Discrepancy and approximation

Nikhil Bansal
Thomas Rothvoss
Outline

Basics + classical results
New Developments

Detailed Application

• Additive $O(\log \text{OPT} \log \log \text{OPT})$ for bin-packing
Combinatorial Discrepancy

Universe: \( U = [1, \ldots, n] \)
Subsets: \( S_1, S_2, \ldots, S_m \)

Find \( \chi: [n] \rightarrow \{-1, +1\} \) to

Minimize \( |\chi(S)|_\infty = \max_S |\sum_{i \in S} \chi(i)| \)

If \( A \) is a \( m \times n \) matrix.

\[
\text{Disc}(A) = \min_{x \in \{-1, 1\}^n} |Ax|_\infty
\]
Applications

CS: Computational Geometry, Approximation, Monte-Carlo simulation, Machine learning, Complexity, Pseudo-Randomness, …

Math: Dynamical Systems, Combinatorics, Mathematical Finance, Number Theory, Ramsey Theory, Algebra, Measure Theory, …
Hereditary Discrepancy

Discrepancy a useful measure of **complexity** of a set system

But not so **robust**

Hereditary discrepancy:

\[
\text{herdisc} (U,S) = \max_{U' \subseteq U} \text{disc} (U', S_{|U'})
\]

**Robust** version of discrepancy

(How to certify \( \text{herdisc} < D \)? In **NP**?)
Connection to Approximation
A Rounding Result

Lovasz-Spencer-Vesztergombi’86: Given any matrix $A$, and $x \in \mathbb{R}^n$, can round $x$ to $\tilde{x} \in \mathbb{Z}^n$ s.t.

$$|Ax - A\tilde{x}|_\infty < \text{Herdisc}(A)$$

**Proof:** Round the bits of $x$ one by one.

$x_1$: blah .0101101 $\leftarrow (-1)$  
$x_2$: blah .1101010  
$\ldots$  
$x_n$: blah .0111101 $\leftarrow (+1)$

Key Point: Low discrepancy coloring guides our updates!

Doerr’01: Error $\leq \left(1 - \frac{1}{m}\right)\text{Herdisc}(A)$

Error $= \text{herdisc}(A) \left(\frac{1}{2k}\right)$. 

Ax=b
Rounding

Totally Unimodular (TU) Matrices
Ghouila-Houri’56: Matrix is TU iff given any subset of columns, there is ±1 coloring s.t. row sum in {-1,0,1}

LSV’86 result guarantees existence of good rounding. How to find it efficiently?

Thm [B’10]. Can round efficiently, so that

\[
\text{Error} \leq O\left(\sqrt{\log m \log n}\right) \text{Herdisc}(A)
\]

Use SDPs + random walks
How to bound discrepancy?

Spencer’85: Any 0-1 matrix (n x n) has disc ≤ $6 \sqrt{n}$
(Random coloring gives $\sqrt{n \log n}$)

Partial Coloring method (Beck’82), refinement: Entropy Method
(very powerful, based on pigeonhole principle)

Almost all combinatorial discrepancy results based on it
(exceptions: Banaszczyk’s result for Komlos conjecture)

Non-algorithmic
Algorithmic variants

B.’10: Algorithmic $O(\sqrt{n})$ for Spencer (SDP + Entropy method)

Lovett-Meka’12: Much simpler + cleaner

Stronger variant of Partial Coloring Lemma
Generalizes iterated rounding.

Rothvoss’13: Bin Packing
B., Charikar, Krishnaswamy, Li’14: Broadcast Scheduling

Extended by Rothvoss’14, Ronen-Singh’14 (Convex Geometry)
A different view

Spencer’85: Any 0-1 matrix (n x n) has disc $\leq 6 \sqrt{n}$

Gluskin’87: Convex geometric approach

Consider polytope $P(t) = -t \mathbf{1} \leq Ax \leq t \mathbf{1}$
P(t) contains a point in $\{-1,1\}^n$ for $t = 6 \sqrt{n}$

Gluskin’87: If $K$ symmetric, convex with large (Gaussian) volume ($> 2^{-n/100}$) then $K$ contains a point with many coordinates {-1,+1}
Algorithmic version

Rothvoss’14:
Pick a random point $y$, and find closest point in $K \cap [-1,1]^n$

Eldan, Singh’14: Pick a random direction; optimize over $K \cap [-1,1]^n$

Extends Lovett-Meka partial coloring lemma
Lower Bounds

Various methods: Spectral, Fourier analytic, …

Determinant lower bound:
\[ \text{detlb}(A) \leq \text{herdisc} (A) \]  [Lovasz et al. 86]

\[ \text{herdisc}(A) \leq \text{polylog}(n,m) \text{ detlb}(A) \]  [Matousek’11]
(SDP duality)

Polylog approximation for \( \text{herdisc}(A) \)
[Nikolov, Talwar, Zhang’13, Nikolov, Talwar’15, Nikolov Matousek’15]
Details + Proof Sketches
Vector Discrepancy

**Exact:** Min $t$

$$-t \leq \sum_j a_{ij} x_j \leq t$$ for all rows $i$

$x_j \in \{-1,1\}$ for each $j$

**SDP:** vecdisc($A$)

$$\min t$$

$$\left| \sum_i a_{ij} v_j \right|_2 \leq t$$ for all rows $i$

$$|v_j|_2 = 1$$ for each $j$
Is vecdisc a good relaxation?

Not directly. vecdisc(A) = 0 even if disc(A) very large

[Charikar, Newman, Nikolov’11]

**NP-Hard**: Whether disc(A) = 0 or Ω(√n) for Spencer’s setting?

Let hervecdisc(A) = \max_S vecdisc( A|_S )

(not clear how to compute this in poly time)

Thm [B’10]: disc(A) = \( O\left(\sqrt{\log m \log n}\right) \) hervecdisc(A)
Algorithm: Correlated random walk (each step round an SDP)
Fix a variable if it reaches -1 or +1.

Analysis: Few steps to reach a vertex (walk has high variance)
Disc($S_i$) does a random walk (with low variance)
An SDP

Hereditary disc. \( \lambda \Rightarrow \) SDP is feasible (for any set of alive variables)

SDP: \( |\sum a_{ij} v_j|^2 \leq \lambda^2 \) for each row \( i \).
\( |v_j|^2 = 1 \) for each active element \( j \).

Project on random Gaussian \( g \) (\( \eta_j = g \cdot v_j \))

**Lemma:** For any \( v \in \mathbb{R}^n \), \( g \cdot v \sim N(0,|v|^2) \)

1. Each \( \eta_j \sim N(0,1) \)
2. For each row \( i \), \( \sum_j a_{ij} \eta_j = g \cdot (\sum_j a_{ij} v_j) \sim N(0, \leq \lambda^2) \)

Use \( \eta_j \) to update the color of element \( j \).
Algorithm Overview

Initially: Start with coloring $x_0 = (0,0,0, \ldots,0)$ at $t = 0$.
Time $t$: Update coloring as $x_t = x_{t-1} + \gamma (\eta_1^t, \ldots, \eta_n^t)$ ($\gamma$ tiny: $1/n$)

Color of element $j$ increment = $\gamma \mathcal{N}(0,1)$
Fixed if reaches -1 or +1.

Disc(row $i$) increment = $\gamma \mathcal{N}(0, \leq \lambda^2)$

At time $O(1/\gamma^2)$
1: Half the elements reaches -1 or +1.
2: Each row has $\leq \lambda$ discrepancy in expectation.
Quite Robust Method

E.g. Can adjust the walk dynamically

For Spencer’s result:
Expected discrepancy for a set $O(n^{1/2})$,
but some random walks will deviate by up to $(\log n)^{1/2}$ factor

Tune down the variance of dangerous sets (not too many)

Another variant by [Harvey, Schwartz, Singh’14]
New Entropy Method
Entropy method

[Beck, Spencer 80’s]: Given an m x n matrix A, there is a partial coloring satisfying \(|a_ix| \leq \lambda_i |a_i|_2\)

provided \(\sum_i g(\lambda_i) \leq \frac{n}{5}\)

\[ g(\lambda_i) \approx \ln \left( \frac{1}{\lambda_i} \right) \quad \text{if} \quad \lambda_i < 1 \]

\[ \approx e^{-\lambda_i^2} \quad \text{if} \quad \lambda_i \geq 1 \]

E.g. Can get 0 discrepancy on \(n/(10 \log n)\) rows

(Non-trivial) Idea: There are lots of colorings \((2^{n/100})\) s.t. for every two x and y \(|a_ix - a_iy| \leq \lambda_i |a_i|_2\) for all row i.

Pick two with large hamming distance and consider \(\frac{1}{2}(x-y)\)
New Condition

[Beck, Spencer 80’s]: Given an m x n matrix A, there is a partial coloring satisfying $|a_ix| \leq \lambda_i |a_i|_2$

provided $\sum_i g(\lambda_i) \leq \frac{n}{5}$

$g(\lambda_i) \approx \ln\left(\frac{1}{\lambda_i}\right)$ if $\lambda_i < 1$

$\approx e^{-\lambda_i^2}$ if $\lambda_i \geq 1$

E.g. Can get 0 discrepancy on $O(n/\log n)$ rows

Lovett-Meka’12: Partial Coloring exists with $|a_ix| \leq \lambda_i |a_i|_2$

if $\sum_i \exp\left(-\lambda_i^2\right) \leq \frac{n}{5}$

Allows 0 discrepancy on $O(n)$ rows

Moreover algorithmic
Lovett Meka Algorithm

Random walk, $\gamma \sim N(0,1)$ in each dimension

a) Fix $j$ if $x_j = \pm 1$

b) If row $A_i$ gets tight ($\text{disc}(A_i) = \lambda_i |a_i|_2$)

Move in space $A_i x = 0$ (not violate discrepancy)

Idea: Walk makes progress as long as dimension $= \Omega(n)$

After $\frac{1}{\gamma^2}$ steps: $\Pr[ \text{Row } A_i \text{ tight}] \approx \exp(-\lambda_i^2)$

As $\sum_i \exp(-\lambda_i^2) \leq \frac{n}{5}$ so $n/5$ tight rows in expectation

Win-Win: Either lots of variables already at $\pm 1$, or else walk makes fast progress
Iterated Rounding

Simple, yet very powerful

Karmarkar Karp’82 (bin packing)
Jain’98 (survivable network design)
Singh Lau’06 (deg bdd spanning trees)
...

Obtain classic results in this framework.
Iterated Rounding

LP: \[ \text{max } cx \]
\[ Ax \leq b \]
\[ 0 \leq x \leq 1 \]

**Obs:** There is an optimum solution with \( \geq n-m \) variables set to 0-1.

**Vertex:** determined by solution of \( n \) (tight) linearly independent constraints
(basic feasible solutions)
Iterated Rounding

LP \quad \text{max } cx
Ax \leq b
0 \leq x \leq 1

n > m \text{ variables (say } n = m+1) \quad x_1, ..., x_n

Key point: If drop another constraint, get another integral variable

Note: objective can only go up during the iterations

Hope: Dropping a constraint does not mess up things much
Iterated rounding in action

cx

drop

fixed Ax b
Iterated rounding in action

Cleverness: Which constraint to drop?
Comparison with iterated rounding

Iterated rounding: LP with m constraints, if only keep $n/2$ constraints then get $n/2$ integral variables (no control on dropped constrains ($a_i \times b_i$), error up to $|a_i|_1$)

Lovett-Meka Lemma: Can find solution with $\geq n/2$ integral variables with error $\leq \lambda_i |a_i|_2$.

Benefits:
1) Can choose $\lambda_i$’s (as long as $\sum_i \exp(-\lambda_i^2) \leq \frac{n}{5}$ )
2) $|a_i|_2$ vs. $|a_i|_1$ (potentially $\sqrt{n}$ vs $n$ if $a_i$ “spread”)

Potential drawback: $O(\log n)$ phases; error could accumulate
Geometric view

Lovett-Meka condition: Precisely that polytope has large volume.

\[ \gamma_n(|a_i x| \leq \lambda_i |a_i|_2) \approx \exp(-\lambda_i^2) \]

Sidak-Khatri Lemma: \( \gamma_n(K \cap Strip) \geq \gamma_n(K) \cdot \gamma_n(Strip) \)

Rothvoss’14 Proof: Random point \( y \)
Far from cube (so far from \( K \cap [-1,1]^n \))
But close to \( K \) (measure concentration)

If very few variables I set to \( \{-1,1\} \).

Closest in \( K \cap [-1,1]^I = \) Closest in \( K \cap [-1,1]^n \)
Contradiction.
Lower Bounds
Lower bounding

If disc(A) > D
\[ |Ax|_\infty \geq D \text{ for all } x \in \{-1,1\}^n \]

(say A: n x n matrix for convenience)

If \( \sigma_{min}(A) \geq D \)
\[ |Ax|_2 \geq \sigma_{min}(A) |x|_2 \]

Could be very weak bound.

Can consider \( \sigma_{min}(PA) \quad P:\text{diag, } \text{tr}(P) = n \)
Determinant Lower Bound

**Thm** (Lovasz Spencer Vesztergombi’86): herdisc(A) $\geq$ detlb(A)

$\text{detlb}(A): \max_k \max_{\{k \times k \text{ submatrix } B \text{ of } A\}} \det(B)^{1/k}$

(simple geometric argument)

**Conjecture (LSV’86):** Herdisc $\leq O(1) \detlb$

**Remark:** For TU Matrices, Herdisc(A) = 1, detlb = 1
(every submatrix has det -1, 0 or +1)

False: $\text{detlb}(A) = 2$ \ herdisc(A) $= \Omega(\log n)$ \ [Hoffman, Palvolgyi]
Matousek’11: \( \text{herdisc}(A) \leq O(\log n \sqrt{\log m}) \) detlb.

**Idea:** Large herdisc \( \rightarrow \) Large hervecdisc

SDP Duality \( \rightarrow \) Dual Witness for large hervecdisc(A).

Dual Witness \( \rightarrow \) Submatrix with large determinant.
SDP Duality

If \( \text{vecdisc}(A) \geq D \), there exist weights \( w_1, \ldots, w_m \geq 0 \) with \( \sum_i w_i \leq 1 \) and \( z_1, \ldots, z_n \geq 0 \) such that \( \sum_j z_j \geq D^2 \)

\[
\sum_i w_i \left( \sum_j a_{ij} x_j \right)^2 \geq \sum_j z_j x_j^2 \quad \text{for all } x \in \mathbb{R}^n.
\]

Why is this a witness?
Proof Sketch

As $\sum_j z_j \geq D^2$

Some set of variables $U$ with $\sum_{j \in U} z_j \geq \frac{D}{\log n}$

and $z_j$ roughly similar

$W^{1/2} A|_U$ has large $\sigma_{\min}$

Use Cauchy-Binet to show that $A$ has large sub-determinant
Concluding Remarks

Algorithmic view has been extremely useful in discrepancy

Some open questions:

Beck Fiala Conjecture
Constructive version of Banaszczyk’s theorem?
Tightness of $\det_{lb}$ bound ($\log$ vs $\log^{3/2}$)
$O(1)$ approximation for herdisc
Thanks!