19th Symposium IMEKO TC 4 Symposium and 17th IWADC Workshop
Advances in Instrumentation and Sensors Interoperability
July 18-19, 2013, Barcelona, Spain

Threshold estimation in least-squares error functions: Application to impedance spectroscopy

Fernando M. Janeiro[1], Pedro M. Ramos[2]

[1] *Universidade de Évora, Instituto de Telecomunicações, Rua Romão Ramalho 59, Évora, Portugal, fmtj@uevora.pt*
[2] *Instituto de Telecomunicações, Instituto Superior Técnico, DEEC, UTL, Av. Rovisco Pais 1, Lisboa, Portugal, pedro.m.ramos@ist.utl.pt*

Abstract - Least-squares error functions are widely used in the determination of the parameters of models of a system. When this procedure is iterative, the problem of deciding when to stop iterating arises. Usually a threshold is defined, so that when the error function falls below this value, the least-squares error procedure can be stopped and the solution is found. However, this threshold depends on the uncertainty of the measurements. In this paper, the cost function PDF (probability density function) is analyzed to obtain a closed-form expression for the estimation of this threshold, in the framework of impedance spectroscopy.

I. Introduction

System identification is of importance in many fields of science since it can provide valuable insight about the process being modeled [1]. Usually it consists on finding a model that reasonably predicts the output of the system as a response to some set of stimulus inputs. Sometimes the model is known and only its parameters need to be searched for, as in the calibration of a stereo vision system using stars [2], or in impedance spectroscopy of impedances with a known topology [3]. Other times, there is no intrinsic knowledge about the system and the only available information is the system response. In these cases, the search for a suitable model can be performed by evolutionary algorithms such as [4, 5], where a population of potential models are tested and evolved until a suitable model is found. This is performed by minimizing a cost function which is usually the sum of the squares of the errors between the system output and the model output, in what is known as least-squares fitting. The parameters and/or model are found when the value of this cost function is below a predefined threshold. However, the optimal threshold value depends on the uncertainty of the measurements, and is usually not know beforehand. Least-squares are universally used: examples include image registration [6, 2], trace gas analysis [7], sine-fitting algorithms [8] and impedance measurements [9].
In this paper, the problem of threshold definition in least-squares problems is studied under the framework of impedance spectroscopy [10]. The probability density function [11] of the cost function is derived and its histograms are obtained by simulating different impedance measurements which include magnitude and phase uncertainty. Closed-form approximate expressions for the threshold value in least-squares problems are also derived.

II. Cost Function Probability Density Function

In circuit identification through impedance spectroscopy, the measured impedance response is fitted to the response of the equivalent circuit topology through a cost function defined, in a least-squares sense, as

$$\varepsilon = \frac{1}{N_f}\sum_{i=1}^{N_f}\left|\frac{Z_m(f_i)-Z_{real}(f_i)}{Z_{real}(f_i)}\right|^2 \tag{1}$$

where $Z_m(f_i)$ is the measured impedance at frequency f_i, with N_f frequency points, and $Z_{real}(f_i)$ is the impedance of the equivalent circuit topology. Considering that there is relative magnitude and absolute phase measurement uncertainty, both normally distributed with zero mean and variance σ_Z^2 for the magnitude and variance σ_ϕ^2 for the phase, the measured impedance can be written as

$$Z_m(\omega_i) = \left|Z_{real}(\omega_i)\right|(1+X_i)\,e^{j(\phi_{real}+Y_i)} \tag{2}$$

ISBN-10: 84-616-5438-2 | ISBN-13: 978-84-616-5438-3

where $X_i \sim N\left(0,\sigma_Z^2\right)=\sigma_Z N\left(0,1\right)$ is the relative magnitude error and $Y_i \sim N\left(0,\sigma_\phi^2\right)=\sigma_\phi N\left(0,1\right)$ is the phase error and both follow a normal distribution with different standard deviations. Substituting (2) into (1) results into

$$\begin{aligned}\varepsilon &= \frac{1}{N_f}\sum_{i=1}^{N_f}\left|\left(1+X_i\right)\mathrm{e}^{jY_i}-1\right|^2 \\ &= \frac{1}{N_f}\sum_{i=1}^{N_f}\left|\left(1+X_i\right)\left[\cos\left(Y_i\right)+j\sin\left(Y_i\right)\right]-1\right|^2 \\ &= \frac{1}{N_f}\sum_{i=1}^{N_f}\left[\left(1+X_i\right)^2+1-2\left(1+X_i\right)\cos\left(Y_i\right)\right].\end{aligned} \tag{3}$$

which is independent from the real impedance values. Assuming that $Y_i \ll 1$, it is possible to substitute $\cos\left(Y_i\right)$ by its series expansion, resulting into

$$\begin{aligned}\varepsilon &\cong \frac{1}{N_f}\sum_{i=1}^{N_f}\left[\left(1+X_i\right)^2+1-2\left(1+X_i\right)\left(1-\frac{Y_i^2}{2}\right)\right] \\ &= \frac{1}{N_f}\sum_{i=1}^{N_f}\left(X_i^2+Y_i^2+X_iY_i^2\right).\end{aligned} \tag{4}$$

Assuming also that $X_i \ll 1$, the term $X_iY_i^2$ in (4) can be neglected since it is of an higher order than terms X_i^2 and Y_i^2, resulting into

$$\varepsilon \cong \frac{1}{N_f}\sum_{i=1}^{N_f}X_i^2+\frac{1}{N_f}\sum_{i=1}^{N_f}Y_i^2\ . \tag{5}$$

Also, each of the elements in (5) can be written as

$$\frac{1}{N_f}\sum_{i=1}^{N_f}X_i^2 \sim \frac{1}{N_f}\sum_{i=1}^{N_f}\sigma_Z^2N^2\left(0,1\right)=\frac{\sigma_Z^2}{N_f}\chi^2\left(N_f\right) \tag{6}$$

$$\frac{1}{N_f}\sum_{i=1}^{N_f}Y_i^2 \sim \frac{1}{N_f}\sum_{i=1}^{N_f}\sigma_\phi^2N^2\left(0,1\right)=\frac{\sigma_\phi^2}{N_f}\chi^2\left(N_f\right) \tag{7}$$

which are χ^2 distributions, with N_f degrees of freedom, and different scaling factors. Thus, in the special cases where only magnitude uncertainty or only phase uncertainty are considered, the cost function (5) follows a scaled χ^2 distribution. However, in general, since the scaling factors are different, cost function (5) does not follow a chi-square distribution, but according to the central limit theorem, it approaches the normal distribution for high N_f (*i.e.*, when enough different impedance frequency measurements are taken). According to (6) and (7), the mean and standard deviation of the cost function (5) are

$$\mu_\varepsilon=\mathrm{E}\left[\varepsilon\right]=\frac{\sigma_Z^2}{N_f}\mathrm{E}\left[\chi^2\left(N_f\right)\right]+\frac{\sigma_\phi^2}{N_f}\mathrm{E}\left[\chi^2\left(N_f\right)\right]=\sigma_Z^2+\sigma_\phi^2, \tag{8}$$

$$\sigma_\varepsilon=\sqrt{\frac{\sigma_Z^4}{N_f^2}\mathrm{Var}\left[\chi^2\left(N_f\right)\right]+\frac{\sigma_\phi^4}{N_f^2}\mathrm{Var}\left[\chi^2\left(N_f\right)\right]}=\sqrt{\frac{2}{N_f}\left(\sigma_Z^4+\sigma_\phi^4\right)}. \tag{9}$$

Thus, for high N_f, the cost function approaches the normal distribution

$$\varepsilon \sim N\left(\mu_\varepsilon,\sigma_\varepsilon^2\right). \tag{10}$$

ISBN-10: 84-616-5438-2 | ISBN-13: 978-84-616-5438-3

III. Numerical Results

In this section, the previously derived results are numerically analyzed. The impedance of a circuit Z_{real} was computed for N_f frequency points in the range $[0.1, 10]$ kHz. To simulate the measured impedance response, uncertainty was included in the magnitude and phase of Z_{real}, with $\sigma_Z = 0.08\%$ and $\sigma_\phi = 0.05^o$ corresponding to the measurement uncertainties of an impedance meter [12]. The cost function was computed using (1) for 10^6 different realizations of simulated measurements. Figure 1 shows the histograms for the cases with only magnitude uncertainty, only phase uncertainty and both magnitude and phase uncertainty for 10 impedance measurements $\left(N_f = 10\right)$. In Figures 1a) and 1b), the histograms approximately follow the scaled χ^2 distributions (6) and (7), respectively, which are shown by the black line in each plot. However, Figure 1c) approximately follows (5) which is not a χ^2 distribution, thus it is approximated by the normal distribution (10), as shown by the black line. In this case, there is a deviation from the histogram and the probability distribution, as expected since the number of degrees of freedom is only $N_f = 10$. Figure 2 presents the same analysis for $N_f = 100$. Due to the higher number of degrees of freedom the probability density functions now resemble normal distributions and the fit between the obtained histograms and the probability density functions are quite good.

When the histogram follows a χ^2 distribution, the threshold TH to be used in equivalent circuit identification can be obtained by the chi-square inverse cumulative distribution

$$TH = \frac{\sigma^2}{N_f} F^{-1}\left(1-p\right) \text{ where } F\left(x\right) = \int_0^x \frac{t^{\frac{N_f-2}{2}} \mathrm{e}^{-\frac{t}{2}}}{2^{\frac{N_f}{2}} \Gamma\left(\frac{N_f}{2}\right)} dt \tag{11}$$

where $\Gamma\left(x\right)$ is the Gama function and p is the percentage of realizations that are allowed to remain above the threshold. When both magnitude and phase uncertainty are included, the threshold TH can be obtained by

$$TH = \bar{\varepsilon} + \sigma_\varepsilon \sqrt{2}\, \mathrm{erf}^{-1}\left(1-2p\right) \tag{12}$$

which is valid for normal distributions. Figures 1 and 2 also present the estimated thresholds for $p = 5\%$, using (11) for cases a) and b), and (12) for case c). With the computed thresholds, the corresponding probability density function areas for cost function values below the threshold were computed from the histograms. For the χ^2 distributions, the areas are all very close to $1-p = 95\%$ as expected. However, when there is magnitude and phase uncertainty, as in Figure 1c) and Figure 2c), the areas are lower but still close to $1-p = 95\%$ even for low N_f, as is the case of Figure 1c).

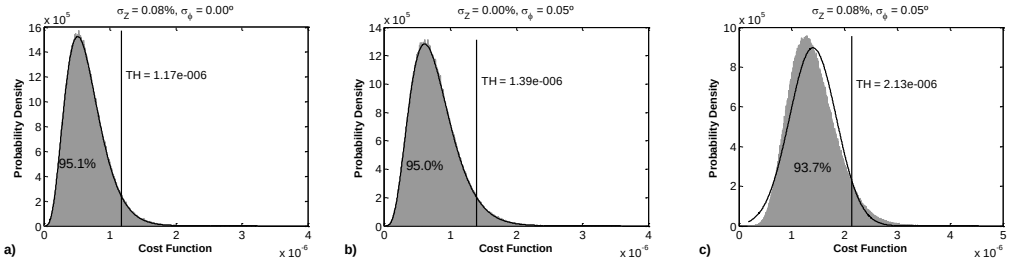

Figure 1. Histogram of the cost function for $N_f = 10$ with: a) only magnitude uncertainty; b) only phase uncertainty; c) both magnitude and phase uncertainty. Black curve corresponds to: a) the χ^2 distribution (6); b) the χ^2 distribution (7); c) the normal distribution (10).

ISBN-10: 84-616-5438-2 | ISBN-13: 978-84-616-5438-3

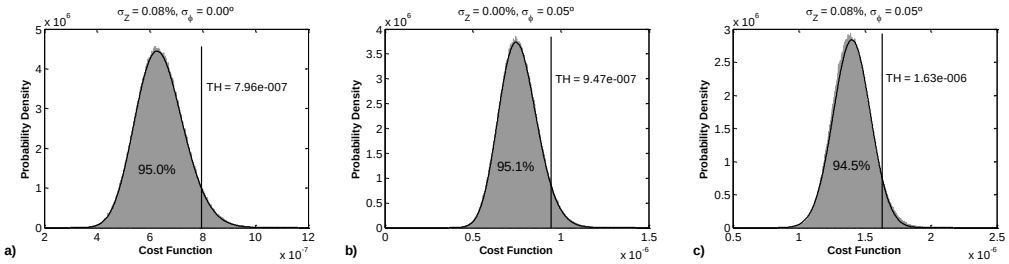

Figure 2. Histogram of the cost function for $N_f = 100$ with: a) only magnitude uncertainty; b) only phase uncertainty; c) both magnitude and phase uncertainty. Black curve corresponds to: a) the χ^2 distribution (6); b) the χ^2 distribution (7); c) the normal distribution (10).

In real measurement conditions, both magnitude and phase uncertainties will be present. Therefore, the cost function distribution is approximated by normal distribution (10).
From the deduction in the previous section, it was concluded that the error function distribution was independent from the impedance under test, as shown in (3). To further illustrate these results, the circuit in Figure 3, corresponding to the equivalent impedance of a humidity sensor [13], was considered. Its impedance magnitude and phase response, in the range $[0.1, 10000]$ Hz is shown in Figure 4.

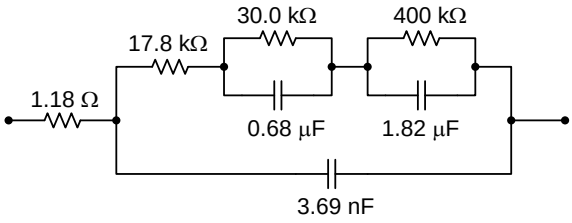

Figure 3. Humidity sensor equivalent circuit for 20 % relative humidity [13].

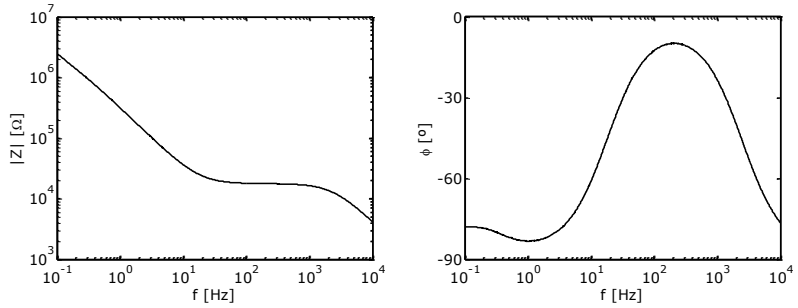

Figure 4. Impedance magnitude and phase response of the circuit in Figure 3.

Once again, magnitude uncertainty $\sigma_Z = 0.08\%$ and phase $\sigma_\phi = 0.05^o$ uncertainty were included in the simulated measurements. The resulting histograms for $N_f = 10$ and $N_f = 100$ are shown in Figure 5 together with the approximate normal distributions, shown by the black curve overlaying the histograms, and the threshold values obtained from (12). The threshold values and corresponding areas are almost the same as the results shown in Figures 1c) and 2c), confirming that the resulting probability density functions depend only on the cost function definition and on the magnitude and phase uncertainty. A further test was performed on the simulated measurements of an impedance consisting on an ideal 1000 Ω resistance. The resulting probability functions are shown in Figure 6 and confirm that they are independent from the impedance values. Notice that the histograms are obtained from (1) using the simulated impedance values and none of the approximations that lead to (4), (5) or (10).

ISBN-10: 84-616-5438-2 | ISBN-13: 978-84-616-5438-3

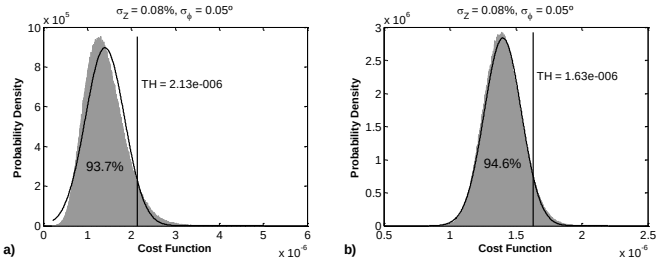

Figure 5. Histogram of the cost function for the humidity equivalent circuit shown in Figure 3 for: a) $N_f = 10$; b) $N_f = 100$. Black curve corresponds to the estimated normal distribution (10).

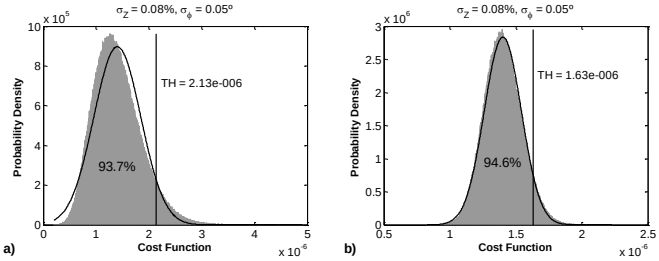

Figure 6. Histogram of the cost function for impedance $Z = 1000\,\Omega$ for: a) $N_f = 10$; b) $N_f = 100$. Black curve corresponds to the estimated normal distribution (10).

V. Conclusions

This paper addressed the problem of defining the appropriate threshold when using iterative least-squares to fit a noisy set of measurements to a model of the system under test. Impedance spectroscopy was used as a case-study but the results can be adapted to any problem of the iterative least-squares class. The cost function PDF for an impedance response with relative Gaussian magnitude uncertainty as well as phase uncertainty has been derived. It was found that when there is only magnitude or only phase uncertainty in the measurements, the resulting PDF is a scaled χ^2 distribution. In the more realistic case where both magnitude and phase have uncertainty, the resulting PDF is not a χ^2, but can be reasonably approximated by a normal distribution even if the number of frequency measurements is as low as 10. Approximate expressions for the average and standard deviation of the approximate normal distribution were also derived. Histograms for different test impedances were simulated and were in good agreement with the predicted probability density functions as shown in the numerical results. The derivation of the PDF of the cost function showed that it is independent of the impedance under test, which was confirmed by the numerical results. Finally, closed-form expressions for the threshold were derived for the case where the PDF is a scaled χ^2 and also for the case where it has to be approximated by a normal curve.

Acknowledgements

This work has been supported by: FCT project PTDC/CTE-ATM/115833/2009 and
Program COMPETE FCOMP-01-0124-FEDER-014508.

References

[1] T. Söderström, P. Stoica, System Identification, Prentice Hall, Upper Saddle River, N.J., 1989
[2] F. M. Janeiro, F. Carretas, K. Kandler, P. M. Ramos, F. Wagner, "Automated Cloud Base Height and Wind Speed Measurement Using Consumer Digital Cameras", XX IMEKO World Congress, Busan, Republic of Korea, 2012.
[3] F. M. Janeiro and P. M. Ramos, "Application of Genetic Algorithms for Estimation of Impedance Parameters of Two-Terminal Networks", I2MTC 2009, Singapore, May 2009.
[4] F. M. Janeiro, J. Santos, P. M. Ramos, "Gene Expression Programming in Sensor Characterization: Numerical Results and Experimental Validation", IEEE Transactions in Instrumentation and Measurement, vol. 62, no. 5, pp. 1373-1381, 2013.
[5] P. Arpaia, "A cultural evolutionary programming approach to automatic analytical modeling of electrochemical phenomena through impedance spectroscopy," Meas. Sci. Technol., vol. 20, no. 6, pp. 1–10, Jun. 2009.
[6] A. Gruen, D. Akca, "Least squares 3D surface and curve matching", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 59, n.º 3, pp. 151–174, 2005.
[7] P. K. Koner, J. R. Drummond, "Atmospheric trace gases profile retrievals using the nonlinear regularized total least squares method", Journal of Quantitative Spectroscopy and Radiative Transfer, vol. 109, n.º 11, pp. 2045–2059, 2008.
[8] T. Radil, P. M. Ramos, A Cruz Serra, "DSP based portable impedance measurement instrument using sine-fitting algorithms", IMTC 2005, Ottawa, Canada, May 2005.
[9] F. M. Janeiro and P. M. Ramos, "Impedance measurements using genetic algorithms and multiharmonic signals," IEEE Trans. Instrum.Meas., vol. 58, no. 2, pp. 383–388, Feb. 2009.
[10] J. R. Macdonald, ed. *Impedance Spectroscopy: Emphasizing solid materials and systems*, Wiley Interscience, 1987.
[11] W. Navidi, *Statistics for Engineers and Scientists*, McGraw-Hill, 2010.
[12] 3522-50/3532-50 LCR HiTester, Hioki E. E. Corporation, 2001.
[13] J. Hoja and G. Lentka, "Method using bilinear transformation for measurement of impedance parameters of a multielement two-terminal network," IEEE Trans. Instrum. Meas., vol. 57, no. 8, pp. 1670–1677, Aug. 2008.

ISBN-10: 84-616-5438-2 | ISBN-13: 978-84-616-5438-3