A cutoff transition for repeated averages

Lingfu Zhang
(Joint work with Sourav Chatterjee, Persi Diaconis, and Allan Sly)

Princeton University
Department of Mathematics

May 7, 2021

Columbia-Princeton Probability Day 2021

arXiv:1911.02756

Consider a following Markov chain on $\mathbb{R}^n$ (Bourgain '80):

Start with $x_0 = (x_{0,1}, \ldots, x_{0,n}) \in \mathbb{R}^n$. At step k, given x_k,

1. pick two distinct coordinates I and J uniformly at random,
2. replace both $x_{k,I}$ and $x_{k,J}$ by $(x_{k,I} + x_{k,J})/2$,
3. keep all other coordinates the same,

to obtain x_{k+1}.

Consider a following Markov chain on $\mathbb{R}^n$ (Bourgain '80):

Start with $x_0 = (x_{0,1}, \ldots, x_{0,n}) \in \mathbb{R}^n$. At step k, given x_k,

1. pick two distinct coordinates I and J uniformly at random,
2. replace both $x_{k,I}$ and $x_{k,J}$ by $(x_{k,I} + x_{k,J})/2$,
3. keep all other coordinates the same,

to obtain x_{k+1}.

Background: quantum computing; distribution of wealth; etc.

Consider a following Markov chain on $\mathbb{R}^n$ (Bourgain '80):

Start with $x_0 = (x_{0,1}, \ldots, x_{0,n}) \in \mathbb{R}^n$. At step k, given x_k,

1. pick two distinct coordinates I and J uniformly at random,
2. replace both $x_{k,I}$ and $x_{k,J}$ by $(x_{k,I} + x_{k,J})/2$,
3. keep all other coordinates the same,

to obtain x_{k+1}.

Background: quantum computing; distribution of wealth; etc.

Related studies/models: [Chatterjee, Seneta, 77] [Feller, 68]; consensus algorithm [Olshevsky, Tsitsiklis, 09] [Shah, 08]; local iterated averaging [Diaconis, Saloff-Coste, 12]; convergence on general graphs [Aldous, Lanoue, 12]; Deffuant model [Häggström, 12] [Lanchier, 12]; Kac walk [Kac, 54] [Pillai, Smith, 17]

The repeated average process

- Asymptotic behaviour?

- Asymptotic behaviour?

This is not irreducible: almost surely converge to $(\bar{x}_0, \ldots, \bar{x}_0)$ for $\bar{x}_0 = \frac{1}{n}\sum_{i=1}^{n} x_{0,i}$.

The repeated average process

- Asymptotic behaviour?

This is not irreducible: almost surely converge to $(\bar{x}_0, \ldots, \bar{x}_0)$ for $\bar{x}_0 = \frac{1}{n}\sum_{i=1}^{n} x_{0,i}$.

- Question: rate of convergence?

- Asymptotic behaviour?

This is not irreducible: almost surely converge to $(\bar{x}_0, \dots, \bar{x}_0)$ for $\bar{x}_0 = \frac{1}{n}\sum_{i=1}^{n} x_{0,i}$.

- Question: rate of convergence?

In terms of L^2: can be explicitly computed

$$\mathbb{E}\left(\sum_{i=1}^{n}(x_{k,i} - \bar{x}_0)^2\right) = \left(1 - \frac{1}{n-1}\right)^k \sum_{i=1}^{n}(x_{0,i} - \bar{x}_0)^2$$

- Asymptotic behaviour?

This is not irreducible: almost surely converge to $(\bar{x}_0, \dots, \bar{x}_0)$ for $\bar{x}_0 = \frac{1}{n}\sum_{i=1}^n x_{0,i}$.

- Question: rate of convergence?

In terms of L^2: can be explicitly computed

$$\mathbb{E}(\sum_{i=1}^{n}(x_{k,i} - \bar{x}_0)^2) = (1 - \frac{1}{n-1})^k \sum_{i=1}^{n}(x_{0,i} - \bar{x}_0)^2$$

More relevant (and difficult): L^1 distance to $(\bar{x}_0, \dots, \bar{x}_0)$?

Consider $T(k) = \sum_{i=1}^n |x_{k,i} - \bar{x}_0|$.

The repeated average process

- Asymptotic behaviour?

This is not irreducible: almost surely converge to $(\bar{x}_0, \ldots, \bar{x}_0)$ for $\bar{x}_0 = \frac{1}{n}\sum_{i=1}^{n} x_{0,i}$.

- Question: rate of convergence?

In terms of L^2: can be explicitly computed

$$\mathbb{E}\Big(\sum_{i=1}^{n}(x_{k,i} - \bar{x}_0)^2\Big) = \Big(1 - \frac{1}{n-1}\Big)^k \sum_{i=1}^{n}(x_{0,i} - \bar{x}_0)^2$$

More relevant (and difficult): L^1 distance to $(\bar{x}_0, \ldots, \bar{x}_0)$?

Consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

The initial condition matters:
take $x_0 = (1, 0, \ldots, 0)$, a worst case by linearity.

Take $x_0 = (1, 0, \ldots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

The Cutoff

Take $x_0 = (1, 0, \ldots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

A pre-cutoff:

The Cutoff

Take $x_0 = (1, 0, \ldots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

A pre-cutoff:

By L^2 decay and Cauchy-Schwarz, for $k = n \log n + cn$ with $c > 0$ we have

$$\mathbb{E}(T(k)) \leq e^{-c/2}.$$

The Cutoff

Take $x_0 = (1, 0, \ldots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

A pre-cutoff:

By L^2 decay and Cauchy-Schwarz, for $k = n \log n + cn$ with $c > 0$ we have

$$\mathbb{E}(T(k)) \leq e^{-c/2}.$$

For $k = (\frac{1}{2} - \epsilon) n \log n$ there are only $o(n)$ number of non-zero entries and hence $T(k) = 2 - o(1)$.

The Cutoff

Take $x_0 = (1, 0, \ldots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

A pre-cutoff:

By L^2 decay and Cauchy-Schwarz, for $k = n \log n + cn$ with $c > 0$ we have

$$\mathbb{E}(T(k)) \leq e^{-c/2}.$$

For $k = (\frac{1}{2} - \epsilon) n \log n$ there are only $o(n)$ number of non-zero entries and hence $T(k) = 2 - o(1)$.

How it decays between $\frac{n}{2} \log n$ and $n \log n$?

The Cutoff

Take $x_0 = (1, 0, \ldots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

The Cutoff

Take $x_0 = (1, 0, \ldots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

Theorem (Chatterjee-Diaconis-Sly-Z. '20)

As $n \to \infty$, we have in probability convergence

$$T(\theta n \log n) \to 2, \qquad \text{for any } \theta < (2 \log 2)^{-1},$$

$$T(\theta n \log n) \to 0, \qquad \text{for any } \theta > (2 \log 2)^{-1}.$$

The Cutoff

Take $x_0 = (1, 0, \dots, 0)$, and consider $T(k) = \sum_{i=1}^{n} |x_{k,i} - \bar{x}_0|$.

Theorem (Chatterjee-Diaconis-Sly-Z. '20)

As $n \to \infty$, we have in probability convergence

$$T(\theta n \log n) \to 2, \quad \text{for any } \theta < (2 \log 2)^{-1},$$

$$T(\theta n \log n) \to 0, \quad \text{for any } \theta > (2 \log 2)^{-1}.$$

This also has a Gaussian cutoff profile.

Theorem (Chatterjee-Diaconis-Sly-Z. '20)

Let $\Phi : \mathbb{R} \to [0, 1]$ be the cumulative distribution function of the standard normal distribution. For any $a \in \mathbb{R}$, as $n \to \infty$ we have

$$T(\lfloor n(\log_2(n) + a\sqrt{\log_2(n)})/2 \rfloor) \to 2\Phi(-a)$$

in probability .

Proof Ideas

Let's run repeated average:

Let's run repeated average:
W.h.p., initially be like

$$1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

Proof Ideas

Let's run repeated average:
W.h.p., initially be like

$$1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

Proof Ideas

Let's run repeated average:
W.h.p., initially be like

$$1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{4}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, \frac{1}{4}, 0, 0, 0, 0, 0,$$

Proof Ideas

Let's run repeated average:
W.h.p., initially be like

$$1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{4}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, \frac{1}{4}, 0, 0, 0, 0, 0,$$

$$\frac{1}{4}, 0, 0, 0, 0, \frac{1}{4}, 0, 0, 0, \frac{1}{4}, 0, 0, 0, \frac{1}{4}, 0,$$

Proof Ideas

Let's run repeated average:
W.h.p., initially be like

$$1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{4}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, \frac{1}{4}, 0, 0, 0, 0, 0,$$

$$\frac{1}{8}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, \frac{1}{4}, 0, \frac{1}{8}, 0, 0, 0,$$

Let's run repeated average:
W.h.p., initially be like

$$1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

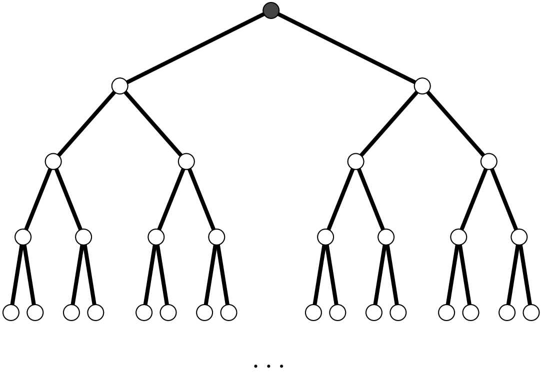
$$\frac{1}{4}, 0, 0, 0, 0, \frac{1}{2}, 0, 0, 0, \frac{1}{4}, 0, 0, 0, 0, 0,$$

$$\frac{3}{8}, 0, 0, 0, 0, \frac{3}{8}, 0, 0, 0, \frac{1}{4}, 0, 0, 0, 0, 0,$$

Unlikely to happen.

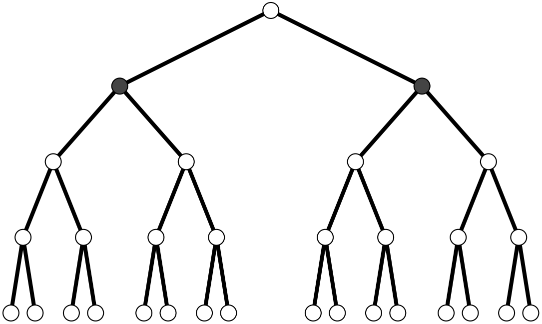
Proof Ideas

Tree structure:



Proof Ideas

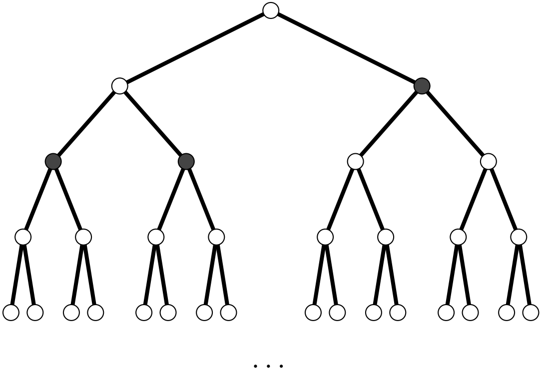
Tree structure:



. . .

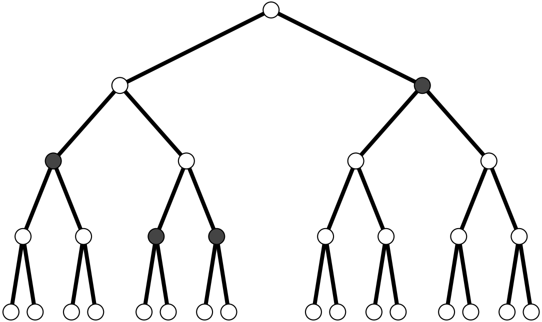
Proof Ideas

Tree structure:



Proof Ideas

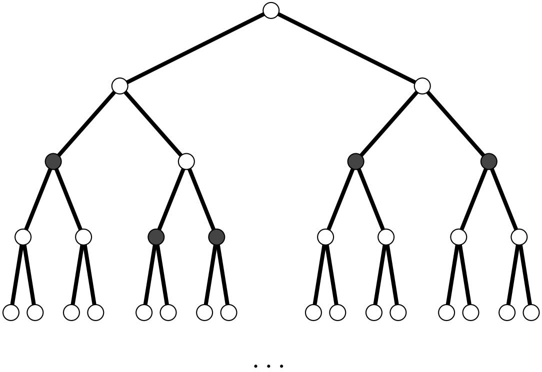
Tree structure:



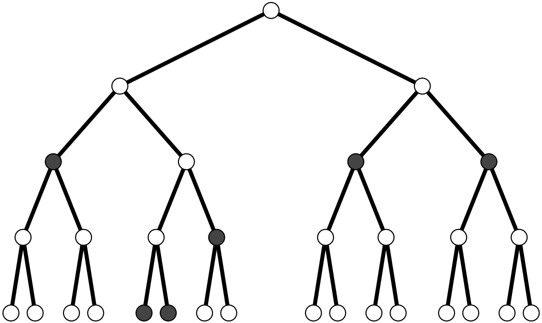
. . .

Proof Ideas

Tree structure:



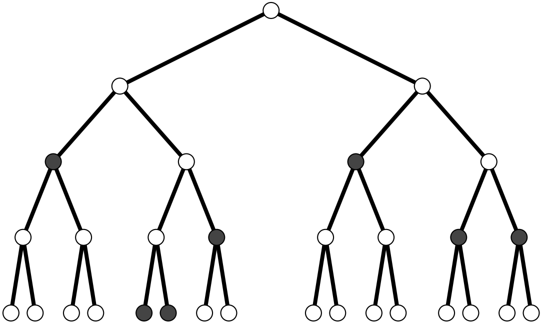
Tree structure:



. . .

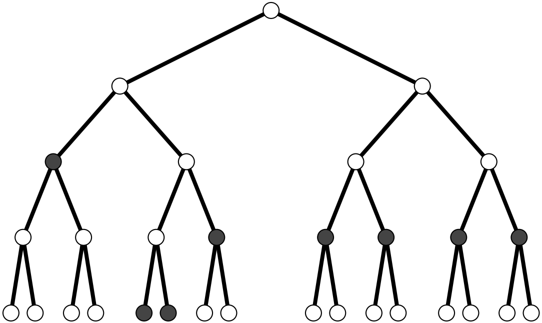
Proof Ideas

Tree structure:



. . .

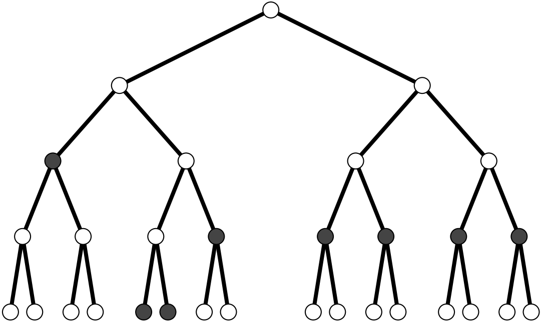
Tree structure:



. . .

Proof Ideas

Tree structure:

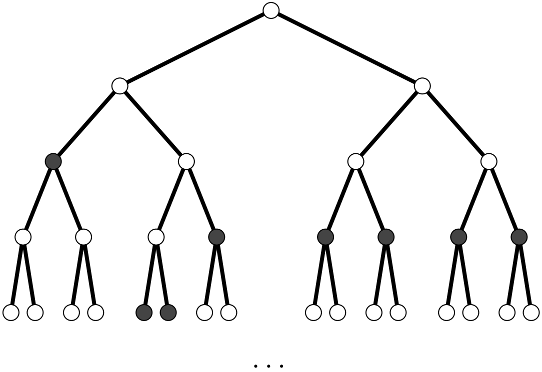


$\cdots$

Each particle at level i corresponds to one 2^{-i}.

Proof Ideas

Tree structure:

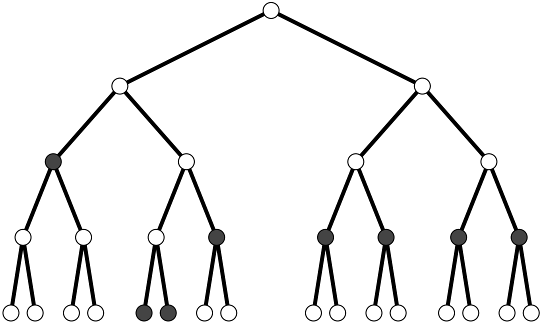


Each particle at level i corresponds to one 2^{-i}.

Critical level $\log_2(n)$: a particle at level $\log_2(n)$ is reached at time $\frac{n}{2}(\log_2(n) + \sqrt{\log_2(n)}\mathcal{N}(0,1))$.

Proof Ideas

Tree structure:

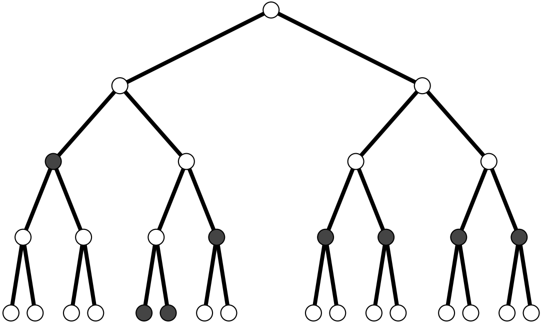


$\cdots$

$\Rightarrow$ For $k < \frac{n}{2}(\log_2(n) - C\sqrt{\log_2(n)})$: most coordinates are 0.

Proof Ideas

Tree structure:



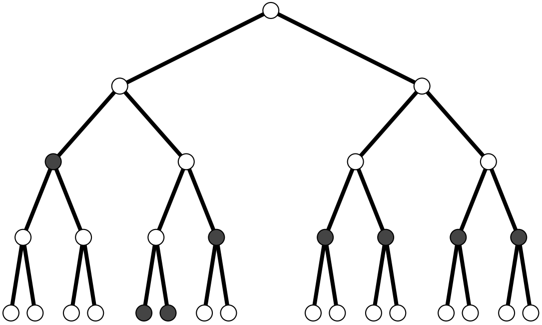
$\cdots$

$\Rightarrow$ For $k < \frac{n}{2}(\log_2(n) - C\sqrt{\log_2(n)})$: most coordinates are 0.

1. $T(k) = 2 - o(1)$.
2. This tree is a good approximation.

Proof Ideas

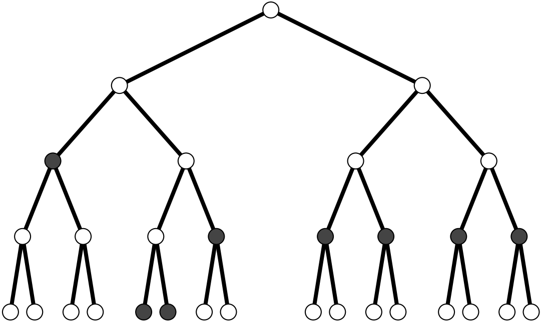
Tree structure:



$\cdots$

For $k \approx \frac{n}{2} \log_2(n)$: most weights are $O(\frac{1}{n})$.

Tree structure:



$\cdots$

For $k \approx \frac{n}{2}\log_2(n)$: most weights are $O(\frac{1}{n})$.

$\Rightarrow$ L^2-distance is of order $O(n^{-1})$; run for Cn more steps to get $o(n^{-1})$, then by Cauchy-Schwarz $T(k) = o(1)$.

Thank you!

ALDOUS, D. and LANOUE, D. (2012). A lecture on the averaging process. *Probab. Surv.,* **9**, 90–102.

CHATTERJEE, S., DIACONIS, S., SLY, A., ZHANG, L. (2018). A phase transition for repeated averages. *Ann. Probab.,* to appear.

CHATTERJEE, S. and SENETA, E. (1977). Towards consensus: some convergence theorems on repeated averaging. *J. Appl. Probab.,* **14** no. 1, 89–97.

DIACONIS, P. and SALOFF-COSTE, L. (2014). Convolution powers of complex functions on $\mathbb{Z}$. *Math. Nachr.,* **287** no. 10, 1106–1130.

FELLER, W. (1968). *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney.

HÄGGSTRÖM, O. (2012). A pairwise averaging procedure with application to consensus formation in the Deffuant model. *Acta Applicandae Math.,* **119** no. 1, 185–201.

KAC, M. (1954) Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, no. 3, 171–197.

LANCHIER, N. (2012). The critical value of the Deffuant model equals one half. *ALEA Lat. Am. J. Probab. Math. Stat.,* **9** no. 2, 383–402.

OLSHEVSKY, A. and TSITSIKLIS, J. N. (2009). Convergence speed in distributed consensus and averaging. *SIAM J. Control Optim.,* **48**, 33–55.

PILLAI, N. S. and SMITH, A. (2017). Kac's walk on n-sphere mixes in $n \log n$ steps. *Ann. Appl. Probab.,* **27** no. 1, 631–650.

SHAH, D. (2008). Gossip algorithms. *Found. Trends Networking,* **3**, 1–125.