Accepted Manuscript

A geometric analysis of time series leading to information encoding and a new entropy measure

Kaushik Majumdar, Srinath Jayachandran

PII: S0377-0427(17)30347-3
DOI: http://dx.doi.org/10.1016/j.cam.2017.07.006
Reference: CAM 11217

To appear in: *Journal of Computational and Applied Mathematics*

Received date : 12 May 2017

Please cite this article as: K. Majumdar, S. Jayachandran, A geometric analysis of time series leading to information encoding and a new entropy measure, *Journal of Computational and Applied Mathematics* (2017), http://dx.doi.org/10.1016/j.cam.2017.07.006

A Geometric Analysis of Time Series Leading to Information Encoding and A New Entropy Measure

Kaushik Majumdar, Systems Science and Informatics Unit, Indian Statistical Institute, 8th Mile, Mysore Road, Bangalore 560059, India; E-mail: kmajumdar@isibang.ac.in

Srinath Jayachandran, Systems Science and Informatics Unit, Indian Statistical Institute, 8th Mile, Mysore Road, Bangalore 560059, India; E-mail: sri9s@yahoo.in

Abstract: A time series is uniquely represented by its geometric shape, which also carries information. A time series can be modelled as the trajectory of a particle moving in a force field with one degree of freedom. The force acting on the particle shapes the trajectory of its motion, which is made up of elementary shapes of infinitesimal neighborhoods of points in the trajectory. It has been proved that an infinitesimal neighborhood of a point in a continuous time series can have at least 29 different shapes or configurations. So information can be encoded in it in at least 29 different ways. A 3-point neighborhood (the smallest) in a discrete time series can have precisely 13 different shapes or configurations. In other words, a discrete time series can be expressed as a string of 13 symbols. Across diverse real as well as simulated data sets it has been observed that 6 of them occur more frequently and the remaining 7 occur less frequently. Based on frequency distribution of 13 configurations or 13 different ways of information encoding a novel entropy measure, called semantic entropy (E), has been defined. Following notion of power in Newtonian mechanics of the moving particle whose trajectory is the time series, a notion of information power (P) has been introduced for time series. E/P turned out to be an important indicator of synchronous behaviour of time series as observed in epileptic EEG signals.

MSC codes: 37M10; 92C55; 97I40

Keywords: Entropy; geometric features; power; synchronization; time series.

1. Introduction

A time series is a collection of observations indexed by the date of each observation (Hamilton 1994, p. 25; Lin et al. 2012, Chapter 28). This is for slowly evolving discrete time series chronicling social events. There are continuous time series like the ECG signal of the heart (Richman and Moorman 2000) or the EEG signals of the brain. In general, any time domain signal is a time series. The term *signal* is generally applied to something that conveys information (Shannon 1948; Oppenheim et al. 1999). Time series are also studied for their underlying information content. In signal processing transformations of the time domain signals to domains other than the time are quite popular. In contrast, time series analyses in