Searching the optimal matrix configurations by Replica-Exchange Monte Carlo methods for matrix models

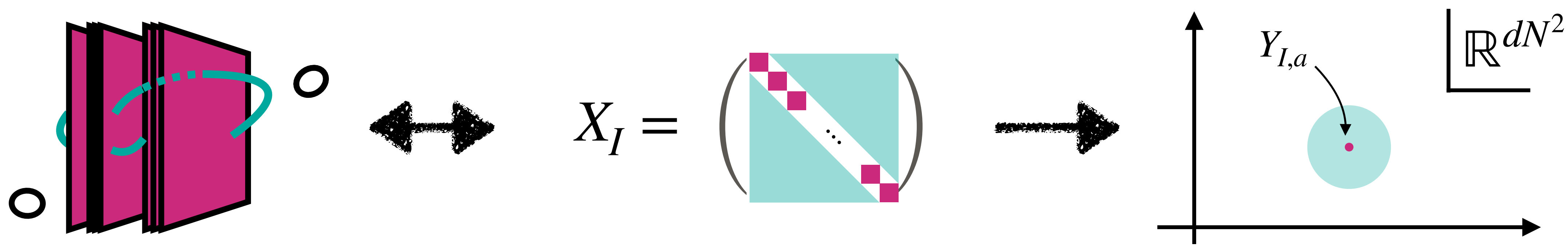
Hiromasa Watanabe（渡辺 展正)
YITP, Kyoto U.

Based on collaboration with
M. Hanada (Queens Mary, London), S. Kanno (Tsukuba), S. Matsuura (Keio)
in progress

2023/09/13 @ 離散的手法による場と時空のダイナミクス 2023, U. Tsukuba

Short summary

To read off geometric information from field theory configs. in string theory, the notion of wave packet in space of matrix plays an essential role.



To determine the wave packet in a high-dim space, we compute a quantity

$$R_\infty(X) := \min_U \left(\max_a \left| \left(U^\dagger X U - Y \right)_a \right| \right)$$

X, Y : $N \times N$ hermitian mat.

U : unitary mat.

which can be translated into an optimization problem.

We employ the Replica-Exchange Monte Carlo methods (REMC) and consider their extensions to solve this problem numerically.

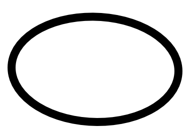
Contents

- D-brane geometry from matrix model
    - Center of wave packet in the space of color dof. and a proposal determining it from matrix configs.
- Monte Carlo method to the minimization problem
    - Basic strategy by Replica-Exchange method and its extensions
- Numerical results
- Summary

Strings and D-branes

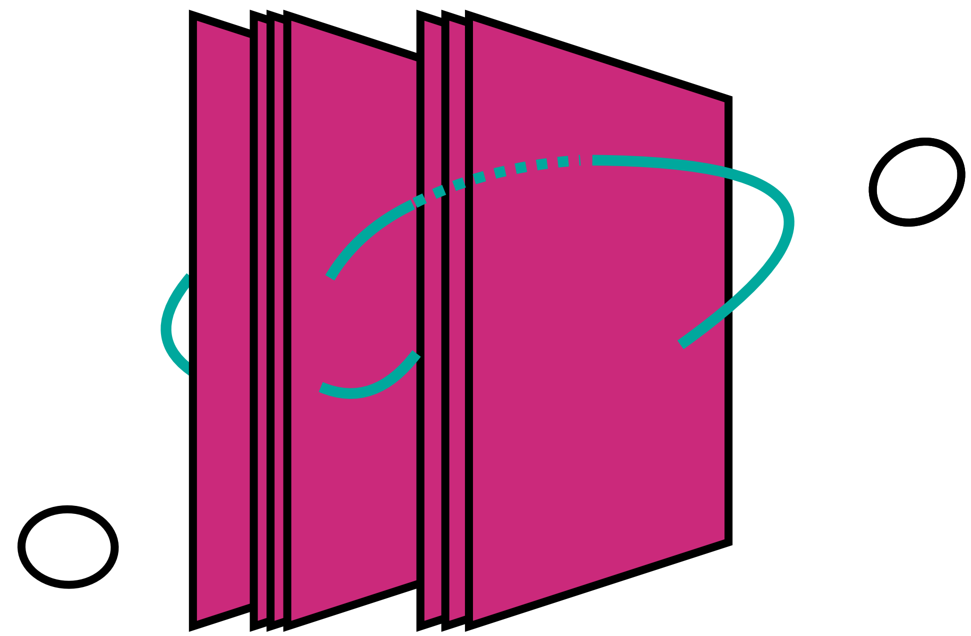
String theory : a candidate for theory of quantum gravity

: open string



: closed string

D-branes

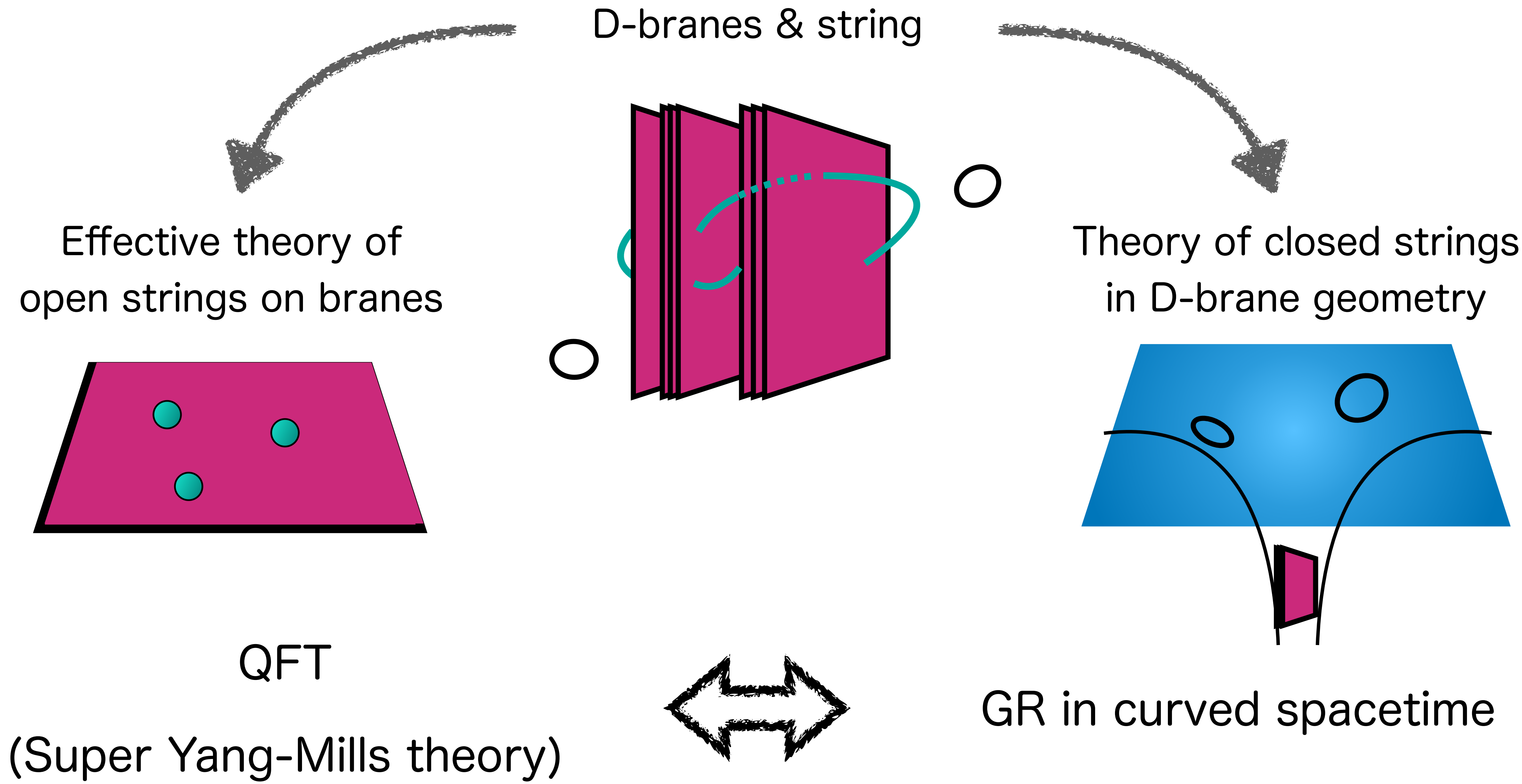


: the endpoints of open strings
can attach.

We want to understand the physical properties of D-branes.

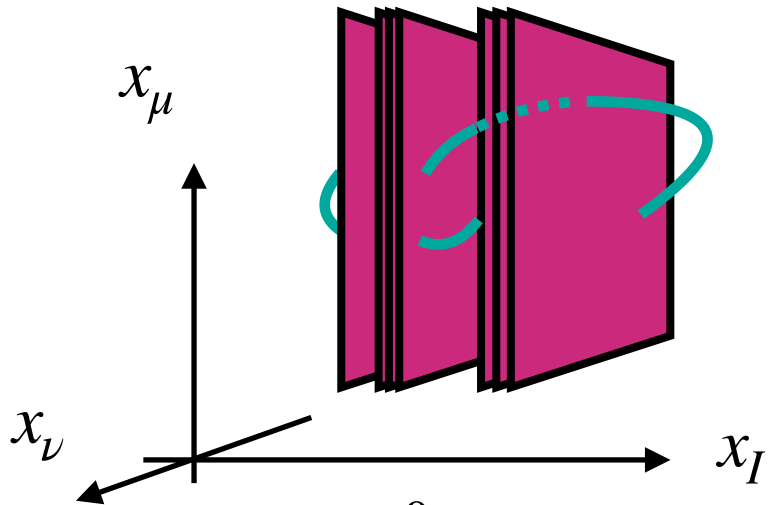
A clue : gauge/gravity duality

conjecture from 2 descriptions of D-branes in string theory;



Expected to capture the nonperturbative aspects of string theory

Position of D-branes & open strings



$\mu, \nu = 0, \cdots, p$

$I, J = p+1, \cdots, 9$

Effective action (: (p+1)-dim. U(N) gauge theory)

$$\int d^{p+1}x \,\mathrm{tr}\left(\frac{1}{4}F_{\mu\nu}^{\ 2} + \frac{1}{2}(D_\mu X_I)^2 + \frac{g^2}{4}[X_I, X_J]^2 + (\text{fermion terms})\right)$$

$X_I(x)$: $N \times N$ hermitian matrices

$N \gg 1$ to satisfy the duality

Some special cases

$\rightarrow$ X : simultaneously diagonal

diagonal : position of D-branes

off-diagonal : open strings among D-branes

[Witten, (1995)]

$\because$) Suppose $X_I = Y_I + \tilde{X}_I, \quad Y = \mathrm{diag}(y_1, \cdots, y_N),$

$$\mathrm{tr}\left[Y_I, X_J\right]^2 = (Y_I X_J - X_I Y_J)^{ij}(Y_I X_J - X_I Y_J)^{ji} \supset -(y_I^i - y_I^j)^2 |X_I^{ij}|^2$$

And remember (open string mass) = (string tension) $\times$ (string length).

How to read "geometry" in general?

← D-brane configuration

Key; Separation of the classical modes and fluctuation around them

$$X_I = Y_I + \tilde{X}_I$$

[Polchinski, (1998/1999) / Susskind, (1999) /⋯]

However, we cannot diagonalize X_I simultaneously.

∵) In the 't Hooft limit,

$$N\left\langle \mathrm{tr}\, X_I^2 \right\rangle, \quad N\left\langle \mathrm{tr}\left[X_I, X_J\right]^2 \right\rangle \sim O(N^2) \quad \Rightarrow \quad (\text{eigenvalue of } X_I) \sim O(N^0)$$

When diagonalizing X_1, $X_{J\neq 1}$ are far from diagonal;

$$\mathrm{tr}\left[X_1, X_{J\neq 1}\right]^2 = \sum_{i,j} \underbrace{(X_1^{ii} - X_1^{jj})^2}_{\sim O(N^0)} \underbrace{\left|X_J^{ij}\right|^2}_{\sim O(N^{-1})}, \qquad \mathrm{tr}\, X_J^2 = \sum_i (X_J^{ii})^2 + \underbrace{\sum_{i\neq j} \left|X_J^{ij}\right|^2}_{\sim O(N)!!}$$

Need to extract a "classical mode" by dropping open string fluctuations.

In Hamilton formalism

Let us consider the Matrix Quantum Mechanics ($p=0$) for simplicity.

→ each matrix element is an **operator**

$$\hat{X}_{I,ij} = \sum_{a=1}^{N^2} \hat{X}_{I,a}\tau^a_{ij}, \quad \hat{P}_{I,ij} = \sum_{a=1}^{N^2} \hat{P}_{I,a}\tau^a_{ij}$$

τ^a : generator of $G = \mathrm{U}(N)$

$$\mathrm{tr}(\tau_a\tau_b) = \delta_{ab}$$

$$\sum_a (\tau_a^{ij}\tau_a^{kl}) = \frac{1}{N}\delta^{ik}\delta^{jl}$$

Uncertainty relation

$$[\hat{X}_{I,a}, \hat{P}_{J,b}] = \mathrm{i}\delta_{IJ}\delta_{ab}$$

- Hilbert space;

$$\mathscr{H} = \mathrm{Span}\left\{ |X\rangle;\ \hat{X}_{I,a}|X\rangle = X_{I,a}|X\rangle \right\} = \mathrm{Span}\left\{ |P\rangle;\ \hat{P}_{I,a}|P\rangle = P_{I,a}|P\rangle \right\}$$

"coordinate basis" "momentum basis"

- Partition function at finite temperature

$$Z(T) = \frac{1}{\mathrm{Vol}G}\int_G \mathrm{d}g\, \mathrm{Tr}_{\mathscr{H}}\left(\hat{g}\, \mathrm{e}^{-\hat{H}/T}\right) = \mathrm{Tr}_{\mathscr{H}_{\mathrm{inv}}}\left(\mathrm{e}^{-\hat{H}/T}\right)$$

Notion of wave packet

[Hanada (2021)]

In the same meaning of the previous slide,

we cannot use the coordinate (or momentum) eigenstate.

$\because$) By $[\hat{X}_{I,a}, \hat{P}_{J,b}] = \mathrm{i}\delta_{IJ}\delta_{ab}$, the eigenstates have infinitely large energy,

containing a lot of quantum fluctuation.

Remember that, in the quantum mechanics,

Closest state to classical state is wave packet.

To identify the geometry, consider the wave packet in dN^2-dim space

✔ $$|\Phi\rangle = \int_{\mathbb{R}^{dN^2}} \mathrm{d}X\,|X\rangle\langle X|\Phi\rangle = \int_{\mathbb{R}^{dN^2}} \mathrm{d}X\,\Phi(X)\,|X\rangle \qquad I = 1,2,\cdots,d$$

and the center of $\Phi(X)$, $Y_{I,a}$ (c.f. coherent state)

Wave packet in color space

[Hanada (2021)]

The center of wave packet Y_I determines the location of D-branes!

$$|\Phi\rangle = |Y;Q\rangle, \qquad \langle\Phi|\hat{X}_I|\Phi\rangle = Y_I, \quad \langle\Phi|\hat{P}_I|\Phi\rangle = Q_I$$

: localized around Y_I and Q_I in each basis.

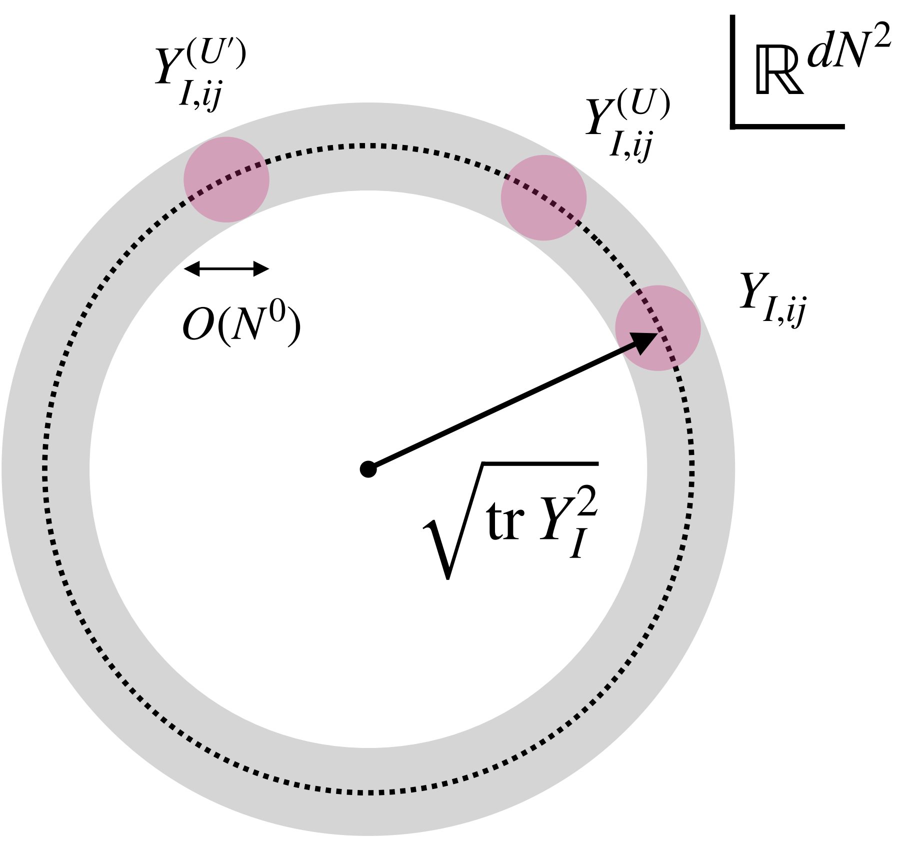
(e.g., for free matrix QM, it corresponds to the coherent state.)

Affected by gauge transformations

$$\hat{X}_{I,ij} \to \left(U\hat{X}_I U^{-1}\right)_{ij} = \sum_{k,l=1}^{N} U_{ik}\hat{X}_{I,kl}U^{-1}_{lj} =: \hat{X}^{(U)}_{I,ij}$$

provides the gauge orbit of $Y_{I,a}$

- position of the wave packet moves
  $\Leftrightarrow$ “diagonalizability” of Y
- But shape of the wave packet are unchanged



for free matrix QM

Determination of wave packet

How to identify the low-energy wave function for generic theory?

Proposal 1 (Hamiltonian formalism) [Hanada (2021)]

$$\min_{\Phi} \langle \Phi | \hat{H} | \Phi \rangle \quad \text{with given} \quad \langle \Phi | \hat{X}_I | \Phi \rangle = Y_I, \quad \langle \Phi | \hat{P}_I | \Phi \rangle = Q_I, \quad \cdots$$

Proposal 2 (Path-integral formalism) [Hanada, Kanno, Matsuura, HW, in progress]

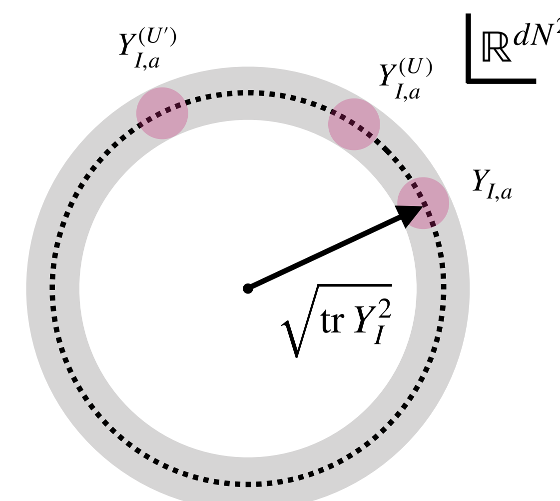
- Prepare $\{X_I\}$, and find a unitary matrix U minimizing R_∞ with given $Y_I^{(\mathrm{trial})}$

$$R_\infty(X, Y^{(\mathrm{trial})}) := \min_U \left(\max_{I,a} \left| \left(X_I^{(U)} - Y_I^{(\mathrm{trial})} \right)_a \right| \right)$$

: L_∞-distance or Chebyshev distance

- Vary $Y_I^{(\mathrm{trial})}$ searching $\min_Y R_\infty(X, Y)$
- Repeat above for different X_I and take average

$\langle R_\infty(X, Y_{\min}) \rangle$ is gauge invariant



: Variational approaches

Contents

- D-brane geometry from matrix model
    - Center of wave packet in the space of color dof. and a proposal determining it from matrix configs.
- Monte Carlo method to the minimization problem
    - Basic strategy by Replica-Exchange method and its extensions
- Numerical results
- Summary

C.f.) Fukuma-san's afternoon talk;

Applying the Worldvolume Hybrid Monte Carlo method to dynamical fermion systems

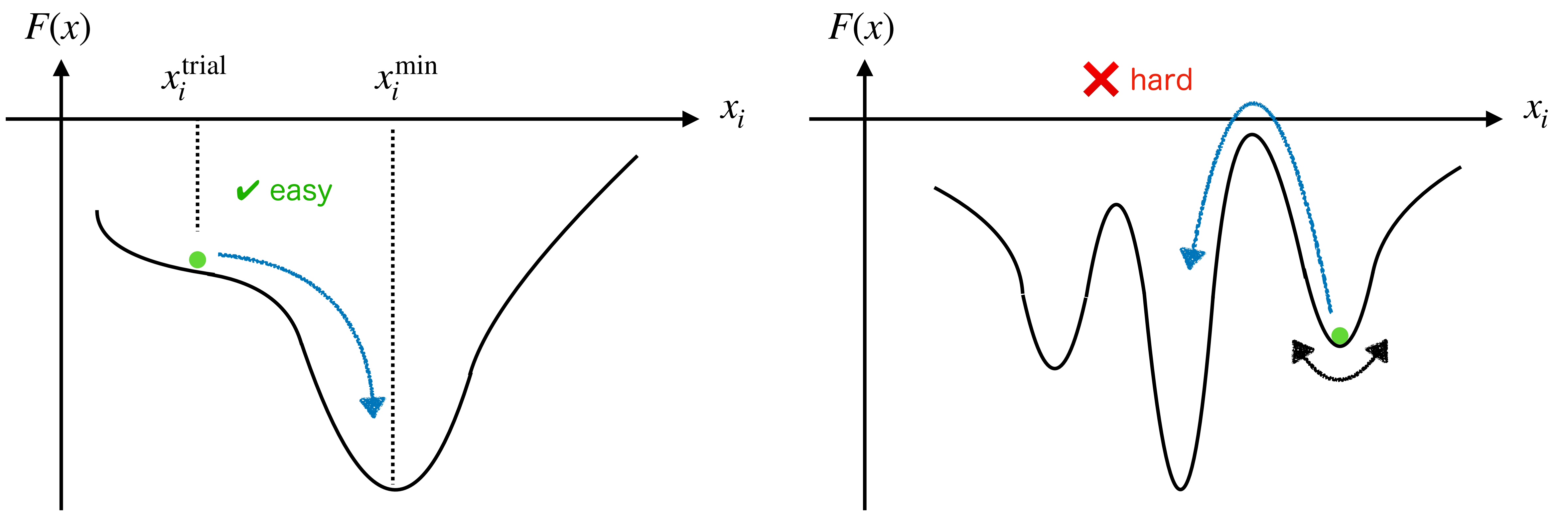
(Several similarities can be found.)

MCMC & minimization

- Importance sampling realized by Markov-chain Monte Carlo method is applicable to generic "potential" $F(x)$, by regarding it as the action $S(x)$

$$P(x) \propto \mathrm{e}^{-F(x)} \qquad x^{(0)} \to x^{(1)} \to x^{(2)} \to \cdots \to x^{(N_\mathrm{s})}$$

- Importance sampling is a powerful tool not only for performing integrals but also to search the minima of $F(x)$, if "tunneling" effect is enhanced.

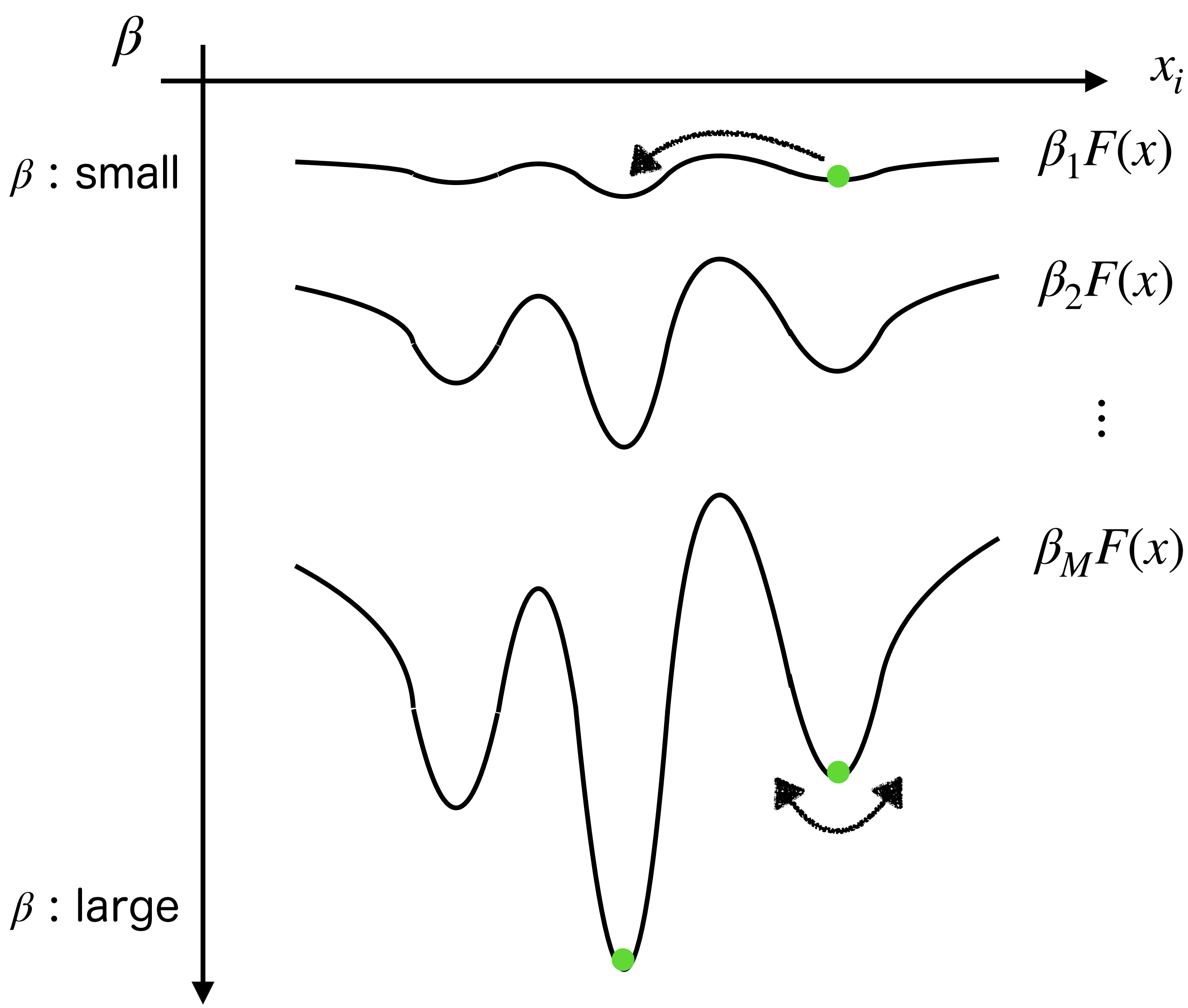


→concept of annealing which introduce a "temperature"

Simulated Annealing (SA)

[Kirkpatrick, Gelatt, Vecchi, (1983)]

known also as the tempering, is a method searching global minimum;



$F(X)$: a function to be minimized

$\beta_1 < \beta_2 < \cdots < \beta_M$: fictitious inverse temp.

scaling the depth of "potential"

- At small β, "tunneling" occurs easily. (various region can be reached)
- At high β, potential depth grows. →Precision of determining optimized config. improves.

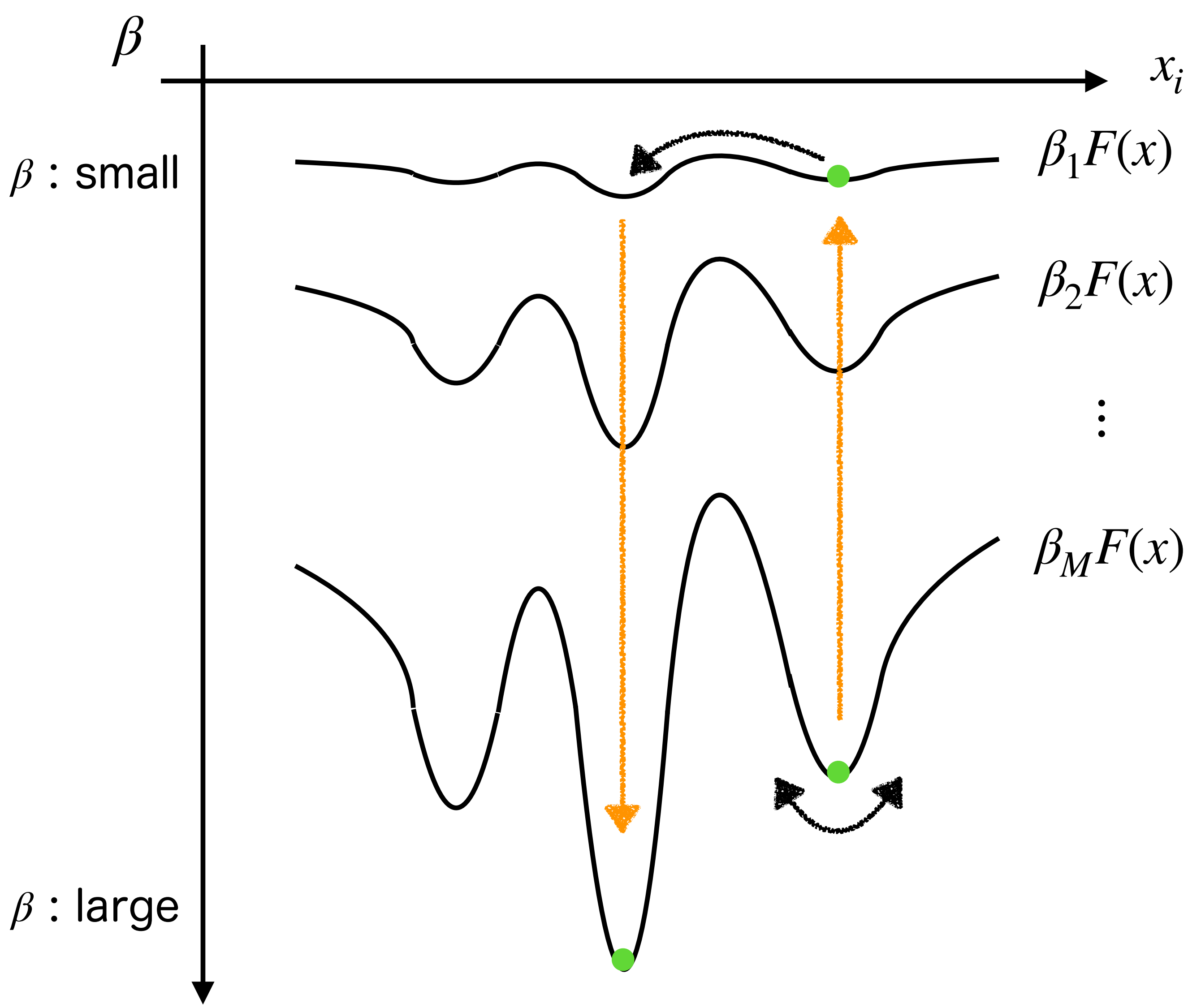
SA : gradually lowering "temperature"

MCMC with $\beta_1 F(x)$ → input: $x_{\rm min}^{(\beta_1)}$ → MCMC with $\beta_2 F(x)$ → ... → input: $x_{\rm min}^{(\beta_{M-1})}$ → MCMC with $\beta_M F(x)$

Replica-Exchange Monte Carlo (REMC)

[Swendsen, Wang, (1986) / Geyer, (1991)]

known also as the parallel tempering, is an upgraded method of SA;



Key:
config. exchange among
the simulations of different βs

- Running the simulations of different "temperature" simultaneously. (hence, it costs a lot of resource)

- Exchange configurations x_m & x_{m+1} ($m = 1, \cdots, M-1$) with weight (:Metropolis test)

$$\Delta S := \beta_m F(x_{m+1}) + \beta_{m+1} F(x_m) - \beta_m F(x_m) - \beta_{m+1} F(x_{m+1})$$

- REMC enhances more tunneling than SA.
- REMC removes an approximation in SA, coming from finite trials in each β.

Still, it was insufficient for us ⋯

Further improvements

In our work,

$$R_\infty(X, Y^{(\mathrm{trial})}) := \min_U \left(\max_{I,a} \left| \left(X_I^{(U)} - Y_I^{(\mathrm{trial})} \right)_a \right| \right)$$

is what we want to compute.

However, we struggled with the following issues;

- Minimization hits the limit quickly due to stuck in local minima.
- REMC costs much, but we need to repeat it for several configs.

Therefore,

- "Regularization" of replica action $F(U) = R_\infty(U; X)$
- Introduction of Replica-Exchange SA (RESA)

Refinement of replica actions

Still severe to minimize the L_∞-distance due to the huge #local minima

original problem

$$R_\infty(U,X) = \max_{I,a} |X_I^{(U)} - Y_I|_a$$

: L_∞-distance

⇒

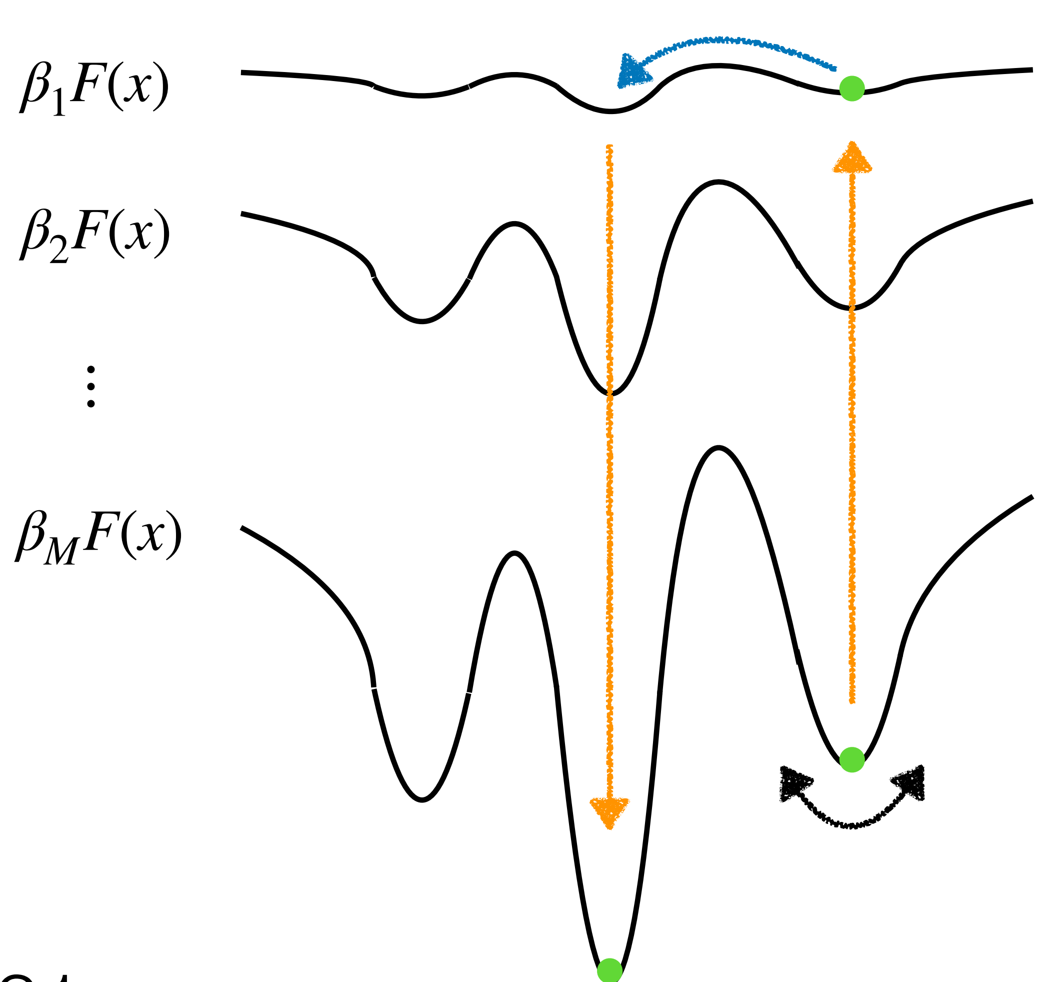
new problem

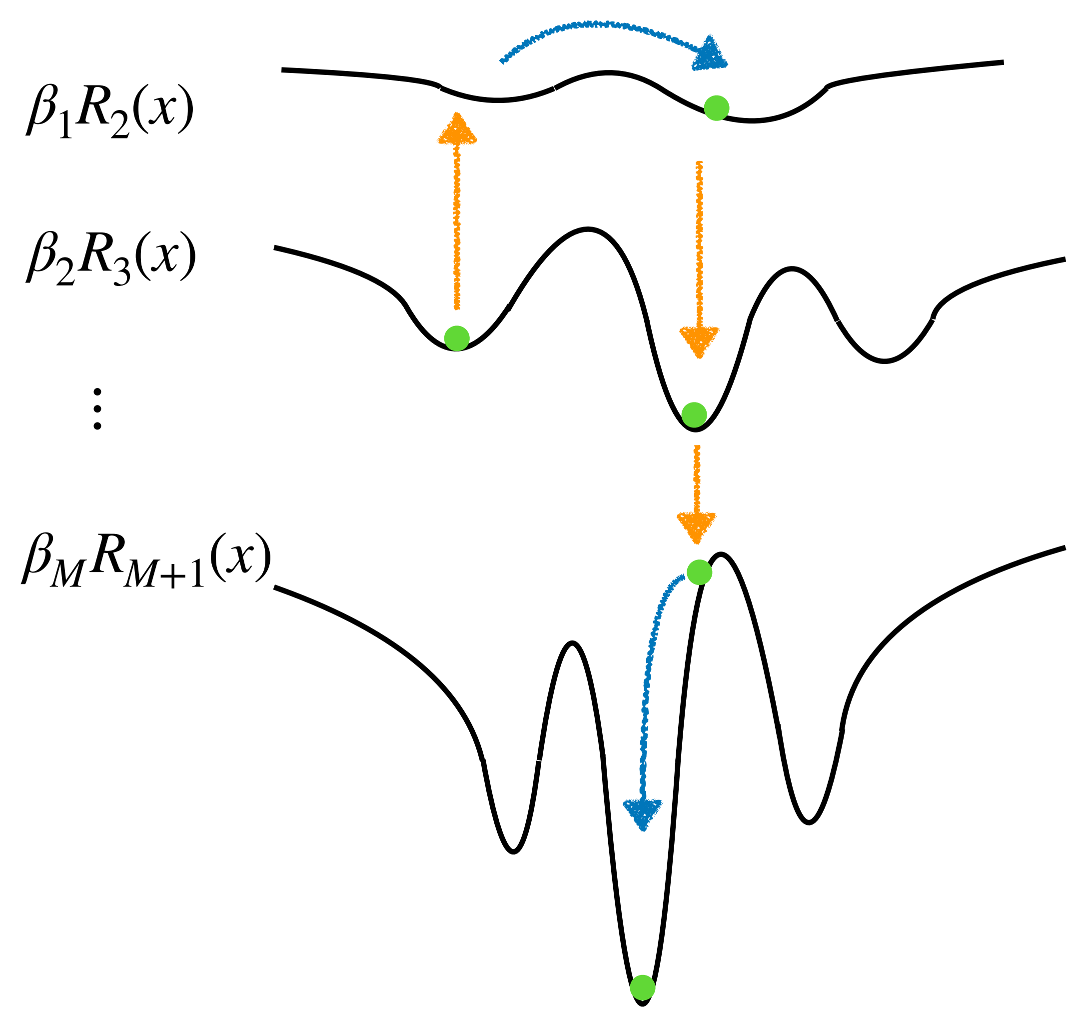
$$R_p(U,X) = \left(\sum_{I,a} |X_I^{(U)} - Y_I|_a{}^p \right)^{1/p}$$

: L_p-distance

[Hanada, Kanno, Matsuura, HW, in progress]

extend by introducing "evaluation function" on each replica





Properties of extended REMC

- MCMC algorithm in each replica → guaranteed it could work
- Different pot. structure among replicas → many minimizing path

  ∵) for an X $(:= X^{(U)})$

  $$R_2(X) \geq R_3(X) \geq \cdots \geq R_\infty(X) \geq 0$$: monotonic series of X

  which implies great acceptance for larger p.

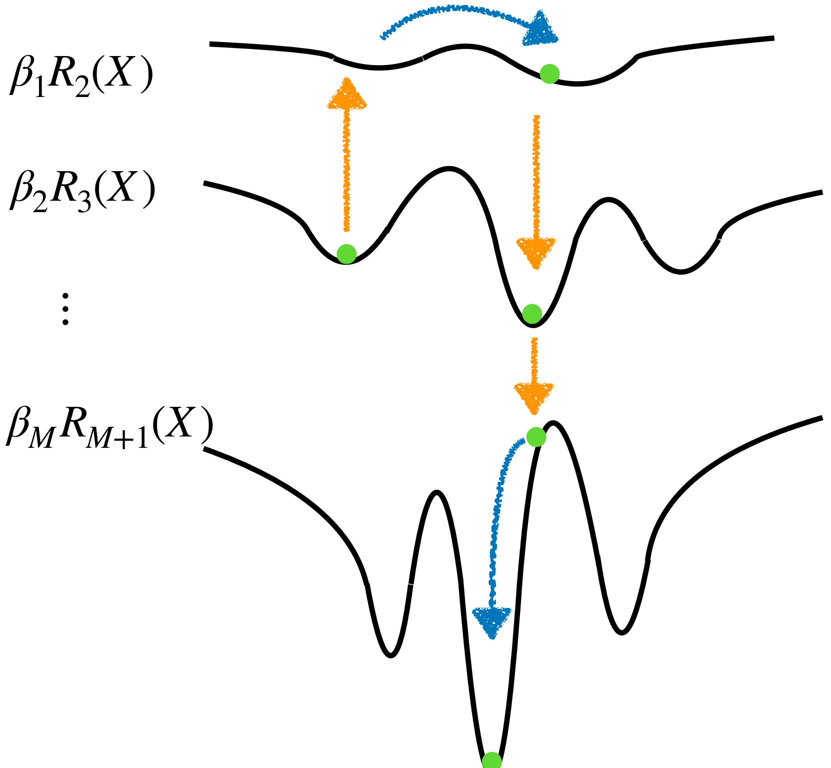
  But not satisfied

  $$\beta_2 R_2(X) \geq \beta_3 R_3(X) \geq \cdots \geq \beta_M R_M(X), \qquad \beta_2 < \beta_3 < \cdots < \beta_M$$

- Less local minima for smaller p

  ∵) $R_2(X^{(U)}, Y_I) = \sqrt{\mathrm{tr}(X_I^{(U)} - Y_I)^2}$ is gauge inv.

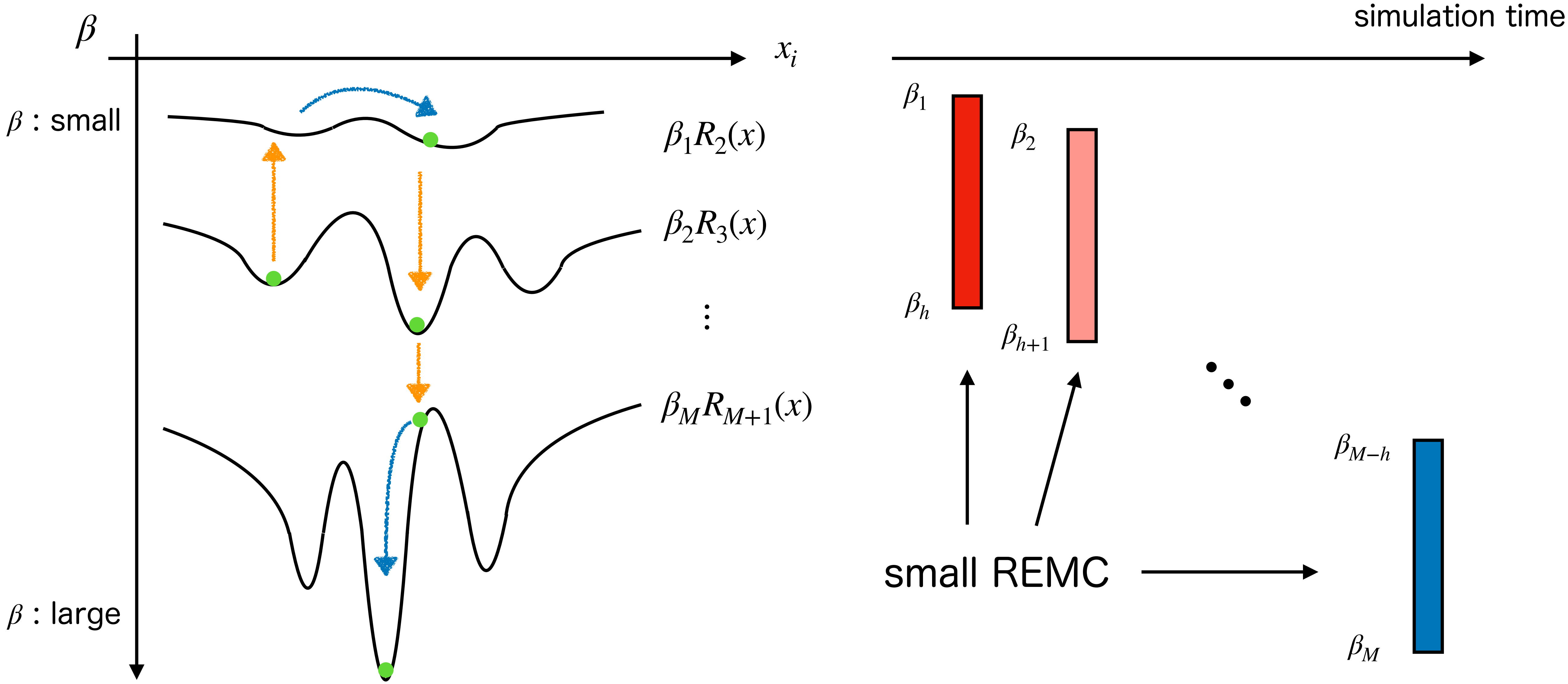
- $R_p(X) \approx R_{p+1}(X)$ for sufficiently large p



Replica-Exchange SA (RESA)

[Hanada, Kanno, Matsuura, HW, in progress]

RESA = Annealing of REMC with small replicas



- Drastic reduction of computational resource and time.
- Combination with refined replica actions may create synergy.

Contents

Prep.: Mock-data analysis

Consider the simple setup ($I = 1$ case)

$$X_{ij} = \sum_{a=1}^{N^2-1} X_a \tau^a_{ij}, \qquad X_a = 1$$

$\mathrm{tr}\, \tau^a \tau^b = \delta^{ab}$

τ^a : $\mathrm{SU}(N)$ generator

and prepare the mock data Z by randomly generating the unitary matrices V

$$Z = VXV^{-1} \qquad (Y = \mathbf{O})$$

We minimize the distance w.r.t. {Z}

$$R_\infty(U, Z) := \max_a \left| Z^{(U)} \right|_a$$

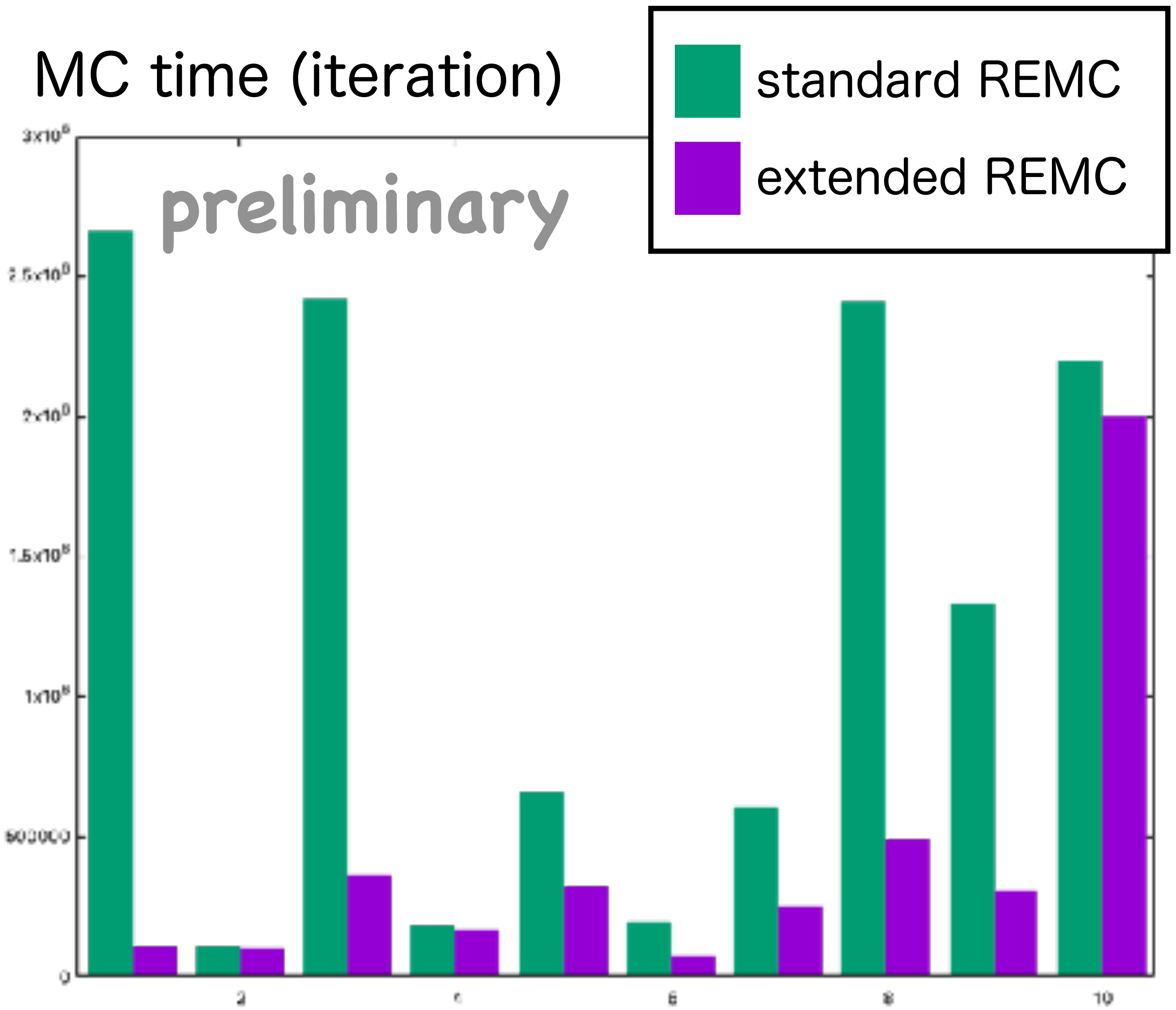
$$R_\infty(U, X) = 1$$

$$\longrightarrow \quad R_p(U, X) = \left(\sum_a |Z^{(U)}|_a{}^p \right)^{1/p}$$

The searching problem of a unitary matrix $U \approx V^{-1}$

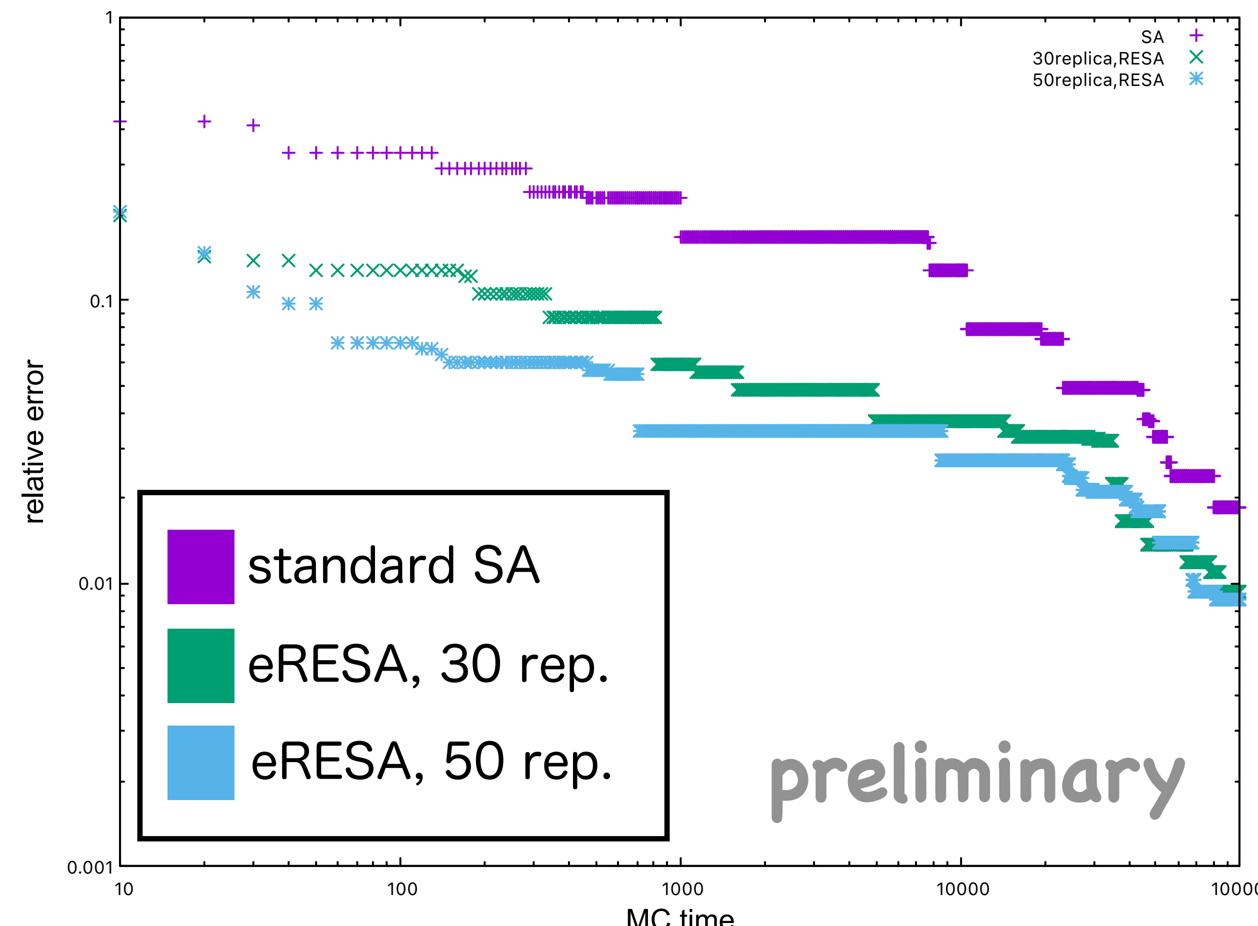
Prep.: Mock-data analysis

Demonstration: 4×4 matrix in which we know the answer



(#replica = 500, (threshold) = 1.025)

Relative error



MC time

→ Minimization by eREMC, eRESA tends faster than standard ones.

Example: One-matrix model

$$S(X) = N\,\mathrm{tr}\left(\frac{m^2}{2}X^2 + \frac{1}{4}X^4\right), \qquad Z = \int \mathrm{d}X\,\mathrm{e}^{-S(X)}$$

:"(0+0)-dim" toy model

and assuming $m^2 < 0$ and $\mathrm{tr}X = 0$. "Classical" minima can be described as

$$C = \mathrm{diag}\,(\underbrace{+c, \cdots, +c}_{\# = N/2}\ \underbrace{-c, \cdots, -c}_{\# = N/2}) \qquad c = \sqrt{-m^2}$$

We prepare {X} by MC simulation and minimize the distance w.r.t. X

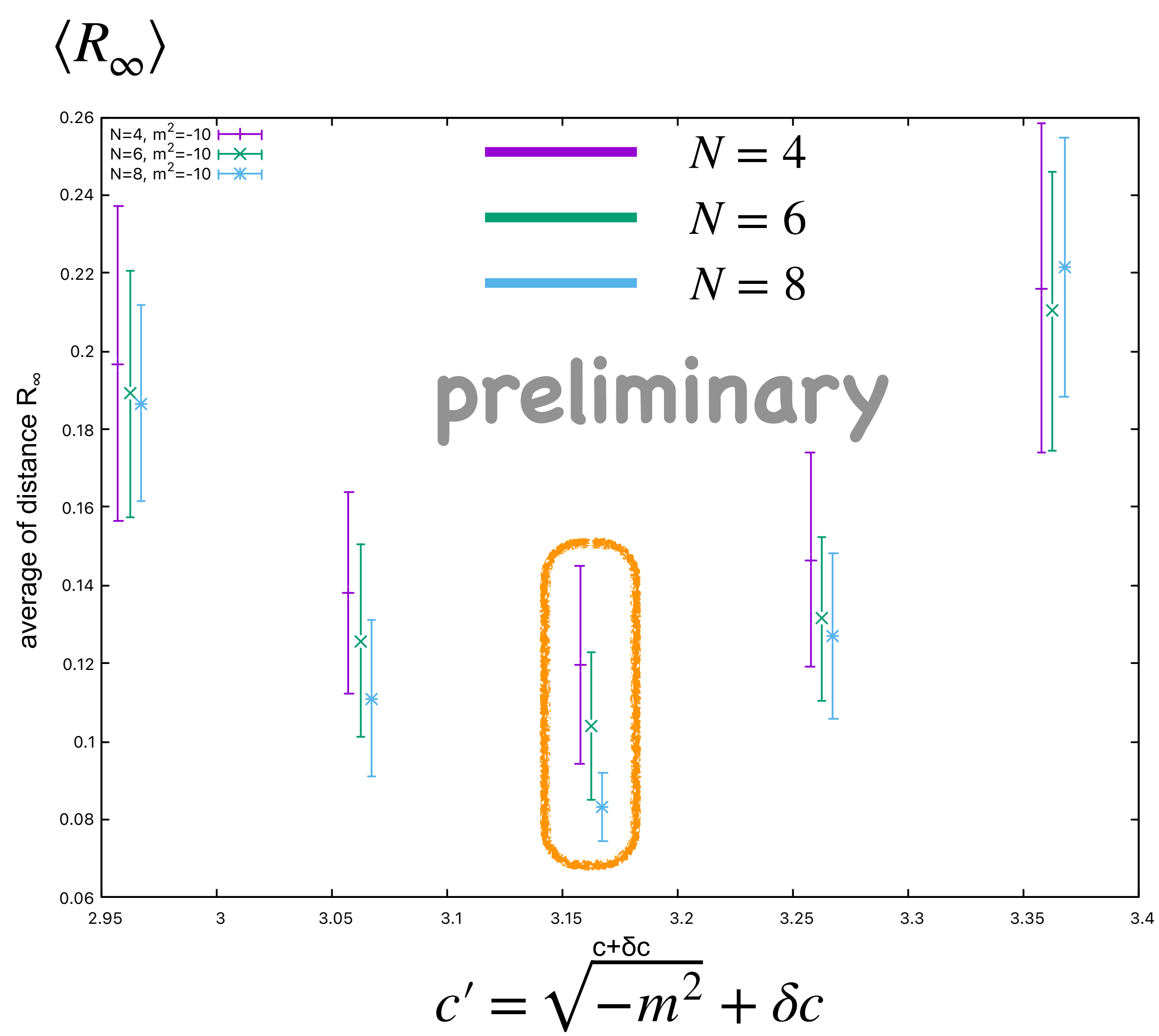
$$R_\infty(U, X) = \max_a |X^{(U)} - C'|_a \longrightarrow R_p(U, X) = \left(\sum_a |X^{(U)} - C'|_a{}^p\right)^{1/p}$$

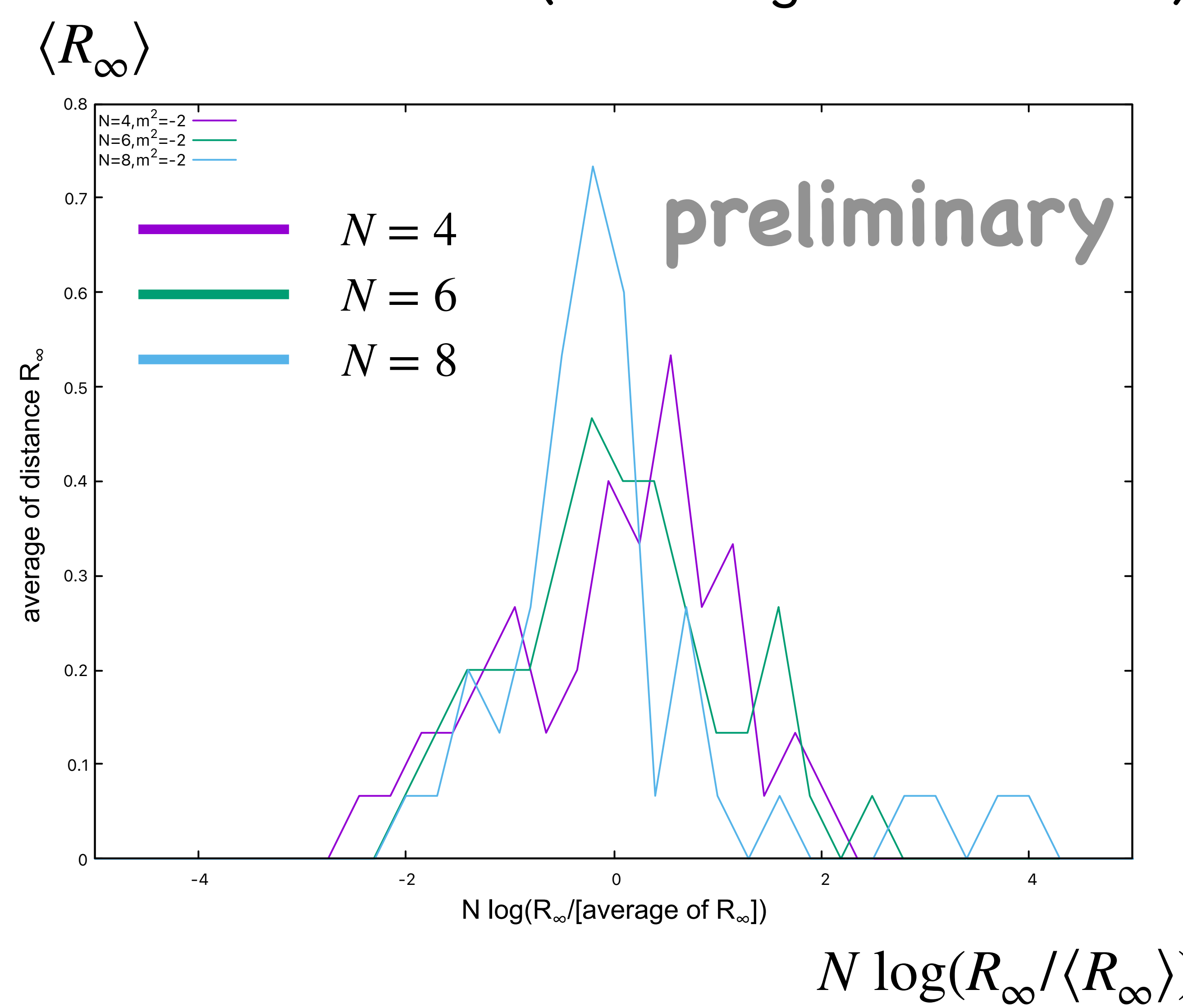
with changing the ansatz

$$C' := \mathrm{diag}(+c', \cdots, +c', -c', \cdots, -c'), \qquad c' = \sqrt{-m^2} + \delta c$$

Result : optimal C

(50 configs. on $m^2 = -10$)

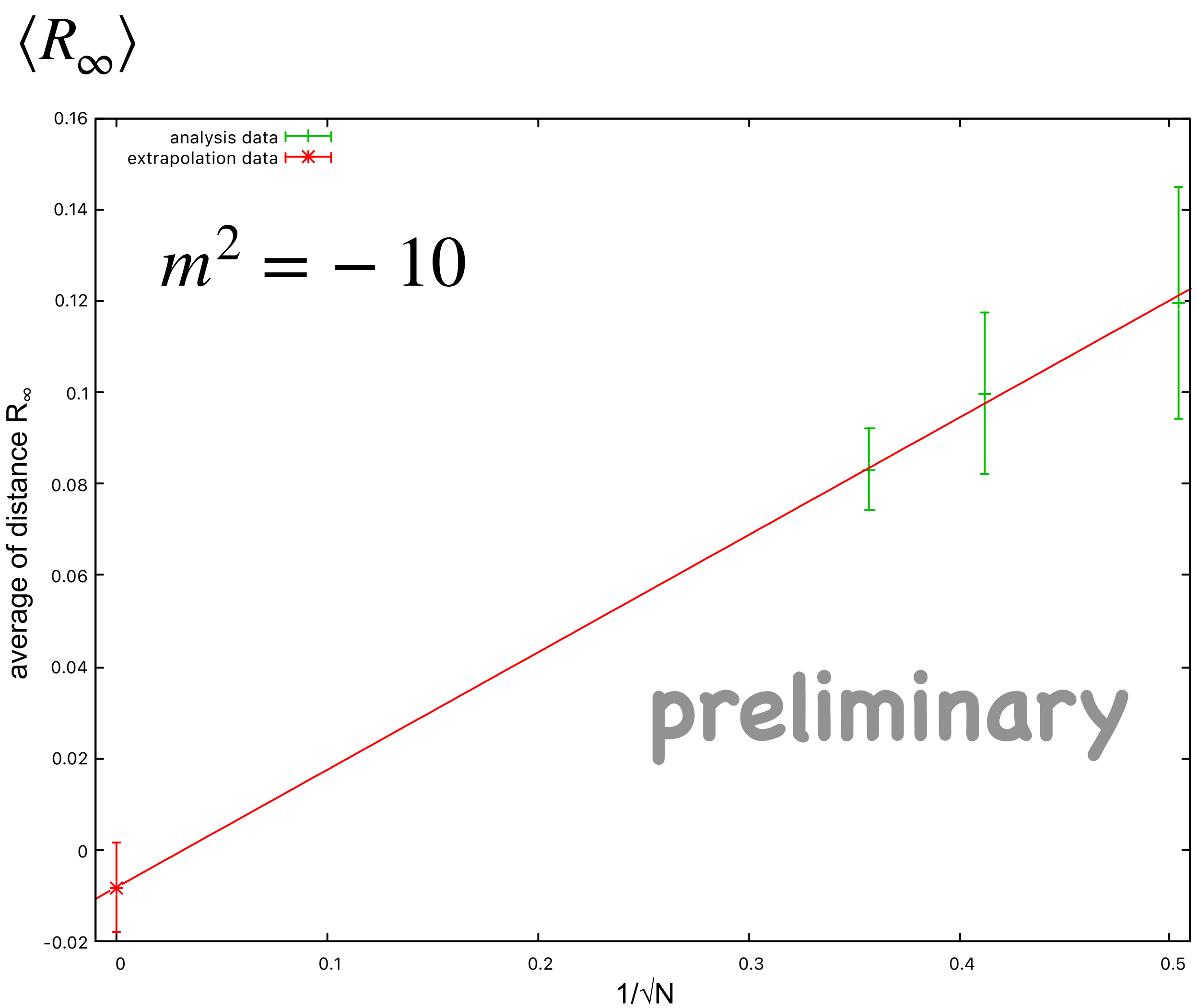


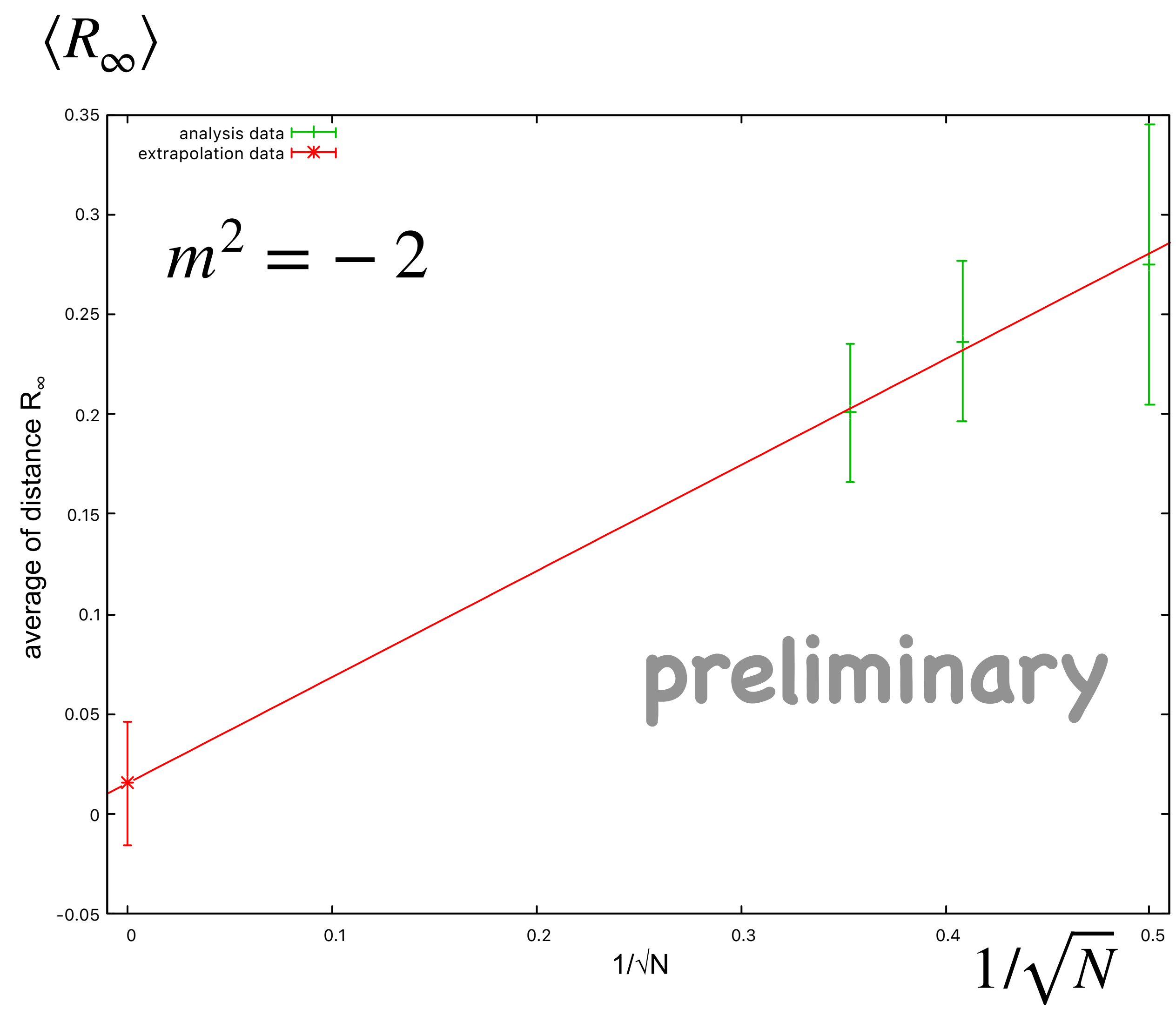


[Left] “Classical” config. minimizes the distance → good candidate for “geometry”

[Right] Histogram of $R_\infty(X)$ shows that width scales by N, as theoretically expected.

Large-N extrapolation



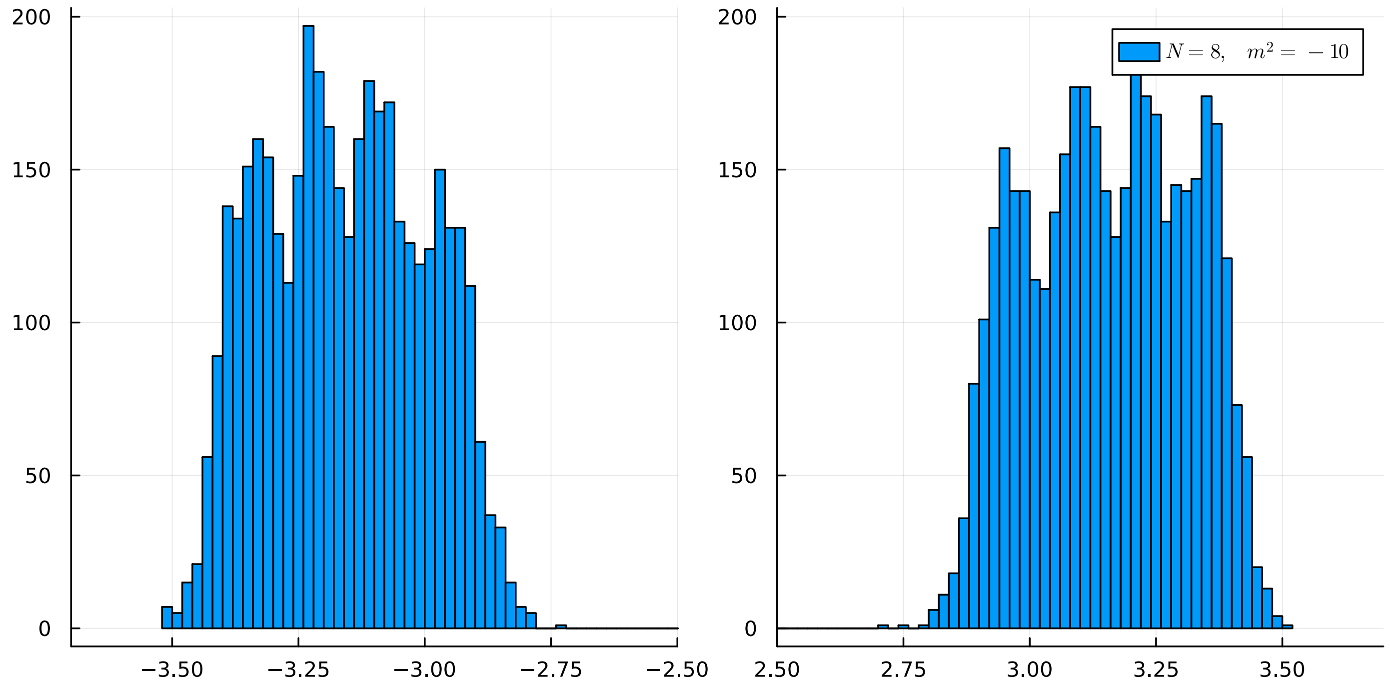


- Large N extrapolation shows an $1/\sqrt{N}$ scaling and convergence to zero, which is predicted from theoretical side!

$$\mathrm{tr}(X_I - Y_I)^2 = \sum_a \left| X_I - Y_I \right|_a^2 \sim O(N), \qquad R_\infty \sim \max_a \left| X_I - Y_I \right|_a \sim O(N^{-1/2})$$

c.f.) Eigenvalue distribution

e.g.) $N = 8$

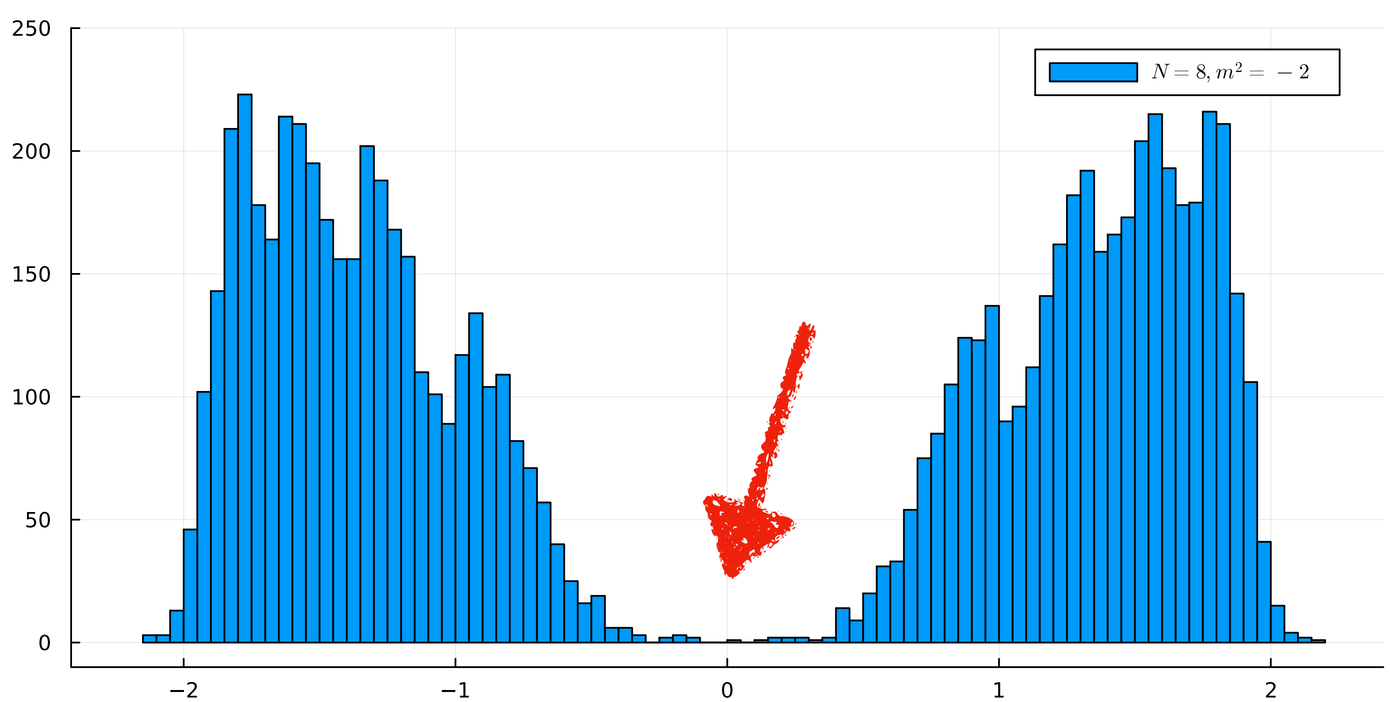


$m^2 = -10$

No distribution in [-2.5, 2.5].

Two bunches are disconnected.

"Higgsed" phase



$m^2 = -2$

Two bunches start to connect

at $m^2 \geq m_c^2$.

Even in this region, we can extract "center of wave packet"!

Example(2): Fuzzy sphere matrix model

[Iso, Kimura, Tanaka, Wakatsuki, (2001)]

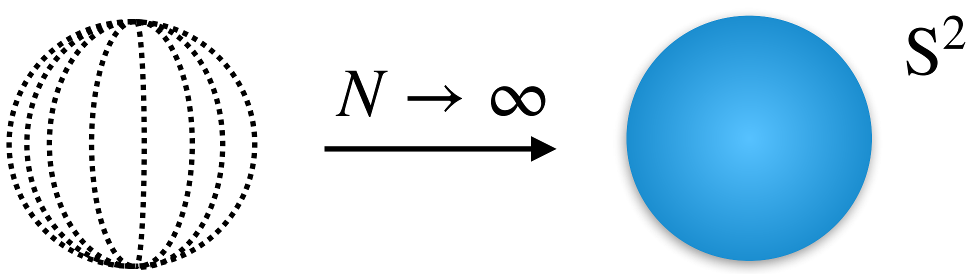
$$S(X_1, X_2, X_3, \psi) = N\mathrm{tr}\left(-\frac{1}{4}[X_I, X_J]^2 + \frac{2i\mu}{3}\epsilon_{IJK}X_I X_J X_K + \frac{1}{2}\bar{\psi}\sigma^I[X_I, \psi] + \mu\bar{\psi}\psi\right)$$

: X_Is are not simultaneously diagonalizable

"Classical" minima : Fuzzy sphere solution

$$X_I^{\mathrm{classical}} = \mu J_I, \qquad [J_I, J_J] = i\epsilon_{IJK}J_K$$

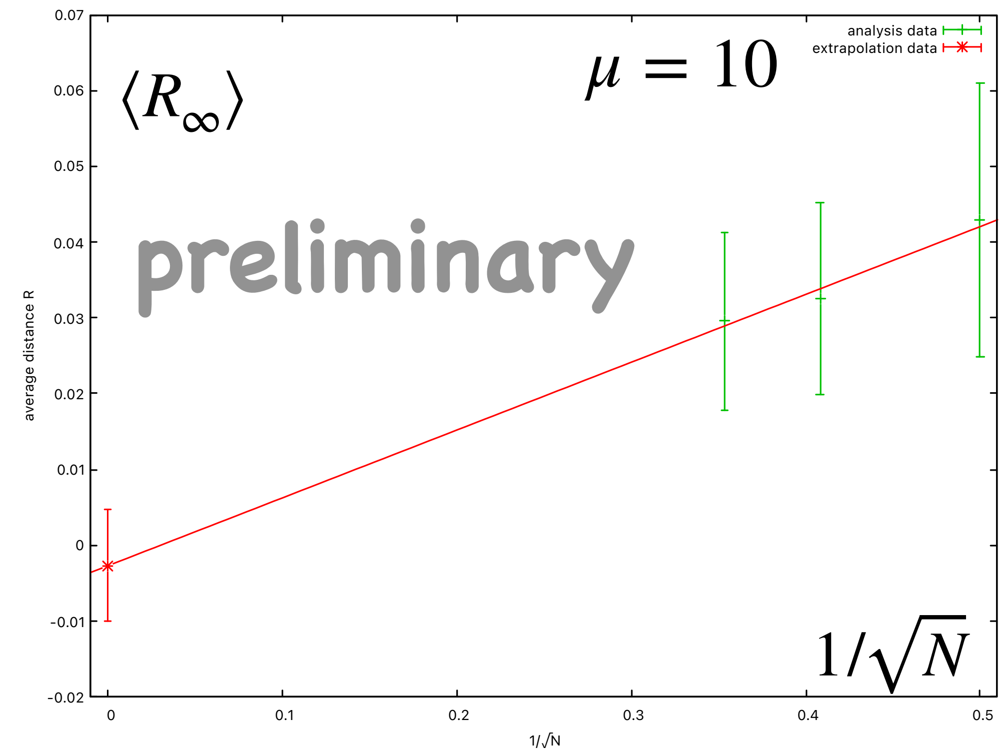
J_I : N-dim. rep. of SU(2) generator



$$R_{\mathrm{FS}}^2 = \frac{1}{N}\,\mathrm{tr}\,X_I^2 = \frac{\mu^2}{4}(N^2 - 1)$$

Minimize the distance w.r.t. U

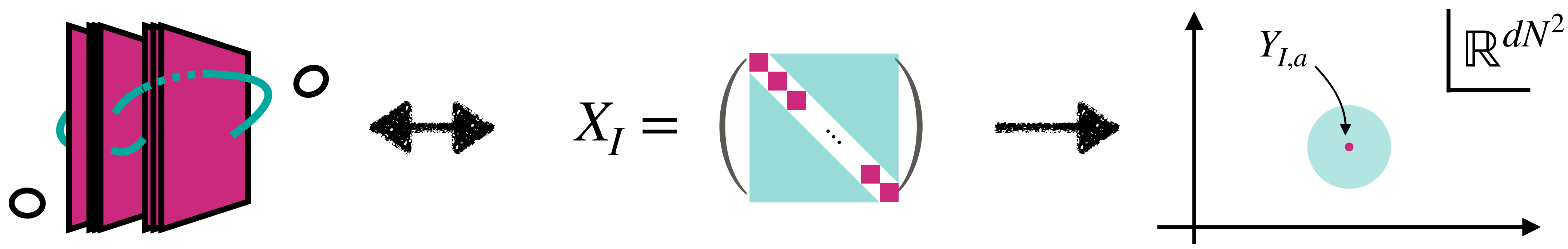
$$R_\infty(U, X) = \max_{I,a} |X_I^{(U)} - X_I^{\mathrm{FS}}|_a$$



Contents

Summary

To read off geometric information from field theory configs. in string theory, the notion of wave packet in space of matrix plays an essential role.



To determine the wave packet in a high-dim space, we compute a quantity

$$R_\infty(X) := \min_U \left(\max_a \left| \left(U^\dagger X U - Y\right)_a \right| \right)$$

X, Y : $N \times N$ hermitian mat.

U : unitary mat.

which can be translated into an optimization problem.

We employ the Replica-Exchange Monte Carlo methods (REMC) and consider their extensions to solve this problem numerically.

Future directions

- More detailed analysis for (bosonic) fuzzy-sphere three-matrix model

$$S(X_1, X_2, X_3, \psi) = N\mathrm{tr}\left(-\frac{1}{4}[X_I, X_J]^2 + \frac{2\mathrm{i}\mu}{3}\epsilon_{IJK}X_IX_JX_K + \frac{1}{2}\bar{\psi}\sigma^I[X_I, \psi] + \mu\bar{\psi}\psi\right)$$

- Compare the obtained results with those in Hamiltonian formalism.
- How should we determine better ansatz Y_I

  ← Essential for analyzing (0+1)d models (e.g. BFSS-type model) and so on.

- Further understanding, generalization, application of the extended REMC
  - Tuning of the loss function, if we say in the language of ML.
  - Combination of ML and RE method?