Meta stable states in Markow chains

Dijana Pratljacic, Niko Perica, Vice Perica

November 28, 2025

1 Introduction

As part of a project for the course Matrix and Tensor Methods in Data Analysis, our team explored methods for identifying metastable states in Markov chains. This report builds on several research papers [1], [2].

In the context of biomolecular dynamics, the kinetic and potential energy landscape gives rise to conformations that behave as almost-invariant dynamical subsets. These are regions of the state space where, once the system enters, it tends to remain for a long time. By discretizing the underlying Markov operator, one obtains a finite dimensional, time homogeneous Markov chain whose structure reflects these metastable conformations.

The goal of our project is to detect such metastable states using algorithms presented in the referenced literature.

2 Problem and Algorithm

2.1 Detecting Metastable Structure by Permuting the Transition Matrix

Given a stochastic transition matrix T representing the discretized states of a biomolecule which may include measurement noise. We seek a permutation matrix P such that

$$PTP^\top = C + E,$$

where C is a block diagonal stochastic matrix and E contains only small off block probabilities. The diagonal blocks of C correspond to metastable states, while E captures noise and the rare transitions between different metastable regions. Importantly, the number of such metastable blocks is not known in advance.

Our goal is therefore to find a permutation that rearranges the transition matrix into a form that is as block diagonally dominant as possible. In particular, we compute the singular vector corresponding to the second largest singular value of T, sort its entries, and apply the resulting permutation to obtain

$$\widetilde{T} = PTP^\top.$$

This reordering separates states that are unlikely to belong to the same metastable group, making it easier to identify the metastable structure. To obtain the final metastable clusters, the algorithm can be applied recursively. After permuting the matrix using the second singular vector, each group of states separated by this vector is treated as a submatrix. The same procedure is applied to each submatrix, splitting it further into smaller clusters if necessary. This recursive application continues until no further meaningful separation is detected, yielding the final set of metastable clusters.

2.2 Algorithm (Sketch)

1. Compute the singular value decomposition $T = U\Sigma V^\top$.

2. Extract the singular vector $u^{(2)}$ corresponding to the second largest singular value.

3. Sort the entries of $u^{(2)}$ to obtain an index permutation π.

4. Construct the permutation matrix P associated with π.

5. Form the permuted matrix $\widetilde{T} = PTP^\top$.

6. Structure of $\widetilde{T}$ can be written as

$$\widetilde{T} = \begin{bmatrix} T_1 & \epsilon_1 \\ \epsilon_2 & T_2 \end{bmatrix},$$

where T_1 and T_2 are diagonal blocks corresponding to candidate metastable groups, and ϵ_1, ϵ_2 are small off-diagonal blocks. T_1 is $k \times k$ matrix, where k is the number of negative elements in $u^{(2)}$. If the diagonal blocks satisfy the chosen threshold (i.e., they are sufficiently large), apply the algorithm recursively on T_1 and T_2 to further separate subgroups.

This procedure typically yields a transition matrix that is nearly block diagonal, making metastable states visually and numerically identifiable.

3 Results

After implementing the algorithm, the only parameter we can adjust is the threshold. The threshold represents a trade off between false positives and "undiscovered" clusters, and its choice depends on the specific problem and desired balance between these factors.

In practice, an intuitive way to select the threshold is by assessing the stability of the resulting clusters. By trying several threshold values, plotting the corresponding metrics, and performing a visual check asking "does this result look sensible?", we can observe how the number of clusters changes with the threshold, as shown in Figure 5. Based on this analysis, we determined that a threshold of 0.4 produces reasonable and stable clusters for our data.
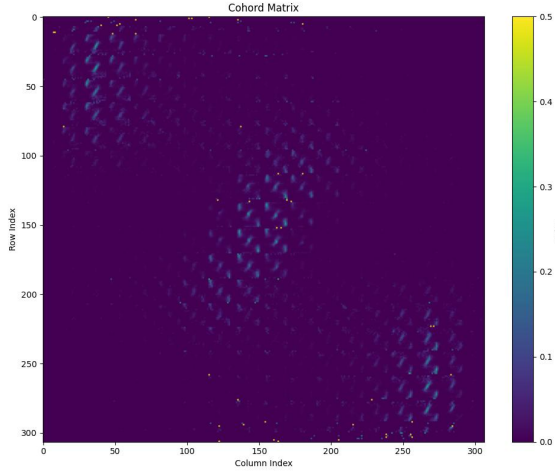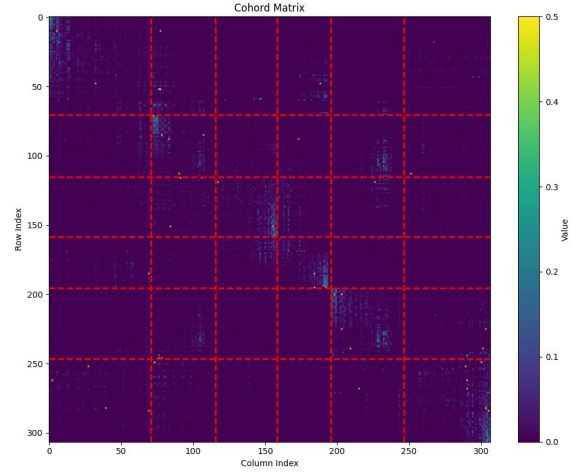

Figure 1: Original Ph500.mat
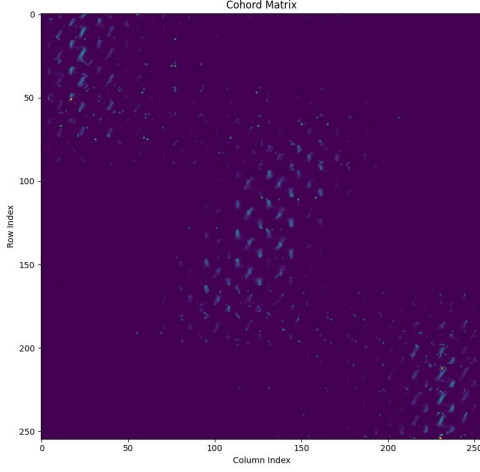

Figure 2: Permutated Ph500.mat
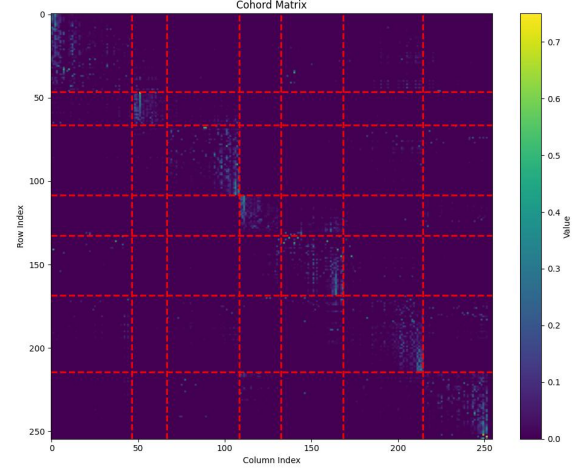
Figure 3: Original Ph300.mat
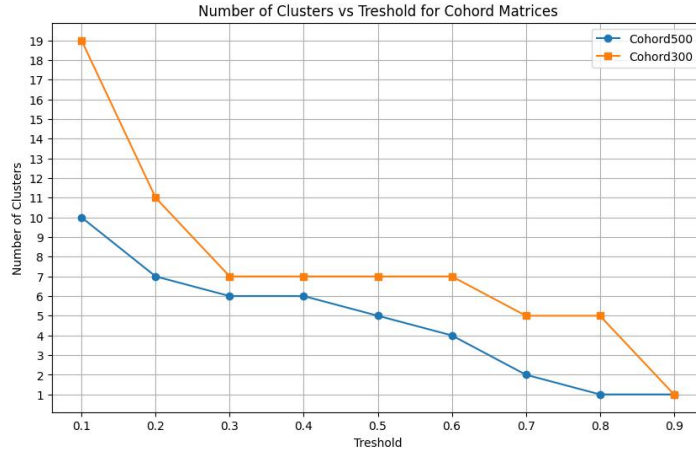

Figure 4: Permutated Ph300.mat

Figure 5: Number of clusters depeding on threshold

4 Additional Improvements

In the referenced papers, it was suggested that using the right singular vector corresponding to the second largest singular value could improve the algorithm. However, the instructions were not detailed, leaving room for potential improvements.

In our implementation, we used the full singular value decomposition (SVD) function from the NumPy library, which is somewhat redundant since we only need the second dominant singular vector. Nevertheless, the algorithm was sufficiently fast for our purposes, so further optimization was not necessary.

There are several ways to improve efficiency, such as computing only the leading singular vectors using iterative methods like Power iteration combined with deflation.

References

[1] P. Deuflhard, W. Huisinga, A. Fischer, and Ch. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled markov chains. *Linear Algebra and its Applications*, 315:39–59, 2000. Submitted by V. Mehrmann.

[2] David Fritzsche, Volker Mehrmann, Daniel B. Szyld, and Elena Virnik. An svd approach to identifying metastable states of markov chains. *Electronic Transactions on Numerical Analysis*, 29:46–69, 2008.