Jerzy AUGUSTYN
KIELCE UNIVERSITY OF TECHNOLOGY, DIVISION OF ELECTRICAL ENGINEERING AND MEASUREMENT SYSTEMS
7 Tysiąclecia Państwa Polskiego Ave., 25-314 Kielce

Measurement Uncertainty In Algorithms Using PCA Method

Abstract

Propagation analysis of selected uncertainty sources in algorithms using the PCA method has been presented. The paper shows uncertainty analysis in algorithms, which use minimization of squared distances technique and maximizing the variance technique. On the basis of simulation tests the influence of the used signal sampling technique on the eigenvalues vector for the sinusoidal signal containing additive white noise, has been compared. Three applied sampling techniques have been analyzed: synchronization of the beginning of the data acquisition, for the successive sequences $\mathbf{x}_i$, immediately after a positive zero-level crossing of the analyzed waveform; separation of the *NM*-element sequence $\mathbf{x}$ to *M* of *N*-element sequences $\mathbf{x}_i$ collected sequentially in subsequent rows of the matrix $\mathbf{X}$; and application of sampling with the fractional delay of $d=T_s/2$, and $\mathbf{X}$ matrix construction from alternating strings: $\mathbf{x}_{2i-1}$ - sampled with a delay of d, and $\mathbf{x}_{2i}$ - sampled with a zero delay from the moment of the zero crossing by analyzed waveform.

Keywords: PCA method, uncertainty analysis, measurement algorithm.

1. Introduction

In the measurements using the algorithms of processing the sequences of the sampled signal values, which are obtained in the multi-channel data acquisition systems, in some cases, the method of principal component analysis (PCA) is used. The application of this method enables to decrease the influence of interference affecting the analyzed measurement signal. In [1] the PCA method has been used in the process of elimination of artifacts from the electroencephalographic signal. In the high accuracy measurements, in which data acquisition system function is performed by voltmeter with an integrating analog-to-digital converter, it can be used to decrease the influence of generator and voltmeter noises. PCA method, combined with fractional delay sampling technique was used in [2] for the estimation of the complex ratio of sinusoidal voltage amplitudes. In [3] a similar method was used to determine the harmonic components of periodic signals. The second group of application of the PCA method is reduction of results matrix dimensionality through rotation of coordinate system in order to eliminate correlation between the measurand components. Hence, in the algorithm for the estimation of impedance components by ellipse fitting method [4], the PCA was used in order to reduce the measurement uncertainty.

Principal Component Analysis enables to reduce the dimensions by linear transformation of a set of correlated variables into smaller in number collection of uncorrelated variables - the main components. The problem of dimension reduction can be solved using two methods:

– by finding a linear subspace of $P < N$ dimension, for which the sum of squared distances of the original points to the points after projection will be the smallest,
– by finding a linear subspace of $P < N$ dimension, which ensures the maximal variances of the primary data projected on the subspace.

Regardless of the method of dimension reduction, the solution is obtained by determining the eigenvectors and eigenvalues of the empirical covariance matrix or correlation matrix.

For the first method, a sought solution is $P–M$ of eigenvectors with the smallest eigenvalues, for the second - a *P*-dimensional subspace defined by *P* eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_P$ of empirical covariance matrix $\mathbf{C}$, corresponding to the *P* largest eigenvalues $\lambda_1, ..., \lambda_P$.

The empirical covariance matrix $\mathbf{C}$ for $N\times M$ dimensional data matrix $\mathbf{X}$ is defined as:

$$\mathbf{C} = \frac{1}{N}(\mathbf{X} - \frac{1}{N}\mathbf{1}_{N\times N}\mathbf{X})(\mathbf{X} - \frac{1}{N}\mathbf{1}_{N\times N}\mathbf{X})^{\mathbf{T}} \in \mathbb{R}^{N\times N}, \qquad (1)$$

where: $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix}$, and $\mathbf{1}_{N\times N}$ is $N\times N$ dimensional matrix of ones.

In (1) the most frequent way of writing the successive measurement results in the matrix $\mathbf{X}$ rows has been applied.

For each real, symmetric matrix $\mathbf{C}$, an orthogonal matrix $\mathbf{V}$ of eigenvectors exists, for which:

$$\mathbf{V}^{\mathbf{T}}\mathbf{C}\mathbf{V} = \mathbf{\Lambda} = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_N] \in \mathbb{R}^{N\times N}, \qquad (2)$$

where: $\lambda_1, \lambda_2, \ldots, \lambda_N$ are ordered ascending, the real eigenvalues of the matrix $\mathbf{C}$, and $\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathbf{T}} = \mathbf{I}_N$.

In some algorithms, instead of the covariance matrix the correlation matrix $\mathbf{R_X}$ is used:

$$\mathbf{R_X} = \mathbf{X}\mathbf{X}^{\mathbf{T}} = \begin{bmatrix} R_x(0) & R_x(1) & \cdots & R_x(N-1) \\ R_x(1) & R_x(0) & \cdots & R_x(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_x(N-1) & R_x(N-2) & \cdots & R_x(0) \end{bmatrix}. \qquad (3)$$

Matrix (3) is a circulant matrix, which can be obtained by permutations of elements from its first row $\mathbf{r}_1=[R_x(0)\ R_x(1)\ ...\ R_x(N-1)]$, hence:

$$\mathbf{R_X} = \mathbf{X}\mathbf{X}^{\mathbf{T}} = \sum_{k=1}^{N} r_{1,k}\mathbf{P}^{\mathbf{k}-1}, \qquad (4)$$

where $N\times N$ dimensional permutation matrix $\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$.

On the other hand, one can perform decomposition of the matrix (3) to components, corresponding a correlation matrix of each column of $N\times M$ dimensional data matrix $\mathbf{X}$:

$$\mathbf{R_X} = \sum_{i=1}^{M} \mathbf{R}_{\mathbf{x}i}. \qquad (5)$$

As discrete Fourier transform (DFT) of $\mathbf{X}$ matrix is equal to

$$\mathrm{DFT_N}\{\mathbf{X}\} = \mathbf{X}_\omega = \mathbf{F_N}\mathbf{X}, \qquad (6)$$

where: $\mathbf{F}_\mathrm{N} \in \mathbb{C}^{N\times N}$ is a complex Fourier matrix, for which $\mathbf{F}_\mathrm{N}^{\mathbf{H}}\mathbf{F}_\mathrm{N} = N$, and the superscript "H" denotes a Hermitian matrix (transpose conjugate).

The power spectral density function can be determined from the formula

$$\mathbf{X}_{\omega}\mathbf{X}_{\omega}^{\mathbf{H}} = \mathbf{F_N}\mathbf{X}\mathbf{X}^{\mathbf{H}}\mathbf{F_N^H} = NM\,\mathrm{diag}[P_x(0), P_x(1),..., P_x(N-1)]\,. \quad (7)$$

Due to the similarity of matrices **C** and $\mathbf{XX^H}$, the comparison of (2) and (7) shows, that the eigenvalues calculated from the formula (2) are proportional to the value of the discrete power spectral density functions (7).

The paper presents the analysis examples of influence of selected uncertainty sources on the measurand value determined in the algorithms using the method of the principal components analysis.

2. Minimization of squared distances

The PCA method has been used in the extended algorithm of ellipse fitting in order to decrease the measurement uncertainty [4]. The reduction of data dimension from the 3D space to 2D plane in this case consists in finding linear subspace with the smallest sum of squared distances of the points after projection to the original points. In the 3-channel data acquisition circuit a three sequences of samples are collected. The first are samples of voltage on the measured impedance $x_1(n)$. Secondly, the samples of voltage $x_2(n)$ proportional to the current, on the standard serial resistor. Finally, closing circuit, the voltage samples of the sinusoidal generator $x_3(n)$, which forces the flow of current in the circuit. For this kind of circuit the voltage equation can be defined as:

$$\mathbf{x}_3 = \mathbf{x}_1 + \mathbf{x}_2\,. \quad (8)$$

On account of interference and noise as well as uncertainties connected with the instrumental errors of analog-to-digital processing system, equation (8) is satisfied only approximately.

The original set of measurement data obtained in 3-channel data acquisition system has been collected in $N\times3$ dimensional matrix $\mathbf{X}=[\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3]$. As on the basis of Kirchhoff's voltage law it is known, that the set of points contained in this matrix should be on a plane, the PCA method was used. The measurement results in subsequent columns of the matrix **X** are characterized by mean values $\bar{x}_1, \bar{x}_2, \bar{x}_3$ and uncertainties, related with each of the three channels (dimensions). In the analyzed algorithm one used different than in (1) way of data matrix arrangement – successive measurement results for the particular channels have been collected in the columns of matrix **X**. Therefore, the covariance matrix $\mathbf{C}_{3D}$ is described as:

$$\mathbf{C}_{3D} = \frac{1}{N}(\mathbf{X} - \frac{1}{N}\mathbf{1}_{N\times N}\mathbf{X})^{\mathbf{T}}(\mathbf{X} - \frac{1}{N}\mathbf{1}_{N\times N}\mathbf{X}) \in \mathbb{R}^{3\times3}. \quad (9)$$

Matrix (9) can be written as:

$$\mathbf{C}_{3D} = \frac{1}{N}\mathbf{X}^{\mathbf{T}}(\mathbf{I}_N - \frac{1}{N}\mathbf{1}_{N\times N})\mathbf{X} = \frac{1}{N}\mathbf{X}^{\mathbf{T}}\mathbf{M}\mathbf{X}. \quad (10)$$

Hence, after taking (2) into account, a plane, which is searched, is described by eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, and searched points are projections of points from the plane $\mathbf{v}_1$, $\mathbf{v}_2$ onto a plane defined by the base vectors of the Cartesian system $[\mathbf{e}_1\ \mathbf{e}_2]$:

$$\mathbf{X}_{2D} = (\mathbf{MXV} + \frac{1}{N}\mathbf{1}_{N\times N}\mathbf{X})[\mathbf{e}_1\,\mathbf{e}_2], \quad (11)$$

where: $[\mathbf{e}_1\,\mathbf{e}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$.

It results from (11) that $\mathbf{X}_{2D}=\mathbf{F}(\mathbf{X}, \mathbf{V})$, wherein, on the basis of (2), $\mathbf{V}=\mathbf{G}(\mathbf{X})$. Estimation of the covariance matrix under the law of uncertainty propagation requires the designation of the corresponding derivatives of the matrix functions **F** and **G** [5]. It can be shown, that the transformation carried out did not change the mean values of primary data and the covariance data matrix on the plane $\mathbf{C}_{2D}$ is equal to

$$\mathbf{C}_{2D} = \begin{bmatrix} \mathbf{v}_1^{\mathbf{T}} \\ \mathbf{v}_2^{\mathbf{T}} \end{bmatrix}\mathbf{C}_{3D}[\mathbf{v}_1\ \mathbf{v}_2] = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \quad (12)$$

It results from the presented in [4] simulation tests that the obtained on the basis of (11) corrected values of signals connected with the tested impedance enable a double decrease uncertainty of estimates the module as well as impedance phase angle, determined by the ellipse fitting method.

3. Maximizing the variance

Diagonalization of the correlation matrix can be used for estimating the power spectral density of periodic signals. An important source of uncertainty are in this case the noise of generator and incoherent sampling. If the tested periodic signal $x(t)$ is accompanied by an additive, uncorrelated with the signal, white noise $s(t)$ with zero expected value and variance σ_s, then, for N-element set of samples $y(n)$, collected synchronously in the period T

$$y(n) = x(n) + s(n), \qquad n = 0,1,\ldots,N-1\,, \quad (13)$$

correlation matrix $\mathbf{R_y}$ in the form of (3) is obtained. It can be decomposed into two components:

$$\mathbf{R_y} = \mathbf{R_x} + \sigma_s^2\mathbf{I}_N\,. \quad (14)$$

After applying this matrix decomposition in terms of its eigenvalues, in accordance with (2)

$$\mathbf{\Lambda_y} = \mathbf{\Lambda_x} + \sigma_s^2\mathbf{I}_N \in \mathbb{R}^{N\times N}. \quad (15)$$

If data are collected synchronously in successive M periods of the analyzed periodic signal form matrix **X**, then the correlation matrix described by the equation (5) takes the form

$$\mathbf{R_X} = \sum_{i=1}^{M}(\mathbf{R}_{\mathbf{x}i} + \sigma_s^2\mathbf{I}_N) = M(\mathbf{R_x} + \sigma_s^2\mathbf{I}_N)\,. \quad (16)$$

Because at the same time

$$\mathbf{R_X} = \mathbf{V\Lambda_X V^H} = \mathbf{V}(\sum_{i=1}^{M}\mathbf{\Lambda}_{\mathbf{x}i} + \sigma_s^2\mathbf{I}_N)\mathbf{V^H} = \sum_{k=1}^{N}\sum_{i=1}^{M}(\lambda_{i,k} + \sigma_s^2)\mathbf{v}_k\mathbf{v}_k^{\mathbf{H}}\,, \quad (17)$$

then in the case of (16) $\mathbf{\Lambda_X} = M(\mathbf{\Lambda_x} + \sigma_s^2\mathbf{I}_N)$.

If data forming the matrix **X** are sampled asynchronously, the correlation function values $\mathbf{R}_{\mathbf{x}i}$, corresponding to successive columns $\mathbf{x}_i$ of the matrix **X** differ - then the general equation (17) should be applied.

Independently of the noise propagation, the phenomenon of spectral leakage is important for an asynchronous sampling. Then, depending on the selected sampling strategy, different averaging level of power density function value determined for subsequent correlation matrix $\mathbf{R}_{\mathbf{x}i}$ is obtained. Three basic techniques for samples acquisition are applied:

I. synchronization of the beginning of the data acquisition, for the successive sequences $\mathbf{x}_i$, immediately after a positive zero-level crossing of the analyzed waveform $x(t)$,

II. separation of the NM-element sequence $\mathbf{x}$ to M of N-element sequences $\mathbf{x}_i$ collected in subsequent row of the matrix $\mathbf{X}\in\mathbb{R}^{N\times M}$,

III. application of sampling with the fractional delay of $d=T_s/2$, and $\mathbf{X}$ matrix construction from alternating strings: $\mathbf{x}_{2i-1}$ - sampled with a delay of d, and $\mathbf{x}_{2i}$ - sampled with a zero delay from the moment of the zero crossing by analyzed waveform.

In order to compare the influence of the acquisition technique on the correlation matrix $\mathbf{R_X}$ form and its eigenvalues $\mathbf{\Lambda_X}$, a simulation experiment has been carried out. It was assumed that a sinusoidal signal with 1V amplitude and f frequency is analyzed. The signal is accompanied by the additive white noise with 0.05 V RMS value. For the calculation, data sequence with NM=1024 length sampled by a 16-bit A/D converter was assumed. The matrix $\mathbf{X}$ was constructed as a result of division of this sequence into M=64 sequences $\mathbf{x}_i$. For synchronous sampling, sampling frequency was $f_s=Mf$. The influence of asynchronous sampling was tested assuming that the signal period increased by 5%. In Fig. 1 and 2 correlation matrix maps for the analyzed sampling techniques was presented.

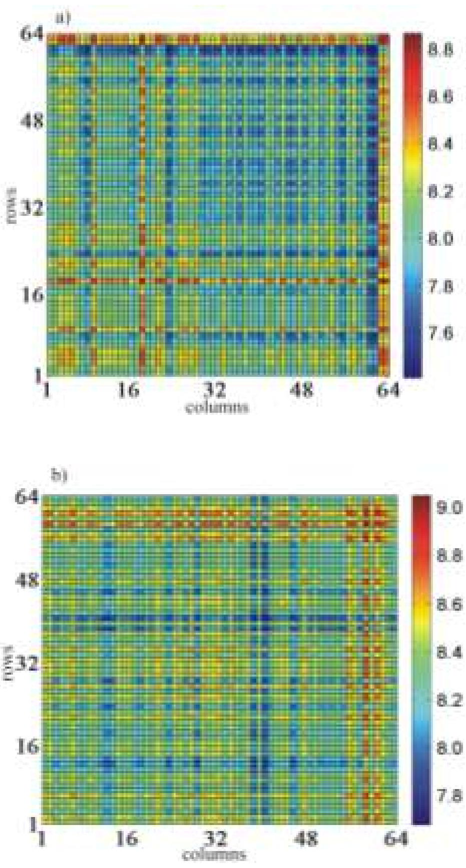

Fig. 1. Correlation matrix map: a) synchronous sampling, b) asynchronous sampling with the synchronization of the beginning of the data acquisition, for the successive sequences $\mathbf{x}_i$, immediately after a positive zero-level crossing of the analyzed waveform (I)

For comparison, the matrix map for the synchronous sampling was shown (Fig. 1a). For the simulated measurement conditions, the correlation matrix maps using synchronization technique of the beginning of sampling immediately after a positive zero-level crossing (Fig. 1b) and sampling with the fractional delay (Fig. 2b) are practically the same – the differences in values of these matrices elements do not exceed 6×10^{-3} V^2.

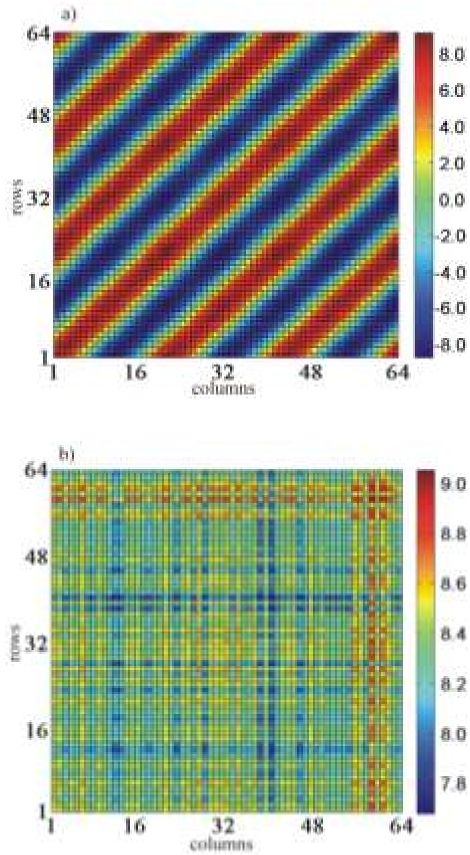

Fig. 2. Correlation matrix map: a) asynchronous sampling with the separation of the NM -element sequences $\mathbf{x}_i$ collected sequentially in the matrix rows (II), b) application of sampling with the fractional delay $d=T_s/2$ (III)

Table 1 presents values of nonzero eigenvector elements corresponding to the analyzed sampling techniques.

Tab. 1. The values of nonzero eigenvalues elements of the covariance matrix for the analyzed sampling techniques

Sampling technique	Desynchronization factor			δP
	0.01	0.05	0.1	
I	$\begin{bmatrix} 0 \\ \vdots \\ 0.5109\times10^{-3} \\ 0.5412\times10^{-3} \\ 1.0061 \end{bmatrix}$	$\begin{bmatrix} 0 \\ \vdots \\ 0.4965\times10^{-3} \\ 0.6057\times10^{-3} \\ 1.0432 \end{bmatrix}$	$\begin{bmatrix} 0 \\ \vdots \\ 0.4824\times10^{-3} \\ 0.5671\times10^{-3} \\ 1.0596 \end{bmatrix}$	4.32×10^{-3}
II	$\begin{bmatrix} 0 \\ \vdots \\ 0.6118\times10^{-3} \\ 0.4021 \\ 0.5938 \end{bmatrix}$	$\begin{bmatrix} 0 \\ \vdots \\ 0.5364\times10^{-3} \\ 0.4727 \\ 0.5339 \end{bmatrix}$	$\begin{bmatrix} 0 \\ \vdots \\ 0.6118\times10^{-3} \\ 0.4616 \\ 0.5414 \end{bmatrix}$	4.33×10^{-3}
III	$\begin{bmatrix} 0 \\ \vdots \\ 0.5122\times10^{-3} \\ 0.5403\times10^{-3} \\ 1.0061 \end{bmatrix}$	$\begin{bmatrix} 0 \\ \vdots \\ 0.4910\times10^{-3} \\ 0.6222\times10^{-3} \\ 1.0430 \end{bmatrix}$	$\begin{bmatrix} 0 \\ \vdots \\ 0.4818\times10^{-3} \\ 0.662\times10^{-3} \\ 1.0601 \end{bmatrix}$	$4,34\times10^{-3}$

For the synchronous sampling, eigenvector contains one nonzero element equal to λ_{64}=1.0000 – for the signal without noise. After adding the additive white noise along with the value λ_{64}=1.0072 eigenvalues from the range $\lambda_{49},...,\lambda_{63}=1.05\times10^{-4},..., 5.64\times10^{-4}$ appear.

Additionally, the table presents values of eigenvector elements for the desynchronization factor 0.01 and 0.1 defined as the fractional part of f_s/f ratio. Due to asynchronous sampling, in the correlation matrix spectrum additional nonzero elements appear. A relative estimation error of signal power can be estimated by calculating the ratio of the sum of K nonzero eigenvector elements resulting from the noise and the asynchronous sampling to the sum of eigenvalues resulting from the properties of the analyzed signal:

$$\delta P = \frac{\sum_{i=1}^{M}\lambda_i - \sum_{i=K+1}^{M}\lambda_i}{\sum_{i=K+1}^{M}\lambda_i}. \qquad (18)$$

4. Conclusion

The paper presents the analysis examples of the influence of the chosen uncertainty sources on the measurand determined in the algorithms using covariance matrix or data correlation matrix diagonalization. In the case of data dimensionality reduction, a reduction of measurement uncertainty value is obtained as a result of the removal of the covariance between the measured values (12). The use of the matrix of eigenvalues to estimate spectral components of the analyzed periodic signal enables averaging of the uncertainty associated with additive, uncorrelated noise, and reducing the impact of asynchronous sampling by applying appropriate sampling techniques. On the basis on the simulation tests it can be assumed that in conditions of synchronous sampling the largest eigenvalue of the correlation matrix is equal to the power of the analyzed sinusoidal signal. The results of these tests confirmed the ability to average the additive noise accompanied by the analyzed signal. Among the analyzed sampling techniques, for the analyzed values of the desynchronization factor, techniques I and III are resistant to lack of the synchronous sampling. The sampling technique II is characterized by different covariance matrix structure (Fig. 2a). Its result is the appearance of two eigenvalues – their sum corresponds to the power of the analyzed signal. A way of the division of data vector in the covariance matrix rows has a significant influence on the covariance matrix structure. The analysis of the influence of the covariance matrix structure on its eigenvalues should indicate the best sampling technique for the specific type of signal.

5. References

[1] Paszkiel S.: Wykorzystanie metody PCA i ICA do analizy sygnału EEG w kontekście usuwania zakłóceń. PAK vol. 59, no 3, pp. 204-207, 2013.

[2] Rochniński P., Kampik M.: Measurement of Amplitude Ratio and Phase Shift between Sinusoidal Voltages with Superimposed Gaussian Noise. 19th IMEKO TC-4 Symposium Measurements of Electrical Quantities, Barcelona, 2013.

[3] de Barros e Vasconcellos R. T., de Campos M. L. R.: Sampling System with Fractional Delay and PCA Applied to High-Accuracy Measurements. IEEE Trans. on Instr. Meas., vol. 62, no. 6, pp. 1407-1412, 2013.

[4] Augustyn J.: Algorytm estymacji składowych impedancji w trójkanałowym układzie akwizycji danych. PAK, vol. 59, no. 1, pp. 3-5, 2013.

[5] Evaluation of measurement data - Supplement 2 to the “Guide to the expression of uncertainty in measurement” - Extension to any number of output quantities. JCGM 102:2011.

Received: 02.08.2015 *Paper reviewed* *Accepted: 02.10.2015*

Prof. Jerzy AUGUSTYN, PhD, DSc, eng.

Graduated from the Faculty of Electrical Engineering Silesian University of Technology in 1978. At the same Faculty received the Ph.D., and D.Sc. degrees in electrical engineering in 1988, and 2008 respectively. Associate Professor in the Division of Electrical Engineering and Measurement Systems at the Faculty of Electrical Engineering, Automation and Informatics Kielce University of Technology. Main research interests: impedance measurements, measurement data processing algorithms, uncertainty in measurement systems, virtual instruments.

e-mail: j.augustyn@tu.kielce.pl