

## PRECISION OF THE SINEFITTING-BASED TOTAL HARMONIC DISTORTION ESTIMATOR

**Francisco C. Alegria**

Universidade de Lisboa, Instituto de Telecomunicações and Instituto Superior Técnico, Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal  
(✉ falegria@lx.it.pt, +351 218 418 376)

### Abstract

An analytical expression for the standard deviation of *Total Harmonic Distortion* (THD) estimation is derived. It applies to the case where the estimator uses sine fitting. It is shown that, in common circumstances, it is inversely proportional to the actual value of THD, the signal-to-noise ratio and the square root of the number of samples. The proposed expression is validated both with numerical simulations and an experimental setup using a Monte Carlo procedure.

Keywords: Total Harmonic Distortion, noise, sine wave fitting.

© 2016 Polish Academy of Sciences. All rights reserved

### 1. Introduction

A sine wave is a signal commonly used in many areas ranging from physics and electronics to natural phenomena (diurnal temperature variation). Real signals, however, are rarely identical with the perfect mathematical model of sinusoid. The amplitude, phase and frequency can vary in unintended ways, other signals become added to it that can have a related frequency (harmonics, for instance), or not (spurious components). And there is the ever present random noise which is, most commonly, additive but which can also affect, for instance, the instantaneous phase of the sinusoid.

These non-idealities are originated mainly by the devices the signals pass through. Every device is, to some extent, non-linear which causes the appearance of harmonics. One way to characterize these devices, which can range from amplifiers to analog-to-digital converters, is to use a sine wave as the input signal and measure the output signal, in order to determine how far it is from the ideal sinusoid [1].

In this paper we focus on characterization of the linearity of a device by a parameter called *Total Harmonic Distortion* (THD), which is defined as the ratio of the sum of the powers of all harmonic components to the power of the fundamental [2]. This can be expressed, in decibel, using the amplitudes of sinusoidal components of the signal ( $A_h$ ), as:

$$\widehat{THD}_{dB} = 10 \log \left( \frac{\sum_{h=2}^H \widehat{A}_h^2}{\widehat{A}_1^2} \right), \quad (1)$$

where:  $H$  is the number of harmonics considered when estimating the *THD* (theoretically it can go to  $\infty$ ). The hats over the  $\widehat{THD}_{dB}$  and  $\widehat{A}$  variables mean that they are estimates of the real values. THD can also be expressed as a percentage instead of dBs but the latter is more common nowadays where the power of the harmonics is usually several orders of magnitude lower than the power of the fundamental. One way of estimating the amplitude of a sine wave is using

a non-linear least squares error algorithm to fit the data to a sinusoidal model. IEEE Standards 1057 [3] and 1241 [4] describe these methods for single tone sine waves. There is a 4-parameter method that is used to estimate the amplitude, initial phase, offset and frequency and a 3-parameter method that assumes the frequency is known.

When using sine waves to characterize a device, the frequency of the output signal is determined by the frequency of the stimulus signal which can be assumed known. If an integer number of periods of the output signal are acquired, then a simple method of estimating the THD is to repeat single tone sine fittings using a 3-parameter method and using for the frequency parameter the frequency of the stimulus signal and its integer multiples ( $h$ ) up to a chosen limit ( $H$ ). From the estimated amplitudes  $A_h$  one can use (1) to estimate the THD.

In the past, sine fitting algorithms have been thoroughly studied. In particular, the precision of the estimated parameters has been determined in the presence of Gaussian [5, 6] and non-Gaussian [7, 8] additive noise. The bias of the estimated sine wave amplitude in the presence of additive noise [9] and jitter [10] have also been analyzed.

In this paper we study a specific effect of additive noise on the precision of THD estimates. The theoretical derivations are validated using a Monte Carlo procedure with numerical simulated data and experimental values obtained from real signals.

Section 2 presents an overview of the Sine Wave Fitting algorithm and its parameters. Section 3 presents a theoretical formulation to determine the relation between THD standard deviation and actual THD, number of samples and *Signal-to-Noise Ratio* (SNR). A simplified analytical expression is proposed for computing the standard deviation of THD estimation. The results are validated in Section 4 using a Monte Carlo simulation, and in Section 5, where an experimental setup is described and field results are presented and compared to the analytical expression. Finally, Section 6 presents the conclusions of this work.

## 2. Sine wave fitting

Consider  $M$  data points  $z_1, z_2, \dots, z_M$  sampled from a signal made up of a fundamental sine wave and its harmonics (distorted sine wave):

$$z_i = C + \sum_{h=1}^{\infty} A_h \cos(h\omega_x t_i + \alpha_h + \varphi), \quad (2)$$

where:  $t_i$  are the sampling instants;  $\varphi$  is the initial phase of the signal;  $\alpha_h$  are the initial phases of harmonics  $h$  and  $\omega_x$  is the angular frequency ( $2\pi f_x$ ). These data are affected by additive white Gaussian noise voltage  $d_i$ , with null mean and standard deviation  $\sigma_v$ :

$$y_i = z_i + d_i. \quad (3)$$

We intend to estimate the THD, in dB, of this signal using three-parameter sine fitting to estimate each harmonic amplitude (including the fundamental one,  $h = 1$ ). For each of the  $H$  signal harmonics being considered in the THD estimation, having an index  $h$ , the frequency used for the sine fitting is a multiple of the frequency ( $\omega_x$ ) of the analyzed signal which is assumed known:

$$\omega_h = h \cdot \omega_x, \quad h = 1 \dots H. \quad (4)$$

Note, that in the general case where the signal frequency is not known the first step is to estimate that frequency. That can be carried out using iterative methods that start with estimating a given initial frequency and then refining it in several steps until a minimum in the sine-fitting error is attained [3] (a 3-parameter sine-fitting procedure is employed in each step). Other methods, for example, estimate the frequency based on zero-crossings and then resample the signal so that coherent sampling is obtained [11]. There are numerous methods of estimating

the frequency, and correctness of the estimation will necessarily affect the uncertainty of the harmonic amplitudes. In this paper, however, we are not considering the effect of a poor frequency estimation on the precision of THD estimation (which is arguable an interesting subject for a future paper). Here, we are just considering the presence of additive noise.

As described in [3], the estimation of sinewave amplitudes can be computed from:

$$\begin{bmatrix} \widehat{A}_{I_h} \\ \widehat{A}_{Q_h} \\ \widehat{C} \end{bmatrix} = (D^T D)^{-1} D^T \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_M \end{bmatrix} \quad (5)$$

with

$$D = \begin{bmatrix} \cos(\omega_h t_1) & \sin(\omega_h t_1) & 1 \\ \cos(\omega_h t_2) & \sin(\omega_h t_2) & 1 \\ \dots & \dots & \dots \\ \cos(\omega_h t_M) & \sin(\omega_h t_M) & 1 \end{bmatrix} \quad (6)$$

and

$$\widehat{A}_h = \sqrt{\widehat{A}_{I_h}^2 + \widehat{A}_{Q_h}^2} . \quad (7)$$

If the samples cover an integer number of sinewave periods, inserting (6) into (5) leads to (as shown in [9]):

$$\begin{bmatrix} \widehat{A}_{I_h} \\ \widehat{A}_{Q_h} \\ \widehat{C} \end{bmatrix} = \begin{bmatrix} \frac{2}{M} \sum_{i=1}^M y_i \cos(\omega_h t_i) \\ \frac{2}{M} \sum_{i=1}^M y_i \sin(\omega_h t_i) \\ \frac{1}{M} \sum_{i=1}^M y_i \end{bmatrix} . \quad (8)$$

From this we can easily obtain the square amplitude of the sine wave of frequency  $\omega_h$  that best fits, in a least square error sense, to these  $M$  points:

$$\widehat{A}_h^2 = \frac{4}{M^2} \sum_{i,j} y_i y_j \cos[\omega_h (t_i - t_j)] , \quad (9)$$

where the indices  $i$  and  $j$  go from 1 to  $M$ .

Here, we assume that the number of acquired samples ( $M$ ) covers exactly an integer number of periods ( $J$ ) of the sine wave we are trying to fit to the data. This means that the sine wave frequency ( $f_h$ ), sampling frequency ( $f_s$ ) and number of samples satisfy:

$$\frac{f_h}{f_s} = \frac{J}{M} , J \text{ integer} . \quad (10)$$

Note, that  $J$  and  $M$  should be mutually prime, so that the  $M$  different samples acquired at  $M$  different time instants correspond to  $M$  different sine wave phases. If not, you will have less than  $M$  different phases which will increase the uncertainty in the estimation of the sine wave parameters. In the case of  $J$  being a multiple of  $M/2$ , the sampling instants will correspond to only 2 different sine wave phases and the matrix  $D^T D$  will be singular and hence not invertible (you cannot estimate the 3 sine wave parameters with only two data points).

Note, that the sampling instants are given by  $t_i = i/f_s$ . The assumption is reasonable because we can choose whatever values we want for those frequencies and the number of samples. In practice, however, due to instrument inaccuracies, the actual values of those frequencies may

not be exactly the chosen values that satisfy (10) but are close enough – considering typical frequency errors smaller than 100 ppm. If a non-integer number of periods is acquired a bias will affect the estimator. In this work, however, we will not consider this scenario.

### 3. Standard deviation of THD estimator

The THD estimator, given by (1), can be written as:

$$\widehat{THD}_{dB} = S_{dB} - F_{dB}, \quad (11)$$

where:

$$S_{dB} = 10 \log(S) = 10 \log\left(\sum_{h=2}^H \widehat{A}_h^2\right) \text{ and } F_{dB} = 10 \log(F) = 10 \log\left(\widehat{A}_1^2\right). \quad (12)$$

These two terms ( $S_{dB}$  and  $F_{dB}$ ) can be considered uncorrelated since the fundamental and harmonics are orthogonal and thus the standard deviation of (11) can be computed with:

$$\sigma_{\widehat{THD}_{dB}} = \sqrt{\sigma_{S_{dB}}^2 + \sigma_{F_{dB}}^2}. \quad (13)$$

The variance of these two terms expressed in decibel, in turn, can be computed from their linear versions by approximating the logarithmic function by a Taylor series and retaining only the first term [12]. Thus:

$$\sigma_{S_{dB}}^2 \approx \left(\frac{10}{\ln(10)}\right)^2 \frac{\sigma_S^2}{\mu_S^2} \text{ and } \sigma_{F_{dB}}^2 \approx \left(\frac{10}{\ln(10)}\right)^2 \frac{\sigma_F^2}{\mu_F^2}, \quad (14)$$

where:  $\mu_S$  and  $\mu_F$  are the mean values of  $S$  and  $F$ , respectively. This leads to:

$$\sigma_{\widehat{THD}_{dB}} = \frac{10}{\ln(10)} \sqrt{\frac{\sigma_S^2}{\mu_S^2} + \frac{\sigma_F^2}{\mu_F^2}}. \quad (15)$$

It has been proven in [9] that the sine-fitting estimation of the squared amplitude is biased in the presence of additive noise. Thus, according to (27) of [9]:

$$E\{\widehat{A}^2\} = A^2 + \frac{4}{M} \sigma_v^2, \quad (16)$$

where  $\sigma_v$  is the additive noise standard deviation. This is valid for the fundamental and, of course, also for the harmonics.

The presence of the logarithmic function, in the definition of THD in decibels, leads us to opt for an approximate expression of our estimator's standard deviation. Proceeding in the same way we will consider that for a typical number of samples (much more than one hundred) the second term in (16) can be neglected. Thus:

$$\mu_s \approx \sum_{h=2}^H A_h^2 \text{ and } \mu_F \approx A_1^2. \quad (17)$$

The variance of the estimated square amplitude was also computed in [9] (52):

$$\sigma_{A^2}^2 = \frac{8}{M} \sigma_v^2 A^2 + \frac{16}{M^2} \sigma_v^4. \quad (18)$$

Here, we are also going to neglect the second term leading to:

$$\sigma_S^2 = \frac{8}{M} \sigma_v^2 \sum_{h=2}^H A_h^2 \text{ and } \sigma_F^2 = \frac{8}{M} \sigma_v^2 A_1^2. \quad (19)$$

Inserting (17) and (19) into (15) leads to:

$$\sigma_{\widehat{THD}_{dB}} = \frac{20}{\ln(10)} \frac{1}{\sqrt{M}} \sqrt{2} \frac{\sigma_v}{A_1} \sqrt{\frac{A_1^2}{\sum_{h=2}^H A_h^2} + 1}. \quad (20)$$

Considering that usually  $THD^2 \ll 1$  we can neglect the +1 under the square root since it is much smaller than the first term which corresponds to  $1/THD^2$ . This leads to:

$$\sigma_{\widehat{THD}_{dB}} = \frac{20}{\ln(10)} \frac{1}{\sqrt{M}} \frac{1}{\sqrt{\sum_{h=2}^H \left( \frac{A_h}{\sqrt{2}\sigma_v} \right)^2}}. \quad (21)$$

Note, that the estimation standard deviation does not depend on the signal's fundamental amplitude ( $A_1$ ) but only on the ratio of harmonic amplitude to noise standard deviation.

Furthermore, making use of:

$$SNR = \frac{A_1}{\sqrt{2}\sigma_v} \quad \text{and} \quad THD = \sqrt{\frac{\sum_{h=2}^H A_h^2}{A_1^2}}, \quad (22)$$

we can write a more compact expression for the estimator's precision:

$$\sigma_{\widehat{THD}_{dB}} \approx \frac{20}{\ln(10) \times SNR \times THD \times \sqrt{M}}. \quad (23)$$

This expression is the main result of the paper. It gives us an approximate expression for the standard deviation of the decibel value of THD. It is especially accurate for the case of a high number of samples, a high signal to noise ratio and a low value of THD (in linear units).

#### 4. Monte Carlo simulations

Computer simulations using the Monte Carlo method were carried out to validate the presented derivations and to determine the range of values of the involved variables for which the proposed analytical expressions are valid.

The first simulation has to do with the dependence on the additive noise standard deviation (Fig. 1). We chose to simulate a signal made up of a sine wave (1 V<sub>rms</sub>) plus the 3rd order harmonic component (5 mV<sub>rms</sub>) which corresponds to a THD of -46 dB. The Monte Carlo procedure was executed with 5000 repetitions ( $R$ ).

It can be seen that the value given by our approximate expression is an upper bound for the actual estimation standard deviation. The approximation is rather accurate for values of SNR higher than 46 dB ( $\sigma_v$  lower than 5 mV and  $A = 1$  V<sub>rms</sub>) for the case of a relatively low value of 100 samples ( $M$ ). In general, the harmonic amplitudes are higher than the noise standard deviation in which case the approximation is very good.

We also plotted, in Fig. 2, the dependence of the  $THD_{dB}$  estimator on the number of data points ( $M$ ). The higher the number of data points the lower the estimator standard deviation (better precision).

While Fig. 2 shows the case for an SNR of 46 dB, Fig. 3 shows a case with more noise, namely with an SNR of 34 dB. The higher amount of noise leads to worse approximation for a low number of samples, in particular for less than 100 samples in this case.

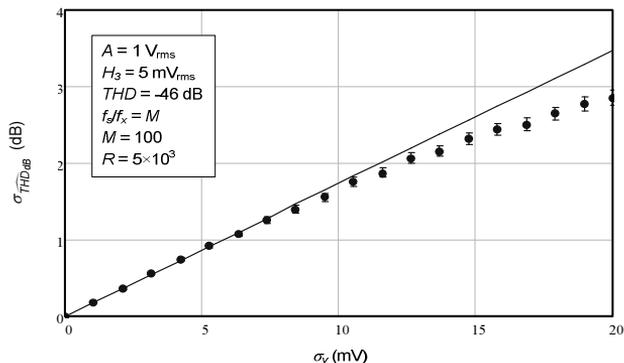


Fig. 1. The standard deviation of the estimated THD in dB as a function of the additive noise standard deviation. The circles represent the values obtained with the Monte Carlo analysis. The confidence intervals for a confidence level of 99.9% are represented by the vertical bars. The solid line represents the theoretical value given by (23).

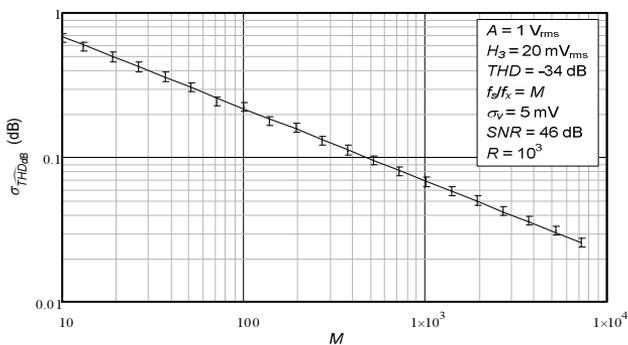


Fig. 2. The standard deviation of the estimated THD in dB as a function of the number of data points for an SNR of 46 dB. The vertical bars represent the values obtained with the Monte Carlo analysis for a confidence level of 99.9%. The solid line represents the theoretical value given by (23).

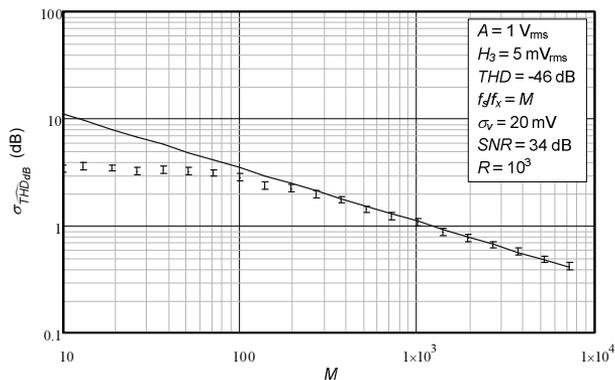


Fig. 3. The standard deviation of the estimated THD in dB as a function of the number of data points for an SNR of 34 dB. The vertical bars represent the values obtained with the Monte Carlo analysis for a confidence level of 99.9%. The solid line represents the theoretical value given by (23).

## 5. Experimental results

All data presented until now are the result of numerical simulation, where signals have been computer-generated in order to verify (23). However, it is also necessary to have experimental data and compare these results to the theoretical expressions in order to prove that the developed formulations truly take into account the phenomena that occur during real experiments.

Thus, an experimental system was assembled in order to create a sine wave which represents the fundamental harmonic with a fixed frequency and amplitude, and another sine wave (the third harmonic) with a lower amplitude. To this signal we added Gaussian white noise created by a separate generator (Fig. 7).

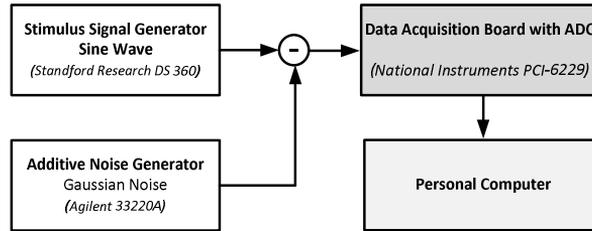


Fig. 4. A scheme of the Test Bench. Combining the sinusoidal stimulus signal and normally distributed noise is carried out inside the data acquisition board by using one of its differential inputs.

National Instruments LabVIEW software provides control of the equipment and collects data from them, resulting in a bidirectional flow of information. There is a function generator (Stanford Research DS 360) responsible for producing a stimulus signal which controls the harmonics (main and third), and another generator (Agilent 33220A) creating additive noise. The output of both generators is connected to an acquisition board (NI PCI-6229), which is plugged into the PC. Table 1 summarizes the main characteristics of the used equipment.

Table 1. Manufacturer specifications for the used instruments and ADC.

Specification	Value
<b>Signal Generator (Stanford Research DS360)</b>	
Sine Wave Amplitude Accuracy	1 %
Sine Wave Offset Accuracy	1% + 25 mV
Sine Wave Frequency Accuracy	25 ppm+4 mHz
Total Harmonic Distortion	-93 dB
<b>Noise Generator (Agilent AG 33220A)</b>	
Noise Bandwidth	≈ 10 MHz
<b>ADC (National Instruments PCI-6229)</b>	
Number of Bits	16
Range	±10 V
Integral Non Linearity (INL)	1.5 LSB
Differential Non Linearity (DNL)	1 LSB
Gain Error	0.015 %
Offset Error	0.5 mV
Bandwidth (small signal)	500 kHz
Quantization Step (Q)	2.44 mV
Noise standard deviation	0.1 LSB
Jitter	Not specified
Timing Accuracy (relative to sample rate)	50 ppm

The standard deviation of noise created by the signal generation system was determined recurring to calibration, using a sine wave of 200 Hz and a fixed noise level of 500 mV. Fig. 5

shows that the calibration constant is 3.32, which corresponds to the value where the experimental value of noise is equal to the desired one (500 mV).

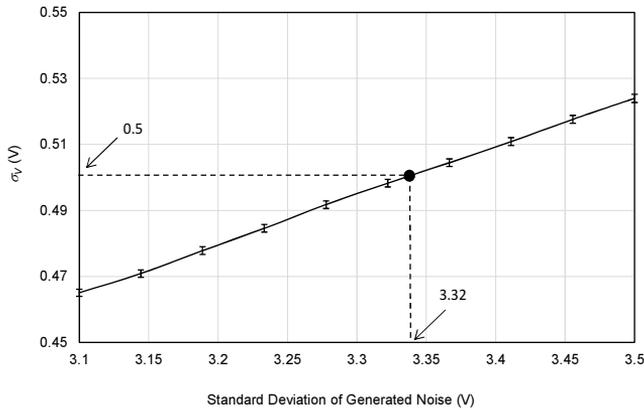


Fig. 5. Mean of the generated RMS signal. The vertical bars correspond to the values obtained experimentally and the solid line represents the constant noise value.

The experimental data were used to validate the models developed in the presented work. The effect of noise, the number of acquired samples and the actual THD (using just one harmonic) was verified separately.

Figure 6 shows the experimental results obtained for different levels of noise. The experimental data show good agreement with the analytical expression. The analytical data clearly follow experimental tendencies, with an expected increase of the THD standard deviation as the noise level grows. These data show that the approximate expression (23) can be used as a THD standard deviation upper bound for a given noise amplitude.

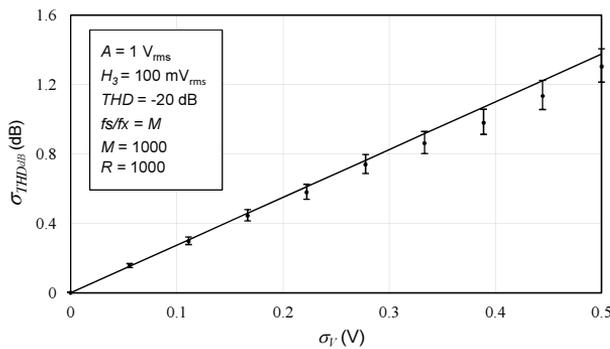


Fig. 6. The standard deviation of the estimated THD in dB as a function of the noise level. The vertical bars represent the values obtained experimentally with the Monte Carlo analysis for a confidence level of 99.9%. The solid line represents the theoretical value given by (23).

The effect of the third harmonic to fundamental amplitudes ratio ( $H_3/A$ ) was studied for a fixed amount of noise (100 mV). Fig. 7 shows that as this ratio increases, the THD standard deviation decreases, according to (22) and (23). For instance, for  $THD = 50\%$ , the THD standard deviation is almost by 80% lower than for  $THD = 10\%$ . This tendency is followed by the experimental results, providing results similar to those obtained by the analytical data provided by (23).

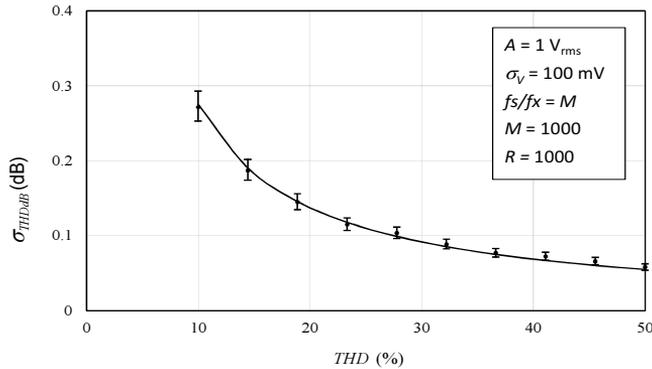


Fig. 7. The standard deviation of the estimated THD in dB as a function of the ratio of the main to third harmonic amplitudes. The vertical bars represent the values obtained experimentally with the Monte Carlo analysis for a confidence level of 99.9%. The solid line represents the theoretical value given by (23).

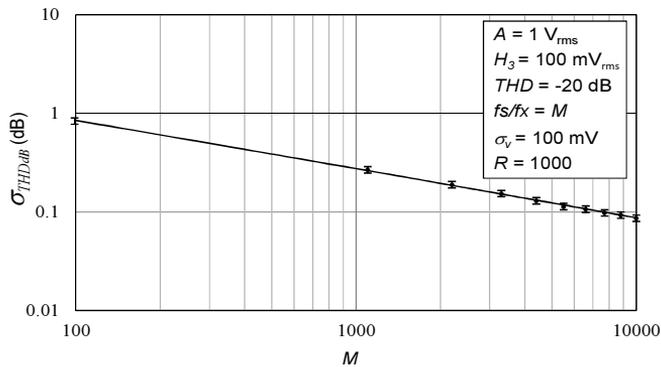


Fig. 8. The standard deviation of the estimated THD in dB as a function of the number of data points. The vertical bars represent the values experimentally obtained with the Monte Carlo analysis for a confidence level of 99.9%. The solid line represents the theoretical value given by (23).

The effect of the number of samples ( $M$ ) on the THD standard deviation shows that, as expected, a higher number of samples is related to a lower standard deviation and higher precision of the results. According to Fig. 8, the experimental data agree with the theoretical analysis, showing the same evolution as  $M$  increases.

## 6. Conclusions

An analytical expression was derived for the standard deviation of the estimated THD, in dBs, of a signal using a least square error 3-parameter sine fitting procedure in the presence of additive noise. The precision of the THD estimates (reciprocal of the standard deviation) is proportional to the actual value of the THD (in linear units), proportional to the signal-to-noise ratio and proportional to the square root of the number of samples.

The expression presented (23) is accurate in typical conditions, *i.e.* for the amount of additive noise smaller than the amplitude of the signal harmonics (and fundamental). Its validity was verified using a Monte Carlo procedure with both simulated and experimental data.

The presented expression is useful in determining the confidence intervals of THD estimates which are important in translating the quality of performed measurements. It can also be used

to determine the minimum number of samples that should be used in order to guarantee a given bound on the THD estimate standard deviation. This in turn is important for optimizing the test run time and memory requirements for measurement systems.

As an example of the application of the presented work consider a case where one wants to estimate the THD with a precision better than 0.1 dB. Taking into account the amount of present noise (for instance 5mV) and the expected value of the harmonics (for instance one single harmonic with 20 mV<sub>rms</sub>) one is able to determine that about 500 samples are required (vide Fig. 2).

## References

- [1] Wigren, T., Händel, P. (1996). Harmonic Signal Modeling Using Adaptive Nonlinear Function Estimation. *Proc. of the Acoustics, Speech, and Signal Processing Conference*, Atlanta, GA, USA, 5, 2952–2955.
- [2] Schmilovitz, D. (2005). On the Definition of Total Harmonic Distortion and Its Effect on Measurement Interpretation. *IEEE Transactions on Power Delivery*, 20(1), 526–528.
- [3] IEEE (2008). IEEE Standard for digitizing waveform recorders – IEEE Std. 1057–2007.
- [4] IEEE (2001). IEEE Standard for Terminology and Test Methods for Analog-to-Digital Converters – IEEE Standard 1241–2000.
- [5] Händel, P. (2000). Properties of the IEEE-STD-1057 four-parameter sine wave fit algorithm. *IEEE Transactions on Instrumentation and Measurements*, 49(6), 1189–1193.
- [6] Andersson, T., Händel, P. (2006). IEEE standard 1057, Cramér-Rao bound and the parsimony principle. *IEEE Transactions on Instrumentation and Measurements*, 55(1), 44–53.
- [7] Moschitta, A., Carbone P. (2004). Cramér-Rao lower bound for parametric estimation of quantized sinewaves. *Proc. of the 21st Instrumentation and Measurement Technology Conference IEEE IMTC 2004*, Como, Italy, 3, 1724–1729.
- [8] Moschitta, A., Carbone, P. (2005). Statistical efficiency of sinewave fitting when using non-linear quantizers. *Proc. IEEE Instrumentation Measurement Technology Conf. IMTC/2005*, Ottawa, Canada, 510–515.
- [9] Alegria, F.C. (2009). Bias of Amplitude Estimation Using Three-Parameter Sine Fitting in the Presence of Additive Noise. *Measurement*, 42(5), 748–756.
- [10] Alegria, F.C., Serra, A.C. (2009). Gaussian Jitter Induced Bias of Sine Wave Amplitude Estimation Using Three-Parameter Sine Fitting. *IEEE Transactions on Instrumentation and Measurement*, 59(9), 2328–2333.
- [11] Borkowski, D. (2005). On-line instantaneous frequency estimation and voltage/current coherent resampling method. *Metrol. Meas. Syst.*, 12(1), 59–75.
- [12] Papoulis, A. (1991). *Probability, Random Variables and Stochastic Processes*. 3rd ed, McGraw-Hill.