Random matrix approach in search for weak signals hidden in noisy data

D. Grech\(^{(a)}\) and J. Miśkiewicz\(^{(a,b)}\)

\(^1\) Institute of Theoretical Physics, University of Wroclaw - pl. M. Borna 9, PL-50-204 Wroclaw, Poland, EU
\(^2\) Department of Physics and Biophysics, Wroclaw University of Environmental and Life Sciences ul. Norwida 25/27, PL-50-375 Wroclaw, Poland, EU

received 7 October 2011; accepted in final form 22 December 2011
published online 6 February 2012

PACS 05.45.Tp – Time series analysis
PACS 95.75.Pq – Mathematical procedures and computer techniques
PACS 95.55.Ym – Gravitational radiation detectors; mass spectrometers; and other instrumentation and techniques

Abstract – We present new, original and alternative method for searching signals coded in noisy data. The method is based on the properties of random matrix eigenvalue density. First, we describe general ideas and support them with results of numerical simulations for basic periodic signals hidden in artificial stochastic noise. Then, the main effort is put to examine the strength of a new method in investigation of data content taken from the real astrophysical NAUTILUS detector designed to search for gravitational waves. Our method discovers some previously unknown problems with data aggregation in this experiment. We provide also the results of new method applied to the entire respond signal from ground-based bar detectors in future experimental activities with reduced background noise level. A good performance of our method is indicated what makes it a positive predictor for further applications in many areas.

One often encounters a problem if in the given noisy experimental data some repeatable weak signal is hidden. If such signal is very weak, even its existence in stochastic background data is difficult to confirm, not mentioning the challenge to describe its detailed properties. This problem is relevant not only in detector data analysis in many fields of physics but it is also important in other areas of human activity where data transmission, data collection and data processing is involved — just to mention: telecommunication, electronics, computer science, genetics, acoustics, astronomy, etc.

The search for long-memory effects in the measured background noise is essential in proper determination of the conceivable periodic or just repeatable signal immersed in this noise. Such search is also the first step to reveal and then to separate other sources of long-memory data which, if added to the measured background noise, may finally change its nature.

In this report we introduce the novel method based on random matrix theory (RMT) approach to examine the presence of long-memory or autocorrelations in time series. Our method uses properties of random matrices ensemble average eigenvalue density [1–3], called shortly eigenvalue spectrum throughout this paper. The RMT approach is applied here in the new context and uses a link between time series and random matrices indicated in different approach in [4].

Usually, one applies RMT to multidimensional time series of data, seeking for correlations between various 1-dimensional subsseries of these time series (see, e.g., [5–9]). This technique analyzes eigenvalues of the correlation matrix between 1-dimensional subsseries with the corresponding eigenvalues of Wishart matrix [10,11]. In our approach, we propose to analyze eigenvalue spectrum for matrix entries built entirely from increments of 1-dimensional time series with no correspondence to Wishart correlation matrix. The construction of our matrix entries runs as follows.

Let \(\{x_i\}, (i = 1, \ldots, N + 1)\) is the 1-dimensional cumulated time series with very large \(N\). The time series of its increments is thus \(\{\Delta x_i\}\) with \(\Delta x_i = x_{i+1} - x_i\). We divide \(\{\Delta x_i\}\) into \([N/L]\) ([\(\cdot\)] symbol stands for the integer part) non-overlapping subsseries \(s_k = \{\Delta x_{(k-1)L+1}, \ldots, \Delta x_{kL}\}\) of length \(L\) each \((k = 1, \ldots, [N/L])\). It is assumed \(N \gg L^2\). Every subsseries is then renormalized according to RMT

\(^{(a)}\)E-mail: dgrech@ift.uni.wroc.pl
\(^{(b)}\)E-mail: jamis@ift.uni.wroc.pl
rules:
\[ s_k \rightarrow \hat{s}_k = \frac{s_k}{\sqrt{L}} \sigma_k, \]  
(1)

where \( \sigma_k \) is the standard deviation of elements in \( k \)-th subseries.

We build \( L \times L \) matrices from \( \hat{s}_k \) subseries treating them as subsequent matrix rows. The first \( L \) subseries (containing \( L^2 \) data) build the first matrix, the second \( L \) subseries build the second matrix, etc. This way, the \((i, j)\) entry of the \( n \)-th matrix \( M^{(n)}_{ij} \) \((n = 1, \ldots, [N/L^2])\) reads in terms of time series increments as follows:
\[ M^{(n)}_{ij} = \frac{\Delta x_{(((n-1)L+i-1)L+j)} \sqrt{\sigma_{(n-1)L+i}}} \sqrt{\sigma_{(n-1)L+j}}. \]  
(2)

Further steps rely on examination of eigenvalue spectrum properties for the ensemble of symmetrized matrices with \((i, j)\) entries \(1/\sqrt{2}(M^{(n)}_{ij} + M^{(n)}_{ji})\) built this way and on comparison of these properties with some a priori known spectrum patterns. The eigenvalues have been calculated by us numerically within Octave 3.2.4 software. Any distortion of the a priori known long-range dependence in the background time series data from the pattern will be classified as the presence of a new signal hidden in these data. The measurement of this distortion in terms of the averaged eigenvalue spectrum for matrices built according to the scheme presented above, leads not only to discovery of a new immersed signal but may give us also some information on its parameters.

Let us look at the beginning, how this scheme works in the simplest case of white noise chosen as the background, enriched with the sinusoidal discrete signal \( s(t) = A_s \sin(2\pi t/T) \) \((t = 1, 2, \ldots, N)\) of period \( T \) added at various signal-to-noise ratios \( s/n \) (amplitude) level. Throughout this paper \( s/n \) has the meaning of ratio between the amplitude \( A_s \) of the signal and the standard deviation \( \sigma_0 \) of the background noise. If \( s/n = 0 \), we simply should reconstruct the Wigner semicircle for the ensemble averaged level density [12]. The crucial observation is that eigenvalue spectrum \( \rho(\lambda) \) (i.e. the probability density function of finding the particular eigenvalue \( \lambda \) in the matrix eigenvalue spectrum normalized according to eq. (1)) obtained from non-correlated data should disappear for \(|\lambda| > 2\) because of Wigner semicircle formula:
\[ \rho(\lambda) = \left\{ \begin{array}{ll} \frac{1}{\pi} \sqrt{4 - \lambda^2}, & \text{for } |\lambda| \leq 2, \\ 0, & \text{for } |\lambda| > 2. \end{array} \right. \]  
(3)

The above relation describes analytically the eigenvalue spectrum of symmetric \( N \times N \) real matrix with entries drawn from normal distribution (so called GOE ensemble) in the limit \( N \rightarrow \infty \). This fact is confirmed numerically in fig. 1 for the ensemble of \( 10^3 \) matrices of size \( 200 \times 200 \) constructed from the pure white noise signal. Increasing \( s/n \) ratio we see the tails of \( \rho(\lambda) \) spectrum becoming much longer. Eventually, they exceed the Wigner limit \(|\lambda| \leq 2\), indicating autocorrelations in data. We have checked also that obtained results do not qualitatively depend on period \( T \) of the added sinusoidal signal.

Let us turn to analyze an example of much more complicated but very realistic noisy data being of wide interest in physics and astrophysics. We examined the background noise data from NAUTILUS experiment [13]. It is one of ground-placed detectors constructed to detect bursts of gravitational radiation coming from sources like spinning irregular neutron stars located in our Galaxy or in the Local Group. The idea of NAUTILUS and similar bar detectors was based on the relative change in macroscopic body sizes (of cylindrical shape for the NAUTILUS experiment) while the hypothetical gravitational wave is passing through it. The relative change in sizes (dimensionless amplitude) up to \( 10^{-20} \) was expected to be observed by the NAUTILUS group. The experiment is no longer running since more modern approaches started data collection like VIRGO, LIGO or are planned (LISA). Nevertheless, the use of data from NAUTILUS seems to be important to show the efficiency of RMT-based method in comparison with the standard Fourier analysis most often applied to examine such complicated signal. The latter analysis is usually the tool of the first choice in examination of periodic signal hidden in background data.

We took 10 different samples of signals, every sample containing around \( 4 \times 10^5 \) recorded experimental data points. Samples of respond signal and the simulated noise signal of the detector were received by courtesy of the NAUTILUS group. First, the samples have been cleared of the obvious technical distortions (major picks) present due to technological reasons, mainly the replacement of liquid argon cooling the cylinder placed inside the apparatus to minimize its thermal noise. The real signal \( n(t) \) which was the subject of further analysis, exceeded \( N \geq 10^6 \) data points after such clearance procedure. All examined signals were searched with matrices \( 200 \times 200 \), but we checked that obtained results do not depend qualitatively on the matrix size one uses.
Random matrix approach in search for weak signals hidden in noisy data

The eigenvalue spectrum of matrices built on \( n(t) \) time series is plotted in fig. 2. It makes the characteristic triangle shape very much different from the semicircle form resulting from eq. (3). As one could have expected, the time series of randomly shuffled (permuted) data \( n_{\text{shuff}}(t) \) of the initial experimental time series signal \( n(t) \) leads to different eigenvalue spectrum shape — the Wigner semicircle. This is also shown in fig. 2 and confirms only the existence of strong correlations in the initial \( n(t) \) series. Generally, the presence of such correlations between matrix entries is expected to produce long tails in the corresponding eigenvalue spectrum [14]. The presence of memory in initial data can be easy explained. Any detector is adjusted to get maximum sensitivity for specific frequencies of measured signal. This adjustment procedure is “seen” by RMT approach as some memory in measured signal leading to the characteristic triangle shape of the corresponding eigenvalue spectrum.

Nevertheless, there is one very striking phenomenon in data analysis shown in details in fig. 3. We compare here eigenvalue spectra for the real detector respond signal \( n(t) \) and the simulated background noise \( n_{\text{sim}}(t) \) of the detector. The latter signal was used by the NAUTILUS group as the reference signal and any deviations from it should be a subject of careful experimental and theoretical analysis. One may see small difference between two spectra, particularly in the tail part. The eigenvalue spectrum for the real signal is wider and heavy tailed comparing with the same spectrum for simulated noise, indicating the presence of additional autocorrelations in \( n(t) \) signal in comparison with the simulated background. It is important to notice that this difference between two signals has not been detected in the standard Fourier analysis performed by the NAUTILUS group. The importance of such difference is visible only within the eigenvalue spectrum approach, which makes an additional strength of the method. This difference should be carefully examined, particularly in the spirit of tails modification caused by the presence of an additional periodic signal and shown for the simplest signals in fig. 1.

Let us concentrate in the beginning on the simulated detector background noise \( n_{\text{sim}}(t) \) enriched with a single sinusoidal signal at different \( \text{s/n ratio} \). The eigenvalue spectrum for such compound is drawn in fig. 4. We notice that long tails appear in spectrum for such modified signal with local maxima whose position does not depend on the period of added signal. However, the shape of the tail part and its length does depend on the \( \text{s/n ratio} \). Simultaneously, the central part of the eigenvalue spectrum (maximum of the spectrum) decreases with decreasing \( \text{s/n ratio} \). The above observation confirms the role of long tails in the eigenvalue spectrum and points to the necessity of their exact test for the presence of additional sinusoidal signals, which after all might be hypothetical gravitational waves. Indeed, the gravitational wave can be described as superposition of many sinusoidal signals with varying amplitude and periods [15–18]. It is described by many parameters defining the mutual angular position and distance of the source with respect to the wave observatory on the ground, orientation of rotation axes, rotation speeds, etc. Most of these parameters are chosen randomly, since \textit{a priori} we do not know the place in the Galaxy from where we can expect the arrival of the signal. Therefore, the exact shape of expected gravitational wave signal and the responded signal of detectors is the subject of Monte
Fig. 4: (Color online) Simulated expected eigenvalue spectra for the NAUTILUS background noise with added sinusoidal signal at various s/n ratios for two different periods of this signal: $T = 10$ (a) and $T = 10^4$ (c). Panels (b), (d) show the zoomed part of both plots, indicating for comparison the actual detector noise level as well (bold line).

Fig. 5: (Color online) Eigenvalue spectra of the detector’s respond signal in the case of strong and weak gravitational wave when background noise is neglected. Shape of the tail part is magnified and shown in the central part of both figures while the inboxes present the corresponding whole spectrum.

Carlosimulation [15–17]. Depending on the choice of free parameters, one can get signals varying very much and having different impact on the relative sizes of a probe body.

We took in our further simulations a model of the signal representing the quadrupole gravitational wave that is emitted by a freely precessing axisymmetric star [15]. Two extreme detector respond signals with significantly different amplitude were generated (amplitude ratio of these signals was about one order of magnitude). Both signals were offered to us by courtesy of A. Królak and P. Jaranowski. They are based on simulation and analysis described in details in a series of their papers [15–18]. We will call these respond signals, with $s/n$ ratios, respectively, $2.17 \cdot 10^{-1}$ and $5.45 \cdot 10^{-2}$, strong and weak later on. Their eigenvalue spectra are shown in fig. 5 for completeness of our reasoning. They are more complicated than those for a single sinusoidal signal but the central pick is well noticed in all cases what reflects the periodic nature of gravitational waves.

Finally, we checked how the presence of such real gravitational wave signal would change the eigenvalue spectrum of the respond signal from the detector if the
Random matrix approach in search for weak signals hidden in noisy data

Fig. 6: (Color online) Eigenvalue spectra of the detector’s respond signal for the case of strong and weak gravitational wave when the background noise (simulated) is added. Magnitude of the tested signal is kept constant, while the background noise is being changed $10^k$ times, $(k = 0, -1, -2, -3)$ accordingly. Spectra corresponding to $k = 0, -1, -2$ are relatively artificially shifted along $\rho$ axis by $\Delta \rho = 0.1$ to show their details.

Simulated background noise of detector is also switched on. The outcomes of this simulation is presented in series of figs. 6–8. The magnitude of gravitational wave signal was kept constant in this simulation, while the level of background noise was the subject of correction $10^k$ times ($k = 0, -1, -2, -3$). We simulated this way the eigenvalue spectrum of the entire respond signal expected from the detector in different cases of background noise reduction. This scenario is expected to come true in ground-based experiments in near future.

It is seen that in approach based on eigenvalue spectrum, the strong signal of gravitational wave passing through the Earth is going to be detected already at the present sensitivity of NAUTILUS detector ($k = 0$) (see figs. 6–8). The weaker respond signal can be found if the background noise level is reduced $10^{-2}$ times (see fig. 8). The positive respond signal of gravitational wave will most easier be seen in both cases by modification of the tail and head part in the eigenvalue spectrum as indicated in figs. 7, 8.

Concluding, we presented a novel approach in detection of periodic signals in noisy data. Moreover, from the construction of this method follows that even non-periodic but repeatable signals hidden in noisy data can be detected this way. Our analysis was focused, as an example, on the real noisy data obtained from NAUTILUS gravitational wave experiment. We believe that in order to get in future unambiguous experimental evidence of gravitational waves, one has to apply independent methods, based on various theoretical philosophies. New approach we proposed is based on RMT techniques. It has revealed an existing subtle difference between the simulated detector noise and truly recorded background signal. This difference was not evident so far with the use of other techniques like Fourier analysis and has to be eliminated if very weak gravitational waves signal is going to be cross-checked by independent and diversified methods of data analysis. We provided also the eigenvalue spectrum of the entire respond signal expected from ground-based detector in future experimental activities, by diminishing its background noise level.

The noise of the current laser interferometer gravitational wave detectors is much more complicated than the bar detectors analyzed in this paper. Nevertheless, we believe that this method can be applied also to laser interferometers with much complicated structure of background noise. It is because the final results in the presented method come from comparison of two signals: one taken from the detector respond signal when no gravitational signal is detected, but all background noises included, and the second one coming from this noise dressed up with new periodic or even non-periodic but repeatable signal.
amending the eigenvalues of respective matrices build from the time series data. We believe that good performance of our method in experiments with bar detectors is also a positive predictor for its further applications in many other areas.

** **

The authors would like to thank A. Królok for helpful discussion and the NAUTILUS group for delivering the samples of experimental data and the background signal of the NAUTILUS detector for the purpose of this elaboration. The authors also thank M. Daszkiewicz for suggesting the problem and contacting us with gravitational wave researchers.

REFERENCES

[1] Mehta M., *Random Matrices* (Academic Press, New York) 1995.
[2] Guhr T., Müller-Groeling A. and Weidenmüller H. A., *Phys. Rep.*, **299** (1998) 190.
[3] Laloux L., Cizeau P., Bouchaud J. P. and Potters M., *Risk Mag.*, **12** (1999) 69.
[4] Relaño A., Gómez J. M. G., Molina R. A., Rетamosa J. and Faleiro E., *Phys. Rev. Lett.*, **89** (2002) 244102.
[5] Laloux L., Cizeau P., Bouchaud J. P. and Potters M., *Phys. Rev. Lett.*, **83** (1999) 1467.
[6] Plerou V., Gopikrishnan P., Rosenov B., Amaral L. N. and Stanley H. E., *Phys. Rev. Lett.*, **83** (1999) 1471.
[7] Plerou V., Gopikrishnan P., Rosenov B., Amaral L. N., Guhr T. and Stanley H. E., *Phys. Rev. E*, **65** (2002) 066126.
[8] Podobnik B., Wang D., Horvatić D., Grosse I. and Stanley H. E., *EPL*, **90** (2010) 68001.
[9] Wang D., Podobnik B., Horvatić D. and Stanley H. E., *Phys. Rev. E*, **83** (2011) 046121.
[10] Wishart J., *Biometrica A*, **20** (1928) 32.
[11] Marčenko V. A. and Pastur L. A., *Math. USSR Sb.*, **1** (1967) 457.
[12] Wigner E. P., *Ann. Math.*, **53** (1951) 36.
[13] http://www.roma1.infn.it/rog/nautilus.
[14] Cizeau P. and Bouchaud J. P., *Phys. Rev. E*, **50** (1994) 1810.
[15] Jaranowski P., Królok A. and Schutz B. F., *Phys. Rev. D*, **58** (1998) 063001.
[16] Jaranowski P. and Królok A., *Phys. Rev. D*, **59** (1999) 063003.
[17] Jaranowski P. and Królok A., *Phys. Rev. D*, **61** (2000) 062001.
[18] Astone P., Borkowski K. M., Jaranowski P. and Królok A., *Phys. Rev. D*, **65** (2002) 042003.