Approximating Hereditary Discrepancy via Small Width Ellipsoids

Aleksandar Nikolov Kunal Talwar

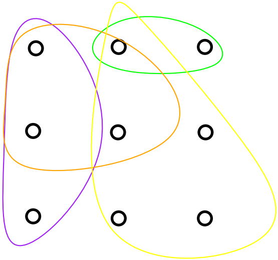
Rutgers University

MSR SVC

Outline

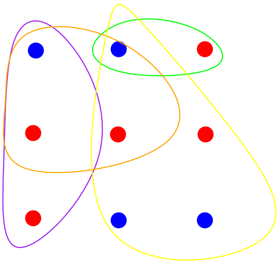
Discrepancy of Set Systems

Given a collection of m subsets $\{S_1, \ldots, S_m\}$ of a size n universe U.



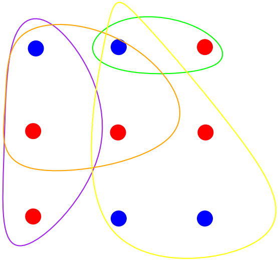
Discrepancy of Set Systems

Color each universe element red or blue, so that each set is as balanced as possible.

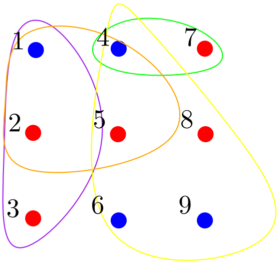


Discrepancy: maximum imbalance (above: 1).

Matrix Representation



Matrix Representation



Matrix Representation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 7 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

Matrix Representation

$$\left(\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 7 & 1 & 1 \end{array}\right) \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

$$\text{disc}(A) = \min_{x \in \{\pm 1\}^n} \|Ax\|_\infty$$

Hereditary Discrepancy

For an $m \times n$ matrix A:

- *Discrepancy*:

$$\text{disc}(A) = \min_{x \in \{\pm 1\}^n} \|Ax\|_\infty$$

- *Hereditary Discrepancy*

$$\text{herdisc}(A) = \max_{S \subseteq [n]} \text{disc}(A|_S)$$

- $A|_S$: submatrix of columns indexed by S
  - corresponds to restricted set system $\{S_1 \cap S, \ldots, S_m \cap S\}$.

Some Applications

- *Rounding*:[Lovász, Spencer, and Vesztergombi, 1986] For any $y \in [-1,1]^n$, there exists $x \in \{\pm 1\}^n$ such that $\|Ax - Ay\|_\infty \leq 2\,\mathrm{herdisc}(A)$.
  - efficient, if discrepancy solutions can be computed efficiently
  - used e.g. in [Rothvoß, 2013].
- *Sparsification*: Constructing ϵ-approximations, and ϵ-nets.
- *Private Data Analysis*:[Nikolov, Talwar, and Zhang, 2013] Lower bounds on the necessary error to prevent a privacy breach.

Classical Results

- [Spencer, 1985] When $A \in [-1,1]^{m\times n}$, $\mathrm{herdisc}(A) = O(\sqrt{n \log \frac{m}{n}})$.
- [Beck and Fiala, 1981] When $A = (a_i)_{i=1}^n$, and $\forall i : \|a_i\|_1 \leq 1$, $\mathrm{herdisc}(A) \leq 2$.
- [Banaszczyk, 1998] When $A = (a_i)_{i=1}^n$, and $\forall i : \|a_i\|_2 \leq 1$, $\mathrm{herdisc}(A) \leq O(\sqrt{\log m})$.
  - Komlos Conjecture: $\mathrm{herdisc}(A) \leq O(1)$.

Hardness

- [Charikar, Newman, and Nikolov, 2011] NP-hard to distinguish between $\mathrm{disc}(A) = 0$ and $\mathrm{disc}(A) = \Omega(\sqrt{n})$ for A and $O(n) \times n$ matrix.
- [Austrin, Guruswami, and Håstad, 2013] NP-hard to approximate herdisc to within a factor of 2.
  - Is there super-constant hardness?
- The problem "$\mathrm{herdisc}(A) \leq t$?" is in Π_2^{P}
  - Is it in NP? Is it Π_2^{P}-hard?

Approximating Discrepancy

- [Bansal, 2010] If $\text{herdisc}(A) \le D$, can find an x such that $\|Ax\|_\infty \le O(D \log m)$.
    - But it's possible that $\|Ax\|_\infty \ll D$
- [Lovász, Spencer, and Vesztergombi, 1986; Matoušek, 2013] A determinant lower bound for $\text{herdisc}(A)$ is tight within a factor of $O(\log^{3/2} m)$. But not efficient!
- [Nikolov, Talwar, and Zhang, 2013] An $O(\log^3 m)$-approximation to $\text{herdisc}(A)$ by relating it to the noise complexity of an efficient differentially private algorithm.

Approximating Discrepancy

- [Bansal, 2010] If $\mathrm{herdisc}(A) \leq D$, can find an x such that $\|Ax\|_\infty \leq O(D \log m)$.
  - But it's possible that $\|Ax\|_\infty \ll D$
- [Lovász, Spencer, and Vesztergombi, 1986; Matoušek, 2013] A determinant lower bound for $\mathrm{herdisc}(A)$ is tight within a factor of $O(\log^{3/2} m)$. But not efficient!
- [Nikolov, Talwar, and Zhang, 2013] An $O(\log^3 m)$-approximation to $\mathrm{herdisc}(A)$ by relating it to the noise complexity of an efficient differentially private algorithm.

This work: An $O(\log^{3/2} m)$-approximation to $\mathrm{herdisc}(A)$.

- Simpler, more direct proof.

Our Result

Theorem

There exists an efficiently computable function f, s.t.

$$\frac{c}{\log m} f(A) \leq \mathrm{herdisc}(A) \leq C\sqrt{\log m}\, f(A),$$

for absolute constants c, C.

- herdisc(A) is a max over 2^n subsets of a min over 2^n colorings
  - No easy to ceritfy *upper* or *lower* bound
- We prove a *simple geometric certificate* gives both upper and lower bounds.
- First (approximate) formulation of herdisc as convex program.

Outline

The Min-Width Ellipsoid

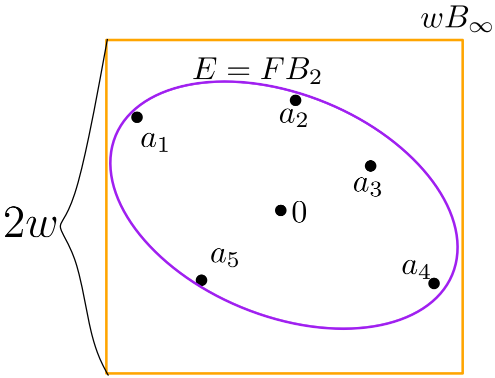
(Centrally symmetric) ellipsoid: $E = FB_2^m$.
Hypercube: $B_\infty^m = [-1, 1]^m$.

Convex Program (MWE): Let $A = (a_1, \dots, a_n)$, $a_i \in \mathbb{R}^m$.

$$f(A) = \min w$$
$$\text{over } E,\ w \text{ subject to}$$
$$\{a_1, \dots, a_m\} \subseteq E \subseteq wB_\infty$$

The Min-Width Ellipsoid

Minimize width w over all E and w s.t. $\{a_1, \ldots, a_m\} \subseteq E \subseteq wB_\infty$



Proof Strategy

- *Upper Bound*: $\operatorname{herdisc}(A) \le C\sqrt{\log m} f(A)$
  - *Banaszczyk's discrepancy theorem*.
- *Lower Bound*: $\frac{c}{\log m} \le \operatorname{herdisc}(A)$
  - Extract a lower bound on $\operatorname{herdisc}(A)$ from any solution to a *convex dual* of the (MWE) program.
  - Bound follows from *strong duality*.

Outline

Banaszczyk's Theorem

Theorem ([Banaszczyk, 1998])

Let $A = (a_1, \ldots, a_n)$*, where* $\|a_i\|_2 \leq 1$ *for all* i*. Let* $K \subseteq \mathbb{R}^m$ *be a convex body so that*

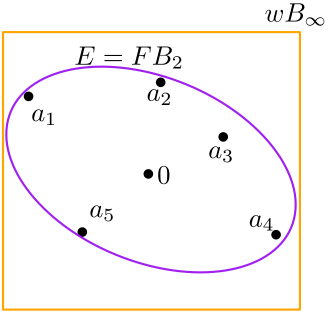
$$\Pr[g \in K] \geq \frac{1}{2},$$

for $g \sim N(0,1)^m$ *a standard guassian. Then* $\exists x \in \{-1,1\}^n$ *so that*

$$Ax \in 10K.$$

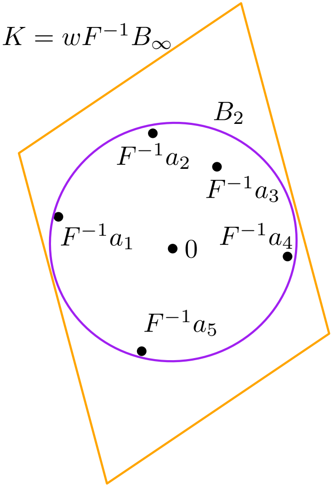
Applying the Theorem

Take some $E = FB_2$ and w s.t. $\{a_1, \ldots, a_m\} \subseteq E \subseteq wB_\infty$.

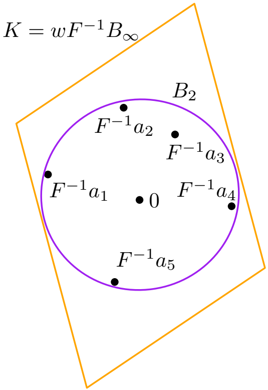


Applying the Theorem

$\{F^{-1}a_1, \ldots, F^{-1}a_m\} \subseteq B_2 \subseteq K.$



Applying the Theorem

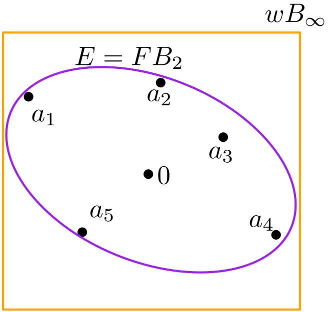


- Every facet of K is at least distance 1 from the origin.
  - Because $B_2 \subseteq K$.
- Chernoff bound + Union bound: $\Pr[g \in C\sqrt{\log m}\ K] \geq \frac{1}{2}$.
- By B.'s Theorem: $\exists x \in \{-1, 1\}^n$, so that $F^{-1}Ax \in K$
  - $\Leftrightarrow Ax \in w \cdot C\sqrt{\log m}\ B_\infty$.
  - $\Leftrightarrow \|Ax\|_\infty \leq w \cdot C\sqrt{\log m}$.
  - $\mathrm{disc}(A) \leq w \cdot C\sqrt{\log m}$.

The Bound is Hereditary

The bound immediately works for $A|_S$:

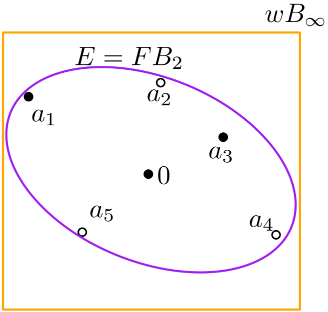
- $\{a_i\}_{i \in S} \subseteq \{a_1 \ldots, a_n\} \subseteq E \subseteq wB_\infty$.
- I.e. E an w are feasible for $A|_S$



The Bound is Hereditary

The bound immediately works for $A|_S$:

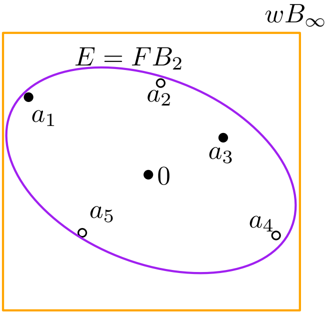
- $\{a_i\}_{i \in S} \subseteq \{a_1 \ldots, a_n\} \subseteq E \subseteq wB_\infty$.
- I.e. E an w are feasible for $A|_S$



The Bound is Hereditary

The bound immediately works for $A|_S$:

- $\{a_i\}_{i \in S} \subseteq \{a_1 \ldots, a_n\} \subseteq E \subseteq wB_\infty$.
- I.e. E an w are feasible for $A|_S$
- $\text{herdisc}(A) \le w \cdot C\sqrt{\log m}$.



Outline

Spectral Lower Bound

Smallest singular value: $\sigma_{\min}(A) = \min_x \frac{\|Ax\|_2}{\|x\|_2}$.

Proposition

For any $m \times n$ matrix A, any diagonal $P \geq 0$, $\mathrm{tr}(P^2) = 1$,

$$\mathrm{disc}(A)^2 \geq n\sigma_{\min}^2(PA).$$

Comes from (the dual of) a convex relaxation of $\mathrm{disc}(A)$.

Spectral Lower Bound

Smallest singular value: $\sigma_{\min}(A) = \min_x \frac{\|Ax\|_2}{\|x\|_2}$.

Proposition

For any $m \times n$ matrix A, any diagonal $P \geq 0$, $\mathrm{tr}(P^2) = 1$,

$$\mathrm{disc}(A)^2 \geq n\sigma_{\min}^2(PA).$$

Comes from (the dual of) a convex relaxation of $\mathrm{disc}(A)$.

Implies for any $S \subseteq [n]$:

$$\mathrm{herdisc}(A)^2 \geq |S|\sigma_{\min}^2(PA|_S).$$

Proof.

$$\mathrm{disc}(A)^2 = \min_{x \in \{-1,1\}^n} \max_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j \right)^2$$

□

Proof.

$$\mathrm{disc}(A)^2 = \min_{x \in \{-1,1\}^n} \max_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j \right)^2$$

$$\geq \min_{x \in \{-1,1\}^n} \sum_{i=1}^{m} P_{ii}^2 \left(\sum_{j=1}^{n} A_{ij} x_j \right)^2 \quad (\text{avaraging})$$

□

Proof.

$$
\begin{aligned}
\mathrm{disc}(A)^2 &= \min_{x \in \{-1,1\}^n} \max_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j \right)^2 \\
&\geq \min_{x \in \{-1,1\}^n} \sum_{i=1}^{m} P_{ii}^2 \left(\sum_{j=1}^{n} A_{ij} x_j \right)^2 \quad \text{(avaraging)} \\
&= \min_{x \in \{-1,1\}^n} \|PAx\|_2^2
\end{aligned}
$$

□

Proof.

$$\mathrm{disc}(A)^2 = \min_{x\in\{-1,1\}^n} \max_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j\right)^2$$

$$\geq \min_{x\in\{-1,1\}^n} \sum_{i=1}^{m} P_{ii}^2 \left(\sum_{j=1}^{n} A_{ij}x_j\right)^2 \quad (\text{avaraging})$$

$$= \min_{x\in\{-1,1\}^n} \|PAx\|_2^2$$

$$\geq n\sigma_{\min}^2(PA) \quad (x \in \{-1,1\}^n \Rightarrow \|x\|_2 = n^{1/2})$$

□

Dual of (MWE)

Primal

$$
\begin{aligned}
f(A) = \min w& \\
\text{subject to}& \\
\{a_1, \ldots, a_m\} \subseteq E \subseteq wB_\infty&
\end{aligned}
$$

Nuclear norm: $\|M\|_{S_1}$ is equal to the sum of singular values of M.

Dual

$$
\begin{aligned}
f(A) = \max \|PAQ\|_{S_1}& \\
\text{subject to}& \\
P, Q \geq 0, \text{ diagonal}& \\
\operatorname{tr}(P^2) = \operatorname{tr}(Q^2) = 1&
\end{aligned}
$$

Spectral LB from the Dual

Lemma

For any feasible P and Q, there exists a set $S \subseteq [n]$ such that

$$|S|\sigma_{\min}(PA|_S)^2 \geq \frac{c^2}{(\log m)^2}\|PAQ\|_{S_1}^2.$$

The set S is efficiently computable.

Spectral lowerbound $\Rightarrow$ herdisc$(A) \geq \frac{c}{\log m} f(A)$.

Restricted Invertibility Principle

Theorem ([Bourgain and Tzafriri, 1987; Spielman and Srivastava, 2010])

Assume that any two nonzero singular values σ_i, σ_j of the $m \times k$ matrix M satisfy $\frac{1}{2} \leq \frac{\sigma_i}{\sigma_j} \leq 2$. Then there exists a subset $S \subseteq [k]$ such that

$$|S|\sigma_{\min}(M|_S)^2 \geq \frac{1}{64k}\|M\|_{S_1}^2$$

Restricted Invertibility Principle

Theorem ([Bourgain and Tzafriri, 1987; Spielman and Srivastava, 2010])

Assume that any two nonzero singular values σ_i, σ_j of the $m \times k$ matrix M satisfy $\frac{1}{2} \leq \frac{\sigma_i}{\sigma_j} \leq 2$. Then there exists a subset $S \subseteq [k]$ such that

$$|S|\sigma_{\min}(M|_S)^2 \geq \frac{1}{64k}\|M\|_{S_1}^2$$

Simple transformations to PAQ to get a matrix M:

- M satisfies the assumption of the restricted invertibility principle
- $\|M\|_{S_1} \geq \frac{\sqrt{k}}{\log m}\|PAQ\|_{S_1}$
  - Captures a large fraction of the dual value
- All columns of M are projections of columns of PA
  - Spectral lower bounds for M lower bound herdisc(A)

Outline

Conclusion

This work:

- $O(\log^{3/2} m)$ approximation for hereditary discrepancy
- *Direct* proof using geometric techniques
- Approximate *characterization* of hereditary discrepancy as a *convex program*
    - Can use tools of convex analysis to understand herdisc.

Conclusion

This work:

- $O(\log^{3/2} m)$ approximation for hereditary discrepancy
- *Direct* proof using geometric techniques
- Approximate *characterization* of hereditary discrepancy as a *convex program*
    - Can use tools of convex analysis to understand herdisc.

Open:

- $2 + \epsilon$ hardness of approximating hereditary discrepancy
- How far can $f(A)$ be from herdisc(A)?
- Constructive proof of Banaszczyk's theorem
- Improve the approximation ratio

Thank you!

Per Austrin, Venkatesan Guruswami, and Johan Håstad. $(2+\epsilon)$-sat is np-hard. *ECCC TR13-159, 2013.*, 2013.

Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4): 351–360, 1998.

N. Bansal. Constructive algorithms for discrepancy minimization. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 3–10. IEEE, 2010.

József Beck and Tibor Fiala. Integer-making theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

J. Bourgain and L. Tzafriri. Invertibility of large submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel journal of mathematics*, 57(2):137–224, 1987.

M. Charikar, A. Newman, and A. Nikolov. Tight hardness results for minimizing discrepancy. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1607–1614. SIAM, 2011.

L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151–160, 1986.

Jiří Matoušek. The determinant bound for discrepancy is almost tight. *Proceedings of the American Mathematical Society*, 141(2):451–460, 2013.

Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, STOC '13, pages 351–360, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2029-0. doi: 10.1145/2488608.2488652. URL http://doi.acm.org/10.1145/2488608.2488652.

Thomas Rothvoß. Approximating bin packing within o (log opt* log log opt) bins. In *Foundations of Computer Science (FOCS), 2013 54th Annual IEEE Symposium on*, 2013.

Joel Spencer. Six standard deviations suffice. *Transactions of the American Mathematical Society*, 289(2):679–706, 1985.

D.A. Spielman and N. Srivastava. An elementary proof of the restricted invertibility theorem. *Israel Journal of Mathematics*, pages 1–9, 2010.