Combinatorial Discrepancy: A Tutorial

Raghu Meka

Abstract

Notes taken by Karel Král, Martin Balko and Vojta Tůma at REU 2013.

Discrepancy theory grew out of the study of irregularities of distributions and number sequences, or, in layman's terms, trying to be an egalitarian economist. From its number theoretic origins in early nineteenth century, discrepancy theory has developed into a rich and beautiful field over the last three decades. Besides being of importance in combinatorics, discrepancy theory has many applications in computational geometry, communication complexity, algorithm design, image processing and more.

In the first session of this two-part tutorial, we cover several classical results in combinatorial discrepancy theory, focusing on Beck's Partial Coloring Method. Along the way, we illustrate the use of linear algebraic and probabilistic methods in combinatorics. In the second session, we cover some recent results with special emphasis on algorithmic aspects and applications, which in turn shed new light and suggest new questions pertaining to existential ones.

Let our universe be $[n] = \{1, 2, 3, \dots, n\}$ and let $\mathcal{S} = \{S_1, S_2, S_3, \dots, S_m\}$ be a set system over $[n]$. Let $\chi\colon [n] \to \{-1, 1\}$ be a coloring of elements of $[n]$. We define the *combinatorial discrepancy* of the set system $\mathcal{S}$ as

$$\operatorname{disc}(\mathcal{S}) = \min_{\chi \in \{-1,1\}^n} \max_{S \in \mathcal{S}} \left| \sum_{i \in S} \chi_i \right| .$$

That is, we wish to color the elements of $[n]$ such that in any $S \in \mathcal{S}$ no color outnumbers the other one too much. The set system $\mathcal{S}$ can be described using a *characteristic matrix* $M = (M_{i,j})$, $i \in [m]$ and $j \in [n]$, where $M_{i,j} = 1$ if $j \in S_i$ and $M_{i,j} = 0$ otherwise. See Figure 1.

	1	2	...	n
S_1	0	1	...	0
S_2	0	0	...	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S_m	1	0	...	1

Figure 1: An example of a characteristic matrix M.

1 Examples of discrepancy bounds

Let $\mathcal{A} = \{\{1, 3, 5, \dots\}, \{1, 5, 9, 13, \dots\}, \dots\}$ be the family of all arithmetic progressions up to n (note that $|\mathcal{A}|$ is approximately n^2). Discrepancy of such families was well studied in the past. In the following three results $\operatorname{disc}(\mathcal{A})$ denotes $\min_\chi \max_{S\in\mathcal{A}} |\sum_{i\in S} \chi(i)|$, where the minimum is taken over all colorings $\chi\colon [n] \to \{-1, 1\}$ and the maximum is taken over all arithmetic progressions in $[n]$.

Theorem 1 ([9])**.** *We have a lower bound* $\operatorname{disc}(\mathcal{A}) = \Omega(n^{1/4})$.

Theorem 2 ([5])**.** *An upper bound* $\operatorname{disc}(\mathcal{A}) = \mathcal{O}(n^{1/4}(\log n)^{5/2})$ *holds.*

Theorem 3 ([8])**.** *The Roth's lower bound is tight up to a constant factor. That is,* $\operatorname{disc}(\mathcal{A}) = \Theta(n^{1/4})$ *holds.*

The last two results were derived using so called Partial Coloring Method [5].

2 Beck-Fiala Theorem

In this section we examine *set systems of bounded degree t*. That is, the set systems $\mathcal{S} = \{S_1, S_2, \dots, S_m\}$ where every element $i \in [n]$ appears in at most t sets $S_j \subseteq [n]$, $j \in [m]$. In other words, every column of the characteristic matrix M contains at most t nonzero entries.

Theorem 4 ([6])**.** *Let $\mathcal{S}$ be a set system over $[n]$ of bounded degree t, then* $\operatorname{disc}(\mathcal{S}) \le 2t-1$.

This result says that for set systems of bounded degree the discrepancy is independent of n. Before showing the proof, which uses so called Linear algebra method, we look at discrepancy of random colorings.

Given a random vector $\chi \in_{\mathcal{U}} \{-1, 1\}^n$, we ask what is the discrepancy of $\mathcal{S}$. It is not difficult to prove that for such χ and $S \subseteq [n]$ we have $\mathbb{E}[|\sum_{i\in[n]} \chi_i|] = \Theta(\sqrt{n})$ and thus $\mathbb{E}[|\sum_{i\in S} \chi_i|] = \Theta(\sqrt{|S|})$.

Using Chernoff bounds for $\chi \in_{\mathcal{U}} \{-1, 1\}^k$ we get that $\Pr[|\sum_{i=1}^k \chi_i| > t\sqrt{k}] \le 2e^{-t^2/4}$, where $\sqrt{k}$ is the standard deviation of the sum. Therefore if we have a set system $\mathcal{S} = \{S_1, \dots, S_m\}$, $S_j \subseteq [n]$ and a random coloring $\chi \in_{\mathcal{U}} \{-1, 1\}^n$, then we obtain $\Pr[|\sum_{i\in S_j} \chi_i| > t\sqrt{|S_j|}] \le 2e^{-t^2/4}$. Since $|S_j| \le n$ for every $j \in [m]$, we get

$$\Pr[|\sum_{i\in S_j} \chi_i| > t\sqrt{n}] \le 2e^{-t^2/4}.$$

Using the Union Bound we derive

$$\Pr[\exists j \in [m] : |\sum_{i\in S_j} \chi_i| > t\sqrt{n}] \le 2me^{-t^2/4}$$

and thus if we set $t := \mathcal{O}(\sqrt{\log m})$, then the probability is strictly less than one. Hence there exists a coloring $\chi \in \{-1, 1\}^n$ such that no set S_j has discrepancy greater than $t\sqrt{n}$. Therefore $\operatorname{disc}_\chi(\mathcal{S}) < \mathcal{O}(\sqrt{n \log m})$.

Proof of Theorem 4. We prove Beck-Fiala Theorem by the method of floating variables. For more information see [4].

Phase 0: Let us relax the problem from $\chi \in \{-1,1\}^n$ to $\chi \in [-1,1]^n$. Afterwards we use rounding to obtain integer coordinates of the vector χ. For the initial vector with small discrepancy we choose $\chi = (\chi_1, \ldots, \chi_n) = (0, \ldots, 0)$.

A variable χ_i, $i \in [n]$, is called *floating* if $\chi_i \notin \{-1,1\}$. Let us denote the set of indices of floating variables as F. For the initial vector χ we have $F = [n]$. We call a set $S_j \in \mathcal{S}$ *dangerous* if $|S_j \cap F| > t$.

Claim 5. *The number of dangerous sets is strictly less than* $|F|$.

Proof. We just use double-counting for the incidence matrix restricted to the set of floating variables F. For each dangerous set we have strictly more than t nonzero entries, but the sum of all entries of the matrix is at most $t|F|$, since $\mathcal{S}$ is a system of the bounded degree t. □

Phase r: In each phase we want to have $\sum_{i \in S \cap F} \chi_i = 0$ where S is a dangerous set. We know that there is a solution (the zero vector), but we would like to have another one.

Claim 6. *Let* $V_r = \{\chi \in [-1,1]^{|F|} \mid \sum_{i \in S \cap F} \chi_i = 0 \text{ holds}\}$, *then* V_r *is a subspace of* $\mathbb{R}^{|F|}$ *and* $\dim(V_r) \geq 1$.

Proof. Every dangerous set gives us at most one linear constraint and we already observed that the number of dangerous sets is strictly less than $|F|$. Since $V_r \subseteq \mathbb{R}^{|F|}$, we obtain the rest. □

The algorithm works as follows. First we pick a line $L \in V_r$ which exists thanks to the previous claim. Then we go along L until we hit the boundary of the cube $[-1,1]^n$. At the boundary we fix at least one floating variable χ_{r+1} to 1 or -1 and set $F_{r+1} = F_r \setminus \{\text{new fixed variables}\}$. We repeat this procedure until $F_r = \emptyset$, then output χ_r.

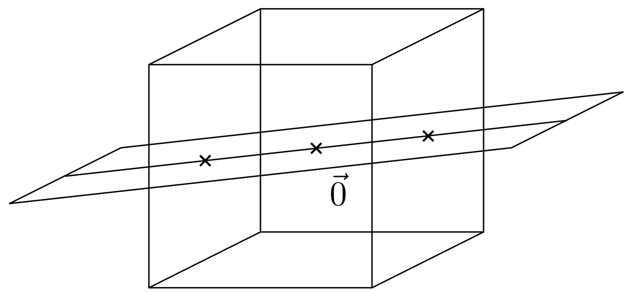

Figure 2: The cube $[-1,1]^n$ with the line L in the subspace V_r.

Claim 7. *Discrepancy of the final output vector is less than* $2t$.

Proof. Observe that once a dangerous set ceases to be dangerous, than it can never become dangerous again. We also know that for every dangerous set S we have an invariant $\sum_{i \in S \cap F} \chi_i = 0$.

Thus when a set stops being dangerous at a phase r, then $\sum_{i \in S \cap F_r} \chi_i = 0$ and we have at most t elements in $S \cap F$ which we can change. Any change of an element contributes by a value less than two, thus the total change is strictly less than $2t$. $\square$

Since the value of discrepancy is an integer, we have $\mathrm{disc}(\mathcal{S}) \leq 2t - 1$.

$\square$

Conjecture 8 (Beck and Fiala, 1981)**.** *Let $\mathcal{S}$ be a set system over $[n]$ of bounded degree t, then* $\mathrm{disc}(\mathcal{S}) = \mathcal{O}(\sqrt{t})$.

This conjecture remains open. The conjectured upper bound matches a known lower bound $\mathrm{disc}(\mathcal{S}) = \Omega(\sqrt{t})$. The current best upper bound $\mathcal{O}(\sqrt{t \log n})$ was proved by Banaszczyk [2], but it depends on the size n of the universe.

3 Six Standard Deviations Theorem

Suppose we have a set system $\mathcal{S} = \{S_1, \ldots, S_n\}$ over $[n]$. That is, the size of the set system $\mathcal{S}$ is the same as the size of the universe $[n]$. The main question of this section is: what is the the maximum discrepancy of such system? Random coloring gives us an upper bound $\mathcal{O}(\sqrt{n \log n})$, but this not the best we can do, as Spencer showed that even "six standard deviations suffice":

Theorem 9 ([10])**.** *Let $\mathcal{S} = \{S_1, \ldots, S_n\}$ be the set system over the universe $[n]$. Then* $\mathrm{disc}(\mathcal{S}) \leq 6\sqrt{n}$.

This result is stronger that the one obtained using random colorings, because for $|\mathcal{S}| = \mathcal{O}(n)$ the discrepancy is just $\mathcal{O}(\sqrt{n})$ whereas for random colorings we had $\mathcal{O}(\sqrt{n \log n})$. The result can be generalized to the case where $|\mathcal{S}| = m$, then we would get $\mathrm{disc}(\mathcal{S}) = \mathcal{O}(\sqrt{n \log \frac{2m}{n}})$. The original proof of this theorem uses a pigeon-hole argument based on the the Partial Coloring Method [5], therefore its proof is not constructive. In fact it was even conjectured that such an algorithm cannot exist [1], however this conjecture was recently disproved by Bansal [3]. Later in this section we introduce a new elementary constructive proof of this result.

Let us now gow back to the Spencer's original approach using the Partial Coloring Method. The main idea is to relax the problem to a *partial coloring* $\chi \in \{-1, 0, 1\}^n$ (note that we have an additional color 0). Now we want a partial coloring $\chi \in \{-1, 0, 1\}^n$ such that for every $S \in \mathcal{S}$ we have $|\sum_{i \in S} \chi_i| \ll \Delta$ for some $\Delta = \mathcal{O}(\sqrt{n})$ and at least half of the χ_i are nonzero.

If we can solve the relaxed problem with discrepancy bound $\mathcal{O}(\sqrt{n})$, then we can recursively use it for the zero columns (we know that at least one half of the columns are nonzero) of the incidence matrix. We follow this procedure until we obtain a complete coloring. Since the partial colorings generated in every step always have at least one half of their columns nonzero, we see that this recursive algorithm bounds discrepancy to

$$\mathcal{O}(\sqrt{n}) + \mathcal{O}(\sqrt{n/2}) + \mathcal{O}(\sqrt{n/4}) + \cdots = \mathcal{O}(\sqrt{n}).$$

In contrast to this technique of Spencer, our approach uses *fractional colorings* which take values in $[-1, 1]$ though many of the colors are close to $\{-1, 1\}$. In the rest of the notes we sketch the main idea of the proof of Theorem 9, for details see [7]. Also, we do not prove the result with the constant 6, but only with some absolute constant C (in order to make the presentation easier).

Proof of Theorem 9. Let $\Delta = \mathcal{O}(\sqrt{n})$ and suppose that $\mathcal{S} = \{S_1, \dots, S_n\}$ is our set system described by the matrix M. Let ϑ_j be the j-th row of M. Then $\langle \vartheta_j, \chi \rangle = \sum_{i \in S_j} \chi_i$ and we want to have $|\langle \vartheta_j, \chi \rangle| \leq \Delta$ for every j.

We start with $\chi \in [-1, 1]^n$. Then the restrictions on $|\langle \vartheta_j, \chi \rangle|$ correspond to hyperplanes cutting the cube $[-1, 1]^n$ and thus we get a polytope $P = \{\chi \in [-1, 1]^n \mid |\langle \vartheta_j, \chi \rangle| \leq \Delta\}$. Our goal is to prove that there exists $\chi \in P$ such that $|\{i \mid |\chi_i| = 1\}| \geq n/2$.

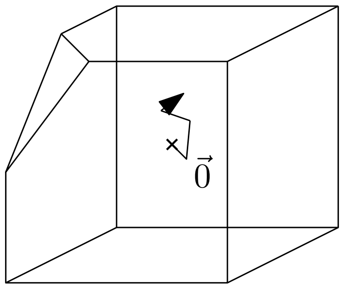

Figure 3: A random walk in the polytope P.

We use a similar idea as before to get rid of floats. We consider a random walk starting at the origin. Then, using tiny steps we are moving in P, until we hit one of its faces for the first time. Then the walk continues, but only inside of this face and we proceed in the same way until we eventually arrive to some vertex of P. The main idea behind this is that we basically want to find a point $x \in P$ which is as far away from origin as possible, because then many of the coordinates of x will be close to one in the absolute value. To analyze this EDGEWALK algorithm [7], we need to know some basic facts about Gaussian distribution. Let $\|\chi\|_2^2$ denote $\sum_{i=1}^n \chi_i^2$.

Fact 10. *Let ϑ be a vector such that $\|\vartheta\|_2 = 1$ and let $X \sim N(0,1)^n$. Then we have $\langle \vartheta, X \rangle \sim N(0, 1)$.*

Fact 11. *Sub-Gaussian Tails: let $X \sim N(0,1)^n$ and let $\vartheta \in \mathbb{R}^n$. Then*

$$\Pr[|\langle \vartheta, X \rangle| > t\|\vartheta\|_2] \leq 2e^{-t^2/4}.$$

For a linear subspace $V \subseteq \mathbb{R}^n$ we use $G \sim N(V)$ to denote the *Standard Gaussian distribution supported on* V : $G = G_1\vartheta_1 + \dots + G_d\vartheta_d$ where $\{\vartheta_1, \dots, \vartheta_k\}$ is an orthonormal basis for V and $G_1, \dots, G_d \sim N(0, 1)$ are independent standard Gaussian variables. Also let us call the constraints of the form $|\langle \vartheta_j, \chi \rangle| \leq \Delta$, $j \in [n]$, as *discrepancy constraints* and the constraints of the form $|\chi_i| \leq 1$, $i \in [n]$, as *coloring constraints.*

The algorithm EDGEWALK runs in phases $t = 0, 1, \dots, T$. For some approximation parameter δ, let us denote as $\mathrm{Color}_t = \{i \mid |(\chi^t)_i| \geq 1 - \delta\}$ and $\mathrm{Disc}_t = \{j \mid |\langle \vartheta_j, \chi^t \rangle| \geq \Delta - \delta\}$ the sets of coloring and discrepancy constraints nearly hit so far. Also let $V_t = \{\chi \mid \chi_i = 0$ for every $i \in \mathrm{Color}_t$ and $\langle \vartheta_j, \chi \rangle = 0\}$ for every $j \in \mathrm{Disc}_t\}$ be the linear subspace orthogonal to the nearly hit coloring and discrepancy constraints. In every step of the algorithm we set χ^{t+1} to $\chi^t + \delta\gamma^t$ where $\gamma^t \leftarrow N(\vartheta_t)$ is the step size in the random walk.

Claim 12. *If $T = c/\gamma^2$, then with probability at least $1/2$ the coloring χ_T satisfies $\chi_T \in P$ and $|\{i \mid |(\chi_T)_i| \geq 1 - \delta\}| \geq n/2$.*

The intuition behind this claim is that the coloring constraints are closer to the origin than the discrepancy constraints, thus the probability that we hit them first is approximately $e^{-1/4}$ which is higher than e^{-6^2}. This is enough to satisfy at least $n/2$ coloring constraints in the end.

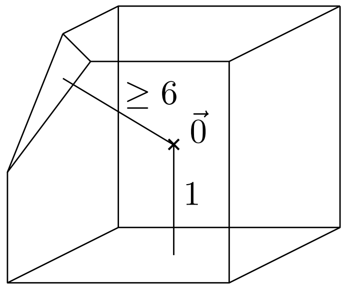

Figure 4: Polytope P with the face $\langle \vartheta_j, \chi \rangle = 6\sqrt{n}$.

Lemma 13. *We have $\chi_T \in P$ with probability at least 0.9.*

Sketch of the proof. It suffices to use tail bound for Gaussians in the case $\gamma \ll \delta$. □

Lemma 14. *The expected value of the number of nearly violated constraints satisfies $\mathbb{E}[|Disc_T|] = 0.01n$.*

Sketch of the proof. It follows from the fact that the probability that any discrepancy constraint is satisfied is at most $2e^{-6^2/4}$. In this part of the proof we use the assumption that $|\mathcal{S}| = n$. □

Lemma 15. *We have $\mathbb{E}[\dim(V_T)] < 0.5n$.*

Sketch of the proof. In every step we move slightly further away from the origin. □

Altogether we have $\chi_T \in P$ and $\mathbb{E}[|\mathrm{Color}_T|] \geq (0.5 - 0.01)n$ and from this we get $\Pr[|\mathrm{Color}_T| \geq n/2] \geq 1/8$. Thus we have just proved that there is an efficient randomized algorithm which can with positive probability find $x \in [-1, 1]^n$ which satisfies the coloring and discrepancy constraints for some approximation parameter δ. This can be understood as a new version of the Partial Coloring Lemma which we now use to finish the proof of Theorem 9.

Starting at $\chi_0 = (0, \dots, 0)$ we apply the new Partial Coloring Lemma to obtain such $\chi \in P$ (we can repeat the algorithm $\mathcal{O}(\log n)$ times to boost the probability of finding x and thus ignore the probability that we do not find such a vector). Then we iteratively apply the same lemma to the restricted system described by the coordinates which were not fixed in the previous steps and concatenate the resulting vectors appropriately. Afterwards we obtain $\chi \in [-1, 1]^n$ such that $\chi_i \geq 1 - \delta$ for every $i \in [n]$ and $|\langle \vartheta_j, \chi \rangle| = \mathcal{O}(\sqrt{n})$ for every $j \in [n]$. Then it suffices to round such χ to get proper coloring $\chi' \in \{-1, 1\}^n$. To do so, we set for $i \in [n]$, $\chi'_i = sign(\chi_i)$ with probability $(1+|\chi_i|)/2$ and $\chi'_i = -sign(\chi_i)$ with probability $(1 - |\chi_i|)/2$ (so that $\mathbb{E}[\chi'_i]) = \chi_i$). Using the Chernoff Bounds it can be shown that the resulting proper coloring χ' still has discrepancy $\mathcal{O}(\sqrt{n})$ (see [7] for details).

$\square$

Before the end of this section we state the following generalization of the statement above which can be used for many applications including counting the discrepancy of arithmetic progressions, half-planes, rectangles and so on.

Theorem 16 (Partial Coloring Lemma)**.** *Let $\vartheta_1, \dots, \vartheta_m \in \mathbb{R}^n$ and $\Delta_1, \dots, \Delta_n \geq 0$. Then there is a vector $\chi \in [-1,1]^n$ such that at least half of its coordinates equals one, and if $\sum_{j=1}^m \exp -\lambda_j^2/4 \leq n/16$ then we have $|\langle \vartheta_j, \chi\rangle| \leq \|\lambda_j\|_2$.*

4 Bio:

Raghu Meka is currently a postdoctoral fellow at the Institute for Advanced Study, Princeton and DIMACS, Rutgers. He received his PhD from the University of Texas at Austin in 2011. He is a recipient of the Bert Kay best dissertation award, the Dean's Excellence award and an MCD fellowship at UT Austin. His main interests are in complexity theory, pseudo-randomness, algorithm design, learning theory and more generally in anything to do with probability.

References

[1] N. Alon and J. Spencer, The Probabilistic Method, in: *Wiley Series in Discrete Mathematics and Optimization*, John Wiley & Sons, 2011.

[2] W. Banaszczyk, Balancing vectors and Gaussian measure of n-dimensional convex bodies, *Random Structures and Algorithm* **12** (1998), 351–360.

[3] N. Bansal, Constructive algorithms for discrepancy minimization, in: *FOCS '10*, Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, IEEE Computer Society, 2010, 3–10.

[4] I. Bárány, On the Power of Linear Dependencies, in: *Building Bridges*, Bolyai Society Mathematical Studies, 19, Springer, Berlin Heidelberg, 2008, 31–45.

[5] J. Beck, Roth's estimate of the discrepancy of integer sequences is nearly sharp, *Combinatorica* **1** (1981), 319–325.

[6] J. Beck and T. Fiala, "Integer-making" theorems, *Discrete Applied Mathematics* **3** (1981), 1–8.

[7] S, Lovett and R. Mekka, Constructive Discrepancy Minimization by Walking on the Edges, in: *FOCS '12*, Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science, IEEE Computer Society, 2012, 61–67.

[8] J. Matoušek and J. H. Spencer, Discrepancy in arithmetic progressions, *Journal of the American Mathematical Society* **9** (1996), 195–204.

[9] K. F. Roth, Remark concerning integer sequences, *Acta Arith.* **9** (1964), 257–260.

[10] J. Spencer, Six standard deviations suffice, *Transactions of the American Mathematical Society* **289** (1985), 679–706.