Bulk of random matrices generated by Markov chains with community structure

Presented at the fourth ZiF Bielefeld summer school (2022)

Alexander Van Werde

Department of Mathematics and Computer science

Motivation

Many real-world systems can be modelled as a Markov chain:

- Sequence of words in text
- DNA
- Taxi pickup locations.

The state space can be very large. This causes problems such as algorithms slowing down.

Solution: cluster together similar states.

Motivation

Many real-world systems can be modelled as a Markov chain:

- Sequence of words in text
- DNA
- Taxi pickup locations.

The state space can be very large. This causes problems such as algorithms slowing down.

Solution: cluster together similar states.

Motivation

Many real-world systems can be modelled as a Markov chain:

- Sequence of words in text
- DNA
- Taxi pickup locations.

The state space can be very large. This causes problems such as algorithms slowing down.

Solution: cluster together similar states.

Provable clustering

For so-called *block Markov chains* the consistency of spectral community detection algorithms has been proved in [Sanders et. al., 2020] and [Zhang et. al., 2020]. An impossibility theorem when insufficient data is available was also proved in [Sanders et. al., 2020].

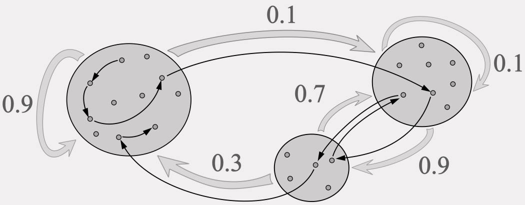
Key references:

- Anru Zhang and Mengdi Wang. Spectral state compression of Markov processes. IEEE Transactions on Information Theory, 66(5):3202–3231, 2020.
- Jaron Sanders, Alexandre Proutière, and Se–Young Yun. Clustering in block Markov chains. The Annals of Statistics, 48(6):3488–3512, 2020.

Block Markov Chain

Fix the following data:

- Number of clusters $K \geq 1$.
- Cluster transition matrix $p \in \mathbb{R}^{K \times K}$.



Then, the *block Markov chain* associated to a partition $\mathcal{V}_1 \cup \ldots \cup \mathcal{V}_K = \{1, \ldots, n\}$ is the Markov chain with transition matrix P given by

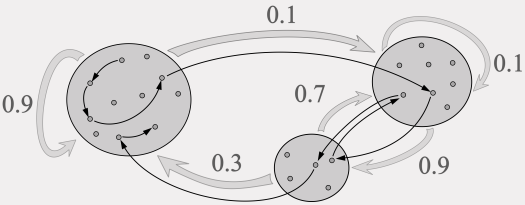
$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x,\ \forall j \in \mathcal{V}_y.$$

We consider the regime where $n \to \infty$ but K and p are fixed.

Block Markov Chain

Fix the following data:

- Number of clusters $K \geq 1$.
- Cluster transition matrix $p \in \mathbb{R}^{K \times K}$.



Then, the *block Markov chain* associated to a partition $\mathcal{V}_1 \cup \ldots \cup \mathcal{V}_K = \{1, \ldots, n\}$ is the Markov chain with transition matrix P given by

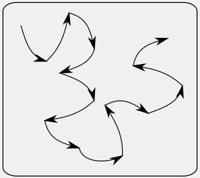
$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x,\ \forall j \in \mathcal{V}_y.$$

We consider the regime where $n \to \infty$ but K and p are fixed.

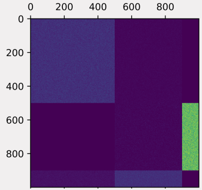
Spectral algorithm

Spectral community detection algorithm:

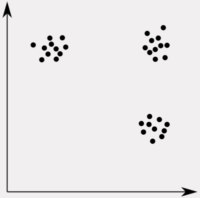
1. Observe a sample path $X_0, \ldots, X_\ell$.
2. Construct $\widehat{P}_X$, the estimator for the transition matrix P.
3. Principal component analysis of $\widehat{P}_X$ yields embedding of state space into $\mathbb{R}^K$.
4. k-means + improvement algorithm.



$\Longrightarrow$



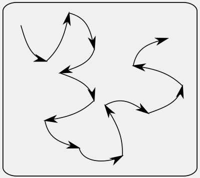
$\Longrightarrow$



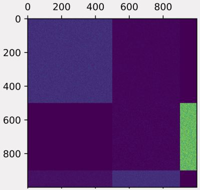
Spectral algorithm

Spectral community detection algorithm:

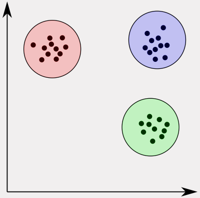
1. Observe a sample path $X_0, \dots, X_\ell$.
2. Construct $\widehat{P}_X$, the estimator for the transition matrix P.
3. Principal component analysis of $\widehat{P}_X$ yields embedding of state space into $\mathbb{R}^K$.
4. k-means + improvement algorithm.



$\Longrightarrow$



$\Longrightarrow$



Motivating question

Note that the spectral algorithm relied on singular vectors of $\widehat{P}_X$. This motivates studying the spectral properties of $\widehat{P}_X$ in general. For instance, what is the limiting empirical singular value distribution?

Key difficulty: $\widehat{P}_X$ was constructed based on a sample path of a Markov chain. This is **dependent** data.

Key contribution: A more natural way to deal with the dependence than was used in previous works about block Markov chains. Namely, a coupling argument.

Motivating question

Note that the spectral algorithm relied on singular vectors of $\widehat{P}_X$. This motivates studying the spectral properties of $\widehat{P}_X$ in general. For instance, what is the limiting empirical singular value distribution?

Key difficulty: $\widehat{P}_X$ was constructed based on a sample path of a Markov chain. This is **dependent** data.

Key contribution: A more natural way to deal with the dependence than was used in previous works about block Markov chains. Namely, a coupling argument.

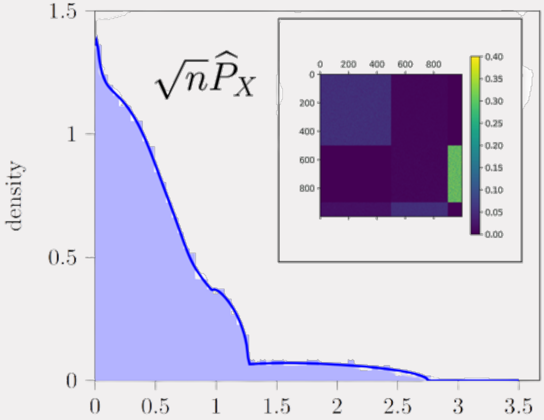
Motivating question

Note that the spectral algorithm relied on singular vectors of $\widehat{P}_X$. This motivates studying the spectral properties of $\widehat{P}_X$ in general. For instance, what is the limiting empirical singular value distribution?

Key difficulty: $\widehat{P}_X$ was constructed based on a sample path of a Markov chain. This is **dependent** data.

Key contribution: A more natural way to deal with the dependence than was used in previous works about block Markov chains. Namely, a coupling argument.

Result



$\sqrt{n}\widehat{P}_X$
density
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3
3.5
0
200
400
600
800
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Theorem

Assume that as the size of the state space n tends to infinity it holds that $\#\mathcal{V}_i = \alpha_i n + o(n)$ and that the length of the sample path satisfies $\ell = \lambda n^2 + o(n^2)$. Assume moreover that p defines an ergodic chain with equilibrium distribution π.

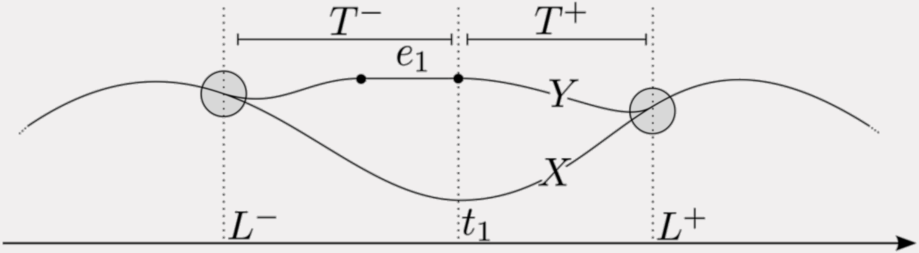
Then, the empirical singular value distribution $\nu_{\sqrt{n}\widehat{P}_X}$ converges weakly in probability to a probability measure ν on $\mathbb{R}_{\geq 0}$ whose symmetrization $\mathrm{sym}(\nu)$ has Stieltjes transform $s(z) = \sum_{i=1}^{K} \alpha_i (a_i(z) + a_{K+i}(z))/2$ with for $i = 1, \dots, K$

$$a_i(z)^{-1} = z - \sum_{j=1}^{K} \lambda^{-1} \pi(i)^{-1} \alpha_i p_{i,j} a_{K+j}(z)$$

$$a_{i+K}(z)^{-1} = z - \sum_{j=1}^{K} \lambda^{-1} \pi(j)^{-1} \alpha_i^{-1} \alpha_j^2 p_{j,i} a_j(z).$$

Key proof technique: Coupling

One can pretend as if the observation at time t is independent of all other observations up to some error term. More precisely, given $t_1 \in \{0, \ldots, \ell\}$ one constructs a path Y with Y independent of X_{t_1} such that $Y_t = X_t$ for most $t \neq t_1$.



Thank you!

The key references for this talk are:

- The current result: *Jaron Sanders and Alexander Van Werde. "Singular value distribution of dense random matrices with block Markovian dependence." arXiv preprint arXiv:2204.13534 (2022).*
- Spectral algorithm and information-theoretic lower bound: *Jaron Sanders, Alexandre Proutière, and Se–Young Yun. Clustering in block Markov chains. The Annals of Statistics, 48(6):3488–3512, 2020.*
- Different perspective on community detection in Markov chains and an algorithm: *Anru Zhang and Mengdi Wang. Spectral state compression of Markov processes. IEEE Transactions on Information Theory, 66(5):3202–3231, 2020.*

A different approach to coupling in dependent random matrices, namely a blocking procedure, can be found in the works of Merlevède and Banna in the period 2015-2016.