Discrepancy and Optimization

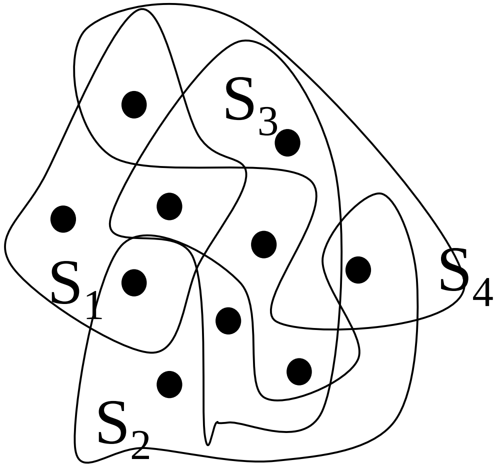
Nikhil Bansal

IPCO Summer School (lecture 2)

www.win.tue.nl/~nikhil/ipco-slides.pdf

(notes coming)

Discrepancy



Universe: U= [1,…,n]
Subsets: $S_1,S_2,\ldots,S_m$

Color elements red/blue so each
set is colored as evenly as possible.

Given χ: $[n] \rightarrow \{-1,+1\}$
Disc $(\chi) = \max_S |\sum_{i\in S} \chi(i)| = \max_S |\chi(S)|$

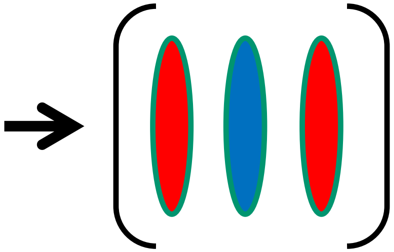
Disc (set system) $= \min_\chi \max_S |\chi(S)|$

Matrix Notation

$$\text{Incidence matrix } A = \begin{pmatrix} 1 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix}$$

Rows: sets
Columns: elements

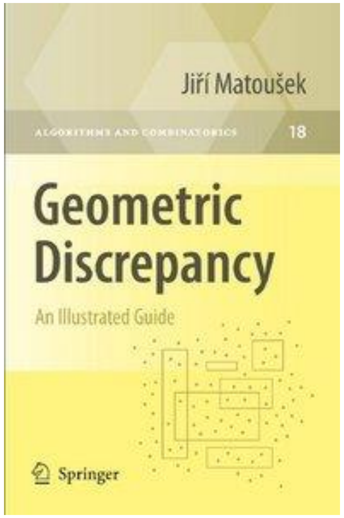
Given any matrix A,
find coloring $x \in \{-1,1\}^n$, to minimize $|Ax|_\infty$



Applications

CS: Computational Geometry, Approximation, Complexity, Differential Privacy, Pseudo-Randomness, …

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



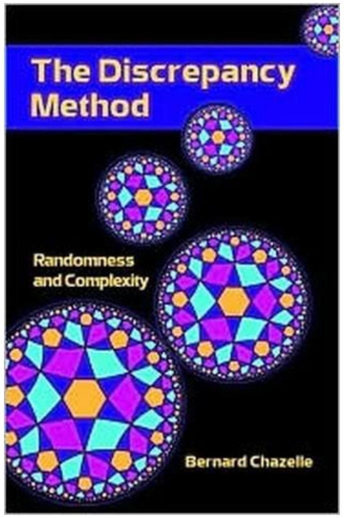
Jiří Matoušek

ALGORITHMS AND COMBINATORICS 18

Geometric Discrepancy

An Illustrated Guide

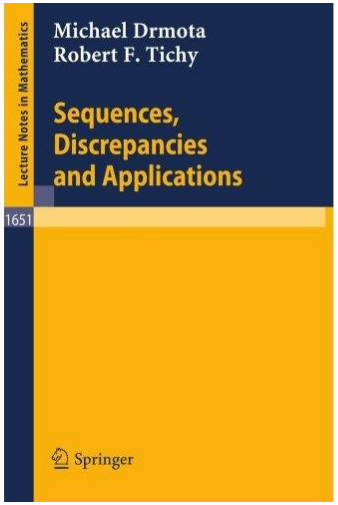
Springer



The Discrepancy Method

Randomness and Complexity

Bernard Chazelle



Lecture Notes in Mathematics

1651

Michael Drmota
Robert F. Tichy

Sequences, Discrepancies and Applications

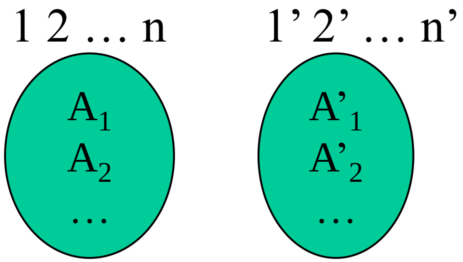
Springer

. . . . .

Hereditary Discrepancy

Discrepancy a useful measure of complexity of a set system

But not so robust



1 2 … n

A_1
A_2
…

1’ 2’ … n’

A'_1
A'_2
…

$S_i = A_i \cup A'_i$

Discrepancy = 0

Hereditary discrepancy:

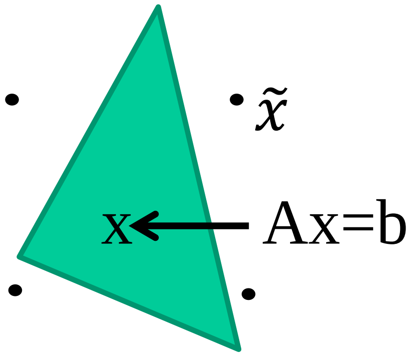
herdisc (U,S) = $\max_{U' \subseteq U}$ disc (U’, $S_{|U'}$)

Robust version of discrepancy
(99% of problems: bounding disc = bounding herdisc)

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)



Intuition: Discrepancy is like rounding ½ integral solution to 0 or 1.

Can do dependent (correlated) rounding based on A.

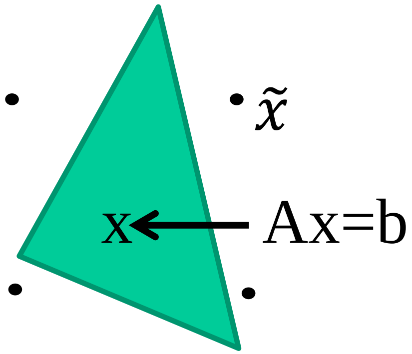
For approximation algorithms: need algorithms for discrepancy

Bin packing: OPT + $\tilde{O}$(log OPT) [Rothvoss'13]

Herdisc(A) = 1 iff A is TU matrix.

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



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

Only shows existence of good rounding

How to actually find it?

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

Ordering with small prefix sums

Vectors $v_1, \ldots, v_n \in R^d$ $|v|_\infty \leq 1$ $\sum_i v_i = 0$

Find a permutation π such that each prefix sum has small norm

i.e. $Max_k \left| v_{\pi(1)} + \ldots + v_{\pi(k)} \right|_\infty$ is minimized

d=1 numbers in [-1,1] e.g. 0.7 -0.2 -0.9 0.8, 0.7 …

What would a random ordering give?

d=2 $\begin{bmatrix} 0.7 \\ -0.4 \end{bmatrix}, \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}, \begin{bmatrix} -0.8 \\ 0.5 \end{bmatrix}, \ldots$ can we get $O(1)$

(Posed by Reimann, solved by Steinitz in 1913, called Steinitz problem)

Steinitz Problem

Given $v_1, \dots, v_n \in R^d$ with $\sum_i v_i = \mathbf{0}$

Find permutation to minimize norm of prefix sums

$$m(\pi) = \max_k \left| v_{\pi(1)} + \ \dots + \ v_{\pi(k)} \right|$$

Discrepancy of prefix sums: Given ordering find signs to minimize norm of signed prefix sums

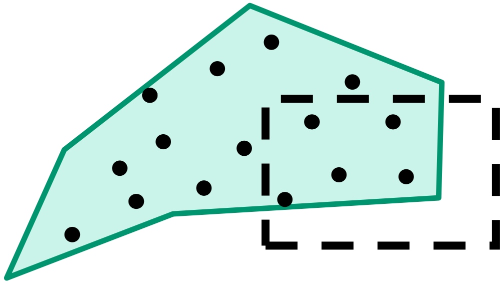
$\pi \quad v_1 v_2 v_3 v_4\ v_5\ v_6\ v_7\ v_8$ $\quad \Longrightarrow \quad$ $v_1 v_3 v_4 v_8 v_7 v_6 v_5 v_2$

\+ - + + - - - +

$$m(\pi) \qquad\qquad \frac{m(\pi) + f(d)}{2}$$

Sparsification

Original motivation: Numerical Integration/ Sampling
How well can you approximate a region by discrete points ?



Discrepancy:
Max over rectangles R
|(# points in R) – (Area of R)|

Use this to sparsify
Quasi-Monte Carlo integration: Huge area (finance, …)

Error MC $\approx \frac{1}{\sqrt{n}}$

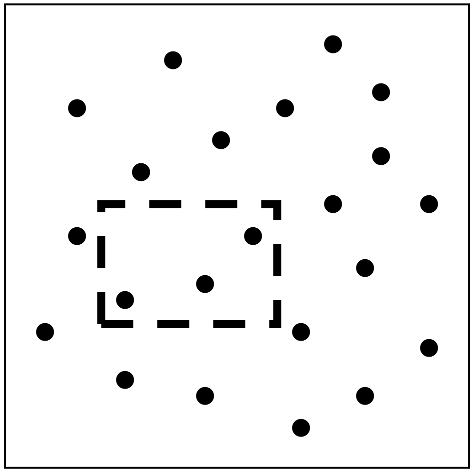
QMC $\approx \frac{disc}{n}$

Tusnady's problem

Input: n points placed arbitrarily in a grid.

Sets = axis-parallel rectangles

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



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

Very long line of work

$O(\log^4 n)$ [Beck 80's]

...

$O(\log^{2.5} n)$ [Matousek'99]

$O(\log^2 n)$ [B., Garg'16]

$O(\log^{1.5} n)$ [Nikolov'17]

Questions around Discrepancy bounds

Combinatorial: Show good coloring exists

Algorithmic: Find coloring in poly time

Lower bounds on discrepancy

Approximating discrepancy

Combinatorial (3 generations)

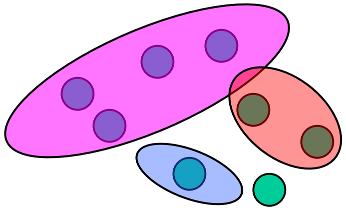
0) Linear Algebra (Iterated Rounding)

[Steinitz, Beck-Fiala, Barany, …]

1) Partial Coloring Method:

Beck/Spencer early 80's: Probabilistic Method + Pigeonhole

Gluskin'87: Convex Geometric Approach



Very versatile (black-box)

Loss adds over O(log n) iterations

2) Banaszczyk'98: Based on a deep convex geometric result

Produces full coloring directly (also black-box)

Brief History (combinatorial)

Method	**Tusnady (rectangles)**	**Steinitz (prefix sums)**	**Beck-Fiala (low deg. system)**
Linear Algebra	$\log^4 n$	d	k
Partial Coloring	$\log^{2.5} n$ [Matousek'99]	$d^{1/2} \log n$	$k^{1/2} \log n$
Banaszczyk	$\log^{1.5} n$ [Nikolov'17]	$(d \log n)^{1/2}$ [Banaszczyk'12]	$(k \log n)^{1/2}$ [Banaszczyk'98]
Lower bound	$\log \quad n$	$d^{1/2}$	$k^{1/2}$

Brief History (algorithmic)

Partial Coloring now constructive

Bansal'10: SDP + Random walk

Lovett Meka'12: Random walk + linear algebra

Rothvoss'14: Sample and Project (geometric)

Many others by now [Harvey, Schwartz, Singh], [Eldan, Singh]

Method	**Tusnady (rectangles)**	**Steinitz (prefix sums)**	**Beck-Fiala (low deg. system)**
Linear Algebra	$\log^4 n$	d	k
Partial Coloring	$\log^{2.5} n$ [Matousek'99]	$d^{1/2} \log n$	$k^{1/2} \log n$
Banaszczyk	$\log^{1.5} n$ [Nikolov'17]	$(d \log n)^{1/2}$ [Banaszczyk'12]	$(k \log n)^{1/2}$ [Banaszczyk'98]

Algorithmic aspects (2)

Beck-Fiala (B.-Dadush-Garg'16) (tailor made algorithm)

General Banaszczyk (B.-Dadush-Garg-Lovett'18)

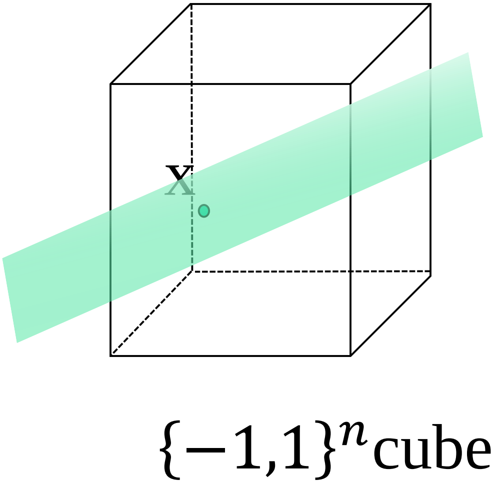
Method	**Tusnady (rectangles)**	**Steinitz (prefix sums)**	**Beck-Fiala (low deg. system)**
Linear Algebra	$\log^4 n$	d	K
Partial Coloring	$\log^{2.5} n$ [Matousek'99]	$d^{1/2} \log n$	$k^{1/2} \log n$
Banaszczyk	$\log^{1.5} n$ $\log^2 n$ [Nikolov'17] [BDG16]	$(d \log n)^{1/2}$ [BDGL] [Banaszczyk'12]	$(k \log n)^{1/2}$ [BDG'16] [Banaszczyk'98]
Lower bound	$\log \quad n$	$d^{1/2}$	$k^{1/2}$

Linear Algebraic approach

Start with x(0) = (0,…,0) coloring.

Update at each step t

If a variable reaches -1 or 1, fixed forever.



x(t) = x(t-1) + y(t)

Update y(t) obtained by solving By(t) = 0

B cleverly chosen.

Beck-Fiala: B = rows with size > k (on floating variables)

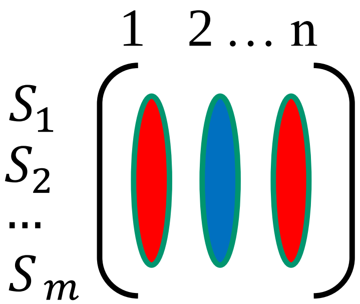
Row has 0 discrepancy as long as it is big.

(no control once it becomes of size <= k).

Partial Coloring

Spencer’s problem

Spencer Setting: Discrepancy of any set system on n elements and m sets?



[Spencer’85]: (independently by Gluskin’87)

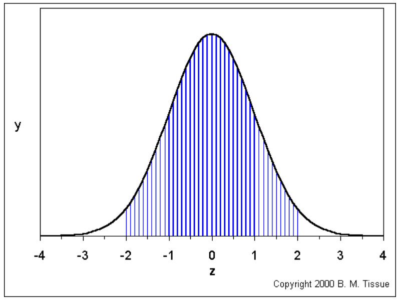
For m = n discrepancy $\leq 6n^{1/2}$

Tight: Cannot beat $0.5\, n^{1/2}$ (Hadamard Matrix).

Random coloring gives $O(n \log n)^{1/2}$

Proof: For set S, Pr [disc(S) $\approx c|S|^{1/2}$] $\approx \exp(-c^2)$

Set c = $O(\log n)^{1/2}$ and apply union bound.



Tight. Random gives $\Omega(n \log n)^{1/2}$ with very high prob.

Beating random coloring

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

$$g(\lambda_i) \approx \ln\left(\frac{1}{\lambda_i}\right) \quad \text{if } \lambda_i < 1$$
$$\approx e^{-\lambda_i^2} \quad \text{if } \lambda_i \geq 1$$

Union bound: $\sum_i e^{-\lambda_i^2} < 1$

n/5 vs 1 very powerful

Can demand discrepancy 0 for $\approx \Omega(n)$ rows.

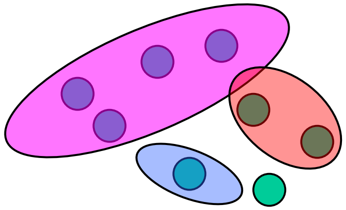
(while still having control on other rows).

Combines strengths of probability + linear algebra

Spencer's O($n^{1/2}$) result

Partial Coloring suffices: For any set system with m sets, there exists a coloring on $\geq$ n/2 elements with discrepancy $\Delta = O(n^{1/2} \log^{1/2} (2m/n))$ [For m=n, disc = O($n^{1/2}$)]

Algorithm for total coloring:



Repeatedly apply partial coloring lemma

Total discrepancy

$O(n^{1/2} \log^{1/2} 2)$ [Phase 1]

$+ O((n/2)^{1/2} \log^{1/2} 4)$ [Phase 2]

$+ O((n/4)^{1/2} \log^{1/2} 8)$ [Phase 3]

$+ \ldots = O(n^{1/2})$

Beck Fiala

Thm: Partial coloring $O\left(k^{1/2}\right)$, so Full coloring $O\left(k^{1/2}\log n\right)$

Total number of 1's in matrix $\leq nk$

Why can we set $\Delta = k^{1/2}$?

$\sum_i g(\lambda_i) \leq \frac{n}{5}$ $\lambda_i = \frac{\Delta}{\sqrt{|S_i|}}$ $g(\lambda_i) \approx \ln\left(\frac{1}{\lambda_i}\right)$ if $\lambda_i < 1/2$

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

n sets of size k	n g(1)	$\approx n$
n/t sets of size tk	$\frac{n}{t} g\left(\frac{1}{t^{\frac{1}{2}}}\right)$	$\approx (n/t)\log t$
tn sets of size k/t	$tn\, g\left(t^{1/2}\right)$	$\approx tn\, e^{-t}$

Proving Partial Coloring Lemma

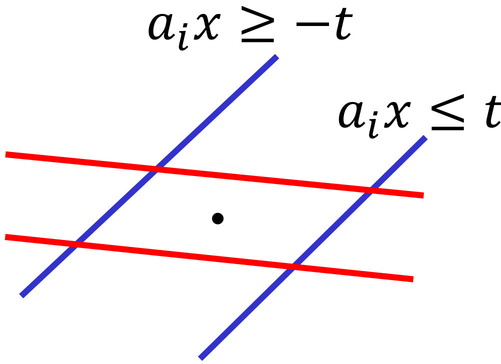
A geometric view

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

Gluskin'87: Convex geometric approach

Consider polytope P(t) $= -t\,\mathbf{1} \le Ax \le t\,\mathbf{1}$

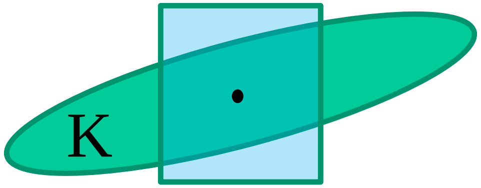
P(t) contains a point in $\{-1,1\}^n$ for t = $6\sqrt{n}$



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

d-dim Gaussian Measure: $\gamma_d(x) = \exp\left(-|x|^2/2\right)(2\pi)^{-d/2}$

$\gamma_d(K)$: Pr[$(y_1, \dots, y_m) \in K$] each y_i iid N(0,1)

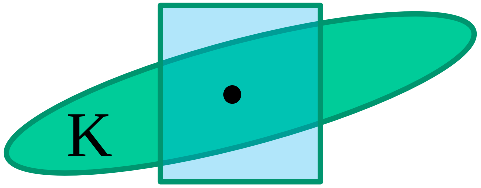


$[-1,1]^n$ cube

What is the Gaussian volume of $[-1,1]^n$ cube

A geometric view

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



Proof: Look at K+x for all $x \in \{-1,1\}^n$

Total volume of shifts = $2^{\Omega(n)}$

$$\gamma_n(K+x) \geq \gamma_n(K)\exp\left(-|x|^2/2\right)$$

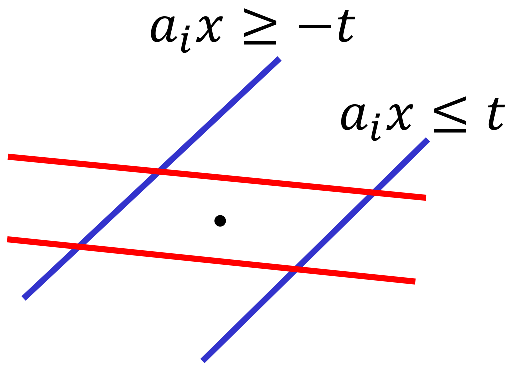
Some point z lies in $2^{\Omega(n)}$ copies

$z = k + x$ and $z = k' + x'$ where x, x' have large hamming distance

Gives $(x - x')/2 = (k - k')/2 \in K$.

Gluskin for Polytopes

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



Consider polytope P $= \{\, |a_i x| \leq \Delta_i,\ i \in [m]\}$

For what Δ_i Gaussian volume large enough?

Sidak's Thm: $\gamma_n(K \cap Slab) \geq \gamma_n(K)\gamma_n(Slab)$

$\gamma_n(P) \geq \Pi_i\, \gamma_n(Slab_i)$ $\qquad Slab_i = |a_i x| \leq t$

Gaussian correlation Thm (Royen'14): Any convex symmetric K, S

$\gamma_n(K \cap S) \geq \gamma_n(K)\gamma_n(S)$

Volume of a slab

Sidak's Thm: $\gamma_n(P) \geq \Pi_i \gamma_n(Slab_i)$ $\quad$ $Slab_i = |a_i x| \leq t$

Useful to normalize $t = \lambda |a_i|_2$

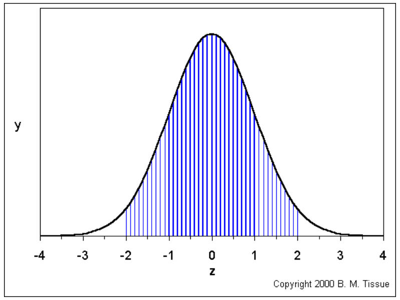
Lemma: $\gamma_n(Slab) =$ exp(-g(λ))

Proof: Can assume $a_i = |a_i| e_1$ (rotational invariance of Gaussian)

$$\Pr[\, |a_i x| \leq \lambda |a_i|_2] = \Pr[\, g_1 \leq \lambda] = 1 - \exp\left(-\lambda^2\right) \qquad \lambda \geq 1$$

$$\approx \lambda \qquad \lambda < 1$$

Sidak's Lemma, $\gamma_n(P) \geq 2^{-n/100}$ gives the result.



Algorithmic Partial Coloring

Useful View

Independent rounding.

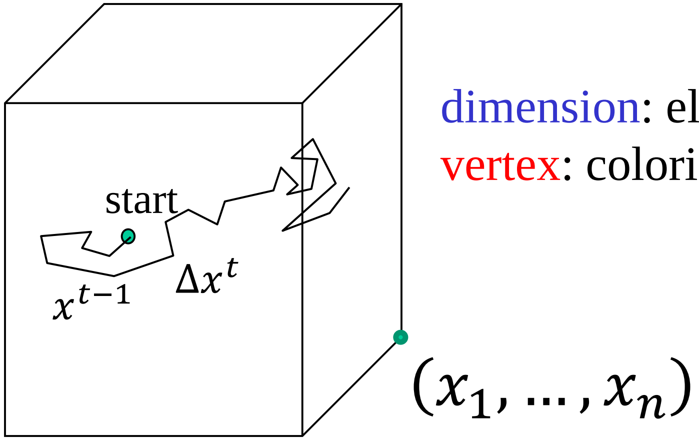
A (complicated) view

Brownian motion in cube.

Same as random coloring

Each coordinate independent

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



dimension: element

vertex: coloring

Useful View

If additional constraints.

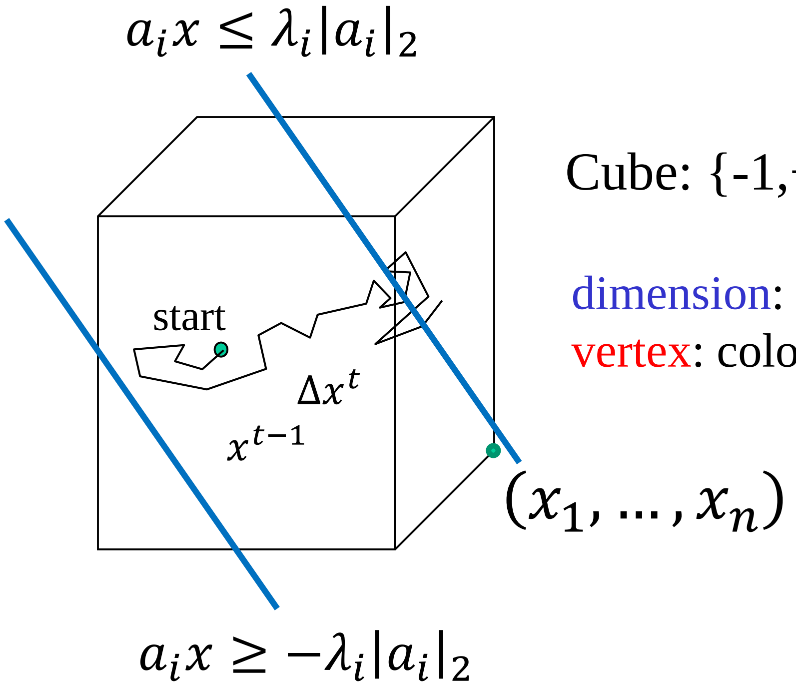
Can tailor walk accordingly.

Pick covariance matrix for Δx^t

(slow down towards bad regions)

Design barrier functions

…



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

dimension: element

vertex: coloring

Lovett Meka Algorithm

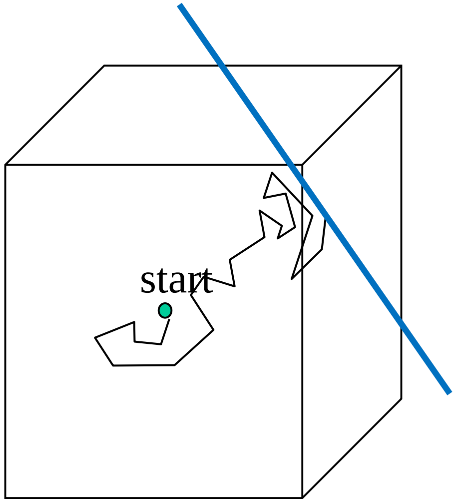
Random walk, γ N(0,1) in each dimension

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

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

Move in subspace $a_i x = \lambda_i |a_i|_2$

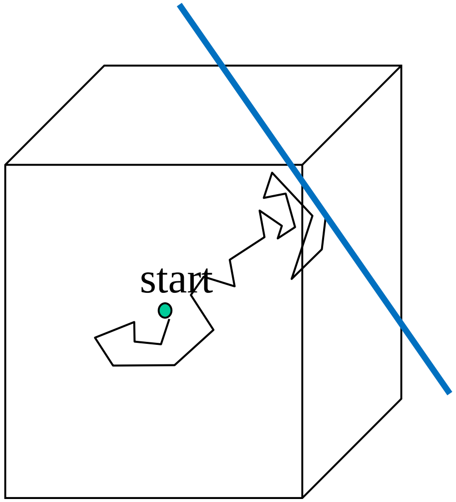
(not violate discrepancy)



Thm [LM'12] : Given an m x n matrix A, can a partial coloring $x \in [-1,1]^n$ with $\Omega(n)$ of them ± 1

$|a_i x| \leq \lambda_i \, |a_i|_2$ for each row i, provided $\sum_i e^{-\lambda_i^2} \leq \frac{n}{5}$

Lovett Meka Algorithm



Random walk, γ N(0,1) in each dimension

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

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

Move in subspace a_ix = $\lambda_i |a_i|_2$

(not violate discrepancy)

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

After $\frac{10}{\gamma^2}$ steps: $\Omega(n)$ variables must have hit ± 1

Pr[Row a_i tight] $\approx \exp\left(-\lambda_i^2\right)$

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

Another Algorithm

(general convex bodies, not just polytopes)

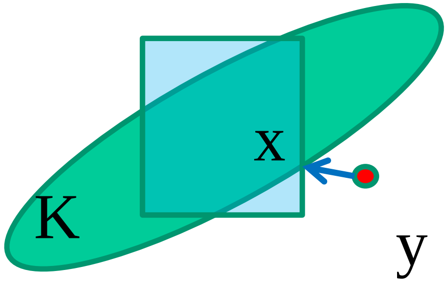
Algorithmic version

Rothvoss'14: Pick a random y, return closest point x in K$\cap$ $[-1,1]^n$

Idea: Measure concentration

If $\gamma_n(K) \geq$ ½

$\gamma_n(K + tB_2) \geq 1 - e^{-t^2/2}$ (halfspace)



$\gamma_n(K) \geq 2^{-\epsilon n}$ dist(y, K) $\approx (\epsilon n)^{1/2}$

dist(y, Cube) $\approx \sqrt{n}$

So dist(y, K$\cap$ $[-1,1]^n$) $\geq \sqrt{n}$

Suppose x has only δn coordinates ± 1.

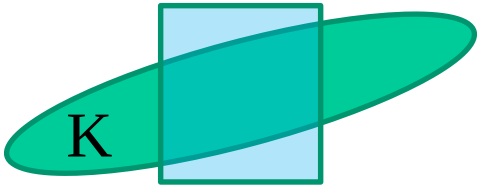
Would get same x if body $K' = K \cap \delta n$ slabs

But by Sidak $\gamma_n(K') \approx 2^{-(\epsilon+\delta)n}$

so $dist(y, K') \approx ((\epsilon + \delta)\, n)^{1/2}$ (gives contradiction)

Partial Coloring

Eldan, Singh'14: Pick a random direction c; optimize max $c \cdot x$ over K$\cap$ $[-1,1]^n$



Approximating Discrepancy

Vector Discrepancy

Exact: Min t

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

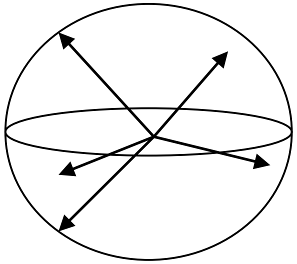
$x_j \in \{-1,1\}$ for each j

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: Whether disc(A) = 0 or $\Omega(\sqrt{n})$ for Spencer's setting?

Also implies vecdisc not a good relaxation.

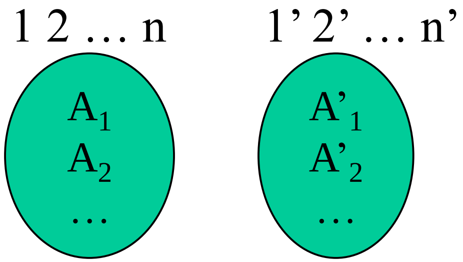
There must exist set systems where disc(A) = $\Omega(\sqrt{n})$

(but any polynomial time computable function returns 0)

Still SDP can be useful

Discrepancy a useful measure of complexity of a set system

But not so robust



1 2 … n

A_1
A_2
…

1' 2' … n'

A'_1
A'_2
…

$S_i = A_i \cup A'_i$

Discrepancy = 0

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

Hervecdisc(A) $\leq$ herdisc(A)

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

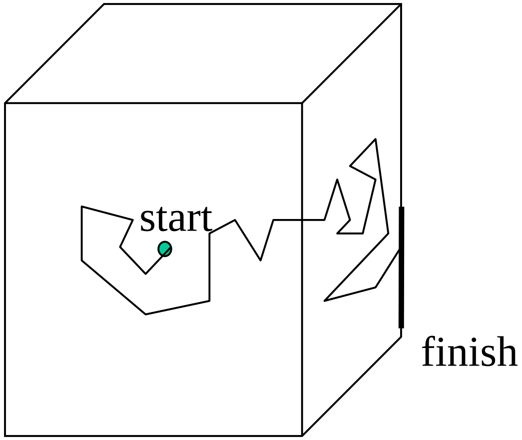
Rounding Application

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)

Gives algorithmic $|Ax - A\tilde{x}|_\infty < O\left(\sqrt{\log m \log n}\right)$ Herdisc(A)

Algorithm (at high level)

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



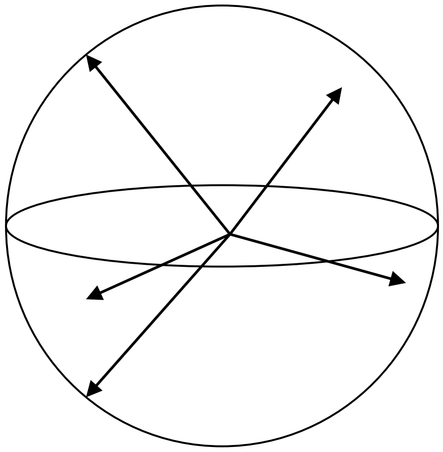
Each dimension: An Element
Each vertex: A Coloring

Algorithm: "Sticky" random walk
Each step generated by rounding a suitable SDP
Move in various dimensions correlated, e.g. $\delta^t_1 + \delta^t_2 \approx 0$

Analysis: Few steps to reach a vertex (walk has high variance)
$\mathrm{Disc}(S_i)$ does a random walk (with low variance)

An SDP

Hereditary disc. $\lambda \Rightarrow$ the following SDP is always feasible



Obtain $v_i \in R^n$

SDP:

Low discrepancy: $|\sum_{i \in S_j} v_i|^2 \leq \lambda^2$

$|v_i|^2 = 1$

Rounding:

Pick random Gaussian $g = (g_1, g_2, \ldots, g_n)$

each coordinate g_i is iid N(0,1)

For each i, consider $\eta_i = g \cdot v_i$

Properties of Rounding

Lemma: If $g \in R^n$ is random Gaussian. For any $v \in R^n$,

$g \cdot v$ is distributed as $N(0, |v|^2)$

Pf: $N(0,a^2) + N(0,b^2) = N(0,a^2+b^2)$ $g\cdot v = \sum_i v(i)\, g_i \sim N(0, \sum_i v(i)^2)$

Recall: $\eta_i = g \cdot v_i$

SDP:

$|v_i|^2 = 1$

$|\sum_{i\in S} v_i|^2 \leq \lambda^2$

1. Each $\eta_i \sim N(0,1)$
2. For each set S,

   $\sum_{i \in S} \eta_i = g \cdot (\sum_{i\in S} v_i) \sim N(0, \leq \lambda^2)$

   (std deviation $\leq\lambda$)

η's mimics a low discrepancy coloring (but is not {-1,+1})

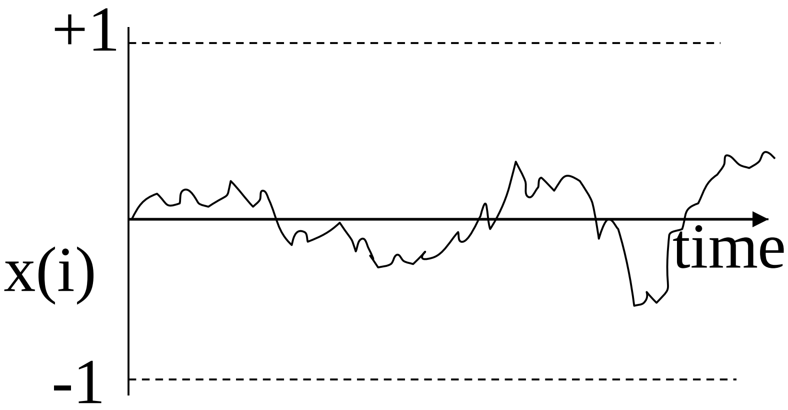
Algorithm Overview

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: 1/n suffices)



$x_t(i) = \gamma (\eta^1_i + \eta^2_i + \ldots + \eta^t_i)$

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

Fixed if reaches -1 or +1.

Set S: $x_t(S) = \sum_{i \in S} x_t(i)$ does a random walk w/ step $\gamma N(0,\leq \lambda^2)$

Analysis

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

Claim 1: With prob. ½, at least n/2 variables reach -1 or +1.
Pf: Each element doing random walk with size $\approx \gamma$.

$\Rightarrow$ Everything colored in O(log n) rounds.

Claim 2: Each set has $O(\lambda)$ discrepancy in expectation per round.
Pf: For each S, $x_t(S)$ doing random walk with step size $\approx \gamma\, \lambda$

Log n rounds + Union bounds over m sets gives
$O(\lambda(\log n \log m)^{1/2})$ bound

Recap

At each step of walk, formulate SDP on unfixed variables.

SDP is feasible

Gaussian Rounding -> Step of walk

Properties of walk:

High Variance -> Quick convergence

Low variance for discrepancy on sets -> Low discrepancy

Approximating Herdisc

CNN’11: Discrepancy was hard to approximate (not very robust)

Can we approximate herdisc(A)

(not even clear if in NP, do to check if herdisc(A) $\leq t$)

$\text{Hervecdisc}(A) \ \leq \text{herdisc}(A) \ \leq O((\log n \log m)^{1/2}) \ \text{Hervecdisc}(A)$

For any restriction $A_{|S}$, can find coloring of S

With discrepancy $O((\log n \log m)^{1/2})$ hervecdisc(A)

But: Not clear how to compute hervecdisc(A) efficiently.

Matousek Lower Bound

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

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

Conjecture (LSV'86): Herdisc ≤ O(1) detlb

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

Detlb

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

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

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

Idea: Algorithm -> hervecdisc is within log of herdisc

SDP Duality -> Dual Witness for large hervecdisc(A).

Dual Witness -> Submatrix with large determinant.

For a matrix A, let r(A) = max row length (ℓ_2 $norm$)

c(A) = max column length

$\gamma_2(A)$ = min r(B) c(C) over all factorizations A= BC

Theorem: $\frac{1}{\log m}$ $\gamma_2(A)$ $\leq$ herdisc(A) $\leq$ $\gamma_2(A)$ $\sqrt{\log m}$

γ_2 is computable using an SDP (can assume r(B) = c(C))

$A_{ij} = w_i \cdot v_j$

$|w_i|_2 \leq t$, $|v_j|_2 \leq t$ for all $i \in [m], j \in [n]$

Beyond Partial Coloring

Annoying loss of $O(\log n)$ to get full coloring

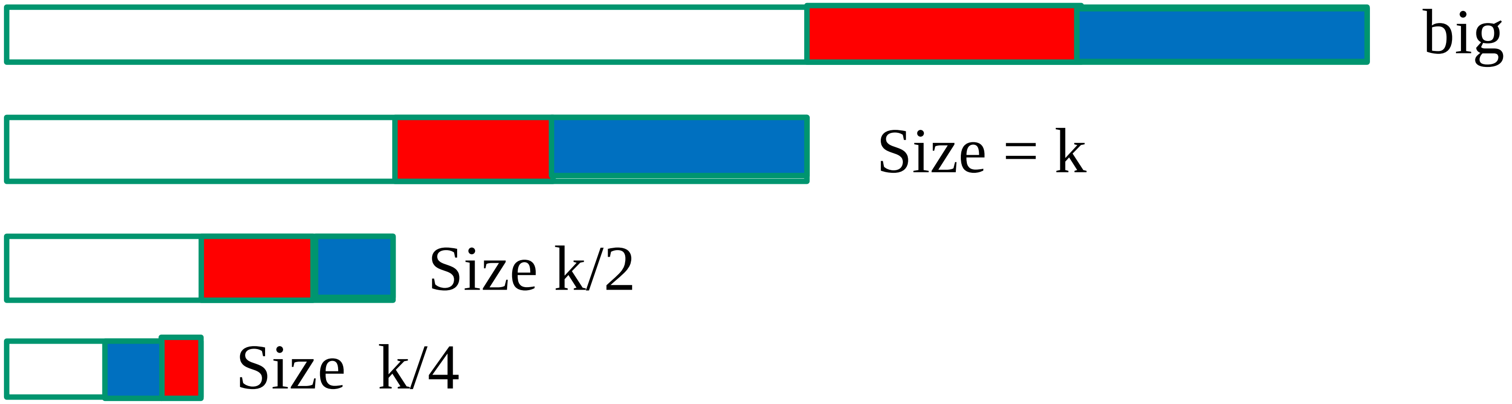
Ideal case

Beck-Fiala Setting: At most n/10 big (>10k) sets

Partial Coloring: 0 for big sets.

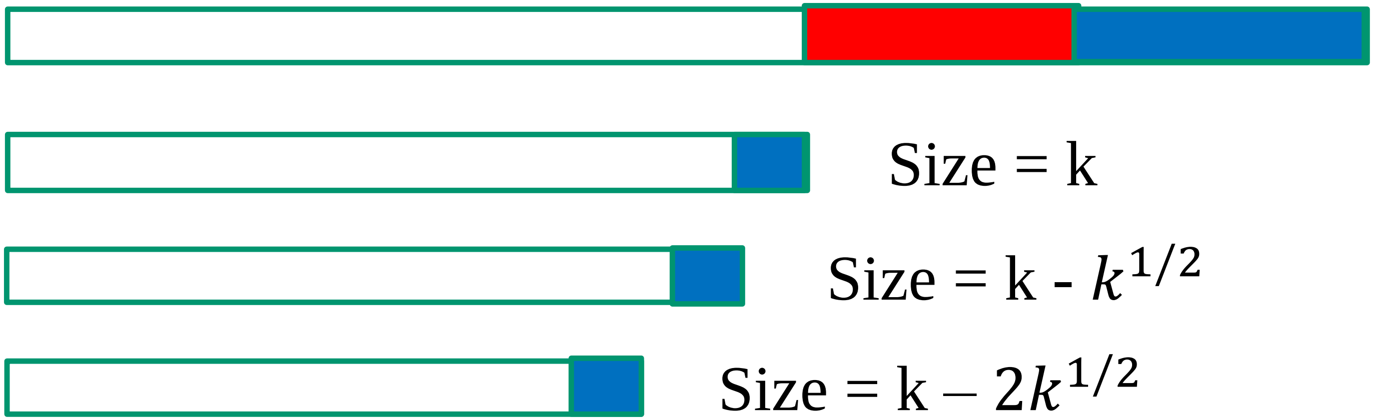
About $s^{1/2}$ for small sets of size s.

“Ideal” life cycle of a set



Ideal case: Discrepancy = $k^{1/2} + (k/2)^{1/2} + (k/4)^{1/2} + \ldots$

What can go wrong



Trouble: A set can get $k^{1/2}$discrepancy, but very few elements colored.

Banaszczyk’s full coloring method

Discrepancy

Given an $m \times n$ matrix A,
find $x \in \{-1,1\}^n$, to minimize
disc(A) = $|Ax|_\infty$

Incidence matrix $A = \begin{pmatrix} 1 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix}$

Rows: sets
Columns: elements

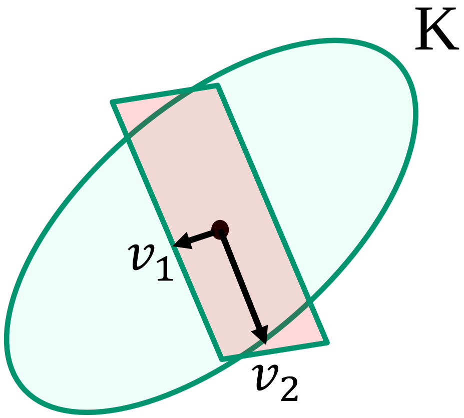
Vector balancing view: Given vectors $v_1, \ldots, v_n \in R^m$
find $x \in \{-1,1\}^n$ to minimize $\left|\sum_i x_i v_i\right|_\infty$

Banaszczyk's Theorem

Thm: Let A have columns $v_1, \ldots, v_n \in R^m$, $|v_i|_2 \leq 1/5$

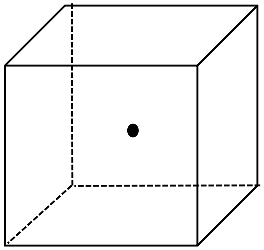
K = symmetric convex body with $\gamma_m(K) \geq \frac{1}{2}$

$\exists\, x \in \{-1,1\}^n$ s.t. $Ax \in K$



Banaszczyk's Theorem

Cube: K = $O(\log m)^{1/2}$ $[-1,1]^m$ $\gamma_{\mathrm{m}}(\mathrm{K}) \geq 1/2$



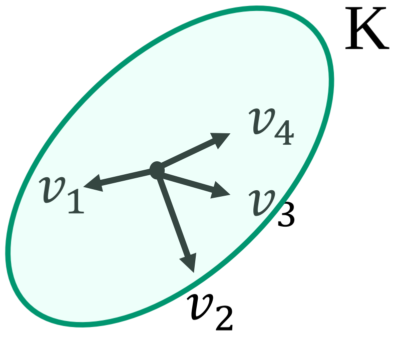
Gives $O(k \log n)^{1/2}$ for Beck-Fiala easily

Scale matrix by $\frac{1}{k^{1/2}}$ (length of columns ≤ 1)

$\exists$ signed sum w/ ℓ_∞-norm $O(\log m)^{1/2}$ (and $m \leq nt$)

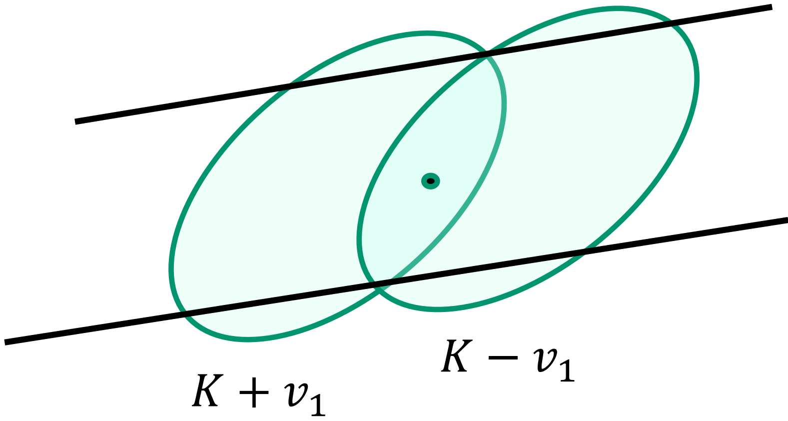
Surprising results for various bodies K.

Proof idea



Given $v_1, \dots, v_n$, each $|v_i| < 1/5$. $\gamma_m(K) \geq \frac{1}{2}$

Goal: Find signing $\sum_i x_i v_i \in K$



Key observation: Signing exists iff

Some signing of $v_2, \dots, v_n$ with sum in

$(K + v_1) \cup (K - v_1)$.

Convexify:

Remove regions of K width $< 2|v_1|$ along v_1

Lose and gain volume.

(non-trivial) computation to show volume stays $\geq$ ½

Algorithmic history

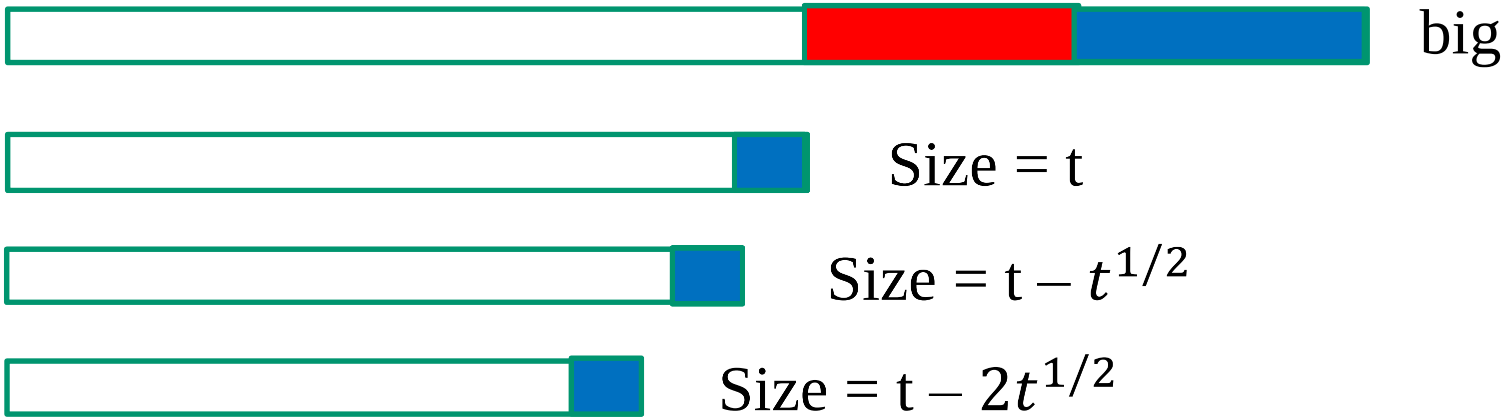
Banaszczyk based approaches:

[B., Dadush, Garg'16]: $O(\log n)^{1/2}$ algorithm for Komlos problem

[B., Dadush, Garg, Lovett 18]: algorithm for general Banaszczyk.

Recall trouble with Partial Coloring

Beck Fiala Setting



Trouble: A set can get $t^{1/2}$ discrepancy, but very few elements colored.

Lovett Meka Algorithm

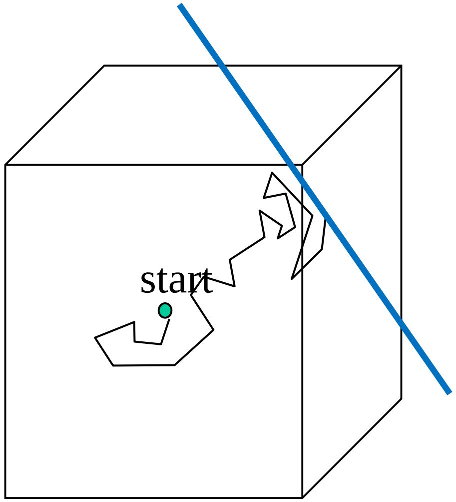
Random walk, γ N(0,1) in each dimension

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

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

Move in subspace $a_i x = \lambda_i |a_i|_2$

(not violate discrepancy)



Correlations in Lovett-Meka

Consider set S = {1,2,…,k}

Ideal case: If randomly color each element

Progress = k discrepancy $\approx k^{1/2}$

Suppose move in subspace $x_1 = x_2 = \cdots = x_k$

E.g. if have constraints x_1 - x_2 = 0, x_2 - x_3= 0, …

Can only color all +1 or all -1.

Progress = k discrepancy = k

In Lovett-Meka, such sets hit subspace at $k^{1/2}$ discrepancy, but progress is only $k^{1/2}$

Suggests a solution

Used for algorithmic $O\left(k^{1/2}\log^{1/2} n\right)$ bound for Beck-Fiala [B., Dadush, Garg'16]

Can we design a walk that moves in some subspace, but still looks quite "random"?

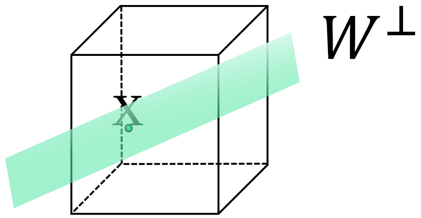
E.g. If constrained to move in subspace $x_1 = x_2 = \cdots = x_k$

Just set $\Delta x_i = 0$ for i=1,2,..,t

Can still do a random walk for i = k+1,..,n.

Smarter covariance matrices

W: arbitrary subspace dim(W) $\leq (1-\delta)n$
Need to walk in $W^{\perp}$



-1/+1 cube

Property 1: $w^T(\Delta x) = 0 \quad \forall w \in W$

$$E\left[\, w^T \Delta x \, \Delta x^T w\right] = 0 \quad \text{or} \quad w^T Y w \;=\; 0$$

Covariance matrix
$Y(i,j) = E\left[\Delta x_i, \Delta x_j\right]$

Property 2: Still looks almost independent.

For any direction $c = (c_1, \dots, c_n)$

$$E[\left(\textstyle\sum_i c_i \Delta x_i\right)^2] \leq \frac{1}{\delta} \textstyle\sum_i c_i^2 \, E\left[\, \Delta x_i^2\right]$$

$$c^T Y\, c \leq \left(\frac{1}{\delta}\right) c^T diag(Y)\, c \qquad \forall c \in R^n.$$

$$Y \preccurlyeq \left(\frac{1}{\delta}\right) diag(Y)$$

Can find such a good walk

Key Thm: If $\dim(W) \leq (1-\delta)n$

There is a non-zero solution Y to the SDP

$$w^T Y w \; = \; 0 \qquad \forall w \; \in W$$

$$Y \preccurlyeq \left(\frac{1}{\delta}\right) diag(Y)$$

$$Y \succcurlyeq 0$$

Proof: Using SDP duality

Use this to design the walk $\Delta x \; = Y^{1/2} g$

Getting Concentration

Thm: Upon termination the 0-1 solution satisfies concentration for every linear constraint

Fix $c = (c_1, \dots, c_n)$. Then cx evolves as a martingale

Key idea: Use sub-isotropic updates to control error during walk

Need “Freedman type” martingale analysis
must use intrinsic variance (avoid dependence on time steps).

Potential: $\sum_i c_i x_i - \lambda \sum_i c_i^2 \left(1 - x_i^2\right)$ evolves nicely.

Algorithm for Beck-Fiala

Time t: If n_t variables alive, at most $n_t/10$ big rows

Pick W = span of these constraints.

Run the SDP walk.

No phases, continue till all variables -1/+1 (i.e. $n_t = 0$).

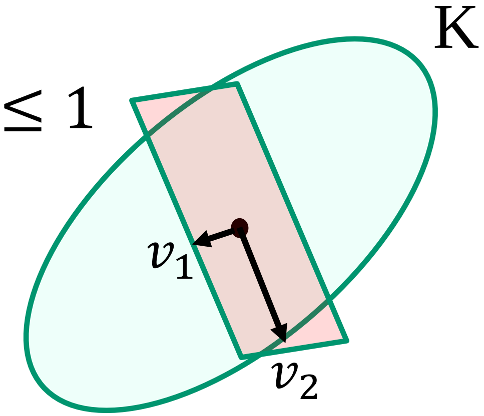
If row big = discrepancy 0

When becomes small, just like a random walk.

"Freedman type" martingale analysis (avoid dependence on time steps), gives the result.

General Banaszczyk

Making Banaszczyk Algorithmic



Thm [Banaszczyk 97]: Input $v_1, \dots, v_n \in R^d$, $|v_i|_2 \leq 1$

$\forall$ convex body K, with $\gamma_d(K) \geq \frac{1}{2}$

$\exists$ coloring $x \in \{-1,1\}^n$ s.t. $\sum_i x(i) v_i \in 5K$

Coloring depends on the convex body K.

How is K specified? (input size could be exponential)

Idea [Dadush, Garg, Lovett, Nikolov'16]: Minimax Thm. (2-player game)

Universal distribution on colorings that works for all convex bodies

Equivalent formulation

Alternate formulation [Dadush, Garg, Lovett, Nikolov'16]:

$\exists$ distribution on colorings $x \in \{-1,1\}^n$,

s.t. Y = $\sum_i x(i) v_i$ is $\approx \underbrace{N(0,1) \text{ in every direction}}_{O(1)\text{ subgaussian}}$

No body K
anymore

$Y \in R^d$ is σ-subgaussian if in all directions $\theta \in R^d, |\theta|_2 = 1,$

$\langle \theta, Y \rangle$ has same tails as $N\left(0, \sigma^2\right)$ i.e. $\Pr[\, |\langle \theta, Y \rangle| \geq \lambda] \leq 2 \exp\left(-\lambda^2/2\sigma^2\right)$

Lemma: $Y \in K$ (for K convex, $\gamma_d(K) \geq \frac{1}{2}$) with constant prob.

Suffices to sample x implicitly from such a distribution.

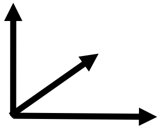
Goal: $\exists$ distribution on colorings $x \in \{-1,1\}^n$,

s.t. random vector $Y = \sum_i x(i) v_i$ is O(1) subgaussian

$\forall \theta \in S^{m-1}$, $\langle Y, \theta \rangle = \sum_i x(i) \langle v_i, \theta \rangle$ decays like N(0,1).

Special cases:

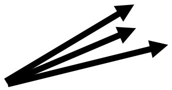
1) v_i are Orthogonal: Random ± coloring x_i works



As $\sum_i c_i x_i \approx N\left(0, \sum_i c_i^2\right)$

Var($\langle Y, \theta \rangle$) = $\sum_i \langle v_i, \theta \rangle^2 \le |\theta|^2 \le 1$

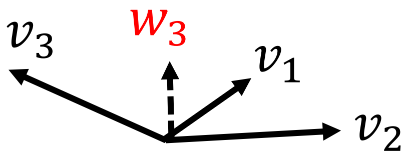
2) All equal vectors



$v_1 = \cdots = v_n = v$ random coloring bad: $\Omega\left(\sqrt{n}\right)$ in direction v

Need dependent coloring: n/2 +1's and n/2 -1's

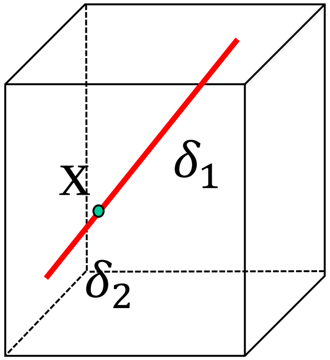
Gram Schmidt Walk



Algorithm: Consider vectors $v_1, \ldots, \boxed{v_n}$

Write $v_n = c_1 v_1 + \ldots c_{n-1} v_{n-1} + w_n$

where $w_n \in span\,(v_1, \ldots, v_{n-1})^{\perp}$



Let direction $c = (c_1, \ldots, c_{n-1}, -1)$

Update coloring x as δc s.t. $E[\delta] = 0$

i.e. $\Delta x = +\delta_1 c$ or $-\delta_2\, c$

Key Point: $\Delta Y = \sum_i \Delta x(i)\, v_i = \delta\left(\sum_{i=1}^{n-1} c_i v_i - v_n\right) = -\delta w_n.$

As $\delta \leq 2$ and $E[\delta] = 0$

$\Delta\langle Y, \theta\rangle$ evolves as a martingale with variance $O(\langle \theta, w_n\rangle^2)$

Proof Idea (ideal case)

$v_1, \ldots, \bigcirc\!\!\!\!\!\!v_n$

Suppose pivot is the one to freeze every time

Pivot v_n ΔY: $\delta_n w_n$

Pivot v_{n-1} ΔY: $\delta_{n-1}\, w_{n-1}$

….

$w_1, \ldots, w_n$ obtained by Gram Schmidt process.

$w_1 = v_1$ $\hat{w}_1 = w_1/|w_1|$

$w_2 = v_2 - \langle v_2, \hat{w}_1 \rangle\, \hat{w}_1$ $\hat{w}_2 = w_2/|w_2|$

$w_3 = v_3 - \langle v_3, \hat{w}_1 \rangle\, \hat{w}_1 - \langle v_3, \hat{w}_2 \rangle\, \hat{w}_2$ $\hat{w}_3 = w_3/|w_3|$

$$Y = \delta_n w_n + \delta_{n-1} w_{n-1} + \cdots + \delta_1\, w_1$$

$$Var\,(\langle Y, \theta \rangle) = \sum_i \delta_i^2 \langle w_i, \theta \rangle^2 \le \sum_i \delta_i^2\, \langle\, \hat{w}_i, \theta\, \rangle^2 \le 4|\theta|^2 = 4$$

Some more details

$v_1, \ldots, \xcancel{v_5}, \ldots, \enclose{circle}{v_n}$ No reason why pivot should get fixed.

Suppose v_5 gets fixed.

w_n becomes w_n' which can be longer.

Proof idea: Can charge increase in $|w_n|^2$ to v_5 disappearing.

Track evolution of $E\left[e^{\lambda\langle\theta,Y\rangle}\right]$ by a suitable potential

and show $E\left[e^{\lambda\langle\theta,Y\rangle}\right] = e^{O(\lambda^2)}$ for each θ, λ

(Recall Z is σ-subgaussian iff $E\left[e^{\lambda Z}\right] = e^{O(\lambda^2\sigma^2)}$ for all λ)

Thanks for your attention!