# Contrastive Divergence by Accelerated Langevin Dynamics

Masayuki Ohzeki

Kyoto University

2015/08/11

Japan-France Joint Seminar
"New Frontiers in Non-equilibrium Physics of Glassy Materials"
This work is in collaboration with
M. Yasuda (Yamagata Univ.) and A. Ichiki (Nagoya Univ.)

QUANTUM ANNEALING MACHINE LEARNING

What is the accelerated stochastic dynamics?

## Ordinary Langevin dynamics

The over-damped $N$-dimensional Langevin dynamics is given by

$$d\mathbf{x} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \sqrt{2T} d\mathbf{W},$$

where $T$ is the temperature and $\mathbf{W}$ is the Wiener process.

## Equilibrium distribution

The equilibrium state is

$$P_{\rm eq}(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{E(\mathbf{x})}{T}\right).$$

Why do you use this dynamics?

- Investigation of the probability distribution in the dynamics
- Simulation of the natural stochastic dynamics

## Ordinary Langevin dynamics

The over-damped $N$-dimensional Langevin dynamics is given by

$$d\mathbf{x} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \sqrt{2T}d\mathbf{W},$$

where $T$ is the temperature and $\mathbf{W}$ is the Wiener process.

## Equilibrium distribution

The equilibrium state is

$$P_{\rm eq}(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{x})}{T}\right).$$

Why do you use this dynamics?

- Investigation of the probability distribution in the dynamics
- Simulation of the natural stochastic dynamics

QUANTUM ANNEALING MACHINE LEARNING

## Ordinary Langevin dynamics

The over-damped $N$-dimensional Langevin dynamics is given by

$$d\mathbf{x} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \sqrt{2T}d\mathbf{W},$$

where $T$ is the temperature and $\mathbf{W}$ is the Wiener process.

## Equilibrium distribution

The equilibrium state is

$$P_{\mathrm{eq}}(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{x})}{T}\right).$$

Why do you use this dynamics?

- Investigation of the probability distribution in the dynamics
- Simulation of the natural stochastic dynamics

QUANTUM ANNEALING MACHINE LEARNING

## Ordinary Langevin dynamics

The over-damped $N$-dimensional Langevin dynamics is given by

$$d\mathbf{x} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \sqrt{2T} d\mathbf{W},$$

where $T$ is the temperature and $\mathbf{W}$ is the Wiener process.

## Equilibrium distribution

The equilibrium state is

$$P_{\mathrm{eq}}(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{E(\mathbf{x})}{T}\right).$$

Why do you use this dynamics?

- Investigation of the probability distribution in the dynamics
- Simulation of the natural stochastic dynamics

QUANTUM ANNEALING MACHINE LEARNING

In order to evaluate the distribution quickly,
we do not necessarily use the natural force

QUANTUM ANNEALING MACHINE LEARNING

## Accelerated Langevin dynamics

Let us find the accelerated Langevin dynamics with the simple form of

$$d\mathbf{x} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \sqrt{2T}d\mathbf{W},$$

where $T$ is the temperature and $d\mathbf{W}$ is the Wiener process.

## Condition

- The steady state has the Gibbs-Boltzmann distribution

$$P_{\text{ss}}(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{x})}{T}\right)$$

What force can hold the same steady state?

## Accelerated Langevin dynamics

Let us find the accelerated Langevin dynamics with the simple form of

$$d\mathbf{x} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \sqrt{2T}d\mathbf{W},$$

where $T$ is the temperature and $d\mathbf{W}$ is the Wiener process.

## Condition

- The steady state has the Gibbs-Boltzmann distribution

$$P_{\mathrm{ss}}(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{x})}{T}\right)$$

What force can hold the same steady state?

QUANTUM ANNEALING MACHINE LEARNING

## Accelerated Langevin dynamics

Let us find the accelerated Langevin dynamics with the simple form of

$$d\mathbf{x} = -\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \sqrt{2T}d\mathbf{W},$$

where $T$ is the temperature and $d\mathbf{W}$ is the Wiener process.

## Condition

- The steady state has the Gibbs-Boltzmann distribution

$$P_{\mathrm{ss}}(\mathbf{x}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{x})}{T}\right)$$

What force can hold the same steady state?

## Nontrivial force [M.Ohzeki and A. Ichiki (2015)]

Find solution of the Fokker-Planck equation

$$\frac{\partial P_t(\mathbf{x})}{\partial t} = -\frac{\partial}{\partial \mathbf{x}}\left(-\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) - T\frac{\partial}{\partial \mathbf{x}}\right) P_t(\mathbf{x})$$

The condition is reduced to

$$0 = -\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{F}(\mathbf{x}) P_{\rm ss}(\mathbf{x})\right)$$

- Equilibrium force $\mathbf{F}(\mathbf{x}) = \mathbf{0}$
- Exponential force $\mathbf{F}(\mathbf{x}) \propto \gamma \exp\left(E(\mathbf{x})/T\right)$
- Rotational force

$$[\mathbf{F}(\mathbf{x})]_{P(i)} = \gamma\left(\left[\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\right]_{P(i-1)} - \left[\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\right]_{P(i+1)}\right)$$

where $P(i)$ is the permutation of indices.

## Nontrivial force [M.Ohzeki and A. Ichiki (2015)]

Find solution of the Fokker-Planck equation

$$\frac{\partial P_t(\mathbf{x})}{\partial t} = -\frac{\partial}{\partial \mathbf{x}}\left(-\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) - T\frac{\partial}{\partial \mathbf{x}}\right)P_t(\mathbf{x})$$

The condition is reduced to

$$0 = -\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{F}(\mathbf{x})P_{\rm ss}(\mathbf{x})\right)$$

- Equilibrium force $\mathbf{F}(\mathbf{x}) = \mathbf{0}$
- Exponential force $\mathbf{F}(\mathbf{x}) \propto \boldsymbol{\gamma}\exp\left(E(\mathbf{x})/T\right)$
- Rotational force

$$[\mathbf{F}(\mathbf{x})]_{P(i)} = \gamma\left(\left[\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\right]_{P(i-1)} - \left[\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\right]_{P(i+1)}\right)$$

where $P(i)$ is the permutation of indices.

QUANTUM ANNEALING MACHINE LEARNING

## Nontrivial force [M.Ohzeki and A. Ichiki (2015)]

Find solution of the Fokker-Planck equation

$$\frac{\partial P_t(\mathbf{x})}{\partial t} = -\frac{\partial}{\partial \mathbf{x}}\left(-\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) - T\frac{\partial}{\partial \mathbf{x}}\right)P_t(\mathbf{x})$$

The condition is reduced to

$$0 = -\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{F}(\mathbf{x})P_{\mathrm{ss}}(\mathbf{x})\right)$$

- Equilibrium force $\mathbf{F}(\mathbf{x}) = \mathbf{0}$
- Exponential force $\mathbf{F}(\mathbf{x}) \propto \boldsymbol{\gamma}\exp\left(E(\mathbf{x})/T\right)$
- Rotational force

$$[\mathbf{F}(\mathbf{x})]_{P(i)} = \gamma\left(\left[\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\right]_{P(i-1)} - \left[\frac{\partial E(\mathbf{x})}{\partial \mathbf{x}}\right]_{P(i+1)}\right)$$

where $P(i)$ is the permutation of indices.

QUANTUM ANNEALING MACHINE LEARNING

## Nontrivial force in duplicate system [M.Ohzeki and A. Ichiki (2015)]

Find solution of the Fokker-Planck equation for a duplicate system

$$
\begin{aligned}
\frac{\partial P_t(\mathbf{x}_1,\mathbf{x}_2)}{\partial t} &= -\frac{\partial}{\partial \mathbf{x}_1}\left(-\frac{\partial E(\mathbf{x}_1)}{\partial \mathbf{x}_1} + \mathbf{F}_1(\mathbf{x}_1,\mathbf{x}_2) - T\frac{\partial}{\partial \mathbf{x}_1}\right) P_t(\mathbf{x}_1,\mathbf{x}_2) \\
&\quad -\frac{\partial}{\partial \mathbf{x}_2}\left(-\frac{\partial E(\mathbf{x}_2)}{\partial \mathbf{x}_2} + \mathbf{F}_2(\mathbf{x}_1,\mathbf{x}_2) - T\frac{\partial}{\partial \mathbf{x}_2}\right) P_t(\mathbf{x}_1,\mathbf{x}_2)
\end{aligned}
$$

The condition is reduced to

$$
0 = -\frac{\partial}{\partial \mathbf{x}_1}\left(\mathbf{F}_1(\mathbf{x}_1,\mathbf{x}_2) P_{\mathrm{ss}}(\mathbf{x}_1) P_{\mathrm{ss}}(\mathbf{x}_2)\right) - \frac{\partial}{\partial \mathbf{x}_2}\left(\mathbf{F}_1(\mathbf{x}_1,\mathbf{x}_2) P_{\mathrm{ss}}(\mathbf{x}_1) P_{\mathrm{ss}}(\mathbf{x}_2)\right)
$$

- Nontrivial force in the duplicate system

$$
\begin{aligned}
\mathbf{F}_1(\mathbf{x}_1,\mathbf{x}_2) &= \gamma \frac{\partial E(\mathbf{x}_2)}{\partial \mathbf{x}_2} \\
\mathbf{F}_2(\mathbf{x}_1,\mathbf{x}_2) &= -\gamma \frac{\partial E(\mathbf{x}_1)}{\partial \mathbf{x}_1}.
\end{aligned}
$$

QUANTUM ANNEALING MACHINE LEARNING

## What does the nontrivial force yield?

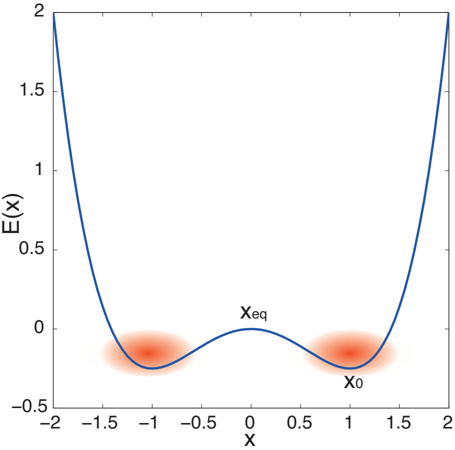
- Violation of the detailed balance condition ($\gamma \neq 0$)
- Convergence to nonequilibrium steady state
- Faster convergence than equilibrium system
  - in analytical way by matrix analysis [A. Ichiki and M. Ohzeki (2013)]

QUANTUM ANNEALING MACHINE LEARNING

## Example: double-valley potential [M. Ohzeki and A. Ichiki (2015)]

We set $N = 1000$ particles in a double-valley potential

$$E(x) = -\frac{1}{2}x^2 + \frac{1}{4}x^4$$



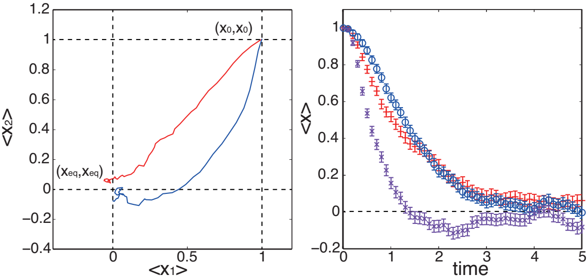

QUANTUM ANNEALING MACHINE LEARNING

## Example: double-valley potential [M. Ohzeki and A. Ichiki (2015)]

We set $N = 1000$ particles in a double-valley potential

$$E(x) = -\frac{1}{2}x^2 + \frac{1}{4}x^4$$

at $t = 5$ in $T = 1$. $\gamma = 0$ (red) vs $\gamma = 1$ (blue and purple).

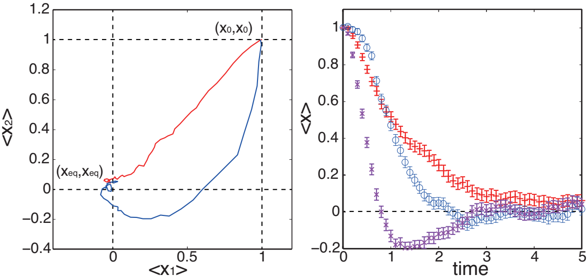


## Example: double-valley potential [M. Ohzeki and A. Ichiki (2015)]

We set $N = 1000$ particles in a double-valley potential

$$E(x) = -\frac{1}{2}x^2 + \frac{1}{4}x^4$$

at $t = 5$ in $T = 1$. $\gamma = 0$ (red) vs $\gamma = 2$ (blue and purple).

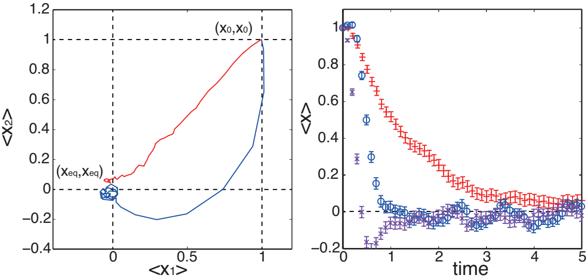


## Example: double-valley potential [M. Ohzeki and A. Ichiki (2015)]

We set $N = 1000$ particles in a double-valley potential

$$E(x) = -\frac{1}{2}x^2 + \frac{1}{4}x^4$$

at $t = 5$ in $T = 1$. $\gamma = 0$ (red) vs $\gamma = 5$ (blue and purple).

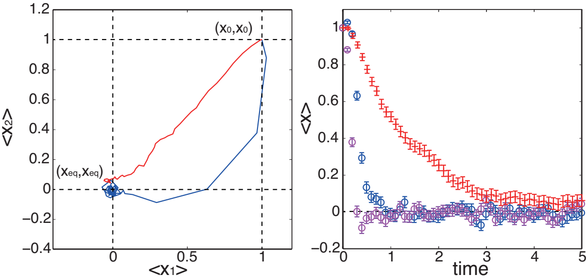


## Example: double-valley potential [M. Ohzeki and A. Ichiki (2015)]

We set $N = 1000$ particles in a double-valley potential

$$E(x) = -\frac{1}{2}x^2 + \frac{1}{4}x^4$$

at $t = 5$ in $T = 1$. $\gamma = 0$ (red) vs $\gamma = 10$ (blue and purple).

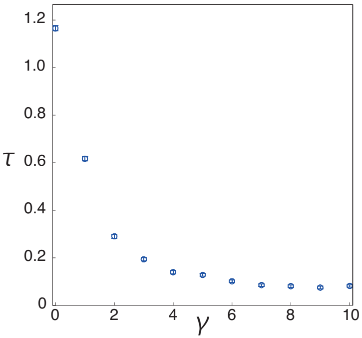


## Example: double-valley potential [M. Ohzeki and A. Ichiki (2015)]

We set $N = 1000$ particles in a double-valley potential

$$E(x) = -\frac{1}{2}x^2 + \frac{1}{4}x^4$$

We confirm reduction of correlation time of x by $\tau = \sum_{t=1}^{\infty} \frac{\langle O_i O_{i+t}\rangle - \langle O_i\rangle^2}{\langle O_i^2\rangle - \langle O_i\rangle^2}$



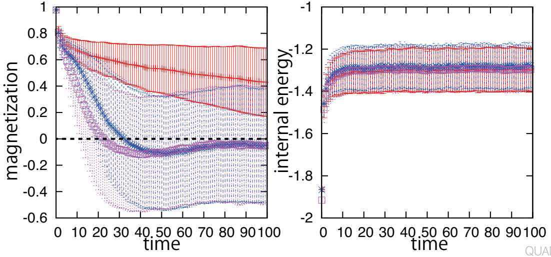
## Example: XY model [M. Ohzeki and A. Ichiki (2015)]

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

$$E(\mathbf{x}) = -\sum_{i=1}\sum_{j\in\partial i}\cos\left(x_i - x_j\right),$$

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

We set $N = 10 \times 10$ spins of independent $N = 1000$ runs and $\gamma = 0$ (Red) and 10 (Blue and Purple) at $T = 0.5$ below $T_{\mathrm{KT}}$.




QUANTUM ANNEALING MACHINE LEARNING

Other accelerated stochastic dynamics

- in MCMC by Suwa-Todo method (optimization of transition matrix) [H. Suwa and S. Todo (2010)]
- in MCMC by Skewed DBC (global flow in a duplicate system) [Y. Sakai and K. Hukushima (2013)]
- in analytical way by optimization of master equation (Brachistochrone) [K. Takahashi and M. Ohzeki, to be submitted]

Our study

Nonequilibrium physics $\longrightarrow$ Machine learning

What is Boltzmann machine learning

QUANTUM ANNEALING MACHINE LEARNING

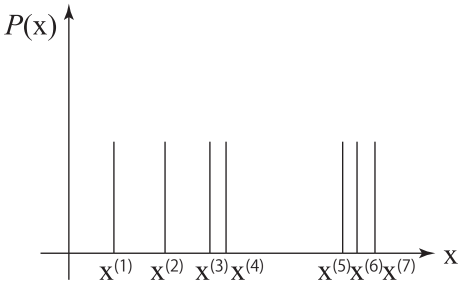
## Aim

- Clarify a generative model of the given high-dimensional data $\mathbf{x}^{(d)} \in \mathbb{R}^N (d = 1, 2, \cdots, D)$

Maximum Likelihood Estimation:

- Learning model

$$P(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(-E(\mathbf{x}|\boldsymbol{\theta})\right)$$



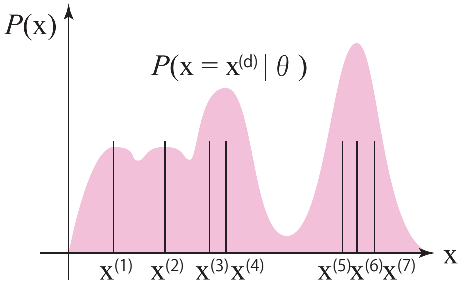
## Aim

- Clarify a generative model of the given high-dimensional data $\mathbf{x}^{(d)} \in \mathbb{R}^N (d = 1, 2, \cdots, D)$

Maximum Likelihood Estimation:

- Learning model

$$P(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(-E(\mathbf{x}|\boldsymbol{\theta})\right)$$



How to perform the maximum likelihood estimation?

- Compute logarithm of likelihood function

$$L_D(\boldsymbol{\theta}) = \frac{1}{D}\sum_{d=1}^{D}\log P(\mathbf{x} = \mathbf{x}^{(d)}|\boldsymbol{\theta})$$

- Use gradient method

$$\frac{\partial L_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{D}\sum_{d=1}^{D}\frac{\partial E(\mathbf{x} = \mathbf{x}^{(d)}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \left\langle \frac{\partial E(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}}$$

  - first term =empirical mean of data
  - second term= thermal average of model $\langle \cdots \rangle_{\boldsymbol{\theta}} = \sum_{\mathbf{x}} P(\mathbf{x}|\boldsymbol{\theta})\times$
- Iterative update to achieve the maximum

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^{t} + \eta\frac{\partial L_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

How to perform the maximum likelihood estimation?

- Compute logarithm of likelihood function

$$L_D(\boldsymbol{\theta}) = \frac{1}{D}\sum_{d=1}^{D} \log P(\mathbf{x} = \mathbf{x}^{(d)}|\boldsymbol{\theta})$$

- Use gradient method

$$\frac{\partial L_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{D}\sum_{d=1}^{D} \frac{\partial E(\mathbf{x} = \mathbf{x}^{(d)}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \left\langle \frac{\partial E(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}}$$

  - first term =empirical mean of data
  - second term= thermal average of model $\langle \cdots \rangle_{\boldsymbol{\theta}} = \sum_{\mathbf{x}} P(\mathbf{x}|\boldsymbol{\theta})\times$

- Iterative update to achieve the maximum

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^{t} + \eta \frac{\partial L_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

QUANTUM ANNEALING MACHINE LEARNING

How to perform the maximum likelihood estimation?

- Compute logarithm of likelihood function

$$L_D(\boldsymbol{\theta}) = \frac{1}{D}\sum_{d=1}^{D}\log P(\mathbf{x}=\mathbf{x}^{(d)}|\boldsymbol{\theta})$$

- Use gradient method

$$\frac{\partial L_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{D}\sum_{d=1}^{D}\frac{\partial E(\mathbf{x}=\mathbf{x}^{(d)}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \left\langle \frac{\partial E(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}}$$

  - first term =empirical mean of data
  - second term= thermal average of model $\langle \cdots \rangle_{\boldsymbol{\theta}} = \sum_{\mathbf{x}} P(\mathbf{x}|\boldsymbol{\theta})\times$
- Iterative update to achieve the maximum

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^{t} + \eta\frac{\partial L_D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

## How to evaluate thermal average

### Approximation or Monte-Carlo simulation

- Markov-Chain Monte-Carlo method

$$\mathbf{x}^{t=0} \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=1} \xrightarrow[\text{MCMC}]{} \cdots \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=T}$$

Slow but asymptotically exact in $T \to \infty$

- Contrastive divergence

$$\mathbf{x}^{(d)} \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=1} \xrightarrow[\text{MCMC}]{} \cdots \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=T}$$

Early stop! but good performance

- Pseudo likelihood estimation, etc
  Asymptotically exact in $D \to \infty$, and less flexibility

QUANTUM ANNEALING MACHINE LEARNING

# How to evaluate thermal average

## Approximation or Monte-Carlo simulation

- Markov-Chain Monte-Carlo method

$$\mathbf{x}^{t=0} \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=1} \xrightarrow[\text{MCMC}]{} \cdots \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=T}$$

  Slow but asymptotically exact in $T \to \infty$

- Contrastive divergence

$$\mathbf{x}^{(d)} \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=1} \xrightarrow[\text{MCMC}]{} \cdots \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=T}$$

  Early stop! but good performance

- Pseudo likelihood estimation, etc
  Asymptotically exact in $D \to \infty$, and less flexibility

QUANTUM ANNEALING MACHINE LEARNING

## How to evaluate thermal average

### Approximation or Monte-Carlo simulation

- Markov-Chain Monte-Carlo method

$$\mathbf{x}^{t=0} \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=1} \xrightarrow[\text{MCMC}]{} \cdots \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=T}$$

  Slow but asymptotically exact in $T \to \infty$

- Contrastive divergence

$$\color{red}{\mathbf{x}^{(d)}} \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=1} \xrightarrow[\text{MCMC}]{} \cdots \xrightarrow[\text{MCMC}]{} \mathbf{x}^{t=T}$$

  Early stop! but good performance

- Pseudo likelihood estimation, etc
  Asymptotically exact in $D \to \infty$, and less flexibility

## Our present study

Let us implement the accelerated Langevin dynamics to

Contrastive divergence

The speedup of the contrastive divergence is essential in machine learning

QUANTUM ANNEALING MACHINE LEARNING

## Our present study

Let us implement the accelerated Langevin dynamics to

Contrastive divergence

The speedup of the contrastive divergence is essential in machine learning

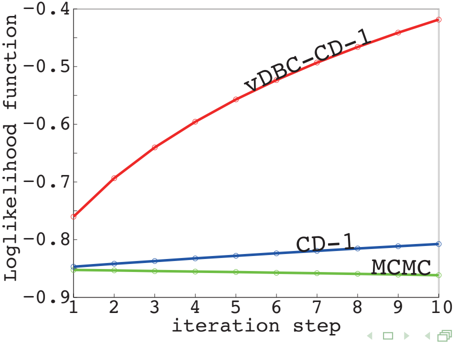
# Preliminary result: Simple Gaussian distribution

We assume that the generative model is

$$P(\mathbf{x}|\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}J\mathbf{x} - \mathbf{h}^{\mathrm{T}}\mathbf{x}\right)$$

We have $D = 1000$ data points to infer the original $J$ and $\mathbf{h}$.

A test in extremely small system $N = 1$. We use $\gamma = 5$.
CD-1 step is defined as the integration time $t = 1$ $(dt = 0.01)$.




QUANTUM ANNEALING MACHINE LEARNING

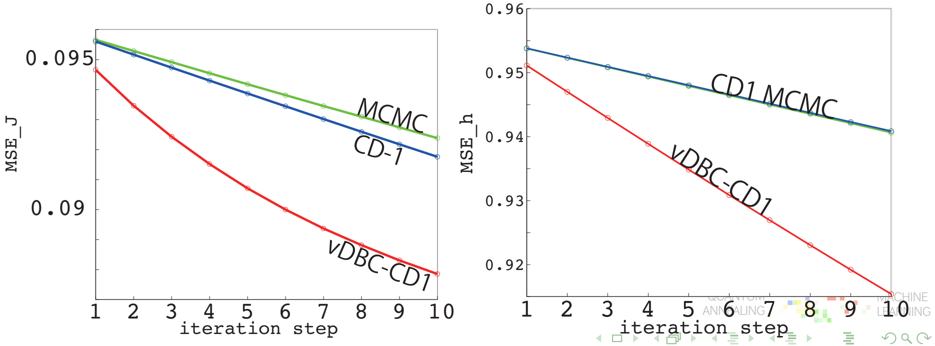
## Preliminary result: Simple Gaussian distribution

We assume that the generative model is

$$P(\mathbf{x}|\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}J\mathbf{x} - \mathbf{h}^{\mathrm{T}}\mathbf{x}\right)$$

We have $D = 1000$ data points to infer the original $J$ and $\mathbf{h}$.

A test in extremely small system $N = 1$. We use $\gamma = 5$.
CD-1 step is defined as the integration time $t = 1$ $(dt = 0.01)$.



## Summary of our present study

We implement the accelerated stochastic dynamics to

Contrastive divergence

- Utilization of violation of the detailed balance condition
- Confirm its efficiency in terms of the log-likelihood function

The speedup of the contrastive divergence is essential in machine learning

QUANTUM ANNEALING MACHINE LEARNING

## Summary of our present study

We implement the accelerated stochastic dynamics to

Contrastive divergence

- Utilization of violation of the detailed balance condition
- Confirm its efficiency in terms of the log-likelihood function

The speedup of the contrastive divergence is essential in machine learning