Algorithmic Spin Glass Theory

Andrea Montanari

Stanford University

June 22, 2022

Ahmed El Alaoui

Mark Sellke

Outline

FOCS 2019, SIAM J on Computing 2020 (arXiv:1812.10897)

Annals of Probability, 2021 (arXiv:2001.00904, w/ El Alaoui, Sellke)

arXiv:2009.11481 (w/ El Alaoui)

An algorithmic question

Sherrington-Kirkpatrick model

$$H_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}}\langle \boldsymbol{\sigma}, \boldsymbol{G}\boldsymbol{\sigma}\rangle\,,$$

$$G_{i,j} \sim_{iid} \mathsf{N}(0,1)\,, \qquad \boldsymbol{\sigma} \in \{+1,-1\}^N\,.$$

Equivalently, centered Gaussian process on $\{+1,-1\}^N$:

$$\mathbb{E}\{H_N(\boldsymbol{\sigma}_1)H_N(\boldsymbol{\sigma}_2)\} = \frac{\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\rangle^2}{N}\,.$$

Sherrington-Kirkpatrick model

$$H_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}} \langle \boldsymbol{\sigma}, \boldsymbol{G}\boldsymbol{\sigma} \rangle \,,$$

$$G_{i,j} \sim_{iid} \mathsf{N}(0,1)\,, \qquad \boldsymbol{\sigma} \in \{+1,-1\}^N \,.$$

Equivalently, centered Gaussian process on $\{+1,-1\}^N$:

$$\mathbb{E}\{H_N(\boldsymbol{\sigma}_1) H_N(\boldsymbol{\sigma}_2)\} = \frac{\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle^2}{N} \,.$$

General p-spin models

$$H_N(\boldsymbol{\sigma}) = \sum_{k=2}^{k_{\max}} \frac{c_k}{N^{(k-1)/2}} \langle \boldsymbol{G}^{(k)}, \boldsymbol{\sigma}^{\otimes k} \rangle$$

$$\langle \boldsymbol{G}^{(k)}, \boldsymbol{\sigma}^{\otimes k} \rangle \equiv \sum_{1 \le i_1, \cdots, i_k \le N} G^{(k)}_{i_1, \cdots, i_k} \sigma_{i_1} \cdots \sigma_{i_k}$$

$$G^{(k)}_{i_1, \cdots, i_k} \sim \mathsf{N}(0, 1)$$

Equivalently, centered Gaussian process on $\{+1, -1\}^N$:

$$\mathbb{E}\{H_N(\boldsymbol{\sigma}_1) H_N(\boldsymbol{\sigma}_2)\} = N \, \xi(\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle / N) \,,$$

$$\xi(t) := \sum_k c_k^2 t^k \,.$$

General p-spin models

$$H_N(\boldsymbol{\sigma}) = \sum_{k=2}^{k_{\max}} \frac{c_k}{N^{(k-1)/2}} \langle \boldsymbol{G}^{(k)}, \boldsymbol{\sigma}^{\otimes k} \rangle$$

$$\langle \boldsymbol{G}^{(k)}, \boldsymbol{\sigma}^{\otimes k} \rangle \equiv \sum_{1 \le i_1, \cdots, i_k \le N} G^{(k)}_{i_1, \cdots, i_k} \sigma_{i_1} \cdots \sigma_{i_k}$$

$$G^{(k)}_{i_1, \cdots, i_k} \sim \mathsf{N}(0, 1)$$

Equivalently, centered Gaussian process on $\{+1, -1\}^N$:

$$\mathbb{E}\{H_N(\boldsymbol{\sigma}_1) H_N(\boldsymbol{\sigma}_2)\} = N\, \xi(\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle / N)\,,$$

$$\xi(t) := \sum_k c_k^2 t^k\,.$$

Optimum value of the Hamiltonian[1]

$$H_N(\boldsymbol{\sigma}) = \sum_{k=2}^{k_{\max}} \frac{c_k}{N^{(k-1)/2}} \langle \boldsymbol{G}^{(k)}, \boldsymbol{\sigma}^{\otimes k} \rangle \,,$$

$$\mathsf{OPT}_N := \frac{1}{N} \mathbb{E} \max_{\boldsymbol{\sigma} \in \{+1,-1\}^n} H_N(\boldsymbol{\sigma}) \,.$$

Gaussian concentration: With high probability

$$\frac{1}{N} \max_{\boldsymbol{\sigma} \in \{+1,-1\}^n} H_N(\boldsymbol{\sigma}) = \mathsf{OPT}_N + O(N^{-1/2}) \,.$$

[1] 'Ground state energy'

Parisi formula

$\gamma : [0,1) \to \mathbb{R}_{\geq 0}$ non-decreasing, right continuous ($\gamma \in \mathscr{U}([0,1])$)

$$\partial_t \Phi_\gamma(t,x) + \frac{1}{2}\xi''(t)\Big(\partial_x^2 \Phi_\gamma(t,x) + \gamma(t)(\partial_x \Phi(t,x))^2\Big) = 0\,, \quad \Phi_\gamma(1,x) = |x|\,,$$

$$\mathsf{P}(\gamma) \equiv \Phi_\gamma(0,0) - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)\,\mathrm{d}t\,.$$

Theorem (Talagrand 2006; Panchenko, 2013; Auffinger, Chen, 2017)

$$\lim_{N\to\infty} \mathsf{OPT}_N = \mathsf{OPT} = \inf_{\gamma \in \mathscr{U}} \mathsf{P}(\gamma)\,.$$

Parisi formula

$\gamma : [0,1) \to \mathbb{R}_{\geq 0}$ non-decreasing, right continuous ($\gamma \in \mathscr{U}([0,1])$)

$$\partial_t \Phi_\gamma(t,x) + \frac{1}{2}\xi''(t)\Big(\partial_x^2 \Phi_\gamma(t,x) + \gamma(t)(\partial_x \Phi(t,x))^2\Big) = 0\,, \quad \Phi_\gamma(1,x) = |x|\,,$$
$$\mathsf{P}(\gamma) \equiv \Phi_\gamma(0,0) - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)\,\mathrm{d}t\,.$$

Theorem (Talagrand 2006; Panchenko, 2013; Auffinger, Chen, 2017)

$$\lim_{N\to\infty} \mathsf{OPT}_N = \mathsf{OPT} = \inf_{\gamma \in \mathscr{U}} \mathsf{P}(\gamma)\,.$$

Question

Algorithm

- Input $(\boldsymbol{G}^{(k)})_{k\le k_{\max}}$
- Output $\boldsymbol{\sigma}^{\rm alg}\in\{+1,-1\}^N$
- Approximation $\rho\in(0,1)$:

$$\lim_{N\to\infty}\mathbb{P}\Big(H_N(\boldsymbol{\sigma}^{\rm alg})\ge\rho\cdot\max_{\boldsymbol{\sigma}\in\{+1,-1\}}H_N(\boldsymbol{\sigma})\Big)=1$$

Sufficient to compute ALG s.t. $H_N(\boldsymbol{\sigma}^{\rm alg})/N=\mathsf{ALG}+o_P(1)$.

CS Theory: Worst case — SK model

$$\begin{aligned} \text{maximize} \quad & H_{\mathbf{N}}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{2N}} \langle \boldsymbol{\sigma}, \boldsymbol{G}\boldsymbol{\sigma} \rangle \,, \\ \text{subj. to} \quad & \boldsymbol{\sigma} \in \{+1, -1\}^N \,. \end{aligned}$$

- NP-hard to approximate within $O(\log^\gamma N)$

  [Arora, Berger, Hazan, Kindler, Safra, 2005]

- Grothendieck ineq: $O(\log N)$ approximation

  [Charikar, Wirth, 2004]

CS Theory: Average case

- Exponential # local maxima

[Addario-Berry et al. 2017]

- Spectral relaxation

$$\frac{1}{\sqrt{2N^3}} \max_{\|\boldsymbol{\sigma}\|_2^2=N} \langle \boldsymbol{\sigma}, \boldsymbol{G}\boldsymbol{\sigma} \rangle \to 1$$

- Semidefinite Programming Relaxation

$$\frac{1}{\sqrt{2N^3}} \max \Big\{ \sum_{i,j=1}^{N} G_{ij} \langle \boldsymbol{s}_i, \boldsymbol{s}_j \rangle , \;\; \boldsymbol{s}_i \in \mathbb{S}^{N-1} \Big\} \to 1 \, ,$$

$$\frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{GW}}) \to \frac{2}{\pi} \approx 0.636619$$

M, Sen, 2016

OPT ≈ 0.763166

CS Theory: Average case

- Exponential # local maxima

[Addario-Berry et al. 2017]

- Spectral relaxation

$$\frac{1}{\sqrt{2N^3}} \max_{\|\boldsymbol{\sigma}\|_2^2=N} \langle \boldsymbol{\sigma}, \boldsymbol{G}\boldsymbol{\sigma} \rangle \to 1$$

- Semidefinite Programming Relaxation

$$\frac{1}{\sqrt{2N^3}} \max \Big\{ \sum_{i,j=1}^{N} G_{ij} \langle \mathbf{s}_i, \mathbf{s}_j \rangle \, , \;\; \mathbf{s}_i \in \mathbb{S}^{N-1} \Big\} \to 1 \, ,$$

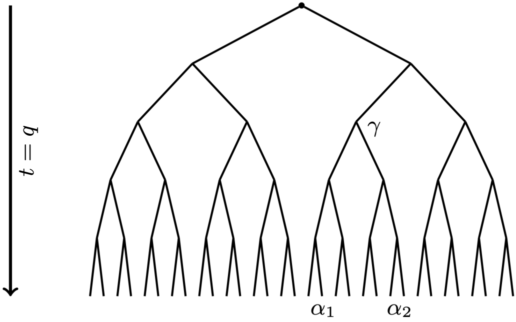
$$\frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{GW}}) \to \frac{2}{\pi} \approx 0.636619$$

M, Sen, 2016

OPT ≈ 0.763166

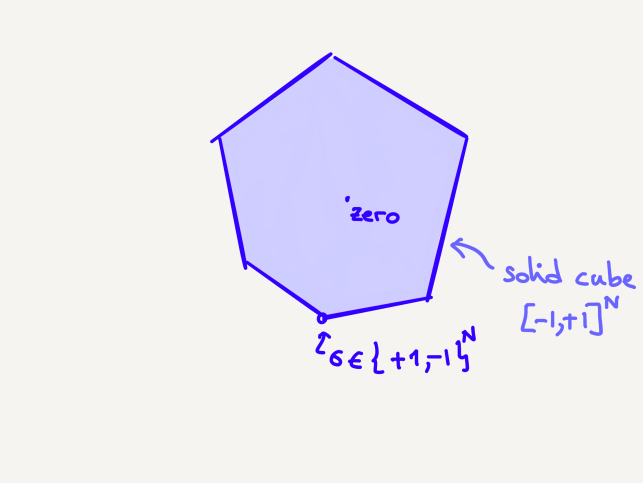
Insights from replica symmetry breaking

Replica symmetry breaking

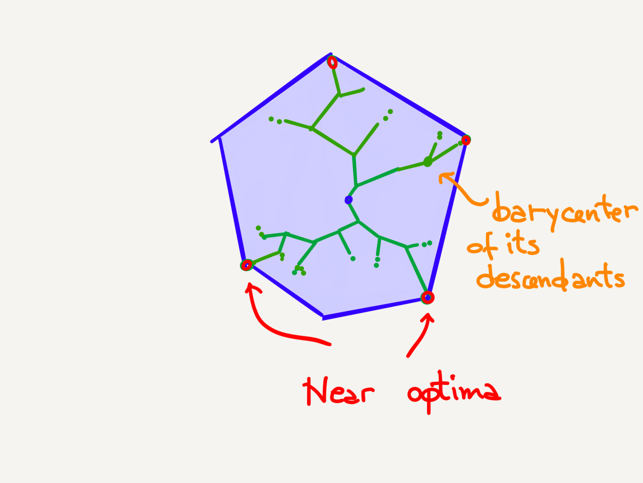


- (Many) Near optima on leaves of a (complete, balanced, metric) tree.
- All nodes at same generation same distance from the root.
- $\|\boldsymbol{\sigma}^{(\alpha)} - \boldsymbol{\sigma}^{(\alpha')}\|_2^2 \approx N\, d_T(\alpha, \alpha')$

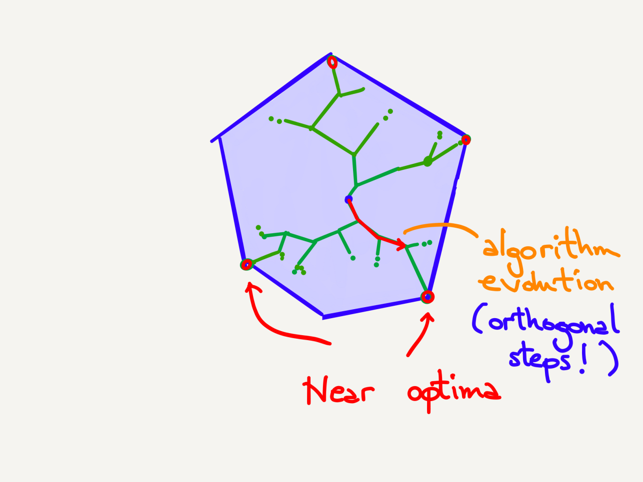
$\{-1, +1\}^N \subseteq [-1, +1]^N$



Tree $\subseteq [-1, +1]^N$



Algorithm Path $\subseteq [-1, +1]^N$



- **Intuition:** Efficient algorithm if tree branches 'continuously'

Algorithm Path $\subseteq [-1, +1]^N$

Algorithm 1: Cartoon

Data: Hamiltonian H_N

Result: Integer solution $\boldsymbol{\sigma}^{\text{alg}} \in \{+1, -1\}^N$

Initialize $\boldsymbol{m}^0 = \boldsymbol{0}$;

for $t \in \{0, \delta, \dots, 1-\delta\}$ **do**

  Compute $\boldsymbol{m}^{t+\delta} = F(\boldsymbol{m}^t, \nabla H_N(\boldsymbol{m}^t), \nabla^2 H_N(\boldsymbol{m}^t))$;

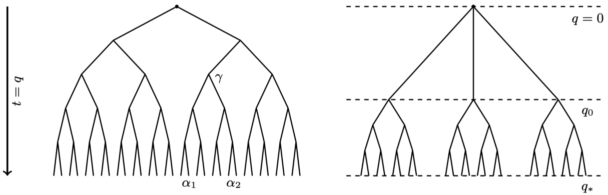
  Ensure that $\|\boldsymbol{m}^{t+\delta}\|_2^2 = N(t+\delta)$, $\|\boldsymbol{m}\|_\infty \leq 1$;

end

return Round$(\boldsymbol{m}^{t=1})$;

- **Intuition:** Efficient algorithm if tree branches ‘continuously’

Possible tree structures



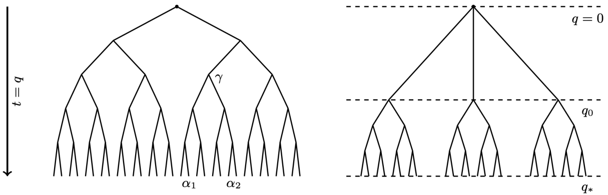
Theorem (Auffinger,Chen 2018)

Assume ξ even, and let $\gamma_* := \arg\min_{\gamma \in \mathscr{U}} \mathsf{P}(\gamma)$. If $q \in [0, 1]$ point of increase of γ_*, then $\forall \varepsilon > 0$, with high probability there exists $\boldsymbol{\sigma}^{(1)}$, $\boldsymbol{\sigma}^{(2)}$ s.t.

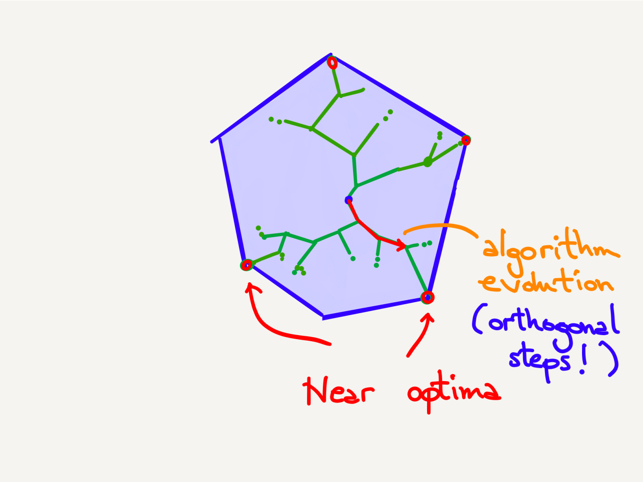
$$H_N(\boldsymbol{\sigma}^{(\alpha)}) \geq N(\mathsf{OPT} - \varepsilon), \quad \alpha \in \{1, 2\},$$

$$\frac{1}{N} \langle \boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)} \rangle \in [q - \varepsilon, q + \varepsilon].$$

Possible tree structures



Continuous tree $\approx$ Strictly increasing γ_*



This idea can be made precise and proved to work!

Algorithms of this type

- Subag, 2018: Spherical model $\boldsymbol{\sigma} \in \mathbb{S}^{N-1}$ (Hessian ascent)
- M 2019: SK model, $\boldsymbol{\sigma} \in \{+1, -1\}^N$ (AMP)
- El Alaoui, M, Sellke 2020: Mixed p-spin model $\boldsymbol{\sigma} \in \{+1, -1\}^N$ (AMP)
- El Alaoui, Sellke 2021: Perceptron (AMP)
- El Alaoui, M, Sellke 2022: Maxcut/Min Bisection (MP)
- …

Algorithms of this type

- Subag, 2018: Spherical model $\boldsymbol{\sigma} \in \mathbb{S}^{N-1}$ (Hessian ascent)
- **M 2019: SK model, $\boldsymbol{\sigma} \in \{+1, -1\}^N$ (AMP)**
- **El Alaoui, M, Sellke 2020: Mixed *p*-spin model $\boldsymbol{\sigma} \in \{+1, -1\}^N$ (AMP)**
- El Alaoui, Sellke 2021: Perceptron (AMP)
- El Alaoui, M, Sellke 2022: Maxcut/Min Bisection (MP)
- …

Algorithm analysis

Strategy

- General class of algorithms parametrized by $v, g : \mathbb{R} \times [0, 1] \to \mathbb{R}$
- Exact analysis as $N \to \infty$: Stochastic differential equation.
- Determine optimal v, g: Stochastic optimal control.

General algorithm (SK: $H_N(\boldsymbol{\sigma}) = \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle/2$, $\boldsymbol{A} = (\boldsymbol{G} + \boldsymbol{G}^{\mathsf{T}})/\sqrt{2N}$)

Structure

$$\boldsymbol{b}(t+\delta) = \boldsymbol{A} f_t(\boldsymbol{b}_{\le t}) - \sum_{s\le t, s\in T(\delta)} \mathsf{c}(s) f_s(\boldsymbol{b}_{\le s})$$

$$\boldsymbol{m}(t) = f_t(\boldsymbol{b}_{\le t}),$$

$$\boldsymbol{b}_{\le t} = [\boldsymbol{b}(0), \dots, \boldsymbol{b}(t)], \quad f_t(\boldsymbol{b}_{\le t})_i \equiv f_t(b_i(0), \dots, b_i(t))$$

- State $\boldsymbol{b}(t) \in \mathbb{R}^N$
- $t \in T(\delta) \equiv \{0, \delta, 2\delta, \dots, 1\}$
- Complexity $O(N^2/\delta)$.
- Deterministic constants $(\mathsf{c}(s))_{s\in T(\delta)}$: explicitly given.

Key insight #1: State evolution (for suitable $\mathsf{c}(s)$)

$$\boldsymbol{b}(t+\delta) = \boldsymbol{A} f_t(\boldsymbol{b}_{\le t}) - \sum_{s\in[0,t)\cap T(\delta)} \mathsf{c}(s) f_s(\boldsymbol{b}_{\le s})$$

$$(b_i(t))_{t\in T(\delta)} \approx \mathsf{N}(0, \boldsymbol{Q})\,,$$

$$\frac{1}{n}\langle \boldsymbol{b}_i(t+\delta), \boldsymbol{b}(s+\delta)\rangle \approx \frac{1}{n}\langle f_t(\boldsymbol{b}_{\le t}) f_s(\boldsymbol{b}_{\le s})\rangle\,.$$

Bolthausen 2014, Bayati, Montanari 2011, Javanmard, Montanari, 2013

Key insight #1: State evolution (for suitable $\mathsf{c}(s)$)

$$\boldsymbol{b}(t+\delta) = \boldsymbol{A} f_t(\boldsymbol{b}_{\le t}) - \sum_{s\in[0,t)\cap T(\delta)} \mathsf{c}(s) f_s(\boldsymbol{b}_{\le s})$$

$$(b_i(t))_{t\in T(\delta)} \approx \mathsf{N}(0,\boldsymbol{Q})\,,$$

$$\frac{1}{\boldsymbol{n}}\langle \boldsymbol{b_i}(t+\delta), \boldsymbol{b}(s+\delta)\rangle \approx \frac{1}{\boldsymbol{n}}\langle f_t(\boldsymbol{b}_{\le t}) f_s(\boldsymbol{b}_{\le s})\rangle\,.$$

Bolthausen 2014, Bayati, Montanari 2011, Javanmard, Montanari, 2013

Key insight #2: Orthogonal updates

$$\frac{1}{\boldsymbol{n}}\langle \boldsymbol{b_i}(t+\delta), \boldsymbol{b}(s+\delta)\rangle \approx \frac{1}{\boldsymbol{n}}\langle f_t(\boldsymbol{b}_{\le t}), f_s(\boldsymbol{b}_{\le s})\rangle\,.$$

- Let $\mathrm{d}\boldsymbol{b}(t) = \boldsymbol{b}(t) - \boldsymbol{b}(t-\delta)$.
- Want: $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx 0$ for all $t \in T(\delta)$
  - Assume $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(s+\delta), \boldsymbol{b}(s)\rangle \approx 0$ for $s \le t-\delta$
  - $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx \frac{1}{n}\langle \mathrm{d}f_t(\boldsymbol{b}_{\le t}), f_{t-\delta}(\boldsymbol{b}_{\le t-\delta})\rangle, \quad \mathrm{d}f_t \equiv f_t - f_{t-\delta}$
  - Define $\mathrm{d}f_t(\boldsymbol{b}_{\le t}) \equiv \tilde{g}(\boldsymbol{b}_{\le t-\delta}) \odot \mathrm{d}\boldsymbol{b}(t)$.

Key insight #2: Orthogonal updates

$$\frac{1}{n}\langle \boldsymbol{b}_i(t+\delta), \boldsymbol{b}(s+\delta)\rangle \approx \frac{1}{n}\langle f_t(\boldsymbol{b}_{\le t}), f_s(\boldsymbol{b}_{\le s})\rangle\,.$$

- Let $\mathrm{d}\boldsymbol{b}(t) = \boldsymbol{b}(t) - \boldsymbol{b}(t-\delta)$.
- Want: $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx 0$ for all $t \in T(\delta)$
  - Assume $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(s+\delta), \boldsymbol{b}(s)\rangle \approx 0$ for $s \le t-\delta$
  - $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx \frac{1}{n}\langle \mathrm{d}f_t(\boldsymbol{b}_{\le t}), f_{t-\delta}(\boldsymbol{b}_{\le t-\delta})\rangle, \quad \mathrm{d}f_t \equiv f_t - f_{t-\delta}$
  - Define $\mathrm{d}f_t(\boldsymbol{b}_{\le t}) \equiv \tilde{g}(\boldsymbol{b}_{\le t-\delta}) \odot \mathrm{d}\boldsymbol{b}(t)$.

Key insight #2: Orthogonal updates

$$\frac{1}{n}\langle \boldsymbol{b}_i(t+\delta), \boldsymbol{b}(s+\delta)\rangle \approx \frac{1}{n}\langle f_t(\boldsymbol{b}_{\le t}), f_s(\boldsymbol{b}_{\le s})\rangle \,.$$

- Let $\mathrm{d}\boldsymbol{b}(t) = \boldsymbol{b}(t) - \boldsymbol{b}(t-\delta)$.
- Want: $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx 0$ for all $t \in T(\delta)$
  - Assume $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(s+\delta), \boldsymbol{b}(s)\rangle \approx 0$ for $s \le t-\delta$
  - $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx \frac{1}{n}\langle \mathrm{d}f_t(\boldsymbol{b}_{\le t}), f_{t-\delta}(\boldsymbol{b}_{\le t-\delta})\rangle, \quad \mathrm{d}f_t \equiv f_t - f_{t-\delta}$
  - Define $\mathrm{d}f_t(\boldsymbol{b}_{\le t}) \equiv \tilde{g}(\boldsymbol{b}_{\le t-\delta}) \odot \mathrm{d}\boldsymbol{b}(t)$.

Key insight #2: Orthogonal updates

$$\frac{1}{n}\langle \boldsymbol{b}_i(t+\delta), \boldsymbol{b}(s+\delta)\rangle \approx \frac{1}{n}\langle f_t(\boldsymbol{b}_{\le t}), f_s(\boldsymbol{b}_{\le s})\rangle \,.$$

- Let $\mathrm{d}\boldsymbol{b}(t) = \boldsymbol{b}(t) - \boldsymbol{b}(t-\delta)$.
- Want: $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx 0$ for all $t \in T(\delta)$
  - Assume $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(s+\delta), \boldsymbol{b}(s)\rangle \approx 0$ for $s \le t - \delta$
  - $\frac{1}{n}\langle \mathrm{d}\boldsymbol{b}(t+\delta), \boldsymbol{b}(t)\rangle \approx \frac{1}{n}\langle \mathrm{d}f_t(\boldsymbol{b}_{\le t}), f_{t-\delta}(\boldsymbol{b}_{\le t-\delta})\rangle, \quad \mathrm{d}f_t \equiv f_t - f_{t-\delta}$
  - Define $\mathrm{d}f_t(\boldsymbol{b}_{\le t}) \equiv \tilde{g}(\boldsymbol{b}_{\le t-\delta}) \odot \mathrm{d}\boldsymbol{b}(t)$.

Achievability for AMP algorithm

$N \to \infty, \delta \to 0$: one dimensional stochastic differential equation

$$\mathrm{d}X_t = v(t, X_t)\,\mathrm{d}t + \sqrt{\xi''(t)}\mathrm{d}B_t\,, \quad \text{with } X_0 = 0\,,$$
$$M_t = \int_0^t \sqrt{\xi''(s)}\,u(s, X_s)\,\mathrm{d}B_s\,.$$

$m_i(t) \stackrel{\mathrm{d}}{\approx} M_t$, $b_i(t) \stackrel{\mathrm{d}}{\approx} B_t$,

Achievability for AMP algorithm

$$\mathrm{d}X_t = v(t, X_t)\,\mathrm{d}t + \sqrt{\xi''(t)}\mathrm{d}B_t\,, \quad \text{with } X_0 = 0\,,$$
$$M_t = \int_0^t \sqrt{\xi''(s)}\,u(s, X_s)\,\mathrm{d}B_s\,.$$

Theorem (El Alaoui, M, Sellke, 2020)

Under regularity conditions on u, v, further assume that $M_1 \in [-1, 1]$ almost surely and $\mathbb{E}[M_t^2] = t$ for all $t \in [0, 1)$.
Then there exists an algorithm with complexity $N^{k_{\max}}/\varepsilon^2$, such that, whp

$$\frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{alg}}) \geq \int_0^1 \xi''(t)\mathbb{E}\{u(t, X_t)\}\,dt - \varepsilon\,,$$

Stochastic optimal control problem

$$
\begin{aligned}
\text{maximize} \quad & \int_0^1 \xi''(t)\mathbb{E}\{u(t,X_t)\}\,\mathrm{d}t \\
\text{subj. to} \quad & \mathbb{E}[M_t^2] = t\,, \quad M_1 \in [-1,1]\,, \\
& \mathrm{d}X_t = v(t,X_t)\,\mathrm{d}t + \sqrt{\xi''(t)}\mathrm{d}B_t\,, \\
& M_t = \int_0^t \sqrt{\xi''(s)}\,u(s,X_s)\,\mathrm{d}B_s\,.
\end{aligned}
$$

- Decision variables $u, v : [0,1] \times \mathbb{R} \to \mathbb{R}$

Stochastic optimal control problem

$$
\begin{aligned}
\text{maximize} \quad & \int_0^1 \xi''(t)\mathbb{E}\{U_t\}\,\mathrm{d}t \\
\text{subj. to} \quad & \mathbb{E}[M_t^2] = t\,, \quad M_1 \in [-1,1]\,, \\
& M_t = \int_0^t \sqrt{\xi''(s)}\,U_s\,\mathrm{d}B_s\,.
\end{aligned}
$$

- This is dual to a modified Parisi formula!

Main result

Theorem (M, 2019; El Alaoui, M, Sellke, 2021)

Let $\mathscr{L} := \{\gamma : [0,1) \to \mathbb{R}_{\geq 0} : \|\xi''\gamma\|_{\mathrm{TV}[0,t]} < \infty\}$*. Assume that the infimum* $\inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$ *is achieved at a function* $\gamma_* \in \mathscr{L}$*.*
Then there exists an algorithm with complexity $C(\varepsilon)N^{\boldsymbol{k}_{\max}}$ *such that, whp*

$$\operatorname*{p-lim}_{N\to\infty} \frac{1}{N} H_N(\boldsymbol{\sigma}^{\mathrm{alg}}) =: \mathsf{ALG} = \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)\,,$$

Comparison

Optimum value

$$\mathsf{OPT} = \inf_{\gamma \in \mathscr{U}([0,1])} \mathsf{P}(\gamma)$$

$$\mathscr{U} := \mathscr{L} \cap \{\gamma \text{ nondecreasing}\},.$$

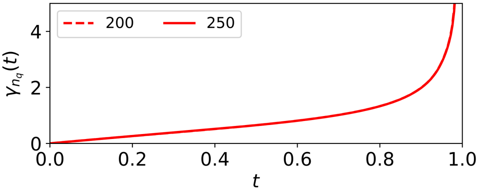
Algorithmic value

$$\mathsf{ALG} = \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)\,,$$

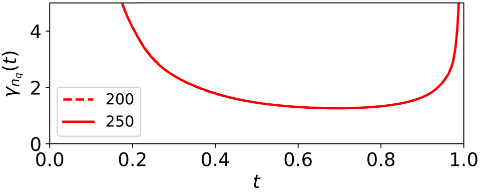
$$\mathscr{L} := \{\gamma : [0,1) \to \mathbb{R}_{\geq 0} : \|\xi''\gamma\|_{\mathrm{TV}[0,t]} < \infty\},.$$

$\gamma_*^{\mathscr{L}} := \arg\min_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$

$\xi(t) = t^2$



$\xi(t) = t^3$



Hardness

Lipschitz algorithms

Input:

$$H_N = (\boldsymbol{G}^{(2)}, \boldsymbol{G}^{(3)}, \dots, \boldsymbol{G}^{(\boldsymbol{k}_{\max})})$$

Algorithm:

$$\mathcal{A} : H_N \mapsto \mathcal{A}(H_N) \in [-1, +1]^N$$

Lipschitz algorithm:

$$\frac{1}{\sqrt{N}} \|\mathcal{A}(H_N) - \mathcal{A}(H_N')\|_2 \le L\, \|H_N - H_N'\|_2$$

Lipschitz algorithms

Input:

$$H_N = (\boldsymbol{G}^{(2)}, \boldsymbol{G}^{(3)}, \dots, \boldsymbol{G}^{(k_{\max})})$$

Algorithm:

$$\mathcal{A} : H_N \mapsto \mathcal{A}(H_N) \in [-1, +1]^N$$

Lipschitz algorithm:

$$\frac{1}{\sqrt{N}} \|\mathcal{A}(H_N) - \mathcal{A}(H_N')\|_2 \le L \, \|H_N - H_N'\|_2$$

Lipschitz algorithms

Input:

$$H_N = (\boldsymbol{G}^{(2)}, \boldsymbol{G}^{(3)}, \dots, \boldsymbol{G}^{(\boldsymbol{k}_{\max})})$$

Algorithm:

$$\mathcal{A} : H_N \mapsto \mathcal{A}(H_N) \in [-1, +1]^N$$

Lipschitz algorithm:

$$\frac{1}{\sqrt{N}} \|\mathcal{A}(H_N) - \mathcal{A}(H_N')\|_2 \le L \, \|H_N - H_N'\|_2$$

Lipschitz algorithms

- Gradient descent for $O(1)$ iterations
- Langevin dynamics for time $O(1)$
- Message passing algorithms for $O(1)$ iterations

Theorem (Huang, Sellke, 2021)

Let $\boldsymbol{m}^{\rm alg}$ be the output of a Lipschits algorithm. Then

$$\frac{1}{N} H_N(\boldsymbol{m}^{\rm alg}) \le \mathsf{ALG} + o_P(1)\,,$$

$$\mathsf{ALG} = \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)\,.$$

Lipschitz algorithms

- Gradient descent for $O(1)$ iterations
- Langevin dynamics for time $O(1)$
- Message passing algorithms for $O(1)$ iterations

Theorem (Huang, Sellke, 2021)

Let $\boldsymbol{m}^{\rm alg}$ be the output of a Lipschits algorithm. Then

$$\frac{1}{N} H_N(\boldsymbol{m}^{\rm alg}) \le \mathsf{ALG} + o_P(1)\,,$$

$$\mathsf{ALG} = \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)\,.$$

Geometric intuition

- Assume algorithm $\mathcal{A}$ achieves energy density e.
- Can be used to construct $\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(M)}$, $M \gg 1$ s.t.
  - $H_N(\boldsymbol{m}^{(\alpha)})/N \approx e \;\forall \alpha \le M$
  - Geometry $\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(M)} \approx$ (any) ultrametric tree
- Impossible above ALG!

Conclusion

Conclusion

- StatPhys $\longrightarrow$ Typical properties of optima

  [Ding, Sly, Sun, 2015; Coja-Oghlan, 2016,…]

- StatPhys $\longrightarrow$ Algorithms with rigorous guarantees

  [This talk, …]

- Many other computational questions. e.g. sampling

  [El Alaoui, M, Sellke, 2022; …]

Conclusion

- StatPhys $\longrightarrow$ Typical properties of optima

  [Ding, Sly, Sun, 2015; Coja-Oghlan, 2016,…]

- **StatPhys $\longrightarrow$ Algorithms with rigorous guarantees**

  [This talk, …]

- **Many other computational questions. e.g. sampling**

  [El Alaoui, M, Sellke, 2022; …]

Conclusion

- StatPhys $\longrightarrow$ Typical properties of optima

  [Ding, Sly, Sun, 2015; Coja-Oghlan, 2016,...]

- **StatPhys $\longrightarrow$ Algorithms with rigorous guarantees**

  [This talk, ...]

- **Many other computational questions. e.g. sampling**

  [El Alaoui, M, Sellke, 2022; ...]

Acknowledgement

- I got interested in StatPhys when I was 20, reading Giorgio's textbooks.
- A never ending source of inspiration since.

Thanks for listening! Thanks Giorgio!

Acknowledgement

- I got interested in StatPhys when I was 20, reading Giorgio's textbooks.
- A never ending source of inspiration since.

Thanks for listening! Thanks Giorgio!