Physics Letters A 379 (2015) 997–1011

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Approximation of diagonal line based measures in recurrence quantification analysis

David Schultz [a,*], Stephan Spiegel [a], Norbert Marwan [b], Sahin Albayrak [a]

[a] DAI-Lab, Berlin Institute of Technology, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
[b] Potsdam Institute for Climate Impact Research, 14412 Potsdam, Germany

ARTICLE INFO

Article history:
Received 25 November 2014
Received in revised form 24 January 2015
Accepted 27 January 2015
Available online 29 January 2015
Communicated by C.R. Doering

Keywords:
Recurrence quantification analysis
Recurrence plot
Determinism
Approximation
Phase space discretization

ABSTRACT

Given a trajectory of length N, recurrence quantification analysis (RQA) traditionally operates on the recurrence plot, whose calculation requires quadratic time and space ($\mathcal{O}(N^2)$), leading to expensive computations and high memory usage for large N. However, if the similarity threshold ε is zero, we show that the recurrence rate (*RR*), the determinism (*DET*) and other diagonal line based RQA-measures can be obtained algorithmically taking $\mathcal{O}(N\log(N))$ time and $\mathcal{O}(N)$ space. Furthermore, for the case of $\varepsilon > 0$ we propose approximations to the RQA-measures that are computable with same complexity. Simulations with autoregressive systems, the logistic map and a Lorenz attractor suggest that the approximation error is small if the dimension of the trajectory and the minimum diagonal line length are small. When applying the approximate determinism to the problem of detecting dynamical transitions we observe that it performs as well as the exact determinism measure.

1. Introduction

Recurrence quantification analysis (RQA), i.e., the quantification of structures in recurrence plots, has established in several fields of research as a powerful tool to investigate recurrence related properties of complex dynamical systems [2]. The popularity of RQA is founded on its simplicity and flexibility to be applied to almost any type of data, including non-stationary processes [3]. In particular, the outstanding role of the RQA-measure *determinism* (DET) has been demonstrated in several applications, including discriminating signals from noise [4], detecting dynamical transitions [5,6], and the recently proposed use for pattern mining and classification [7]. A comprehensive overview of recurrence plots and its applications is given in [1].

The computation and quantification of recurrence plots generally involves operations with quadratic time and space complexity ($\mathcal{O}(N^2)$). This computational complexity leads to strongly increasing computation times and memory consumption for long time series (longer than 100,000 data points). Recurrence analysis of long time series, such as audio data [8], epileptic seizures [9], material damage detection [10], or hourly weather variability [11], is, therefore, limited. Another application that can be limited by the high computational complexity is online monitoring of data streams, e.g., for video surveillance [12], monitoring social interactions [13], or assessing driving behavior [7]. This is also crucial for medical applications. For example, monitoring the brain activity of epilepsy patients by multichannel electroencephalography (EEG) can help to identify early signs of a coming epileptic seizure. This provides the opportunity to initiate an EEG-triggered on-demand countermeasure just before the epileptic seizure (e.g., by an electrical stimulation in order to reset the brain dynamics) and, hence, to improve the life quality of such patients [14,15]. Recurrence plot based measures are promising candidates for such purpose [16–19]. Similar efforts can also be found for early detection of life-threating cardio-vascular diseases, such as ventricular tachycardia [20] or obstructive sleep apnea [21]. Parallel computing approaches (e.g., using GPU calculations [11,22]) can accelerate computation but do not reduce the computational cost.

In this Letter we show the following. If the similarity threshold ε is zero, then the recurrence rate, the determinism and other diagonal line based RQA-measures are in the computational complexity class $\mathcal{O}(N\log(N))$, whereas space complexity is $\mathcal{O}(N)$. We use this

* Corresponding author.
E-mail address: schultz@dai-lab.de (D. Schultz).

http://dx.doi.org/10.1016/j.physleta.2015.01.033
0375-9601/

observation in order to propose approximations to these measures for the case of $\varepsilon > 0$. The (approximative) measures are obtained algorithmically, without having to calculate the recurrence plot.

2. Motivation

Recent work has introduced recurrence plot-based distance measures, which can be utilized for mining (multi-dimensional) time series with nonlinear dynamics [23,24]. However, the quadratic time and space complexity of computation and quantification of recurrence plots makes distance calculations for relatively long time series and online processing of fast time series streams intractable. For these purposes we aim to approximate the proposed recurrence plot-based distance measures in such a way as to reduce the computational complexity while maintaining the classification accuracy.

3. Recurrence quantification analysis

For a given d-dimensional phase space trajectory $\vec{x}$ (reconstructed from a time series x, e.g., by time-delay embedding [25]) of length N and similarity threshold $\varepsilon \geq 0$ the recurrence plot of $\vec{x}$ is an illustration of the binary recurrence matrix $\mathbf{R}$, given by

$$R_{i,j} = \Theta\big(\varepsilon - \|\vec{x}_i - \vec{x}_j\|\big), \quad i, j = 1, \dots, N,$$

where $\|\cdot\|$ is a norm in the phase space of $\vec{x}$ and Θ is the Heaviside step function, defined by $\Theta(y) = 1$ if $y \geq 0$ and $\Theta(y) = 0$ if $y < 0$. Thus Θ indicates whether $\vec{x}_i$ and $\vec{x}_j$ are in ε-proximity (also denoted as similar) or not, i.e., $R_{i,j} = 1$ if $\|\vec{x}_i - \vec{x}_j\| \leq \varepsilon$ and $R_{i,j} = 0$ if $\|\vec{x}_i - \vec{x}_j\| > \varepsilon$. This relation is essential for the study of recurrence plots and will be used extensively in this Letter. The recurrence plot contains the *line of identity (LOI)*, which means that each entry on the main diagonal of $\mathbf{R}$ is 1. Structures parallel to the main diagonal, referred to as diagonal lines, are caused by similarly evolving epochs of the phase space trajectory $\vec{x}$.

Recurrence quantification analysis was developed in order to quantitatively describe recurrence plots. For this purpose, small scale structures, such as recurrence points or diagonal lines in the recurrence plot are used [26]. The fraction of recurrence points in the recurrence plot is measured by the *recurrence rate*,

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}, \tag{1}$$

which is interpreted as the probability to find a recurrence of trajectory $\vec{x}$. A more sophisticated RQA-measure is the *determinism*, which is defined for a given minimum diagonal line length μ as

$$DET^{(\mu)} = \frac{\sum_{l=\mu}^{N} l \cdot P(l)}{\sum_{i,j=1}^{N} R_{i,j}}, \tag{2}$$

where $P(l)$ is the number of diagonal lines of length l in $\mathbf{R}$. DET can be interpreted as the probability that a recurrence point belongs to a diagonal line. The parameter μ is usually set to 2. This choice is sufficient for most applications. However, in particular cases, larger values of μ can be necessary, e.g., reducing effects of tangential motion (oversampling), noise, or embedding effects [1].

As already mentioned, a phase space trajectory of a univariate time series can be reconstructed by time delay embedding [25]. We call this procedure *time series embedding*, since it is applied to the time series. In the sequel we will apply the method of time delay embedding to the trajectory $\vec{x}$ (that possibly was created by time series embedding for reconstruction purposes), but with the intention of quantifying diagonal structures in $\mathbf{R}$. In order to distinguish that from the time series embedding, we will denote this as *trajectory embedding*. More precisely, for a fixed time delay 1 and embedding dimension ν, we consider the trajectory embedding vectors

$$\vec{x}_j^{\nu} = (\vec{x}_j, \vec{x}_{j+1}, \dots, \vec{x}_{j+\nu-1}), \tag{3}$$

which are of dimension $d \cdot \nu$, provided that the trajectory $\vec{x}$ is d-dimensional. The trajectory embedding of $\vec{x}$ is then defined to be the sequence $\vec{x}^{\nu} = (\vec{x}_j^{\nu})_{j=1,\dots,N-\nu+1}$, which can be imagined as a trajectory in a $(d \cdot \nu)$-dimensional phase space. In Section 4.2 we show that information about $P(l)$ can be extracted by these representations leading to a surprising identity for the determinism.

4. *RR* and *DET* identities

We deduce identities for RR and $DET^{(\mu)}$, which allow fast calculation (without computing the recurrence plot) if the similarity threshold ε is zero. The identity for RR does hold for $\varepsilon = 0$ only. The identity for $DET^{(\mu)}$ is first shown for arbitrary $\varepsilon \geq 0$ and the assumption that the phase space norm is the maximum norm $\|\cdot\|_\infty$. However, in the special case of $\varepsilon = 0$, we will argue that the restriction to the $\|\cdot\|_\infty$-norm becomes redundant. Consequently it follows the important fact that the recurrence rate and the determinism are in $\mathcal{O}(N \log(N))$ if $\varepsilon = 0$, whereas the computational complexity of the classical methods that quantify the recurrence plot is $\mathcal{O}(N^2)$.

4.1. Recurrence rate identity

Given the trajectory embedding $\vec{x}^{\nu}$, Eq. (3), in analogy to Eq. (1) we define

$$\mathcal{PP}^{(\nu)} := \sum_{i,j=1}^{N-\nu+1} \Theta\big(\varepsilon - \|\vec{x}_i^{\nu} - \vec{x}_j^{\nu}\|\big), \tag{4}$$