A Non-Parametric Test for Detecting the Complex-Valued Nature of Time Series

Temujin Gautama[1], Danilo P. Mandic[2], and Marc M. Van Hulle[1] *

[1] Laboratorium voor Neuro- en Psychofysiologie
K.U.Leuven, Campus Gasthuisberg, Herestraat 49
B-3000 Leuven, BELGIUM
Email: {temu,marc}@neuro.kuleuven.ac.be
[2] Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
Exhibition Road, SW7 2BT, London, UK
Email: d.mandic@ic.ac.uk

Abstract. The emergence of complex-valued signals in natural sciences and engineering has been highlighted in the open literature, and in cases where the signals have complex-valued representations, the complex-valued approach is likely to exhibit advantages over the more convenient real-valued bivariate one. It remains unclear, however, whether and when the complex-valued approach should be preferred over the bivariate one, thus, clearly indicating the need for a criterion that addresses this issue. To this cause, we propose a statistical test, based on the local predictability in the complex-valued phase space, which discriminates between the bivariate or complex-valued nature of time series. This is achieved in the well-established surrogate data framework. Results on bothe the becnhmark and real-work IPIX complex radar data support the approach.

1 Introduction

Recently, the use of complex-valued signals has shown many advantages over real-valued bivariate ones, and are an increasingly popular topic in many branches of physics and DSP. Consequently, considerable research effort has been devoted to the extension of nonlinear modelling and filtering approaches towards complex-valued signals [2–4], the applicability of which has been demonstrated,

* T.G. was supported by a scholarship from the Flemish Regional Ministry of Education (GOA 2000/11) and a research grant from the Fund for Scientific Research (G.0248.03). D.P.M. was supported by QLG3-CT-2000-30161 and a visiting fellowship from the K.U.Leuven (F/01/079) while at the Laboratorium voor Neuro- & Psychofysiologie, K.U.Leuven, Belgium. M.M.V.H. was supported by research grants received from the Fund for Scientific Research (G.0185.96N; G.0248.03), the National Lottery (Belgium) (9.0185.96), the Flemish Regional Ministry of Education (Belgium) (GOA 95/99-06; 2000/11), the Flemish Ministry for Science and Technology (VIS/98/012), and the European Commission, 5th framework programme (QLG3-CT-2000-30161 and IST-2001-32114).

among others in radar, sonar and phase-only DSP. The surrogate data method, as originally proposed by Theiler *et al.* [1], has evolved into a standard technique to test for the presence of nonlinearity in a real-valued time series. In the field of nonlinear signal processing, such tests are indispensable, since, in principle, signal nonlinearity should be assessed *prior* to the utilisation of nonlinear models, the parameters of which are more mathematically involved to determine than those of linear models. A reliable, statistical test for assessing the complex-valued nature of a signal, however, is still lacking.

To that cause, we extend the iterative Amplitude Adjusted Fourier Transform (iAAFT) approach [5] toward complex-valued signals (Section 2). A null hypothesis of a complex-valued linear system underlying the time series under study i utilised. Next, a novel methodology is proposed for characterising (Section 3), and statistically testing (Section 4) the complex-valued nature of a time series. Simulations which support the analysis are performed on both the benchmark and real-world complex valued data.

2 Surrogate Data

The surrogate data method computes test statistics on the original time series and a number of so-called 'surrogates', which are realisations of a certain null hypothesis, H_0. These are further used for bootstrapping the distribution of the test statistic under the assumption of H_0. In this section, a surrogate data generation procedure known as the (real-valued) iterative Amplitude Adjusted Fourier Transform (iAAFT) method [5] is shortly introduced, after which an extension of this method towards complex-valued signals is proposed.

2.1 Real-Valued iAAFT Method

The iAAFT method generates a surrogate for a real-valued (univariate) time series under the null hypothesis that the original time series is generated by a Gaussian linear process, followed by a, possibly nonlinear, static (memoryless) observation function, $h(\cdot)$. The surrogates have their signal distributions identical to that of the original signal, and amplitude spectra that are approximately identical, or *vice versa*. Let $\{|S_k|\}$ be the Fourier amplitude spectrum of the original time series, s, and $\{c_k\}$ the (signal value) sorted version of the original time series. Note that k denotes the frequency index for the amplitude spectrum, whereas for a time series, it denotes the time index. At every iteration j of the algorithm, two time series are calculated, namely $r^{(j)}$, which has an signal distribution identical to that of the original, and $s^{(j)}$, which has an amplitude spectrum identical to the original. The iAAFT iteration starts with $r^{(0)}$ a random permutation of the time samples:

Repeat:
1. compute the phase spectrum of $r^{(j-1)} \rightarrow \{\phi_k\}$
2. compute $s^{(j)}$ as the inverse transform of $\{|S_k| \exp(i\phi_k)\}$

3. compute $r^{(j)}$ by rank-ordering $s^{(j)}$ to match $\{c_k\}$, *i.e.*, sort $\{s_k^{(j)}\}$ in ascending order and set $r_k^{(j)} = c_{\text{rank}(s_k^{(j)})}$

Until error convergence

The modelling error can be quantified as the mean-square-error (MSE) between $\{|S_k|\}$ and the amplitude spectrum of $r^{(j)}$. The algorithm was extended towards the multivariate case in [5], yielding surrogates that retain not only the amplitude spectra of the variates separately, but also the cross-correlation spectrum. This was done by modifying the phase adjustment step (step 1): the cross-correlation between the variates can be retained if the relative phases between the frequency components remains intact. For details we refer to [5]. Figure 1B shows a real-valued bivariate iAAFT realisation of the Ikeda Map (shown in Fig. 1A).

2.2 Complex-Valued iAAFT Method

A straightforward extension of the univariate iAAFT-method towards complex-valued signals would be if the desired amplitude spectrum is replaced by the amplitude spectrum of the original complex-valued signal. In the next step, the desired signal distribution needs to be imposed on the surrogate in the time domain (step 3 in the iAAFT-procedure). This can be achieved by applying the rank-ordering procedure to the real and imaginary parts of the signal separately. However, in practice for complex-valued signals, it is more important to impose equal empirical distributions on the moduli of the complex-valued samples, rather than on the real and imaginary parts separately. Therefore, we subsequently perform a rank-ordering procedure on the moduli, so as to match the moduli of the original time series. The underlying null hypothesis is that the time series is generated by a linear complex-valued process, driven by Gaussian white noise, followed by a (possibly nonlinear) static observation function, $h(\cdot)$, which operates on the *moduli* of the complex-valued time samples. We propose the following complex-valued iAAFT (CiAAFT) procedure, using the same conventions as in the iAAFT case, namely $\{|S_k|\}$ is the Fourier amplitude spectrum of the original time series, $\{c_k\}$ is the modulus sorted version of the time series, $r^{(j)}$ and $s^{(j)}$ are time series at iteration j with a modulus distribution, respectively, amplitude spectrum identical to the original time series:

Repeat:

1. compute the phase spectrum of $r^{(j-1)} \rightarrow \{\phi_k\}$
2. compute $s^{(j)}$ as the inverse transform of $\{|S_k| \exp(i\phi_k)\}$
3. rank-order the real and imaginary parts of $r^{(j)}$ to match the real and imaginary parts of $\{c_k\}$
4. rank-order the moduli of $r^{(j)}$ to match the modulus distribution of $\{c_k\}$

Until error convergence

The iteration is started with $r^{(0)}$ a random permutation of the complex-valued time samples. Convergence can be monitored as the MSE computed between $\{|S_k|\}$ and the amplitude spectrum of $r^{(j)}$. Simulations suggest that the iteration can be terminated when the MSE decrement is smaller than 10^{-5}, which typically occurs after fewer than 100 iterations. Figure 1C shows a CiAAFT realisation of the Ikeda Map, for which the error curve is shown in Fig. 2A.

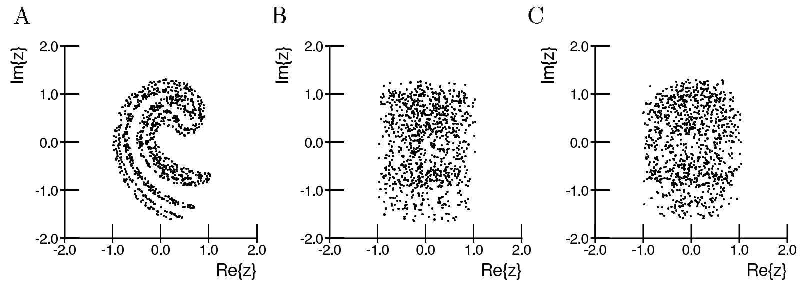

Fig. 1. Realisation of the Ikeda Map (A), iAAFT (B) and CiAAFT (C) surrogates.

3 Delay Vector Variance Method

We have used a complex-valued variant of the Delay Vector Variance (DVV) method [6] for characterising the time series based on its local predictability in phase space over different scales. For a given embedding dimension m and resulting time delay embedding representation (*i.e.*, a set of delay vectors (DV), $\mathbf{x}(k) = [x_{k-m}, \ldots, x_{k-1}]^T$), a measure of unpredictability, $\sigma^{*2}(r_\mathrm{d})$, is computed for a standardised range of degrees of locality, r_d:

- The mean, μ_d, and standard deviation, σ_d, are computed over all pairwise Euclidean distances between DVs, $\|\mathbf{x}(i) - \mathbf{x}(j)\| = \sqrt{\sum_{n=1}^{m} |x_{i-n} - x_{j-n}|^2}$ $(i \neq j)$.
- The sets $\Omega_k(r_\mathrm{d})$ are generated such that $\Omega_k(r_\mathrm{d}) = \{\mathbf{x}(i) | \, \|\mathbf{x}(k) - \mathbf{x}(i)\| \leq r_d\}$. The range r_d is taken from the interval $[\max\{0, \mu_\mathrm{d} - n_\mathrm{d}\sigma_\mathrm{d}\}; \mu_\mathrm{d} + n_\mathrm{d}\sigma_\mathrm{d}]$, *e.g.*, uniformly spaced, where n_d is a parameter controlling the span over which to perform the DVV analysis.
- For every set $\Omega_k(r_\mathrm{d})$, the variance of the corresponding targets, $\sigma_k^2(r_\mathrm{d})$, is computed as the sum of the variances of the real and imaginary parts. The average over all sets $\Omega_k(r_\mathrm{d})$, normalised by the variance of the time series, σ_x^2, yields the 'target variance', $\sigma^{*2}(r_\mathrm{d})$:

$$\sigma^{*2}(r_\mathrm{d}) = \frac{\frac{1}{N}\sum_{k=1}^{N} \sigma_k^2(r_\mathrm{d})}{\sigma_x^2}. \tag{1}$$

Note that the computation of the Euclidean distance between complex-valued DVs is equivalent to considering real and imaginary parts as separate dimensions. Since for bivariate time series, a delay vector is generated by concatenating time delay embedded versions of the two dimensions, the complex-valued and bivariate versions of the DVV method are equivalent, and can be conveniently compared when the variance of a bivariate variable is computed as the sum of the variances of each variate.

A DVV plot, $\mathcal{D}$, is obtained by plotting the target variance, $\sigma^{*2}(r_\mathrm{d})$, as a function of the standardised distance, $\frac{r_d - \mu_d}{\sigma_d}$. The DVV plots for a 1000 sample realisation of the Ikeda Map, $\mathcal{D}$, and the two types of surrogates, $\mathcal{D}^\mathrm{b}$ and $\mathcal{D}^\mathrm{c}$, generated using the iAAFT and CiAAFT method, respectively, are shown in Fig. 2B, using $m = 3$ and $n_d = 3$.

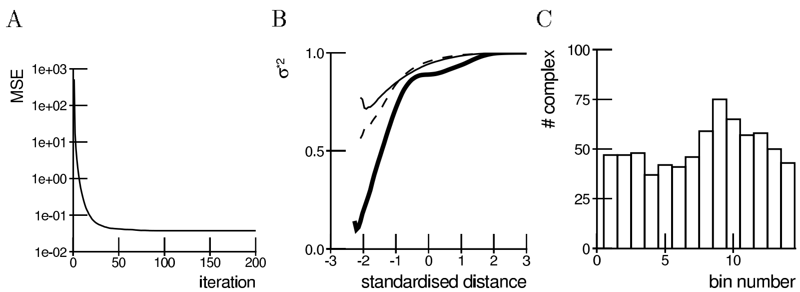

Fig. 2. A) Convergence of the CiAAFT algorithm; B) DVV plots for the Ikeda Map (thick solid), for an iAAFT (thin dashed) and a CiAAFT surrogate (thin solid). C) Number of time series that were judged complex-valued by the proposed method, for every rangebin in the IPIX radar data set.

4 Statistical Testing

In the framework of surrogate data testing, as introduced by Theiler *et al.* [1], a time series is characterised by a certain test statistic, which is compared to an empirical distribution of test statistics, computed for a number of surrogates generated under the assumption of a null hypothesis. A significant difference between the two then indicates that the null hypothesis can be rejected. In the CiAAFT case, a rejection of the null hypothesis that the signal is complex-valued and linear, could be due to a deviation from either of the two properties. Therefore, we propose a different approach: rather than comparing the original time series to the surrogates, we compare surrogates generated under different null hypotheses, namely that of a linear and bivariate time series, H_0^{b}, and that of a linear and complex-valued time series, H_0^{c}. The respective surrogates are generated using the bivariate iAAFT [5] and the proposed CiAAFT method. All time series are characterised using the DVV method, and a significant difference between the two sets of surrogates is an indication that the original time series is complex-valued. The proposed methodology is the following.

1. Generate $N_{\mathrm{s,ref}}$ CiAAFT surrogates and the average DVV plot $\rightarrow \mathcal{D}^0$;
2. Generate N_{s} iAAFT surrogates and corresponding DVV plots $\rightarrow \{\mathcal{D}^{\mathrm{b}}\}$;
3. Generate N_{s} CiAAFT surrogates and corresponding DVV plots $\rightarrow \{\mathcal{D}^{\mathrm{c}}\}$;
4. Compare $(\mathcal{D}^0 - \{\mathcal{D}^{\mathrm{b}}\})$ and $(\mathcal{D}^0 - \{\mathcal{D}^{\mathrm{c}}\})$.

To perform the final step in a statistical manner, the (cumulative) empirical distributions of root-mean-square distances between $\{\mathcal{D}^{\mathrm{b}}\}$ and $\mathcal{D}^0$, and between $\{\mathcal{D}^{\mathrm{c}}\}$ and $\mathcal{D}^0$, are compared using a Kolmogorov-Smirnoff (K-S) test. This way, the different types of *linearisations* (bivariate, $\{\mathcal{D}^{\mathrm{b}}\}$, and complex-valued, $\{\mathcal{D}^{\mathrm{c}}\}$) are compared to the 'reference' linearisation given a complex-valued nature of the time series, $\mathcal{D}^0$. The distributions of the test statistics (the root-mean-square distances) under the different null hypotheses are, in fact, bootstrapped using the proposed methodology. If the two distributions of test statistics are significantly different at a certain level α, the original time series is complex-valued. Note that assumptions regarding the possible nonlinearity of the signal are avoided.

5 Simulations

5.1 Synthetic Time Series

The proposed algorithm was tested on five sets of synthetically generated benchmark signals containing $N = 1000$ time samples. The two linear sets contained time series 1) consisting of time samples that are drawn from a bivariate Gaussian distribution, $\mathcal{N}([0,0],[1,2])$, rotated over an angle of $\frac{\pi}{3}$ (**linear bivariate**, "LB"), and 2) generated by considering a Gaussian 'amplitude spectrum', adding random phase and computing the inverse FFT (**linear complex**, "LC"). The two nonlinear sets were generated by a nonlinear system described in [7]:

$$y_k = \gamma \, \frac{x_{k-1}\, x_{k-2}\, (x_{k-1} + 2.5)}{1 + x_{k-1}^2 + x_{k-2}^2} + x_k, \tag{2}$$

where γ is a parameter controlling the prevalence of the nonlinear over the linear part of the signal, which was set to $\gamma = 0.6$, unless stated otherwise. In the first nonlinear set (**nonlinear bivariate**, "NLB"), both dimensions of "LB" were separately passed through the nonlinear system, and in the second set (**nonlinear complex**, "NLC"), x represents the complex-valued time series "LC". The final set contained realisations of the Ikeda Map (an example is shown in Fig. 1A).

For each of the five sets, 100 realisations of time series were generated, to each of which the proposed test was applied. For the bivariate sets (LB and NLB), the number of (erroneous) rejections were of the order expected from the $\alpha = 0.05$ level (5/100 and 1/100). The proposed test did not perform well on the LC set: only 16/100 the time series were correctly judged to be complex-valued. However, this is not surprising, since any linear complex-valued system has a bivariate equivalent, though not *vice versa*. Consequently, the iAAFT method can represent these time series equally well as the CiAAFT method. For the NLC set, the proposed test correctly judged the time series to be complex-valued in 62/100 cases (the performance increased to 79/100 with $\gamma = 1.0$), and in all of the Ikeda Map realisations.

5.2 Radar Data

We further considered real-world data taken from in-phase and quadrature components from the Dartmouth 1993 IPIX radar data, which is publicly available (http://soma.crl.mcmaster.ca/ipix). We have arbitrarily selected data set #17, which was recorded during a higher sea state, with the waves moving away from the radar. It consisted of 14 rangebins containing a time series of 131,072 complex-valued samples. In the ninth rangebin (and, to a lesser degree in rangebins 8, 10 and 11), a target was present. The remaining bins only contained so-called '*sea clutter*', *i.e.*, radar backscatter from the ocean surface. From every bin, we considered time segments of $N = 1000$ samples (one second), and generated 100 non-overlapping time segments, on each of which the proposed test was applied.

The number of time series which were judged to be complex-valued, are shown in Fig. 2C for every bin. On average, 51/100 time series in every bin were found to be complex-valued, and there were stronger indications of a complex-valued nature in those bins in which a target was present (bins 8–11, but also in bin 12). The increased complex-valued nature in the presence of a target was consistent over different data sets from the same database (results not shown).

6 Conclusions

We have introduced a novel methodology for statistically testing whether or not the processing of a bivariate time series could benefit from a complex-valued representation. We have proposed a novel procedure, the Complex iterative Amplitude Adjusted Fourier Transform (CiAAFT) method, for generating surrogate time series under the null hypothesis of a linear and complex-valued system underlying the time series. Consequently, surrogates generated using the traditional iAAFT method for bivariate time series can be compared to those generated using the CiAAFT method. Both types of surrogates have been characterised using a complex-valued extension of the Delay Vector Variance (DVV) method, allowing for a statistical comparison between the two types of surrogates. If the difference is significant, the time series is judged complex-valued, and it is judged bivariate otherwise.

The methodology was validated on synthetically generated time series, and applied to real-world data obtained from the IPIX radar. The latter data has been frequently addressed in the open literature (for an overview, see [8]), and it has been shown that short time segments can be modelled adequately by a complex-valued autoregressive (AR) model. It was demonstrated that 50 % of the time series from the radar data showed a complex-valued nature, and, furthermore, that this proportion increased in the presence of a target.

References

1. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.: Testing for nonlinearity in time series: The method of surrogate data. Physica D **58** (1992) 77–94
2. Leung, H., Haykin, S.: The complex backpropagation algorithm. IEEE Trans. Signal Processing **39** (1991) 2101–2104
3. Nitta, T.: An analysis of the fundemental structure of complex-valued neurons. Neural Processing Letters **12** (2000) 239–246
4. Kim, T., Adali, T.: Universal approximation of fully complex feed-forward neural networks. In: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP). (2002)
5. Schreiber, T., Schmitz, A.: Surrogate time series. Physica D **142** (2000) 346–382
6. Gautama, T., Mandic, D., Van Hulle, M.: Indications of nonlinear structures in brain electrical activity. Phys. Rev. E (2003) in press.
7. Narendra, K., Parthasarathy, K.: Identification and control of dynamical systems using neural networks. IEEE Trans. Neural Networks **1** (1990) 4–27
8. Haykin, S., Bakker, R., Currie, B.: Uncovering nonlinear dynamics: The case study of sea clutter. In: Proc. of IEEE, Special Issue. (2002) submitted