Efficient computation of Recurrence Quantitative Analysis

Tomáš Martinovič

National Supercomputing Center IT4Innovations

tomas.martinovic@vsb.cz

November 1, 2017

Motivation

Analysis of long time series

- Recurrence quantitative analysis is used in analysis of many areas.
- Necessity to analyse long time series for robustness.
- Necessity to provide scalable parallel algorithm.

The objective

- Determine algorithm for effective computation of RQA.
- Describe advantages of the algorithm.

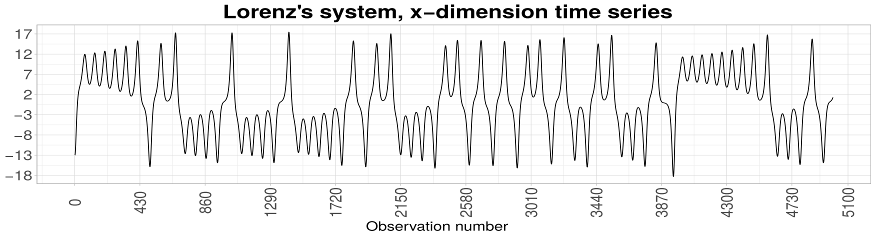
Takens embedding

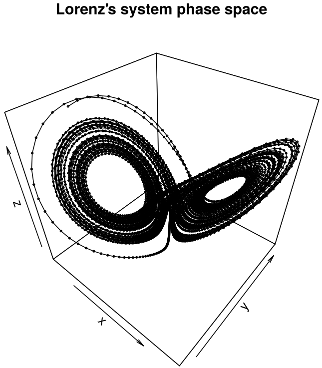
- Experimental time-series is one dimensional map of the multidimensional dynamical system.
- We can reconstruct the original trajectories by phase space reconstruction.

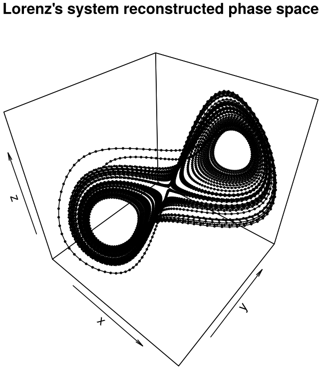
Definition

Let x be the univariate time series of length n. Then let m be embedding dimension and l be time delay. Then we create $n-((m-1)l)$, embedding vectors $X_1, \ldots, X_{n-((m-1)l)}$ in the following way

$$
\begin{aligned}
X_1 &= \{x_1, x_{1+l}, \ldots, x_{1+(m-1)l}\} \\
&\vdots \\
X_{n-(m-1)l} &= \{x_{n+m\cdot l}, x_{n+(m+1)\cdot l}, \ldots, x_n\}
\end{aligned}
$$


Lorenz's system, x−dimension time series
17
12
7
2
−3
−8
−13
−18
0
430
860
1290
1720
2150
2580
3010
3440
3870
4300
4730
5100
Observation number


Lorenz's system phase space
z
x
y


Lorenz's system reconstructed phase space
z
x
y

Recurrence plot

Let D be distance matrix such as

$$D_{i,j} = d(X_i, X_j) = \sqrt{\sum_{k=0}^{m-1} (x_{i+kl} + x_{j+kl})^2}, \quad (1)$$

for all the pairs $X(i), X(j)$, where $i, j \in \{1, 2, \ldots, n - (m-1)l\}$. Then let recurrence plot for given ε be

$$RP^{\varepsilon}_{i,j} = \begin{cases} 1, & D_{i,j} < \varepsilon \\ 0, & \text{otherwise}, \end{cases} \quad (2)$$

for $i, j \in \{1, 2, \ldots, n - (m-1)l\}$.

Example

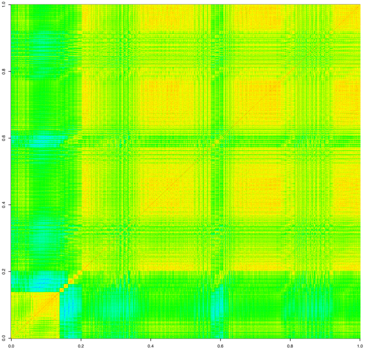


Figure: Unthresholded RP

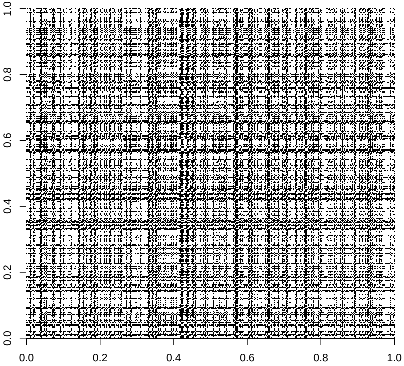
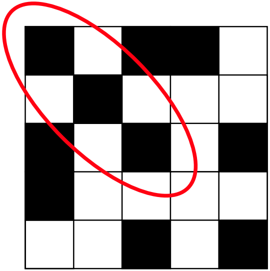
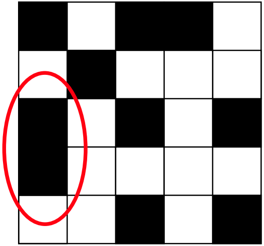
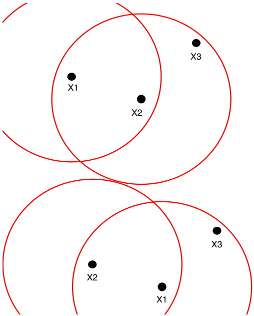


Figure: Thresholded RP

RQA







Diagonal lines

Line length	1	2	3
Count	7	0	1

Vertical lines

Line length	1	2	3
Count	8	1	0

RQA

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} P(l)}, \tag{3}$$

where l is the length of diagonal lines and $P(\cdot)$ is the histogram of diagonal lines.

$$ENTR = -\sum_{l=l_{min}}^{N} p(l) \ln p(l), \tag{4}$$

where $p(l)$ is probability that diagonal line has length l.

Computation process

Classical workflow

- Embed time series and store it.
- Compute windowed RP and store it.
- Compute RQA.

New workflow

- Compute histogram of diagonal anad vertical lines length directly from the time series.
- Compute RQA directly from histograms.

Direct computation

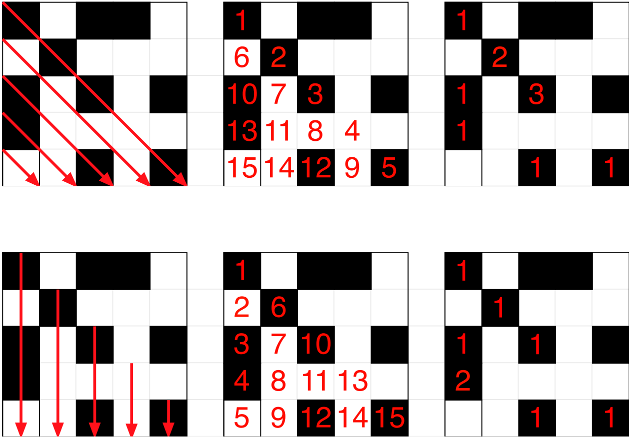
To compute histograms directly from the time series it is necessary to exploit the definition of distance matrix.

$$D_{i,j} = d(X_i, X_j) = \sqrt{\sum_{k=0}^{m-1} (x_{i+kl} + x_{j+kl})^2}, \qquad (5)$$

for all the pairs X_i, X_j, where $i, j \in \{1, 2, \ldots, n-(m-1)l\}$.

x represent elements of the time series $\rightarrow$ it is enough to be able to compute the indices $i + kl$ and $j + kl$ to compute the distance directly from the time series.
To compute the histograms of diagonal/horizontal lines the order in which we compute the distances is important.

Computation order



Scaling

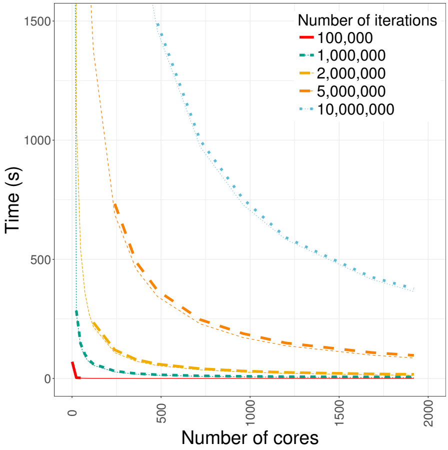


Figure: Strong scaling

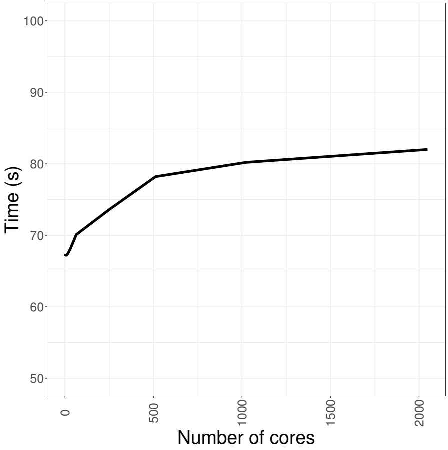


Figure: Weak scaling

Conlusion

Conclusion

- Recurrence quantitative analysis may be computed directly from the time series, without storing the recurrence plot.
- Such computation decreases memory complexity of the algorithm.
- Computation of histograms in diagonal and vertical order allows easy parallelization.

Future work

- Application of RQA for time series analysis.
- Optimizing parallel implementation of the algorithm.