Efficient approximation of BRW Gibbs measures

Fu-Hsuan Ho, joint work with **Pascal Maillard**

Random Networks and Interacting Particle Systems (webinar)
10/09/2021

Institut de Mathématiques de Toulouse

Statistical physics and disordered system

A fast recap of statistical physics

A generic model in statistical physics can be described by

- configurations/states σ of "spins". For example, $\sigma \in \{-1, +1\}^N$ for Ising spins, $\sigma \in \mathbb{S}^N$ for spherical spins, or σ might be a leaf of a binary tree $\mathbb{T}_N$.
- Hamiltonian $H_N(\sigma)$: the energy of the configuration σ.
- Gibbs measure: $\mu_{\beta,N}(\sigma) \propto e^{\beta H_N(\sigma)}$, where β is the inverse temperature.

A generic model in statistical physics can be described by

- configurations/states σ of "spins". For example, $\sigma \in \{-1, +1\}^N$ for Ising spins, $\sigma \in \mathbb{S}^N$ for spherical spins, or σ might be a leaf of a binary tree $\mathbb{T}_N$.
- Hamiltonian $H_N(\sigma)$: the energy of the configuration σ.
- Gibbs measure: $\mu_{\beta,N}(\sigma) \propto e^{\beta H_N(\sigma)}$, where β is the inverse temperature.

Example: the Curie–Weiss model.

- Ising spins: $\sigma \in \{\pm 1\}^N$.
- Hamiltonian: $H_N(\sigma) = \frac{1}{N}\sum_{i,j=1}^{N} \sigma_i \sigma_j$.
- Gibbs measure:

$$\mu_{\beta,N}(\sigma) = \frac{1}{Z_{\beta,N}} \exp(\beta H_N(\sigma)), \quad \text{where} \quad Z_{\beta,N} = \sum_{\sigma \in \{\pm 1\}^N} \exp(\beta H_N(\sigma)).$$

- There exists a critical temperature $T_c = 1/\beta_c$ (Curie temperature) such that the model exhibits a **phase transition**.
- To be precise, if we consider the magnetization density of the Curie–Weiss model

$$m_N = \frac{1}{N}\sum_{i=1}^{N}\sigma_i,$$

then the magnetization concentrates at zero when $\beta \le \beta_c$ (paramagnetic), and it is bounded away from zero when $\beta > \beta_c$ (ferromagnetic).

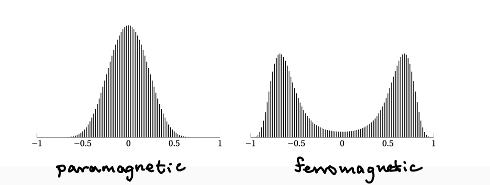
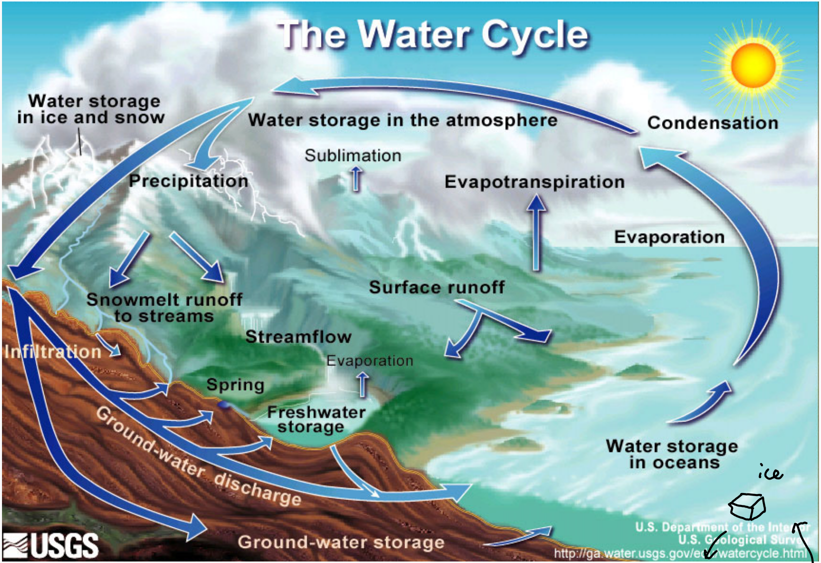


Figure 1: Distribution of the magnetization $\mu_{\beta,N}(m_N = \cdot)$. Credit: S. Friedli and Y. Velenik, Figure 2.2.

Phase transition in everyday life



ice

water

steam

For the Curie–Weiss model, the Hamiltonians are deterministic. But physicists are also interested in systems where the Hamiltonians are random (disordered).

Example: The Sherrington–Kirkpatrick (SK) model.

- Ising spins $\sigma \in \{\pm 1\}^N$
- Hamiltonian
$$H_N(\sigma) = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij}\sigma_i\sigma_j,$$
where $A \in \mathrm{GOE}(N)$. That is, A is a symmetric matrix where the upper-triangular entries are i.i.d., $A_{ii} \sim N(0, 2/N)$, and $A_{ij} \sim N(0, 1/N)$ for $i < j$.
- The SK model is an example of *spin glasses*.

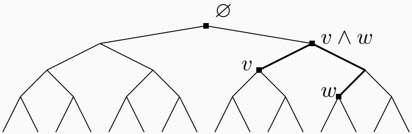
The continuous random energy model (CREM)

Physicists came up with other models which have more tractable correlation structures.

- The CREM was proposed by [Bovier and Kurkova ’04], which generalized Derrida’s random energy model formulated in the ’80s.
- $(X_v)_{v\in\mathbb{T}_N}$: centered Gaussian process indexed by a binary tree $\mathbb{T}_N$ of depth N with correlation function

$$\mathbb{E}\left[X_v X_w\right] = N \cdot A\left(\frac{|v \wedge w|}{N}\right),$$

  where $A : [0,1] \to [0,1]$ is a continuous non-decreasing function such that $A(0) = 0$ and $A(1) = 1$.
- The CREM can be seen as a time-inhomogeneous binary **branching random walk** with Gaussian increments.



Disordered systems and algorithms

- Disordered systems originated from the study of glassy materials, but then the research on this direction had a far-reaching influence on other fields of science.
  - Disordered systems and neural networks: Lenka Zdeborová, *What has statistical physics to say about machine learning?*
- Algorithmic aspect of disordered systems:
  - How to find a near optimal state/ground state of the system?
  - How to sample a typical state from the Gibbs measure?
  - Do these algorithms always work efficiently?

Finding ground states of the CREM [Addario-Berry and Maillard '19]

- **Goal**: Fix an energy level $x \in \mathbb{R}$. Find a leaf v such that $X_v \geq xN$.
- **Results**: there exists a threshold

$$x_* = x_*(A) = \sqrt{2 \log 2} \int_0^1 \sqrt{A'(t)} \, \mathrm{d}t$$

such that

1. for all $x < x_*$, there exists a linear-time algorithm which can achieve the goal with high probability.
2. for all $x > x_*$, the goal is exponentially hard to accomplish.

※ assumption in the paper : $A(t) = \int_0^t a(s) \, ds$ where $a(s)$ is Riemann integrable.

Comments on [Addario-Berry and Maillard '19]

- The ground state energy x_s is defined by

$$x_s = \lim_{N\to\infty} \frac{1}{N} \max_{|v|=N} X_v \overset{\text{a.s.}}{=} \sqrt{2\log 2}\int_0^1 \sqrt{\hat{a}(t)}\,\mathrm{d}t\,,$$

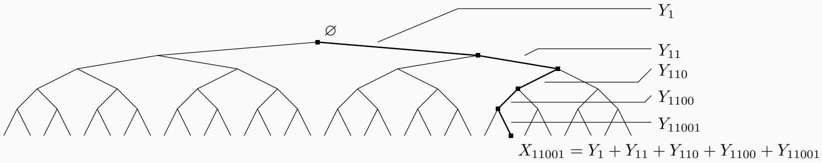
  where $\hat{a}$ is the left-derivative of the concave hull of A.
- The ground state equals to x_* if and only if A is concave. Thus,

  convexity of $A \Leftrightarrow \exists$ efficient algorithms finding the near ground state.
- Parallel works on spin glasses:
  - [Montanari '18] for the SK model
  - [Subag '18] for mixed spherical p-spin model
  - [Gamarnik and Jagannath '19] for pure p-spin models.
- The near ground state particles are rare. How to sample a **typical** particle from the Gibbs measure of the CREM?
  The rest of the talk: answering the special case where $A(t) = t$.

Branching random walk and the algorithm

- Rooted d-ary tree: $\mathbb{T} = \{\varnothing\} \cup \bigcup_{n=1}^{\infty}\{0, \dots, d-1\}^n$
- Depth of $v \in \mathbb{T} = |v|$.
- Independent increments: $(\mathsf{Y}^v)_{v \in \mathbb{T}\setminus\{\varnothing\}}$ i.i.d. copies of $\mathsf{Y} = (Y_0, \dots, Y_{d-1})$.
- **branching random walk** $\mathsf{X} = (X_v)_{v\in\mathbb{T}}$ with increments Y is a process defined by

$$\begin{cases} X_\varnothing = 0, \\ X_v = \sum_{\varnothing < w \le v} Y_w, \quad |v| \ge 1. \end{cases}$$



- The Gibbs measure of the CREM is defined by

$$\mu_{\beta,N}(u) = \frac{1}{Z_{\beta,N}} e^{\beta X_u}, \quad \text{where} \quad Z_{\beta,N} = \sum_{|u|=N} e^{\beta X_u}.$$

- For a given node v and $|u| = m$, we define

$$\mu^v_{\beta,m}(u) = \frac{1}{Z^v_{\beta,N}} e^{\beta(X_{vu} - X_v)}, \quad \text{where} \quad Z^v_{\beta,N} = \sum_{|u|=m} e^{\beta(X_{vu} - X_v)}.$$

Questions: how to sample the Gibbs measure efficiently? If we can, is it possible to do it for all $\beta > 0$?

Consider the following algorithm

---

set $v = \varnothing$;
while $|v| < N$ **do**
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;
  replace v with vw;
end
output v.

Recursive sampling on M-renormalized tree

Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

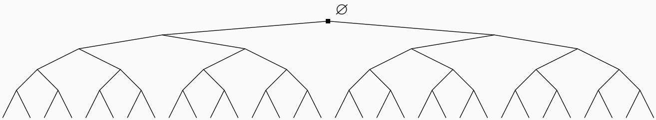
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;

  replace v with vw;

end

output v.

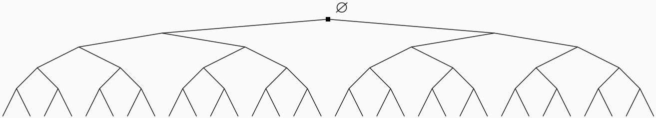
Example: binary tree, $M = 2$ and $N = 5$.

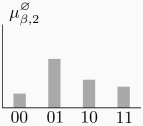


Consider the following algorithm

set $v = \varnothing$;
while $|v| < N$ **do**
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;
  replace v with vw;
end
output v.

Example: binary tree, $M = 2$ and $N = 5$.





Consider the following algorithm

set $v = \varnothing$;

while $|v| < N$ **do**

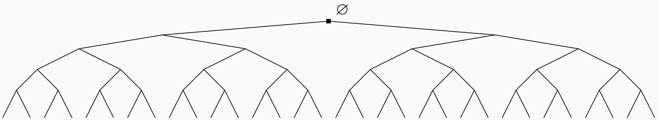
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;

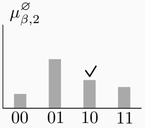
  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.





Recursive sampling on M-renormalized tree

Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

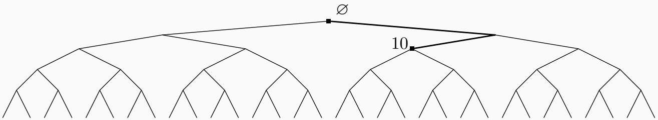
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;

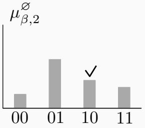
  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.





Recursive sampling on M-renormalized tree

Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

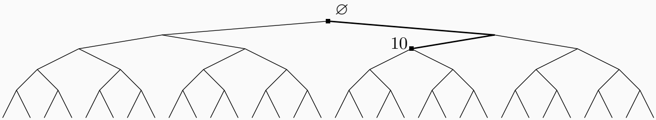
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;

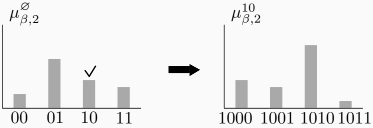
  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.





Recursive sampling on M-renormalized tree

Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

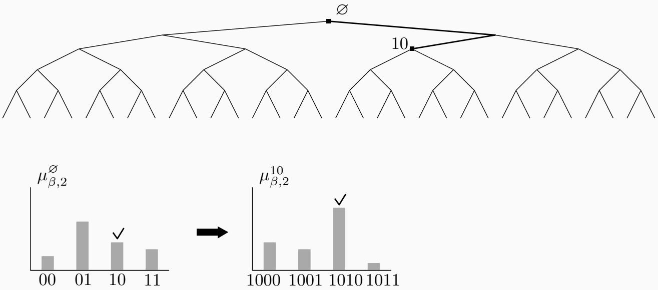
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M \wedge (N-|v|)}$;

  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.



Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

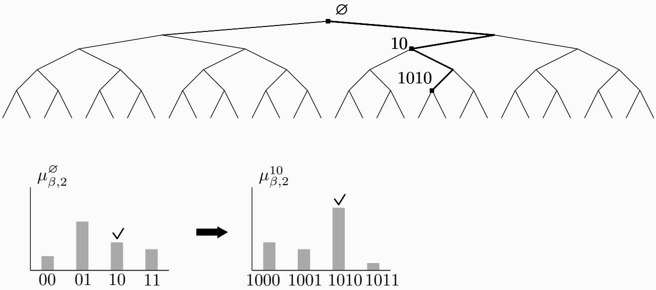
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;

  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.



Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

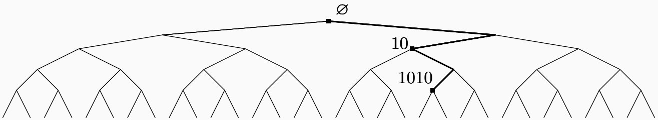
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;

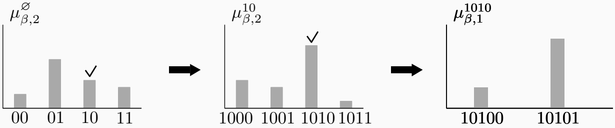
  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.





Recursive sampling on M-renormalized tree

Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

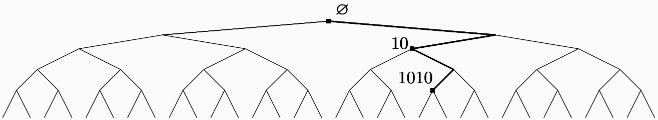
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M\wedge(N-|v|)}$;

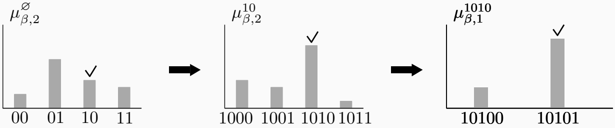
  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.





Recursive sampling on M-renormalized tree

Consider the following algorithm

---

set $v = \varnothing$;

while $|v| < N$ **do**

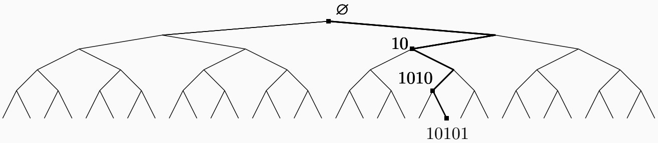
  choose w with $|w| = M \wedge (N - |v|)$ according to the Gibbs measure $\mu^{v}_{\beta, M \wedge (N-|v|)}$;

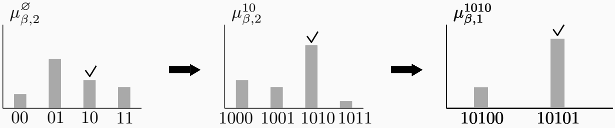
  replace v with vw;

end

output v.

Example: binary tree, $M = 2$ and $N = 5$.





Remarks

- The running time of the algorithm is deterministic and bounded by $\lceil N/M \rceil 2^M$.
- Moreover, the law of the algorithm is a random probability measure $\mu_{\beta,M,N}$ defined by

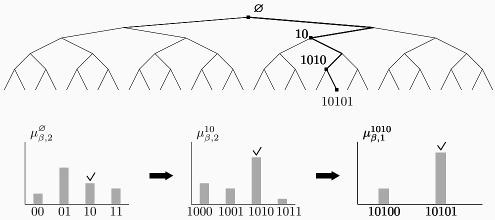
$$\mu_{\beta,M,0}(\varnothing) = 1$$

$$\mu_{\beta,M,N\wedge(K+1)M}(vw) = \mu_{\beta,M,KM}(v) \cdot \mu^{v}_{\beta,M\wedge(N-KM)}(w)$$

for all $|v| = KM$, $|w| = M \wedge (N-KM)$ and $0 \leq K \leq \lfloor \frac{N}{M} \rfloor$.

Example: binary tree, $M = 2$ and $N = 5$. On this configuration,

$$\mu_{\beta,5,2}(10101) = \mu_{\beta,2}(10) \times \mu^{10}_{\beta,2}(10) \times \mu^{1010}_{\beta,2}(1)$$



Question: does $\mu_{\beta,M,N}$ approximate the Gibbs measure?

Main results

The entropy of $\mu_{\beta,N}$ and the Kullback–Leibler divergence from $\mu_{\beta,N}$ to $\tilde{\mu}_{\beta,N}$ are respectively defined by

$$H(\mu_{\beta,N}) = \sum_{|u|=N} \mu_{\beta,N}(u) \cdot \log\left(\frac{1}{\mu_{\beta,N}(u)}\right)$$

$$\mathrm{d}(\tilde{\mu}_{\beta,N} \mid\mid \mu_{\beta,N}) = \sum_{|u|=N} \tilde{\mu}_{\beta,N}(u) \cdot \log\left(\frac{\tilde{\mu}_{\beta,N}(u)}{\mu_{\beta,N}(u)}\right).$$

Definition (Approximation)

Let $\beta \geq 0$. We say that a sequence of random probability measures $(\tilde{\mu}_{\beta,N})_{N\geq 1}$ *approximates* the Gibbs measure $\mu_{\beta,N}$ if

$$\frac{\mathrm{d}(\tilde{\mu}_{\beta,N} \mid\mid \mu_{\beta,N})}{H(\mu_{\beta,N})} \to 0, \quad \text{in probability as } N \to \infty.$$

Main results I: efficient approximation in the subcritical regime

Theorem (Approximation bounds)

Let $M \in \mathbb{N}$ and $\beta \in [0, \beta_c)$. Then for all $p \geq 1$, there exists a constant $C_1(p) > 0$ such that

$$\|\mathsf{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N})\|_p \leq C_1(p) \cdot \left\lfloor \frac{N}{M} \right\rfloor. \tag{1}$$

Moreover, for all $p \geq 1$, there exists a constant $C_1(p) > 0$ such that

$$\|\mathsf{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N}) - \mathbb{E}\left[\mathsf{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N})\right]\|_p \leq C_1(p). \tag{2}$$

Corollary (Complexity upper bound)

If $\beta \in [0, \beta_c)$, then there exists a polynomial-time algorithm such that for every $p > 1$, denoting by $\tilde{\mu}_{\beta,N}$ the law of its output,

$$\frac{1}{N}\|\mathsf{d}(\tilde{\mu}_{\beta,N} \mid\mid \mu_{\beta,N})\|_p \to 0 \tag{3}$$

as $N \to \infty$. In particular, $\tilde{\mu}_{\beta,N}$ approximates the Gibbs measure $\mu_{\beta,N}$.

Theorem (Complexity lower bound)

Assume $\beta_c \in \mathcal{D}(\varphi)^\circ$. Let $\beta > \beta_c$. Suppose that $v = (v(k))_{k \geq 0}$ a random algorithm with law $\tilde{\mu}_N$ such that $\tilde{\mu}_N$ approximates the Gibbs measure $\mu_{\beta,N}$, and τ is the running time of the algorithm. Then for every $\delta > 0$, there exists $z > 0$, such that for large enough N,

$$\mathbb{P}\left(\tau \geq e^{z\sqrt{N}}\right) \geq 1 - \delta.$$

Remark. Note that $\beta_c \in \mathcal{D}(\varphi)^\circ$ implies in particular that $\beta_c < \infty$.

※ $\mathcal{D}(\varphi) = \{\beta \in \mathbb{R} : \varphi(\beta) < \infty\}$ effective domain of φ.

Main ingredients of the proofs

Ingredients of proving the approximation bound

Homogeneity of the BRW $\rightsquigarrow$ factorization of the KL divergence

$$\mathrm{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N}) = \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathrm{d}(\mu^{u}_{\beta,M} \mid\mid \mu^{u}_{\beta,N-KM}).$$

Calculation of the expectation $\rightsquigarrow$ generalizing it to p-norms.

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N})\right] \\
&= \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \sum_{|u|=KM} \mathbb{E}\left[\mu_{\beta,M,KM}(u) \cdot \mathrm{d}(\mu^{u}_{\beta,M} \mid\mid \mu^{u}_{\beta,N-KM})\right] \\
&= \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \sum_{|u|=KM} \mathbb{E}\left[\mu_{\beta,M,KM}(u) \cdot \mathbb{E}\left[\mathrm{d}(\mu^{u}_{\beta,M} \mid\mid \mu^{u}_{\beta,N-KM}) \,\middle|\, \mathcal{F}_{KM}\right]\right] \\
&= \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \sum_{|u|=KM} \mathbb{E}\left[\mu_{\beta,M,KM}(u) \cdot \mathbb{E}\left[\mathrm{d}(\mu_{\beta,M} \mid\mid \mu_{\beta,N-KM})\right]\right] \\
&= \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \mathbb{E}\left[\mathrm{d}(\mu_{\beta,M} \mid\mid \mu_{\beta,N-KM})\right].
\end{aligned}$$

Ingredients of proving the complexity upper bound

- We showed the approximation bound

$$\|\mathsf{d}(\tilde{\mu}_{\beta,N} \,||\, \mu_{\beta,N})\|_p \leq C_1(p) \cdot \left\lfloor \frac{N}{M} \right\rfloor. \tag{4}$$

- By choosing $M(N) = O(\log N)$, we have that the running time $\lceil N/M(N) \rceil 2^{M(N)}$ is of polynomial, and that

$$\frac{1}{N}\|\mathsf{d}(\tilde{\mu}_{\beta,N} \,||\, \mu_{\beta,N})\|_p \to 0, \quad N \to \infty.$$

- (Folklore) In the subcritical regime $\beta \in [0, \beta_c)$, $H(\mu_{\beta,N})/N$ converges in probability to a positive constant as $N \to \infty$.

$$\Rightarrow \frac{\mathsf{d}(\tilde{\mu}_{\beta,N} \,||\, \mu_{\beta,N})}{H(\mu_{\beta,N})} \xrightarrow{\mathbb{P}} 0, \quad N \to \infty.$$

Ingredients of proving the complexity lower bound

- One can show that with high probability, there exists $z > 0$ such that for large enough N, the algorithm finds a vertex u whose ancestor w at generation $\lfloor N/2 \rfloor$ is z-exceptional, that is,

$$X_u - X_w - \varphi'(\beta_c)N/2 > z\sqrt{N}.$$

- For all $|w| = \lfloor N/2 \rfloor$, define

$$A_w = \{w \text{ is } z\text{-exceptional}\}.$$

  The events are independent by the branching property. Moreover, the probability $\mathbb{P}(A_w)$ are exponentially small in $z\sqrt{N}$ by tail estimate of the maximal particle of BRW.

- The proof follows from adapting the argument of [Addario-Berry and Maillard '19].nea

Conclusions and Outlook

- Disordered systems give rise up to many interesting algorithmic problems, such that sampling the ground states or the typical states of the system.
- We provided an algorithm that can approximate the Gibbs measure of BRW in the subcritical regime, and we showed that the Gibbs measure is algorithmically hard to approximate in the supercritical regime. These results cover the case of the CREM where $A(t) = t$.
- Ongoing work of us is trying to extend the results to general CREM.

Merci ! / Thank you! / 謝謝 !

Appendix

Corollary (Complexity upper bound)

If $\beta \in [0, \beta_c)$, then there exists a polynomial-time algorithm such that for every $p > 1$, denoting by $\tilde{\mu}_{\beta,N}$ the law of its output,

$$\frac{1}{N}\|\mathsf{d}(\tilde{\mu}_{\beta,N} \mid\mid \mu_{\beta,N})\|_p \to 0 \tag{5}$$

as $N \to \infty$. In particular, $\tilde{\mu}_{\beta,N}$ approximates the Gibbs measure $\mu_{\beta,N}$.

- In the physics literature, one would say that $\tilde{\mu}_{\beta,N}$ and $\mu_{\beta,N}$ are *measure equivalent* or *equivalent in the sense of specific relative entropy*.
- Equation (5) implies the following: if $(A_N)_{N\geq 1}$ is a sequence of sets such that $\mu_{\beta,N}(A_N)$ convergences to 0 exponentially fast as $N \to \infty$, then one can show that $\tilde{\mu}_{\beta,N}(A_N) \to 0$ (ex: via Birgé's inequality).

Ideas to prove the approximation bound (1) and (2)

1. Homogeneity of the BRW $\rightsquigarrow$ factorization of the KL divergence

$$\mathrm{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N}) = \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor -1} \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathrm{d}(\mu^u_{\beta,M} \mid\mid \mu^u_{\beta,N-KM}).$$

2. Biggins' theoerm + negative moments estimation of $W_{\beta,n}$
$\rightsquigarrow$ for all $p \geq 1$, we have

$$\sup_{n \geq 0} \|\log W_{\beta,n}\|_p < \infty \quad \text{which will give us} \quad \|\mathbf{d}(\mu_{\beta,M} \mid\mid \mu_{\beta,N})\|_p \leq C(p).$$

⚠ Item 2 holds as long as $\beta < \beta_c$.

3. Item 1 + Item 2 $\rightsquigarrow$ approximation bound (1)

$$\|\mathbf{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N})\|_p \leq \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor -1} C(p) = \left\lfloor \frac{N}{M} \right\rfloor \cdot C(p).$$

4. Concentration inequalities + estimate of the square sums of $\mu_{\beta,M,N}(\cdot) \rightsquigarrow$ approximation bound (2)

$$\|\mathbf{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N}) - \mathbb{E}\left[\mathbf{d}(\mu_{\beta,M,N} \mid\mid \mu_{\beta,N})\right]\|_p \leq C_p \cdot r^K.$$

Proof of the first statement:

1. Choose $M = M(N) \to \infty$ as $N \to \infty$, and set $\tilde{\mu}_{\beta,N} = \mu_{\beta,M,N}$.
2. By the approximation bound (1), we obtain

$$\frac{1}{N}\|\mathbf{d}(\tilde{\mu}_{\beta,N} \,||\, \mu_{\beta,N})\|_p = \frac{1}{N}\left\|\mathbf{d}(\mu_{\beta,M(N),N} \,||\, \mu_{\beta,N})\right\|_p \le \frac{1}{N} \cdot C_1(p) \cdot \left\lfloor \frac{N}{M(N)} \right\rfloor \to 0, \quad N \to \infty. \tag{6}$$

Proof of the second statement:

1. Assuming moreover that $M = O(\log N)$. In particular, the running time is of polynomial.
2. We have the following folklore lemma

Lemma. Asymptotics of the entropies for $\beta < \beta_c$

If $\beta \in [0, \beta_c)$, then $H(\mu_{\beta,N})/N$ converges in probability to a positive constant as $N \to \infty$.

3. Then by (6) and the lemma above, we conclude that

$$\frac{\mathbf{d}(\tilde{\mu}_{\beta,N} \,||\, \mu_{\beta,N})}{H(\mu_{\beta,N})} \to 0, \quad \text{in probability as } N \to \infty.$$

Steps of proving the complexity lower bound

1. If $\beta > \beta_c$ and if $\beta_c \in \mathcal{D}(\varphi)^\circ$, then $H(\mu_{\beta,N}) = O_{\mathbb{P}}(1)$, as $N \to \infty$.
2. By [Chen, Madaule and Mallein '19], there exists a RV $Z > 0$ with continuous distribution function such that
$$\frac{X_u - X_{u_{N/2}} - \varphi'(\beta_c) \cdot N/2}{\sqrt{N}} \overset{\text{law}}{\to} Z, \quad N \to \infty.$$
3. Assume $\tilde{\mu}_{\beta,N}$ is the law of an algorithm that approximates the Gibbs measure $\mu_{\beta,N}$. Step 1 then implies that $\mathsf{d}(\tilde{\mu}_{\beta,N} \mid\mid \mu_{\beta,N}) \overset{\mathbb{P}}{\to} 0$ as $N \to \infty$.
4. Pinsker's inequality implies that the total variation between $\tilde{\mu}_{\beta,N}$ and $\mu_{\beta,N} \overset{\mathbb{P}}{\to} 0$ as $N \to \infty$. Thus Step 2 holds as well for $\tilde{\mu}_{\beta,N}$.
5. Fix $\delta > 0$. There exists $z > 0$ such that for large enough N, the algorithm finds a vertex u whose ancestor w at generation $\lfloor N/2 \rfloor$ is z-exceptional, that is,
$$X_u - X_w - mN/2 > z\sqrt{N},$$
with probability at least $1 - \delta$.

Steps of proving the complexity lower bound

6. For all $|w| = \lfloor N/2 \rfloor$, define

$$A_w = \{w \text{ is } z\text{-exceptional}\}.$$

   The events are independent by the branching property.
7. By tail estimate of the maximal particle of BRW, the probability $\mathbb{P}(A_w)$ are exponentially small in $z\sqrt{N}$. Then one can adapt the argument of [Addario-Berry and Maillard '19] to derive the result.