Large time behaviour and the second eigenvalue problem for finite state mean-field interacting particle systems

Sarath Yasodharan
Joint work with Rajesh Sundaresan

ECE Department, Indian Institute of Science

29 Feb 2020

System model

- N particles. State space: a finite set $\mathcal{Z}$.
- State of the nth particle at time t is $X_n^N(t) \in \mathcal{Z}$.
- Certain allowed transitions: specified by a directed graph $(\mathcal{Z}, \mathcal{E})$
- Empirical measure at time t

$$\mu_N(t) = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(t)} \in M_1(\mathcal{Z}).$$

- For each $(z, z') \in \mathcal{E}$, we have a function $\lambda_{z,z'} : M_1(\mathcal{Z}) \to [0, +\infty)$.
- Particle transitions: at time t, a $z \to z'$ transition occurs at rate $\lambda_{z,z'}(\mu_N(t))$. Mean-field interaction.
- $\{(X_n^N(t), 1 \leq n \leq N), t \geq 0\}$ is a Markov process on $\mathcal{Z}^N$.

The empirical measure process μ_N

- $\{\mu_N(t), t \geq 0\}$ is a Markov process on $M_1(\mathcal{Z})$ with infinitesimal generator

$$L^N f(\xi) = N \sum_{(z,z')\in\mathcal{E}} \xi(z)\lambda_{z,z'}(\xi)\left[f\left(\xi + \frac{\delta_{z'}}{N} - \frac{\delta_z}{N}\right) - f(\xi)\right].$$

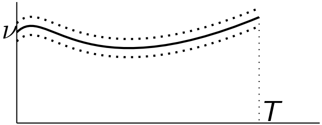
The empirical measure process μ_N

- $\{\mu_N(t), t \geq 0\}$ is a Markov process on $M_1(\mathcal{Z})$ with infinitesimal generator

$$L^N f(\xi) = N \sum_{(z,z')\in\mathcal{E}} \xi(z)\lambda_{z,z'}(\xi) \left[f\left(\xi + \frac{\delta_{z'}}{N} - \frac{\delta_z}{N} \right) - f(\xi) \right].$$

- Typical behaviour of μ_N (mean-field limit):
Let $\mu_N(0) \to \nu$ weakly as $N \to \infty$. Then $\{\mu_N(t), t \geq 0\}$ converges in probability (in $D([0,T], M_1(\mathcal{Z}))$ to the solution to the McKean-Vlasov equation:

$$\dot{\mu}(t) = \Lambda^*(\mu(t))\mu(t),\ \mu(0) = \nu.$$



- Our interest: study of the large time behaviour of μ_N when the above ODE has multiple stable equilibria.

An Example: Interaction in WiFi networks

- N nodes accessing a common wireless medium.
- Interaction among nodes via the distributed MAC protocol.
- State $X_n^N(t)$ represents aggressiveness of packet transmission.
-

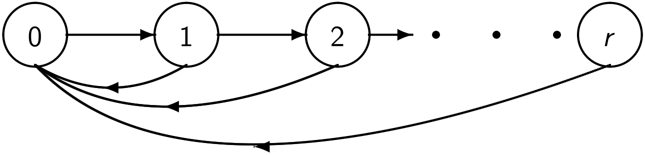


Figure: Set of allowed transitions in WiFi example

- State evolution:
  - Becomes less aggressive after a collision.
  - Moves to the most aggressive state after a successful packet transmission.

A sample path of μ_N in WiFi example

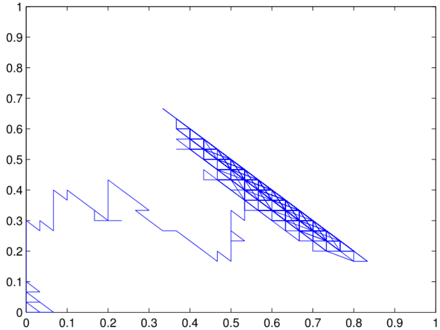


Figure: Evolution of states in a WiFi network under the MAC protocol

- Multiple stable regions in the system. Transition between two stable region occur over large time durations.
- Metastability: system exhibits very different behaviour over multiple time scales.

Large deviations

- S: a metric space. $\{X_N\}_{N\geq 1}$ is a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-NI_A\}$ where $I_A = \inf_{x\in A} I(x)$.

Large deviations

- S: a metric space. $\{X_N\}_{N\geq 1}$ is a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-NI_A\}$ where $I_A = \inf_{x\in A} I(x)$.
- $\{X_N\}_{N\geq 1}$ is said to satisfy the large deviation principle (LDP) with rate function $I : S \to [0, +\infty]$ if
    - for each $M > 0$, $\{x \in S : I(x) \leq M\}$ is a compact subset of S,
    - for each open set $G \subset S$,

$$\liminf_{N\to\infty} \frac{1}{N} \log P(X_N \in G) \geq -\inf_{x\in G} I(x),$$

    - for each closed set $F \subset S$,

$$\limsup_{N\to\infty} \frac{1}{N} \log P(X_N \in F) \leq -\inf_{x\in F} I(x).$$

Large deviations: contraction principle

- S, T are metric spaces. $f : S \to T$ is continuous.
- $\{X_N\}$s are S- valued random variables. Define $Y_N = f(X_N)$.
- If $\{X_N\}$ satisfies the LDP with rate function I, then $\{Y_N\}$ satisfies the LDP with rate function

$$J(y) = \inf_{x \in S : y = f(x)} I(x).$$

  - Compactness of level sets:
    $\{y \in T : J(y) \leq M\} = f(\{x \in S : I(x) \leq M\})$.
  - Upper and lower bounds:
    $P(Y_N \in A) = P(X_N \in f^{-1}(A))$

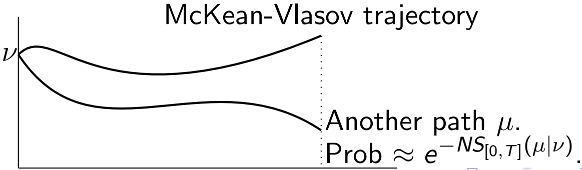
Large deviations of μ_N

Theorem (Léonard (1995), Borkar and Sundaresan (2012))

Let $\nu_N \to \nu$ weakly. Then μ_N satisfies the LDP in $D([0,T], M_1(\mathcal{Z}))$ with rate function $S_{[0,T]}(\cdot|\nu)$ defined as follows. If $\mu(0) = \nu$ and $[0,T] \ni t \mapsto \mu_t \in M_1(\mathcal{Z})$ is absolutely continuous,

$$S_{[0,T]}(\mu|\nu) = \int_{[0,T]} \sup_{\alpha \in \mathbb{R}^{|\mathcal{Z}|}} \Big\{ \langle \alpha, \dot{\mu}_t - \Lambda^*_{\mu_t}\mu_t \rangle - \sum_{(z,z') \in \mathcal{E}} \tau(\alpha(z') - \alpha(z)) \lambda_{z,z'}(\mu_t)\mu_t(z) \Big\},$$

else $S_{[0,T]}(\mu|\nu) = +\infty$.



Large deviations of $\mu_N(T)$

- The mapping $\pi_T : D([0, T], M_1(\mathcal{Z})) \to M_1(\mathcal{Z})$ is continuous. Use the contraction principle.

Theorem (Borkar and Sundaresan (2012))

Let $\nu_N \to \nu$ in weakly. Then $\{\mu_N(T)\}_{T\geq 1}$ satisfies the LDP on $M_1(\mathcal{Z})$ with rate function

$$S_T(\xi|\nu) = \inf\{S_{[0,T]}(\mu|\nu) : \mu(0) = \nu, \mu(T) = \xi, \mu \in \mathcal{AC}[0, T]\}.$$

- Quasipotential (Freidlin and Wentzell (1984)):

$$V(\nu, \xi) = \inf\{S_T(\xi|\nu), T > 0\}.$$

"Cost" to go from ν to ξ in arbitrary time.

- We say that $\nu \sim \xi$ if $V(\nu, \xi) = 0$ and $V(\xi, \nu) = 0$.

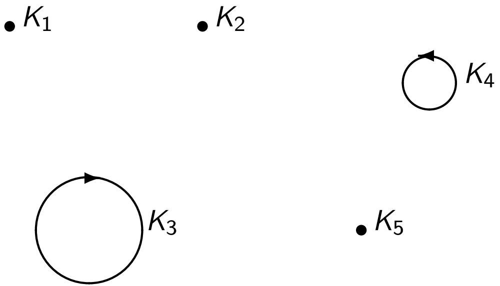
Some notations

- Assumptions:
  - There exists a finite number of compact sets $K_1, K_2, \ldots, K_l$ such that
    - For each $i = 1, 2, \ldots l$, $\nu_1, \nu_2 \in K_i$ implies $\nu_1 \sim \nu_2$.
    - For each $i \neq j$, $\nu_1 \in K_i$ and $\nu_2 \in K_j$ implies $\nu_1 \nsim \nu_2$.
    - Every ω-limit set of the McKean-Vlasov equation lies completely in one of the compact sets K_i.
- Cost of transport from K_i to K_j without touching the others:

$$\tilde{V}(K_i, K_j) = \inf\{S_{[0,T]}(\mu|\nu) : \nu \in K_i, \mu(t) \notin \cup_{k \neq i,j} K_k \\ \text{for all } 0 \leq t \leq T, \mu(T) \in K_j, T > 0\}.$$

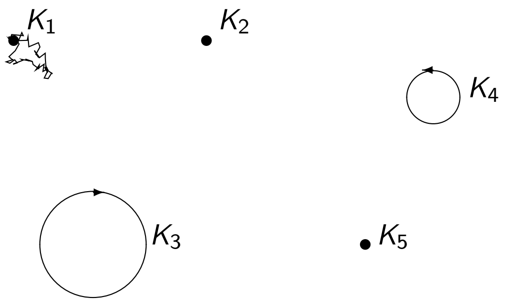
- $G(W)$: W-graph, $W \subset \{1, 2, \ldots, l\}$. Each node in $\{1, 2, \ldots, l\} \setminus W$ has exactly one outgoing arrow and there are no cycles.

Some notations - example

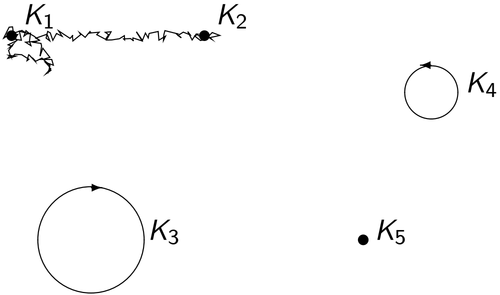


- $G = \{(1,2),(2,3),(3,4)\}$ is an example of a $\{4,5\}$-graph.
- $\tilde{V}$ values: $\begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$

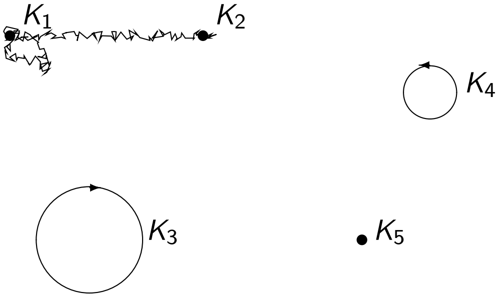
Approximation of μ_N using a discrete chain



Approximation of μ_N using a discrete chain



Approximation of μ_N using a discrete chain



- τ_n: hitting time of μ_N in a given neighbourhood of K_i's.
- Hitting time chain: $Z_n^N = \mu_N(\tau_n)$, $n \geq 1$

Estimates on one step transition probability

Lemma (Borkar and Sundaresan (2012))

Given $\varepsilon > 0$, there exist $\rho_0 > 0$ and $N_0 \geq 1$ such that, for any $\rho_2 < \rho_0$, there exists $\rho_1 < \rho_2$ such that for any $\nu \in [K_i]_{\rho_2} \cap M_1^N(\mathcal{Z})$ and $N \geq N_0$, the one-step transition probability of the chain Z^N satisfies

$$\exp\{-N(\tilde{V}(K_i, K_j) + \varepsilon)\} \leq P(\nu, \gamma_j) \leq \exp\{-N(\tilde{V}(K_i, K_j) - \varepsilon)\}.$$

- Can reconstruct the invariant measure of μ_N from that of Z^N - Khasminskii formula.
- Stationary LDP for $\mu_N(\infty)$ with rate function

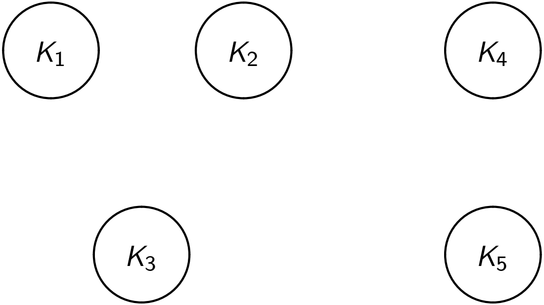
$$s(\xi) = \min_{1 \leq i \leq l} \{W(i) + V(K_i, \xi)\} - \min_{1 \leq j \leq l} W(j),$$

where

$$W(i) = \min_{g \in G(i)} \sum_{(m,n) \in g} \tilde{V}(m, n).$$

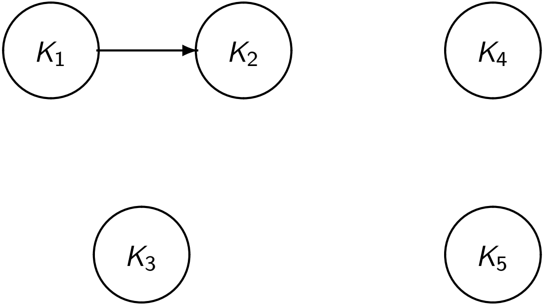
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



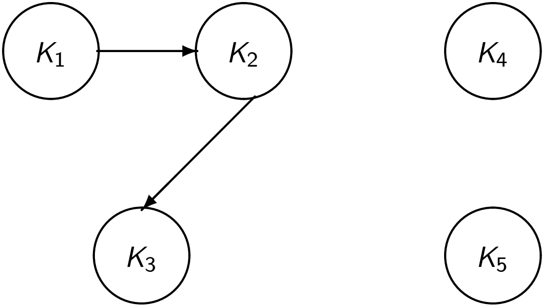
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



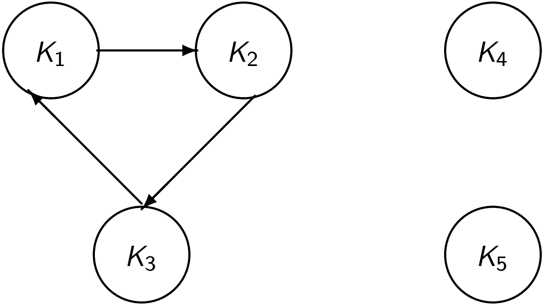
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



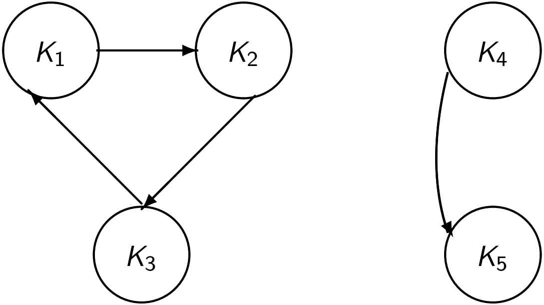
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



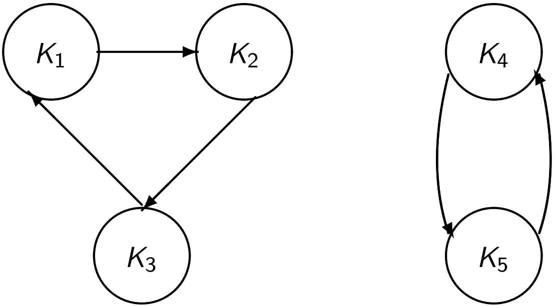
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



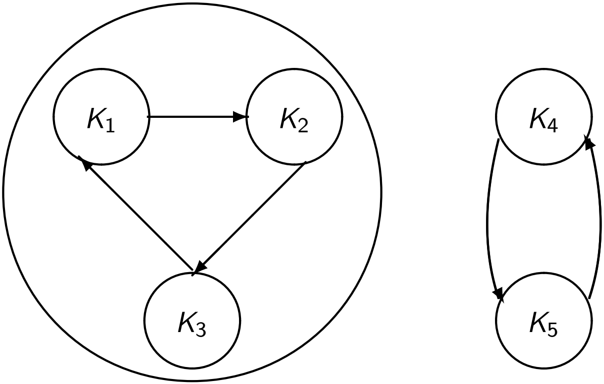
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



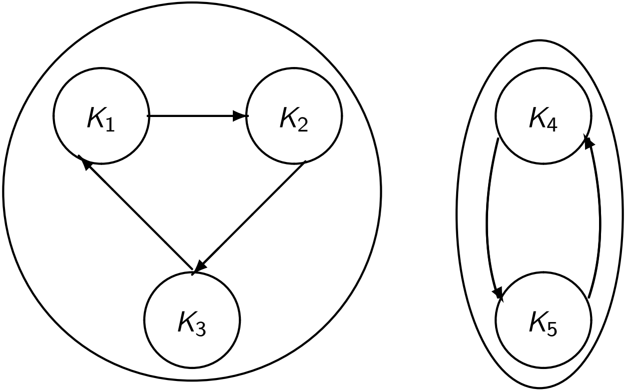
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



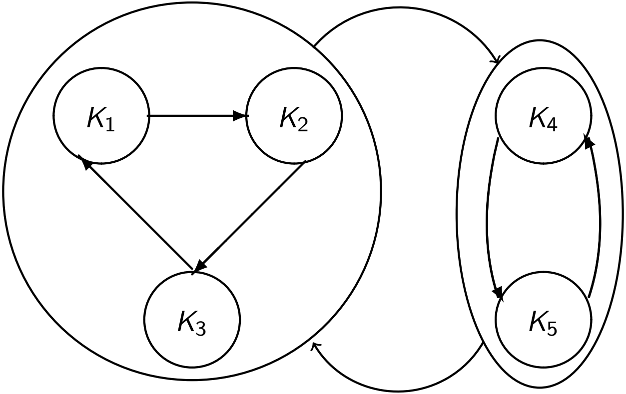
Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



Decomposition into cycles - example

$$\tilde{V} \text{ values: } \begin{pmatrix} 0 & 4 & 9 & 13 & 12 \\ 7 & 0 & 5 & 10 & 11 \\ 6 & 8 & 0 & 17 & 15 \\ 3 & 6 & 8 & 0 & 2 \\ 5 & 7 & 10 & 3 & 0 \end{pmatrix}$$



Decomposition into cycles

- Level-0 cycle: K_i's.
- Define $\tilde{V}(K_i) = \min_m \tilde{V}(K_i, K_m)$. We say that $i \to j$ if $\tilde{V}(K_i, K_j) = \tilde{V}(K_i)$.
- $i \Rightarrow j$ if $i \to i_1 \to \cdots i_k \to j$.

Definition (Hwang and Sheu (1990))

A cycle π is a subgraph of L satisfying

1. $i \in \pi$ and $i \Rightarrow j$ implies $j \in \pi$.
2. For any $i \neq j$ in π, we have $i \Rightarrow j$ and $j \Rightarrow i$.

- Similarly, we can define a hierarchy of cycles.

Large time behaviour

- Cycles are "very stable" subsets of $M_1(\mathcal{Z})$.

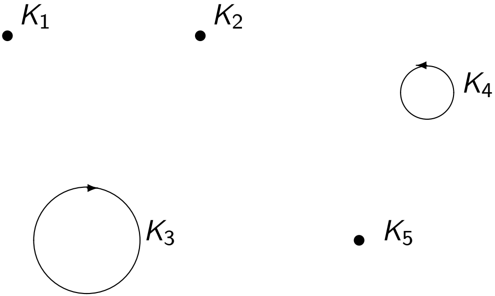
Theorem

Let π_1^k, π_2^k be k-cycles, $\pi_1^k \neq \pi_2^k$, and $K_i \in \pi_1^k$. Let $W = L \setminus \pi_1^k$. Given $\varepsilon > 0$, there exist $\rho > 0$ and $N_0 \geq 1$ such that for all $\rho_1 \leq \rho$, $\nu \in \gamma_i \cap M_1^N(\mathcal{Z})$ and $N \geq N_0$, we have

$$\exp\{-N(\tilde{V}(\pi_1^k, \pi_2^k) - \tilde{V}(\pi_1^k) + \varepsilon)\} \leq P_\nu(\mu_N(\hat{\tau}_W) \in \gamma_{\pi_2^k}) \\ \leq \exp\{-N(\tilde{V}(\pi_1^k, \pi_2^k) - \tilde{V}(\pi_1^k) - \varepsilon)\}.$$

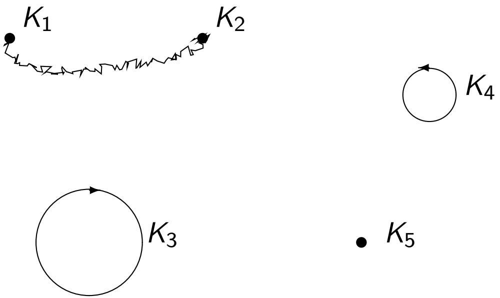
- Lower bound: construct a specific path.
- Upper bound: Use strong Markov property and the uniform LDP of μ_N.

Mixing of μ_N



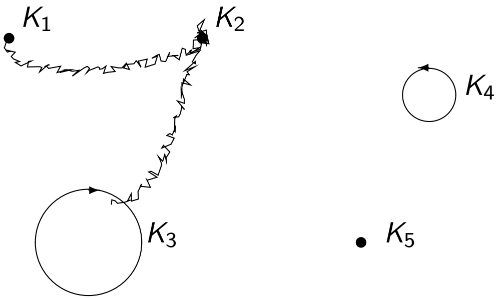
- Suppose that $W(K_3) = \min_{i \neq 3} W(K_i)$.
- μ_N mixes well if it reaches a neighbourhood of K_3.

Mixing of μ_N



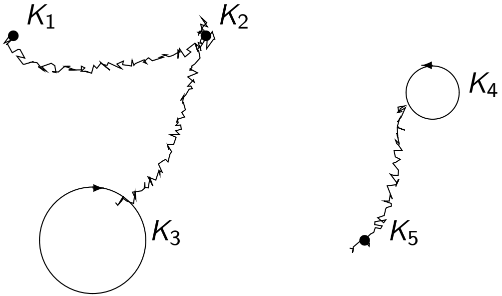
- Suppose that $W(K_3) = \min_{i \neq 3} W(K_i)$.
- μ_N mixes well if it reaches a neighbourhood of K_3.

Mixing of μ_N



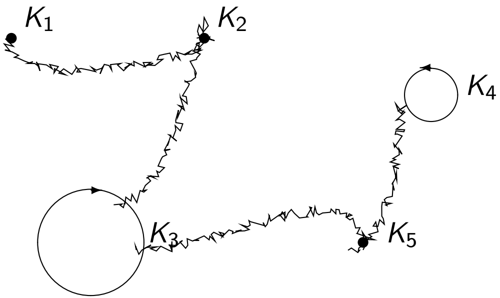
- Suppose that $W(K_3) = \min_{i \neq 3} W(K_i)$.
- μ_N mixes well if it reaches a neighbourhood of K_3.

Mixing of μ_N



- Suppose that $W(K_3) = \min_{i \neq 3} W(K_i)$.
- μ_N mixes well if it reaches a neighbourhood of K_3.

Mixing of μ_N



- Suppose that $W(K_3) = \min_{i \neq 3} W(K_i)$.
- μ_N mixes well if it reaches a neighbourhood of K_3.

Mixing of μ_N

- Define

$$\Lambda = \min\{\tilde{V}(g) : g \in G(i), i \in L\} \\ - \min\{\tilde{V}(g) : g \in G(i,j), i,j \in L, i \neq j\}.$$

Theorem

Given $\varepsilon > 0$, there exist $\delta_0 > 0$, $\rho > 0$ and $N_0 \geq 1$ such that for all $\rho_1 \leq \rho$, $N \geq N_0$, $\nu \in M_1^N(\mathcal{Z})$, we have

$$P_{T_0}(\nu, \gamma_{i_0}) \geq \exp\{-N\varepsilon\},$$

where $T_0 = \exp\{N(\Lambda - \delta_0)\}$.

- Proof via the estimates of large time behaviour.

Convergence to the invariant measure

Theorem

Given $\delta > 0$, *there exist* $\varepsilon > 0$ *and* $N_0 \geq 1$ *such that for all* $\nu \in M_1^N(\mathcal{Z})$ *and* $N \geq N_0$

$$|E_\nu(f(\mu_N(T))) - \langle f, \wp_N \rangle| \leq \|f\|_\infty \exp\{-\exp(N\varepsilon)\},$$

where $T = \exp\{N(\Lambda + \delta)\}$ *and* $f \in B(M_1(\mathcal{Z}))$.

Asymptotics of the second eigenvalue

- μ_N is a Markov process with infinitesimal generator

$$L^N f(\xi) = N \sum_{(z,z')\in\mathcal{E}} \xi(z)\lambda_{z,z'}(\xi)\left[f\left(\xi + \frac{\delta_{z'}}{N} - \frac{\delta_z}{N}\right) - f(\xi)\right].$$

- If μ_N is reversible (i.e. L^N is self-adjoint in $L^2(\wp_N)$), spectral decomposition of L^N tells us that

$$E_\nu f(\mu_N(t)) = \langle f, \wp_N\rangle + \sum_{k\geq 2} e^{-t\lambda_k^N}(f, u_k^N)u_k^N(\nu),$$

- Mixing time of μ_N is governed by λ_2^N.

Theorem

$$\lim_{N\to\infty} \frac{1}{N}\log\lambda_2^N = -\Lambda.$$

Convergence to the global minimum

- Fix $c > 0$. Start with $N_0 = \min\{n \in \mathbb{N} : \exp\{nc\} - 2 \geq 0\}$ particles.
- Let $t_{N_0} = 0$. Add a particle at times $t_N = \exp\{Nc\} - 2$, $N > N_0$, with a certain state.
- $\bar{\mu}$: the resulting process.
- Small c: particles are added too frequently; $\bar{\mu}$ could get trapped in a local minimum of s depending on $\bar{\mu}(0)$.
- Large c: sufficient time for exploration, $\bar{\mu}$ converges to i_0.

Theorem

For $c > c^*$ and any $\rho_1 > 0$,

$$P_{0,\nu}(\bar{\mu}(t) \in \gamma_{i_0}) \to 1$$

as $t \to \infty$, uniformly for all $\nu \in M_1^{N_0}(\mathcal{Z})$.

Conclusion

- Study of large time behaviour of finite state mean-field interacting particle systems
  - Exit time estimates. Decomposition into cycles.
  - Convergence of μ_N to its invariant measure.
- Scaling of $\lambda_2^N \sim \exp\{-N\Lambda\}$ when μ_N is reversible.
- Convergence of a controlled process to the global minimum of a certain entropy functional.

Reference: arXiv:1909.03805

Thank You