# HPC ORIENTED ALGORITHM FOR COMPUTATION OF RECURRENCE QUANTITATIVE ANALYSIS

Tomáš Martinovič, Georg Zitzlsberger
IT4Innovations, VŠB – Technical University of Ostrava, Czech Republic

Objectives

Recurrence quantitative analysis (RQA) is a quantification of the dynamical properties of time series. This method is used in many disciplines, such as medicine, biology, speech and vocalization research. The main drawback of this method is its computational complexity. The first step for the RQA is computation of the distance matrix of subsequences of the time series. Afterwards, this distance matrix is thresholded, setting 1 if the distance is less than a parameter ε and 0 otherwise. These are called recurrence plots. The histogram of consecutive 1s on the diagonals is computed. This histogram is then used to compute RQA. We present an algorithm for the computation of RQA directly from the input data. This algorithm allows easy parallelization of the computation with minimal spatial complexity.
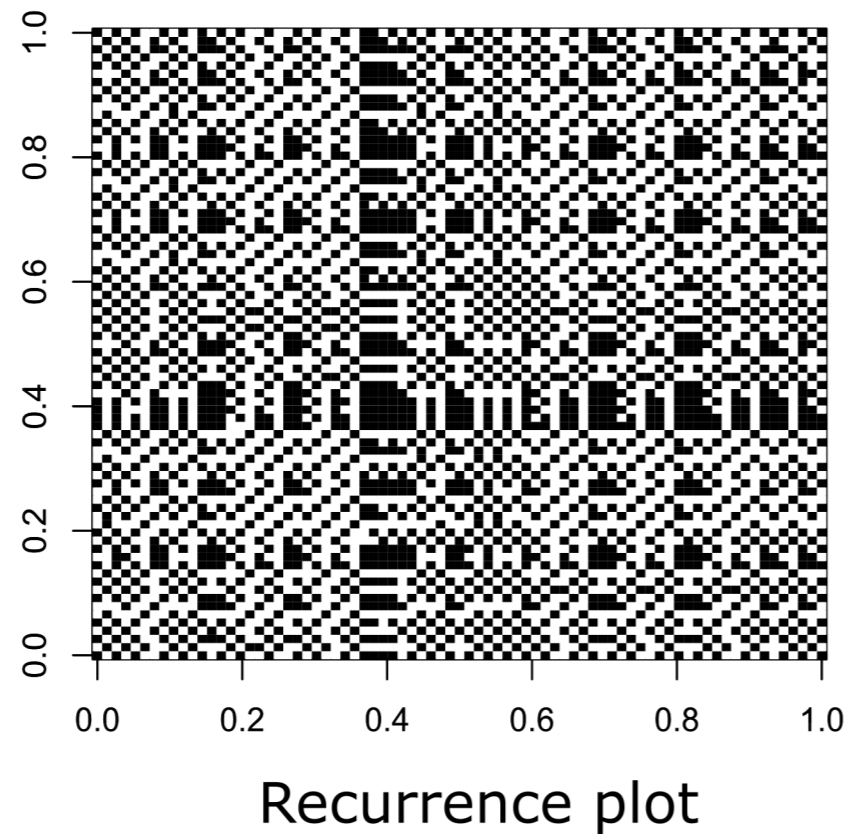
Recurrence plot

Algorithm

To compute the recurrence plot, it is necessary to set three parameters: embedding dimension m, lag l, and threshold ε. The first two set the length and step of the subsequences of the input time series. Distances of these subsequences are computed and then thresholded. Computation of the distance between two subsequences may be written as:

$$D(t_1, t_2) = \sqrt{\sum_{k=0}^{m-1} \Big(x(t_1 + kl) - x(t_2 + kl) \Big)^2}$$

And the thresholding is computed as follows:

$$RP^\varepsilon(i,j) = \begin{cases} 1, & D(i,j) < \varepsilon \\ 0, & \text{otherwise} \end{cases}$$
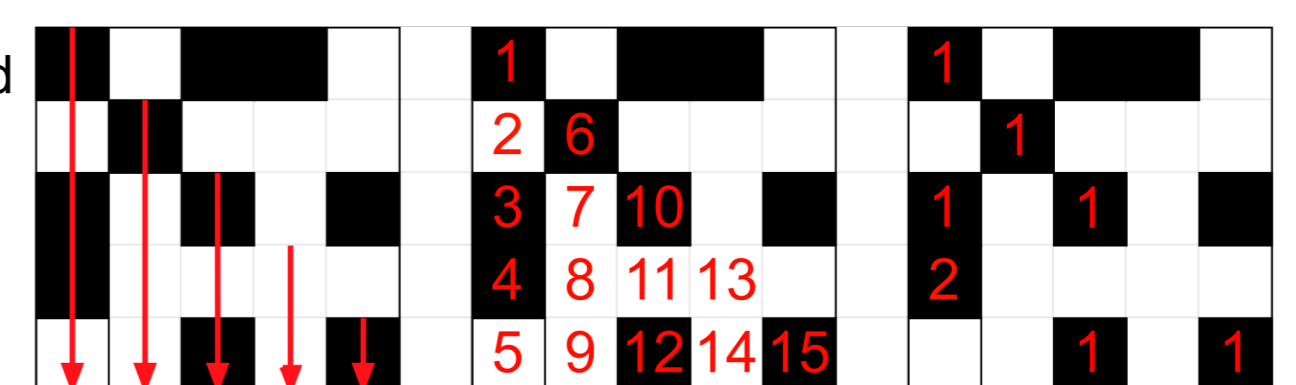
RQA is then computed from the histogram of diagonal and vertical lines found in the recurrence plot. For example the determinism is the ratio of diagonal lines of minimal length l_{min} to all the recurrence points in the recurrence plot. This measure indicates how often long similar subsequences are found in the input time series. Laminarity is computed as the ratio of vertical lines of minimal length v_{min} to the number of recurrences and it stands for how similar subsequences are repeated in succession. If P(l) and P(v) are histograms of the the diagonal and vertical lines respectively, then we can define determinism and laminarity as:

$$DET = \frac{\sum_{l=l_{min}}^{N} l P(l)}{\sum_{i,j=1}^{N} RP(i,j)} \quad LAM = \frac{\sum_{v=v_{min}}^{N} v P(v)}{\sum_{i,j=1}^{N} RP(i,j)}$$

Since the subsequences dimension and lag is known, it is possible to compute the distance of subsequences directly from the time series and only the histograms of the diagonal/vertical lines are needed to compute the RQA. It is possible to compute RQA directly from the input time series and independently for each diagonal/vertical in the recurrence plot. There is similar approach to the massive parallelization of the RQA computation using the divide and recombine method (https://pypi.org/project/PyRQA/). The main difference our method is no need for the carryover buffer.
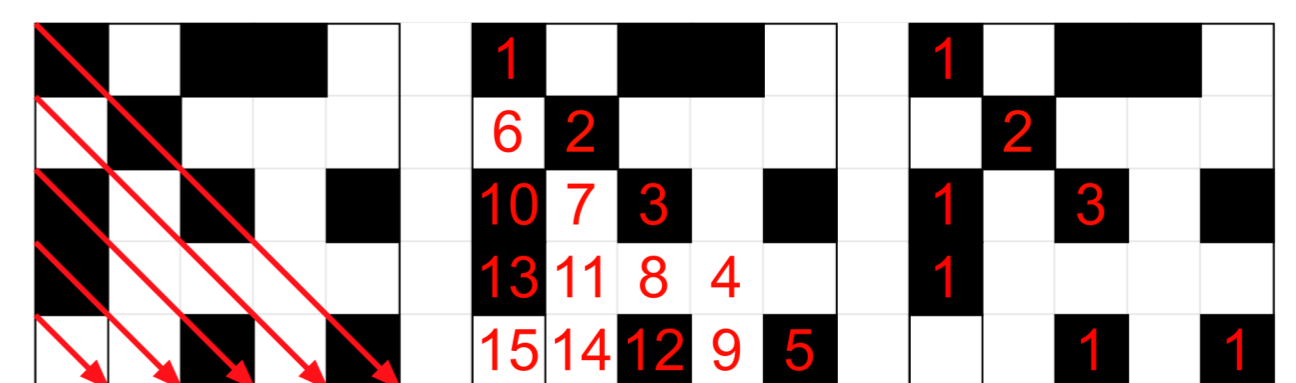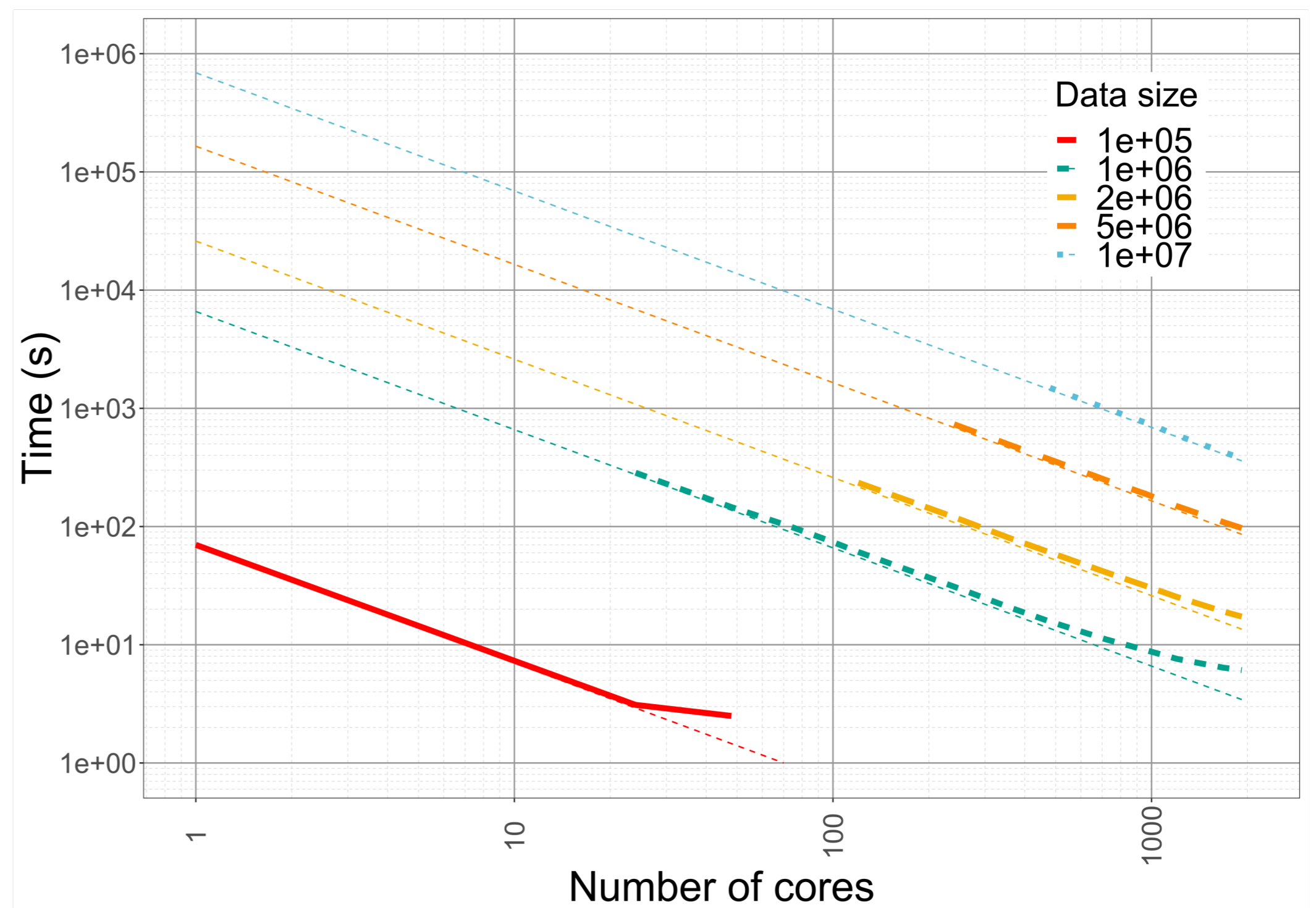

Diagonal histogram computation
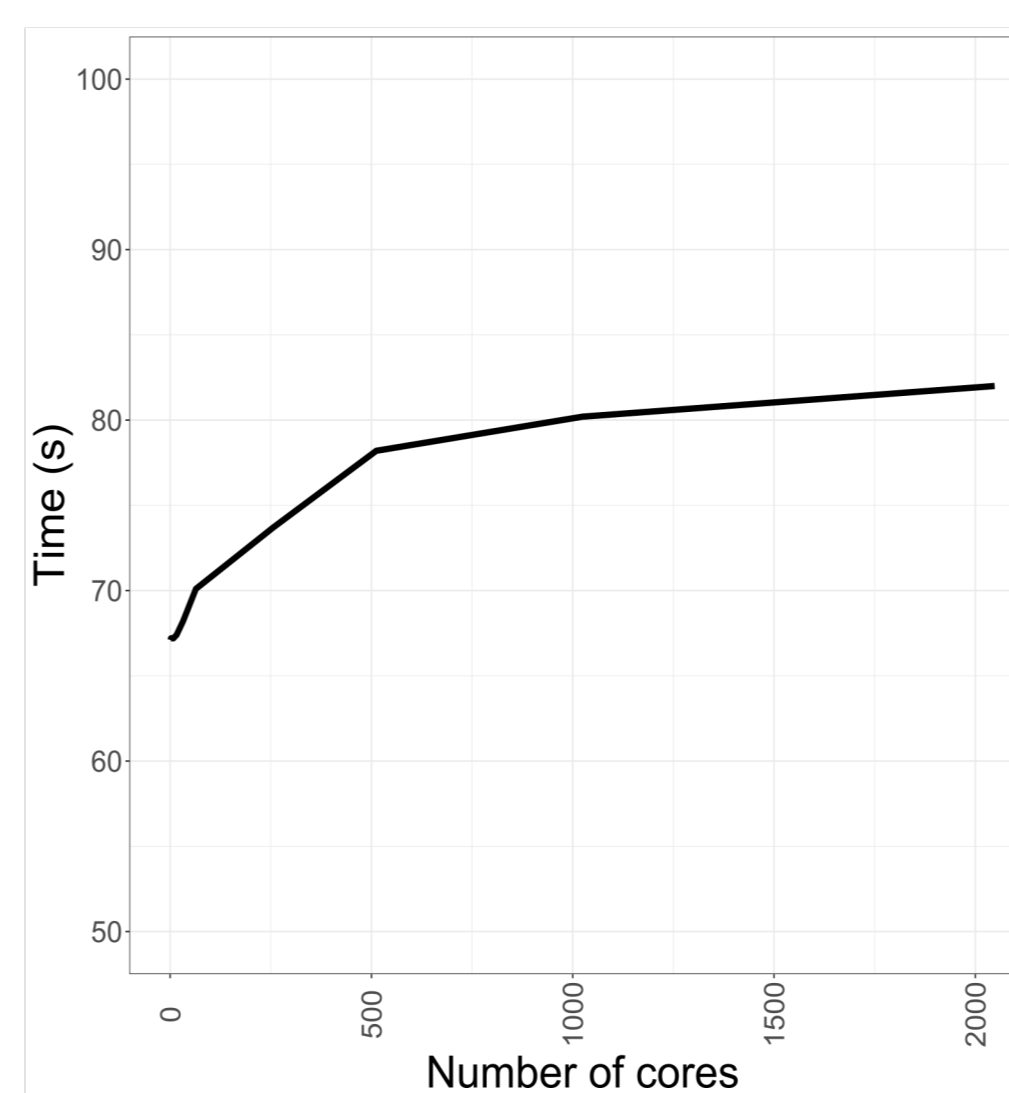
Each arrow in the figures stands for one independent thread computation and results are gathered at the end of the computation.

Vertical histogram computation

Results

Strong scaling

Weak scaling

The proposed algorithm was tested with the MPI implementation using up to 2048 cores on Salomon Cluster (https://docs.it4i.cz/salomon/hardware-overview/) using Intel MPI and Intel compiler. Figures show the

Acknowledgement

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER