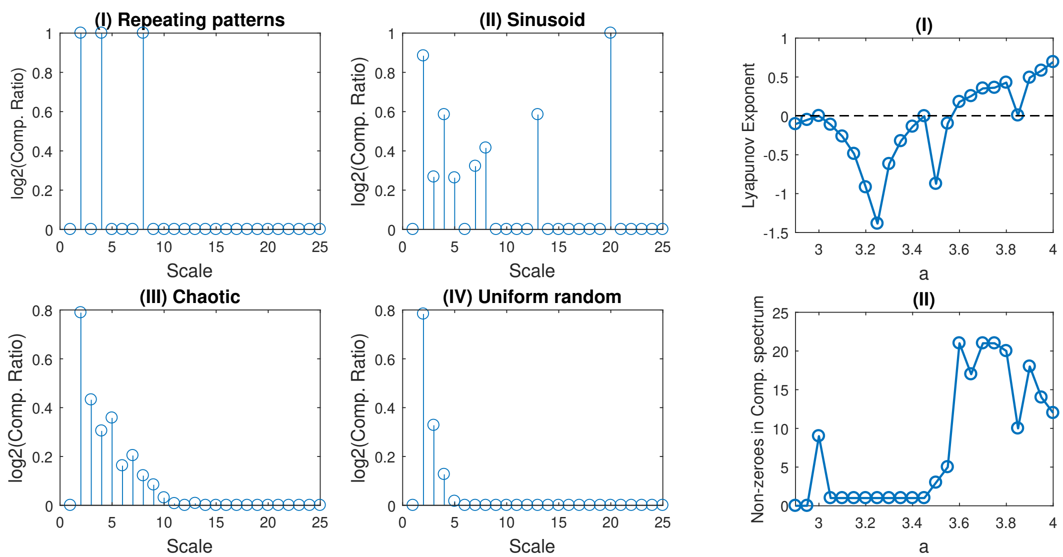
Spectral estimation based on compressibility

Aditi Kathpalia[a] and **Nithin Nagaraj**[b]
[a] *Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic (Email: kathpalia@cs.cas.cz)*
[b] *Consciousness Studies Programme, National Institute of Advanced Studies, Bengaluru, India (Email: nithin@nias.res.in)*

Signal processing and information theory [1] are two disparate fields used for characterizing signals for various scientific and engineering applications. Spectral analysis (a subset of signal processing) helps estimation of power at different frequency components present in the signal. Characterizing a time-series based on its average amount of information is useful for estimating its complexity and compressibility (eg., for communication applications). Information theory doesn't deal with spectral content while signal proessing doesn't consider the information content or compressibility of the signal. Data compression is closely tied to information theory and allows for an algorithmic approach to estimate complexity and compressibility of time series. In this work, we attempt to bring the fields of signal processing and information theory together by using a lossless data compression algorithm to estimate the amount of information or 'compressibility' of time series at different scales or frequencies. We employ the Effort-to-Compress (ETC) algorithm [2] to obtain a *Compression Spectrum*.

The compression spectrum for a number of simulated signals (length= 2000, 8 bins for quantizing the signal to obtain the symbolic sequence) is shown in Fig. 1a. Fig. 1a(I) displays the spectrum for a signal X composed of repeating patterns of length 8: $[1, 2, 3, 4, 5, 6, 7, 8]$. Similarly, Fig. 1a(II) shows the spectrum for $X(t) = \sin(2\pi wt)$, where $w = 50$ Hz and sampling frequency is 1 kHz; 1a(III) is the spectrum for the time series from the chaotic logistic map: $X(n) = aX(n-1)(1-X(n-1))$, where $a = 4.0$; and 1a(IV) is for a noise signal comprised of entries drawn from a uniform random distribution $U(0, 1)$.

In Fig. 1a(I), compressibility at patterns corresponding to scales 2, 4 and 8 can be observed. In 1a(II), the highest scale (at which compressibility is also maximum) can be observed at 20, which is the no. of samples in one cycle of the sinusoid. There are other lower scales at which compressibility is non-zero as there is some periodicity at those scales. Further, both 1a(III) and 1a(IV) have a decaying spectrum as there is no periodicity in the signal and the information content is limited to lower scales. However, the spectrum of (III) is broader than that of (IV). This is because, a random noise signal is expected to have negligible information or compressibility at any specific scale.



(a) (b)

Figure 1: (a) Compression spectra for different types of signals. (b) (I) Lyapunov exponent and (II) number of non-zero values in the compression spectrum of logistic maps for varying bifurcation parameter a.

Further, we analyse the variation of compression spectra as the level of chaos is varied (Fig. 1b(I)) in case of a logistic map time series (bifurcation paramter a varied from 2.9 to 4.0). Fig. 1b(II) shows the number of non zero values in the corresponding compression spectrum, which is broadest at intermediate values of chaos. Spectral analysis based on compressibility can be a useful tool for nonlinear time series analysis where traditional Fourier analysis has limitations. It can provide insight into the information content of signals at different scales.

References

[1] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[2] Nithin Nagaraj, Karthi Balasubramanian, and Sutirth Dey. A new complexity measure for time series analysis and classification. *The European Physical Journal Special Topics*, 222(3-4):847–860, 2013.