest and is small when compared to the amplitude. If unluckily it were too large, we would just have to take for the new origin the average value of the signal; the residual shifting of the zero is then small enough to verify condition (11).

The effect of the random part $N(\tilde{\omega})$ is described in Appendix C, where we show that the signal-to-noise ratio increases like $T^{1/2}$, i.e., like the square root of the number of observations, which is a classical result.

3.3. *Description of the Method*

Now we are able to describe the proposed method.

The first step consists in determining the frequency range $[\omega_A, \omega_B]$ on which $\mathscr{S}(\omega)$ has only one extremum which is its absolute minimum. It is performed by giving arbitrary values to $\tilde{\mathscr{A}}$, $\tilde{\phi}$, and $\tilde{C}$ and determining the maximum of $F(\tilde{\omega})$. If necessary, we first take as the new origin the average value of the signal to satisfy the condition (11). As we must avoid the case $|\phi - \tilde{\phi}| = \pi/2$, we compute $|F(\tilde{\omega})|$, at intervals of $2\pi/(3T)$ in frequency, for two values of the estimated phase $\tilde{\phi}$ differing by $\pi/2$. Maxima are found,respectively, at $\tilde{\omega}_1$ and $\tilde{\omega}_2$. We know that if conditions (10) and (11) are verified, at least one of these extrema belongs to the convex domain surrounding the main extremum of F, but we cannot be quite sure it corresponds to the higher value. The situation is described in Fig. 1, where we have represented the isocurves of $|D(\tilde{\omega})|$ and schematically shown computations. To remove the ambiguity, we compute $|F(\tilde{\omega}_1, \tilde{\phi}_j)|$ and $|F(\tilde{\omega}_2, \tilde{\phi}_j)|$ with a sampling in $\tilde{\phi}$ with a step of $\pi/10$ between $-\pi/2$ and $+\pi/2$ and look again for the extrema. Thus, we get a point which is less than $\pi/(3T)$ in frequency and $\pi/20$ in phase from the main extremum, and in general it will give a value of F larger than the secondary extrema.

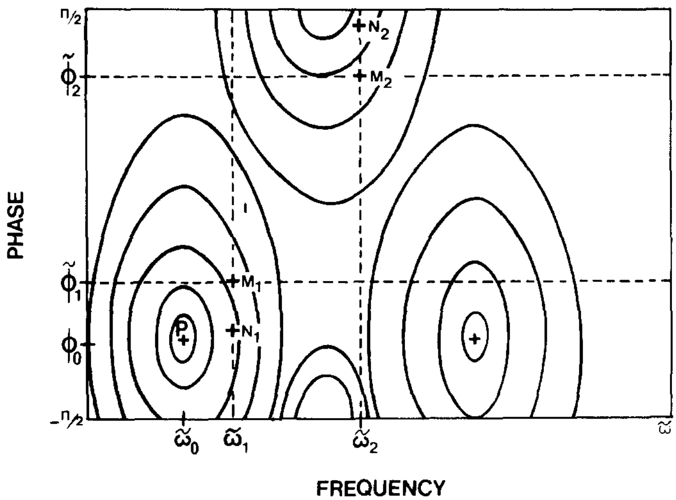

FIG. 1. Schematic representation of the method. We have represented the isocurves of $|F(\tilde{\omega}, \tilde{\phi})|$. P is the main extremum. $|F(\tilde{\omega})|$ is tabulated at three points per pseudo period for $\tilde{\phi} = \tilde{\phi}_1$ and $\tilde{\phi} = \tilde{\phi}_2$, giving respectively maxima M_1 ($\tilde{\omega} = \tilde{\omega}_1$) and M_2 ($\tilde{\omega} = \tilde{\omega}_2$). We then tabulate $|F(\tilde{\phi})|$ for $\tilde{\omega} = \tilde{\omega}_1$ and $\tilde{\omega} = \tilde{\omega}_2$, giving maxima N_1 and N_2. If conditions (12) and (13) hold, the larger maximum (N_1 on the figure) belongs to the convex domain surrounding P.

Let $\tilde{\omega}_i$ be the frequency of that extremum. We look for the minimum of $\mathcal{S}(\tilde{\omega})$ belonging to the interval $[\omega_A, \omega_B]$ with

$$\omega_A = \tilde{\omega}_i - (\pi/T),$$
$$\omega_B = \tilde{\omega}_i + (\pi/T).$$

During this second step, we compute the root of $\mathcal{S}'(\omega) = 0$ numerically, as described in Section 3.1. Apart from noise effect, it may be determined with arbitrary precision, for instance, by the halving method.

At this point, two theoretical problems remain. The first one is to establish properties of the estimator $\tilde{\omega}$, if we do not want to content ourselves with its asymptotic properties.

The second point is that, to have a linear problem, we estimate $\tilde{A}$ and $\tilde{B}$ of model equation (2) when the physical parameters are $\tilde{\mathcal{A}}$ and $\tilde{\phi}$ of model equation (1) which are computed by the formula,

$$\tilde{\mathcal{A}} = (\tilde{A}^2 + \tilde{B}^2)^{1/2},$$
$$\tilde{\phi} = \arctan(-\tilde{B}/\tilde{A}). \tag{12}$$

To be rigorous, we would derive the distribution functions of $\tilde{\mathcal{A}}$ and $\tilde{\phi}$ from those of $\tilde{A}$ and $\tilde{B}$. In fact, computations to get the qualities of the estimators are very dependent on the assumptions concerning the noise. Moreover, even with simplifying assumptions such as gaussian noise, they are nearly inextricable. This is why we preferred to test the algorithm on simulated data rather than to perform a difficult and yet limited analytical study.

4. DESCRIPTION OF THE ALGORITHM

The flow chart of the computation is shown in Fig. 2. Let us comment on it.

The starting table contains sampled data $y_i = y(t_i)$ ($i = 1$ to N). The first step consists in determining the frequency range $[\omega_A, \omega_B]$ surrounding the true value. Two starting values ω_1, ω_2 are used such that we have

$$\omega_1 < \omega_A < \omega_B < \omega_2.$$

Without other information, we know that the frequency must be less than the Nyquist frequency $f_c = \frac{1}{2} f_e = 1/(2\Delta t)$ (f_e being the sampling frequency) and that the data length T must be equal to or greater than, say, half a period; we can thus assume

$$\omega_1 \simeq \pi/T,$$
$$\omega_2 \simeq \omega_c = \pi/\Delta t. \tag{13}$$

We give arbitrary values to $\tilde{\mathcal{A}}$, $\tilde{\phi}$, and $\tilde{C}$, for instance, $\tilde{\mathcal{A}} = 1$, $\tilde{\phi} = 0$, $\tilde{C} = 0$.

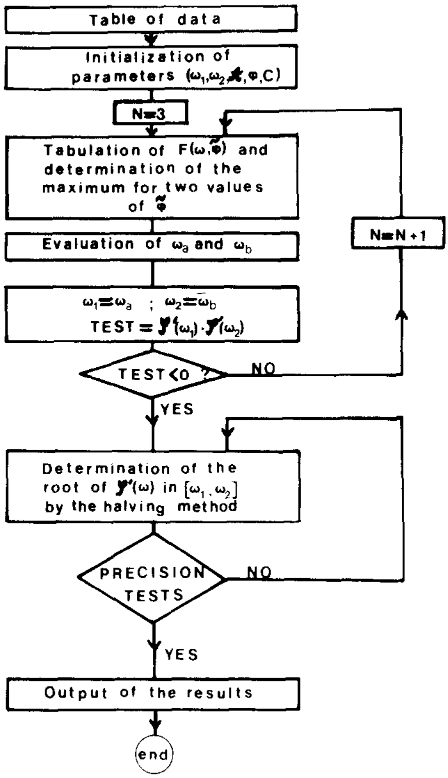

FIG. 2. Flow chart of the program.

Starting from ω_2, we compute

$$F(\omega_j) = \sum_{i=1}^{N} y(t_i)[\tilde{\mathscr{A}} \cos(\omega_j t_i + \tilde{\phi}) + \tilde{C}]$$

with three points per pseudo period, i.e., for

$$\omega_j = \omega_2 - \frac{2\pi}{3T}(j-1) \qquad \text{such that} \quad \omega_j \in [\omega_1, \omega_2]$$

and evaluate the value ω_k for which $|F(\omega_j)|$ is maximum. The same computation is made with $\tilde{\phi}' = \tilde{\phi} + (\pi/2)$ and gives a maximum for another frequency ω_k'. We know that at least one of these frequencies should be at a distance from ω less than π/T, if T and C are such that conditions (10) and (11) are verified (both values may be right, if neither $\tilde{\phi}$ nor $\tilde{\phi}'$ differs from ϕ by about $\pi/2$).

To determine the best value, we use the sweeping in phase described in the discussion of Section 3.3.

We compute $|F(\omega_k, \phi_j)|$ and $|F(\omega'_k, \phi_j)|$ with ϕ_j varying with a step of $\pi/10$ between $-\pi/2$ and $+\pi/2$. We keep the value of ω which leads to the maximum value. Let it be ω_0.

The frequency range surrounding the solution is taken as

$$\begin{aligned} \omega_A &= \omega_0 - \pi/T, \\ \omega_B &= \omega_0 + \pi/T. \end{aligned} \tag{14}$$

We have now to find the zero of $\mathscr{S}'(\omega)$ belonging to the interval $[\omega_A, \omega_B]$. We first check that we have, as expected, $\mathscr{S}'(\omega_A) \cdot \mathscr{S}'(\omega_B) < 0$; otherwise, we can repeat the first step of computation with a better description of $F(\omega)$ (more than three points per pseudo period).

The zero can then be determined by any numerical method. Although it is probably not the most efficient, we have used the halving method which allows a knowledge of the precision with which it is known. We have

$$\mathscr{S}(\omega) = \sum_{i=1}^{N} [(y_i - A(\omega) \cos \omega t_i - B(\omega) \sin \omega t_i - C(\omega)]^2 \tag{15}$$

and taking into account that $A(\omega)$, $B(\omega)$, and $C(\omega)$ are solutions of the system (9).

$$\begin{aligned} \frac{d\mathscr{S}}{d\omega} = 2 \sum_{i=1}^{N} & [y_i - A(\omega) \cos \omega t_i - B(\omega) \sin \omega t_i - C(\omega)] \\ & \cdot [A(\omega)\, t_i \sin \omega t_i - B(\omega)\, t_i \cos \omega t_i]. \end{aligned}$$

The calculation is stopped when we have simultaneously

$$\begin{aligned} |\mathscr{S}'(\omega_n)| &< \varepsilon, \\ |\omega_n - \omega_{n-1}| &< \varepsilon_\omega, \end{aligned}$$

with ε and ε_ω two arbitrary parameters of precision. The estimations of the parameters of the model are

$$\begin{gathered} \tilde{\omega} = \omega_n, \\ \tilde{A} = \tilde{A}(\omega_n) \qquad \tilde{B} = \tilde{B}(\omega_n) \qquad \tilde{C} = \tilde{C}(\omega_n), \\ \tilde{\mathscr{A}} = (\tilde{A}^2 + \tilde{B}^2)^{1/2} \qquad \tilde{\phi} = \arctan(-\tilde{B}/\tilde{A}). \end{gathered}$$

5. Application of the Algorithm

5.1. *Application to Simulated Data*

In order to test the applicability of the algorithm, we applied it so simulated data, the parameters of which are known and controlled. The simulated data are of the form

$$y(t) = s(t) + n(t),$$

with $s(t) = \mathscr{A} \cos(\omega t + \phi) + C$, where

$$\mathscr{A} = 1,$$
$$\omega = 2\pi,$$
$$\phi = 35° = 0.61 \text{ rad},$$
$$C = 0,$$

and $n(t)$ is noise of various types.

All computations have been made with double precision on an IBM 370/168. We can therefore look at the effect of a change of

— the noise level,
— the length of the signal,
— the sampling frequency.

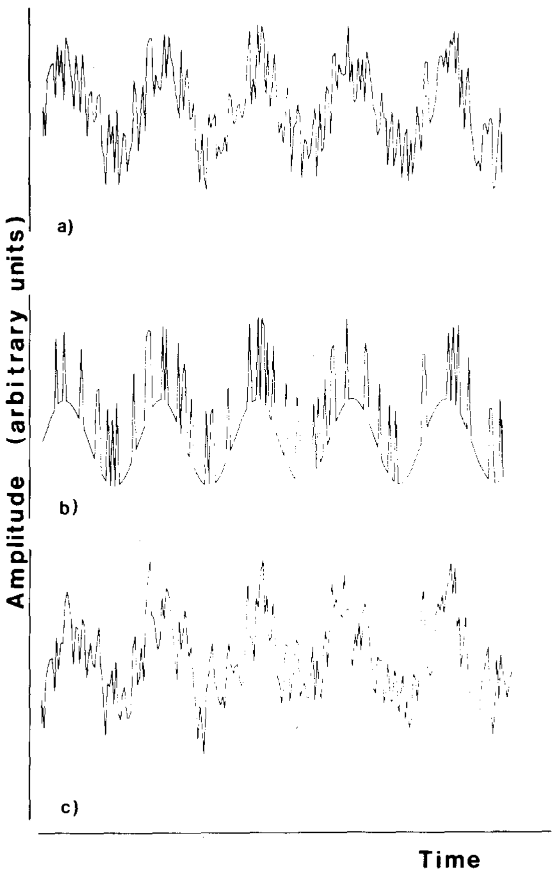

FIG. 3. (a) Sinusoidal signal with gaussian noise (zero mean, $\sigma = 1$). (b) Sinusoidal signal with positive impulses of noise. (c) Sinusoidal signal with noise presenting memory.

TABLE I

Effect of Increasing the Noise Level, $T = 10$ sec

σ	$\tilde{\mathscr{A}}$	$\tilde{\phi}(°)$	$\tilde{\omega}/2\pi$	$\tilde{C}$	
0	1	35	1	0	
0.5	0.96	36	0.999	−0.01	
1	0.92	39	0.998	−0.03	$f_e = 51.11$ Hz
2	0.84	43	0.995	−0.07	
2.5	0.81	46	0.994	−0.09	
3	0.97	64	0.98	10^{-4}	
4	0.988	77	0.978	10^{-3}	$f_e = 91.11$ Hz
4.5	0.988	84	0.97	10^{-3}	

1. *Sinusoidal Signal and White Gaussian Noise* (*Fig.* 3a). The most systematic work has been done for white gaussian noise, with zero mean and standard deviation σ, as considered in the theoretical analysis presented in Sections 2 and 3.

The length of the signal, T, has been varied from 1 to 17 periods, the sampling rate from 9 to 100 points per period, the standard deviation of the noise σ from 0 to 4 times the amplitude of the signal.

It is clear that the relation which may exist between the signal frequency and the sampling frequency is not without any effect on the estimation. For instance, if the sampling frequency is a perfect multiple of the signal frequency the amelioration of the stimation due to an increase of T will be less. We did not try to study this effect in detail, but, in order to avoid it as much as possible, sampling frequencies have never been multiples of the signal frequency.

We present below some results showing the behaviour of the algorithm.

Effect of the noise level: Estimation was made on a signal length of 10 periods with σ varying from 0 to 4.5. The sampling rate is 51 points per period up to $\sigma = 2.5$. Above this level of noise, estimation is no longer possible with this sampling rate and we used 91 points per period. The results are given in Table I.

As expected, estimation becomes poorer when the noise increases, particularly that of the phase; the frequency is rather good in any case.

Effect of the signal length: The length of the signal has been varied from 1 to 17 periods, with a fixed sampling rate of 41 points per period, and two levels of noise: a moderate one ($\sigma = 1$) and a high level ($\sigma = 4$). The results are given in Table II.

In both cases, the estimation is better when T increases, the improvement in the estimation of the phase being the lowest.

Note that when $\sigma = 4$, the results of the algorithm are not reliable for $T < 10$.

Effect of the sampling rate: With $T = 5$ and $\sigma = 1$, we have varied the sampling rate from 10 to 100 points per period. The results are in Table III.

There is an improvement of the estimation of $\tilde{\omega}$ and $\tilde{\phi}$ when f_e increases. But a definite improvement of the estimation of $\tilde{\mathscr{A}}$ is not apparent.

TABLE II

Effect of Increasing the Signal Length, $f_e = 41.1$ Hz

	$\sigma = 1$				$\sigma = 4$			
T	$\tilde{\mathscr{A}}$	$\tilde{\phi}$	$\tilde{\omega}/2\pi$	$\tilde{C}$	$\tilde{\mathscr{A}}$	$\tilde{\phi}$	$\tilde{\omega}/2\pi$	$\tilde{C}$
1	No convergence of the method				2.7	96	0.86	0.9
2	1.18	41	1.03	0.11	2.25	45	1.10	0.5
5	1.04	56	0.98	0.05	1.17	−1	2.82	0.24
10	1.05	44	0.996	0.03	1.25	63	0.991	−0.12
17	0.992	44	0.997	0.01	1.004	67	0.990	−0.03

TABLE III

Effect of Increasing the Sampling Rate, $T = 5$, $\sigma = 1$

f_e	$\tilde{\mathscr{A}}$	$\tilde{\phi}$	$\tilde{\omega}/2\pi$	$\tilde{C}$
10.24	1.02	61	0.98	0.08
20.48	0.8	42	0.996	0.10
41.11	1.04	56	0.98	0.05
61.46	0.94	43	0.992	−0.003
102.43	1.02	33	0.999	−0.04

2. *Other Types of Noise.* In the previous sections, we limited ourselves to white gaussian noise. Without undertaking an exhaustive study, we now want to see how the algorithm works in the presence of different types of noise.

Gaussian noise is theoretically justified by the central limit theorem as the result of a large number of independent sources. To simulate a predominant source of noise, for instance, electrical interferences due to the operation of a relay or some other electrical device, we have added impulses of large amplitudes distributed according to a Poisson process to the signal, both with random sign (Table IV, column 2) and with a systematic sign (Table IV, column 3 and Fig. 3b).

Moreover, a data recorder is always of limited frequency bandwidth. As a consequency, white noise is physically only an approximation. To investigate the effect on the estimation of a time persistence of the noise, we have used the model of noise:

$$b(t_i) = \sum_{j=0}^{m} a_j n(t_{i-j}),$$

with $n(t_k)$ a sample from a normal population with zero mean and standard deviation $\sigma = 1$ (Fig. 3c). Two such models have been considered, differing by the a_j parameters and corresponding to different degrees of persistence. The results are in

TABLE IV

Influence of the Type of Noise, $T = 5, f_e = 41.111$

	1	2	3	4	5
$\tilde{\mathscr{A}}$	1.04	1.007	0.993	0.755	0.843
$\tilde{\phi}$	56	35.88	33.17	30.28	32.52
$\tilde{\omega}/2\pi$	0.98	0.999	0.996	1.009	1.009
$\tilde{C}$	0.05	0.01	0.21	−0.07	−0.06

Notes:

1. White gaussian noise with zero mean and standard deviation $\sigma = 1$ (cf. Fig. 3a).
2. Impulses with gaussian distribution of amplitudes (mean, 1; standard deviation, 0.25) occurring according to a Poisson process, and of random sign.
3. Same noise as 2 but impulses with positive sign (cf. Fig. 3b).
4. $b(t_i) = n(t_i) + \sum_{k=1}^{3} n(t_{i-k}) \exp(-k\lambda)$ with $\lambda = 2.3$ (cf. Fig. 3c).
5. $b(t_i) = n(t_i) + \sum_{k=1}^{10} n(t_{i-k}) \exp(-k\lambda)$ with $\lambda = 0.69$.

Table IV (columns 4 and 5). Column 1 gives the case of white gaussian noise; $(m = 0, \sigma = 1)$.

For all types of noise, we have $T = 5$ and a sampling rate of 41 points per period. It is clear that estimations are rather good in all cases, showing the robustness of the algorihtm. We can notice, however, that impulses with a given sign affect the estimation of the shifting, and noise with memory that of the amplitude.

3. *Conclusions.* Application of our algorithm to simulated data has shown the following results:

At a given level of noise, the estimation is better when the number of observations is larger (increase of T or f_e);

as far as we have tested it, the method is rather robutst;

the best estimate is that of $\tilde{\omega}$, which is in agreement with the theoretical results of Walker ($\mathrm{Var}(\tilde{\omega}) \sim 1/N^3$) (Eq. (6));

the poorer estimate is that of $\tilde{\phi}$, which is due to the arctan transform to get $\tilde{\phi}$ from $\tilde{A}$ and $\tilde{B}$.

When we have no a priori knowledge of the signal-to-noise ratio; it is not possible to know the quality of the stimation we obtain from the algorithm. Only a comparison of the stimation with the data shows if the estimation is acceptable or not. If it is not, a new estimation can be looked for after the sampling rate and/or the length of data used has been increased. For instance, with $\sigma = 4$ and $f_e = 41$ Hz, we have found (Table II) that the estimation given by the algorithm is correct only when $T > 10$.

5.2. *Application to Experimental Data*

As an illustration, we have applied the algorithm to data from a balloon experiment.

Details of the experiment are given in [13]. For our purpose here, it is sufficient to say there was a measure of the horizontal electric field, by the method of the double-probe, and a magnetometer for the attitude restitution. The payload was rotated by a motor, thus eliminating offsets as DC shifts on a sinusoidal signal. In fact, the rotation speed is not uniform and it is thus necessary to make adjustments on rather short lengths of signals.

The magnetic field data and the estimated signal are given in Fig. 4a, and the electric field data and its estimation in Fig. 4b. It is clear that the magnetic field data look much more like a sinusoid than the electric field. This is due to the fact that the vertical electric field, much more intense than the horizontal electric field, appears as noise when the payload, due to its move, is not perfectly horizontal. Nevertheless, the estimation is very good on both signals.

5.3. *Comparison with the M.E.M. algorithm*

An alternative to our method is the determinatiohn of the frequency first by a maximum entropy method (M.E.M.), and the determination of the other parameters by a linear regression.

Without undertaking a complete comparison of both methods, we applied them to

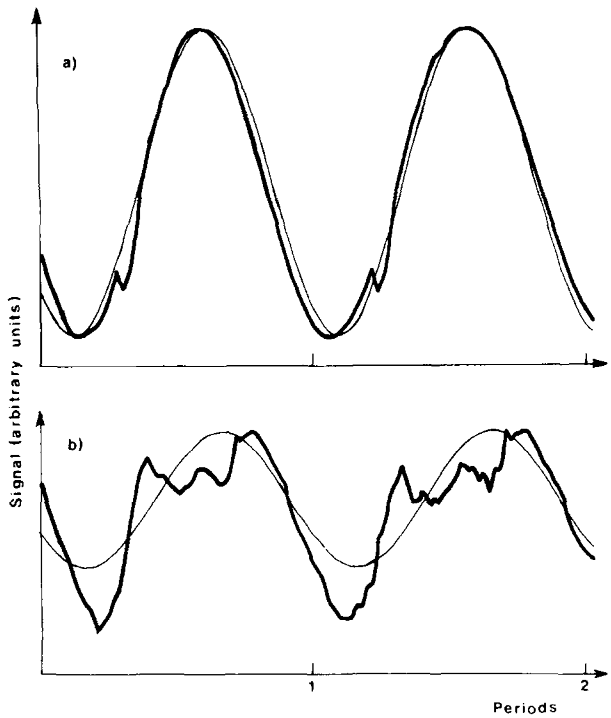

FIG. 4. Application of the method to experimental data: (a) magnetometer signal; (b) horizontal electric field.

TABLE V

Comparison of Our Method with an M.E.M. Algorithm, $T = 10$ sec[c] $f_e = 51.2$ Hz[a]

	Proposed method					M.E.M.algorithm				
σ	$\tilde{\mathscr{A}}$	$\tilde{\phi}$	f^a	$\tilde{C}$	t^b sec	$\tilde{\mathscr{A}}$	$\tilde{\phi}$	f^a	$\tilde{c}$	t^b (sec)
0	1.000	35.00	1.000	0.000	4	1.000	35.00	1.000	0.000	5
0.5	1.024	30.37	1.001	−0.014	7	1.025	32.03	1.000	−0.013	5
1	1.051	25.54	1.002	−0.029	6	1.051	29.75	1.000	−0.027	5
2	1.112	15.42	1.005	−0.059	7	1.067	48.76	0.987	−0.041	5
2.5	1.146	10.26	1.006	−0.075	7	1.046	59.14	0.981	−0.048	5

[a] The M.E.M. algorithm has been used with 100 filter coefficients and the determination of the maximum of the spectrum with a step of 0.003 Hz.

[b] $f = \tilde{\omega}/2\pi$.

[c] t: computing time (in seconds).

the same simulated data, consisting of 10 periods of signal with white gaussian noise, sampled at about 50 points per period. The true values of the parameters of the signal are $\mathscr{A} = 1, f = 1, \phi = 35°, C = 0$. The results given by both methods, for noise with a standard deviation σ that varies from 0 to 2.5 are given in Table V, where the computation time is also indicated. We have used the M.E.M. program MESA from Claerbout [14].

According to our results, both methods give rather equivalent estimations with computation times of the same order of magnitude. For $\sigma > 2$, our method is slower but gives a somewhat more accurate estimation of the frequency.

6. Conclusions

The algorithm we have described allows accurate determination of the parameters of a sinusoidal signal with noise, according to the least-squares method without the need for a rather precise set of starting values, as is the case with classical nonlinear least-squares algorithms.

A limitation (not specific to this algorithm) is the impossibility of obtaining a test of the precision of the estimation, and it is always necessary to compare the estimation with the data. But the algorithm is robust and leads to bad estimations only for very large noise-to-signal ratios. This method is competitive with the use of the M.E.M. followed by a linear regression.

APPENDIX A: PRELIMINARY RESULT

If ωT is large enough, finding the minimum of $S(\tilde{\omega}) = \int_0^T [y(t) - \tilde{s}(t)]^2\, dt$ is equivalent to finding the maximum of $F(\tilde{\omega}) = \int_0^T y(t)\, \tilde{s}(t)\, dt$.

Let $\tilde{s}$ and $\tilde{s}'$ be two models differing only by the values of the estimated parameters. To say that $\tilde{s}$ is the best estimate of the model means that, whatever $\tilde{s}'$ is different from s may be, we have

$$S(\tilde{s}) = \int_0^T [y(t) - \tilde{s}(t)]^2\, dt \leqslant S(\tilde{s}') = \int_0^T [y(t) - \tilde{s}'(t)]^2\, dt, \tag{A.1}$$

so that

$$\int_0^T \tilde{s}(t)^2\, dt - 2\int_0^T \tilde{s}(t)\, y(t)\, dt \leqslant \int_0^T \tilde{s}'(t)^2\, dt - 2\int_0^T \tilde{s}'(t)\, y(t)\, dt. \tag{A.2}$$

Substituting (1) for $\tilde{s}$, one has

$$\begin{aligned} I &= \int_0^T \tilde{s}(t)^2\, dt \\ &= \left(\frac{\tilde{\mathscr{A}}^2}{2} + \tilde{C}^2\right) T + \frac{1}{\tilde{\omega}}\left[\frac{\tilde{\mathscr{A}}^2}{4}(\sin(2\tilde{\omega}T + \tilde{\phi}) - \sin 2\tilde{\phi}) + 2\tilde{C}(\sin(\tilde{\omega}T + \tilde{\phi}) - \sin\tilde{\phi})\right]. \end{aligned}$$

Let us now consider a variation of $\tilde{\omega}$ with $\tilde{\mathscr{A}}$, $\tilde{C}$, and $\tilde{\phi}$ fixed. Then I is the sum of two terms: the first one is independent of the value of $\tilde{\omega}$; the second is a function of $\tilde{\omega}$ but with an absolute value less than

$$\frac{1}{\tilde{\omega}}\left|\frac{\tilde{\mathscr{A}}^2}{4} + 4\tilde{\mathscr{A}}\tilde{C}\right|.$$

The ratio of the second term to the first is thus

$$\frac{1}{\tilde{\omega}T}\,\frac{\tilde{\mathscr{A}}^2/2 + 4\tilde{\mathscr{A}}\tilde{C}}{\tilde{\mathscr{A}}^2/2 + \tilde{C}^2} = \frac{1}{\tilde{\omega}T}\,\frac{x^2 + 8x}{x^2 + 2} \leqslant \frac{3.38}{\tilde{\omega}T}, \tag{A.3}$$

where $x = \tilde{\mathscr{A}}/\tilde{C}$ and when $\tilde{\omega}T$ is large enough, the first term is dominant, and, as a consequence, I is practically no longer dependent on $\tilde{\omega}$. (A.2) then reduces to

$$\int_0^T \tilde{s}(t)\, y(t)\, dt \geqslant \int_0^T \tilde{s}'(t)\, y(t)\, dt \tag{A.4}$$

or $F(\tilde{s}) \geqslant F(\tilde{s}')$, which is the announced result.

APPENDIX B: Study of $D(\tilde{\omega}) = \int_0^T s(t)\,\tilde{s}(t)\,dt$

With $s(t) = \mathscr{A}\cos(\omega T + \phi)$ and $\tilde{s}(t) = \tilde{\mathscr{A}}\cos(\tilde{\omega}T + \tilde{\phi})$ we have

$$\begin{aligned} D(\tilde{\omega}) &= \int_0^T s(t)\,\tilde{s}(t)\,dt \\ &= \frac{\mathscr{A}\tilde{C}}{\omega}\,[\sin(\omega T + \phi) - \sin\phi] + C\tilde{C}T + \frac{\tilde{\mathscr{A}}C}{\tilde{\omega}}\,[\sin(\tilde{\omega}T + \tilde{\phi}) - \sin\tilde{\phi}] \\ &+ \frac{\mathscr{A}\tilde{\mathscr{A}}}{2}\left\{\frac{1}{(\tilde{\omega} - \omega)}\,[\sin((\tilde{\omega} - \omega)\,T + \tilde{\phi} - \phi) - \sin(\tilde{\phi} - \phi)]\right. \\ &+ \left.\frac{1}{\omega + \tilde{\omega}}\,[\sin((\tilde{\omega} + \omega)\,T + \tilde{\phi} + \phi) - \sin(\phi + \tilde{\phi})]\right\}. \end{aligned} \tag{B.1}$$

The first two terms do not depend on $\tilde{\omega}$. All periodic terms in $\tilde{\omega}$ have period $2\pi/T$. As a consequence, $D(\tilde{\omega})$ is a function with pseudo period $2\pi/T$.

The term

$$\frac{\mathscr{A}\tilde{\mathscr{A}}}{2}\,\frac{1}{\tilde{\omega} - \omega}\,[\sin((\tilde{\omega} - \omega)\,T + \tilde{\phi} - \phi) - \sin(\tilde{\phi} - \phi)],$$

except if $(\tilde{\phi} - \phi) = \pi/2$, has its extremum when $\tilde{\omega} = \omega$, and its extremal value is

$$\frac{\mathscr{A}\tilde{\mathscr{A}}}{2}\,T\cos(\tilde{\phi} - \phi). \tag{B.2}$$

The term

$$\frac{\mathscr{A}\tilde{\mathscr{A}}}{2}\,\frac{1}{\tilde{\omega} + \omega}\,[\sin((\tilde{\omega} + \omega)\,T + \tilde{\phi} + \phi) - \sin(\tilde{\phi} + \phi)]$$

has an absolute value which remains less than $\mathscr{A}\tilde{\mathscr{A}}/(\omega + \tilde{\omega})$.

The term

$$\frac{\tilde{\mathscr{A}}C}{\tilde{\omega}}\,[\sin(\tilde{\omega}T + \tilde{\phi}) - \sin\tilde{\phi}]$$

is extremum when

$$\frac{T}{\tilde{\omega}}\cos(\tilde{\omega}T + \tilde{\phi}) - \frac{1}{\tilde{\omega}^2}\,[\sin(\tilde{\omega}T + \tilde{\phi}) - \sin\tilde{\phi}] = 0$$

or

$$\tilde{\omega}T = \frac{\sin(\tilde{\omega}T + \tilde{\phi}) - \sin\tilde{\phi}}{\cos(\tilde{\omega}T + \tilde{\phi})}. \tag{B.3}$$

The value of the extremum is $\tilde{\mathscr{A}} CT\cos(\tilde{\omega}_m T+\tilde{\phi})$ where $\tilde{\omega}_m$ is the root of (B.2)

As a consequence of these results, we see that the function $D(\tilde{\omega})$ attains its maximum near the true value of the frequency when

$$\frac{\mathscr{A}\tilde{\mathscr{A}}}{2} T\,|\cos(\tilde{\phi}-\phi)| \gg \frac{\mathscr{A}\tilde{\mathscr{A}}}{2\omega}$$

or

$$\omega T\,|\cos(\tilde{\phi}-\phi)| \gg 1 \tag{B.4}$$

and

$$\frac{\mathscr{A}\tilde{\mathscr{A}}}{2} T\,|\cos(\tilde{\phi}-\phi)| \gg \tilde{\mathscr{A}} CT\,|\cos(\tilde{\omega}_m T+\phi)|. \tag{B.5}$$

To the second condition can be substituted

$$\frac{\mathscr{A}\tilde{\mathscr{A}}}{2} T\,|\cos(\tilde{\phi}-\phi)| \gg \tilde{\mathscr{A}} CT$$

or

$$\frac{\mathscr{A}}{2}\,|\cos(\tilde{\phi}-\phi)| \gg C. \tag{B.6}$$

APPENDIX C: INFLUENCE OF THE RANDOM PART OF $F(\tilde{\omega})$

The random part of $F(\tilde{\omega})$ is

$$N(\tilde{\omega}) = \int_0^T n(t)\,\tilde{s}(t)\,dt = \tilde{\mathscr{A}} \int_0^T n(t)\cos(\tilde{\omega} t+\tilde{\phi})\,dt + \tilde{C}\int_0^T n(t)\,dt. \tag{C.1}$$

To study its effects, we have to specify the statistical properties of the noise. The analysis has been made for a white gaussian stationary noise with zero mean. This noise is entirely determined by its mean value

$$\langle n(t)\rangle = 0 \tag{C.2}$$

and its autocorrelation function

$$R(t_1, t_2) = R(t_2 - t_1) = R(\tau) = \sigma^2 \delta_0(\tau). \tag{C.3}$$

Computation of the mean of $N(\tilde{\omega})$ is immediate

$$\langle N(\tilde{\omega})\rangle = \left\langle \int_0^T n(t)\,\tilde{s}(t)\,dt \right\rangle = \int_0^T s(t)\langle n(t)\rangle\,dt = 0. \tag{C.4}$$

Its variance is

$$\mathrm{Var}(N(\tilde{\omega})) = \left\langle \int_0^T \int_0^T n(t_1)\, n(t_2)\, \tilde{s}(t_1)\, \tilde{s}(t_2)\, dt_1\, dt_2 \right\rangle.$$

The computation, more tedious but without difficulty, gives us

$$\mathrm{Var}(N(\tilde{\omega})) = \sigma^2 \left\{ \tilde{C}^2 T + 2\, \frac{\tilde{\mathscr{A}}\tilde{C}}{\tilde{\omega}} [\sin(\tilde{\omega}T + \tilde{\phi}) - \sin \tilde{\phi})] + \frac{\mathscr{A}^2}{2\tilde{\omega}} \left[T + \frac{1}{2\omega} (\sin 2(\tilde{\omega}T + \tilde{\phi}) - \sin 2\tilde{\phi}) \right] \right\}. \qquad \text{(C.5)}$$

If $\tilde{\omega}T$ is large enough, we can use the approximate expression

$$\mathrm{Var}(N(\tilde{\omega})) \simeq \left(\frac{\mathscr{A}^2}{2} + \tilde{C}^2 \right) T\sigma^2 \qquad \text{(C.6)}$$

and the standard deviation is then

$$\sigma_F = \sigma \left(\frac{\mathscr{A}^2}{2} + \tilde{C}^2 \right)^{1/2} T^{1/2}. \qquad \text{(C.7)}$$

It increases like the square root of T when the maximum of D (for $\tilde{\omega} \simeq \omega$) increases like T (B.2). Hence, the signal-to-noise ratio increases like $T^{1/2}$.

References

1. A. Papoulis, "The Fourier Integral and Its Applications," McGraw–Hill, New York, 1962.
2. G. M. Jenkins and D. G. Watts, "Spectral Analysis and Its Application," Holden Day, San Francisco, Calif., 1968.
3. T. J. Ulrych, *J. Geophys. Res.* **77** (1972), 1936–1940.
4. T. J. Ulrych and T. N. Bishop, *Rev. Geophys. Space Phys.* **13** (1975), 183–200.
5. W. Y. Chen and G. R. Stegen, *J. Geophys. Res.* **79** (1974), 3019–3022.
6. D. W. Marquardt, *SIAM J.* **2** (1963), 431–441.
7. M. R. Osborne, *SIAM J.* **12** (1975), 571.
8. A. M. Walker, *Biometrika* **58** (1971), 21–36.
9. Y. Bard, "Nonlinear Parameter Estimation," Academic Press, New York/London, 1974.
10. D. W. Marquardt, *in* "Optimization Techniques with FORTRAN" (J. L. Kuester and J. H. Mize, Eds.), pp. 240–250 McGraw–Hill, New York, 1973.
11. W. H. Lawton and E. A. Sylvestre, *Technometrics* **13** (1971), 461–467.
12. G. H. Golub and V. Pereyra *SIAM J. Numer. Anal.* **10** (1973), 413–432.
13. J. J. Berthelier, J. Y. Delahaye, G. Gogly, and F. S. Mozer, *Sparmo Bull.* **5** (1974), 281.
14. J. F. Claerbout, "Fundamentals of Geophysical Data Processing with Applications to Petroleum Prospecting," McGraw–Hill, New York, 1976.