Mathematical understanding of violation of detailed balance condition and its application to Langevin dynamics

Masayuki Ohzeki[1] Akihisa Ichiki [2]

[1]Department of Systems Science, Kyoto University

[2]Green Mobility Collaborative Research Center, Nagoya University

2014/12/01

References: Phys. Rev. E 88, 020101(R) (2013) and cond-mat/1307.0434

Self Introduction

Ohzeki Masayuki (大関 真之 in Chinese characters)
Meaning of the family name "Ohzeki"

- A famous company of Sake
- The second grade of Sumo wrestler
  Just below "Yokozuna" in the ranking. Until the Yokozuna rank was introduced, Ohzeki was the highest rank attainable.
- Great guard in front of castle
  My ancestor was indeed a strong samurai But!! He attacked to Edo castle! In order to make revolution!

1. Review of sampling of the desired distribution
   - Aim
   - Markov-Chain Monte-Carlo and detailed balance condition
   - Violation of DBC
2. Langevin dynamics
   - Langevin equation and its corresponding Fokker-Planck equation
   - Violation of DBC: introducing an additional force
   - Example: washboard
   - Example: XY model
3. Mathematical assurance
   - Rough sketch of proof of accelerated relaxation
   - Asymmetric matrix/operator
4. Conclusion

Aim

We would like to generate the desired distribution in order to

- investigate the property of the equilibrium of many-body system

How to generate?

- Markov-Chain Monte-Carlo method (MCMC)
  Mainly for discrete and continuous variables and discrete time
- Langevin dynamics
  Mainly for continuous variables and continuous time

We demand faster convergence to a desired distribution.

Markov-Chain Monte-Carlo method

We generate a sequence of the stochastic dynamics following the master equation.

$$P_{t+1}(x) = \sum_{y} P(x|y)P_t(y), \tag{1}$$

where $P_t(x)$ is an instantaneous distribution.
We use the transition matrix $P(x|y)$ to imitate the stochastic dynamics:

$$P(x_t|x_{t-1})P(x_{t-1}|x_{t-2})\cdots P(x_2|x_1)P(x_1) \to P^{(\mathrm{ss})}(x_t) \tag{2}$$

After a sufficient iteration, we obtain the desired distribution.

Detailed balance condition

In order to obtain desired distribution, we often offer that the transition matrix satisfies

$$\frac{P(y|x)}{P(x|y)} = \frac{P^{(\mathrm{eq})}(y)}{P^{(\mathrm{eq})}(x)}, \tag{3}$$

where $P^{(\mathrm{eq})}(x)$, namely $\exp(-\beta E(x) - \beta F)$.

Flexibility

If we design the energy function $E(x)$, we generate the desired distribution.

Why use DBC?

It is simple to construct the transition matrix.

Detailed balance condition

In order to obtain desired distribution, we often offer that the transition matrix satisfies

$$\frac{P(y|x)}{P(x|y)} = \frac{P^{(\mathrm{eq})}(y)}{P^{(\mathrm{eq})}(x)}, \tag{3}$$

where $P^{(\mathrm{eq})}(x)$, namely $\exp(-\beta E(x) - \beta F)$.

Flexibility

If we design the energy function $E(x)$, we generate the desired distribution.

Why use DBC?

It is simple to construct the transition matrix.

Detailed balance condition

In order to obtain desired distribution, we often offer that the transition matrix satisfies

$$\frac{P(y|x)}{P(x|y)} = \frac{P^{(\mathrm{eq})}(y)}{P^{(\mathrm{eq})}(x)}, \tag{3}$$

where $P^{(\mathrm{eq})}(x)$, namely $\exp(-\beta E(x) - \beta F)$.

Flexibility

If we design the energy function $E(x)$, we generate the desired distribution.

Why use DBC?

It is simple to construct the transition matrix.

Balanced condition

In order to assure to generate the desired distribution in the steady state, we must demand the balanced condition.

$$\sum_{x} P(y|x)P^{(\mathrm{ss})}(x) = P^{(\mathrm{ss})}(y) \tag{4}$$

Flexibility

If we set $P^{(\mathrm{ss})}(x) \propto \exp(-\beta E(x))$, we can offer the desired distribution.

Why not use BC (Violation of DBC)?

It is “not” simple to construct the transition matrix.

Balanced condition

In order to assure to generate the desired distribution in the steady state, we must demand the balanced condition.

$$\sum_x P(y|x)P^{(\mathrm{ss})}(x) = P^{(\mathrm{ss})}(y) \quad (4)$$

Flexibility

If we set $P^{(\mathrm{ss})}(x) \propto \exp(-\beta E(x))$, we can offer the desired distribution.

Why not use BC (Violation of DBC)?

It is "not" simple to construct the transition matrix.

Balanced condition

In order to assure to generate the desired distribution in the steady state, we must demand the balanced condition.

$$\sum_x P(y|x)P^{(\mathrm{ss})}(x) = P^{(\mathrm{ss})}(y) \tag{4}$$

Flexibility

If we set $P^{(\mathrm{ss})}(x) \propto \exp(-\beta E(x))$, we can offer the desired distribution.

Why not use BC (Violation of DBC)?

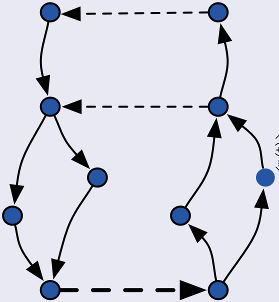
It is “not” simple to construct the transition matrix.

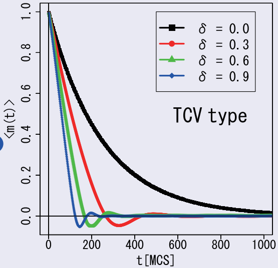
Recent development of violation of DBC

Several types without DBC are proposed.

Skewed DBC (2011)

DBC in the replicated system. We violate DBC for each individual system. The composite system does not violate BC.





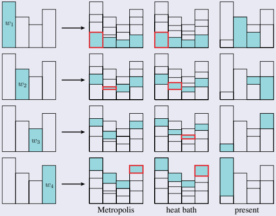
[K. S. Turitsyn, M. Chertkov and M. Vucelja: Physica D **240** (2011) 410.]
[Sakai and Hukushima: JPSJ **82** (2013) 064003.]

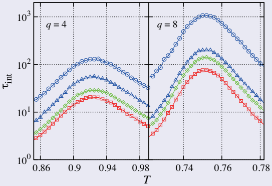
Recent development of violation of DBC

Several types without DBC are proposed.

Suwa-Todo method (2010)

A tricky algorithm based on BC.
While satisfying BC, they allocate the weight neglecting DBC.





[H.Suwa and S.Todo: Phys. Rev. Lett. **105** (2010) 120603.]

Similarly, is there a modification of the Langevin dynamics, which can accelerate the convergence to the desired distribution?

Langevin equation

The N-dimensional Langevin equation in an inverse temperature $\beta = 1/T$ is

$$d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \sqrt{\frac{2}{\beta}}d\mathbf{W} \tag{5}$$

where $\mathbf{A}(\mathbf{x}, t)$ is a force and $\mathbf{W}$ is the Wiener process.

corresponding Fokker Planck equation

The corresponding Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{6}$$

where $\mathbf{J}(\mathbf{x}, t)$ is the probability flow defined as

$$\mathbf{J}(\mathbf{x}, t) = \mathbf{A}(\mathbf{x}, t)P(\mathbf{x}, t) - T\mathrm{grad}P(\mathbf{x}, t) \tag{7}$$

The divergence is $\mathrm{div} = \sum_i \partial/\partial x_i$ and the gradient is $[\mathrm{grad}]_i = \partial/\partial x_i$.

Langevin equation

The N-dimensional Langevin equation in an inverse temperature $\beta = 1/T$ is

$$d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \sqrt{\frac{2}{\beta}}d\mathbf{W} \tag{5}$$

where $\mathbf{A}(\mathbf{x}, t)$ is a force and $\mathbf{W}$ is the Wiener process.

corresponding Fokker Planck equation

The corresponding Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{6}$$

where $\mathbf{J}(\mathbf{x}, t)$ is the probability flow defined as

$$\mathbf{J}(\mathbf{x}, t) = \mathbf{A}(\mathbf{x}, t)P(\mathbf{x}, t) - T\mathrm{grad}P(\mathbf{x}, t) \tag{7}$$

The divergence is $\mathrm{div} = \sum_i \partial/\partial x_i$ and the gradient is $[\mathrm{grad}]_i = \partial/\partial x_i$.

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{8}$$

Equilibrium system

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = 0$.

Solution

The force is given as

$$\mathbf{A}(x, t) = -\mathrm{grad}E(\mathbf{x}) \tag{9}$$

- No flow appears in the equilibrium state.

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{8}$$

Equilibrium system

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = 0$.

Solution

The force is given as

$$\mathbf{A}(x, t) = -\mathrm{grad}E(\mathbf{x}) \tag{9}$$

- No flow appears in the equilibrium state.

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{8}$$

Equilibrium system

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = 0$.

Solution

The force is given as

$$\mathbf{A}(x, t) = -\mathrm{grad}E(\mathbf{x}) \tag{9}$$

- No flow appears in the equilibrium state.

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{10}$$

Nonequilibrium system as a lesson

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = \gamma\mathbf{1}$.

Solution

The force is given as

$$\mathbf{A}(\mathbf{x}, t) = -\mathrm{grad}E(\mathbf{x}) + \gamma\mathbf{1}\exp\left(\beta E(\mathbf{x})\right) \tag{11}$$

- A constant flow appears in the steady state.
- However the force is unidirectional and exponentially large.

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{10}$$

Nonequilibrium system as a lesson

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = \gamma \mathbf{1}$.

Solution

The force is given as

$$\mathbf{A}(\mathbf{x}, t) = -\mathrm{grad}E(\mathbf{x}) + \gamma \mathbf{1} \exp\left(\beta E(\mathbf{x})\right) \tag{11}$$

- A constant flow appears in the steady state.
- However the force is unidirectional and exponentially large.

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{10}$$

Nonequilibrium system as a lesson

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = \gamma\mathbf{1}$.

Solution

The force is given as

$$\mathbf{A}(\mathbf{x}, t) = -\mathrm{grad}E(\mathbf{x}) + \gamma\mathbf{1}\exp\left(\beta E(\mathbf{x})\right) \tag{11}$$

- A constant flow appears in the steady state.
- However the force is unidirectional and exponentially large.

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{12}$$

Nonequilibrium divergence-free system

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = \mathbf{B}(\mathbf{x}, t) P(\mathbf{x}, t)$.

Solution

When we define

$$[\mathbf{B}(\mathbf{x}, t)]_i = \gamma \left(\frac{\partial E(\mathbf{x})}{\partial x_{i-1}} - \frac{\partial E(\mathbf{x})}{\partial x_{i+1}} \right), \; x_{N+1} = x_1 \text{ and } x_0 = x_N. \tag{13}$$

Then $\mathrm{div}\mathbf{J}(\mathbf{x}, \mathrm{t}) = 0$ and $[\mathbf{A}(\mathbf{x})]_i = -\frac{\partial E(\mathbf{x})}{\partial x_i} + [\mathbf{B}(\mathbf{x}, t)]_i$.

- The force is free from the exponentially large term.
- (Similar to Suwa-Todo method??).

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x}, t), \tag{12}$$

Nonequilibrium divergence-free system

We impose $P(\mathbf{x}, t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x}, t) = \mathbf{B}(\mathbf{x}, t) P(\mathbf{x}, t)$.

Solution

When we define

$$[\mathbf{B}(\mathbf{x}, t)]_i = \gamma \left(\frac{\partial E(\mathbf{x})}{\partial x_{i-1}} - \frac{\partial E(\mathbf{x})}{\partial x_{i+1}} \right), \; x_{N+1} = x_1 \text{ and } x_0 = x_N. \tag{13}$$

Then $\mathrm{div}\mathbf{J}(\mathbf{x}, \mathrm{t}) = 0$ and $[\mathbf{A}(\mathbf{x})]_i = -\frac{\partial E(\mathbf{x})}{\partial x_i} + [\mathbf{B}(\mathbf{x}, t)]_i$.

- The force is free from the exponentially large term.
- (Similar to Suwa-Todo method??).

Reference equation

The Fokker-Planck equation is

$$\frac{\partial P(\mathbf{x},t)}{\partial t} = -\mathrm{div}\mathbf{J}(\mathbf{x},t), \tag{12}$$

Nonequilibrium divergence-free system

We impose $P(\mathbf{x},t) \propto \exp(-\beta E(\mathbf{x}))$ and $\mathbf{J}(\mathbf{x},t) = \mathbf{B}(\mathbf{x},t)P(\mathbf{x},t)$.

Solution

When we define

$$[\mathbf{B}(\mathbf{x},t)]_i = \gamma\left(\frac{\partial E(\mathbf{x})}{\partial x_{i-1}} - \frac{\partial E(\mathbf{x})}{\partial x_{i+1}}\right), \ x_{N+1} = x_1 \text{ and } x_0 = x_N. \tag{13}$$

Then $\mathrm{div}\mathbf{J}(\mathbf{x},\mathrm{t}) = 0$ and $[\mathbf{A}(\mathbf{x})]_i = -\frac{\partial E(\mathbf{x})}{\partial x_i} + [\mathbf{B}(\mathbf{x},t)]_i$.

- The force is free from the exponentially large term.
- (Similar to Suwa-Todo method??).

Divergence free **replicated** system

We consider a composite system as $P(\mathbf{x}_1, \mathbf{x}_2, t)$. The Langevin equations are

$$d\mathbf{x}_i = \mathbf{A}_i(\mathbf{x}_1, \mathbf{x}_2, t) + \sqrt{\frac{2}{\beta}} d\mathbf{W}_i, \tag{14}$$

and the corresponding Fokker-Planck equations are

$$\frac{\partial P(\mathbf{x}_1, \mathbf{x}_2, t)}{\partial t} = -\sum_{i=1}^{2} \mathrm{div}_i \mathbf{J}_i(\mathbf{x}_1, \mathbf{x}_2, t), \tag{15}$$

Additional bidirectional forces

We impose the forces as

$$\mathbf{A}_1(\mathbf{x}_1,\mathbf{x}_2,t) = -\mathrm{grad}_1 E(\mathbf{x}_1) + \gamma \mathrm{grad}_2 E(\mathbf{x}_2) \tag{16}$$

$$\mathbf{A}_2(\mathbf{x}_1,\mathbf{x}_2,t) = -\mathrm{grad}_2 E(\mathbf{x}_2) - \gamma \mathrm{grad}_1 E(\mathbf{x}_1) \tag{17}$$

- Each flow is given by
  $\mathbf{J}_i(\mathbf{x}_1,\mathbf{x}_2,t) = \pm\gamma\left(\mathrm{grad}_j E(\mathbf{x}_j)\right) P(\mathbf{x}_1,\mathbf{x}_2,t)$ $(i \neq j)$
  but $\sum_i \mathrm{div}_i \mathbf{J}_i = 0$ (In this sense, similar to the skewed DBC.)
- The steady state is $P^{(\mathrm{ss})}(\mathbf{x}_1,\mathbf{x}_2) \propto \exp\left(-\beta E(\mathbf{x}_1) - \beta E(\mathbf{x}_2)\right)$.
- Implimentation is very simple.
- We remove exponentially large force.

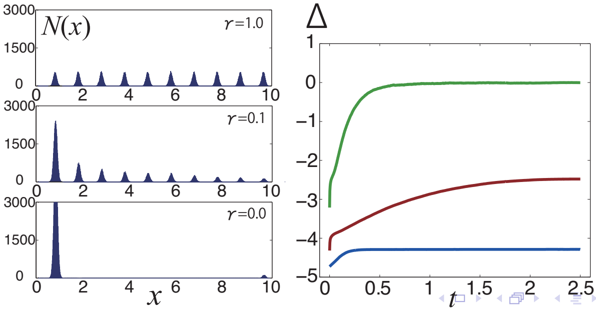
Example: washboard

We set $N = 10,000$ particles in the potential energy, which has a washboard shape characterized by a function,

$$E(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N}(1+\sin(2\pi x_i)) \tag{18}$$

for $x_i \in [0, 10)$.

at $t = 2.5$ in $\beta = 10$.



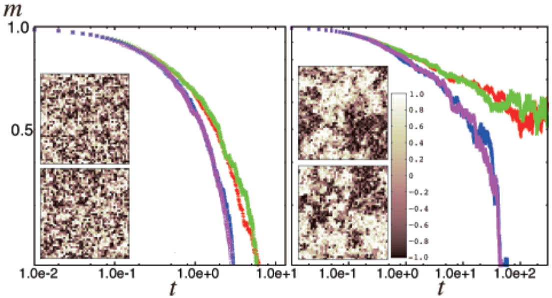
Example: XY model

We employ the XY model as an interacting many-body system

$$E(\mathbf{x}) = -\sum_{j\in\partial i} \cos\left(x_i - x_j\right), \tag{19}$$

Note that x_i here denotes the spin direction such that $x_i \in [0, 2\pi)$.

We set $N = 50 \times 50$ spins and $\gamma = 0$ and 5 from top to bottom at $T = 1.0$ above T_{KT} (left) and $T = 0.5$ below T_{KT} (right).



Why does the convergence seem to be fastened?

Operator form of the Fokker-Planck eq.

Let us consider the operator form of the Fokker-Planck equation as

$$\frac{\partial \tilde{P}(\mathbf{x}_1, \mathbf{x}_2, t)}{\partial t} = L(\{a_i\}, \{\hat{a}_i\})\tilde{P}(\mathbf{x}_1, \mathbf{x}_2, t) \tag{20}$$

where

$$L(\{a_i\}, \{\hat{a}_i\}) = -\sum_{i=1}^{2} \hat{a}_i a_i - \gamma \left(\hat{a}_1 a_2 - a_1 \hat{a}_2\right), \tag{21}$$

a_i and $\hat{a}_i$ are the operator satisfying $[a_i, \hat{a}_i] = -\mathrm{div}_i \mathrm{grad}_i E(\mathbf{x}_i)$, where $\hat{P}(\mathbf{x}_1, \mathbf{x}_2, t) = P(\mathbf{x}_1, \mathbf{x}_2, t)/\sqrt{P^{(\mathrm{ss})}(\mathbf{x}_1, \mathbf{x}_2)}$.

- The first eigenstate for $\sum_{i=1}^{2} \hat{a}_i a_i$ has a vanishing eigenvalue.
- Another term holds the first eigenstate with a vanishing eigenvalue
- When $\gamma = 0$, $L(\{a_i\}, \{\hat{a}_i\})$ is symmetric, otherwise asymmetric.

Rough sketch of proof

The eigenvalue of L characterizes the relaxation speed as $P(\mathbf{x}_1, \mathbf{x}_2, t) \sim \exp(\lambda^L t) P(\mathbf{x}_1, \mathbf{x}_2, 0)$.

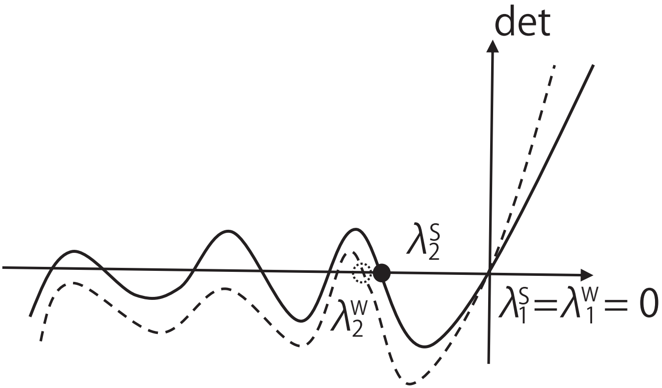
- Let us decompose $L = S + \Gamma$, where S is symmetric and Γ is anti-symmetric.
- We prove that $\mathrm{Re}(\lambda^L) - \lambda^S \leq 0$ for a fixed S.

Ostrowski-Taussky inequality

If $A + A^\dagger$ is positive-definite,

$$|\det A| \geq \det \frac{A + A^\dagger}{2} \tag{22}$$

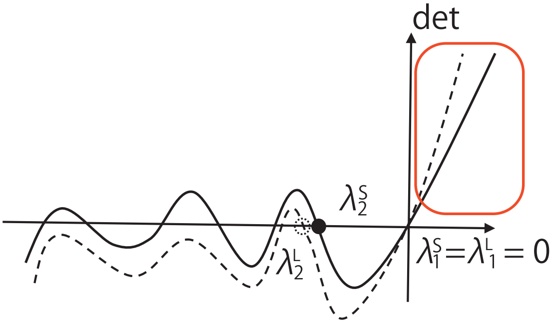
Generalization of $|z| \geq |\mathrm{Re} z|$ for a complex number z to matrices.



Slope of the characteristic polynomial

By use of the OT inequality, we find

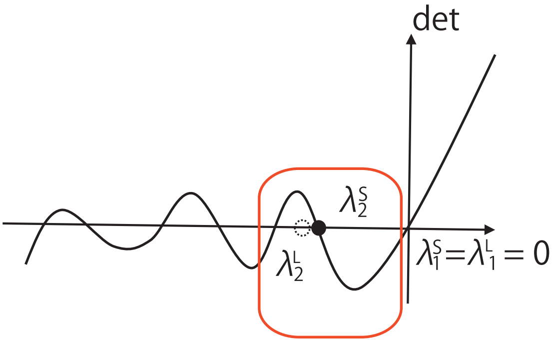
$$\left.\frac{d}{d\lambda}\det(\lambda - L)\right|_{\lambda=0} \geq \left.\frac{d}{d\lambda}\det(\lambda - S)\right|_{\lambda=0} \quad (23)$$



Slope of the characteristic polynomial

By use of the OT inequality, we find

$$\left.\frac{d}{d\lambda}\det(\lambda - L)\right|_{\lambda=0} \geq \left.\frac{d}{d\lambda}\det(\lambda - S)\right|_{\lambda=0} \tag{23}$$



$\lambda_2^L \leq \mathrm{Re}(\lambda) \leq \lambda_1^L = 0$

Division by λ removes the negative sign of the polynomial and OT inequality leads

$$\left| \frac{1}{\lambda} \det(\lambda - L) \right| \geq \frac{1}{\lambda} \det(\mathrm{Re}(\lambda) - S) \tag{24}$$

Conclusion

We propose an additional force to the Langevin dynamical system

- Replicated divergence-free system (Skewed DBC?)
  Implementation is very simple. Confirmed in XY model.
- Single divergence-free system (Suwa-Todo method?)
  Implementation is very simple. Not yet confirmed.

Mathematical assurance

- Nonzero eigenvalues are less than the ordinary Langevin dynamics.

Applications

- Glassy dynamics and molecule dynamics (Physics and Chemistry).
- Stochastic gradient method to infer the inherent parameter generating huge number of data (Machine Learning).