World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:1, No:10, 2007

Tests for Gaussianity of a Stationary Time Series

Adnan Al-Smadi

***Abstract*—**One of the primary uses of higher order statistics in signal processing has been for detecting and estimation of non-Gaussian signals in Gaussian noise of unknown covariance. This is motivated by the ability of higher order statistics to suppress additive Gaussian noise. In this paper, several methods to test for non-Gaussianity of a given process are presented. These methods include histogram plot, kurtosis test, and hypothesis testing using cumulants and bispectrum of the available sequence. The hypothesis testing is performed by constructing a statistic to test whether the bispectrum of the given signal is non-zero. A zero bispectrum is not a proof of Gaussianity. Hence, other tests such as the kurtosis test should be employed. Examples are given to demonstrate the performance of the presented methods.

***Keywords*—**Non-Gaussian, bispectrum, kurtosis, hypothesis testing, histogram.

I. Introduction

UNDOUBTEDLY, the most widely used model for the distribution of a random variable is the Gaussian distribution. That is because it is simple, tractable, and fairly realistic model; *i.e.*, the Gaussian process has many properties that make analytic results possible. It also describes several types of physical phenomena that are usually confirmed by experiments. Furthermore, the central limit theorem provides the mathematical justification for using the Gaussian distribution as a model for a large number of different physical phenomena in which the observed random variable is the result of a large number of individual random processes [1]. These reasons make the Gaussian process very fundamental and important in engineering and science problems.

A random process is Gaussian if every finite set of $\{y(n)\}$ is a Gaussian (Normal) random vector. Normal probabilistic distribution in many cases describes what normally happens, especially when sums of large numbers of random variables are involved. Gaussian entities are the foundation of basic stochastic signal processing. The slope of the Gaussian is the proverbial bell curve.

The Gaussian random process is known as a second order process because its probability density function (PDF) and therefore all its statistical properties are completely determined by the first and second moments; that is, by the mean and the variance which are the sole parameters of the process. Hence, the information contained in the power spectrum is essentially that which is present in the autocorrelation sequence. The first[1] and second order statistics are popular signal processing tools. These tools have been used extensively for the analysis of process data. It is a well-known fact that second order statistics are phase-blind; that is, they are able to describe minimum-phase systems only.

The second order measures work fine if the signal has a Gaussian PDF. The information contained in the second order statistics (SOS), or autocorrelation, suffices for the statistical description of Gaussian process. However, many real-life signals are non-Gaussian. For example, the electromagnetic environment encountered by receiver systems is often non-Gaussian in nature. However, the receiving systems are designed to perform in white Gaussian noise [2]. Also, acoustic noise is in many cases highly non-Gaussian. In addition, biological signals typified by electroencephalograms (EEG) or electromyograms (EMG) are non-Gaussian [13]. Hence, in practice, there are situations where we must look beyond the autocorrelation of the available data to suppress additive noise, extract phase information, or obtain information regarding deviations from Gaussianness. This necessitates the use of higher order statistics (HOS) tools. HOS techniques were first proposed over four decades ago [3, 4].

While the Gaussian random process still plays a great and significant role in stochastic signal processing, non-Gaussian random processes and HOS, or cumulants, are of increasing importance to the researchers. Higher order (≥ 3) cumulants of non-Gaussian measurements contain not only the amplitude but also phase information of the unknown system. Furthermore, they are insensitive to Gaussian noise since all higher order (≥ 3) cumulants of Gaussian random processes are equal to zero. HOS measures are extensions of second order measures to higher orders; *i.e.,* extension of autocorrelation for multiple lags. Applications of HOS have been found in diverse of fields such as speech, seismic data processing, plasma physics, optics, and economics [5]. As the field of HOS progresses, more accurate and sophisticated algorithms and techniques are revealed. In fact, the use of HOS may well be one of the new frontiers in signal processing, communications, statistical data analysis, and many other related fields.

The study of detection and estimation in non-Gaussian process is important for many applications. Examples include radars which must operate in high clutter environments and sonar systems operating in the presence of high reverberation [14].

Manuscript received September 26, 2005.
A. Al-Smadi is with the Department of Electronics Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Irbid 21163, Jordan (e-mail: smadi98@yahoo.com or smadi98@yu.edu.jo).

In this paper, several types of tests are applied to the given sequence such as histogram plot, kurtosis test, and hypothesis testing using cumulants and bispectrum of the available sequence. The basic idea for hypothesis testing for non-Gaussianity test is as follows. If the third order cumulant of a process is zero, then its bispectrum is zero. If the bispectrum is not zero, then the process is non-Gaussian. The paper is organized as follows. Section 2 contains the problem formulation for detection and estimation of non-Gaussian signals. Several examples are presented in Section 3. Finally, concluding remarks are presented in Section 4.

II. Problem Formulation

Consider the real autoregressive moving average (ARMA) stable process $y(k)$ described by the linear difference equation

$$\sum_{i=0}^{p} a_i y(k-i) = \sum_{i=0}^{q} b_i x(k-i) \tag{1}$$

where $\{x(k)\}$ is the driving process, assumed to be white with zero mean and variance σ_x^2, and k denotes the iteration. The transfer function of the process is then

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_q z^{-q}}{a_0 + a_1 z^{-1} + \cdots + a_p z^{-p}} \tag{2}$$

The coefficients $\{a_i\}$ and $\{b_i\}$ are referred to as autoregressive (AR) and moving average (MA) parameters respectively, with $a_0 = 1$ and $b_0 = 1$. Furthermore, $\underline{\theta} = [\underline{a}^T\ \underline{b}^T]^T$ represents the parameter vector of ARMA coefficients, where $\underline{a} = [a_0\ a_1\ \ldots\ a_p]^T$ and $\underline{b} = [b_0\ b_1\ \ldots\ b_q]^T$. The observed output data is represented by $y(k)$. The integers p and q are known as the model order.

Typically, the time series $y(k)$ is observed in additive noise, $d(k)$. That is

$$y_0(k) = y(k) + d(k) \tag{3}$$

The power spectrum is the frequency domain representation of the second order moment. It represents the decomposition or spread of the signal energy over the frequency channels obtained from the Fourier transform.

The k^{th}-order cumulant of $y(k)$ is defined as the joint k^{th}-order cumulant of the random variable $y(k)$, $y(k+\tau_1)$, ..., $y(k+\tau_{n-1})$ [3]. That is,

$$R_{ky}(\tau_1, \tau_2, \ldots \tau_{n-1}) = cum\{y(k), y(k+\tau_1), \ldots, y(k+\tau_{n-1})\} \tag{4}$$

which depends on the time difference $\tau_1, \tau_2, \ldots \tau_{n-1}$. The first order cumulant is the mean value while the second order cumulant is the covariance sequence. The third order cumulant of $y(k)$ is defined as [6].

$$R_{3y}(\tau_1, \tau_2) = E[y(k)y(k+\tau_1)y(k+\tau_2)] \tag{5}$$

In practice, there is only a sample sequence of the data available. Cumulants involve expectations and can not be computed in an exact manner from real data. Hence, as in case of correlation, they must be approximated. Therefore, we replace the true cumulants by their sample averages. We compute the sample estimate of $R_{3y}(\tau_1, \tau_2)$ as follows.

$$R_{3y}(\tau_1, \tau_2) = \frac{1}{N_R} \sum_{n \in R} y(k)y(k+\tau_1)y(k+\tau_2) \tag{6}$$

where N_R is the number of samples in region R. The k^{th}-order cumulant spectra $\Phi_n^y(w_1, \cdots, w_{n-1})$ of $y(k)$ is defined as the (n-1) dimensional Fourier transform of the n^{th}-order cumulant sequence [7]

$$\Phi_n^y(w_1, \cdots, w_{n-1}) = \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{n-1}=-\infty}^{\infty} R_{ny}(\tau_1, \cdots, \tau_{n-1}) \, EXP(-j[w_1\tau_1 + \cdots + w_{n-1}\tau_{n-1}] \tag{7}$$

A particular case of higher order spectra is the third-order spectrum which is known as bispectrum. The bispectrum is defined either as the two-dimensional Fourier transform of the third order cumulant sequence or as the mathematical expectation of the triple product of the Fourier coefficients [15]. That is

$$\Phi_2^y(w_1, w_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} R_{3y}(\tau_1, \tau_2) EXP(-j[w_1\tau_1 + w_2\tau_2]) \tag{8}$$

Or [8]

$$\Phi_2^y(w_1, w_2) = E\{Y(w_1)\, Y(w_2)\, Y^*(w_1+w_2)\} = \mu_3 H(w_1)\, H(w_2)\, H^*(w_1+w_2) \tag{9}$$

where $|w_1| \le \pi, |w_2| \le \pi, |w_1 + w_2| \le \pi$, $Y(w)$ is the Fourier transform of the series $y(k)$, μ_3 is the third order moment of the input, and '*' denotes the complex conjugate.

$$Y(w) = \sum_{n=-\infty}^{n=\infty} y(n) e^{-jwn} \tag{10}$$

The pair (w_1, w_2) is the bispectrum in the bifrerquency (w_1+ w_2). Therefore, if $\mu_3 \neq 0$, then $\Phi_2^y(w_1, w_2) \neq 0$.

Next, several types of tests will be applied to the given sequence. These tests include hypothesis testing using cumulants and bispectrum, the kurtosis test, and the histogram plot of the desired sequence.

A. Hypothesis Testing

Many problems in engineering and science require that we decide whether to accept or reject a statement about some parameters. Assumptions, statements, or guesses about some parameters that may or may not be true are known as hypotheses [9]. The objective is to make decisions about these statements. The procedures and techniques leading to a decision about a particular hypothesis are referred to as hypothesis testing. Hypothesis testing is a statistical decision making technique. These techniques rely on using the information in a random sample from the population of interest. If this information is consistent with the hypothesis, then the decision will be that the hypothesis is true. On the other hand, if this information is inconsistent with the hypothesis, then the decision will be that the hypothesis is false. It must be emphasized that the truth of falsity of a particular hypothesis can never be known with certainty. Hypothesis testing can be two-sided alternative or one-sided hypothesis. In either case, two statements are claimed: Ψ_0 and Ψ_1. The value Ψ_0 is referred to as the null hypothesis, while Ψ_1 is referred to as the alternative hypothesis. The decision will be either reject or fail to reject the null hypothesis Ψ_0.

In this paper, the basic idea for hypothesis testing for non-Gaussianity test is as follows. If the third order cumulant of a process is zero, then its bispectrum is zero. If the bispectrum is not zero, then the process is non-Gaussian. Thus, we have a hypothesis-testing problem for non-Gaussianity. That is we wish to test,

Ψ_0 : the bispectrum of the sequence is zero.

Ψ_1 : the bispectrum of the sequence is nonzero. (11)

A rejection of the null hypothesis implies a rejection of the hypothesis that the signal is Gaussian. It should be pointed out that a zero bispectrum is not proof of Gaussianity. That is because if a random process is symmetrically distributed, its third order cumulant is zero. For example, Uniform and Gaussian distributions are symmetric, whereas Exponential distribution is nonsymmetrical. Hence, other tools become necessary to utilize the process. Histogram plot and the kurtosis test are some of these important tools

.

B. Histogram

Histogram is widely used as a simple, but informative, method of data display. It provides a visual impression of the shape of the distribution of the measurements, as well as information about the scatter or dispersion of the data [10]. The histogram does not have the individual observations. It describes the number of times the estimator produces a given range of values. Hence, a histogram is an approximation to the PDF. Therefore, the histogram can be used as aids to selecting probabilistic model; *i.e.,* Gaussian or non-Gaussian.

C. Kurtosis

The kurtosis is an important property of the density function. It measures the degree of peakedness or flatness of a distribution. The narrower the distribution, the larger the kurtosis becomes. If a distribution has a positive kurtosis, then it is referred to as leptokurtic. If it has a negative kurtosis, it is referred to as platykurtic. If the kurtosis is 0, then the distribution is referred to as mesokurtic [11] and is a normal distribution. The kurtosis is formally given by [12]

$$Kurt(y) = \frac{E\{y^4\}}{[E\{y^2\}]^2} - 3 = \frac{E\{y^4\}}{\sigma_y^4} - 3 \tag{12}$$

where $[.]^2$ means taking the square of each element of the vector [.]. Table I shows the kurtosis for a number of common distributions.

TABLE I
THE KURTOSIS FOR SOME COMMON DISTRIBUTIONS

Distribution	Kurtosis
Exponential	6
Gaussian	0
Laplace	3
Uniform	$-\frac{6}{5}$

III. SIMULATION EXAMPLES

The purpose of the following examples is to demonstrate the performance of several techniques used in diagnosing whether a given signal is Gaussian or not. The first step in doing so is to compute the third order cumulants. Then, the bispectrum is calculated. If the bispectrum is non-zero, then the signal is non-Gaussian. However, if the bispectrum is zero, then the signal could be Gaussian but not necessarily true. To find out if the signal is Gaussian or not, we apply the kurtosis test and we display the histogram of the given data sequence.

Example* 1*: The time series to be considered is given by

$$y(k) + 0.39y(k-1) + 0.3y(k-2) = x(k) - 0.9x(k-1) \tag{13}$$

This is an ARMA model. The modeling data were generated by exciting this system with zero-mean exponential distribution. In system identification problems, the input sequence is unknown in some cases. Hence, it becomes necessary to estimate the input data. The method in [5] was used to obtain an estimate of the input data. Now, we are going to test the estimated input signal for Gaussianity. The cumulant of the sequence was calculated. The highest value of the cumulants was 288.412 on the average of 100 runs different seeds. Fig. 1 displays the cumulants. Then the bispectrum was computed. Fig. 2 displays the bispectrum of the data. Fig. 3 shows the histogram for the estimated signal. The histogram, the kurtosis, and the hypothesis testing approaches were applied. The kurtosis was calculated 100 times using different seeds. The kurtosis was about 5.632 on the average. The results of the three measures indeed show that the excitation is non-Gaussian.

Example* 2*: The same example was considered. However, the modeling data were generated by exciting the system with a Gaussian distribution. The same procedure was taken to detect whether the signal was Gaussian or non-Gaussian. The cumulant of the sequence was calculated. The highest value of the cumulants was 4.941 on the average of 100 runs different seeds. The cumulants are displayed in Fig. 4. The bispectrum is shown in Fig. 5. Since the cumulants are approximately zeros, the kurtosis was obtained and the histogram was plotted. The kurtosis was calculated 100 times using different seeds. The kurtosis was about 0.0004 on the average. The histogram is displayed in Fig. 6. The results of the three measures indeed show that the excitation is Gaussian.

Example 3: The model is the same as in Example 1; i.e, the exciting signal is a zero-mean, exponentially distributed. However, the output of the filter was corrupted with colored additive Gaussian noise at the same SNR on the output sequence. The colored Gaussian noise $d(k)$ was obtained by passing a zero-mean random white Gaussian distribution through the following *sinc* function.

$$h(k) = 0.3sinc(0.01k) \qquad -5 \leq k \leq 5 \qquad (14)$$

Again, we estimated the input signal and tested it for Gaussianity. The cumulant of the sequence was calculated. The highest value of the cumulants was 4.846 on the average of 100 runs different seeds. The cumulants are displayed in Fig. 7. The bispectrum is shown in Fig. 8. The kurtosis was about 5.6424 on the average. Fig. 9 displays the histogram of the data. The results of the three measures indeed show that the excitation is non-Gaussian.

IV. Conclusion

We have considered estimation and detection of Gaussian signals in non-Gaussian processes of unknown PDF. Several methods to test for Gaussianity of a time series have been presented. Namely, hypothesis testing using HOS cumulants, kurtosis test, and histogram plot are used. We have exploited the fact that cumulants of non-Gaussian processes may be estimated in the presence of additive Gaussian signals of unknown covariance. It should be pointed out that a zero bispectrum is not proof of Gaussianity. While the histogram provides a visual impression of the shape of the distribution of the measurements, the kurtosis measures the degree of peakedness or flatness of a distribution. Simulation examples were presented to demonstrate the performance of these methods.

References

[1] A. Al-Smadi and M. Smadi," Study of the reliability of a binary symmetric channel under non-Gaussian disturbances," *International Journal of Communication Systems*, vol. 16, no. 10, pp. 865-973, December 2003.

[2] S. Zabin and D. Furbeck, "Efficient identification of non-Gaussian mixtures," *IEEE Trans. Comm*., vol. 48, pp. 106-117, January 2000.

[3] J. M Mendel," Tutorial on higher order-statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, pp. 278-305, March 1991.

[4] C.L. Nikias and M.R. Raghveer, "Bispectrum estimation: A digital signal processing framework," *Proceedings of the IEEE*, vol.75, no. 7, pp. 869-891, July, 1987.

[5] A. Al-Smadi & D. M. Wilkes, "Robust and Accurate ARX and ARMA Model Order Estimation of Non-Gaussian Processes", *IEEE Trans. On Signal Processing*, vol. 50, no. 3, pp. 759 –763, March 2002.

[6] A. Al-Smadi, "Cumulant based approach to FIR system identification", *International Journal of Circuit Theory and Applications*, vol. 31, no. 6, pp. 625-636, November, 2003.

[7] D.R. Brillinger, "An Introduction to Polyspectra," *Ann. Math. Statist.*, vol. 36, pp. 1351-1374, 1965.

[8] M.Choudhury, S. Shah, and N. Thornhill, "Diagnosis of poor control-loop performance using higher order statistics," *Automatica*, vol. 40, pp. 1719-1728, 2004.

[9] D. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, New York, 2000.

[10] D.C. Montgomery and G. C. Runger, *Applied Statistics and probability for Engineers*, John Wiley, New York, 1999.

[11] R. Pond, *Fundamentals of Statistical Quality Control*, Macmillan College Publishing, New York, 1994.

[12] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Mag.,* vol. 13, no. 5, pp. 24-42, 1996.

[13] Y.Nishiguchi, N. Toda, and S. Usui, "Parametric estimation of Higher-order spectra by tensor product expansion with coordinate transformation," *Electronics and Communications in Japan*, vol. 87, no. 1, pp. 75-83, 2004.

[14] S.A. Kassam, *Signal detection in non-Gaussian noise*, Springer-Verlag, New York, 1988.

[15] J. Caillec and R. Garello, "Comparison of statistical indices using third order statistics for nonlinearity detection," Signal Processing, vol. 84, pp. 499-525, 2004.

Adnan Al-Smadi received B.S. degree (*Magna Cum Laude*) and the M.S. degree in electrical engineering from Tennessee State University, Nashville, TN, USA in 1987 and 1990, respectively. He received the Ph.D. degree in electrical engineering from Vanderbilt University, Nashville, TN, in 1995. From 1989 to 1991, he was an instructor of Mathematics at Tennessee State University, and from 1991 to 1992, he was an instructor of Mathematics at Fisk University, Nashville, TN, USA. From 1992 to 1993, he was an instructor of Electrical Engineering at Tennessee State University. From 1993 to 1996, he was an Assistant Professor of Aeronautical and Industrial Technology at Tennessee State University. In addition, from 1995 to 1997, he was an adjunct Assistant Professor of Electrical Engineering at Vanderbilt University. From 1996 to 1997, he served as Interim Department Head of the Aeronautical and Industrial Technology Department at Tennessee State University. From August 1997 to July 2002, he was an Assistant Professor of Electronics Engineering at Yarmouk University, Jordan and since July 2002 he has been an Associate Professor of Electronics Engineering at Yarmouk University, Jordan. From September 2002 to September 2004, he served as the chairman of the Department of Electronics Engineering at Yarmouk University. Dr. Al-Smadi has been the Dean of Hijjawi Faculty for Engineering Technology since September 2004. His research interests include digital signal processing, system identification and modeling, spectrum estimation, communication systems, wavelets, and higher order statistics.

Dr. Al-Smadi received the Research of Excellence Award for the Distinguished Researchers form the Higher Education & Scientific Research Ministry-Jordan in August 2004. He is a member of Eta Kappa Nu in the USA.

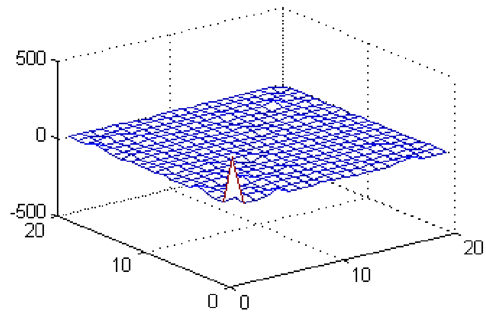

Fig. 1 Cumulants for the signal in Example 1

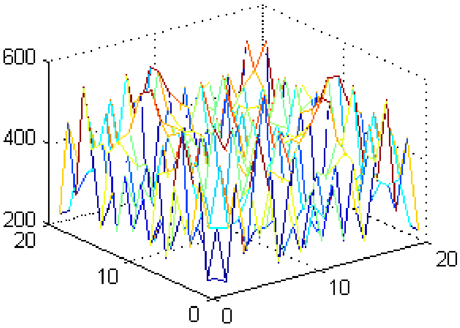

Fig. 2 Bispectrum for the signal in Example 1

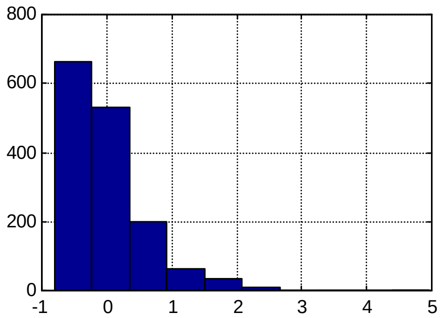

Fig. 3 Histogram for the signal in Example 1

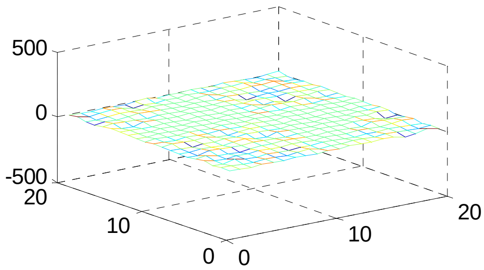

Fig. 4 Cumulants for the signal in Example 2

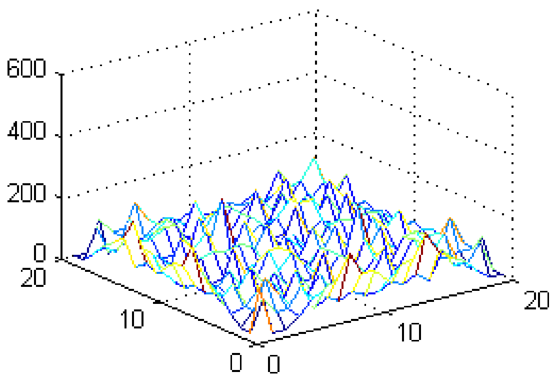

Fig. 5 Bispectrum for the signal in Example 2

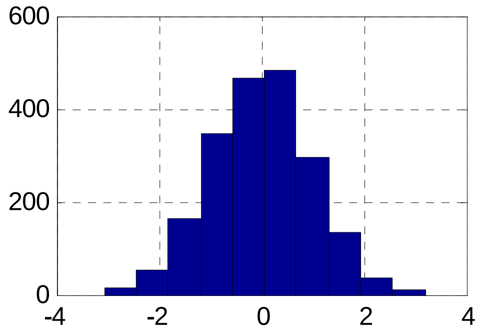

Fig. 6 Histogram for the signal in Example 2

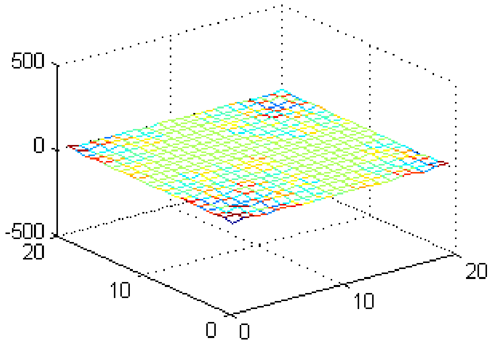

Fig. 7 Cumulants for the signal in Example 3

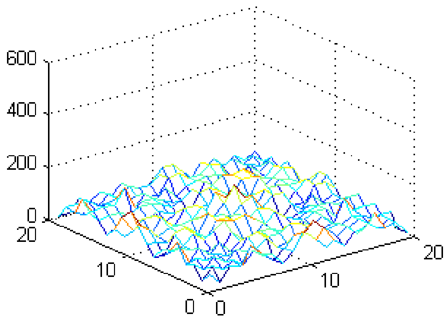

Fig. 8 Bispectrum for the signal in Example 3

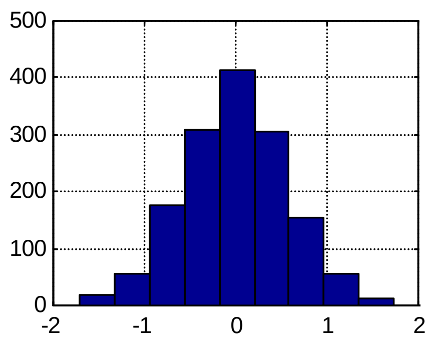

Fig. 9 Histogram for the signal in Example 3