Algorithmic Threshold for Optimizing Spin Glasses

Brice Huang (MIT)

Cargèse Institute 2023
Statistical Physics & Machine Learning Back Together Again

Joint work with Mark Sellke (Harvard)

Mean Field Spin Glasses

Polynomials $H_N : \mathbb{R}^N \to \mathbb{R}$ with **random** coefficients, e.g. random cubic

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} \cdot \sigma_{i_1}\sigma_{i_2}\sigma_{i_3} \qquad g_{i_1,i_2,i_3} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

Mean Field Spin Glasses

Polynomials $H_N : \mathbb{R}^N \to \mathbb{R}$ with **random** coefficients, e.g. random cubic

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} \cdot \sigma_{i_1}\sigma_{i_2}\sigma_{i_3} = \frac{1}{N}\langle \boldsymbol{G}^{(3)}, \boldsymbol{\sigma}^{\otimes 3}\rangle \qquad g_{i_1,i_2,i_3} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

Mean Field Spin Glasses

Polynomials $H_N : \mathbb{R}^N \to \mathbb{R}$ with **random** coefficients, e.g. random cubic

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} \cdot \sigma_{i_1}\sigma_{i_2}\sigma_{i_3} = \frac{1}{N}\langle \boldsymbol{G}^{(3)}, \boldsymbol{\sigma}^{\otimes 3}\rangle \qquad g_{i_1,i_2,i_3} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

More generally, mix different degrees. For $\gamma_2, \gamma_3, \ldots \geq 0$,

$$H_N(\boldsymbol{\sigma}) = \sum_{p=2}^{P} \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ i.i.d. } \mathcal{N}(0,1)\text{s}$$

Mean Field Spin Glasses

Polynomials $H_N : \mathbb{R}^N \to \mathbb{R}$ with **random** coefficients, e.g. random cubic

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} \cdot \sigma_{i_1}\sigma_{i_2}\sigma_{i_3} = \frac{1}{N}\langle \boldsymbol{G}^{(3)}, \boldsymbol{\sigma}^{\otimes 3}\rangle \qquad g_{i_1,i_2,i_3} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

More generally, mix different degrees. For $\gamma_2, \gamma_3, \ldots \geq 0$,

$$H_N(\boldsymbol{\sigma}) = \sum_{p=2}^{P} \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ i.i.d. } \mathcal{N}(0,1)\text{s}$$

Goal: algorithmically optimize H_N over sphere $S_N = \sqrt{N}\mathbb{S}^{N-1}$

Mean Field Spin Glasses

Polynomials $H_N : \mathbb{R}^N \to \mathbb{R}$ with **random** coefficients, e.g. random cubic

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} \cdot \sigma_{i_1}\sigma_{i_2}\sigma_{i_3} = \frac{1}{N}\langle \boldsymbol{G}^{(3)}, \boldsymbol{\sigma}^{\otimes 3}\rangle \qquad g_{i_1,i_2,i_3} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

More generally, mix different degrees. For $\gamma_2, \gamma_3, \ldots \geq 0$,

$$H_N(\boldsymbol{\sigma}) = \sum_{p=2}^{P} \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ i.i.d. } \mathcal{N}(0,1)\text{s}$$

Goal: algorithmically optimize H_N over sphere $S_N = \sqrt{N}\mathbb{S}^{N-1}$

Mixture function: $\xi(q) = \sum_{p=2}^{P} \gamma_p^2 q^p$. Cubic above: $\xi(q) = q^3$

Motivation

- Origin: diluted magnetic alloys (Sherrington-Kirkpatrick 75)

Motivation

- Origin: diluted magnetic alloys (Sherrington-Kirkpatrick 75)
- Natural **high-dimensional, non-convex** random optimization problem

Motivation

- Origin: diluted magnetic alloys (Sherrington-Kirkpatrick 75)
- Natural **high-dimensional, non-convex** random optimization problem
- Random MaxCut and MaxSAT with many constraints (Dembo-Montanari-Sen 17, Panchenko 18)
- Tensor PCA log-likelihood in null model (Ben Arous-Mei-Montanari-Nica 17)
- Neural networks, high-dimensional statistics (Hopfield 82, Gardner-Derrida 87/88, Talagrand 00/02, Choromanska-Henaff-Mathieu-Ben Arous-LeCun 15, Ding-Sun 18, Fan-Mei-Montanari 21)

The maximum of H_N

Two basic questions for any random optimization problem:

- OPT: maximum value that **exists**?
- ALG: maximum value found by **efficient algorithm**?

The maximum of H_N

Two basic questions for any random optimization problem:

- OPT: maximum value that **exists**?
- ALG: maximum value found by **efficient algorithm**?

Theorem (Parisi 82, Talagrand 06/10, Panchenko 14, Auffinger-Chen 17)

The limiting maximum value

$$\mathrm{OPT} = \operatorname*{p-lim}_{N\to\infty} \frac{1}{N} \max_{\boldsymbol{\sigma}\in S_N} H_N(\boldsymbol{\sigma})$$

*exists and is given by the **Parisi formula*** $\mathsf{P}(\xi)$.

Efficient Optimization

- Today's goal: understand power of **efficient** algorithms $\mathcal{A}$ to optimize H_N. For $\boldsymbol{\sigma} = \mathcal{A}(H_N)$, what is max of

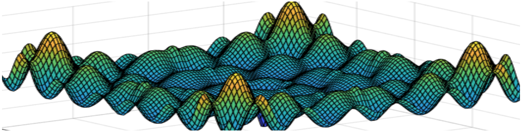
$$E = \lim_{N\to\infty} \frac{1}{N} H_N(\boldsymbol{\sigma}) \text{ ?}$$

Efficient Optimization

- Today's goal: understand power of **efficient** algorithms $\mathcal{A}$ to optimize H_N. For $\boldsymbol{\sigma} = \mathcal{A}(H_N)$, what is max of

$$E = \lim_{N \to \infty} \frac{1}{N} H_N(\boldsymbol{\sigma}) \ ?$$

- Rich landscapes challenging for algorithms
  - e^{cN} bad local maxima well below OPT (Auffinger-Ben Arous-Černy 13, Subag 17)



- Worst-case lower bounds overly pessimistic
  - Adversarial H_N: $(\log^c N)$-approximation NP-hard (ABEKS 05, BBHKSZ 12)

Informal Result

We determine sharp threshold ALG for a class of $O(1)$-**Lipschitz** algorithms

- A Lipschitz algorithm attains ALG, and this is the best known
- No Lipschitz algorithm surpasses ALG

Informal Result

We determine sharp threshold ALG for a class of $O(1)$-**Lipschitz** algorithms

- A Lipschitz algorithm attains ALG, and this is the best known
- No Lipschitz algorithm surpasses ALG

Result holds for yet more general **multi-species spin glasses**

Informal Result

We determine sharp threshold ALG for a class of $O(1)$-**Lipschitz** algorithms

- A Lipschitz algorithm attains ALG, and this is the best known
- No Lipschitz algorithm surpasses ALG

Result holds for yet more general **multi-species spin glasses**

$O(1)$-Lipschitz algorithms include:

- Gradient descent, AMP, or any constant order method for $O(1)$ rounds
- Langevin dynamics for $e^{\beta H_N}$ for $O(1)$ time

Informal Result

We determine sharp threshold ALG for a class of $O(1)$-**Lipschitz** algorithms

- A Lipschitz algorithm attains ALG, and this is the best known
- No Lipschitz algorithm surpasses ALG

Result holds for yet more general **multi-species spin glasses**

$O(1)$-Lipschitz algorithms include:

- Gradient descent, AMP, or any constant order method for $O(1)$ rounds
- Langevin dynamics for $e^{\beta H_N}$ for $O(1)$ time
- Not: low degree polynomials, SOS

Informal Result

We determine sharp threshold ALG for a class of $O(1)$-**Lipschitz** algorithms

- A Lipschitz algorithm attains ALG, and this is the best known
- No Lipschitz algorithm surpasses ALG

Result holds for yet more general **multi-species spin glasses**

$O(1)$-Lipschitz algorithms include:

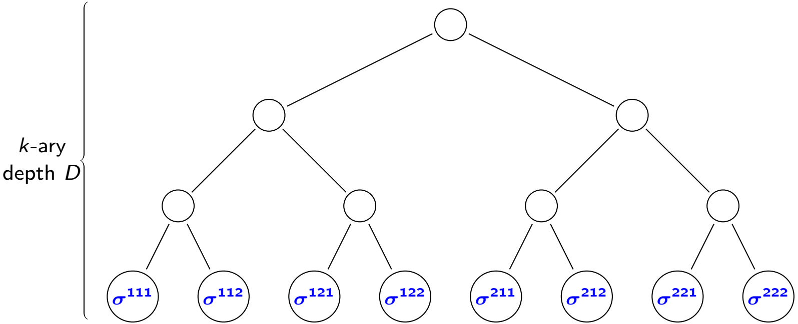
- Gradient descent, AMP, or any constant order method for $O(1)$ rounds
- Langevin dynamics for $e^{\beta H_N}$ for $O(1)$ time
- Not: low degree polynomials, SOS

Hardness result holds for more general *overlap concentrated* algorithms

Densely Branching Ultrametric Trees

Hierarchically clustered constellation of points

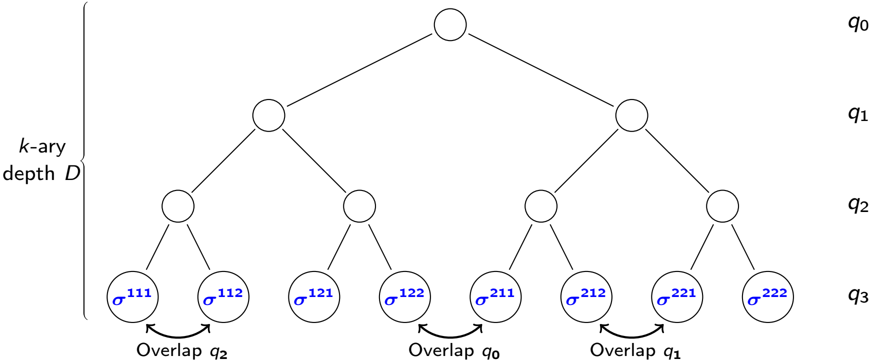
Overlap: $R(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \langle \boldsymbol{\sigma}, \boldsymbol{\rho} \rangle / N \in [-1, 1]$



Densely Branching Ultrametric Trees

Hierarchically clustered constellation of points

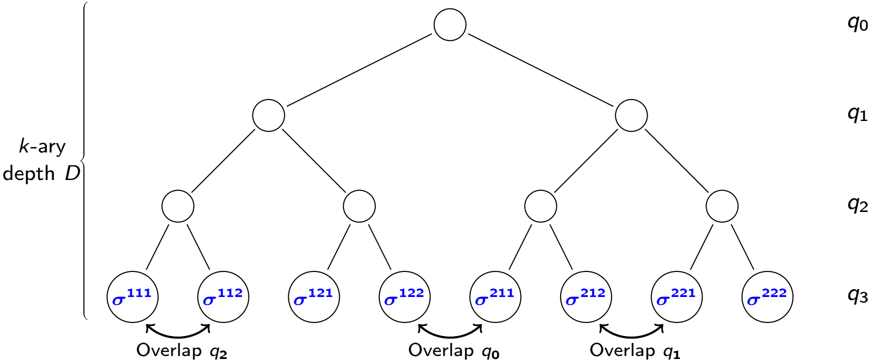
Overlap: $R(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \langle \boldsymbol{\sigma}, \boldsymbol{\rho} \rangle / N \in [-1, 1]$



Densely Branching Ultrametric Trees

Hierarchically clustered constellation of points

Overlap: $R(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \langle \boldsymbol{\sigma}, \boldsymbol{\rho} \rangle / N \in [-1, 1]$



$k, D \in \mathbb{N}$ large, $(q_0, \ldots, q_D) = (0, \frac{1}{D}, \ldots, 1)$ (**dense branching**)

Geometric description of ALG

Largest value whose super-level set contains densely branching ultrametric tree

- Tree exists $\Rightarrow$ algorithm can climb tree
- Tree absent $\Rightarrow$ hard by Branching OGP

Geometric description of ALG

Largest value whose super-level set contains densely branching ultrametric tree

- Tree exists $\Rightarrow$ algorithm can climb tree
- Tree absent $\Rightarrow$ hard by Branching OGP

- Comparison with Gibbs/OPT ultrametricity: ALG trees must branch continuously, Gibbs trees might not

Geometric description of ALG

Largest value whose super-level set contains densely branching ultrametric tree

- Tree exists $\Rightarrow$ algorithm can climb tree
- Tree absent $\Rightarrow$ hard by Branching OGP

- Comparison with Gibbs/OPT ultrametricity: ALG trees must branch continuously, Gibbs trees might not

- Consistent with algorithmic solutions forming dense well-connected cluster
  (Baldassi et. al. 15, Abbe-Li-Sly 21)

Overlap Gap Property

💡 solution geometry **clustering** $\Rightarrow$ rigorous hardness for **stable** algorithms

Overlap Gap Property

💡 solution geometry **clustering** ⇒ rigorous hardness for **stable** algorithms

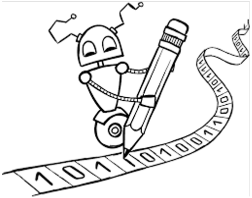
- Max ind. set (Gamarnik-Sudan 14, Rahman-Virág 17, Gamarnik-Jagannath-Wein 20, Wein 20)
- Random (NAE-)k-SAT (Gamarnik-Sudan 17, Bresler-H. 21)
- Hypergraph maxcut (Chen-Gamarnik-Panchenko-Rahman 19)
- Symmetric binary perceptron (Gamarnik-Kızıldağ-Perkins-Xu 22)
- Mean field spin glass (Gamarnik-Jagannath 19, Gamarnik-Jagannath-Wein 20)

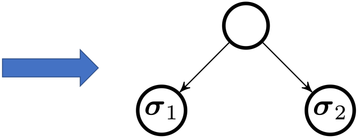
Overlap gap: no high-value $\boldsymbol{\sigma}, \boldsymbol{\rho}$ have **medium** overlap $\in [\nu_1, \nu_2]$

- Intuition: high-value points close together or far apart

Classic OGP (Gamarnik-Sudan 14)

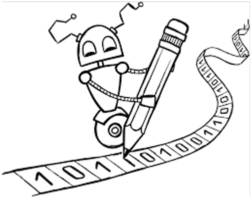
1. Stable algorithm $\mathcal{A}$ reaching $E \Rightarrow$ 2 points of value E with **medium overlap**

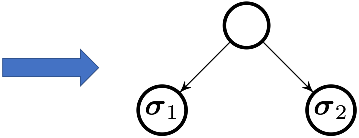




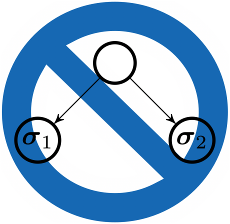
Classic OGP (Gamarnik-Sudan 14)

1. Stable algorithm $\mathcal{A}$ reaching E $\Rightarrow$ 2 points of value E with **medium overlap**

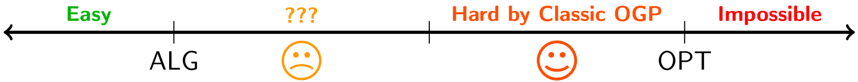




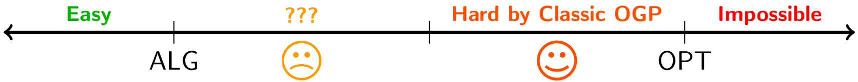
2. **Overlap gap** $\Rightarrow$ this pair does not exist. So $\mathcal{A}$ cannot reach E



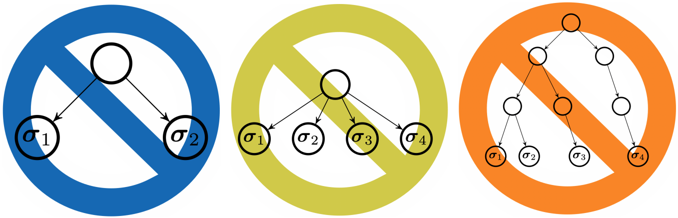
Classic OGP to Multi-OGP



Classic OGP to Multi-OGP

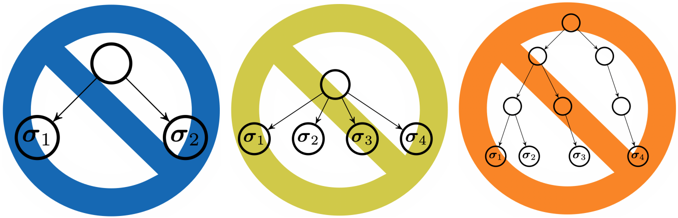


Multi-OGP: more complex forbidden structure (Rahman-Virág 17, Wein 20, ...)



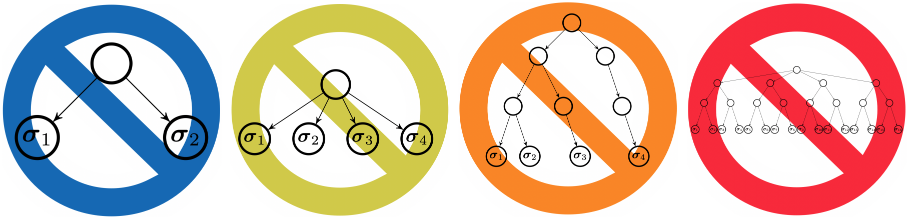
Classic OGP to Multi-OGP

Multi-OGP: more complex forbidden structure (Rahman-Virág 17, Wein 20, ...)



Classic OGP to Multi-OGP

Multi-OGP: more complex forbidden structure (Rahman-Virág 17, Wein 20, ...)



Can we push hardness to ALG? Yes, by **Branching OGP**.

Main Result: Algorithmic Threshold

Theorem (Subag 18)

An efficient algorithm finds $\boldsymbol{\sigma}$ such that

$$\frac{1}{N} H_N(\boldsymbol{\sigma}) \geq \mathsf{ALG} \equiv \int_0^1 \xi''(q)^{1/2} \mathrm{d}q.$$

Main Result: Algorithmic Threshold

Theorem (Subag 18)

An efficient algorithm finds $\boldsymbol{\sigma}$ such that

$$\frac{1}{N} H_N(\boldsymbol{\sigma}) \geq \mathsf{ALG} \equiv \int_0^1 \xi''(q)^{1/2} \mathrm{d}q.$$

Theorem (H.-Sellke 21)

*If ξ **even**, no $O(1)$-Lipschitz algorithm beats ALG with probability e^{-cN}.*

Tight answer for even models, but brittle proof using **Guerra's interpolation**

Main Result: Algorithmic Threshold

Theorem (Subag 18)

An efficient algorithm finds $\boldsymbol{\sigma}$ such that

$$\frac{1}{N} H_N(\boldsymbol{\sigma}) \geq \mathsf{ALG} \equiv \int_0^1 \xi''(q)^{1/2} \mathrm{d}q.$$

Theorem (H.-Sellke 21)

If ξ **even**, *no* $O(1)$*-Lipschitz algorithm beats* ALG *with probability* e^{-cN}.

Tight answer for even models, but brittle proof using **Guerra's interpolation**

Theorem (H.-Sellke 23)

For ***all*** ξ, *no* $O(1)$*-Lipschitz algorithm beats* ALG *with probability* e^{-cN}.

Main Result: Algorithmic Threshold

Theorem (Subag 18)

An efficient algorithm finds $\boldsymbol{\sigma}$ such that

$$\frac{1}{N} H_N(\boldsymbol{\sigma}) \geq \mathsf{ALG} \equiv \int_0^1 \xi''(q)^{1/2} \mathrm{d}q.$$

Theorem (H.-Sellke 21)

*If ξ **even**, no $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}.*

Tight answer for even models, but brittle proof using **Guerra's interpolation**

Theorem (H.-Sellke 23)

*For **all** ξ, no $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}.*

- New proof of Branching OGP avoids Guerra's interpolation

Main Result: Algorithmic Threshold

Theorem (Subag 18)

An efficient algorithm finds $\boldsymbol{\sigma}$ such that

$$\frac{1}{N} H_N(\boldsymbol{\sigma}) \geq \mathrm{ALG} \equiv \int_0^1 \xi''(q)^{1/2} \mathrm{d}q.$$

Theorem (H.-Sellke 21)

*If ξ **even**, no $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}.*

Tight answer for even models, but brittle proof using **Guerra's interpolation**

Theorem (H.-Sellke 23)

*For **all** ξ, no $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}.*

- New proof of Branching OGP avoids Guerra's interpolation
- Same method works for multi-species spin glasses
  - In these models, OPT not always known! (Because Guerra's interpolation fails)

Subag's Algorithm (Hessian Ascent)

Constant step size $\delta = 1/D$. $\boldsymbol{x}^0 = 0 \in \mathbb{R}^N$:

Subag's Algorithm (Hessian Ascent)

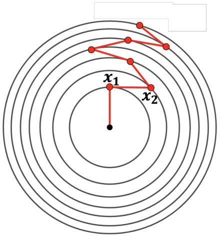
Constant step size $\delta = 1/D$. $\boldsymbol{x}^0 = 0 \in \mathbb{R}^N$:

1. Take $\boldsymbol{v}^t$ the top eigenvector of tangential Hessian $\nabla^2 H_N(\boldsymbol{x}^t)|_{(\boldsymbol{x}^t)^\perp}$

Subag's Algorithm (Hessian Ascent)

Constant step size $\delta = 1/D$. $\boldsymbol{x}^0 = 0 \in \mathbb{R}^N$:

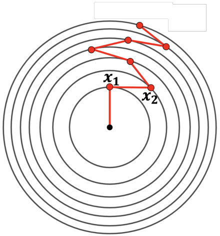
1. Take $\boldsymbol{v}^t$ the top eigenvector of tangential Hessian $\nabla^2 H_N(\boldsymbol{x}^t)|_{(\boldsymbol{x}^t)^\perp}$
2. Explore outward by small orthogonal steps: $\boldsymbol{x}^{t+1} = \boldsymbol{x}^t \pm \sqrt{\delta N}\boldsymbol{v}^t$.
   (Since $\boldsymbol{v}^t \perp \boldsymbol{x}^t$, $\|\boldsymbol{x}^t\|_2^2 = t\delta N$)



Subag's Algorithm (Hessian Ascent)

Constant step size $\delta = 1/D$. $\boldsymbol{x}^0 = 0 \in \mathbb{R}^N$:

1. Take $\boldsymbol{v}^t$ the top eigenvector of tangential Hessian $\nabla^2 H_N(\boldsymbol{x}^t)|_{(\boldsymbol{x}^t)^\perp}$
2. Explore outward by small orthogonal steps: $\boldsymbol{x}^{t+1} = \boldsymbol{x}^t \pm \sqrt{\delta N}\boldsymbol{v}^t$.
   (Since $\boldsymbol{v}^t \perp \boldsymbol{x}^t$, $\|\boldsymbol{x}^t\|_2^2 = t\delta N$)

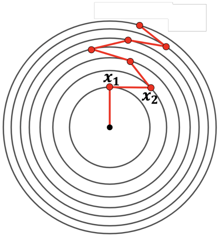


3. Output $\boldsymbol{\sigma} = \boldsymbol{x}^D \in S_N$

Subag's Algorithm (Hessian Ascent)

Constant step size $\delta = 1/D$. $\boldsymbol{x}^0 = 0 \in \mathbb{R}^N$:

1. Take $\boldsymbol{v}^t$ the top eigenvector of tangential Hessian $\nabla^2 H_N(\boldsymbol{x}^t)|_{(\boldsymbol{x}^t)^\perp}$
2. Explore outward by small orthogonal steps: $\boldsymbol{x}^{t+1} = \boldsymbol{x}^t \pm \sqrt{\delta N}\boldsymbol{v}^t$.
   (Since $\boldsymbol{v}^t \perp \boldsymbol{x}^t$, $\|\boldsymbol{x}^t\|_2^2 = t\delta N$)

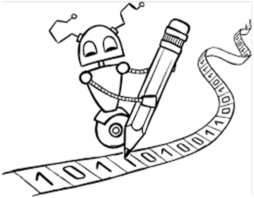


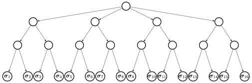
3. Output $\boldsymbol{\sigma} = \boldsymbol{x}^D \in S_N$

Can be implemented as $O(1)$-Lipschitz algorithm (El Alaoui-Montanari-Sellke 20)

Branching OGP (H.-Sellke 21)

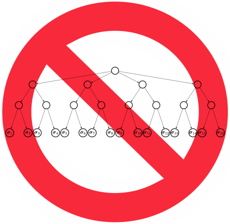
1. $O(1)$-Lipschitz algorithm $\mathcal{A}$ reaching $E \Rightarrow$ ultrametric of points of value E





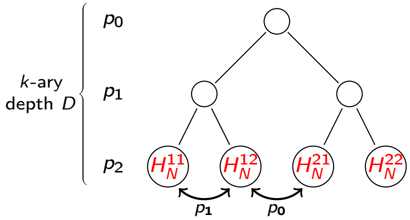
(with respect to a correlated Hamiltonian ensemble)

2. Constellation does not exist for $E \geq \mathrm{ALG} + \varepsilon$. So $\mathcal{A}$ cannot beat ALG



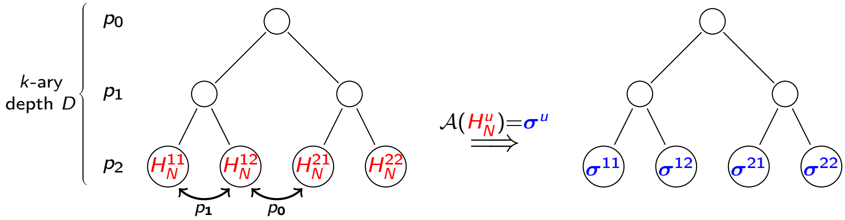
Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E



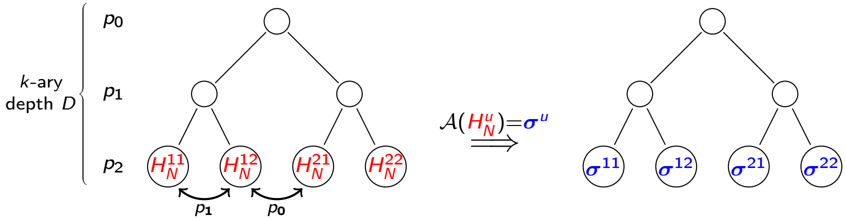
Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E



Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E

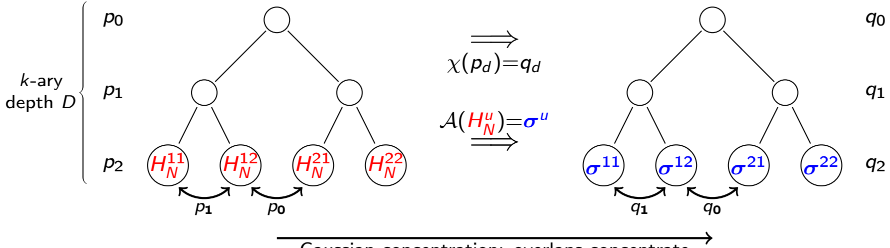


Gaussian concentration: overlaps concentrate

e.g. $R(\boldsymbol{\sigma}^{\mathbf{11}}, \boldsymbol{\sigma}^{\mathbf{12}}) \approx \mathbb{E}R(\boldsymbol{\sigma}^{\mathbf{11}}, \boldsymbol{\sigma}^{\mathbf{12}}) = \chi(p_{\mathbf{1}}) \equiv q_{\mathbf{1}}$

Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E

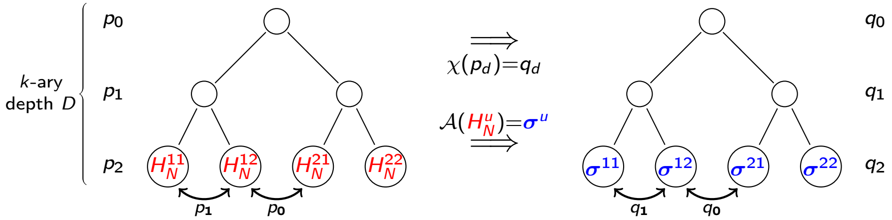


Gaussian concentration: overlaps concentrate

e.g. $R(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{12}) \approx \mathbb{E} R(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{12}) = \chi(p_1) \equiv q_1$

Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E



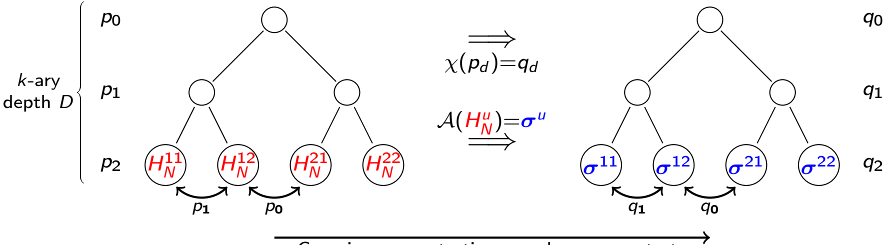
Gaussian concentration: overlaps concentrate

e.g. $R(\boldsymbol{\sigma}^{\mathbf{11}}, \boldsymbol{\sigma}^{\mathbf{12}}) \approx \mathbb{E}R(\boldsymbol{\sigma}^{\mathbf{11}}, \boldsymbol{\sigma}^{\mathbf{12}}) = \chi(p_{\mathbf{1}}) \equiv q_{\mathbf{1}}$

$(\boldsymbol{\sigma}^u)_{u\in[k]^D}$ is approximately ultrametric!

Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E



Gaussian concentration: overlaps concentrate

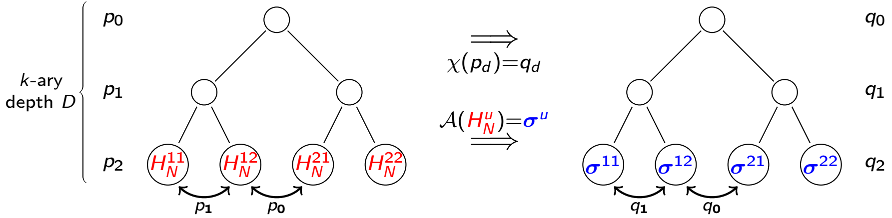
e.g. $R(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{12}) \approx \mathbb{E} R(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{12}) = \chi(p_1) \equiv q_1$

$(\boldsymbol{\sigma}^u)_{u \in [k]^D}$ is approximately ultrametric!

χ continuous $\Rightarrow$ can choose $\vec{p}$ so $(q_0, \ldots, q_D) = (0, \frac{1}{D}, \ldots, 1)$

Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E



Gaussian concentration: overlaps concentrate

e.g. $R(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{12}) \approx \mathbb{E}R(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{12}) = \chi(p_1) \equiv q_1$

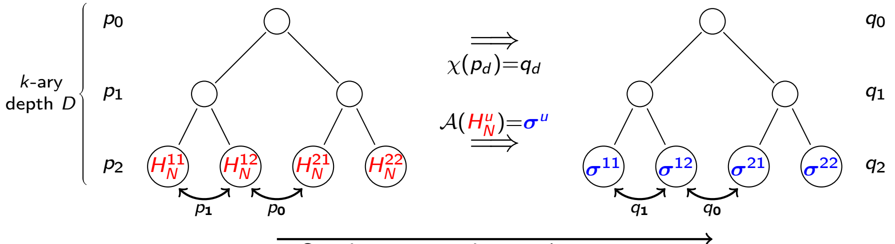
$(\boldsymbol{\sigma}^u)_{u \in [k]^D}$ is approximately ultrametric!

χ continuous $\Rightarrow$ can choose $\vec{p}$ so $(q_0, \ldots, q_D) = (0, \frac{1}{D}, \ldots, 1)$

$\mathcal{A}$ attains $E \quad \Rightarrow \quad \frac{1}{N} H_N^u(\boldsymbol{\sigma}^u) \geq E$ for all $u \in [k]^D$

Lipschitz Algorithms to Dense Ultrametric Trees

Suppose $O(1)$-Lipschitz algorithm $\mathcal{A}$ attains value E



Gaussian concentration: overlaps concentrate

e.g. $R(\boldsymbol{\sigma}^{\mathbf{11}}, \boldsymbol{\sigma}^{\mathbf{12}}) \approx \mathbb{E}R(\boldsymbol{\sigma}^{\mathbf{11}}, \boldsymbol{\sigma}^{\mathbf{12}}) = \chi(p_{\mathbf{1}}) \equiv q_{\mathbf{1}}$

$(\boldsymbol{\sigma}^u)_{u\in[k]^D}$ is approximately ultrametric!

χ continuous $\Rightarrow$ can choose $\vec{p}$ so $(q_0, \ldots, q_D) = (0, \frac{1}{D}, \ldots, 1)$

$\mathcal{A}$ attains $E \;\Rightarrow\; \frac{1}{N} H_N^u(\boldsymbol{\sigma}^u) \geq E$ for all $u \in [k]^D \;\Rightarrow\; \frac{1}{k^D N}\sum_{u\in[k]^D} H_N^u(\boldsymbol{\sigma}^u) \geq E$

Upper Bounding the Constellation Value

We will show: $$\max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q} = (0, 1/D, \ldots, 1)}} \frac{1}{k^D N} \sum_{u \in [k]^D} H_N^u(\boldsymbol{\sigma}^u) \le \mathsf{ALG}$$

Upper Bounding the Constellation Value

We will show: $$\max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q} = (0, 1/D, \ldots, 1)}} \frac{1}{k^D N} \sum_{u \in [k]^D} H_N(\boldsymbol{\sigma}^u) \le \mathsf{ALG}$$

Upper Bounding the Constellation Value

We will show: $$\max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q} = (0, 1/D, \ldots, 1)}} \frac{1}{k^D N} \sum_{u \in [k]^D} H_N(\boldsymbol{\sigma}^u) \le \mathsf{ALG}$$

New proof idea: **optimal constellation $(\boldsymbol{\sigma}^u)$ is greedy**

Upper Bounding the Constellation Value

We will show: $$\max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q} = (0, 1/D, \ldots, 1)}} \frac{1}{k^D N} \sum_{u \in [k]^D} H_N(\boldsymbol{\sigma}^u) \le \mathrm{ALG}$$

New proof idea: **optimal constellation $(\boldsymbol{\sigma}^u)$ is greedy**

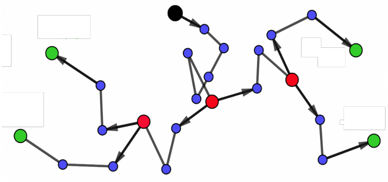
- Can **branch** Subag's algorithm by taking top k eigenvectors

Upper Bounding the Constellation Value

$$\text{We will show:} \quad \max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q} = (0, 1/D, \ldots, 1)}} \frac{1}{k^D N} \sum_{u \in [k]^D} H_N(\boldsymbol{\sigma}^u) \le \mathsf{ALG}$$

New proof idea: **optimal constellation $(\boldsymbol{\sigma}^u)$ is greedy**

- Can **branch** Subag's algorithm by taking top k eigenvectors
- Multi-valued algorithm. Outputs $(\boldsymbol{\sigma}^u)_{u \in [k]^D}$ ultrametric, all achieve $\approx \mathsf{ALG}$

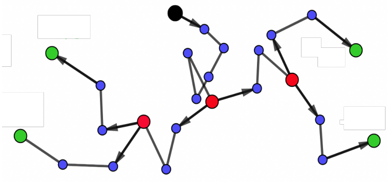


Upper Bounding the Constellation Value

We will show: $\displaystyle \max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q}=(0,1/D,\dots,1)}} \frac{1}{k^D N} \sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u) \le \mathrm{ALG}$

New proof idea: **optimal constellation $(\boldsymbol{\sigma}^u)$ is greedy**

- Can **branch** Subag's algorithm by taking top k eigenvectors
- Multi-valued algorithm. Outputs $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$ ultrametric, all achieve $\approx$ ALG



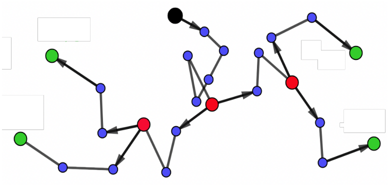
- Main claim: this greedy construction is optimal

Upper Bounding the Constellation Value

We will show: $$\max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q} = (0, 1/D, \ldots, 1)}} \frac{1}{k^D N} \sum_{u \in [k]^D} H_N(\boldsymbol{\sigma}^u) \le \mathrm{ALG}$$

New proof idea: **optimal constellation** $(\boldsymbol{\sigma}^u)$ **is greedy**

- Can **branch** Subag's algorithm by taking top k eigenvectors
- Multi-valued algorithm. Outputs $(\boldsymbol{\sigma}^u)_{u \in [k]^D}$ ultrametric, all achieve $\approx$ ALG



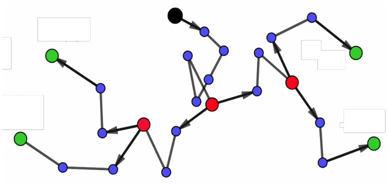
- Main claim: this greedy construction is optimal
    - Consequence: ALG = highest super-level set containing dense ultrametric

Upper Bounding the Constellation Value

We will show: $$\max_{\substack{(\boldsymbol{\sigma}^u) \text{ ultrametric with} \\ \vec{q}=(0,1/D,\dots,1)}} \frac{1}{k^D N} \sum_{u \in [k]^D} H_N(\boldsymbol{\sigma}^u) \le \mathrm{ALG}$$

New proof idea: **optimal constellation $(\boldsymbol{\sigma}^u)$ is greedy**

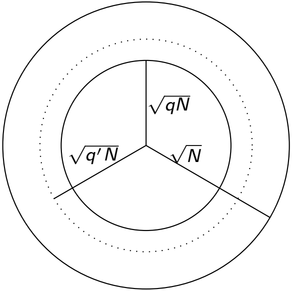
- Can **branch** Subag's algorithm by taking top k eigenvectors
- Multi-valued algorithm. Outputs $(\boldsymbol{\sigma}^u)_{u \in [k]^D}$ ultrametric, all achieve $\approx$ ALG



- Main claim: this greedy construction is optimal
  - Consequence: ALG $=$ highest super-level set containing dense ultrametric
  - "Can't plan ahead" – formalized by **uniform concentration**

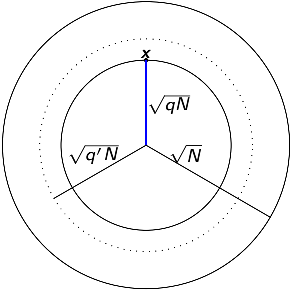
Uniform Concentration

Configuration $\boldsymbol{x}, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k$:



Uniform Concentration

Configuration $\boldsymbol{x}, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k$:

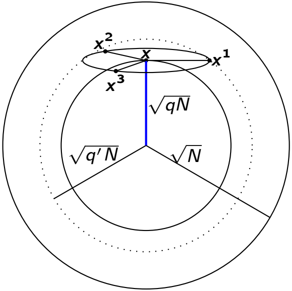


Radius:

$$\|\boldsymbol{x}\|_2 = \sqrt{qN}$$

Uniform Concentration

Configuration $\boldsymbol{x}, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k$:

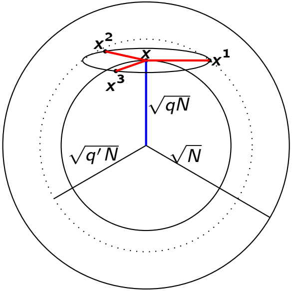


Radius:

$$\|\boldsymbol{x}\|_2 = \sqrt{qN}$$
$$\|\boldsymbol{x}^i\|_2 = \sqrt{q'N}$$

Uniform Concentration

Configuration $\boldsymbol{x}, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k$:



Radius:

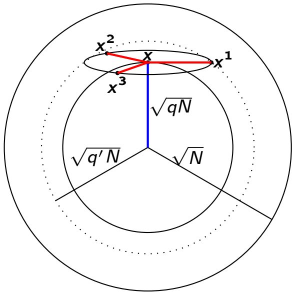
$$\|\boldsymbol{x}\|_2 = \sqrt{qN}$$
$$\|\boldsymbol{x}^i\|_2 = \sqrt{q'N}$$

Increment orthogonality:

$$\boldsymbol{x}^i - \boldsymbol{x} \perp \boldsymbol{x}^j - \boldsymbol{x} \perp \boldsymbol{x}$$

Uniform Concentration

Configuration $\boldsymbol{x}, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k$:



Radius:

$$\|\boldsymbol{x}\|_2 = \sqrt{qN}$$
$$\|\boldsymbol{x}^i\|_2 = \sqrt{q'N}$$

Increment orthogonality:

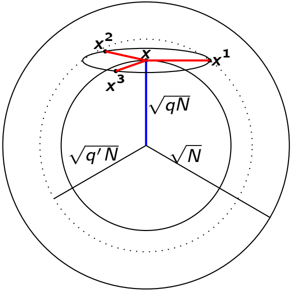
$$\boldsymbol{x}^i - \boldsymbol{x} \perp \boldsymbol{x}^j - \boldsymbol{x} \perp \boldsymbol{x}$$

$$F(\boldsymbol{x}) = \max_{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^k} \frac{1}{kN} \sum_{i=1}^{k} (H_N(\boldsymbol{x}^i) - H_N(\boldsymbol{x}))$$

"Improvement in H_N from $\boldsymbol{x}$ to its children"

Uniform Concentration

Configuration $\boldsymbol{x}, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k$:



Radius:

$$\|\boldsymbol{x}\|_2 = \sqrt{qN}$$
$$\|\boldsymbol{x}^i\|_2 = \sqrt{q'N}$$

Increment orthogonality:

$$\boldsymbol{x}^i - \boldsymbol{x} \perp \boldsymbol{x}^j - \boldsymbol{x} \perp \boldsymbol{x}$$

$$F(\boldsymbol{x}) = \max_{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^k} \frac{1}{kN} \sum_{i=1}^{k} (H_N(\boldsymbol{x}^i) - H_N(\boldsymbol{x}))$$

"Improvement in H_N from $\boldsymbol{x}$ to its children"

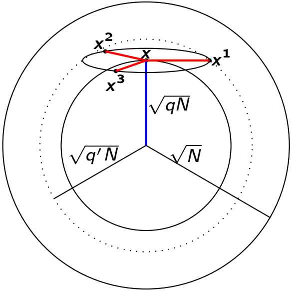
Lemma (Uniform Concentration, Subag 18)

For any $\eta > 0$, sufficiently large $k \geq k_0(\eta)$,

$$\mathbb{P}\left[|F(\boldsymbol{x}) - \mathbb{E}\, F(\boldsymbol{x})| \leq \eta \ \forall \|\boldsymbol{x}\|_2 = \sqrt{qN}\right] \geq 1 - e^{-cN}$$

Uniform Concentration

Configuration $\boldsymbol{x}, \boldsymbol{x}^1, \dots, \boldsymbol{x}^k$:



Radius:

$$\|\boldsymbol{x}\|_2 = \sqrt{qN}$$
$$\|\boldsymbol{x}^i\|_2 = \sqrt{q'N}$$

Increment orthogonality:

$$\boldsymbol{x}^i - \boldsymbol{x} \perp \boldsymbol{x}^j - \boldsymbol{x} \perp \boldsymbol{x}$$

$$F(\boldsymbol{x}) = \max_{\boldsymbol{x}^1, \dots, \boldsymbol{x}^k} \frac{1}{kN} \sum_{i=1}^{k} (H_N(\boldsymbol{x}^i) - H_N(\boldsymbol{x}))$$

"Improvement in H_N from $\boldsymbol{x}$ to its children"

Lemma (Uniform Concentration, Subag 18)

For any $\eta > 0$, sufficiently large $k \geq k_0(\eta)$,

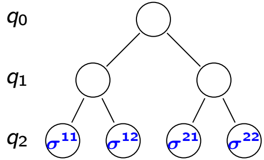
$$\mathbb{P}\left[|F(\boldsymbol{x}) - \mathbb{E}\, F(\boldsymbol{x})| \leq \eta \ \forall \|\boldsymbol{x}\|_2 = \sqrt{qN}\right] \geq 1 - e^{-cN}$$

No $\|\boldsymbol{x}\|_2 = \sqrt{qN}$ is unusually good for building a tree, so might as well be greedy.

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u \in [k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

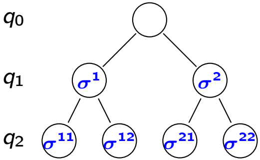
$$\boldsymbol{\sigma}^u = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{\sigma}^{ui}$$



Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u \in [k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

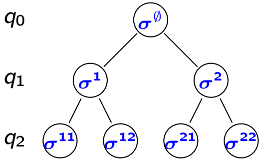
$$\boldsymbol{\sigma}^u = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{\sigma}^{ui}$$



Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

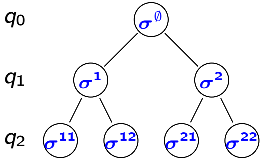
$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^k \boldsymbol{\sigma}^{ui}$$



Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

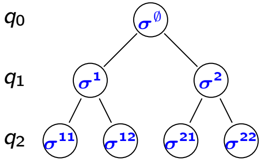
$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

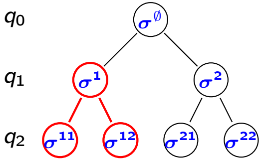
Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

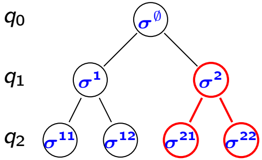
Equals telescoping sum of increments

$$\frac{1}{kN}\sum_{i=1}^{k}(H_N(\boldsymbol{\sigma}^{ui}) - H_N(\boldsymbol{\sigma}^u))$$

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

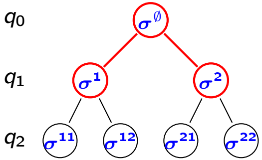
Equals telescoping sum of increments

$$\frac{1}{kN}\sum_{i=1}^{k}(H_N(\boldsymbol{\sigma}^{ui}) - H_N(\boldsymbol{\sigma}^u))$$

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

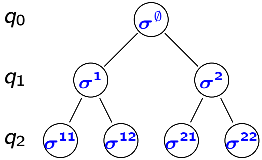
Equals telescoping sum of increments

$$\frac{1}{kN}\sum_{i=1}^{k}(H_N(\boldsymbol{\sigma}^{ui}) - H_N(\boldsymbol{\sigma}^u))$$

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^k \boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

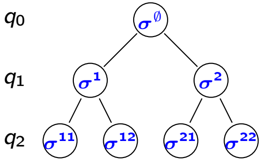
Equals telescoping sum of increments

$$\frac{1}{kN}\sum_{i=1}^k (H_N(\boldsymbol{\sigma}^{ui}) - H_N(\boldsymbol{\sigma}^u)) \le F(\boldsymbol{\sigma}^u)$$

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

Equals telescoping sum of increments

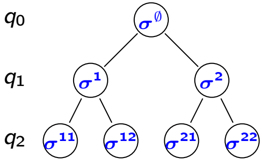
$$\frac{1}{kN}\sum_{i=1}^{k}(H_N(\boldsymbol{\sigma}^{ui}) - H_N(\boldsymbol{\sigma}^u)) \le F(\boldsymbol{\sigma}^u)$$

$F(\boldsymbol{\sigma}^u) \approx \mathbb{E}F(\boldsymbol{\sigma}^u)$ by uniform concentration!

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

Equals telescoping sum of increments

$$\frac{1}{kN}\sum_{i=1}^{k}(H_N(\boldsymbol{\sigma}^{ui}) - H_N(\boldsymbol{\sigma}^u)) \le F(\boldsymbol{\sigma}^u)$$

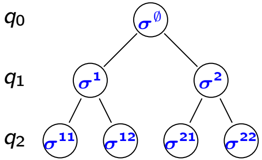
$F(\boldsymbol{\sigma}^u) \approx \mathbb{E}F(\boldsymbol{\sigma}^u)$ by uniform concentration!

Bounds match greedy algorithm, sum to ALG

Upper Bounding the Constellation Value

Given ultrametric $(\boldsymbol{\sigma}^u)_{u\in[k]^D}$, let interior $\boldsymbol{\sigma}^u$ be recursive barycenters

$$\boldsymbol{\sigma}^u = \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{\sigma}^{ui}$$



Satisfy orthogonality relations approximately if k large:

$$\|\boldsymbol{\sigma}^u\|_2 \approx \sqrt{q_{|u|}N}$$

$$\boldsymbol{\sigma}^{ui} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^{uj} - \boldsymbol{\sigma}^u \perp \boldsymbol{\sigma}^u$$

Want to upper bound:

$$\frac{1}{k^D N}\sum_{u\in[k]^D} H_N(\boldsymbol{\sigma}^u)$$

Equals telescoping sum of increments

$$\frac{1}{kN}\sum_{i=1}^{k}\left(H_N(\boldsymbol{\sigma}^{ui}) - H_N(\boldsymbol{\sigma}^u)\right) \le F(\boldsymbol{\sigma}^u)$$

$F(\boldsymbol{\sigma}^u) \approx \mathbb{E}F(\boldsymbol{\sigma}^u)$ by uniform concentration!

Bounds match greedy algorithm, sum to ALG

Correlated H_N^u: similarly bound

$$\frac{1}{kN}\sum_{i=1}^{k}\left(H_N^{ui}(\boldsymbol{\sigma}^{ui}) - H_N^{u}(\boldsymbol{\sigma}^u)\right)$$

Multi-Species Spin Glasses

- Up to now: polynomials in variables $x_1, \ldots, x_N$ that **all look alike**
- Multi-species models: $r = O(1)$ different "variable types"

Multi-Species Spin Glasses

- Up to now: polynomials in variables $x_1, \ldots, x_N$ that **all look alike**
- Multi-species models: $r = O(1)$ different "variable types"
- Formally, each coordinate part of a **species** $s \in \mathscr{S} = \{1, \ldots, r\}$

$$[N] = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_r, \qquad |\mathcal{I}_s| = \lambda_s N$$

Multi-Species Spin Glasses

- Up to now: polynomials in variables $x_1, \ldots, x_N$ that **all look alike**
- Multi-species models: $r = O(1)$ different "variable types"
- Formally, each coordinate part of a **species** $s \in \mathscr{S} = \{1, \ldots, r\}$

$$[N] = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_r, \qquad |\mathcal{I}_s| = \lambda_s N$$

- Interaction weights $\gamma_2, \gamma_3, \ldots$ now $(\gamma_{s_1,s_2})_{s_1,s_2 \in \mathscr{S}}, (\gamma_{s_1,s_2,s_3})_{s_1,s_2,s_3 \in \mathscr{S}}, \ldots$

Multi-Species Spin Glasses

- Up to now: polynomials in variables $x_1, \dots, x_N$ that **all look alike**
- Multi-species models: $r = O(1)$ different "variable types"
- Formally, each coordinate part of a **species** $s \in \mathscr{S} = \{1, \dots, r\}$

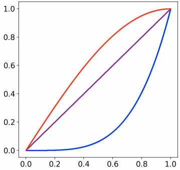
$$[N] = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_r, \qquad |\mathcal{I}_s| = \lambda_s N$$

- Interaction weights $\gamma_2, \gamma_3, \dots$ now $(\gamma_{s_1,s_2})_{s_1,s_2 \in \mathscr{S}}, (\gamma_{s_1,s_2,s_3})_{s_1,s_2,s_3 \in \mathscr{S}}, \dots$
- Goal: optimize H_N over **product of spheres**

$$\mathbb{T}_N = \left\{ \boldsymbol{\sigma} \in \mathbb{R}^N : \|\boldsymbol{\sigma}_{|\mathcal{I}_s}\|_2^2 = \lambda_s N \quad \forall s \in \mathscr{S} \right\}$$

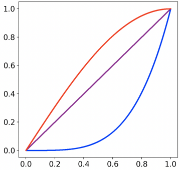
Multi-Species Algorithms

- Optimizing on product of spheres $\Rightarrow$ track radius for each species
  - 2 species: **radius schedule** is up-right path from $(0,0)$ to $(1,1)$



Multi-Species Algorithms

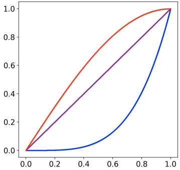
- Optimizing on product of spheres $\Rightarrow$ track radius for each species
  - 2 species: **radius schedule** is up-right path from $(0, 0)$ to $(1, 1)$



- In general, radius schedule is coordinate-increasing $\Phi : [0, 1] \to [0, 1]^{\mathscr{S}}$

Multi-Species Algorithms

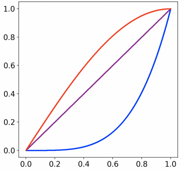
- Optimizing on product of spheres $\Rightarrow$ track radius for each species
  - 2 species: **radius schedule** is up-right path from $(0, 0)$ to $(1, 1)$



- In general, radius schedule is coordinate-increasing $\Phi : [0, 1] \to [0, 1]^{\mathscr{S}}$
- Each Φ gives algorithm taking small orthogonal steps **in each species**

Multi-Species Algorithms

- Optimizing on product of spheres $\Rightarrow$ track radius for each species
  - 2 species: **radius schedule** is up-right path from $(0,0)$ to $(1,1)$



- In general, radius schedule is coordinate-increasing $\Phi : [0,1] \to [0,1]^{\mathscr{S}}$
- Each Φ gives algorithm taking small orthogonal steps **in each species**
- Algorithm value

$$\mathbb{A}(\Phi) \equiv \sum_{s \in \mathscr{S}} \int_0^1 \sqrt{\lambda_s (\partial_s \xi \circ \Phi)'(q) \Phi_s'(q)} \, \mathrm{d}q$$

(ξ now multivariate polynomial in $|\mathscr{S}|$ variables)

Multi-Species Algorithmic Threshold

Theorem (H.-Sellke 23)

Define

$$\mathrm{ALG} = \sup_{\substack{\Phi:[0,1]\to[0,1]^{\mathscr{S}} \\ \textit{increasing, differentiable}}} \sum_{s\in\mathscr{S}} \int_0^1 \sqrt{\lambda_s(\partial_s\xi \circ \Phi)'(q)\Phi_s'(q)}\,\mathrm{d}q$$

Multi-Species Algorithmic Threshold

Theorem (H.-Sellke 23)

Define

$$\mathsf{ALG} = \sup_{\substack{\Phi:[0,1]\to[0,1]^{\mathscr{S}} \\ \text{increasing, differentiable}}} \sum_{s\in\mathscr{S}} \int_0^1 \sqrt{\lambda_s(\partial_s\xi\circ\Phi)'(q)\Phi_s'(q)}\,\mathrm{d}q$$

- *An explicit $O(1)$-Lipschitz algorithm achieves* ALG *w.h.p.*
- *No $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}.*

Multi-Species Algorithmic Threshold

Theorem (H.-Sellke 23)

Define

$$\mathrm{ALG} = \sup_{\substack{\Phi:[0,1]\to[0,1]^{\mathscr{S}} \\ \textit{increasing, differentiable}}} \sum_{s\in\mathscr{S}} \int_0^1 \sqrt{\lambda_s(\partial_s\xi \circ \Phi)'(q)\Phi_s'(q)}\, \mathrm{d}q$$

- *An explicit $O(1)$-Lipschitz algorithm achieves* ALG *w.h.p.*
- *No $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}.*

Theorem (H.-Sellke 23)

The variational formula has a maximizer Φ, which solves an explicit ODE.

Multi-Species Algorithmic Threshold

Theorem (H.-Sellke 23)

Define

$$\mathsf{ALG} = \sup_{\substack{\Phi:[0,1]\to[0,1]^{\mathscr{S}} \\ \textit{increasing, differentiable}}} \sum_{s\in\mathscr{S}} \int_0^1 \sqrt{\lambda_s(\partial_s\xi \circ \Phi)'(q)\Phi_s'(q)}\, \mathrm{d}q$$

- *An explicit $O(1)$-Lipschitz algorithm achieves* ALG *w.h.p.*
- *No $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}.*

Theorem (H.-Sellke 23)

The variational formula has a maximizer Φ, which solves an explicit ODE.

Theorem (H.-Sellke 23)

*For **pure** models $\xi(\vec{q}) = q_1^{b_1} q_2^{b_2} \cdots q_r^{b_r}$,* $\mathsf{ALG} = E_\infty$.

Summary

- We determine algorithmic threshold of $O(1)$-Lipschitz algorithms for optimizing multi-species spherical spin glasses

Summary

- We determine algorithmic threshold of $O(1)$-Lipschitz algorithms for optimizing multi-species spherical spin glasses
- Geometric description of ALG: largest value whose super-level set contains densely-branching ultrametric tree

Summary

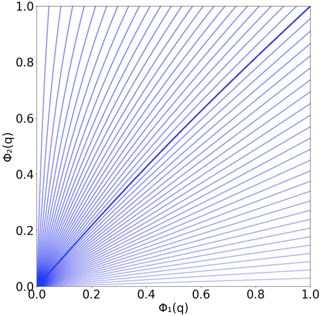
- We determine algorithmic threshold of $O(1)$-Lipschitz algorithms for optimizing multi-species spherical spin glasses
- Geometric description of ALG: largest value whose super-level set contains densely-branching ultrametric tree

Thank you!

Variational Problem Example

Consider $(\lambda_1, \lambda_2) = (1/3, 2/3)$ and

$$\xi(q_1, q_2) = (\lambda_1 q_1)^2 + (\lambda_1 q_1)(\lambda_2 q_1) + (\lambda_2 q_1)^2 + (\lambda_1 q_1)^4 + (\lambda_1 q_1)(\lambda_2 q_2)^3$$

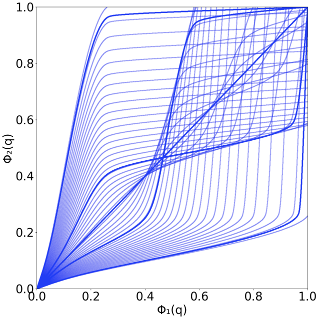


Some ODE solutions. Optimal $\Phi : [0,1] \to [0,1]^2$ in bold

Algorithmic Symmetry Breaking

Optimal Φ may be asymmetric, even when model is symmetric!

$$\lambda_1 = \lambda_2 = \frac{1}{2}, \quad \xi(q_1, q_2) = (3q_1)^2 + (3q_1)(3q_2) + (3q_2)^2 + (3q_1)^4 + (3q_2)^4$$



The plot thickens...

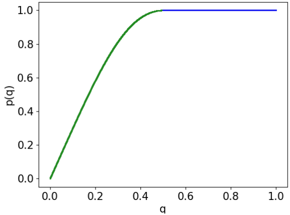
Models with Linear Terms

Suppose model has 1-spin interaction (external field)

$$H_N(\boldsymbol{\sigma}) = \sum_{p=\color{red}{1}}^{P} \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle \qquad \xi(q) = \sum_{p=\color{red}{1}}^{P} \gamma_p^2 q^p$$

Then

$$\mathsf{ALG} = \mathsf{BOGP} = \sup_{\substack{p:[0,1]\to[0,1] \\ \text{increasing, differentiable}}} \int_0^1 \sqrt{(p\xi')'(q)}\, \mathrm{d}q$$



Optimal p for $\xi(q) = q^4 + q$

Multi-Species Algorithmic Threshold with Linear Terms

Theorem (H.-Sellke 23)

Define

$$\mathsf{ALG} = \sup_{\substack{p:[0,1]\to[0,1] \\ \Phi:[0,1]\to[0,1]^{\mathscr{S}} \\ \textit{increasing, differentiable}}} \sum_{s\in\mathscr{S}} \lambda_s \int_0^1 \sqrt{(p \times \xi^s \circ \Phi)'(q)\Phi_s'(q)}\,\mathrm{d}q$$

- *An explicit $O(1)$-Lipschitz algorithm achieves* ALG *w.h.p.*
- *No $O(1)$-Lipschitz algorithm beats* ALG *with probability e^{-cN}*

Theorem (H.-Sellke 23)

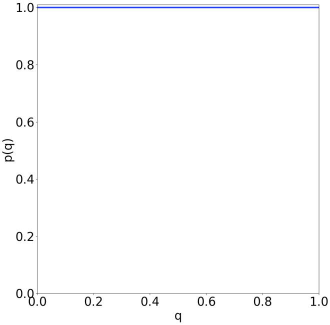
This variational problem has a maximizer (p, Φ).

- *The maximizer solves an explicit ODE.*
- *If ξ has no 1-spin interactions, then $p \equiv 1$.*

Variational Problem Example: No Linear Term

Consider $(\lambda_1, \lambda_2) = (1/3, 2/3)$

$$\xi(q_1, q_2) = (\lambda_1 q_1)^2 + (\lambda_1 q_1)(\lambda_2 q_1) + (\lambda_2 q_1)^2 + (\lambda_1 q_1)^4 + (\lambda_1 q_1)(\lambda_2 q_2)^3$$



Optimal $p : [0,1] \to [0,1]$

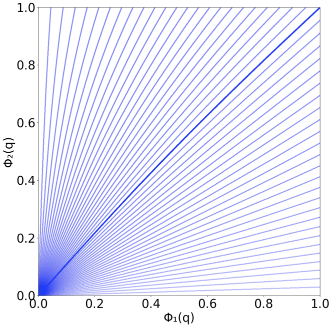
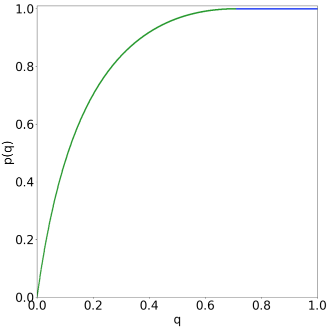


Image of optimal $\Phi : [0,1] \to [0,1]^2$ in bold

Variational Problem Example: Small Linear Term

Consider $(\lambda_1, \lambda_2) = (1/3, 2/3)$

$$\xi(q_1, q_2) = (\lambda_1 q_1)^2 + (\lambda_1 q_1)(\lambda_2 q_1) + (\lambda_2 q_1)^2 + (\lambda_1 q_1)^4 + (\lambda_1 q_1)(\lambda_2 q_2)^3 + 0.05(\lambda_1 q_1) + 0.5(\lambda_2 q_2)$$



Optimal $p : [0,1] \to [0,1]$

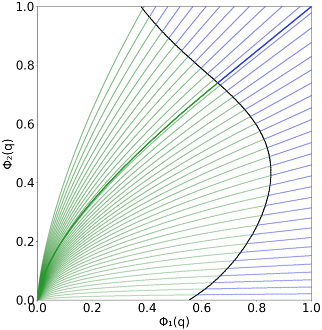
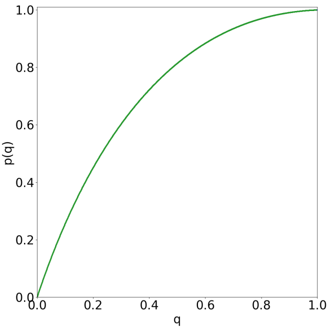


Image of optimal $\Phi : [0,1] \to [0,1]^2$ in bold

Variational Problem Example: Large Linear Term

Consider $(\lambda_1, \lambda_2) = (1/3, 2/3)$

$$\xi(q_1, q_2) = (\lambda_1 q_1)^2 + (\lambda_1 q_1)(\lambda_2 q_1) + (\lambda_2 q_1)^2 + (\lambda_1 q_1)^4 + (\lambda_1 q_1)(\lambda_2 q_2)^3 + 0.2(\lambda_1 q_1) + 1.8(\lambda_2 q_2)$$



Optimal $p : [0,1] \to [0,1]$

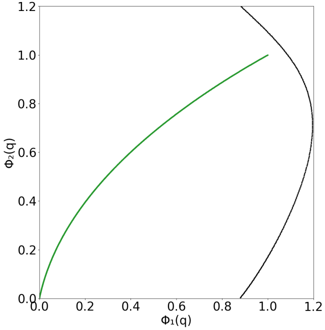


Image of optimal $\Phi : [0,1] \to [0,1]^2$