Distance between configurations in MCMC simulations

Masafumi Fukuma (Dept of Phys, Kyoto Univ)

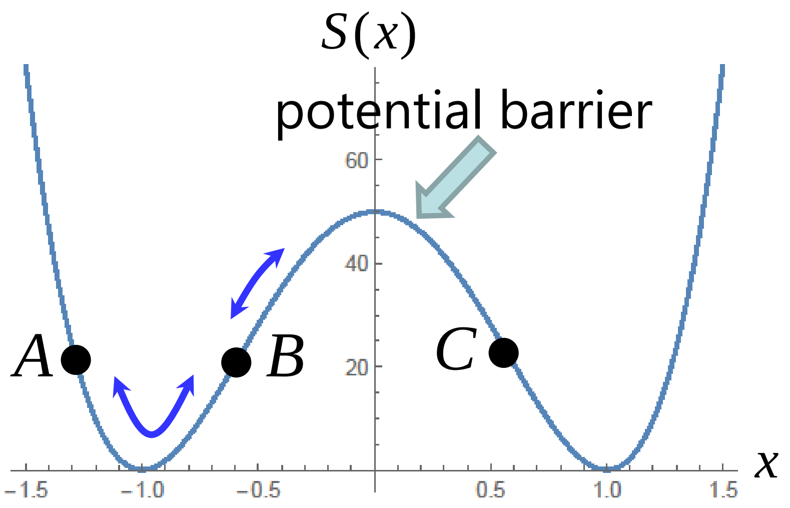
@APCTP, Sep 22, 2017
“Discrete Approaches to the Dynamics
of Fields and Space-Time”（離散研究会）

based on work with
N. Matsumoto and N. Umeda (Kyoto Univ)

[arXiv:1705.0609 (FMN1) + paper in preparation (FMN2)]

1. Introduction

Motivation



Consider the action

$$S(x)=\frac{\beta}{2}(x^2-1)^2 \quad (\beta \gg 1)$$

Separation of $A-B$ and that of $B-C$ are almost the same in x space.

However, in Markov chain Monte Carlo (MCMC) simulations,

A can be reached from B easily ⇐ “close” in MC

C cannot be reached from B easily ⇐ “far” in MC

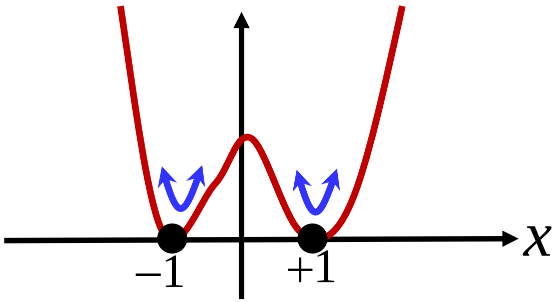
⇓

Can one enumerate this distance?

Main results

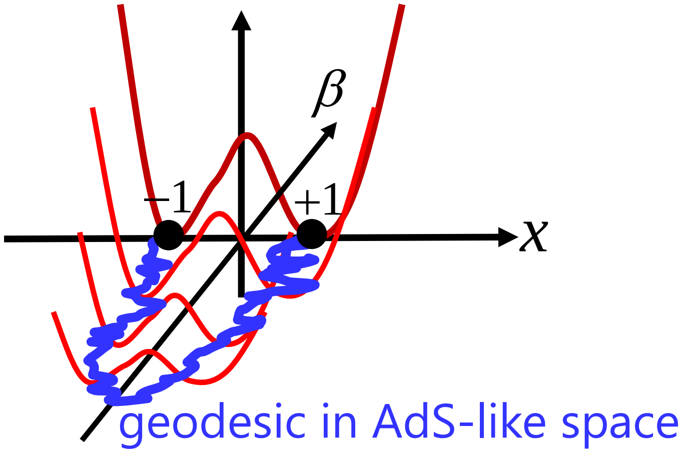
- We introduce a "distance between configurations" which satisfies desired properties as distance
- This definition is universal for MCMC algorithms that generate local moves in configuration space
- The distance gives an AdS-like geometry when a simulated tempering is implemented for multimodal distributions

original config space $\{x\}$



$$S(x)=\frac{\beta}{2}(x^2-1)^2 \quad (\beta \gg 1)$$

extended config space $\{(x,\beta)\}$



Plan

1. Introduction (done)
2. Definition of distance
   - preparation
   - definition of distance
   - universality of distance
3. Examples
   - unimodal case
   - multimodal case
4. Distance for simulated tempering
   - simulated tempering
   - emergence of AdS-like geometry
5. Conclusion and outlook

Plan

1. Introduction
2. Definition of distance
   - preparation
   - definition of distance
   - universality of distance
3. Examples
   - unimodal case
   - multimodal case
4. Distance for simulated tempering
   - simulated tempering
   - emergence of AdS-like geometry
5. Conclusion and outlook

Preparation 1: MCMC simulation (1/3)

$\mathcal{M} = \{x\}$: configuration space

$S(x)$: action

We want to estimate VEVs of operators $\mathcal{O}(x)$:

$$\langle \mathcal{O}(x) \rangle \equiv \frac{1}{Z} \int dx\, e^{-S(x)}\, \mathcal{O}(x) \quad \left(Z = \int dx\, e^{-S(x)} \right)$$

In MCMC simulations:

- Regard $p_{\text{eq}}(x) \equiv \frac{1}{Z} e^{-S(x)}$ as a PDF
- Introduce a Markov chain

$$p_{n-1}(x) \to p_n(x) = \int dy\, P(x \,|\, y)\, p_{n-1}(y) = \int dy\, P^n(x \,|\, y)\, p_0(y)$$

s.t. $p_n(x)$ converges uniquely to $p_{\text{eq}}(x)$ in the limit $n \to \infty$

$\left[\text{i.e., } P^n(x \,|\, y) \simeq p_{\text{eq}}(x) \quad (n \geq N_0) \right]$

Preparation 1: MCMC simulation (2/3)

- Starting from an initial value x_0, generate $x_1, x_2, \ldots$ following the transition matrix $P(x_i \mid x_{i-1})$

$$\underbrace{x_0 \xrightarrow{P} x_1 \xrightarrow{P} \cdots \xrightarrow{P} x_{N_0}}_{\text{relaxation}} \xrightarrow{P} \underbrace{\underset{\equiv y_1}{x_{N_0+1}} \xrightarrow{P} \underset{\equiv y_2}{x_{N_0+2}} \xrightarrow{P} \cdots \xrightarrow{P} \underset{\equiv y_N}{x_{N_0+N}} \xrightarrow{P} \cdots}_{\text{generated with } \sim p_{\text{eq}}(x)}$$

- After the system is well relaxed, take a sample $\{y_i\}_{i=1,\ldots,N}$

- Estimate VEVs of operators $\mathcal{O}(x)$ as

$$\langle \mathcal{O}(x) \rangle \simeq \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(y_i)$$

We first would like to establish a mathematical framework which enables the systematic understanding of relaxation

Preparation 1: MCMC simulation (3/3)

We assume that

(1) $P(x\,|\,y)$ satisfies the detailed balance condition:

$$P(x\,|\,y)\,p_{\mathrm{eq}}(y) = P(y\,|\,x)\,p_{\mathrm{eq}}(x)\left(\Leftrightarrow P(x\,|\,y)e^{-S(y)} = P(y\,|\,x)e^{-S(x)}\right)$$

(2) all of the eigenvalues of P are positive

NB : (1) can be written as

$$\hat{P}e^{-S(\hat{x})} = e^{-S(\hat{x})}\,\hat{P}^T \qquad \left(\begin{array}{l} P(x\,|\,y) = \langle x\,|\,\hat{P}\,|\,y\rangle \\ \hat{x} \equiv \int dx\, x\,|\,x\rangle\langle x\,| \end{array}\right)$$

NB : (2) is not too restrictive

In fact, if P has negative eigenvalues,

then we instead can use P^2 as the elementary transition matrix,

for which

- all the eigenvalues are positive
- the same detailed balance condition is satisfied as P :

$$P^2(x\,|\,y)e^{-S(y)} = P^2(y\,|\,x)e^{-S(x)}$$

Preparation 2: Transfer matrix (1/2)

[MF-Matsumoto-Umeda1]

We introduce the "transfer matrix" :

$$\boxed{\hat{T} \equiv e^{S(\hat{x})/2}\,\hat{P}\,e^{-S(\hat{x})/2}} \quad \left(\Leftrightarrow T(x\,|\,y) = e^{S(x)/2}\,P(x\,|\,y)\,e^{-S(y)/2} \right)$$

properties:

(1) $\hat{T} = \hat{T}^T \quad \left(\Leftrightarrow \hat{P}e^{-S(\hat{x})} = e^{-S(\hat{x})}\hat{P}^T \right)$

(2) same eigenvalue set as $\hat{P}$ (thus all positive)

We order the EVs as

$$\lambda_0 = 1 > \lambda_1 \geq \lambda_2 \geq \cdots > 0$$

spectral decomposition:

$$\hat{T} = \sum_{k\geq 0} \lambda_k\,|k\rangle\langle k| = |0\rangle\langle 0| + \sum_{k\geq 1} \lambda_k\,|k\rangle\langle k|$$

where

$$\langle x\,|\,0\rangle = \frac{1}{\sqrt{Z}}e^{-S(x)/2}$$

Preparation 2: Transfer matrix (2/2)

Note that $\hat{P}^n \Leftrightarrow \hat{T}^n = |0\rangle\langle 0| + \sum_{k\geq 1} \lambda_k^n |k\rangle\langle k| \quad (\lambda_0 = 1 > \lambda_1 \geq \lambda_2 \geq \cdots > 0)$

⇓

relaxation to equilibrium

$\Leftrightarrow$ relaxation of $\hat{T}^n$ to $|0\rangle\langle 0|$ in the limit $n \to \infty$

$\Leftrightarrow$ decoupling of modes $|k\rangle$ with $k \geq 1$

NB:

decoupling occurs earlier for higher modes (i.e. for larger k)

NB:

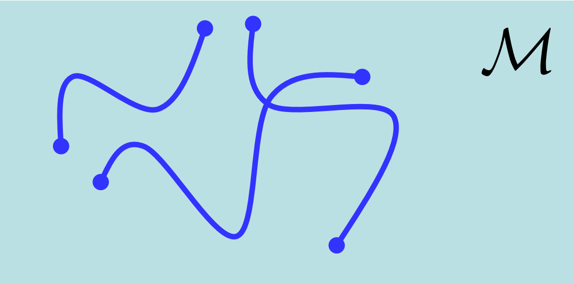
relaxation time τ can be estimated from $\lambda_1 \sim e^{-1/\tau}$

slow relaxation $\Leftrightarrow$ $\lambda_1 \sim 1$

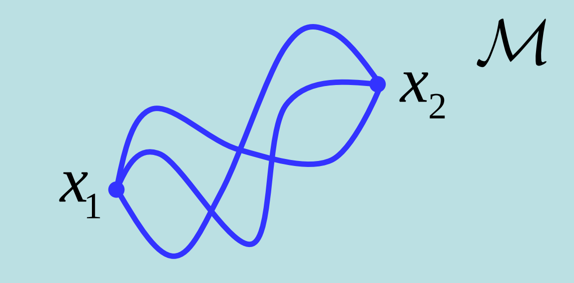
Preparation 3: Connectivity between configs (1/3)

[MF-Matsumoto-Umeda1]

$$\mathbf{X}_n \equiv (\text{set of sequences of } n \text{ processes in } \mathcal{M})$$



$$\mathbf{X}_n(x_1, x_2) \equiv \begin{pmatrix} \text{set of sequences of } n \text{ processes in } \mathcal{M} \\ \text{that start from } x_2 \text{ and end at } x_1 \end{pmatrix}$$



We define the connectivity between two configs as

$$\boxed{f_n(x_1, x_2) \equiv \frac{|\mathbf{X}_n(x_1, x_2)|}{|\mathbf{X}_n|}}$$

$$= (\text{prob to obtain } x_1 \text{ from } x_2) \times (\text{prob to have } x_2)$$

$$= P^n(x_1 \,|\, x_2) \frac{1}{Z} e^{-S(x_2)} \left(\overset{\text{det balance}}{=} P^n(x_2 \,|\, x_1) \frac{1}{Z} e^{-S(x_1)} = f_n(x_2, x_1) \right)$$

Preparation 3: Connectivity between configs (2/3)

normalized connectivity ("half-time overlap"):

$$\left(\begin{aligned} & f_n(x_1,x_2) \\ & = P^n(x_1\,|\,x_2)\frac{1}{Z}e^{-S(x_2)} \\ & = P^n(x_2\,|\,x_1)\frac{1}{Z}e^{-S(x_1)} \end{aligned} \right)$$

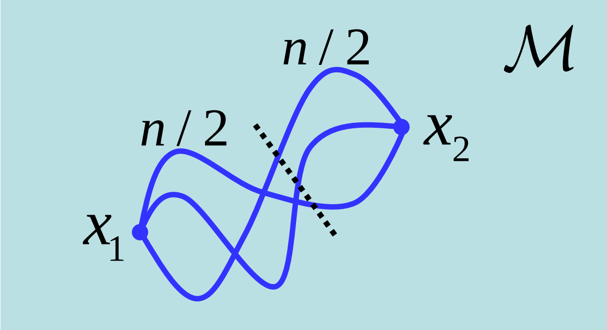
$$\boxed{\begin{aligned} F_n(x_1,x_2) &\equiv \frac{f_n(x_1,x_2)}{\sqrt{f_n(x_1,x_1)\,f_n(x_2,x_2)}} \\ &= \sqrt{\frac{P^n(x_1\,|\,x_2)P^n(x_2\,|\,x_1)}{P^n(x_1\,|\,x_1)P^n(x_2\,|\,x_2)}} = \frac{K_n(x_1,x_2)}{\sqrt{K_n(x_1,x_1)K_n(x_2,x_2)}} \end{aligned}}$$

$$\left(K_n(x_1,x_2) \equiv \langle x_1\,|\,\hat{T}^n\,|\,x_2\rangle \right)$$

$F_n(x_1,x_2)$ is actually the overlap between two normalized "half-time" elapsed states:

$$F_n(x_1,x_2) \equiv \frac{\langle x_1, n/2 | x_2, n/2\rangle}{\| |x_1,n/2\rangle \| \, \| |x_2,n/2\rangle \|}$$

$$\left(|x,n/2\rangle \equiv \hat{T}^{n/2}\,|\,x\rangle \right)$$



Preparation 3: Connectivity between configs (3/3)

properties of $F_n(x_1, x_2)$

$$\begin{cases} (1) \quad F_n(x_1,x_2) = F_n(x_2,x_1) \\ (2) \quad 0 \le F_n(x_1,x_2) \le 1 \\ (3) \quad F_n(x_1,x_2) = 1 \Leftrightarrow x_1 = x_2 \quad \left(\text{when } n \text{ is finite}\right) \\ (4) \quad \lim_{n\to\infty} F_n(x_1,x_2) = 1 \quad \left(\forall x_1, x_2\right) \end{cases}$$

+

(A) If x_1 can be easily reached from x_2 in n steps,
then $F_n(x_1,x_2) \simeq 1$

(B) If x_1 and x_2 are separated by high potential barriers,
then $F_n(x_1,x_2) \ll 1$

proof of (4):

In the limit $n \to \infty$, $\hat{T}^n \to |0\rangle\langle 0|$, and thus,

$$K_n(x_1,x_2) = \langle x_1 | \hat{T}^n | x_2 \rangle \to \langle x_1 | 0 \rangle \langle 0 | x_2 \rangle = \sqrt{K_n(x_1,x_1) K_n(x_2,x_2)}.$$

Definition of distance

[MF-Matsumoto-Umeda1]

$$\boxed{\theta_n(x_1,x_2) \equiv \arccos(F_n(x_1,x_2))}$$

properties of $\theta_n(x_1,x_2)$

(1) $\theta_n(x_1,x_2) = \theta_n(x_2,x_1)$

(2) $\theta_n(x_1,x_2) \geq 0$

(3) $\theta_n(x_1,x_2) = 0 \Leftrightarrow x_1 = x_2$ (when n is finite)

(4) $\lim_{n\to\infty} \theta_n(x_1,x_2) = 0 \ (\forall x_1, x_2)$

(5) $\theta_n(x_1,x_2) + \theta_n(x_2,x_3) \geq \theta_n(x_1,x_3)$

✚

(A) If x_1 can be easily reached from x_2 in n steps,
then $\theta_n(x_1,x_2)$: small

(B) If x_1 and x_2 are separated by high potential barriers,
then $\theta_n(x_1,x_2)$: large

Alternative definition of distance

Instead of $\theta_n(x_1,x_2) = \arccos(F_n(x_1,x_2))$,
one can also use the following as distance:

$$\boxed{d_n^2(x_1,x_2) \equiv -2\ln F_n(x_1,x_2)}$$

we will mainly use this

or $$\boxed{D_n^2(x_1,x_2) \equiv 2[1-\ln F_n(x_1,x_2)]}$$

$$\left(\begin{array}{l} F_n(x_1,x_2) \\ = \cos\theta_n(x_1,x_2) = e^{-(1/2)d_n^2(x_1,x_2)} = 1-\dfrac{1}{2}D_n^2(x_1,x_2) \\ \text{They agree when } \theta_n \approx 0 \end{array}\right)$$

NB: analogy in quantum information

$$\begin{cases} \theta_n(x_1,x_2) : \text{Bures length} \\ D_n(x_1,x_2) : \text{Bures distance} \end{cases}$$ for two pure states $\rho_{1,2} = \dfrac{|\,x_{1,2},n/2\rangle\langle x_{1,2},n/2\,|}{\||\,x_{1,2},n/2\rangle\|\,\||\,x_{1,2},n/2\rangle\|}$

Universality of distance (1/4)

[MF-Matsumoto-Umeda1]

The above distance is expected to be universal
for MCMC algorithms that generate local moves in config space.

("universal" in the sense that differences of distance
between two such local MCMC algorithms
can alway be absorbed into a rescaling of n)

In fact,

universality of $d_n^2(x_1,x_2) \Leftrightarrow$ univ. of $K_n(x_1,x_2)=\langle x_1|\hat{T}^n|x_2\rangle$
$\Leftrightarrow$ univ. of $\hat{T}\equiv e^{-\epsilon\hat{H}}$

and,

If algorithms are sufficiently local,
then $\hat{H}$ are expected to be local operators
acting on functions over $\mathcal{M}$ in almost the same way.

⇩

The wave functions $\langle x|k\rangle$ must be almost the same for small k

Universality of distance (2/4)

This expectation can be explicitly checked using a simple model.

algorithm 1: Langevin

$$x_{n+1} = x_n + \sqrt{\epsilon}\,\nu_n - \epsilon\, S'(x_n) \quad \text{with } \langle \nu_n \nu_m \rangle_\nu = 2\delta_{n,m}$$

$$\Rightarrow \quad \langle x | \hat{T} | y \rangle = \langle x | e^{-\epsilon \hat{H}} | y \rangle \simeq \frac{1}{\sqrt{4\pi\epsilon}}\, e^{-\frac{1}{4\epsilon}(x-y)^2 - \epsilon V\left(\frac{x+y}{2}\right)}$$

$$\text{with } V(x) = (1/4)\big(S'(x)\big)^2 - (1/2)S''(x)$$

algorithm 2: Metropolis (with Gaussian proposal of variance σ^2)

$$\begin{aligned}
\langle x | \hat{T} | y \rangle &= \langle x | \hat{P} | y \rangle \times e^{S(x)/2 - S(y)/2} \\
&= \min\left(1, e^{-S(x)+S(y)}\right) \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-y)^2} \times e^{S(x)/2 - S(y)/2} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-y)^2 - \frac{1}{2}\left|S(x)-S(y)\right|}
\end{aligned}$$

Universality of distance (3/4)

With the identification $\sigma^2 \sim \epsilon$,

both Hamiltonians $\hat{H} \left(\equiv -\frac{1}{\epsilon} \ln \hat{T} \right)$ become local in the limit $\epsilon \to 0$,

and have the same tendency to enhance transitions
when $|x-y|$ and $|S(x)-S(y)|$ are small.

⇓

The low energy structure of $\hat{H}$ should be almost the same.

⇓

The global structure of distance should be almost the same.

(The argument for universality are more trustworthy
as the DOF of the system become larger.)

In fact,
the universality actually holds more than expected
even for a single DOF

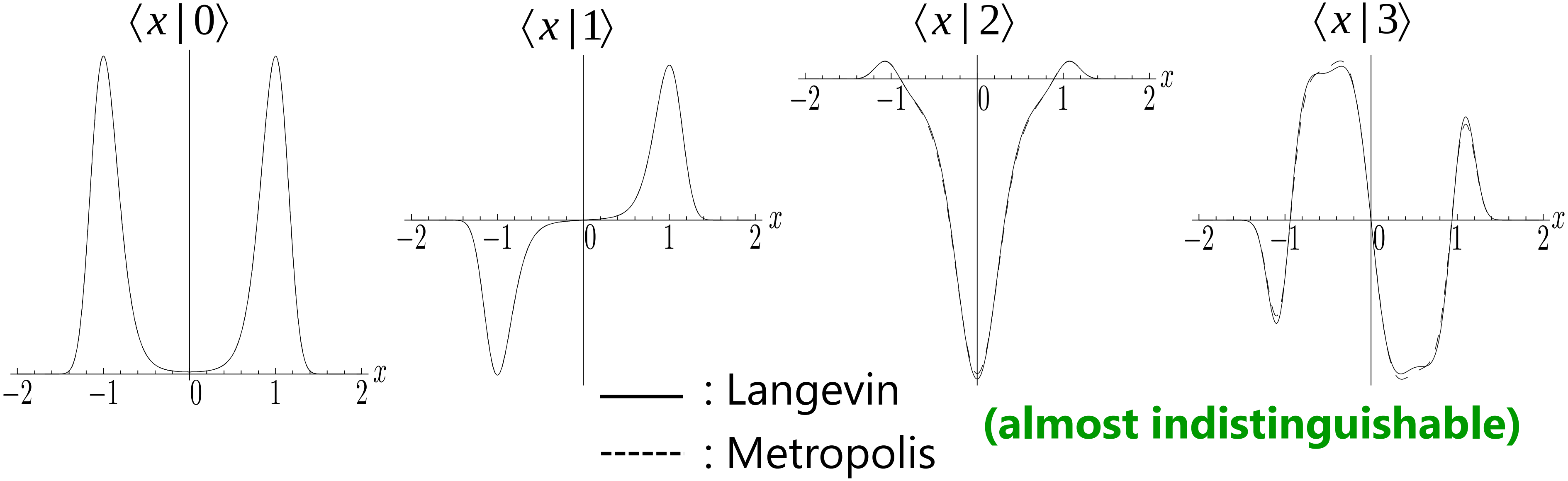
Universality of distance (4/4)

eigenvalues :

$$S(x)=\frac{\beta}{2}(x^2-1)^2 \quad (\beta=20)$$

	E_k (Lang)	E_k/E_1 (Lang)	E_k (Met)	E_k/E_1 (Met)
0	0	0	0	0
1	7.81 x 10^(-4)	1	7.62 x 10^(-4)	1
2	36.2	4.63 x 10^4	34.2	4.49 x 10^4
3	58.2	7.45 x 10^4	54.7	7.17 x 10^4

eigenfunctions :



(almost indistinguishable)

Plan

1. Introduction
2. Definition of distance
   - preparation
   - definition of distance
   - universality of distance
3. Examples
   - unimodal case
   - multimodal case
4. Distance for simulated tempering
   - simulated tempering
   - emergence of AdS-like geometry
5. Conclusion and outlook

Transfer matrix for Langevin

Langevin equation (continuum)

$$\dot{x}_t = \nu_t - S'(x_t) \text{ with} \begin{cases} x_{t=0} = x_0 \\ \langle \nu_t \nu_{t'} \rangle_\nu = 2\delta(t-t') \end{cases}$$

⇒ $$x_t = x_t(x_0, [\nu])$$

⇒ $$P_t(x\,|\,x_0) \equiv \Big\langle \delta\big(x - x_t(x_0,[\nu])\big) \Big\rangle_\nu = \langle x\,|\,e^{-t\hat{H}_{\text{FP}}}\,|\,x_0 \rangle$$

with $\hat{H}_{\text{FP}} = -2\partial_x[\partial_x + S'(x)]$

⇒ $$K_t(x,y) = e^{S(x)/2} P_t(x\,|\,y) e^{-S(y)/2} = \langle x\,|\,e^{-\epsilon\hat{H}}\,|\,y \rangle$$

with $\hat{H} = e^{S(\hat{x})/2}\hat{H}_{\text{FP}} e^{-S(\hat{x})/2} = -\partial_x^2 + V(\hat{x})$

$$\Big[V(x) = (1/4)\big(S'(x)\big)^2 - (1/2)S''(x) \Big]$$

⇒ $$F_t(x_1,x_2) = \frac{K_t(x_1,x_2)}{\sqrt{K_t(x_1,x_1)K_t(x_2,x_2)}} = e^{-\frac{1}{2}d_t^2(x_1,x_2)}$$

Example 1: Unimodal distribution (Gaussian)

$$S(x) = \frac{\omega}{2} x^2$$

subtracts zero-point energy

$$\Rightarrow \quad \hat{H} = -\partial_x^2 + V(\hat{x}) \quad \text{with } V(x) = \frac{\omega^2}{4} x^2 - \frac{\omega}{2}$$

$$\Rightarrow \quad K_t(x,y) = \langle x | e^{-t\hat{H}} \, | \, y \rangle$$

$$= \sqrt{\frac{\omega}{2\pi(1-e^{-2\omega t})}} \exp\left[-\frac{\omega}{4\sinh \omega t} [(x_1^2 + x_2^2)\cosh \omega t - 2x_1 x_2] \right]$$

$$\Rightarrow \quad d_t^2(x_1, x_2) = \frac{\omega}{2\sinh \omega t} | x_1 - x_2 |^2 \sim e^{-\omega t} | x_1 - x_2 |^2$$

We see that:

- geometry is flat and translationally invariant
- relaxation time τ is given by $\tau \sim 1/\omega \ \left[\omega^2 \sim V''(x) \right]$

Example 2: Unimodal dist. (non-Gaussian)

$$S(x) = \frac{\omega}{2} x^2 + \frac{\lambda}{4} x^4$$

perturbative expansion in λ :

$$\begin{aligned} d_t^2(x_1, x_2) = |\, x_1 - x_2 \,|^2 \Big\{ \frac{\omega}{2s} - \frac{\lambda}{8\omega s^4} [& 12(s^3 - 3s^2 c + 3\omega t + 2\omega t s^3 - \omega t s^2 c) \\ & + \omega(s^3 + 3s - 3\omega t c)(x_1 - x_2)^2 \\ & + 3\omega(s^3 + 3s - 3\omega t c + 3\omega t - 3sc + 2\omega t s^2)(x_1 + x_2)^2] + O(\lambda^2) \Big\} \end{aligned}$$

$$\left(c \equiv \cosh \omega t,\ s \equiv \sinh \omega t \right)$$

We see that:

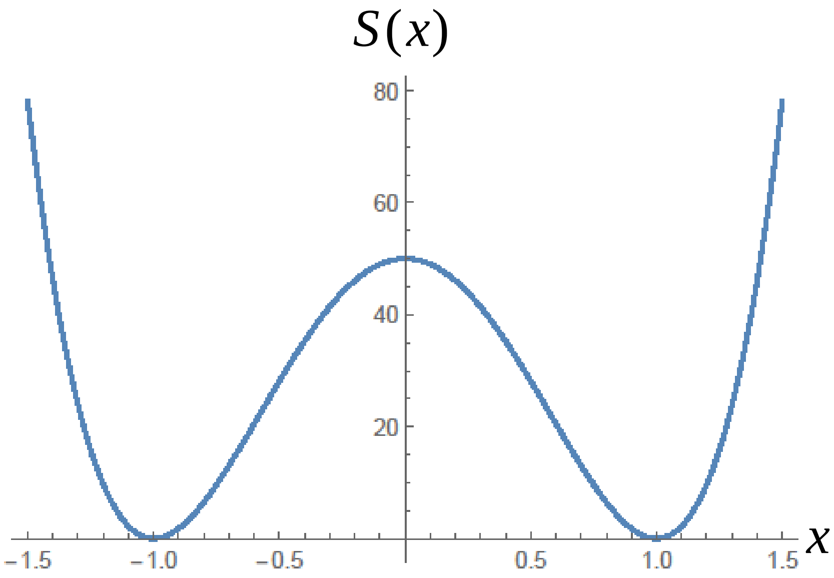
- geometry is no longer flat or translationally invariant
- relaxation time τ is again given by $\tau \sim 1/\omega \ \left[\omega^2 \sim V''(x) \right]$

Example 3: Multimodal dist. (double well) (1/2)

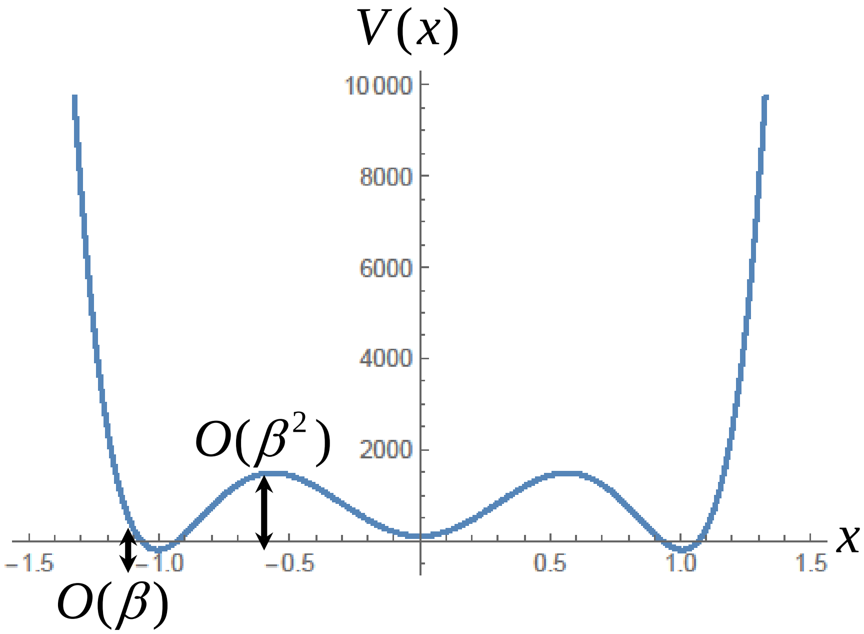
$$S(x) = \frac{\beta}{2}(x^2 - 1)^2 \qquad (\beta \gg 1)$$

⇒ $\hat{H} = -\partial_x^2 + V(\hat{x})$

with $$V(x) = \beta^2 x^6 - 2\beta^2 x^4 + (\beta^2 - 3\beta)x^2 + \beta$$
$$= \beta^2 x^2 (x^2 - 1)^2 + O(\beta)$$



⇒



Example 3: Multimodal dist. (double well) (2/2)

For $\beta = 20$:

$E_0 = 0$

$E_1 = 7.81 \times 10^{-4}$

instanton $e^{-O(\beta)}$

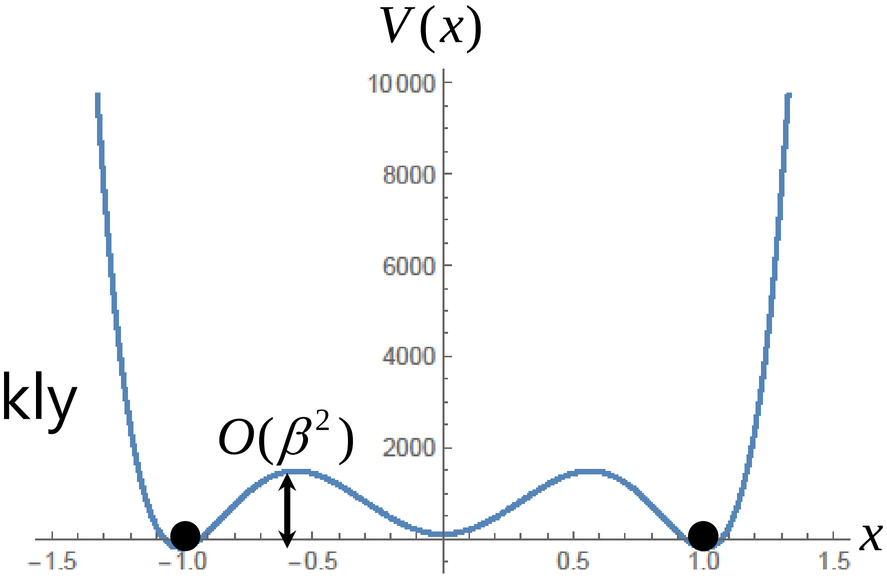
$E_2 = 36.2$

$E_3 = 58.2$

$\langle x | k \rangle \ \left(k = 2, 3, \ldots\right)$: decouple quickly

$\langle x | 1 \rangle$: decouples very slowly

$$\hat{H} = -\partial_x^2 + V(\hat{x})$$



In fact,

n	$d_n^2(-1,+1)$
10	39.1
50	19.2
100	16.9
500	13.2
1,000	11.7
5,000	8.46

decreases only very slowly

Plan

1. Introduction
2. Definition of distance
   - preparation
   - definition of distance
   - universality of distance
3. Examples
   - unimodal case
   - multimodal case
4. Distance for simulated tempering
   - simulated tempering
   - emergence of AdS-like geometry
5. Conclusion and outlook

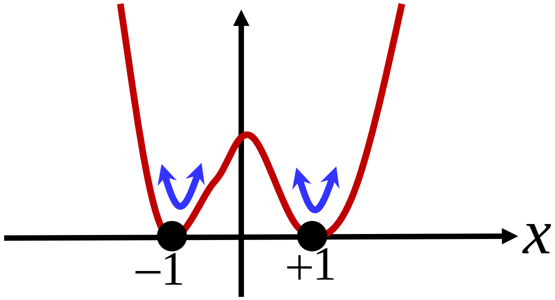
Simulated tempering (1/3)

Basic idea of tempering : **[Marinari-Parisi]**

Even when the original action $S(x;\beta_0)$ is multimodal, it often happens that $S(x;\beta)$ becomes less multimodal if we take smaller β.

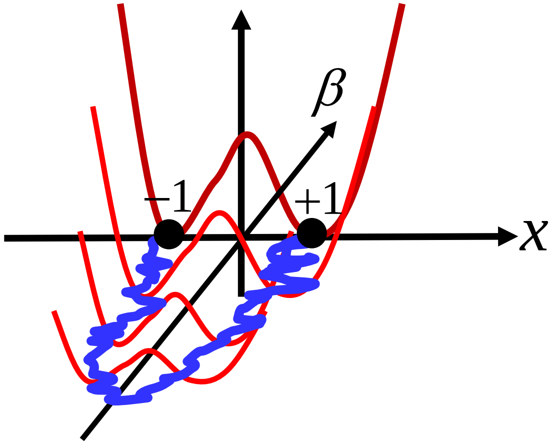
We extend the configuration space s.t. configurations in different modes can be reached from each other by passing through small β's.

original config space $\{x\}$

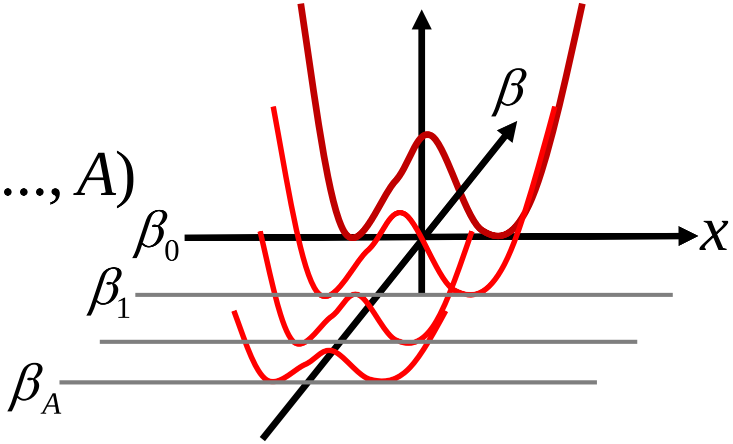


$$S(x;\beta_0)=\frac{\beta_0}{2}(x^2-1)^2 \quad (\beta_0 \gg 1)$$

extended config space $\{(x,\beta)\}$



Simulated tempering (2/3)



Realization

- Extend the config space $\mathcal{M}=\{x\}$
  to $\mathcal{M}\times\mathcal{A}=\{X=(x,\beta_a)\}(x\in\mathcal{M};\ a=0,1,\ldots,A)$

- Introduce a stochastic process

$$P_n(X)\to P_{n+1}(X)$$

s.t. $P_n(X)\xrightarrow{n\to\infty}P_{\text{eq}}(X)=P_{\text{eq}}(x,\beta_a)=w_a e^{-S(x;\beta_a)}$

- Estimate the VEV by only using the subsample with $\beta_{a=0}$

NB : (appearance probability of a-th subsample)

$$=\int dx\, P_{\text{eq}}(x,\beta_a)=w_a Z_a\quad \left(Z_a=\int dx\, e^{-S(x;\beta_a)}\right)$$

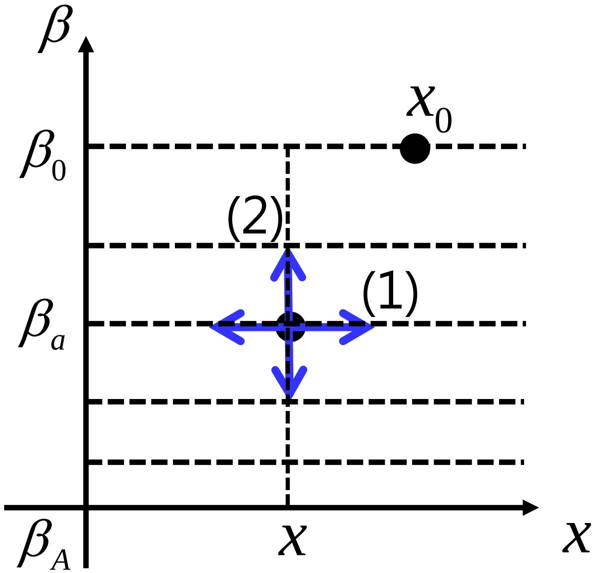
⇩

w_a is often set as $w_a\propto 1/Z_a$,
which ensures that the desired 0-th configs appear
with nonvanishing probability $(=1/(A+1))$

consideration not necessary for parallel tempering
⇨ Umeda's talk

Simulated tempering (3/3)

Algorithm



(1) Generate a transition in the x direction,

$$X=(x,\beta_a)\rightarrow X'=(x',\beta_a)$$

with some proper algorithm

(such as Langevin or Metropolis)

(2) Generate a transition in the β direction,

$$X=(x,\beta_a)\rightarrow X'=(x,\beta_{a'=a\pm1})$$

with the probability $\min\left(1,\dfrac{w_{a'}\,e^{-S(x,\beta_{a'})}}{w_a e^{-S(x,\beta_a)}}\right)$

(3) Extract a subsample with $\beta_{a=0}$, $\{(x_i,\beta_0)\}\ \ (i=1,\ldots,N)$

(4) Evaluate VEVs as $\langle\mathcal{O}(x)\rangle_{\beta_0}\simeq\dfrac{1}{N}\sum_{i=1}^{N}\mathcal{O}(x_i)$

NB : a-dependence of β_a should be chosen s.t. the transition in the β-direction is easy. This adjustment is usually done manually or adaptively.

Distance for simulated tempering

[MF-Matsumoto-Umeda1]

The introduction of tempering should be seen as the reduction of distance.

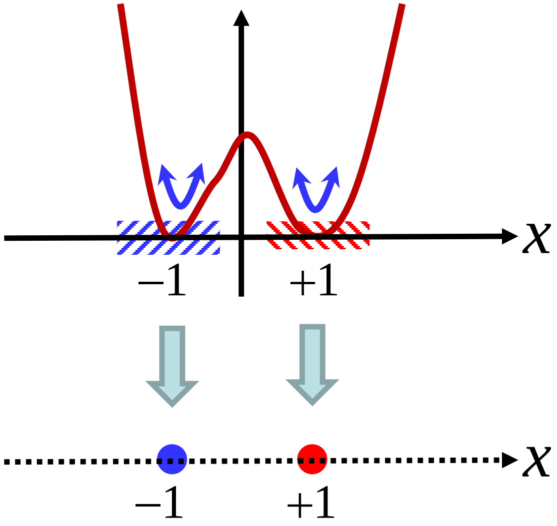
In fact,

	w/o tempering	w/ tempering
n	$d_n^2(-1,+1)$	$d_n^2(-1,+1)$
10	39.1	26.5
50	19.2	7.16
100	16.9	4.35
500	13.2	0.708
1,000	11.7	0.106
5,000	8.46	2.78 x 10^(-8)

rapid decreasing

Emergence of AdS-like geometry (1/3)

[MF-Matsumoto-Umeda1,2]



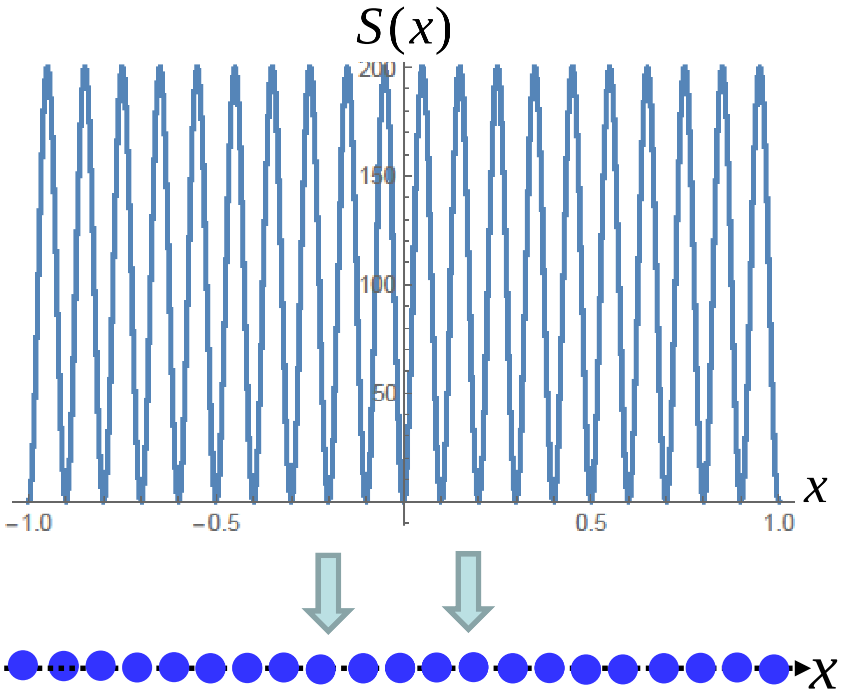
In MCMC simulations,
the most expensive part is the transitions between configs in different modes,
and thus, configs in the same mode can be effectively treated as a point.

This leads us to the idea of "coarse-grained config space" $\overline{\mathcal{M}}$

We would like to show that

when the original config space is multimodal
with high degeneracy,
the extended coarse-grainined config space $\overline{\mathcal{M}}\times\mathcal{A}$
naturally has an AdS-like geometry

Emergence of AdS-like geometry (2/3)



action: $S(x;\beta_0)=\beta_0\left[1-\cos\left(\dfrac{2\pi x}{\epsilon}\right)\right]$

original config space: $\mathcal{M}=\mathbb{R}$

coarse-grained config space:

$\overline{\mathcal{M}}=$(1D lattice with cutoff ϵ)

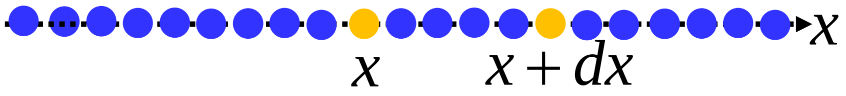
+ sim temp

extended coarse-grained config space:

$\overline{\mathcal{M}}\times\mathcal{A}=\{X=(x,\beta_a)\}$ $[x\in$(1D lattice with cutoff ϵ)$]$

Emergence of AdS-like geometry (3/3)

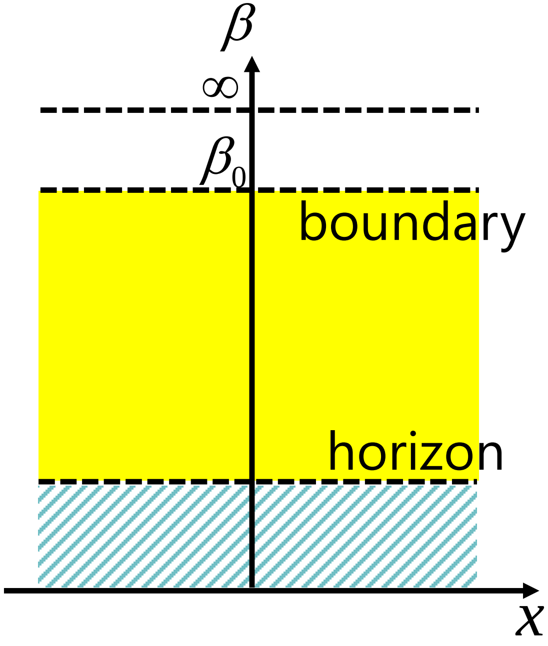
We find:



$$d_n^2\left((x,\beta),(x+dx,\beta)\right)=\text{const.}\,\beta\,dx^2$$

If we set

$$d_n^2\left((x,\beta),(x,\beta+d\beta)\right)=f(\beta)\,d\beta^2 \quad \text{---- (\#)}$$



then we have

$$ds^2\equiv d_n^2\left((x,\beta),(x+dx,\beta+d\beta)\right)$$
$$=\text{const.}\,\beta\,dx^2+f(\beta)\,d\beta^2$$

If (#) is scale invariant (i.e., $f(\beta)\propto 1/\beta^2$), this gives an AdS metric:

$$ds^2=\text{const.}\,\beta\,dx^2+\text{const.}\frac{d\beta^2}{\beta^2}=\frac{\ell^2}{z^2}\left(dx^2+dz^2\right)\qquad\left(\beta\propto 1/z^2\right)$$

(This is actually an AdS BH)

AdS geometry as a result of optimization (1/2)

[MF-Matsumoto-Umeda2]

If $\beta_a \ (a=0,1,\ldots,A)$ is chosen as

$$\beta_a = \beta_0 \left(\frac{\beta_A}{\beta_0} \right)^{a/A}, \qquad \text{--------------------} \ (\#\#)$$

one can show that geometry in β direction becomes scale invariant, so that we will obtain an AdS geometry, as we saw in the previous slide.

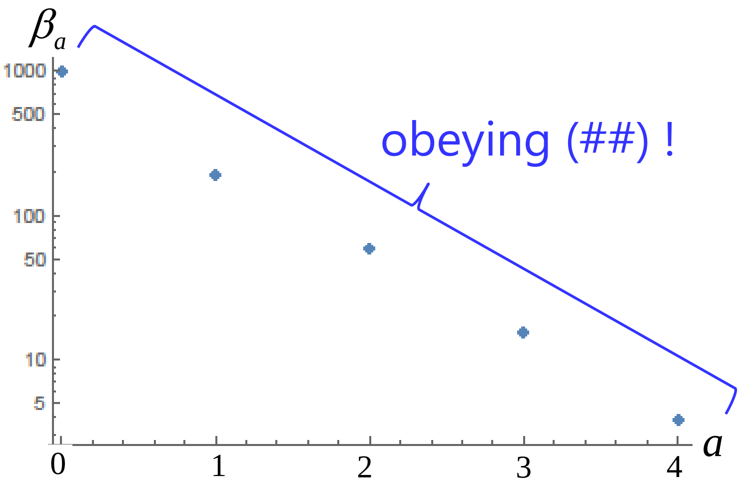
One can actually confirm that (##) is the best choice for minimizing the distance in simulated tempering:

Consider the action :

$$S(x;\beta_0) = \frac{\beta_0}{2} \sum_{\mu=1}^{2} (x_\mu^2 - 1)^2$$

$$\left(\beta_0 = 10^3 \right)$$

Search for $\{\beta_1, \beta_2, \beta_3, \beta_4\}$ that minimize $d_n^2((-1,\beta_0),(+1,\beta_0))$



obeying (##) !

AdS geometry as a result of optimization (2/2)

That is,

optimize β_a s.t. the distance is minimized

$$\beta_a = \beta_0 \left(\frac{\beta_A}{\beta_0} \right)^{a/A} \qquad (a = 0, 1, \ldots, A)$$

AdS metric :

$$ds^2 = \text{const.}\, \beta\, dx^2 + \text{const.} \frac{d\beta^2}{\beta^2} = \frac{\ell^2}{z^2}\left(dx^2 + dz^2 \right) \quad \left(\beta \propto 1/z^2 \right)$$

This is the first example of the “emergence of AdS geometry” in nonequilibrium systems.

Plan

1. Introduction
2. Definition of distance
   - preparation
   - definition of distance
   - universality of distance
3. Examples
   - unimodal case
   - multimodal case
4. Distance for simulated tempering
   - simulated tempering
   - emergence of AdS-like geometry
5. Conclusion and outlook

Conclusion and outlook

What we have done:

- We introduced the concept of "distance between configs" in MCMC simulations
- The distance satisfies desired properties as distance
- This may be used for the optimization of parameters

Future work:

- Establish a systematic method for optimization (such as $I[\beta_a]$)
- Investigate whether such distance can also be introduced to systems with complex actions
- Extend the framework to general nonequilibrium systems, and compare the obtained dynamics with GR.

"What if our world is in the process of relaxation of some unknown dynamics, and if we recognize distance as the extent of difficulty of communication?"

Thank you.