IMPROVING CONVERGENCE OF SINE FITTING ALGORITHMS

M. Fonseca da Silva, A. Cruz Serra

Instituto de Telecomunicações / DEEC, IST, Lisbon Technical University, Portugal
Phone: 351-218418474, Fax: 351-218417672
e-mail: fonseca.silva@lx.it.pt, acserra@ist.utl.pt

Abstract *–Signal-processing applications frequently use least-square sine fitting algorithms. The parameter estimation provided by these algorithms is exposed to errors due to different causes. Good results in sinewave estimation may be achieved when the frequency is unknown by applying the new method presented in this paper.*

Keywords - analog-to-digital converters, sine fitting, parameter estimation, IEEE STD 1057, IEEE STD 1241.

1. INTRODUCTION

In this paper a new technique, to estimate frequency, amplitude, phase and offset values to be used in the initial iteration of the four parameter sine fitting algorithm, is presented. It assures convergence to the global minimum of the error function, even in the presence of saturation, noise or distortion. The restrictive exigency of IEEE Standards 1057 [1] and 1241 [2], of acquisition of a number of records containing a minimum of five cycles may be reduced to about only one period, which means that the domain of its applications may be extended without loss of quality.

We will analyze some aspects related with the sine fitting algorithms, suggested in IEEE Standards 1057 and 1241, which estimate the parameters of a sine wave, defined as $A_0\cos(2\pi f_0 t_n)+B_0\sin(2\pi f_0 t_n)+C_0$, that may fit a set of M samples, $y_1,\ldots,y_M$, acquired at a frequency $f_s=1/T_s$.

The residuals, r_n, of the fit are given by

$$r_n = y_n - A_0\cos(2\pi f_0 t_n) - B_0\sin(2\pi f_0 t_n) - C_0 \qquad (1)$$

In the graphics and tables we will refer the error relative to the amplitude of the ac component $\varepsilon_A = \varepsilon_{rms}/\sqrt{A_0^2+B_0^2}$, where $\varepsilon_{rms} = \sqrt{\sum_{n=1}^{M} r_n^2/M}$ is the rms error.

2. SINE FITTING ALGORITHMS

The knowledge of the signal frequency makes things easier for sine fitting, but normally one has to determine it and some precautions must be taken, otherwise errors will appear. Furthermore, one is frequently dealing with a small amount of samples and periods of a sine wave that may also be degraded by noise and/or distortion.

The importance of accurate frequency estimation may be shown in simple cases involving pure sine waves. Increasing the number of samples can be a bad solution when the frequency is not accurately estimated. The errors will even increase, as shown in Fig. 1, where the results of the three parameter algorithm described in [1,2] for three different frequencies are depicted. The error ε_A reaches the maximum of 100% at approximately a number of samples $N=f_s/\Delta f_0$, where Δf_0 is the absolute error of the estimation of the frequency f_0.

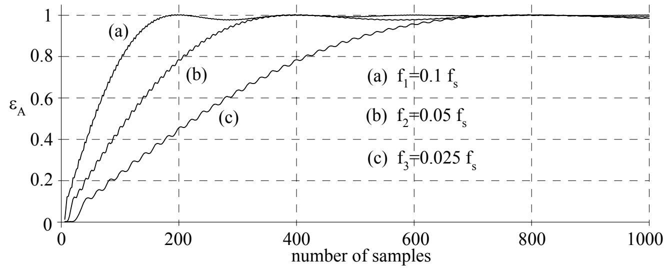

Fig. 1 – Three parameter algorithm error ε_A as a function of the number of samples for three different signal frequencies evaluated with a frequency error Δf_0=5%f_k.

When the estimated frequency diverts from the correct frequency, local minimums are found. Their location depends on the number of samples, M:

– Increasing M increases the number of minimums, decreasing the gap between minimums.

– If the sampling interval contains a large number of periods of the signal under test, the gap between the global minimum and the adjacent maximums is approximately $f=f_0\pm kf_s/M$ [3,4,6], with k a positive integer.

In Fig. 2 the influence of the number of samples is exemplified for the frequency f_0=0.14247f_s ($T_0\cong$7 samples).

It shows the evolution of the relative error of the three parameter algorithm as a function of the number of samples and the estimated frequency, evidencing the presence of the absolute minimum and a large number of local minimums. It is important to notice that in the vicinity of f_0, the sensitivity of the algorithm increases with the number of samples.

Exemplifying for several number of samples the maximums of ε_A near the correct frequency, f_0=0.14247f_s, are listed in Table I. Notice that for a large number of samples (>100) the maximums occur for $f\cong f_0\pm kf_s/M$ whereas for smaller number of samples this rule is not fulfilled.

The main conclusion from the previous considerations is that if the frequency is imperfectly known one risk to get very bad results from the three parameter algorithm, and therefore, it is recommended to use the four parameter algorithm.

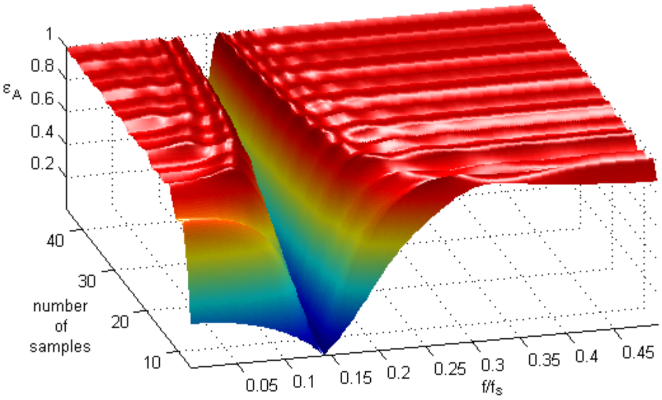

Fig. 2 – Three parameter algorithm error ε_A as a function of the number of samples and the estimated frequency, for f_0=0.14247f_s.

Table I –Maximums of ε_A near the correct frequency, f_0=0.14247f_s, referred to f_s, for different number of samples.

M	1/M	Left maximums		Right maximums	
20	0.05	0.04734	0.09457	0.1908	0.23935
40	0.025	0.09198	0.1173	0.16756	0.19261
50	0.02	0.10201	0.12226	0.16264	0.18278
59	0.0169	0.10885	0.12566	0.15927	0.17606
80	0.0125	0.11765	0.13006	0.15488	0.16728
100	0.01	0.12253	0.1325	0.15244	0.11624
200	0.005	0.13247	0.13747	0.14747	0.15247
500	0.002	0.13847	0.14047	0.14447	0.14647
1000	0.001	0.14047	0.14147	0.14347	0.14447

The four parameter algorithm [1,2] seeks solutions of a nonlinear system of equations, which must be solved in an iterative way. From initial estimated values for the frequency and the other three parameters, A_0, B_0 and C_0, the algorithm produces a new set of values A_i, B_i, C_i and a correction Δf_i to the frequency to be used in the next iteration.

The main problem of this algorithm is that it is highly dependent on the number of samples and especially on the initial estimated values, including naturally the frequency, presenting local minimums in predictable but almost surprising places [3-6]. Occasionally the algorithm generates corrections to the frequency that lead to erroneous solutions making convergence impossible and this problem has not been solved yet!

The phase dependency is very interesting: the same number of samples of the same signal, but retained with different initial phases produce distinct evolutions when submitted to the four parameter algorithm, as it is shown in Fig. 3 which represent the frequency correction (a) and the error (b), resulting from the first iteration of the four parameters algorithm using as initial values, previously estimated by the three parameters algorithm, for frequencies in the vicinity of the correct frequency, f_0=0.14247f_s and 59 samples (≅8.4 periods).,

To illustrate this behavior we have simulated an ADC, with ±10V range and infinite number of bits, sampling the sine wave (10.5sin(2πf_0+φ)+0.2)V with f_0=0.14247f_s and 59 samples: this means that the set of samples will reflect small sections of saturation.

Here is important to recall that the initial estimate frequency by a DFT was f_{DFT}≅0.13559f_s what would be considered a good estimation since the difference to f_0 was only 0.406f_s/M!

The initial phase φ was varied within ±180º with 1 degree steps; all sets of samples converged to the global minimum, using a small amount of iterations (10 or less) with three exceptions for –25º, –26º and 162º: in these three cases, after the 9th iteration, the resulting frequencies alternate around the nearest locals minimums of the global minimum of the error ε_A of the three parameter algorithm, to the left, 0.11819f_s (φ=–26º and 162º) or to the right, 0.16636f_s (φ=–25º) as shown in Fig. 4(a).

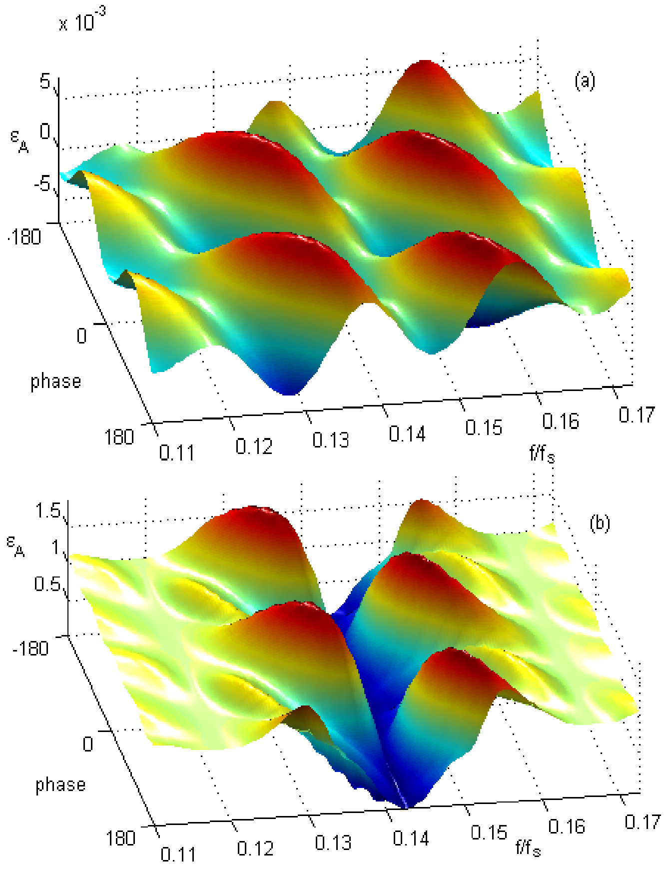

Fig. 3 – Correction of frequency (a) and error ε_A (b) from the 1st iteration of the four parameter algorithm as a function of the initial signal phase and the estimated frequency, for 59 samples and f_0=0.14247f_s.

Repeating the same procedure for an 8 bit ADC, the cases for non-convergence to the global minimum increase significantly as shown in Fig. 4(b). However, one of the previous cases of non-convergence succeed (φ=–26º) and two cases (φ=–76º and 161º) presented frequency errors even bigger, around 0.183632f_s.

In reality, one should not say that the algorithm may, in certain cases, converge to local minimums since the values of the local minimums in the four parameter algorithm vary slightly from iteration to iteration due to frequency, phase and amplitude correction. In fact the resulting frequency alternates around local minimums of the three parameter algorithm ε_A without reaching any stabilized value.

In order to avoid these undesirable situations some improvements were suggested in [5]:

- restricting the corrections to be made to the frequency, assuring that the frequency is not "over-corrected",
- and applying the algorithm with an increasing number of samples, assuring that the frequency it is modified towards the absolute minimum.

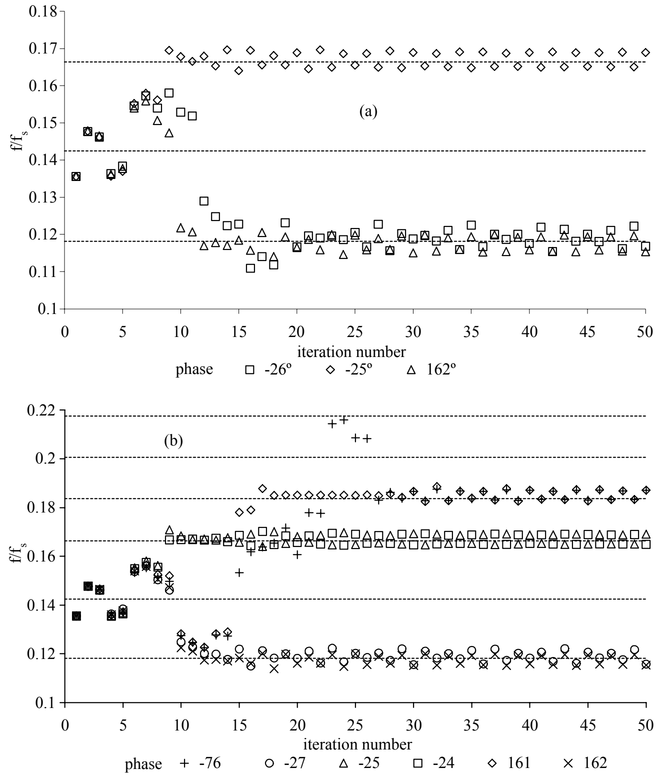

Fig. 4 – Frequency results for the first 50 iterations of the four parameter algorithm applied to 59 samples of the same signal for different initial phase values and f_0=0.14247f_s sampled by an ADC, with an infinite number of bits (a) or 8 bits (b). Dashed lines represent the minimums of the error ε_A of the three parameter algorithm.

Applying this method of progressive subsets of samples to the previous situations, successful convergence has been achieved as shown in Fig. 5 where the evolution of the frequency (a) is shown in conjunction with the "useful" number of samples (b) in each iteration.

Starting from a small number of samples (8/9) and a significant frequency error each new iteration produced a better value of the frequency while the number of samples was slightly increased.

The total set of samples was used when the frequency was almost with the correct value, and the frequency correction given by the algorithm was very small, assuring a good convergence.

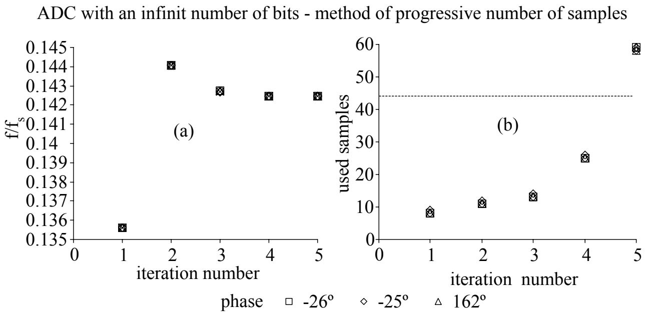

Fig. 5 – Frequency results (a) and number of samples used (b) in the first 5 iterations of the four parameter algorithm applied to an increasing number of samples of the same signal for different initial phase values and f_0=0.14247f_s sampled by an ADC, with an infinite number of bits.

However this method has a weakness: assuring convergence may also signify small increases of the number of samples between iterations and demanding more time to reach the correct values for the frequency and for the other parameters, extending excessively the number of iterations.

3. THE NEW PROPOSAL

The relative error ε_A of the three parameter algorithm as a function of the estimated frequency may be represented, schematically but not scaled, as in Fig. 6. Assuming that the set of samples contains at least one period, the closest maximums in relation to the correct frequency occur approx. for $f_0 \pm f_s/M$ (points L and R) and beyond these maximums the error $\varepsilon_A \cong 1$; but between these points, the curve of ε_A may be approximated by two straight lines intercepting in a point which abscissa is the correct frequency and the coordinate is approximately the total distortion of the sinewave; the bigger the number of samples the better will be the approximation.

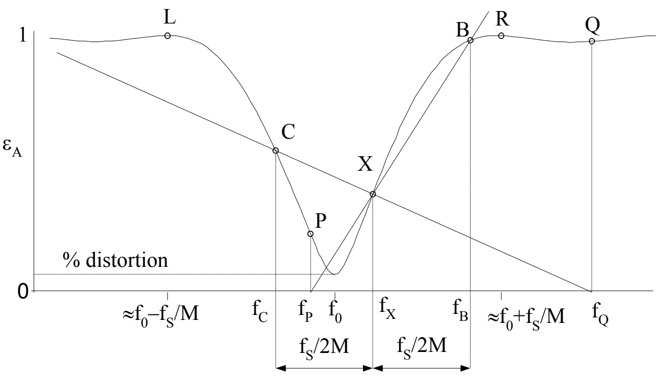

Fig. 6 – Three parameter algorithm error ε_A as a function of the estimated frequency (not scaled) – the point *P* obtained by a linear regression between points *X* (frequency estimated by a DFT or FFT) and *B* (f_B=f_X+f_s/2*M*) helps to find a better assessment of the correct frequency, f_0.

The new procedure we propose to estimate the frequency to be used in the initial iteration of the four parameter sine fitting algorithm is the following:

1) By applying a DFT to the acquired samples, perform a first estimation of the sinewave frequency.

2) Use this frequency as input to the three parameter algorithm. This will result in a certain amplitude error, since the frequency was evaluated with a maximum error of f_s/M. Reporting to Fig. 6, it means that the frequency value would be located in the interval bounded by the frequencies correspondent to points L and R. Let us assume that the first estimate frequency is f_X, and the three parameter algorithm error is ε_X, represented by the point X.
3) Repeat the process for frequencies half way the maximum frequency error, that is f_X–$f_s/2M$ and f_X+$f_s/2M$, obtaining, respectively, points B and C.
4) Perform a linear regression with the two pair of points XB and XC. Let us consider the first pair, located on the same side of the curve relatively to the correct frequency. By the interception with the horizontal axis (ε_A=0) a new frequency f_P is obtained; due to the shape of the concavity of the function of the error the correct frequency will be located between f_A and f_P. The error of the three parameter algorithm at point P will be ε_P eventually greater than ε_X but surely smaller than ε_B.
5) Estimate the frequency by evaluating the weighted average:

$$f_{est} = \frac{f_X \varepsilon_P + f_P \varepsilon_X}{\varepsilon_P + \varepsilon_X}, \quad (2)$$

6) Finally evaluate the remaining initial parameters (A0, B0, and C0) by applying this frequency to the three parameter algorithm.

Applying this set of values to the four parameter algorithm very small corrections will be done and within 2 or 3 iterations all the parameters will stabilize in their corrected values.

If instead of pair XB, we have used the pair XC, a frequency f_Q would have been obtained; this new frequency should be eliminated because $|f_X–f_Q|>f_s/M$ or/and the three parameter amplitude error associated, ε_Q, is probably greater than ε_C but surely greater than ε_X and ε_P.

As an alternative to the evaluation of the amplitude errors associated to the frequencies $f_X \pm f_s/2M$ a thinner interval may be used, specially in the presence of high distortion.

Performing these steps with the three parameters algorithm the control of the frequency is maintained moving towards the global minimum. After this, it is safe to apply the four parameters algorithm, obtaining in only one or, at the most, two steps convergence.

Applying this methodology to the examples earlier referred, as Fig. 7 shows, the six initial iterations use the three parameter algorithm to find the best estimate frequency (clearly the result of iteration 4 should be rejected because the difference to f_X is greater than $1/M$); the next iterations used the four parameter algorithm and the convergence was assured. The frequencies corrections obtained were: 7th iteration ~10^{-4}, 8th iteration ~10^{-6}, and 9th iteration ~10^{-9}.

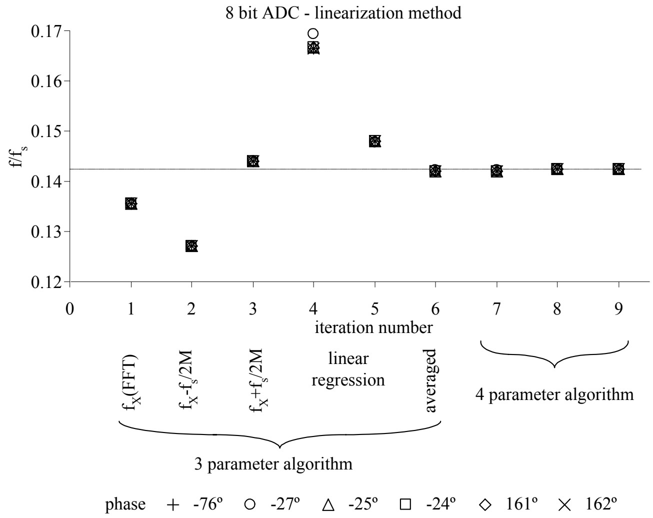

Fig. 7 –Frequency results for the first iterations of the using the three and four parameter algorithm applied to 59 samples of the same signal for different initial phase values and f_0=0.14247f_s sampled by an ADC, with 8 bits using the linearization method

4. CONCLUSIONS

The three and four parameter sine fitting algorithms [1] can perform good results in adverse cases like small number of samples, small number of samples per period, noise and distortion. However they must be used with some precautions. The importance of accurate frequency and other parameters estimations have been discussed.

Using the suggested methods, convergence to the global minimum is assured, avoiding convergence to the neighborhood of local minimums and reducing computation time, even in cases where the algorithms start with rough estimations of the initial parameters. The domain of application of the referred algorithms may be extended to cases where less that 5 periods of the input sine wave are acquired.

ACKNOWLEDGMENT

This paper was sponsored by the Portuguese research project PCTI/ESE/32698/1999 entitled “New Measurement Methods in Analog to Digital Converters Testing”, whose support the authors gratefully acknowledge.

REFERENCES

[1] IEEE Standard for Digitizing Waveform Recorders, IEEE Std 1057-1994.

[2] IEEE Standard for Digitizing Waveform Recorders, IEEE Std 1241 – Draft, v. 050699, May 1999

[3] Haddadi, D., Dallet, D., Marchegay, P., “A Globally Convergent Newton Method for Characterizing A/D Converters”, 4th Workshop on ADC Modeling and Testing, Bordeaux, pp 72-77, 1999.

[4] Haddadi,D.,Dallet,D.,Marchegay, P.,“Time-domain of A/D converters”, 5th Workshop on ADC Modeling and Testing, Vienna, pp 72-77, 2000.

[5] Project Dynad –SMT4-CT98-2214, Draft standard v1.0, pp 65-74.

[6] Fonseca da Silva, M., Cruz Serra, A. “Errors in parameter estimation of quantized sinusoids”, Imeko 2000 World Congress, Vienna, Austria, pp83-88.