INSTITUT de MATHEMATIQUES de TOULOUSE

Probing the transition from polynomial to exponential complexity in spin glasses through N-particle branching Brownian motions

Pascal Maillard (Université de Toulouse)

Nov 4, 2024

Workshop “Recent developments beyond classical regimes in statistical learning”, Toulouse

Overview

1. **Introduction: optimization of random functions**

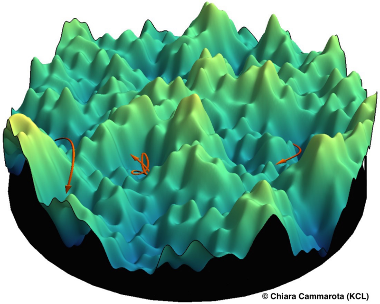
2. **Derrida's CREM and time-inhomogeneous BBM**

3. **Time-inhomogeneous N-BBM: near the hardness threshold**

Introduction: optimization of random functions

Optimization of random functions

- Ubiquitous computational task: optimization of a highly non-convex function on high-dimensional space (machine learning, combinatorial optimization,...).
- Average-case complexity *vs.* worst-case complexity
- Theoretical framework: spin glasses in statistical mechanics


© Chiara Cammarota (KCL)

Spin glasses

Typical statistical mechanics model described by (Boltzmann, Gibbs):

- State space Σ
- Hamiltonian $H : \Sigma \to \mathbb{R}$
- Gibbs measure $\mu_\beta(\sigma) \propto e^{-\beta H(\sigma)}$, $\beta \geq 0$ (inverse temperature)

Example: Ising model (on graph $G = (V, E)$, without magnetic field):
$\Sigma = \{-1, 1\}^V, H(\sigma) = -\sum_{v \sim w} \sigma_v \sigma_w$

Spin glass: the Hamiltonian H is itself random.

Example: (Ising) p-spin model ($p \geq 2$, without magnetic field):

$$\Sigma_n = \{-1, 1\}^n,\ H_n(\sigma) = n^{-\frac{p-1}{2}} \sum_{i_1, \ldots, i_p = 1}^{n} J_{i_1, \ldots, i_p} \sigma_{i_1} \cdots \sigma_{i_p},$$

where $J_{i_1, \ldots, i_p}$ are iid standard Gaussian random variables.

Parisi measure and Parisi ultrametricity

Overlap between $\sigma, \sigma' \in \Sigma_n$: $R(\sigma,\sigma') = \frac{1}{n}\langle \sigma, \sigma' \rangle = \frac{1}{n}\sum_{i=1}^n \sigma_i \sigma_i' \in [-1,1]$.

Mean overlap measure:

$$\nu_{\beta,n}(dt) = \mathbb{E}\left[\sum_{\sigma,\sigma' \in \Sigma_n} \mathbf{1}_{(R(\sigma,\sigma')\in dt)} \mu_{\beta,n}(\sigma)\mu_{\beta,n}(\sigma')\right], \quad t \in \mathbb{R}.$$

Parisi measure: $\nu_\beta := \lim_{n\to\infty} \nu_{\beta,n}$.

Parisi ultrametricity (Parisi 1980's)

Emerging hierarchical structure whose statistics are completely determined (in the limit) by the Parisi measure ν_β.

Optimization algorithms, overlap gap property

Basic question: is it possible to find an approximate ground state, i.e., for a given $\varepsilon > 0$, to find a state $\sigma \in \Sigma_n$ such that

$$\left| \frac{H_n(\sigma)}{\min_{\sigma' \in \Sigma_n} H_n(\sigma')} - 1 \right| \leq \varepsilon,$$

in a time polynomial in n, with high probability?

Folklore conjecture (Gamarnik 21) for a wide class of models: Possible if (and only if) the overlap gap property does not hold.

Addario-Berry–M. 19, Subag 21, Montanari 21, Gamarnik–Jagannath 21, Sellke 24,...

Overlap gap property

We say that the model exhibits the overlap gap property (OGP), if the support of the Parisi measure ν_β is *not* an interval for sufficiently large β.

Derrida's CREM and time-inhomogeneous BBM

Derrida's continuous random energy model

The continuous random energy model (CREM) Derrida 1985, Bovier-Kurkova 2004

- a certain Gaussian field indexed by a tree
- a spin glass model with explicit hierarchical structure
- amenable to quite explicit (asymptotic) analysis

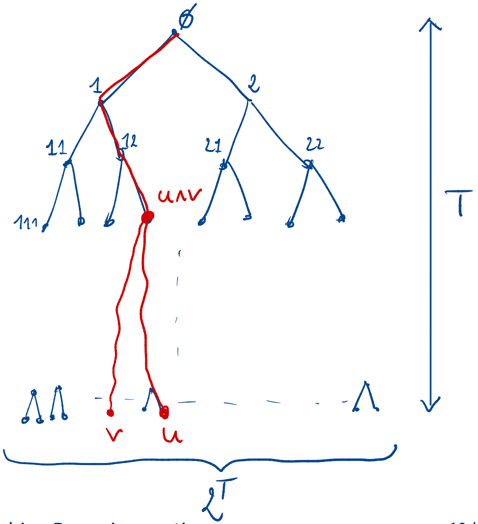
Focus of today's talk: Intrinsic barriers for the efficiency of algorithms for optimizing the CREM Hamiltonian.

joint work with: Louigi Addario-Berry Alexandre Legrand

Continuous random energy model (CREM)

- $T \in \mathbb{N}$ (large)
- $\mathbb{T}_T$: rooted binary tree of depth T
- $(X_u)_{u \in \mathbb{T}_T}$: centered Gaussian field
- $A : [0,1] \to [0,1]$ non-decreasing, $A(0) = 0, A(1) = 1$
- $|u| = \mathrm{dist}(\emptyset, u)$
- $u \wedge v$: most recent common ancestor of u and v
- Covariance matrix:

$$\mathrm{Cov}(X_u, X_v) = T \cdot A\left(\frac{|u \wedge v|}{T}\right)$$



Time-inhomogeneous branching Brownian motion

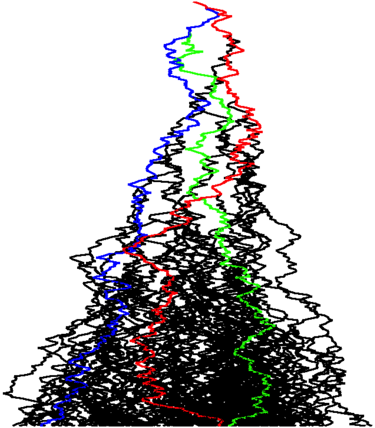
Continuous-time version of CREM. Parameters:

- $T > 0$ (large)
- $\sigma^2 : [0,1] \to \mathbb{R}_+$

The time-inhomogeneous branching Brownian motion (BBM) is a particle system where particles

- diffuse according to independent (time-inhomogeneous) Brownian motions, sped up by a factor $\sigma^2(t/T)$ at time t
- split into two particles at (constant) rate $1/2$

Essentially equivalent to CREM with $A(t) = \int_0^t \sigma^2(s)\, ds$.



Eric Brunet

Optimization problem

Known (Bovier–Kurkova 2004): First order of ground state of CREM (here, maximum instead of minimum):

$$\lim_{T\to\infty} \frac{1}{T} \max_{|u|=T} X_u = \sqrt{2\log 2} \int_0^1 \sqrt{\hat{a}(t)}\, dt,$$

where $\hat{a}$: left-derivative of $\hat{A}$, the concave hull of A.

Optimization problem

Given $x > 0$, is it possible to find vertices u with $X_u \geq xT$ within a time of order poly(T) with high probability? In particular, is there a poly(T)-time algorithm which finds an approximate ground state, i.e. a vertex u with $X_u \geq (1-\varepsilon)\sqrt{2\log 2}\int_0^1 \sqrt{\hat{a}(t)}\, dt \times T$, for every $\varepsilon > 0$?

Remark: The related problem of approximately sampling the Gibbs measure was treated in the thesis of Fu-Hsuan Ho.

Algorithmic model

Pemantle 09: An algorithm is a random sequence of vertices $(v(n))_{n\geq 1}$, such that $v(n+1)$ depends only on

- $v(1), \ldots, v(n)$
- $X_{v(1)}, \ldots, X_{v(n)}$
- possibly some additional randomness (e.g., $U_1, \ldots, U_{n+1}$, where $(U_n)_{n\geq 1}$ is a sequence of iid uniformly distributed r.v., independent of $(X_u)_{u\in\mathbb{T}_N}$)

In other words, $v(n)_{n\geq 1}$ is a predictable process w.r.t. the filtration

$$\mathscr{F}_n = \sigma\left(v(1), \ldots, v(n), X_{v(1)}, \ldots, X_{v(n)}, U_1, \ldots, U_{n+1}\right).$$

A stopping time τ is in this context also called the running time of the algorithm. The output of the algorithm is the vertex $v(\tau)$.

Optimization problem: threshold

$A(t) = \int_0^t a(s)\,ds$, with a Riemann-integrable.

Define $x_* = \sqrt{2\log 2}\int_0^1 \sqrt{a(t)}\,dt$ (algorithmic hardness threshold).

Theorem (Addario-Berry–M. 2021)

1. *For $x < x_*$, there exists an algorithm with $O(T)$ runtime, which finds a vertex u with $X_u \geq xT$ with high probability.*
2. *For $x > x_*$ every algorithm, which finds a vertex u with $X_u \geq xT$, has runtime at least $e^{\gamma T}$ with high probability, for some $\gamma = \gamma(x) > 0$.*

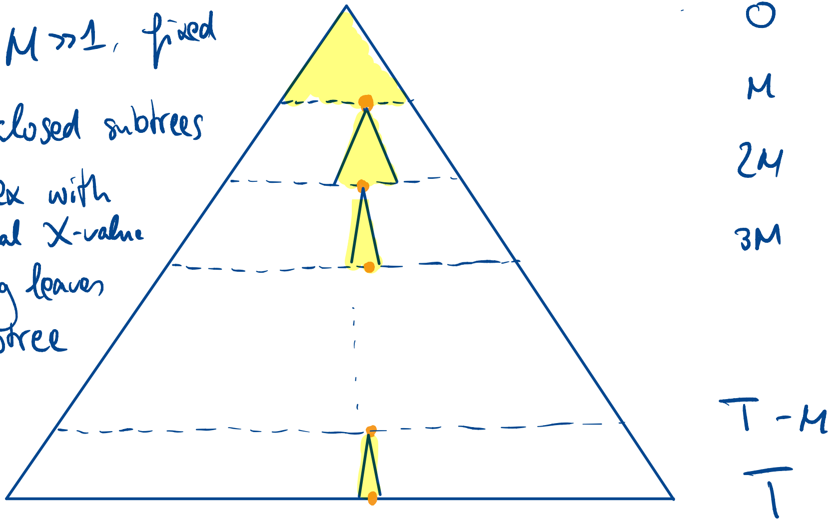
Corollary

One can approximate the ground state in a time $\mathrm{poly}(T)$ if and only if A is concave.

$M \gg 1$, fixed

: disclosed subtrees

: vertex with maximal X-value among leaves in subtree



Threshold and overlap gap property

Corollary

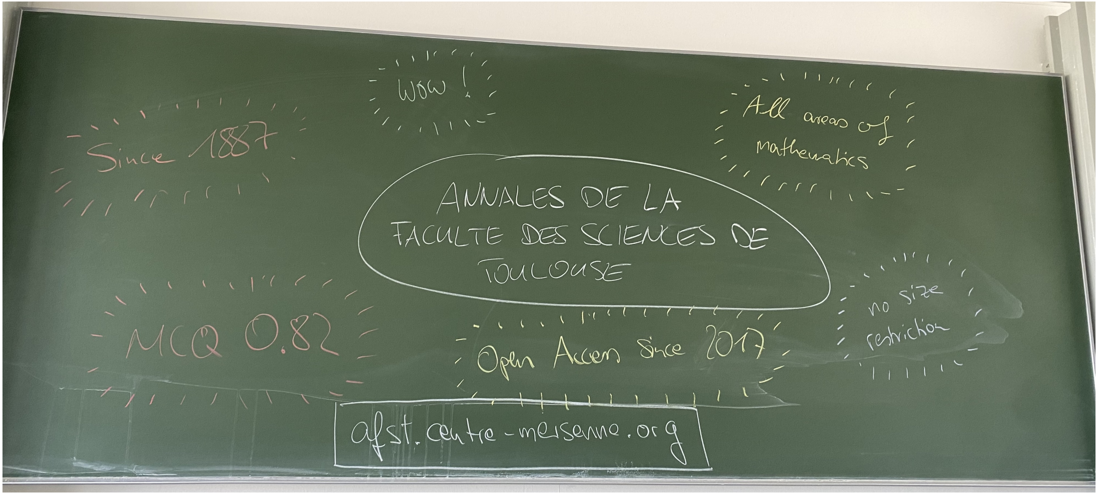
One can approximate the ground state in a time $\mathrm{poly}(T)$ if and only if A is concave.

Known (Bovier-Kurkova 2004-07): Support of Parisi measure (for sufficiently large β) is the set of extremal points of the concave hull of A. Hence, we have the equivalence:

$$\text{no overlap gap} \Leftrightarrow A \text{ strictly concave}$$

Hence, we confirm the fact that the overlap gap property is necessary and sufficient for hardness of approximating the ground state, except for boundary cases.

Time-inhomogeneous N-BBM: near the hardness threshold



https://afst.centre-mersenne.org/

Optimization problem: near the threshold

Q: what happens near the threshold x_*? ("phase transition")

Proposed algorithm to probe this: beam-search of beam width $N = N(T)$:

- follow (at most) N paths of vertices down the tree
- at every step, paths split into two, only keep the N paths with highest (terminal) value, discard the others.

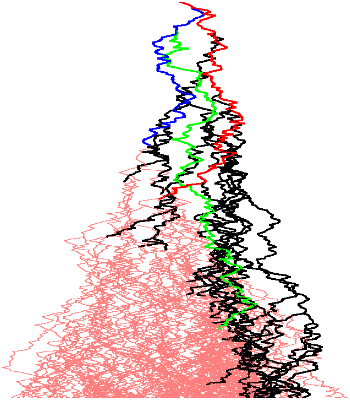
Complexity: $N \times T$.

Interesting regime: $\log T \ll \log N \ll T$ (transition from polynomial to exponential complexity).

Time-inhomogeneous N-BBM

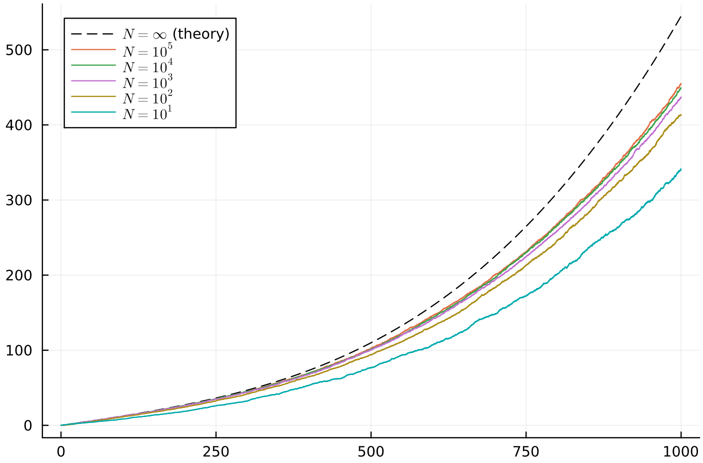
Time-inhomogeneous N-particle branching Brownian motion (N-BBM):

- particle system evolving in continuous time as follows:
- particles diffuse according to independent (time-inhomogeneous) Brownian motions, sped up by a factor $\sigma^2(t/T)$ at time t
- particles split, or "branch" into two particles at (constant) rate $1/2$
- at every branching event, only keep N particles at highest positions



Eric Brunet

M_T: maximum position at time T.



Running maximum of simulations of time-inhomogeneous N-BBM with varying N.
Parameters: $T = 1000$, $\sigma(t) = 0.125 + t^2$.

Time-inhomogeneous N-BBM: main result

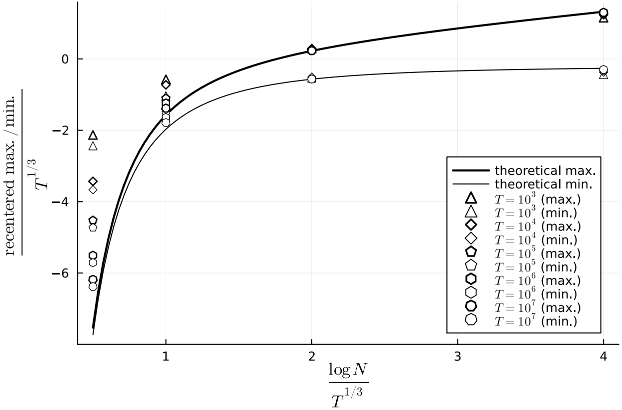
Assume σ^2 smooth, bounded away from 0 and ∞. Set $v := \int_0^1 \sigma(t)\,dt$.

Theorem (Legrand-M. (2024+))

1. *(subcritical phase)* $\log N \ll T^{1/3}$:
$$M_T = vT\left(1 - \frac{\pi^2}{2(\log N)^2}\right) + \cdots.$$
2. *(supercritical phase)* $\log N \gg T^{1/3}$:
$$M_T = vT + \int_0^1 (\sigma'(t))^+\,dt \times \log N + \cdots$$
3. *(critical phase)* $\log N \asymp T^{1/3}$:
$$M_T = vT + \Phi((\log N)/T^{1/3}; \sigma)T^{1/3} + \cdots,$$
for some explicit function $\Phi(\cdot;\sigma)$.

Same result holds also for CREM.

Numerical experiments



Numerical experiments on a discrete model (time-inhomogeneous N-particle branching random walk with Bernoulli increments) with varying N.

Subcritical phase ($\log N \ll T^{1/3}$)

Recall result:

$$M_T = vT\left(1 - \frac{\pi^2}{2(\log N)^2}\right) + \cdots$$

Reminiscent of Brunet-Derrida correction.

Theorem (Brunet-Derrida 1997, Bérard-Gouéré 2010)

Assume $\sigma^2 \equiv 1$ (homogeneous N-BBM). Then,

$$\lim_{T\to\infty} \frac{M_T}{T} = 1 - \frac{\pi^2}{2(\log N)^2} + \cdots .$$

Time-inhomogeneous N-BBM behaves like a concatenation of homogeneous N-BBM living each on a time scale of order $o(T)$.

Supercritical phase ($\log N \gg T^{1/3}$)

Recall result:

$$M_T = vT + \int_0^1 (\sigma'(t))^+ \, dt \times \log N + \cdots$$

Why second term of order $\log N$?

- Particles in the N-BBM are atypical (large deviation event needed for a trajectory to survive)
- As a consequence, particle density decreases exponentially.
- When N large enough, expect a logarithmic increase in the maximum as a function of N.

Critical phase ($\log N \asymp T^{1/3}$)

Recall result:

$$M_T = vT + \Phi((\log N)/T^{1/3};\sigma) \times T^{1/3} + \cdots ,$$

for some explicit functional $\Phi(\cdot;\sigma)$.

Why $T^{1/3}$? Match corrections in subcritical and supercritical phases:

$$\frac{T}{(\log N)^2} \asymp \log N \iff \log N \asymp T^{1/3}$$

Expression of $\Phi(\cdot;\sigma)$ involving a function Ψ defined in Mallein 2015:

$$\Phi(\alpha;\sigma) = \int_0^1 \frac{\sigma(u)}{\alpha^2}\Psi\Big(-\alpha^3\frac{\sigma'(u)}{\sigma(u)}\Big)\mathrm{d}u, \quad \Psi(q)\begin{cases}\sim -q, & q\to-\infty\\ = -\frac{\pi^2}{2}, & q=0\\ \sim -\frac{\mathrm{a}_1 q^{2/3}}{2^{1/3}}, & q\to+\infty\end{cases},$$

where $-\mathrm{a}_1 = -2.33811\ldots$ is the largest root of the Airy function Ai.

$T^{1/3}$ scaling in branching Brownian motion

$T^{1/3}$ scaling appears in many articles involving extremal particles of branching Brownian motion/branching random walks

Kesten 1978, Aldous 1998, Pemantle 2009, Fang–Zeitouni 2010, Faraud–Hu–Shi 2012, Jaffuel 2012, Mallein 2015, M.–Zeitouni 2016,...

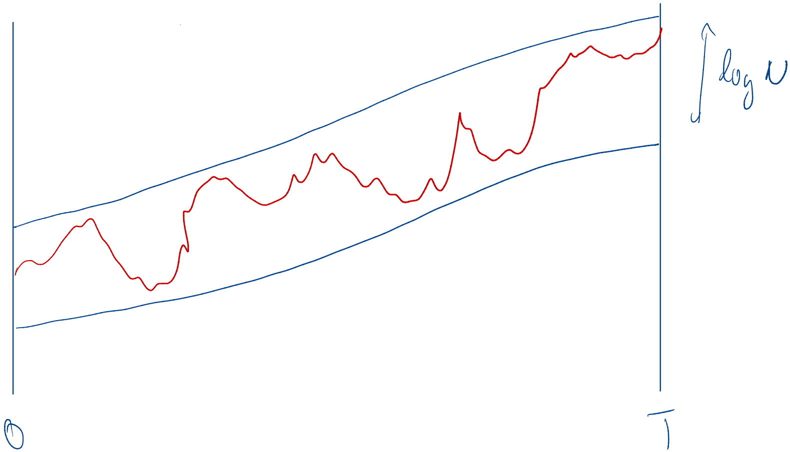
But it appears to our knowledge here for the first time for non-extremal particles.

Proof methods

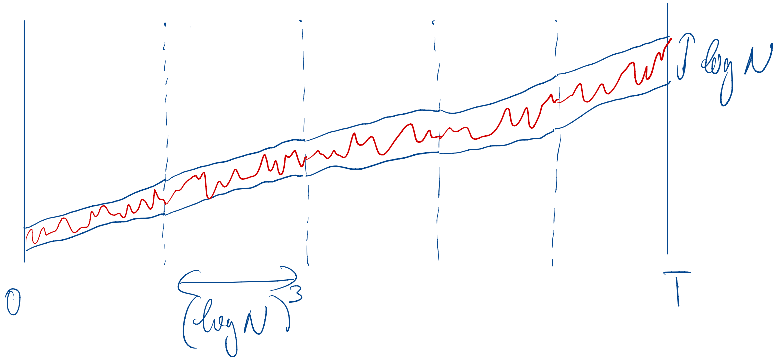
Guiding principle: make use of trajectories!

1. Comparison of the N-BBM with BBM killed outside some well-chosen space-time tube ("barrier method"), over time scale T (critical, supercritical phases) or over a smaller time scale (subcritical phase)
2. Estimates on number of particles staying inside such tubes (truncation!) through first- and second moment estimates
3. Critical phase: eigenvalue problem of Laplacian killed outside an interval in a linear (Airy) potential, Mallein 2015

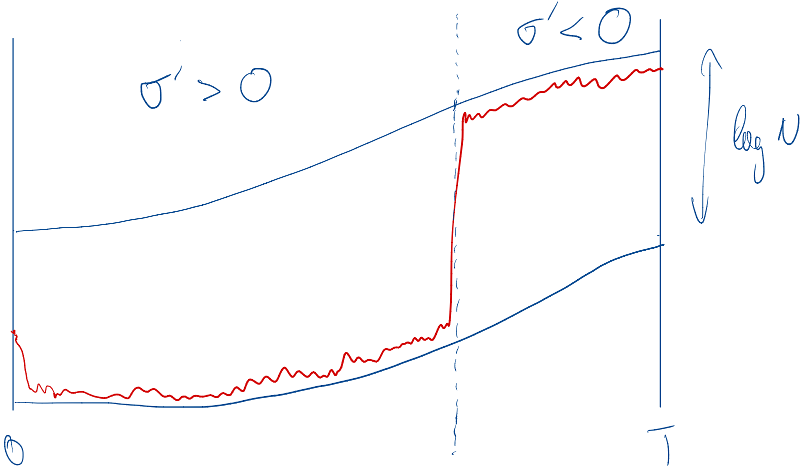
$\log N \simeq T^{1/3}$



$\log N \ll T^{1/3}$



$$\log N \gg T^{1/3}$$



Conclusion

- We have introduced a beam search algorithm for the CREM and a continuous-time counterpart.
- We have rigorously studied the performance of the algorithm when T and the beam width N are large
- Critical phase: $\log N \asymp T^{1/3}$. Below this critical phase, the gain in the performance when increasing the beam width is notable, above the critical phase it becomes negligible (logarithmic increase in N)
- Results quite precise, but still rough for BBM standards.

Open problems:

- Prove algorithmic lower bound for a wide class of algorithms
- Study similar behavior in “true” models (spin glasses, combinatorial optimization,...)

Thank you for your attention!