Analysis and optimization of certain parallel Monte Carlo methods in the low temperature limit

Guo-Jhen Wu* and Paul Dupuis **

RESIM 2021, May 19, 2021

* Department of Mathematics, KTH Royal Institute of Technology

** Division of Applied Mathematics, Brown University

Outline

1. Problem of interest

2. Metastability and accelerated Monte Carlo

3. Performance measure

4. Large deviation properties of empirical measure of metastable diffusion

5. Optimality in two-well model and multi-well model

Problem of interest

RARE EVENT PROBABILITIES AND MCMC

Compute probability $\mu^\varepsilon(A)$ with respect to a Gibbs measure of the form

$$\mu^\varepsilon(dx) = e^{-V(x)/\varepsilon}dx \Big/ Z(\varepsilon),$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is the potential of a complex physical system, ε is the temperature of the system, and A does not contain the global minimum of V.

RARE EVENT PROBABILITIES AND MCMC

Compute probability $\mu^\varepsilon(A)$ with respect to a Gibbs measure of the form

$$\mu^\varepsilon(dx) = e^{-V(x)/\varepsilon}dx \Big/ Z(\varepsilon),$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is the potential of a complex physical system, ε is the temperature of the system, and A does not contain the global minimum of V.

Well-known: $\mu^\varepsilon(dx)$ is the unique invariant distribution of the diffusion process $\{X(t)\}_t$ satisfying

$$dX(t) = -\nabla V(X(t))\, dt + \sqrt{2\varepsilon} dW(t).$$

Rare event probabilities and MCMC

Compute probability $\mu^{\varepsilon}(A)$ with respect to a Gibbs measure of the form

$$\mu^{\varepsilon}(dx) = e^{-V(x)/\varepsilon}dx \Big/ Z(\varepsilon),$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is the potential of a complex physical system, ε is the temperature of the system, and A does not contain the global minimum of V.

Well-known: $\mu^{\varepsilon}(dx)$ is the unique invariant distribution of the diffusion process $\{X(t)\}_t$ satisfying

$$dX(t) = -\nabla V(X(t))\,dt + \sqrt{2\varepsilon}dW(t).$$

Markov Chain Monte Carlo (MCMC)

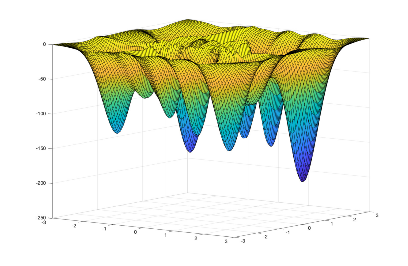
The empirical measure over a large time T:

$$\lambda^T(dx) = \frac{1}{T}\int_0^T \delta_{X(t)}(dx)dt \in \mathcal{P}(\mathbb{R}^d).$$

Use $\lambda^T(A)$ for some large T as an estimate of $\mu^{\varepsilon}(A)$.

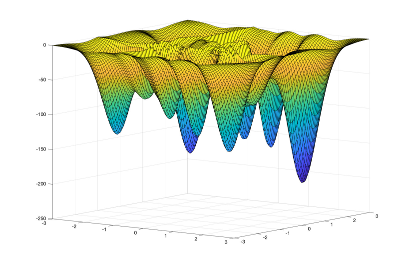
Metastability

EXPONENTIAL EXIT TIME



In general V contains many deep and shallow local minima.

Exponential exit time



In general V contains many deep and shallow local minima.

Exponential exit time: Mean transition time from one local minimum to another is roughly $\exp(h/\varepsilon)$ when the temperature ε is small, where h is the barrier height.

Parallel tempering (two temperatures)

Besides $\varepsilon_1 = \varepsilon$, introduce higher temperature $\varepsilon_2 = \varepsilon/\alpha$ with $\alpha \in (0,1)$.

$$dX_1 = -\nabla V(X_1)dt + \sqrt{2\varepsilon_1}dW_1$$
$$dX_2 = -\nabla V(X_2)dt + \sqrt{2\varepsilon_2}dW_2,$$

with W_1 and W_2 independent. Then allow "swaps" with rate

$$ag(x_1,x_2) = a\left(1 \wedge e^{-\left[\frac{V(x_1)}{\varepsilon_1}+\frac{V(x_2)}{\varepsilon_2}\right]+\left[\frac{V(x_2)}{\varepsilon_1}+\frac{V(x_1)}{\varepsilon_2}\right]}\right).$$

Parallel tempering (two temperatures)

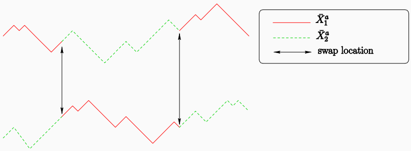
Besides $\varepsilon_1 = \varepsilon$, introduce higher temperature $\varepsilon_2 = \varepsilon/\alpha$ with $\alpha \in (0,1)$.

$$dX_1 = -\nabla V(X_1)dt + \sqrt{2\varepsilon_1}dW_1$$
$$dX_2 = -\nabla V(X_2)dt + \sqrt{2\varepsilon_2}dW_2,$$

with W_1 and W_2 independent. Then allow "swaps" with rate

$$ag(x_1,x_2) = a\left(1 \wedge e^{-\left[\frac{V(x_1)}{\varepsilon_1}+\frac{V(x_2)}{\varepsilon_2}\right]+\left[\frac{V(x_2)}{\varepsilon_1}+\frac{V(x_1)}{\varepsilon_2}\right]}\right).$$

Particle swapped process: (X_1^a, X_2^a)



$\mu^{\varepsilon_1}(dx_1)\mu^{\varepsilon_2}(dx_2)$ is the unique invariant distribution of (X_1^a, X_2^a).

INFINITE SWAPPING PROCESS (TWO TEMPERATURES)

INS process *(limit process as swap rate $a \to \infty$)*:

$$dY_1 = -\nabla V(Y_1)dt + \sqrt{2\varepsilon_1 \rho(Y_1, Y_2) + 2\varepsilon_2 \rho(Y_2, Y_1)}dW_1$$
$$dY_2 = -\nabla V(Y_2)dt + \sqrt{2\varepsilon_2 \rho(Y_1, Y_2) + 2\varepsilon_1 \rho(Y_2, Y_1)}dW_2,$$

where

$$\rho(x_1, x_2) = e^{-\left[\frac{V(x_1)}{\varepsilon_1} + \frac{V(x_2)}{\varepsilon_2}\right]} \Big/ Z_\rho(x_1, x_2),$$
$$Z_\rho(x_1, x_2) = e^{-\left[\frac{V(x_1)}{\varepsilon_1} + \frac{V(x_2)}{\varepsilon_2}\right]} + e^{-\left[\frac{V(x_2)}{\varepsilon_1} + \frac{V(x_1)}{\varepsilon_2}\right]}.$$

The unique invariant distribution of (Y_1, Y_2) becomes $[\mu^{\varepsilon_1}(dx_1)\mu^{\varepsilon_2}(dx_2) + \mu^{\varepsilon_2}(dx_1)\mu^{\varepsilon_1}(dx_2)]/2$.

Weighted empirical measure:

$$\eta^T(dx) = \frac{1}{T}\int_0^T \left[\rho(Y_1, Y_2)\delta_{(Y_1,Y_2)} + \rho(Y_2, Y_1)\delta_{(Y_2,Y_1)}\right] dt,$$

Use $\eta^T(A \times \mathbb{R}^d)$ as an estimate of $\mu^\varepsilon(A)$.

K-TEMPERATURE INS ALGORITHM

K-temperature INS process $\{\boldsymbol{X}^{\varepsilon}(t)\}_{t\geq 0} = \{(X_1^{\varepsilon}(t), \ldots, X_K^{\varepsilon}(t))\}_{t\geq 0}$ for a given $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ with $1 = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_K$:

$$\begin{cases} dX_1^{\varepsilon} = -\nabla V\left(X_1^{\varepsilon}\right) dt + \sqrt{2\varepsilon}\sqrt{\rho_{11}^{\varepsilon}/\alpha_1 + \rho_{12}^{\varepsilon}/\alpha_2 + \cdots + \rho_{1K}^{\varepsilon}/\alpha_K} dW_1 \\ \vdots \\ dX_K^{\varepsilon} = -\nabla V\left(X_K^{\varepsilon}\right) dt + \sqrt{2\varepsilon}\sqrt{\rho_{K1}^{\varepsilon}/\alpha_1 + \rho_{K2}^{\varepsilon}/\alpha_2 + \cdots + \rho_{KK}^{\varepsilon}/\alpha_K} dW_K \end{cases},$$

where

$$\rho_{ij}^{\varepsilon} \doteq \sum_{\sigma:\sigma(j)=i} w^{\varepsilon}\left(\boldsymbol{x}_{\sigma}; \boldsymbol{\alpha}\right), \quad w^{\varepsilon}\left(\boldsymbol{x}; \boldsymbol{\alpha}\right) \doteq \frac{\exp[-\frac{1}{\varepsilon}\sum_{\ell=1}^{K}\alpha_\ell V\left(x_\ell\right)]}{\sum_{\sigma\in\Sigma_K}\exp[-\frac{1}{\varepsilon}\sum_{\ell=1}^{K}\alpha_\ell V(x_{\sigma(\ell)})]}.$$

INS estimator of $\mu^{\varepsilon}(A)$ is defined as

$$\theta_{\text{INS}}^{\varepsilon,T} \doteq \frac{1}{T}\int_0^T \sum_{\sigma\in\Sigma_K} w^{\varepsilon}\left(\boldsymbol{X}_{\sigma}^{\varepsilon}\left(t\right); \boldsymbol{\alpha}\right) 1_A(X_{\sigma(1)}^{\varepsilon}(t))dt.$$

K-TEMPERATURE INS ALGORITHM

K-temperature INS process $\{\boldsymbol{X}^{\varepsilon}(t)\}_{t\geq 0} = \{(X_1^{\varepsilon}(t), \ldots, X_K^{\varepsilon}(t))\}_{t\geq 0}$ for a given $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ with $1 = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_K$:

$$\begin{cases} dX_1^{\varepsilon} = -\nabla V\left(X_1^{\varepsilon}\right) dt + \sqrt{2\varepsilon}\sqrt{\rho_{11}^{\varepsilon}/\alpha_1 + \rho_{12}^{\varepsilon}/\alpha_2 + \cdots + \rho_{1K}^{\varepsilon}/\alpha_K} dW_1 \\ \vdots \\ dX_K^{\varepsilon} = -\nabla V\left(X_K^{\varepsilon}\right) dt + \sqrt{2\varepsilon}\sqrt{\rho_{K1}^{\varepsilon}/\alpha_1 + \rho_{K2}^{\varepsilon}/\alpha_2 + \cdots + \rho_{KK}^{\varepsilon}/\alpha_K} dW_K \end{cases},$$

where

$$\rho_{ij}^{\varepsilon} \doteq \sum_{\sigma:\sigma(j)=i} w^{\varepsilon}\left(\boldsymbol{x}_{\sigma};\boldsymbol{\alpha}\right), \quad w^{\varepsilon}\left(\boldsymbol{x};\boldsymbol{\alpha}\right) \doteq \frac{\exp[-\frac{1}{\varepsilon}\sum_{\ell=1}^{K}\alpha_{\ell}V\left(x_{\ell}\right)]}{\sum_{\sigma\in\Sigma_K}\exp[-\frac{1}{\varepsilon}\sum_{\ell=1}^{K}\alpha_{\ell}V(x_{\sigma(\ell)})]}.$$

INS estimator of $\mu^{\varepsilon}(A)$ is defined as

$$\theta_{\text{INS}}^{\varepsilon,T} \doteq \frac{1}{T}\int_0^T \sum_{\sigma\in\Sigma_K} w^{\varepsilon}\left(\boldsymbol{X}_{\sigma}^{\varepsilon}\left(t\right);\boldsymbol{\alpha}\right) 1_A(X_{\sigma(1)}^{\varepsilon}(t))dt.$$

Question: How to choose $\boldsymbol{\alpha}$?

Performance measure

UNBIASEDNESS AND DECAY RATE OF VARIANCE

Time scale: Good estimation requires $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > 0$.

DEFINITION

An estimator $\theta^{\varepsilon,T^\varepsilon}$ of $\mu^\varepsilon(A)$ is called **essentially unbiased** if there is $c_0 \in (0,\infty)$ such that

$$\liminf_{\varepsilon\to 0} -\varepsilon \log \left| E\theta^{\varepsilon,T^\varepsilon} - \mu^\varepsilon(A) \right| \geq \lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(A) + c_0.$$

DEFINITION

The decay rate of the variance (per unit time) of $\theta^{\varepsilon,T^\varepsilon}$ is defined as

$$\lim_{\varepsilon\to 0} -\varepsilon \log \left(\mathrm{Var}\left(\theta^{\varepsilon,T^\varepsilon}\right) T^\varepsilon \right).$$

UNBIASEDNESS AND DECAY RATE OF VARIANCE

Time scale: Good estimation requires $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > 0$.

DEFINITION

An estimator $\theta^{\varepsilon,T^\varepsilon}$ of $\mu^\varepsilon(A)$ is called **essentially unbiased** if there is $c_0 \in (0,\infty)$ such that

$$\liminf_{\varepsilon\to 0} -\varepsilon \log \left| E\theta^{\varepsilon,T^\varepsilon} - \mu^\varepsilon(A) \right| \geq \lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(A) + c_0.$$

DEFINITION

The decay rate of the variance (per unit time) of $\theta^{\varepsilon,T^\varepsilon}$ is defined as

$$\lim_{\varepsilon\to 0} -\varepsilon \log \left(\operatorname{Var}\left(\theta^{\varepsilon,T^\varepsilon}\right) T^\varepsilon \right).$$

- Performance benchmark is $2\lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(A)$.

UNBIASEDNESS AND DECAY RATE OF VARIANCE

Time scale: Good estimation requires $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > 0$.

DEFINITION

An estimator $\theta^{\varepsilon,T^\varepsilon}$ of $\mu^\varepsilon(A)$ is called **essentially unbiased** if there is $c_0 \in (0,\infty)$ such that

$$\liminf_{\varepsilon\to 0} -\varepsilon \log \left| E\theta^{\varepsilon,T^\varepsilon} - \mu^\varepsilon(A) \right| \geq \lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(A) + c_0.$$

DEFINITION

The decay rate of the variance (per unit time) of $\theta^{\varepsilon,T^\varepsilon}$ is defined as

$$\lim_{\varepsilon\to 0} -\varepsilon \log \left(\mathrm{Var}\left(\theta^{\varepsilon,T^\varepsilon}\right) T^\varepsilon \right).$$

- Performance benchmark is $2\lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(A)$.
- Not the best possible decay rate, but the best practically achievable decay rate.

UNBIASEDNESS AND DECAY RATE OF VARIANCE

Time scale: Good estimation requires $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > 0$.

DEFINITION

An estimator $\theta^{\varepsilon,T^\varepsilon}$ of $\mu^\varepsilon(A)$ is called **essentially unbiased** if there is $c_0 \in (0,\infty)$ such that

$$\liminf_{\varepsilon\to 0} -\varepsilon \log \left| E\theta^{\varepsilon,T^\varepsilon} - \mu^\varepsilon(A) \right| \geq \lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(A) + c_0.$$

DEFINITION

The decay rate of the variance (per unit time) of $\theta^{\varepsilon,T^\varepsilon}$ is defined as

$$\lim_{\varepsilon\to 0} -\varepsilon \log \left(\mathrm{Var}\left(\theta^{\varepsilon,T^\varepsilon}\right) T^\varepsilon \right).$$

- Performance benchmark is $2\lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(A)$.
- Not the best possible decay rate, but the best practically achievable decay rate.
- Optimize decay rate among essentially unbiased estimators.

UNBIASEDNESS AND DECAY RATE OF VARIANCE

Time scale: Good estimation requires $T^{\varepsilon} = e^{\frac{1}{\varepsilon}c}$ for some $c > 0$.

DEFINITION

An estimator $\theta^{\varepsilon,T^{\varepsilon}}$ of $\mu^{\varepsilon}(A)$ is called **essentially unbiased** if there is $c_0 \in (0,\infty)$ such that

$$\liminf_{\varepsilon\to 0} -\varepsilon \log \left| E\theta^{\varepsilon,T^{\varepsilon}} - \mu^{\varepsilon}(A) \right| \geq \lim_{\varepsilon\to 0} -\varepsilon \log \mu^{\varepsilon}(A) + c_0.$$

DEFINITION

The decay rate of the variance (per unit time) of $\theta^{\varepsilon,T^{\varepsilon}}$ is defined as

$$\lim_{\varepsilon\to 0} -\varepsilon \log \left(\mathrm{Var}\left(\theta^{\varepsilon,T^{\varepsilon}}\right) T^{\varepsilon} \right).$$

- Performance benchmark is $2\lim_{\varepsilon\to 0} -\varepsilon \log \mu^{\varepsilon}(A)$.
- Not the best possible decay rate, but the best practically achievable decay rate.
- Optimize decay rate among essentially unbiased estimators.
- Conflict between improving the decay rate and achieving essential unbiasedness is insignificant.

LD properties

Small-noise diffusion and quasipotential

Consider $\{X_t^\varepsilon\}_{0\le t\le T}$ satisfies

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x.$$

Let $\{O_i\}_{i\in L}$ be all the equilibrium points of $\dot{x}_t = b(x_t)$ and $\{X_t^\varepsilon\}$ has an unique invariant distribution μ^ε satisfying

$$\lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(O_1) < \lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(O_i).$$

Under some conditions, $\{X_t^\varepsilon\}$ satisfies a large deviation principle with rate function I_T for any $T\in(0,\infty)$.

SMALL-NOISE DIFFUSION AND QUASIPOTENTIAL

Consider $\{X_t^\varepsilon\}_{0\le t\le T}$ satisfies

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x.$$

Let $\{O_i\}_{i\in L}$ be all the equilibrium points of $\dot{x}_t = b(x_t)$ and $\{X_t^\varepsilon\}$ has an unique invariant distribution μ^ε satisfying

$$\lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(O_1) < \lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(O_i).$$

Under some conditions, $\{X_t^\varepsilon\}$ satisfies a large deviation principle with rate function I_T for any $T \in (0,\infty)$. The quasipotential is as

$$Q(x,y) \doteq \inf\left\{I_T(\phi) : \phi(0) = x, \phi(T) = y, T < \infty\right\}.$$

Small-noise diffusion and quasipotential

Consider $\{X_t^\varepsilon\}_{0\le t\le T}$ satisfies

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x.$$

Let $\{O_i\}_{i\in L}$ be all the equilibrium points of $\dot{x}_t = b(x_t)$ and $\{X_t^\varepsilon\}$ has an unique invariant distribution μ^ε satisfying

$$\lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(O_1) < \lim_{\varepsilon\to 0} -\varepsilon \log \mu^\varepsilon(O_i).$$

Under some conditions, $\{X_t^\varepsilon\}$ satisfies a large deviation principle with rate function I_T for any $T \in (0,\infty)$. The quasipotential is as

$$Q(x,y) \doteq \inf\left\{I_T(\phi) : \phi(0) = x, \phi(T) = y, T < \infty\right\}.$$

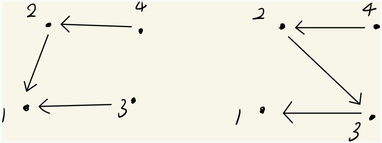
DEFINITION

Given a subset $W \subset L$, a directed graph consisting of arrows $i \to j$ $(i \in L\setminus W, j \in L, i \neq j)$ is called a W**-graph on** L if

1. every point $i \in L \setminus W$ is the initial point of exactly one arrow.
2. for any point $i \in L \setminus W$, there exists a sequence of arrows leading from i to some point in W.

W-GRAPHS

Example: $L = \{1, 2, 3, 4\}$ and $W = \{1\}$.



Denote the set of all W-graphs by $G(W)$.

DEFINITION

For all $i \in L$,

$$W(O_i) \doteq \min_{g \in G(i)} \left[\sum_{(m \to n) \in g} Q(O_m, O_n) \right]$$

and

$$W(O_1 \cup O_i) \doteq \min_{g \in G(1,i)} \left[\sum_{(m \to n) \in g} Q(O_m, O_n) \right].$$

Generalization of Freidlin-Wentzell

Freidlin-Wentzell proved that

$$\lim_{\varepsilon \to 0} -\varepsilon \log \mu^{\varepsilon}(B_{\delta}(x)) = W(x) - W(O_1),$$

where $W(x) \doteq \min_{i \in L}[W(O_i) + Q(O_i, x)]$.

Theorem (Dupuis and Wu, 2020)

Let $T^{\varepsilon} = e^{\frac{1}{\varepsilon}c}$ for some $c > h \vee w$. Given a continuous function $f : \mathbb{R}^d \to \mathbb{R}$ and any compact set $A \subset \mathbb{R}^d$,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left| E\left(\frac{1}{T^{\varepsilon}} \int_0^{T^{\varepsilon}} e^{-\frac{1}{\varepsilon} f(X_t^{\varepsilon})} 1_A(X_t^{\varepsilon})\, dt \right) - \int_{\mathbb{R}^d} e^{-\frac{1}{\varepsilon} f(x)} 1_A(x)\, \mu^{\varepsilon}(dx) \right|$$
$$\geq \inf_{x \in A} [f(x) + W(x)] - W(O_1) + c - (h \vee w),$$

with $h \doteq \min_{i \in L \setminus \{1\}} Q(O_1, O_i)$ and $w \doteq W(O_1) - \min_{i \in L \setminus \{1\}} W(O_1 \cup O_i)$.

Decay rate of variance

Theorem (Dupuis and Wu, 2020)

Under the same conditions,

$$\liminf_{\varepsilon\to 0} -\varepsilon \log \left(T^{\varepsilon} \cdot \mathrm{Var}\left(\frac{1}{T^{\varepsilon}} \int_0^{T^{\varepsilon}} e^{-\frac{1}{\varepsilon} f(X_t^{\varepsilon})} 1_A\left(X_t^{\varepsilon}\right) dt \right) \right)$$

$$\geq \min_{i\in L} \left(R_i^{(1)} \wedge R_i^{(2)} \wedge R_i^{(3)} \right),$$

where

$$R_i^{(1)} \doteq \inf_{x\in A} \left[2f(x) + Q(O_i, x)\right] + W(O_i) - W(O_1),$$

$$R_1^{(2)} \doteq 2 \inf_{x\in A} \left[f(x) + Q(O_1, x)\right] - h,$$

and for $i \in L \setminus \{1\}$

$$R_i^{(2)} \doteq 2 \inf_{x\in A} \left[f(x) + Q(O_i, x)\right] + W(O_i) - 2W(O_1) + W(O_1 \cup O_i),$$

$$R_i^{(3)} \doteq 2 \inf_{x\in A} \left[f(x) + Q(O_i, x)\right] + 2W(O_i) - 2W(O_1) - w.$$

Optimality

Double well

Theorem (Dupuis and Wu, 2020)

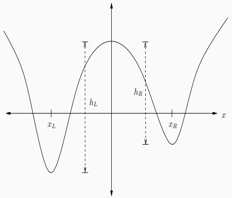
$\theta_{\mathrm{INS}}^{\varepsilon,T^{\varepsilon}}$ is an essentially unbiased estimator of $\mu^{\varepsilon}(A)$. Moreover,

$$\liminf_{\varepsilon\to 0} -\varepsilon \log\left(\mathrm{Var}\left(\theta_{\mathrm{INS}}^{\varepsilon,T^{\varepsilon}}\right)T^{\varepsilon}\right) \geq \begin{cases} r_1(\boldsymbol{\alpha}) \wedge r_3(\boldsymbol{\alpha}), & \text{if } A \subset (-\infty, 0] \\ r_1(\boldsymbol{\alpha}) \wedge r_2(\boldsymbol{\alpha}), & \text{if } A \subset [0, \infty) \end{cases},$$

where $r_3(\boldsymbol{\alpha}) \doteq 2V(A) - \alpha_K h_L$ with $V(A) \doteq \inf_{x\in A} V(x)$ and

$$r_1(\boldsymbol{\alpha}) \doteq \inf_{\boldsymbol{x}\in A\times\mathbb{R}^{K-1}} \left[2\sum_{\ell=1}^{K} \alpha_\ell V(x_\ell) - \min_{\sigma\in\Sigma_K}\left\{\sum_{\ell=1}^{K}\alpha_\ell V(x_{\sigma(\ell)})\right\}\right],$$

$$r_2(\boldsymbol{\alpha}) \doteq \min_{i\in\{2,\ldots,K+1\}}\left\{2V(A) + \left[\sum_{\ell=1}^{i-2}\alpha_{K-\ell+1} - \alpha_{K-i+2}\right](h_L - h_R)\right\} - \alpha_K h_R.$$



Double well

Theorem (Dupuis and Wu, 2020)

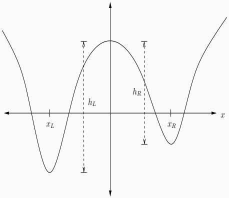
$\theta^{\varepsilon,T^\varepsilon}_{\text{INS}}$ is an essentially unbiased estimator of $\mu^\varepsilon(A)$. Moreover,

$$\liminf_{\varepsilon\to 0} -\varepsilon \log\left(\text{Var}\left(\theta^{\varepsilon,T^\varepsilon}_{\text{INS}}\right) T^\varepsilon\right) \geq \begin{cases} r_1(\boldsymbol{\alpha}) \wedge r_3(\boldsymbol{\alpha}), & \text{if } A \subset (-\infty, 0] \\ r_1(\boldsymbol{\alpha}) \wedge r_2(\boldsymbol{\alpha}), & \text{if } A \subset [0,\infty) \end{cases},$$

where $r_3(\boldsymbol{\alpha}) \doteq 2V(A) - \alpha_K h_L$ with $V(A) \doteq \inf_{x\in A} V(x)$ and

$$r_1(\boldsymbol{\alpha}) \doteq \inf_{\boldsymbol{x}\in A\times\mathbb{R}^{K-1}} \left[2\sum_{\ell=1}^{K} \alpha_\ell V(x_\ell) - \min_{\sigma\in\Sigma_K}\left\{\sum_{\ell=1}^{K}\alpha_\ell V(x_{\sigma(\ell)})\right\}\right],$$

$$r_2(\boldsymbol{\alpha}) \doteq \min_{i\in\{2,\ldots,K+1\}}\left\{2V(A) + \left[\sum_{\ell=1}^{i-2}\alpha_{K-\ell+1} - \alpha_{K-i+2}\right](h_L - h_R)\right\} - \alpha_K h_R.$$



- Optimal $\boldsymbol{\alpha}^* = (1, 1/2, \ldots, (1/2)^{K-2}, \alpha^*_K)$, where α^*_K is determined by $V(A), h_L$ and h_R.

Double well

Theorem (Dupuis and Wu, 2020)

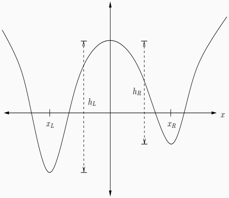
$\theta^{\varepsilon,T^\varepsilon}_{\text{INS}}$ is an essentially unbiased estimator of $\mu^\varepsilon(A)$. Moreover,

$$\liminf_{\varepsilon\to 0} -\varepsilon \log\left(\text{Var}\left(\theta^{\varepsilon,T^\varepsilon}_{\text{INS}}\right) T^\varepsilon\right) \geq \begin{cases} r_1(\boldsymbol{\alpha}) \wedge r_3(\boldsymbol{\alpha}), & \text{if } A \subset (-\infty, 0] \\ r_1(\boldsymbol{\alpha}) \wedge r_2(\boldsymbol{\alpha}), & \text{if } A \subset [0, \infty) \end{cases},$$

where $r_3(\boldsymbol{\alpha}) \doteq 2V(A) - \alpha_K h_L$ with $V(A) \doteq \inf_{x\in A} V(x)$ and

$$r_1(\boldsymbol{\alpha}) \doteq \inf_{\boldsymbol{x}\in A\times\mathbb{R}^{K-1}} \left[2\sum_{\ell=1}^{K} \alpha_\ell V(x_\ell) - \min_{\sigma\in\Sigma_K}\left\{\sum_{\ell=1}^{K}\alpha_\ell V(x_{\sigma(\ell)})\right\}\right],$$

$$r_2(\boldsymbol{\alpha}) \doteq \min_{i\in\{2,\dots,K+1\}}\left\{2V(A) + \left[\sum_{\ell=1}^{i-2}\alpha_{K-\ell+1} - \alpha_{K-i+2}\right](h_L - h_R)\right\} - \alpha_K h_R.$$



- Optimal $\boldsymbol{\alpha}^* = (1, 1/2, \dots, (1/2)^{K-2}, \alpha^*_K)$, where α^*_K is determined by $V(A), h_L$ and h_R.
- Supremum always $\geq 2V(A) - (1/2)^{K-2}V(A)$.

MULTI-WELL

THEOREM (DUPUIS AND WU, 2021)

There exists $B \in (0,\infty)$ such that the following hold. Consider any $\boldsymbol{\alpha}$ and let $T^{\varepsilon} = e^{\frac{1}{\varepsilon}c}$ for some $c > \alpha_K B$. Then $\theta_{\text{INS}}^{\varepsilon,T^{\varepsilon}}$ is essentially unbiased, and

$$\liminf_{\varepsilon\to 0} -\varepsilon \log \left(\text{Var}(\theta_{\text{INS}}^{\varepsilon,T^{\varepsilon}})T^{\varepsilon}\right) \geq r(\boldsymbol{\alpha}) - \alpha_K B,$$

where

$$r(\boldsymbol{\alpha}) \doteq \inf_{x\in A\times\mathbb{R}^{d(K-1)}} \left\{ 2\sum_{\ell=1}^{K} \alpha_\ell V(x_\ell) - \min_{\sigma\in\Sigma_K} \left\{ \sum_{\ell=1}^{K} \alpha_\ell V(x_{\sigma(\ell)}) \right\} \right\}.$$

THEOREM (DUPUIS AND WU, 2021)

For any closed set A, and any $\alpha_K \in (0,(1/2)^{K-1}]$,

$$\sup_{(\alpha_2,\ldots,\alpha_{K-1})\in[\alpha_K,1]^{K-2}} r(\alpha_1,\alpha_2,\cdots,\alpha_{K-1},\alpha_K) = (2+\alpha_K-(1/2)^{K-2})V(A).$$

The supremum is achieved at $(\alpha_1^*,\ldots,\alpha_{K-1}^*)$ with $\alpha_\ell^* = (1/2)^{\ell-1}$ for all ℓ.

SUMMARY

- "Metastability" present a particular challenge for the design of efficient Monte Carlo methods.
- As such, it is natural to use various asymptotic theories to understand issues of algorithm design.
- Have presented one use of large deviation ideas in the context of infinite swapping (and parallel tempering) algorithms to understand the mechanisms that produce variance reduction.
- INS process with a geometric sequence of temperatures explore landscape in a organized and meaningful way. (Ongoing work)

References

Infinite swapping as a limit of parallel tempering:

- ◇ "On the infinite swapping limit for parallel tempering", Dupuis, Liu, Plattner and Doll, *SIAM J. on MMS*, 10, 986–1022, 2012.

Large deviation estimates:

- ◇ "Large Deviation Properties of the Empirical Measure of a Metastable Small Noise Diffusion", Dupuis and **Wu**, *J Theo. Prob.*, 2020.

Analysis of INS algorithm:

- ◇ "Analysis and optimization of certain parallel Monte Carlo methods in the low temperature limit", Dupuis and **Wu**, submitted, 2021.

Contact information: Guo-Jhen Wu; gjwu@kth.se