A New Framework for Matrix Discrepancy: Partial Coloring Bounds via Mirror Descent

Daniel Dadush (CWI), Haotian Jiang (UW) and Victor Reis (UW)

Introduction to vector balancing

- Given $A, B \subseteq \mathbb{R}^d$, $\mathrm{vb}_n(A,B) :=$ smallest $C > 0$ so that for any $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in A$ there are signs $\boldsymbol{x} \in \{\pm 1\}^n$ with
$$\sum_{i=1}^n x_i \boldsymbol{v}_i \in C \cdot B.$$
- Applications to scheduling [BRS22] and RCTs [HSSZ19]
- Random signs give $\mathrm{vb}_n(B_\infty^n, B_\infty^n) \lesssim \sqrt{n \log(2n)}$ by Chernoff + union bound

Spencer's "Six Standard Deviations Suffice" Theorem ('85)

$\mathrm{vb}_n(B_\infty^n, B_\infty^n) \leqslant 6\sqrt{n}$ and more generally $\mathrm{vb}_n(B_\infty^m, B_\infty^m) \lesssim \sqrt{n \log(2m/n)}$.

Matrix Spencer Conjecture

- Generalization for matrices: $S_p^m := \{\text{symmetric } \boldsymbol{A} \in \mathbb{R}^{m\times m} : \|\lambda(\boldsymbol{A})\|_p \leqslant 1\}$
- For diagonal matrices $\|\boldsymbol{A}\|_{S_p^m} = \|\mathrm{diag}(\boldsymbol{A})\|_p$

Matrix Spencer Conjecture (Zouzias' 12, Meka '14)

$\mathrm{vb}_n(S_\infty^n, S_\infty^n) \lesssim \sqrt{n}$ and in general $\mathrm{vb}_n(S_\infty^m, S_\infty^m) \lesssim \sqrt{n \log(2\max(m,n)/n)}$

Theorem (Dadush, Jiang, Reis '22; Hopkins, Raghavendra, Shetty '22)

$\mathrm{vb}_n(S_\infty^m, S_\infty^m) \lesssim \sqrt{n \log(2m^2/n)}$ for all $1 \leqslant n \leqslant m^2$.

$$\boldsymbol{A}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{A}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \boldsymbol{A}_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } \boldsymbol{A}_4 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

- Generalizes for $n = m^2$: $\sqrt{n}$ lower bound

Partial Colorings

- Given $A, B \subseteq \mathbb{R}^d$ and $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in A$, define the *discrepancy body*
$$K := \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i \boldsymbol{v}_i \in B \right\}$$
- Upper bound $\mathrm{vb}_n(A,B)$ by iteratively constructing *partial colorings*
$$\boldsymbol{x} \in (C \cdot K) \cap [-1,1]^n : |\{i : |x_i| = 1\}| \gtrsim n$$
- Can find partial colorings if $\gamma_n(C \cdot K) := \Pr_{\boldsymbol{g} \sim N(0, \boldsymbol{I}_n)}[\boldsymbol{g} \in C \cdot K] \geqslant 2^{-O(n)}$
- Suffices to cover *polar body* $(C \cdot K)^\circ$ with $2^{O(n)}$ balls $\frac{1}{\sqrt{n}} B_2^n$ or cubes $\frac{1}{n} B_\infty^n$
- Construct such covering via optimization

From covering to optimization

- $f_{\boldsymbol{U}}(\boldsymbol{X}) := \max_{i \in [n]} |\langle \boldsymbol{A}_i, \boldsymbol{X} - \boldsymbol{U} \rangle|$ and $\boldsymbol{A}(\boldsymbol{U}) := (\langle \boldsymbol{A}_1, \boldsymbol{U}\rangle, \langle \boldsymbol{A}_2, \boldsymbol{U}\rangle, \dots, \langle \boldsymbol{A}_n, \boldsymbol{U}\rangle)$
- $K^\circ = \{\boldsymbol{A}(\boldsymbol{U}) : \boldsymbol{U} \in S_1^m\}$ and
$$\boldsymbol{A}(\boldsymbol{U}) \in C \cdot B_\infty^n + \boldsymbol{A}(\tilde{\boldsymbol{U}}) \iff f_{\boldsymbol{U}}(\tilde{\boldsymbol{U}}) \leqslant C$$
- Covering $\boldsymbol{A}(\boldsymbol{U})$ with a cube $\iff$ Minimizing $f_{\boldsymbol{U}}$

Mirror descent framework

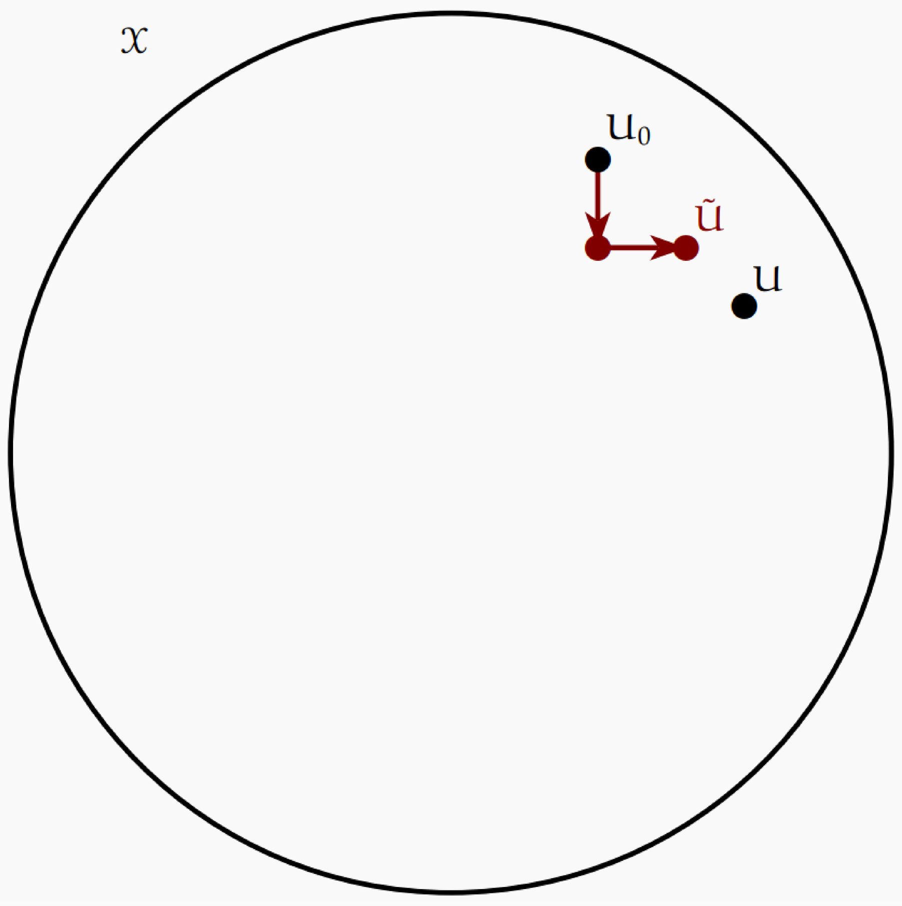
- Fix a domain $\mathcal{D} \subseteq \mathbb{R}^m$ and a subset $\mathcal{X} \subseteq \overline{\mathcal{D}}$, start $\boldsymbol{x}_0 \in \mathcal{X} \cap \mathcal{D}$
- L-Lipschitz convex $f : \mathcal{X} \to \mathbb{R}$ and a ρ-strongly convex $\Phi : \mathcal{D} \to \mathbb{R}$
- Iterate
$$\nabla \Phi(\boldsymbol{y}_{t+1}) := \nabla \Phi(\boldsymbol{x}_t) - \eta \boldsymbol{g}_t$$
$$\boldsymbol{x}_{t+1} := \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{D}} D_\Phi(\boldsymbol{x}, \boldsymbol{y}_{t+1}),$$
where $\boldsymbol{g}_t \in \partial f(\boldsymbol{x}_t)$ and $D_\Phi(\boldsymbol{x}, \boldsymbol{y}) := \Phi(\boldsymbol{x}) - \Phi(\boldsymbol{y}) - \langle \nabla \Phi(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle$.

Mirror descent (Bubéck '15)

After T iterations with $\eta := \frac{1}{L}\sqrt{\frac{2\rho D_\Phi(\boldsymbol{x}^*, \boldsymbol{x}_0)}{T}}$ we get a sequence $\boldsymbol{x}_t$ with
$$\min_{t<T} f(\boldsymbol{x}_t) - f(\boldsymbol{x}^*) \leqslant L \cdot \sqrt{\frac{2 D_\Phi(\boldsymbol{x}^*, \boldsymbol{x}_0)}{\rho T}},$$
where $\boldsymbol{x}^* := \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{D}} f(\boldsymbol{x})$.

Mirror descent: Matrix Spencer setting

- $\mathcal{X} := \{\boldsymbol{X} \in \mathbb{R}^{m\times m} : \boldsymbol{X} \succeq 0, \mathrm{tr}(\boldsymbol{X}) = 1\}$ and $\Phi(\boldsymbol{X}) := \mathrm{tr}(\boldsymbol{X} \log \boldsymbol{X})$
- $f_{\boldsymbol{U}}(\boldsymbol{X}) := \max_{i \in [n]} |\langle \boldsymbol{A}_i, \boldsymbol{X} - \boldsymbol{U}\rangle|$ is $\|\cdot\|_{S_1^m}$-Lipschitz for $\boldsymbol{A}_i \in S_\infty^m$
- $D_\Phi(\boldsymbol{X}, \boldsymbol{Y}) = S(\boldsymbol{X} \| \boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{X}(\log \boldsymbol{X} - \log \boldsymbol{Y}))$



- Closed formula:
$$\boldsymbol{U}_{t+1} = \frac{\exp\left(\log(\boldsymbol{U}_0) - \eta \sum_{i=0}^t \boldsymbol{G}_i\right)}{\mathrm{tr}\left(\exp\left(\log(\boldsymbol{U}_0) - \eta \sum_{i=0}^t \boldsymbol{G}_i\right)\right)},$$
where $\boldsymbol{G}_t \in \partial f(\boldsymbol{U}_t) \subseteq \{\pm \boldsymbol{A}_1, \dots, \pm \boldsymbol{A}_n\}$.
- Key observation: $\boldsymbol{U}_t$ does not depend on the *order* of the gradients!
- After $T := n$ iterations, we have at most
$$\sum_{t=0}^n \binom{t+2n-1}{2n-1} \leqslant (n+1) \cdot \binom{3n}{n} \leqslant 2^{O(n)} \text{ centers } \boldsymbol{U}_t.$$
- Each cube is scaled by
$$\sqrt{\frac{S(\boldsymbol{U}\|\boldsymbol{U}_0)}{n}} \leqslant \sqrt{\frac{\log n}{n}},$$
giving a $\sqrt{n \log n}$ partial coloring bound.
- Pick $2^{O(n)}$ starting points $\boldsymbol{U}_0$, total of $2^{O(n)} \cdot 2^{O(n)} \leqslant 2^{O(n)}$ centers!

Theorem (Dadush, Jiang, Reis '22)

There exists a net of $|\mathcal{N}| = 2^{O(n)}$ starting points so that for any $\boldsymbol{U} \in \mathcal{X}$,
$$\min_{\boldsymbol{U}_0 \in \mathcal{N}} S(\boldsymbol{U}\|\boldsymbol{U}_0) \lesssim \log(m^2/n).$$

Quantum relative entropy net

Key lemma (Dadush, Jiang, Reis '22)

Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{X}$ satisfy $\|\boldsymbol{X} - \boldsymbol{Y}\|_{S_\infty^m} \leqslant \varepsilon$ for some $\varepsilon \geqslant 1/m$. Then
$$S(\boldsymbol{X}\|\boldsymbol{Y}') \leqslant \log(2m\varepsilon),$$
where $\boldsymbol{Y}' := \frac{1}{2} \cdot \boldsymbol{Y} + \frac{1}{2} \cdot \frac{\boldsymbol{I}_m}{m}$.

Covering S_1^m with S_∞^m [HPV17]

S_1^m can be covered with $2^{O(n)}$ translates of $\frac{m}{n} S_\infty^m$.

- Combining the two results gives a net with quantum relative entropy
$$\log(2m \cdot m/n) = \log(2m^2/n).$$

Further applications

Theorem (Dadush, Jiang, Reis '22)

$\mathrm{vb}_n(S_\infty^{m,h}, S_\infty^{m,h}) \lesssim \sqrt{n \log(hm/n)}$ for h-block diagonal matrices.

- Sketch: interpolate between two nets
$$N\left(S_1^m, \frac{m}{n} S_\infty^m\right) \leqslant 2^{O(n)} \text{ and } N\left(B_1^m, \frac{\log(2m/n)}{n} B_\infty^m\right) \leqslant 2^{O(n)}.$$

Theorem (Dadush, Jiang, Reis '22)

$\mathrm{vb}_n(S_p^m, S_q^m) \lesssim \sqrt{n \cdot \min(p, \log(2m^2/n))} \cdot \min(1, m/n)^{1/p - 1/q}$.

- Sketch: use a different mirror map $\Phi(\boldsymbol{X}) := \frac{1}{2(p-1)} \|\boldsymbol{X}\|_{S_p^m}^2$.