Algorithmic Obstructions in the Random Number Partitioning Problem

David Gamarnik [1] Eren C. Kızıldağ [2]

[1]MIT ORC [2]MIT EECS

Random Number Partitioning Problem (NPP)

Setup: Given n items $X_1,\dots,X_n \in \mathbb{R}$; partition them into two "bins" with total weights as close as possible: $\min_{A\subset[n]} \left|\sum_{i\in A} X_i - \sum_{i\in A^c} X_i\right|$. Equivalently,

$$\min_{\sigma\in\mathcal{B}_n} \left|\langle \sigma, X\rangle\right|, \quad \text{where} \quad \mathcal{B}_n = \{-1,1\}^n \quad \text{and} \quad \langle \sigma, X\rangle = \sum_{1\le i\le n} \sigma_i X_i.$$

Our focus: X_i are i.i.d. standard normal: $X_i \stackrel{d}{=} \mathcal{N}(0,1)$.

Applications

Randomized controlled trials: Gold standard for clinical trials (Krieger et al.'19; Harshaw et al.'19).

- n persons with covariate info (age, weight, height,...) $X_i \in \mathbb{R}^d$, $1\le i\le n$.
- Split into two groups (*treatment* and *control*) with similar "features":

$$\min_{\sigma\in\mathcal{B}_n} \|X\sigma\|_\infty, \quad \text{where} \quad X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^{d\times n}.$$

- **Goal.** Accurate inference for a treatment effect.

Many more applications: *multiprocessor scheduling, VLSI design, cryptography,...*

Available Guarantees

Existential: Let $X_i \stackrel{d}{=} \mathcal{N}(0,1)$, $1\le i\le n$ i.i.d. Then,

$$\min_{\sigma\in\mathcal{B}_n} |\langle\sigma, X\rangle| = \Theta\left(\sqrt{n}2^{-n}\right), \quad \text{w.h.p. as } n\to\infty.$$

Non-constructive. Extends to high dimensions: $\Theta\left(\sqrt{n}2^{-n/d}\right)$ for $2\le d\le o(n)$ (Turner et al.'20).

Algorithmic (Polynomial-Time): *Largest Differencing Method (LDM)* by Karmarkar and Karp'82.

For $d=1$ and $X_i \stackrel{d}{=} \mathcal{N}(0,1)$, $1\le i\le n$ i.i.d.; returns a $\sigma_{\rm ALG}\in\mathcal{B}_n$ such that

$$\left|\langle \sigma_{\rm ALG}, X\rangle\right| = 2^{-\Theta(\log^2 n)} \quad \text{w.h.p. as } n\to\infty.$$

Extends to high dimensions: $\exp\left(-\Omega\left(\log^2 n/d\right)\right)$ for $2\le d\le O(\sqrt{\log n})$ (Turner et al.'20).

A *Statistical-to-Computational Gap*

Gap between existential guarantees and what polynomial-time algorithms can promise.

- **Our focus:** Dimension $d=1$. For $X_i \stackrel{d}{=} \mathcal{N}(0,1)$, $1\le i\le n$ i.i.d.

$$\min_{\sigma\in\mathcal{B}_n} |\langle \sigma, X\rangle| = \Theta(\sqrt{n}2^{-n}) \quad \text{vs} \quad |\langle\sigma_{\rm ALG}, X\rangle| = 2^{-\Theta(\log^2 n)}.$$

- Ignoring $\sqrt{n}$, a striking gap: 2^{-n} vs $2^{-\Theta(\log^2 n)}$.

Source of this gap/hardness?

Study of Statistical-to-Computational Gap

Common feature in many algorithmic problems in high-dimensional statistics & random combinatorial structures: *Random k-SAT, optimization over random graphs, p-spin model, planted clique, matrix PCA, linear regression, spiked tensor, largest submatrix problem...*

Average-Case Problems: No analogue of worst-case theory (such as $P\ne NP$). Various Forms of **Rigorous Evidences of Hardness:** *low-degree methods, reductions from the planted clique, failure of MCMC, failure of BP/AMP, SoS lower bounds,...*

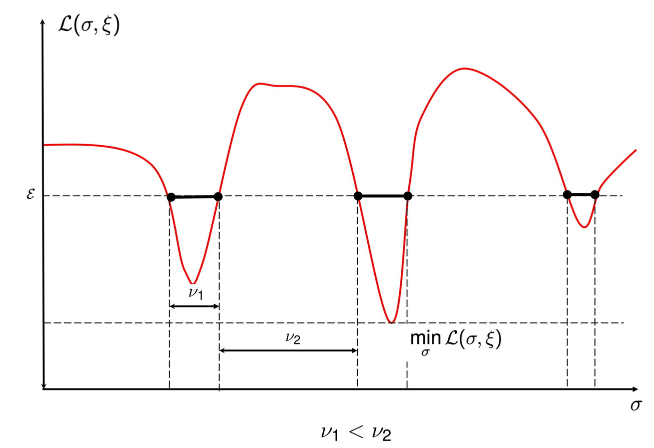
Overlap Gap Property (OGP)

Another approach from *spin glass theory*: **Overlap Gap Property (OGP).**

- Generic optimization problem with random ξ: $\min_{\theta\in\Theta}\mathcal{L}(\sigma,\xi)$.
- (Informally) OGP for **energy** $\mathcal{E}$ if $\exists 0<\nu_1<\nu_2$ s.t. w.h.p. over ξ, $\forall \sigma_1,\sigma_2\in\Theta$,

$$\mathcal{L}(\sigma_j,\xi)\le\mathcal{E} \implies \text{distance}(\sigma_1,\sigma_2)<\nu_1 \quad \text{or} \quad \text{distance}(\sigma_1,\sigma_2)>\nu_2.$$

- Any two **near optimal** σ_1,σ_2 are either *too similar* or *too dissimilar*.



First algorithmic implication: Maximum independent set in $\mathbb{G}_d(n)$ and $\mathbb{G}(n,\frac{d}{n})$ (Gamarnik and Sudan'13). **Many problems with OGP:** *random k-SAT, NAE-k-SAT, p-spin model, sparse PCA, largest submatrix problem, max-CUT, planted clique,...*

OGP as a Provable Barrier to Algorithms: *WALKSAT, local algorithms, stable algorithms, low-degree polynomials, AMP, MCMC, low-depth circuits...*

Landscape Results: Presence of OGP

Theorem. $\forall\epsilon\in(1/2,1)$, $\exists\rho:=\rho(\epsilon)\in(0,1)$ such that if $\sigma,\sigma'\in\mathcal{B}_n$ achieve

$$|\langle\sigma,X\rangle| = O(\sqrt{n}2^{-\epsilon n}) \quad \text{and} \quad |\langle\sigma',X\rangle| = O(\sqrt{n}2^{-\epsilon n})$$

then either $\sigma=\sigma'$ or $n^{-1}|\langle\sigma,\sigma'\rangle|\le\rho$ w.h.p. That is, $n^{-1}|\langle\sigma,\sigma'\rangle|\notin(\rho,\frac{n-2}{n}]$.

- Partitions achieving better than $2^{-\frac{n}{2}}$ are isolated vectors separated by $\Theta(n)$ distance.
- Yields existence of a *Free Energy Well (FEW)*: failure of **Glauber dynamics**.

Still large gap between $2^{-\frac{n}{2}}$ and $2^{-\Theta(\log^2 n)}$. **Idea:** Inspect $m-$tuples instead.

- Interpolate $Y_i(\tau)=\sqrt{1-\tau^2}X_0+\tau X_i$, where $X_0,\dots,X_m \stackrel{d}{=}\mathcal{N}(0,I_n)$ i.i.d.
- Study $m-$tuples $\sigma_i\in\mathcal{B}_n$, $1\le i\le m$, each near-optimal w.r.t. $Y_i(\tau_i)$ (Ensemble $m-$OGP).
- Reduce thresholds further, and rule out sufficiently stable algorithms.

Theorem. $\forall\epsilon>0$, $\forall\mathcal{I}\subset[0,1]$ with $|\mathcal{I}|=2^{o(n)}$, $\exists m\in\mathbb{N}$, $\exists 1>\beta>\eta>0$ s.t. if

$$|\langle\sigma_i, Y_i(\tau_i)\rangle| = O\left(\sqrt{n}2^{-\epsilon n}\right), \quad \tau_i\in\mathcal{I}, \quad 1\le i\le m$$

then w.h.p. $\exists 1\le i<j\le m$ such that $n^{-1}|\langle\sigma_i,\sigma_j\rangle|\notin(\beta-\eta,\beta)$.

Still striking gap between $2^{-\epsilon n}$ and $2^{-\Theta(\log^2 n)}$. Unfortunately, $m-$OGP (with $m=O(1)$) absent for $2^{-o(n)}$. **New Idea:** Study $m-$tuples with $m=\omega_n(1)$.

Theorem. $\forall\omega(\sqrt{n\log n})\le E_n\le o(n)$, $\forall\mathcal{I}\subset[0,1]$ with $|\mathcal{I}|=n^{O(1)}$, $\exists m_n\in\mathbb{N}$, $\exists 1>\beta_n>\eta_n>0$ s.t. if

$$|\langle\sigma_i, Y_i(\tau_i)\rangle| \le \sqrt{n}2^{-E_n}, \quad \tau_i\in\mathcal{I}, \quad 1\le i\le m_n$$

then w.h.p. $\exists 1\le i<j\le m_n$ such that $n^{-1}\langle\sigma_i,\sigma_j\rangle\notin(\beta_n-\eta_n,\beta_n)$.

Algorithmic Hardness Results

Algorithm $\mathcal{A}:\mathbb{R}^n\to\mathcal{B}_n$, potentially randomized.

Stable Algorithms. Informally, $\mathcal{A}$ is stable if small change in X yields small change in $\mathcal{A}(X)$.
Success:

$$\mathbb{P}\left(n^{-\frac{1}{2}}|\langle X,\mathcal{A}(X)\rangle|\le E\right)\ge 1-p_f.$$

Stability: $\exists\rho\in(0,1]$, $X,Y\stackrel{d}{=}\mathcal{N}(0,I_n)$ with ${\rm Cov}(X,Y)=\rho I_n$;

$$\mathbb{P}\left(d_H\left(\mathcal{A}(X),\mathcal{A}(Y)\right)\le f+L\|X-Y\|_2^2\right)\ge 1-p_{\rm st}.$$

Stable algorithms include approximate message passing (AMP) type algorithms (Gamarnik and Jagannath'21) and the low-degree polynomials (Gamarnik, Jagannath, and Wein'20).

A Conjecture (verified by simulations): Largest differencing (LDM) algorithm is stable.

Theorem. Stable algorithms can't achieve value better than

$$\exp\left(-\omega\left(\frac{n}{\log^{1/5}n}\right)\right).$$

Semi-formally, $\forall\epsilon\in(0,1/5)$, $\forall\omega(n\log^{-1/5+\epsilon}n)\le E_n\le o(n)$, there is no stable $\mathcal{A}$ that w.h.p. returns a σ with energy 2^{-E_n} (with appropriate $f,\rho',p_f,p_{\rm st}$).

- For **extreme** case, $E_n=\Theta(n)$: rule out $p_f,p_{\rm st}=O(1)$.
- **Proof Idea.** By contradiction. Suppose $\exists\mathcal{A}$.
  - m-OGP: a structure occurs with *vanishing probability*.
  - Run $\mathcal{A}$ on correlated instances. Show that w.p. >0, *forbidden* structure occurs.
- Rate $2^{-\omega(n\log^{-1/5}n)}$: Via **Ramsey Theory**.

Failure of MCMC: Let $X\stackrel{d}{=}\mathcal{N}(0,I_n)$; define **Hamiltonian** $H(\sigma)\triangleq n^{-\frac{1}{2}}|\langle\sigma,X\rangle|$, and consider the **Gibbs distribution** at inverse temperature $\beta>0$ on $\mathcal{B}_n$: $\pi_\beta(\sigma)\propto\exp\left(-\beta H(\sigma)\right)$.

Construct $\mathbb{G}=(V,E)$ with $V=\mathcal{B}_n$ and $(\sigma,\sigma')\in E \iff d_H(\sigma,\sigma')=1$; and consider any nearest neighbor MC $(X_t)_{t\ge0}$ on $\mathbb{G}$ reversible w.r.t. π_β.

Let $(\pm)\sigma^*=\min_{\sigma\in\mathcal{B}_n}H(\sigma)$, and for $\epsilon\in(1/2,1]$, $\rho:=\rho(\epsilon)$ be the $2-$OGP parameter. Set

$$I_1=\left\{\sigma:-\rho\le\frac{1}{n}\langle\sigma,\sigma^*\rangle\le\rho\right\},\quad I_2=\left\{\sigma:\rho\le\frac{1}{n}\langle\sigma,\sigma^*\rangle\le\frac{n-2}{n}\right\},\quad\text{and}\quad I_3=\{\sigma^*\}.$$

Theorem. For $\beta=\Omega(n2^{n\epsilon})$, w.h.p. (w.r.t. $X\stackrel{d}{=}\mathcal{N}(0,I_n)$), $\min\left\{\pi_\beta(I_1),\pi_\beta(I_3)\right\}\ge e^{\Omega(n)}\pi_\beta(I_2)$.

I_2 is a FEW with exponentially small **Gibbs** mass separating I_3 and $I_1\cup\overline{I_2}\cup\overline{I_3}$. *Exit time* from well is *exponential*: **Slow mixing.**

Future Directions

- Formally verifying stability of **LDM**.
- Proving algorithmic hardness all the way to $2^{-\omega(\sqrt{n\log n})}$. Rate $2^{-\omega(n\log^{-1/5}n)}$ **unimprovable by Ramsey**.
- Still a significant gap $2^{-\omega(\sqrt{n\log n})}$ vs $2^{-\Theta(\log^2 n)}$.
  - Either prove **hardness** for $2^{-\omega(\log^2 n)}$: **OGP not applicable**.
  - Or devise a **better** (polynomial-time) **algorithm** achieving $2^{-\omega(\log^2 n)}$.
- *Slow mixing* for **higher** temperatures (smaller β); or for **different** initialization, e.g. *uniform* case.

*Can OGP rule out **all** polynomial-time algorithms? Is there a problem **with** OGP **yet** admitting a polynomial-time algorithm?*

Cornell ORIE Young Researchers Workshop; October 7-9, 2021.

kizildag@mit.edu