

Distance between configurations in MCMC simulations and the geometric optimization of the tempering algorithms

Nobuyuki Matsumoto (Kyoto University)

based on work with Masafumi Fukuma (Kyoto University)

Naoya Umeda (PwC)

JHEP 1712 001 (2017) [arXiv:1705.06097],

JHEP 1811 060 (2018) [arXiv:1806.10915],

[arXiv:1906.04243], and work in progress

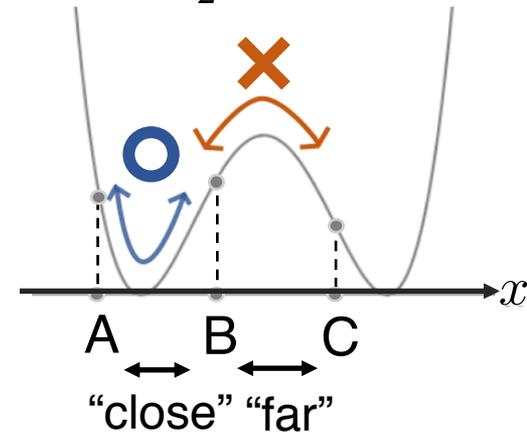
Short summary

- In Markov chain Monte Carlo (MCMC) simulations, transition between configs which belong to different modes is difficult.

➡ We enumerate this difficulty of transition as *distance between configurations*.

[Fukuma, NM, Umeda (2017)]

e.g. $S(x) = \frac{\beta}{2}(x^2 - 1)^2$ [$\beta \gg 1$]



- We consider an extremely multimodal system with the simulated tempering, and show:

- ① the geometry of the **extended**, **coarse-grained** config space is **asymptotically Euclidean AdS**
- ② optimized form of the tempering parameter is obtained by **a flat metric** in the **extended** direction.

[Fukuma, NM, Umeda (2017, 2018)]

- We also discuss the optimized form of the **tempering parameter in the tempered Lefschetz thimble method** (TLTM, an algorithm for the sign problem)

[Fukuma, Umeda (2017), Fukuma, NM, Umeda (2019)]

1. Definition of distance
2. Distance for the simulated tempering
3. Emergence of AdS geometry
4. Geometrical optimization of the tempering parameter
5. Comment: application to TLTM

1. Definition of distance

[Fukuma, NM, Umeda (2017)]

Setup

- defs
- $\mathcal{M} \equiv \{x\}$: config space
 - $P(x|y)$: (1-step) transition probability from y to x
 - $\{x_i\}$: Markov chain with the transition matrix P
$$\left[\cdots \xleftarrow{P} x_{n+1} \xleftarrow{P} x_n \xleftarrow{P} \cdots \xleftarrow{P} x_1 \xleftarrow{P} x_0 \right]$$

assumption for P

- ① unique convergence to the equilibrium distribution $p_{eq}(x) \equiv \frac{1}{Z} e^{-S(x)}$
- ② detailed balance condition : $P(x|y) p_{eq}(y) = P(y|x) p_{eq}(x)$
$$\left[\begin{array}{l} S(x) : \text{action} \\ Z \equiv \int dx e^{-S(x)} \end{array} \right]$$
- ③ positive definiteness $\left(\text{If } P \text{ has negative eigenvalues, we instead use } P^2. \right)$

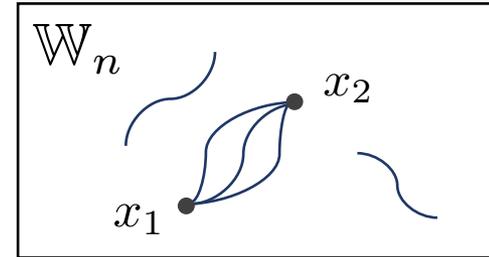
defs

- bra-ket notation : $P(x|y) \equiv \langle x | \hat{P} | y \rangle$
- “transfer matrix” :
$$\hat{T} \equiv e^{S(\hat{x})/2} \cdot \hat{P} \cdot e^{-S(\hat{x})/2} \left(\begin{array}{l} \text{positive symmetric} \\ \because \text{the above assumptions } \textcircled{1} \sim \textcircled{3} \end{array} \right)$$

Definition of distance (1/2)

- Suppose that the Markov chain is in equilibrium.
- We note that a history of n-step transition makes a path in \mathcal{M} .

$$\left[\cdots \leftarrow \underbrace{x_{m+n} \leftarrow \cdots \leftarrow x_{m+1} \leftarrow x_m}_{\text{n-step transition}} \leftarrow \cdots \right]$$



- Let $\left\{ \begin{array}{l} \mathbb{W}_n : \text{set of all n-step transition paths} \\ \mathbb{W}_n(x_1, x_2) \subset \mathbb{W}_n : \text{subset of paths from } x_2 \text{ to } x_1. \end{array} \right.$

defs • connectivity :

$$f_n(x_1, x_2) \equiv \frac{|\mathbb{W}_n(x_1, x_2)|}{|\mathbb{W}_n|} \left[\begin{array}{l} \text{probability of finding an n-step path} \\ \text{from } x_2 \text{ to } x_1 \text{ in equilibrium} \end{array} \right]$$

$$= \langle x_1 | \hat{P}^n | x_2 \rangle p_{\text{eq}}(x_2) = f_n(x_2, x_1)$$

• normalized connectivity :

$$F_n(x_1, x_2) \equiv \frac{f_n(x_1, x_2)}{\sqrt{f_n(x_1, x_1) f_n(x_2, x_2)}} = \frac{\langle x_1 | \hat{T}^n | x_2 \rangle}{\sqrt{\langle x_1 | \hat{T}^n | x_1 \rangle \langle x_2 | \hat{T}^n | x_2 \rangle}}$$

$$\left[\begin{array}{l} 0 \leq F_n(x_1, x_2) \leq 1 \\ \because F_n(x_1, x_2) = \langle x_1, n/2 | x_2, n/2 \rangle \text{ with } |x, n/2\rangle \equiv \frac{\hat{T}^{n/2} |x\rangle}{\|\hat{T}^{n/2} |x\rangle\|} \\ \text{[inner product of normalized states]} \end{array} \right]$$

Definition of distance (2/2)

def • distance : [Fukuma, NM, Umeda (2017)]

$$d_n(x_1, x_2) \equiv \sqrt{-2 \ln F_n(x_1, x_2)} \left[= \sqrt{-\ln \left(\frac{\langle x_1 | \hat{T}^n | x_2 \rangle^2}{\langle x_1 | \hat{T}^n | x_1 \rangle \langle x_2 | \hat{T}^n | x_2 \rangle} \right)} \right]$$

note

• $0 \leq F_n(x_1, x_2) \leq 1 \implies d_n(x_1, x_2) \geq 0$

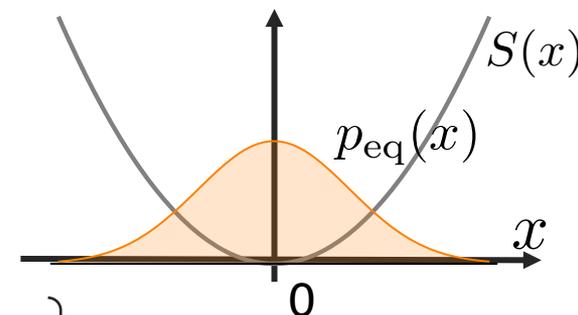
In particular,

$$\left[\begin{array}{l} x_1, x_2 \text{ easily communicate} \iff F_n(x_1, x_2) \sim 1 \iff d_n(x_1, x_2) \sim 0 \\ x_1, x_2 \text{ hardly communicate} \iff F_n(x_1, x_2) \sim 0 \iff d_n(x_1, x_2) \gg 1 \end{array} \right.$$

$\implies d_n(x_1, x_2)$ enumerates the difficulty of transition.

Example 1, Gaussian (unimodal)

- action : $S(x) = \frac{\beta}{2} x^2$



- transfer matrix : [algorithm : Langevin, ϵ : time increment]

$$\langle x_1 | \hat{T} | x_2 \rangle = \sqrt{\frac{\beta}{2\pi(1-e^{-2\beta\epsilon})}} \exp \left[-\frac{\beta}{4 \sinh(\beta\epsilon)} [(x_1^2 + x_2^2) \cosh(\beta\epsilon) - 2x_1x_2] \right]$$

- distance :

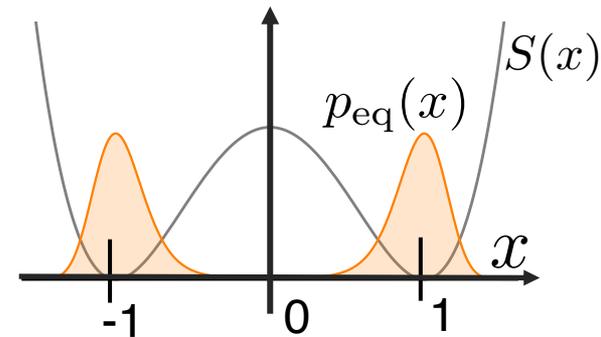
$$d_n^2(x_1, x_2) = \frac{\beta}{2 \sinh(\beta n \epsilon)} |x_1 - x_2|^2 \left[\begin{array}{l} \propto e^{-\beta n \epsilon} |x_1 - x_2|^2 \\ \text{[for } \beta \gg 1 \text{]} \end{array} \right]$$

[flat and translationally invariant geometry]

Example 2, Double-well (multimodal)

- action : $S(x) = \frac{\beta}{2}(x^2 - 1)^2$

- For $\beta \gg 1$, [algorithm : Langevin, ϵ : time increment]



- ① $d_n(x_1, x_2) = O(e^{-\beta n \epsilon / 2})$ when x_1, x_2 are in the same mode
 - ② $d_n(x_1, x_2) \propto \beta$ when x_1, x_2 are in different modes

\therefore When $\beta \gg 1$, we can approximate local equilibrium distributions by Gaussian.
With this approximation,

- ① can be shown by using a quantum mechanical argument with linear superpositions.
- ② can be shown by using an instanton analysis.

2. Distance for the simulated tempering

[Fukuma, NM, Umeda (2017)]

Simulated tempering [Marinari, Parisi (1992)]

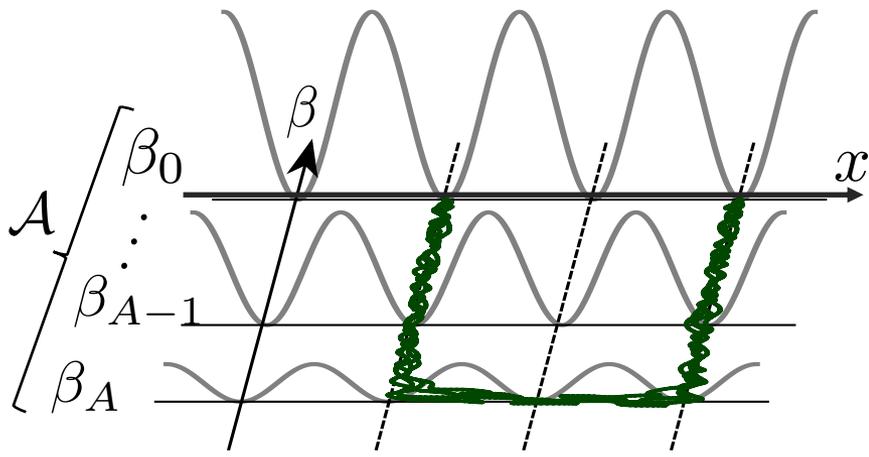
[Aim: speed up relaxation to equilibrium by making a bypass]

① Pick up a parameter in the action as a tempering parameter β
 [e.g. inverse temperature]

② Prepare a parameter set for β :

$$\mathcal{A} \equiv \{\beta_0, \beta_1, \dots, \beta_A\} = \{\beta_a\}_{a=0, \dots, A}$$

$$\left[\begin{array}{l} \beta_0 : \text{original parameter} \\ \beta_0 > \beta_1 > \dots > \beta_A \end{array} \right]$$



③ **Extend** the config sp in the β direction:

$$\mathcal{M} = \{x\} \longrightarrow \mathcal{M} \times \mathcal{A} = \{(x, \beta_a)\}$$

and set up a Markov chain on $\mathcal{M} \times \mathcal{A}$ s.t. it realizes the global equilibrium:

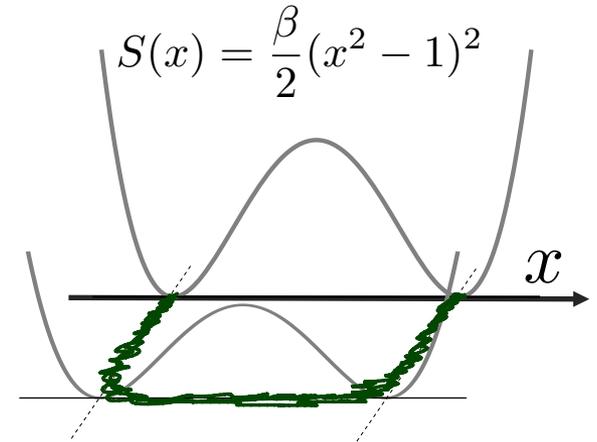
$$P_{eq}(x, \beta_a) \equiv w_a e^{-S(x; \beta_a)} \quad \left[\begin{array}{l} \{w_a\}: \text{weights, we set them to} \\ w_a = \frac{1}{(A+1) Z_a} \left[Z_a \equiv \int dx e^{-S(x; \beta_a)} \right] \end{array} \right]$$

[Transition between different modes easily occurs
 by passing through configs at lower β .]

④ After realizing global equilibrium,
 retrieve configs generated on the subspace $\beta = \beta_0$.

Decrease of distance

- The acceleration of the relaxation due to the tempering can be seen as a significant decrease of the distance.



[double-well case]

w/o tempering [$\beta_0 = 20$]

n (number of steps)	$d_n^2(-1, 1)$
10	39.1
50	19.2
100	16.9
500	13.2
1000	11.7
5000	8.46

significant decrease

w/ tempering [$A = 1, \beta_0 = 20, \beta_1 = 1$]

n (number of steps)	$d_n^2(-1, 1)$
10	26.5
50	7.16
100	4.35
500	0.708
1000	0.106
5000	7.81×10^{-4}



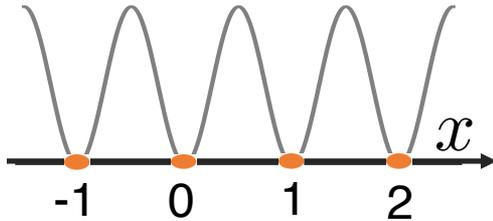
3. Emergence of AdS geometry

[Fukuma, NM, Umeda (2017, 2018)]

Coarse-grained configuration space

- We consider an extremely multimodal system.
As a typical example,

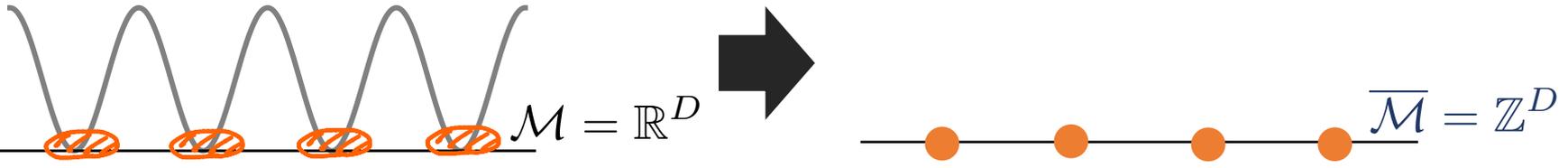
$$S(x; \beta) \equiv \beta \sum_{\mu=1}^D (1 - \cos(2\pi x_{\mu})) \quad \left[\begin{array}{l} \text{minima of the action} \\ = \text{a set of integer points} \end{array} \right]$$



- We note that

$$\begin{cases} d_n(x_1, x_2) \text{ is negligibly small when } x_1, x_2 \text{ are in the same mode.} \\ d_n(x_1, x_2) \text{ is large when } x_1, x_2 \text{ are in different modes.} \end{cases}$$

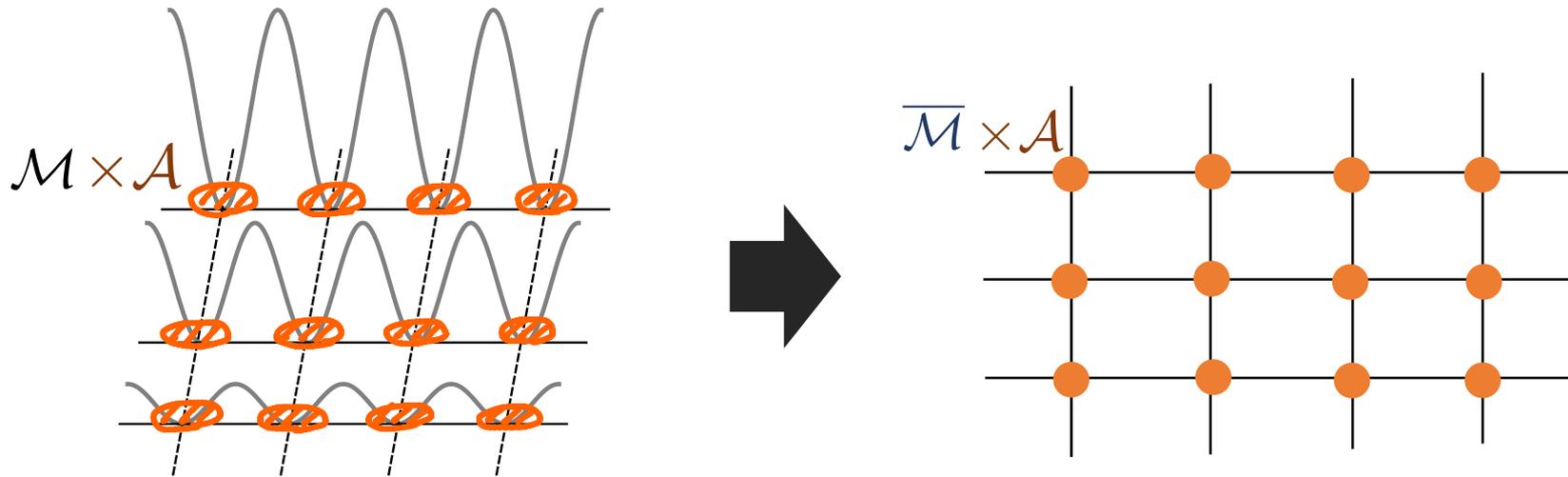
- When investigating the large scale structure of \mathcal{M} , we can identify configs in the same mode as a single config.



We write the obtained coarse-grained config space as $\overline{\mathcal{M}}$. $\left[\begin{array}{l} \text{for the cosine model,} \\ \overline{\mathcal{M}} = \mathbb{Z}^D \end{array} \right]$

Extended, coarse-grained configuration space

- When we implement the simulated tempering to this system, we can similarly coarse-grain the **extended** config space.



We write the obtained **extended, coarse-grained** config space as $\overline{\mathcal{M}} \times \mathcal{A}$.

Emergence of AdS geometry (1/2)

geometry of $\overline{\mathcal{M}} \times \mathcal{A}$

- ① The minima of the action are uniformly distributed in $\overline{\mathcal{M}}$.
 ➔ The metric does not depend on x .
- ② Transition in x direction is difficult for larger β
 ➔ We assume that the distance in x direction takes the form:

$$d_n^2((x, \beta), (x + dx, \beta)) = (\text{const.}) \beta^q \sum_{\mu=1}^D dx_{\mu}^2 \left[\text{power-like increasing function of } \beta \right]$$

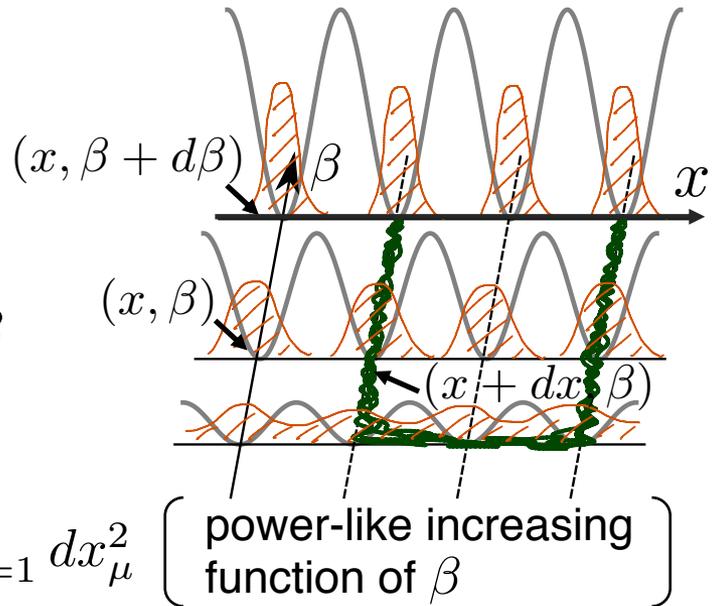
- ③ We approximate local equilibrium distributions by the Gaussian for $\beta \gg 1$. Then we can show that the distance in the β direction, $d_n((x, \beta), (x, \beta + d\beta))$ [$\beta \gg 1$] is invariant under rescaling $\beta \rightarrow \lambda\beta$

➔ $d_n^2((x, \beta), (x, \beta + d\beta)) \propto \frac{d\beta^2}{\beta^2}$

From ① - ③, we deduce that

$$ds^2 \equiv d_n^2((x, \beta), (x + dx, \beta + d\beta)) = (\text{const.})\beta^q \sum_{\mu=1}^D dx_{\mu}^2 + (\text{const.})\frac{d\beta^2}{\beta^2} \quad [\beta \gg 1]$$

(asymptotically AdS metric!) ➔ $= \frac{(\text{const.})}{z^2} (\sum_{\mu=1}^D dx_{\mu}^2 + dz^2) [z \propto \beta^{-q/2}]$ 13/19

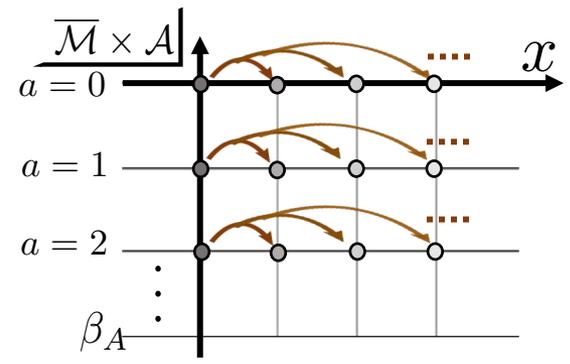


Emergence of AdS geometry (2/2)

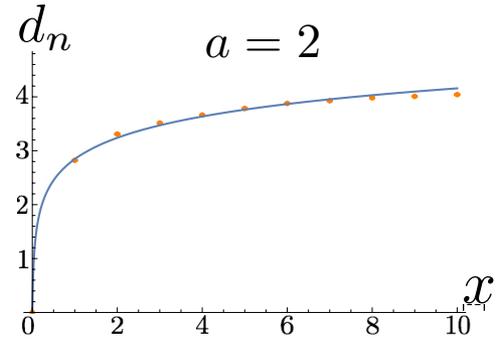
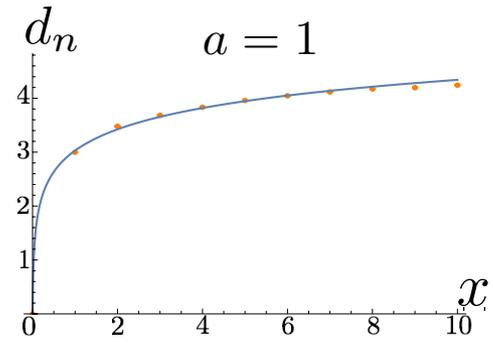
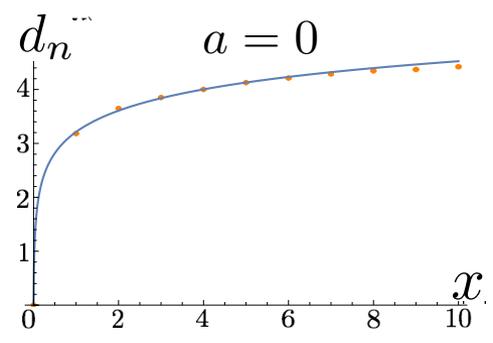
numerical verification

- We numerically calculate the distance $d_n((0, \beta_a), (x, \beta_a))$ ($a = 0, 1, 2$)
- We make the χ^2 -fit by using as the fitting function the geodesic distance calculated from the metric

$$ds^2 = l^2 \left(\alpha \beta^q \sum_{\mu=1}^D dx_\mu^2 + \frac{d\beta^2}{\beta^2} \right)$$



fit result



$$\begin{cases} l = 0.0404(14) \\ \alpha = 2.34(48) \times 10^5 \\ q = 0.289(12) \end{cases}$$

[orange pts : $d_n((0, \beta_a), (x, \beta_a))$ (numerical)
] blue lines : fitted geodesic distance
[w/ $\sqrt{\chi^2/(30 - 3)} = 2.7$]

4. Geometric optimization

[Fukuma, NM, Umeda (2018)]

Geometric optimization

- Our aim is to optimize the functional form of $\beta = \beta_a$, so that the transition in the extended direction becomes smooth.
- Since it is the parameter a that is directly dealt with by MCMC simulations, we expect that the smooth transition corresponds to the flat metric in the a direction: $(\text{const}) \times da^2$
- Since the geometry of $\overline{\mathcal{M}} \times \mathcal{A}$ is asymptotically AdS :

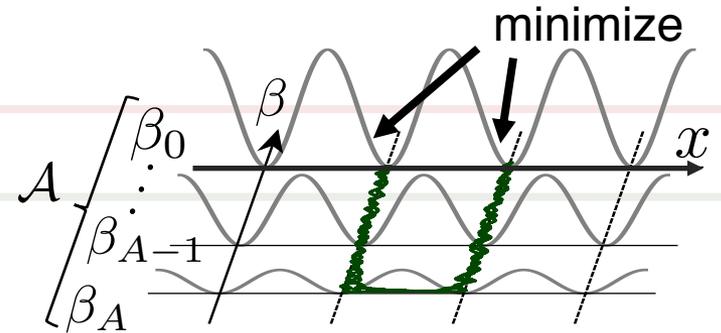
$$ds^2 = (\text{const.})\beta^q \sum_{\mu=1}^D dx_{\mu}^2 + (\text{const.})\frac{d\beta^2}{\beta^2} \quad [\beta \gg 1]$$

we require $\frac{d\beta^2}{\beta^2} \propto da^2$.

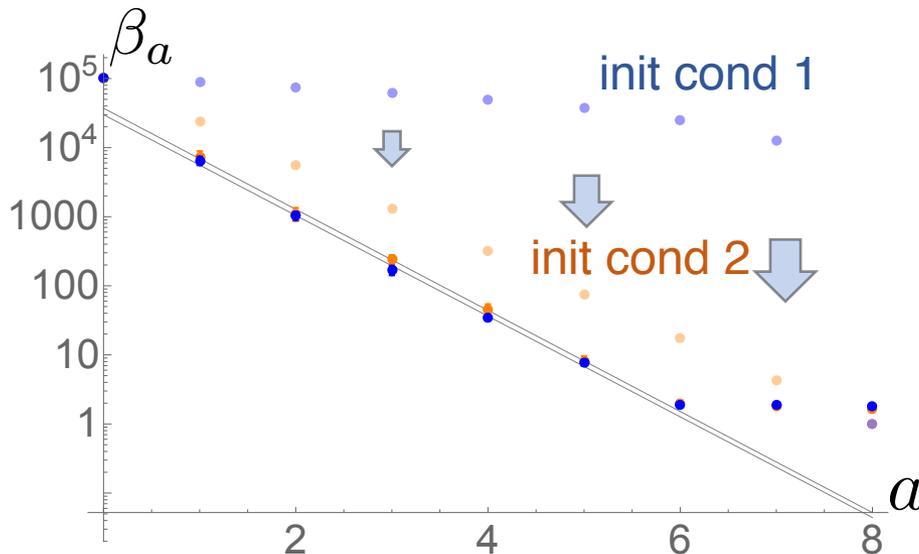
$\Rightarrow \beta_a = \beta_0 R^{-a} \left[\text{exponential} \right]$

Numerical confirmation

- We confirm the expectation by numerically checking the exponential form; we optimize the functional form of β_a by minimizing the distance between different modes.



result



the optimized form of β_a :
$$\beta_a = \beta_0 R^{-a} \text{ [exponential !]}$$

5. Comment: application to TLTM

[Fukuma, NM, Umeda (2019)]

Tempered Lefschetz thimble method

tempered Lefschetz thimble method (TLTM): an algorithm for the sign problem

[Fukuma, Umeda (2017), Fukuma, NM, Umeda (2019)]

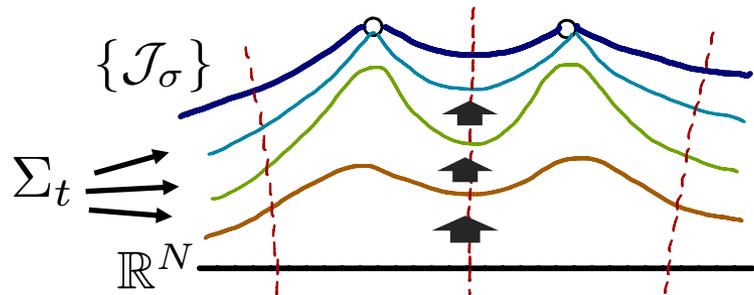
- To tame the oscillatory integrals, we deform the integration region by using the antiholomorphic gradient flow.

[Alexandru, Basar, Bedaque (2016)]

$$\text{def } \begin{array}{ccc} \mathbb{R}^N & \rightarrow & \mathbb{C}^N \\ x & \rightarrow & z_t(x) \end{array} \quad \left\{ \begin{array}{l} \dot{z}_t(x) \equiv \overline{\partial_z S(z_t(x))} \\ z_{t=0}(x) \equiv x \end{array} \right.$$

property

- as $t \rightarrow \infty$, $\Sigma_t \equiv z_t(\mathbb{R}^N) \rightarrow \{\mathcal{J}_\sigma\}$ $\left(\begin{array}{l} \text{a set of Lefschetz thimbles,} \\ \text{on each of which } \text{Im}S(z) = \text{const} \end{array} \right)$



- To take account of multi thimbles, we implement a tempering algorithm by choosing the flow time as the tempering parameter : $\mathcal{A} \equiv \{t_a\}_{a=0, \dots, A}$.

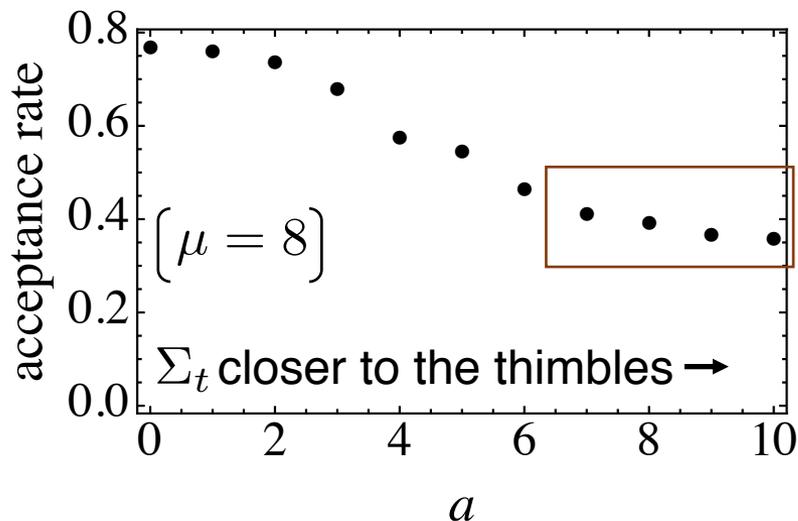
Optimized form for the flow time

the tempering parameter in TLTM

- At large t , $\text{Re}S(z_t(x))$ increases exponentially as $\beta_t \propto e^{(\text{const}) \times t}$
- As in the simulated tempering, we expect that the optimal form of β_{t_a} is an exponential function of a .
- ➔ t_a should be a linear function of a .

(see also discussion in [Alexandru, Basar, Bedaque, Warrington (2017)])

- In the application of TLTM to the Hubbard model, [Fukuma, NM, Umeda (2019)] we confirmed this choice works well :



- Here, t_a is taken to be a piecewise linear function of a . [the gradient is changed at $a = 5$.]
- This choice results in the acceptance rates being roughly above 0.4.
- In particular, the acceptance rate becomes constant for larger t , where Σ_t get close to the thimbles.

Summary

- We defined the **distance between configs**, which enumerates the difficulty of transition.
- We discussed that an **asymptotically AdS geometry** emerges in the **extended** and **coarse-grained** config space.
- We showed that the optimization of the tempering parameter can be easily done **geometrically**.
The result coincides with the one obtained by minimizing the distance.
- We further argued how to determine the optimized form of the **tempering parameter in TLTM**.

Future work

- Consider the distance in the Yang-Mills theory, and investigate the geometry of the config space by **identifying configs with the same topological charge as a single config**.
- Apply the distance to models whose DOF can be interpreted as spacetime coordinates (e.g. matrix models), and identify the geometry of the config space as the spacetime geometry. This would give a way to define **quantum gravity**.

Thank you