Discrepancy and SDPs

Nikhil Bansal
(TU Eindhoven, Netherlands)

Outline

Discrepancy Theory

- What is it
- Basic Results (non-constructive)

SDP connection

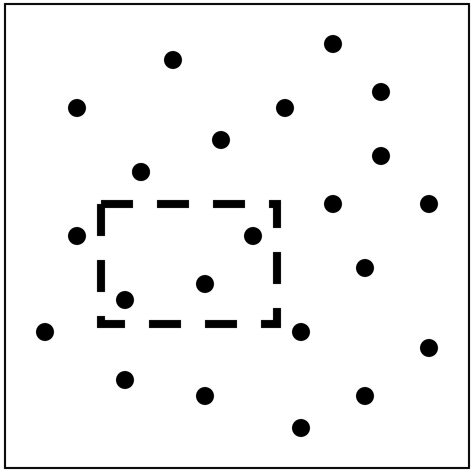
- Algorithms for discrepancy
- New methods in discrepancy (upper/lower bounds)
- Approximation

Discrepancy: Example

Input: n points placed arbitrarily in a grid.

Color them red/blue such that each axis-parallel rectangle is colored as evenly as possible

Discrepancy: max over rect. R (| # red in R - # blue in R |)



Random has about $O(n^{1/2} \log^{1/2} n)$

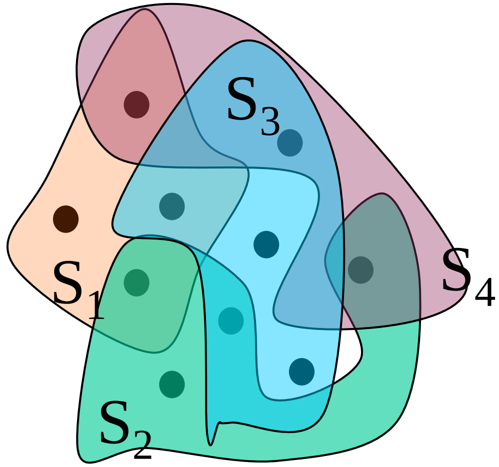
Can achieve $O(\log^{2.5} n)$

Why do we care?

Combinatorial Discrepancy

Universe: U= [1,…,n]

Subsets: $S_1, S_2, \ldots, S_m$

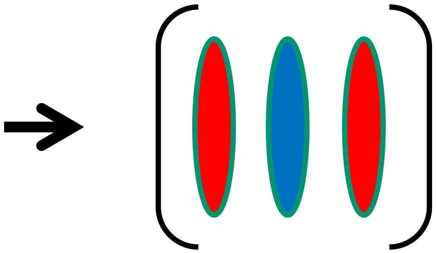


Find χ: [n] ! {-1,+1} to

Minimize $|\chi(S)|_1 = \max_S | \sum_{i \ 2 \ S} \chi(i) |$

If A is a $m \times$ n matrix.

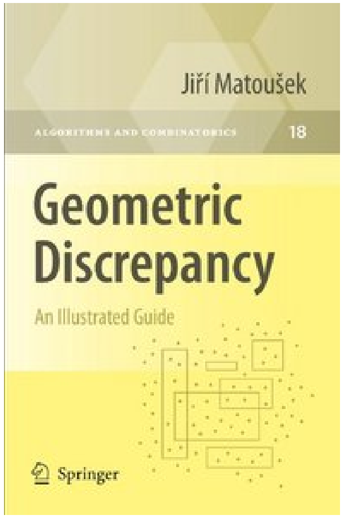
$$\text{Disc}(A) = \min_{x \in \{-1,1\}^n} |Ax|_\infty$$

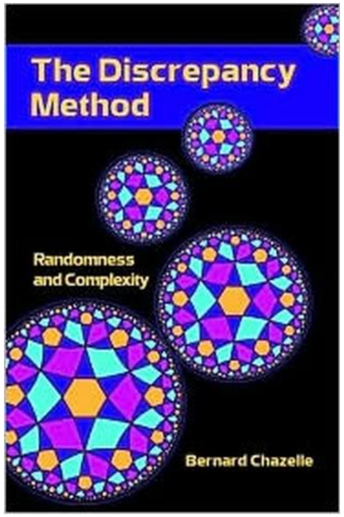


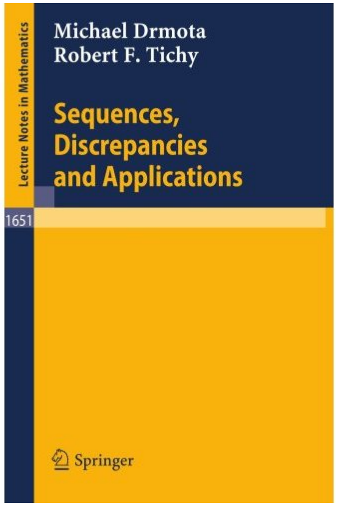
Applications

CS: Computational Geometry, Comb. Optimization, Monte-Carlo simulation, Machine learning, Complexity, Pseudo-Randomness, …

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





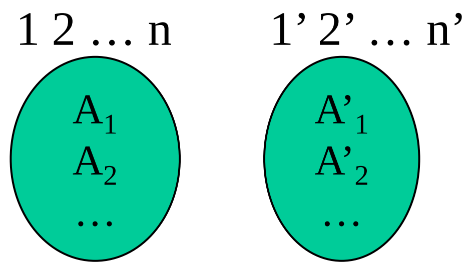


• • • • •

Hereditary Discrepancy

Discrepancy a useful measure of complexity of a set system

But not so robust



$S_i = A_i \cup A'_i$

Discrepancy = 0

Hereditary discrepancy:

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

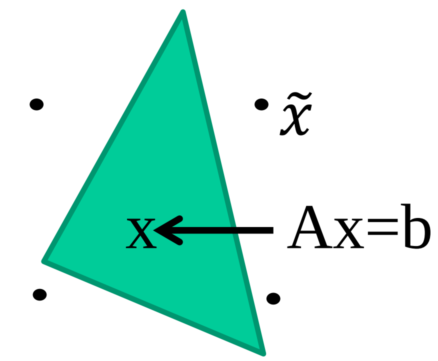
Robust version of discrepancy

(How to certify herdisc < D? In NP?)

Some Applications

Rounding

Lovasz-Spencer-Vesztermgombi'86: Given any matrix A, and $x \in R^n$, can round x to $\tilde{x} \in Z^n$ s.t.
$|Ax - A\tilde{x}|_\infty <$ Herdisc(A)



Proof: Round the bits of x one by one.

x_1: blah .0101101 ← (-1)

x_2: blah .1101010

…

x_n: blah .0111101 ← (+1)

Key Point: Low discrepancy coloring guides our updates!

Error = herdisc(A) ($\frac{1}{2^k}$

Rounding

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, basic method

Refinements

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

Non-constructive Entropy method (very powerful technique)

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

Lovett-Meka'12: (much simpler)

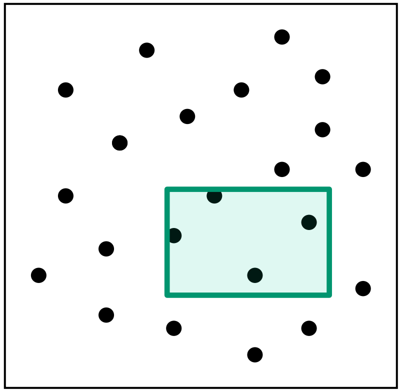
Better variant of "entropy method"

Extends iterated rounding.

Bin-packing: Rothvoss'13: Alg $\leq$ LP + O(log OPT log log OPT)

Karmarkar-Karp'82: Alg $\leq$ LP + O($\log^2$ OPT)

Dynamic Data Structures



N weighted points in a 2-d region.

Weights updated over time.

Query: Given an axis-parallel rectangle R, determine the total weight on points in R.

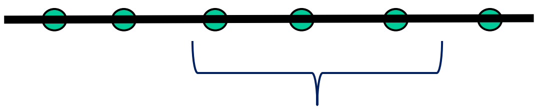
Goal: Preprocess (in a data structure)

1) Low query time
2) Low update time (upon weight change)

Example

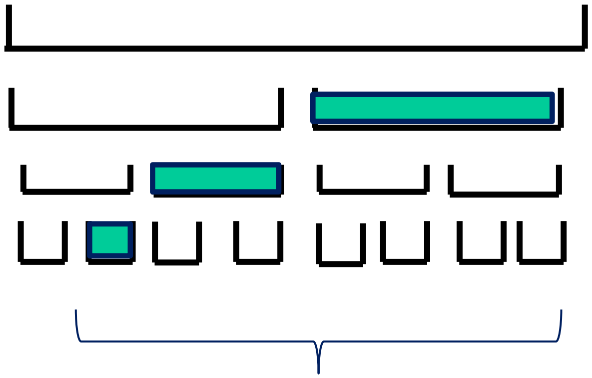
Line: Interval queries

Trivial: Query Time= O(n) Update Time = 1



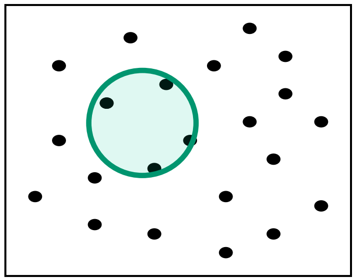
Query time= 1 Update time= O(n^2) (Table of entries W[a,b])

Query = O(log n) Update = O(log n)

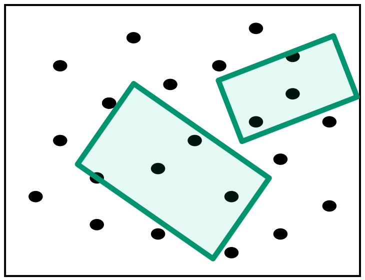


Recursively for 2-d.
$O(\log^2 n, \log^2 n)$

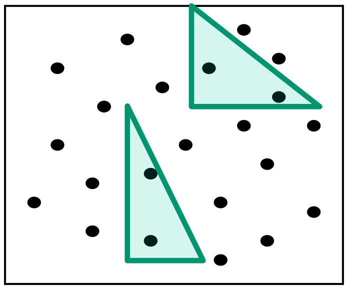
What about other queries?



Circles



arbitrary rectangles



aligned triangle

Turns out $t_q t_u \geq \frac{n^{1/2}}{\log^2 n}$

Reason: Set system S formed by query sets & points has large discrepancy (about $n^{1/4}$)

Larsen-Green'11: $t_q t_u \geq \frac{disc(S)^2}{\log^2 n}$

Lower Bounds

Various methods: Spectral, Fourier analytic, …

Determinant lower bound:

detlb(A) $\leq$ herdisc (A) [Lovasz et al. 86]

herdisc(A) $\leq$ polylog(n,m) detlb(A) [Matousek'11]

(SDP duality)

Polylog approximation for herdisc(A) [Nikolov, Talwar, Zhang'13]

SDP Connection

Vector Discrepancy

Exact: Min t

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

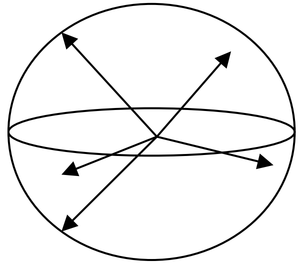
$x_j \in \{-1,1\}$ for each j

SDP: vecdisc(A)

min t

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

$\left| v_j \right|_2 = 1$ 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: disc(A) = 0 or disc(A) very large ?

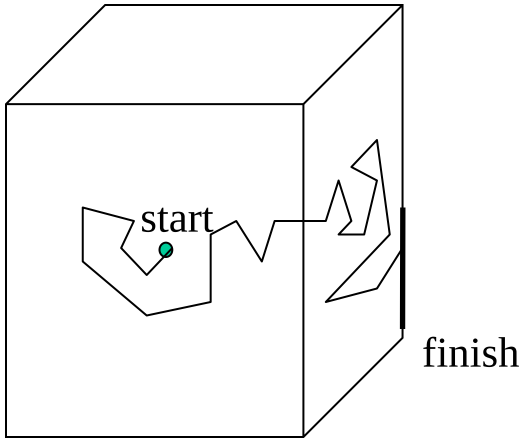
Let hervecdisc(A) = $\max_{S}$ vecdisc($A_{|S}$)

Thm [B'10]: disc(A) = $O\left(\sqrt{\log m \log n}\right)$ hervecdisc(A)

Pf: Algorithm

Algorithm (at high level)

Cube: $\{-1,+1\}^n$



Each dimension: An Element
Each vertex: A Coloring

Algorithm: "Sticky" random walk

Each step generated by rounding a suitable SDP

Move in various dimensions correlated, e.g. $\delta^t_1 + \delta^t_2$ ¼ 0

Analysis: Few steps to reach a vertex (walk has high variance)

Disc(S_i) does a random walk (with low variance)

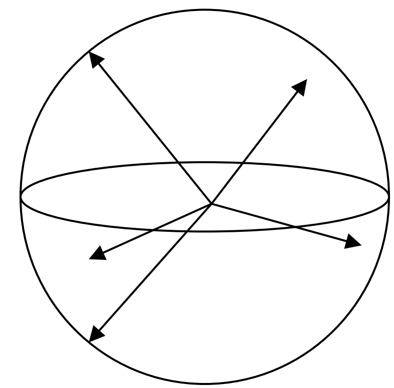
An SDP

Hereditary disc. λ) the following SDP is feasible

SDP: Low discrepancy

$|\sum a_{ij} v_j|^2 \cdot \lambda^2$ for each row i.

$|v_j|^2 = 1$ for each element j.



Obtain v_j 2 R^n

Perhaps v_j can guide us how to update color of element j ?

Trouble: v_j is a vector. Need a real number.

Project on random vector vector g ($\eta_i = g$ ¢v_i)

Seems promising: $\sum_j a_{ij}\, \eta_j = g \cdot \left(\sum_j a_{ij}\, v_j\right)$

Properties of Rounding

Lemma: If g 2 R^n is a random Gaussian, for any v 2 R^n,

g ¢v is distributed as N(0, $|v|^2$)

SDP:

$|v_j|^2$ = 1

$|\sum_j a_{ij} v_j|^2$ · λ^2

1. Each η_j » N(0,1)
2. For each row i,

   $\sum_j a_{ij} \eta_j$ = g ¢($\sum_j a_{ij} v_j$) » N(0, · λ^2)

   (std deviation · λ)

η's will guide our updates to x.

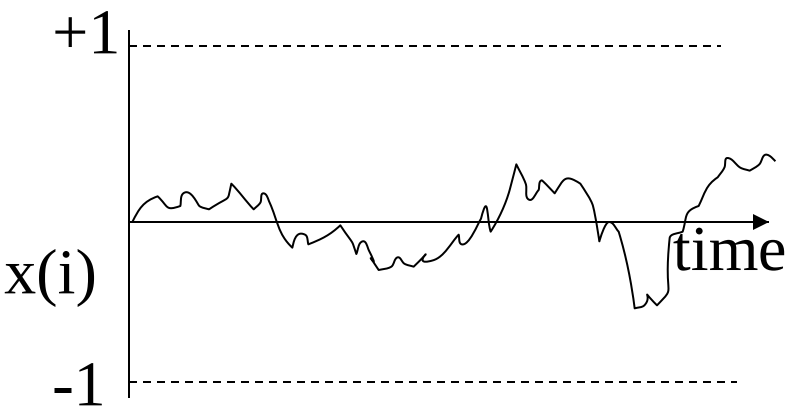
Algorithm Overview

Construct coloring iteratively.

Initially: Start with coloring $x_0 = (0,0,0, \ldots,0)$ at $t = 0$.

At Time t: Update coloring as $x_t = x_{t-1} + \gamma\,(\eta^t_1,\ldots,\eta^t_n)$

(γ tiny: say $1/n$)



$x_t(j) = \gamma\,(\eta^1_j + \eta^2_j + \ldots + \eta^t_j)$

Color of element j: Does random walk over time with step size ¼ $\gamma\,N(0,1)$

Fixed if reaches -1 or +1.

Disc(row i): $\sum_j a_{ij}\, x_t(j)$ does a random walk w/ step $\gamma\,N(0,\cdot\ \lambda^2)$

Analysis

At time T = $O(1/\gamma^2)$

1: With prob. ½, an element reaches -1 or +1.

2: Each row has discrepancy in expectation.

At time T = $O((\log n)/\gamma^2)$

1. Most likely all elements fixed
2. Expected discrepancy for a row = $\lambda\sqrt{\log n}$

(By Chernoff, all have discrepancy $O(\lambda\sqrt{\log n \log m}$)

New Entropy Method

Entropy method

Very powerful method to prove discrepancy upper bounds

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

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

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

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

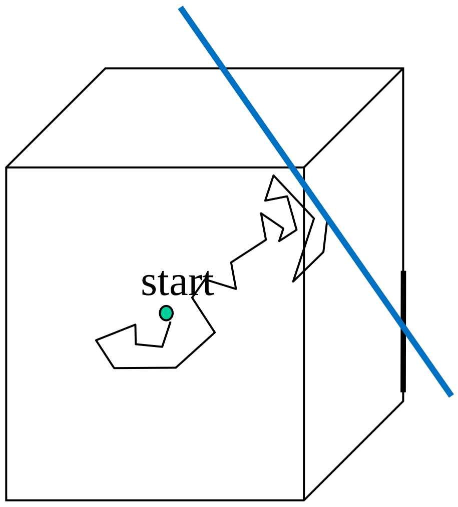
E.g. can ask for partial coloring with 0 discrepancy on n/log n rows, and reasonable amount on others.

Lovett Meka Algorithm

Do a sticky random walk.

If some row A_i gets tight (disc(A_i) $= \lambda_i |a_i|_2$)

Move in space A_i x = 0



Make progress as long as dimension = $\Omega(n)$

$$\sum_i \exp(-\lambda_i^2) \leq \frac{n}{2}$$

(better than entropy method)

Guarantees a partial coloring, even if n/4 λ_i's are 0.

Comparison with iterated rounding

Fact: If n variables, but $\leq n/2$ constraints $Ax = b$ (rest: $0 \leq x \leq 1$)
There exists a basic feasible solution with $> n/2$ variables 0-1.

Iterated rounding: LP with m constraints, drop except n/2.
(no control on dropped constrains (a_i x $\leq b_i$), error up to $|a_i|_1$)

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

E.g. Can set n/10 constraints to have 0 error, and controlled bounds on others

Lower Bounds

Discrepancy

If disc(A) > D

$|Ax|_\infty \geq D$ 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 σ_{min}(PA)   P:diag, tr(P) = n

Determinant Lower Bound

Thm (Lovasz Spencer Vesztergombi'86): $\mathrm{herdisc}(A) \geq$ detlb(A)

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 ≤ O(1) detlb

Remark: For TU Matrices, Herdisc(A) =1, detlb = 1

(every submatrix has det -1,0 or +1)

Hoffman's example

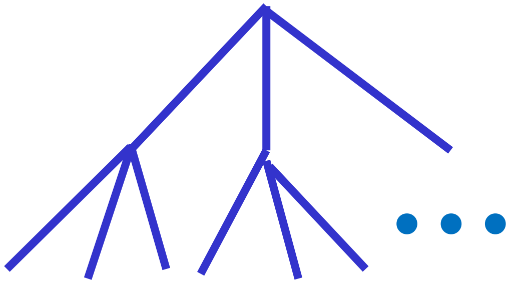
Hoffman: Detlb(A) ≤ 2 herdisc$(A) \geq \left(\frac{\log n}{\log\log n}\right)$

Palvolgyi'11: $\Omega(\log n)$ gap

T: k-ary tree of depth k. $\approx k^k$ nodes.

S: All edges out of a node.

S': All leaf to root paths.



Both S and S' are TU.

Claim: $Detlb(S \cup S')$ ≤ 2 (Expand determinant)

Herdisc($S \cup S'$) = k

Matousek’11: herdisc(A) ≤ $O(\log n \sqrt{\log m})$ detlb.

Idea: SDP Duality -> Dual Witness for large herdisc(A).

Dual Witness -> Submatrix with large determinant.

Other implications:

$$\text{herdisc}(A_1 \cup \cdots \cup A_t) \le O\left(\log n \sqrt{\log m}\right)\sqrt{t}\ \max_i \text{herdisc}(A_i)$$

Matousek's result

Thm: Herdisc(A) = O(log n $\sqrt{\log m}$) detlb(A)

Pf: Recall, Disc$(A) = O\left(\sqrt{\log m \log n}\right)$ Hervecdisc(A)

Some S, s.t. Vecdisc $\left(A_{|S}\right)$ = Herdisc(A)/ $O\left(\sqrt{\log m \log n}\right)$

Will show: vecdisc$\left(A_{|S}\right) \leq O\left(\sqrt{\log \mathrm{n}}\right)$ detlb$\left(A_{|S}\right)$

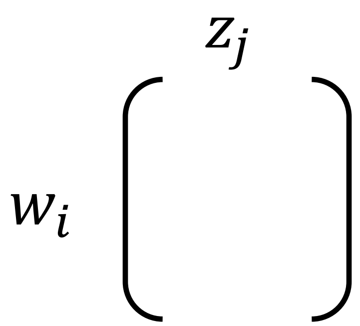
Let us use A for $A_{|S}$

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 $\sum_j z_j \geq D^2$



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

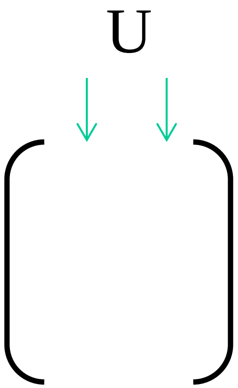
Why is this a witness?

Proof Sketch

As $\sum_j z_j \geq D^2$

There exists a subset of variables U such that

$$z_j \in \left[\frac{D^2}{4|U|\log n}, \frac{D^2}{2|U|\log n}\right]$$



$$\sum_i w_i \left(\sum_{j \in U} a_{ij} x_j\right)^2 \geq \frac{D^2}{4|U|\log n} \sum_{j \in U} x_j^2 \quad \text{for all } x \in R^n.$$

$W^{1/2} A_{|U}$ has large $\sigma_{\min}$

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

Concluding Remarks

SDP view has been extremely useful in discrepancy

Various open problems in discrepancy

- Beck Fiala Conjecture
- Discrepancy of points and rectangles
- Constructive version of Banasczcyk's theorem?
- …

Tightness of detlb bound (log vs $\log^{3/2}$)

O(1) approximation for herdisc

Thanks!