Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών

Τμήμα Μαθηματικών
Μεταπτυχιακό Προγραμμα Λογικής και Αλγορίθμων

The discrepancy problem

Διπλωματική Εργασία

του

Ιωάννη Παναγέα

Επιβλέπων: Ευστάθιος Ζάχος
Καθηγητής Ε.Μ.Π.

Αθήνα, Δεκέμβριος 2011

Η παρούσα Διπλωματική Εργασία

εκπονήθηκε στα πλαίσια των σπουδών

για την απόκτηση του

Μεταπτυχιακού Διπλώματος Ειδίκευσης

στη

Λογική και Θεωρία Αλγορίθμων και Υπολογισμού

που απονέμει το

Τμήμα Μαθηματικών

του

Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών

Εγκρίθηκε την 28η Δεκεμβρίου 2011 από Εξεταστική Επιτροπή αποτελούμενη από τους:

Ονοματεπώνυμο	Βαθμίδα	Υπογραφή
1. : Ε. Ζάχος	Καθηγητής	……………………
2. : Α. Παγουρτζής	Επίκ. Καθηγητής	……………………
3. : Δ. Φωτάκης	Λέκτορας	……………………

Abstract

In this thesis we examine the combinatorial discrepancy problem. We have a ground set of elements $\Omega = \{1, 2, ..., n\}$ and a family of sets $\mathcal{A}$ and we prove that there exists a coloring $\chi : \Omega \to \{-1, +1\}$ such that $disc(\mathcal{A})$ is $\Theta(\sqrt{n})$. To prove the upper bound of this result, we use the well-known entropy method. Afterwards, since the proof is not constructive, we give an algorithm that finds such a coloring with probability $\frac{1}{\log n}$. Finally, we prove the lower bound using techniques from Linear Algebra and we also mention some modifications of the combinatorial discrepancy based on the structure of $\mathcal{A}$ and the number of sets it contains.

Keywords Combinatorial Discrepancy, Partial Coloring, Entropy Method

Περίληψη

Στην παρούσα διπλωματική εργασία εξετάζουμε το combinatorial discrepancy problem. Έχουμε ένα σύνολο $\Omega = \{1, 2, ..., n\}$ και μία οικογενεια συνόλων $\mathcal{A}$ και αποδεικνύουμε ότι υπάρχει χρωματισμός $\chi : \Omega \to \{-1, +1\}$ τέτοιος ώστε $disc(\mathcal{A})$ είναι $\Theta(\sqrt{n})$. Για να αποδείξουμε το άνω φράγμα αυτού του αποτελέσματος, χρησιμοποιούμε τη entropy method. Έπειτα, αφού η απόδειξη δεν είναι κατασκευαστική, δινουμε έναν αλγόριθμο που βρίσκει τέτοιο χρωματισμό με πιθανότητα $\frac{1}{\log n}$. Τέλος, αποδεικνύουμε το κάτω φράγμα χρησιμοποιώντας τεχνικές από γραμμική άλγεβρα και επίσης αναφέρουμε κάποιες παραλλαγες του combinatorial discrepancy που βασίζονται στη δομή του $\mathcal{A}$ και στο πληθος των συνόλων που περιέχει.

Λέξεις κλειδιά Combinatorial Discrepancy, Partial Coloring, Entropy Method

Acknowledgements

I would like to deeply thank my three teachers and members of the committee Prof. S.Zachos, Prof. A. Pagourtzis and Prof. D. Fotakis. I feel very lucky I had the chance to be taught by them; I feel they have influenced me in a very positive way not only academically but also on a personal level. Particularly I want to thank Prof. S. Zachos for his guidance and his persistence in helping me make this thesis better.

Many thanks to the members of the Corelab team, namely Evangelos, Georgia, George, Andreas, Paris, Thanasis, Eleni, Eirini, Themis, Paris, Christos, Andreas for the interesting discussions we made. Finally, i want to thank my parents for their support.

Ioannis Panageas

Contents

Chapter 0

Preliminaries

0.1 Introduction

In this thesis we examine the "classical" versions of combinatorial discrepancy. Generally, we are given a set $\Omega = \{1, ..., m\}$ of elements and also a family of n sets $\mathcal{A} = \{A_1, ..., A_n\}$ where $A_i \subseteq \Omega$. A coloring is considered a function $\chi : \Omega \rightarrow \{-1, +1\}$ and discrepancy of a set A_i is just $disc_\chi(A_i) = |\sum_{s \in A_i} \chi(s)|$. Discrepancy of $\mathcal{A}$ is defined as

$$disc(\mathcal{A}) = \min_{\chi} \max_{A_i \in \mathcal{A}} disc_\chi(A_i) \tag{1}$$

This problem is very important and lots of researchers have focused on it. There is a branch of mathematics that is called discrepancy theory that studies the inevitable irregularities of distributions. Discrepancy has a lot of applications in different fields, such as communication complexity, high-dimensional algorithms, computational geometry, etc. We are going to prove lower and upper bounds for $disc(\mathcal{A})$, using very powerful techniques such as entropy method, SDP and Hadamard matrices. First of all, we will show the well-known result of Spencer [11] where he proved that in case $m = n$ it holds that $disc(\mathcal{A}) = O(\sqrt{n})$ up to constant factors. In his paper he proved that $disc(\mathcal{A}) \leq 6\sqrt{n}$. However, this kind of proof is not constructive. In words, he shows the existence of a coloring χ, such that $disc_\chi(A_i) \leq 6\sqrt{n}$ for $1 \leq i \leq n$ without finding it, and his proof is based on properties of entropy as subadditivity and pigeonhole principle. Twenty five years later, Bansal in his paper [3], gave an elegant randomized algorithm that succeeds in finding a coloring that upper bounds the discrepancy from $O(\sqrt{n})$ with probability at least $\frac{1}{\log n}$. Finally, it can be proven using techniques from Linear Algebra or Probabilistic Methods that this bound is tight. Namely, we will construct sets A_i such that $disc(\mathcal{A}) = \Omega(\sqrt{n})$. To begin with, we discuss below some basic definitions and theorems that are important for the proofs that follow in the next chapters.

0.1.1 An example - Discrepancy of lights and switches

Before starting with the basics for the combinatorial discrepancy problem and the techniques we will need, we will give an example of discrepancy in matrices, to see how powerful technique is the probabilistic method. However, as it will become clear later, probabilistic method is not enough when we deal with combinatorial discrepancy as was defined in (1). The following problem is called Discrepancy of matrix of lights and switches.

We are given a $n \times n$ matrix $A = \{a_{ij} = \pm 1\}$ (these are the lights), row switches x_i and column switches y_j $(i, j = 1, ..., n)$. The objective is to find $x_i, y_j \in \{\pm 1\}$ so as to maximize

$|\sum_{i,j} a_{ij}x_iy_j|$. We will prove that there exists a matrix A such that

$$\max_{x_i,y_j} |\sum_{i,j} a_{ij}x_iy_j| < cn^{3/2}$$

To prove this, assume that each entry a_{ij} of matrix A takes values uniformly at random from $\{-1,+1\}$. We also consider that we know a priori the values of x_i, y_j (we suppose that are fixed). Then, obviously $\Pr[A_{ij}x_iy_j = -1] = \Pr[a_{ij}x_iy_j = +1] = \frac{1}{2}$ (A_{ij} is a random variable that corresponds to the value of a_{ij}). Therefore $A_{ij}x_iy_j$ are independent random variables and also $E[A_{ij}x_iy_j] = 0$. Hence from chernoff bounds, given that $S_{n^2} = \sum_{i,j} A_{ij}x_iy_j$ we have that

$$\Pr[|S_{n^2}| > a] < 2e^{-\frac{a^2}{2n^2}} \quad (i)$$

Hence, using union bound we have that the probability some configuration of the switches gives discrepancy more than a is upper bounded by $2^{2n} \cdot 2e^{-\frac{a^2}{2n^2}}$ (we have 2^{2n} possible configurations for the switches). Letting $2^{2n} \cdot 2e^{-\frac{a^2}{2n^2}} = 1$ (ii), it follows that there is a matrix A, such that for each configuration of the switches, the discrepancy is less than a. Finally solving the equality ii, we have that $a^2 = 2n^2(2n+1)\ln 2 \Rightarrow a = 2n\sqrt{\ln 2}\sqrt{n+1/2}$. By picking $c = 2\sqrt{\ln 2}$ the result follows.

0.1.2 Easy Bound

Before continuing with the definitions and techniqual points of the proofs, let's examine an easy upper bound which can be derived with the use of elementary ideas. It is rather common in probabilistic methods, to color the elements independently and uniformly at random. So, considering that χ takes values $-1, 1$ with probability $\frac{1}{2}$ and independently for each element, we get that $E[\chi(a)] = 0$ for each $a \in \Omega$, so using Chernoff Bounds it follows that

$$Pr[|\chi(A_i)| > a] < 2e^{-\frac{a^2}{2|A_i|}} \leq 2e^{-\frac{a^2}{2m}}$$

where $\chi(A_i) = \sum_{s \in A_i} \chi(s)$. Finally, by choosing $a^2 = 2m \ln 2n$ and using union bound it follows that

$$\Pr_{\chi}[disc(A) > \sqrt{2m\ln(2n)}] \leq \sum_{i=1}^{n} \Pr[|\chi(A_i)| > \sqrt{2m\ln(2n)}] < 1 \quad (2)$$

Therefore, there exists coloring χ such that $disc_{\chi}(A_i) \leq \sqrt{2m\ln(2n)}$ for every i and hence $disc(\mathcal{A}) \leq \sqrt{2m\ln(2n)}$.

It is remarkable that for $m = n$ it occurs that $disc(\mathcal{A})$ is $O(\sqrt{2n\ln(2n)})$. On the other hand, as we promised before, we will prove something slightly better. Another thing that we should mention is that the standard deviation of the random variable discrepancy (we consider random coloring) for each set A_i is $\sqrt{A_i}$ and so $O(\sqrt{n})$ (assuming $m = n$) so we expect from discrepancy to have lower upper bound, closer to the standard deviation. In a sense, it looks that uniform the random coloring is far away from the best we can do (since Spencer [11] proved that $disc(\mathcal{A})$ is $O(\sqrt{n})$). Trying to do a cleverer random coloring, for example splitting uniformly at random Ω in disjoint pairs and give each pair opposite signs, we can improve $\sqrt{2m\ln(2n)}$ by a constant factor, so even that it is not enough.

0.2 Properties of Entropy

For the purpose of this thesis, we use the notion of *binary entropy*, measure that was introduced by Shannon [10]. Let X be a random variable and S be the range of X.

Definition 0.1. *Binary entropy* is denoted by

$$\sum_{i \in S} \Pr[X = i] \log \frac{1}{\Pr[X = i]} \tag{3}$$

Intuitively, the entropy of a random variable X measures the amount of information X encodes. We recommend [2] for furher studying on entropy.

0.2.1 Good chance and Uniformity optimal

Lemma 0.1. *(Good chance) Let X be a (discrete) random variable such that $H(X) \leq K$ for some K. Then there exists a value x in the range of X such that $\Pr[X = x] \geq 2^{-K}$*

Proof. Assume that $\Pr[X = x] < 2^{-K}$ for all x, then $\Pr[X = x] \log \frac{1}{\Pr[X=x]} > K \Pr[X = x]$ and hence $H(X) > \sum_x K \Pr[X = x] = K$ (contradiction) □

Lemma 0.1 is very useful for proving Spencer's Result, because in order to prove that a random variable X takes a value with sufficient probability p, it suffices to prove that $H(X) \leq \log \frac{1}{p}$.

Lemma 0.2. *(Uniformity optimal) Let X be a (discrete) random variable which takes at most k distinct values. Then $H(X) \leq \log_2 k$.*

Proof. Let $f(x) = \log_2 x$. Then $f''(x) < 0$ and hence f is concave. From Jensen's Inequality we get that

$$\begin{aligned}\sum_x \Pr[X = x] \log_2 \left(\frac{1}{\Pr[X = x]}\right) &\leq \log_2 \left(\sum_x \frac{\Pr[X = x]}{\Pr[X = x]}\right) \\ &\leq \log_2 k\end{aligned}$$

□

0.2.2 Subadditivity

Theorem 0.1. *Let $X_1, ..., X_n$ be random variables and $X = (X_1, ..., X_n)$ be a random variable which is the cartesian product. The following holds:*

$$H(X) \leq \sum_{i=1}^{n} H(X_i) \tag{4}$$

Proof. We prove the following claim and then it comes from induction on n.

Claim: Let X, Y random variables. Then $H(X, Y) \leq H(X) + H(Y)$.

Let S, T be the range of X, Y respectively. Then we get that

$$H(X) = \sum_{i \in S} \Pr[X = i] \log \frac{1}{\Pr[X = i]} = \sum_{i \in S} \sum_{j \in T} \Pr[X = i, Y = j] \log \frac{1}{\Pr[X = i]}$$

$$H(Y) = \sum_{j \in T} \Pr[Y = j] \log \frac{1}{\Pr[Y = j]} = \sum_{j \in T} \sum_{i \in S} \Pr[X = i, Y = j] \log \frac{1}{\Pr[Y = j]}$$

$$H(X, Y) = \sum_{i \in S} \sum_{j \in T} \Pr[X = i, Y = j] \log \frac{1}{\Pr[X = i, Y = j]}$$

Therefore, $H(X)+X(Y)-H(X,Y)=\sum_{i\in S}\sum_{j\in T}\Pr[X=i,Y=j]\log\frac{\Pr[X=i,Y=j]}{\Pr[X=i]\Pr[Y=j]}$. Using the fact that $f(x)=x\log x$ is convex (it comes from $(x\log x)''=\frac{1}{x}>0$), from Jensen's inequality it follows that

$$\begin{aligned}&H(X)+H(Y)-H(X,Y)\\&=\sum_{i\in S}\sum_{j\in T}\Pr[X=i]\Pr[Y=j]\frac{\Pr[X=i,Y=j]}{\Pr[X=i]\Pr[Y=j]}\log\frac{\Pr[X=i,Y=j]}{\Pr[X=i]\Pr[Y=j]}\\&\geq 1\log 1=0\end{aligned}$$

□

0.3 Entropy Method and Pigeonhole Principle

0.3.1 Partial Coloring

A very powerful method for achieving lower bounds on discrepancy is the partial coloring method. In almost every result is used as one of the basic ideas.

Definition 0.2. *Partial coloring* of a set Ω is a function $\chi:\Omega\to\{-1,0,+1\}$. An element x is colored if $\chi(x)\neq 0$ else it is uncolored.

For example, in Spencer's proof, the idea is to find a partial coloring χ such that the number of uncolored elements is bounded by a constant percentage of the number of elements and also the discrepancy of $\mathcal{A}$ for the given χ to be as small as possible. This kind of procedure is repeated (find a new partial coloring χ' with again the same properties, for the uncolored elements instead of dealing with all the elements etc, for the proof see chapter 1).

Theorem 0.2. *(Partial Coloring Lemma) Let $\mathcal{B}$ and $\mathcal{C}$ be disjoint subsets of $\mathcal{A}$ such that $|C_i|\leq s$ for every $C_i\in\mathcal{C}$ and*

$$\prod_{B_i\in\mathcal{B}}(|B_i|+1)\leq 2^{(n-1)/5}$$

Then there exists a partial coloring χ such that at least $\frac{n}{10}$ elements are colored, $disc_\chi(B_i)=0$ for every B_i and $disc_\chi(C_i)\leq\sqrt{2s\ln(4|\mathcal{C}|)}$ for every C_i.

The proof of the theorem 0.2 has 2 main steps. To give some intuition, family of sets $\mathcal{B}$ stands for few sets where we insist that discrepancy is 0 and $\mathcal{C}$ plays the role of the rest sets. The number $\sqrt{2s\ln(4|\mathcal{C}|)}$ has been taken from the easy bound we proved earlier. The constant factor is 4, in order to accomplish that at least half of the random colorings work for the constraint $disc_\chi(C_i)\leq\sqrt{2s\ln(4|\mathcal{C}|)}$. On the other hand, in order to satisfy the constraint $disc_\chi(B_i)=0$, it suffices to take two colorings χ_1,χ_2 that satisfy the first constraint, define the same discrepancy for every B_i, namely $disc_{\chi_1}(B_i)=disc_{\chi_2}(B_i)$ and consider the partial coloring (it is clear why it is partial, notice that χ_1,χ_2 are not partial)

$$\chi'=\frac{\chi_1-\chi_2}{2}\tag{5}$$

This proposition, namely the existence of χ_1,χ_2 that satisfy the first constraint and $disc_{\chi_1}(B_i)=disc_{\chi_2}(B_i)$ comes from the hypothesis for B_i and Pigeonhole Principle. Finally, to certify that the partial coloring χ' leaves uncolored for example $\frac{9n}{10}$ elements, it suffices to choose χ_1,χ_2 such that they differ in at least $n-\frac{9n}{10}=\frac{n}{10}$ elements.

Below we discuss a refinement of the Partial coloring lemma, which uses entropy and it can found in bibliography with the term entropy method. The proof of spencer's result will actually be a corollary of that method, which will be formally proved in the next chapter. It is remarkable that the proof is not algorithmic because of the use of Pigeonhole Principle (similar use of Pigeonhole Principle to the Partial coloring Lemma).

0.3.2 Entropy Method

We will use the notation that Bansal [3] used in his paper. We recommend you to read also [7].

Theorem 0.3. *(Quantitative version) Let $\mathcal{A}$ be a family of n sets A_i, $\Omega = \{1, 2, ..., n\}$ and $\Delta_{A_i} \geq 2\sqrt{|A_i|}$ be a given number for every set A_i. Suppose that*

$$\sum_{A_i \in \mathcal{A}} K e^{-\Delta_{A_i}^2/(4|A_i|)} \log_2 \left(2 + \frac{\sqrt{|A_i|}}{\Delta_{A_i}} \right) \leq \frac{n}{5} \tag{6}$$

with K a constant. Then there exists a partial coloring $\chi : \Omega \to \{-1, 0, +1\}$ such that $disc_\chi(A_i) < \Delta$ for all $A_i \in \mathcal{A}$ and the number of colored elements is at least $\frac{n}{10}$.

This theorem is the key to prove Spencer's result (the existence of a coloring χ which certifies that $disc(\mathcal{A})$ is $O(\sqrt{n})$ (using this theorem, the proof is just three lines). Bansal also was based on this theorem to give a randomized algorithm for finding the coloring. The proof will be postponed until the next chapter.

0.4 Bounded degree case and other versions

In this final section of chapter 0, we will mention some different versions of the discrepancy problem, such as the bounded degree case and the permutation problem. For the former version, we consider that each element of Ω can be used at most k times. This setting can be seen as a hypergraph, where the vertices are the elements of Ω and have degree at most k and each edge corresponds to a set A_i. The goal is to find a coloring with values $\{-1, +1\}$ such that the number of -1's and 1's are close to each other for each set in $\mathcal{A}$. Beck and Fiala [4] proved that $disc(\mathcal{A}) \leq 2k - 1$ using techniques from Linear Algebra. A proof of their result can be found in chapter 2.

For the latter version, commonly known in bibliography as k-permutation problem, the setting is the following: Let $\pi_1, ..., \pi_k$ be k permutations of the elements of Ω, $A_{ij} = \{\pi_i(1), ..., p_i(j)\}$ for $1 \leq j \leq n$ and $1 \leq i \leq k$ and $\mathcal{A}$ be the collection of all A_{ij}. An example for $k = 2$ is $\mathcal{A} = \{\{\pi_1(1)\}, \{\pi_1(1), \pi_1(2)\}, ..., \{\pi_1(1), \pi_1(2), ..., \pi_1(n)\}, \{\pi_2(1)\}, \{\pi_2(1), \pi_2(2)\}, ..., \{\pi_2(1), \pi_2(2), ..., \pi_2(n)\}\}$. It has been proven by Srinivasan [13] that k-permutation problem is $O(\sqrt{k} \log n)$ and recently Newman et al [9] proved that discrepancy is unbounded even for the case $k = 3$, it is of order $\Omega(\log n)$ (the conjecture of Beck that the discrepancy of 3-permutation problem is $O(1)$ isn't true).

Chapter 1

Algorithmic vs Non-Algorithmic

1.1 Introduction

The first half of this chapter concerns the entropy method, which will be more clear with the proof of Spencer's result at section 1.2. The second half concerns Bansal's algorithmic result. It is remarkable that for more than 25 years, it wasn't known any algorithmic guarantee better than the random coloring we presented in the previous chapter (section 0.1.2). Before we continue with the two amazing results - proofs, we mention some necessary theorems that we are going to use.

1.1.1 Gaussian random variables and tail bounds

Definition 1.1. Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2 has probability distribution function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

An important property (additivity) that we will use is the fact that a linear combination of two random variables $x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ follows gaussian distribution too, $a_1x_1 + a_2x_2 \sim \mathcal{N}(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$.

Definition 1.2. A martingale is a sequence $X_0, ..., X_m$ of random variables so that for $0 \leq i < m$

$$E[X_{i+1}|X_i, ..., X_0] = X_i$$

Theorem 1.1. *(Tail bounds) Let $0 = X_0, X_1, ..., X_n$ be a martingale with increments $Y_i = X_i - X_{i-1}$. Suppose that $Y_i|(X_{i-1}, ..., X_0)$ has distribution c_iG where G is the gaussian distribution $\mathcal{N}(0, 1)$ and $|c_i| \leq 1$ constant. Then it holds that*

$$\Pr[|X_n| \geq a\sqrt{n}] \leq 2e^{-a^2/2} \tag{1.1}$$

Proof. Let $\lambda > 0$, then we get that $E[e^{\lambda Y_i}|X_{i-1}, ..., X_0] \leq e^{\lambda^2 c_i^2/2} \leq e^{a^2/2}$ (i). Using properties of conditional expectation (see appendix) we get that

$$\begin{aligned} E[e^{\lambda X_n}] &= E[e^{\lambda Y_n} e^{\lambda X_{n-1}}] \\ &= E[e^{\lambda X_{n-1}} E[e^{\lambda Y_n}|X_{n-1}, ..., X_0]] \\ &\overbrace{\leq}^{\text{(i)}} e^{\lambda^2/2} E[e^{\lambda X_{n-1}}] \end{aligned}$$

Therefore taking the telescopic product, it follows that $E[e^{\lambda X_n}] \leq e^{\lambda^2 n/2}$. Hence from Markov's inequality (see appendix) it occurs that

$$\Pr[X_n \geq a\sqrt{n}] = \Pr[e^{\lambda X_n} \geq e^{\lambda a\sqrt{n}}] \leq \frac{E[e^{\lambda X_n}]}{e^{\lambda a\sqrt{n}}} \leq e^{\lambda^2 n/2 - \lambda a\sqrt{n}}$$

Finally substituting $\lambda = \frac{a}{\sqrt{n}}$ and by symmetry the result follows. □

1.1.2 Semidefinite Programming

Let A_i and C be $n \times n$ symmetric matrices and b_i be real numbers. A general SDP has the following form:

$$\begin{aligned} &\max C \bullet X \\ \text{s.t} \quad & A_i \bullet X \leq b_i \\ & X \succeq 0 \\ & X \text{ is symmetric} \end{aligned}$$

where $A \bullet B = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}$. SDP is important in convex optimization and it is known that it can be solved efficiently (see [6]). For our purpose (for Bansal's result), we are interested in finding efficiently a feasible solution to the SDP and not the optimal.

1.2 Spencer's result

1.2.1 Proof of Entropy Method

In this section we will prove theorem 0.3. Let χ be a coloring uniformly at random and $b_i(\chi) = \left\lfloor \frac{\sum_{s \in A_i} \chi(s)}{2\Delta_{A_i}} \right\rceil$ (random variable) where $\lfloor u \rceil$ denotes the nearest integer to u. Using Chernoff bounds (see appendix) we get that

$$\Pr[b_i(\chi) = 0] \geq 1 - 2e^{-5} \tag{1.2}$$

$$\Pr[b_i(\chi) = s] = \Pr[b_i(\chi) = -s] \leq e^{-5(2s-1)^2} \tag{1.3}$$

Let $p_s = \Pr[b_i(\chi) = s]$. Then from elementary calculus and using the fact that $-x \log x$ is nondecreasing on $(0, \frac{1}{e})$ and that $\log(1-x) \leq -x$ we get that

$$\begin{aligned} H(b_i) &= -p_0 \log p_0 - \sum_{s \geq 1} p_s \log p_s \\ &\leq 4e^{-5} \sum_{s \geq 1} \frac{5(2s-1)^2}{\ln 2} e^{-5(2s-1)^2} \\ &< 14e^{-5} < 0.1 \end{aligned}$$

Hence the constraint of the theorem 0.3 holds since the maximum of left hand side of the inequality happens when $\Delta_{A_i} \approx 2\sqrt{\Delta_{A_i}}$ and $14e^{-5} < \frac{1}{5}$.

Moreover we use theorem 0.1 (subadditivity) and we get that

$$H(b) \leq 14ne^{-5}$$

where b is the random variable $(b_1, ..., b_n)$. Additionally from lemma 0.1 (good chance) and the inequality above, it occurs that there exists a b_0 such that

$$\Pr[b = b_0] \geq 2^{-\epsilon n}$$

with $\epsilon = 14e^{-5}$ (this is the step that makes the proof non-constructive, pigeonhole in a sense). Since χ was taken uniformly at random, it follows that there exists a set of colorings C such that $b_i(\chi_1) = b_i(\chi_2)$ for all $1 \leq i \leq n$ and $\chi_1, \chi_2 \in C$ and $|C| \geq 2^{(1-\epsilon)n}$. To finish the proof we will prove that there exist colorings $\chi_1, \chi_2 \in C$ which differ in at least $\frac{n}{10}$ elements. To do this, let N be the number of colorings that differ in less that $\frac{n}{10}$ elements. Clearly we have that

$$\begin{aligned} N = \sum_{q < \frac{n}{10}} \binom{n}{q} 2^q &< \left(\frac{2en}{n/10}\right)^{n/10} \\ &= (20e)^{n/10} \\ &< 2^{6n/10} \\ &< 2^{(1-\epsilon)n} = |C| \end{aligned}$$

and thus there exist $\chi_1, \chi_2 \in C$ that differ in at least $\frac{n}{10}$ elements. Hence the partial coloring $\chi' = \frac{\chi_1 - \chi_2}{2}$ colors at least $\frac{n}{10}$ elements and the discrepancy of each set A_i is at most Δ_{A_i} (since $b_i(\chi_1) = b_i(\chi_2)$ for all i, they lie in the same interval $[2\Delta_{A_i}(k - 0.5), 2\Delta_{A_i}(k + 0.5))$ for some integer k and hence $disc_{\chi'}(A_i) = \frac{|\sum_{s \in A_i} \chi_1(s) - \sum_{s \in A_i} \chi_2(s)|}{2} \leq \Delta_{A_i}$) and the proof finishes.

1.2.2 Corollary - Spencer's result

Theorem 1.2. *(J.H.Spencer, [11]) Let $\mathcal{A}$ be a family of n subsets of a set $\Omega = \{1, 2, ..., n\}$. Then*

$$disc(\mathcal{A}) \leq c\sqrt{n} \tag{1.4}$$

for c constant (Spencer proved it for $c = 6$).

Proof. It is clear that for c sufficiently large, we can use the theorem 0.3, therefore there exists a partial coloring χ_1 which leaves uncolored at most $\frac{9n}{10}$ elements and $disc_{\chi_1}(A_i) \leq c\sqrt{n}$. Having obtained χ_1 coloring, let Ω_2 be the set of uncolored elements and $\mathcal{A}_{\in}$ be the "induced" family of sets on Ω_2. Similarly from theorem 0.3, there exists a partial coloring χ_2 which leaves uncolored at most $\frac{9^2 n}{10^2}$ elements $disc_{\chi_1}(A_{2\ i}) \leq c\sqrt{n}$ (and so on). Hence at step t, we have at most $(\frac{9}{10})^t n$ elements uncolored. We may stop at some step k such that the uncolored elements are sufficiently small (say constant, hence $k = O(\ln n)$ and the discrepancy will be at most $\sum_{j=1}^{k} c\sqrt{(\frac{9}{10})^j n \ln(2(\frac{10}{9})^j)} = O(\sqrt{n})$. □

1.3 Bansal's result

1.3.1 Sketch of the algorithm

The algorithm has some similarities with the proof of Beck-Fiala in the next chapter. We consider a vector $\chi \in [-1, 1]^n$ (*fractional coloring*) which is initially $\chi_0 = (0, 0, ..., 0)$. At step t, vector $\chi_t = \chi_{t-1} + c_t$ where c_t is a tiny value. If $\chi_t(i)$ (at step t) is very close to 1 or -1 then we set $\chi_t(i) = 1$ or -1 respectively with probability very close to 1 and with probability very close to 0, we will set $\chi_t(i)$ the opposite value. The choice of the updates c_t depends on a certain SDP and is made so as to be able to use theorem 0.3 (entropy method). In words, the variables that are floating (have not been fixed yet), do a random walk certifying that at each time the hypothesis of entropy method holds.

The main theorem from which follows as a corollary an algorithm that finds a proper coloring with discrepancy $O(\sqrt{n})$ is the following:

Theorem 1.3. *([3]) Let $x \in [-1,1]^n$ be a fractional coloring with at most a floating variables (not fixed yet). Then there exists an algorithm that with probability at least $\frac{1}{2}$, produces a fractional coloring y with at most $a/2$ floating variables and the discrepancy of any set increases by $O(\sqrt{a}\ln(2n/a))$.*

We will postpone theorem's 1.3 proof sketch for later. Let us first prove the main theorem (below) using theorem 1.3.

Theorem 1.4. *There exists an algorithm that finds a proper coloring $\chi : \Omega \to \{-1,+1\}$ with discrepancy $O(\sqrt{n})$. The algorithm succeeds with probability at least $\frac{1}{\log n}$.*

Proof. Starting with coloring $\chi = (0,0,...,0)$, we apply theorem 1.3 for $k = \log\log n$ steps and hence with probability at least $\frac{1}{\log n}$ we have a fractional coloring with at most $n/\log n$ floating variables and as in Spencer's proof with discrepancy $O(\sqrt{n})$. For the floating variables, we set $\chi(i) = 1$ with probability $\frac{1+y(i)}{2}$ and $\chi(i) = -1$ with probability $\frac{1-y(i)}{2}$ independently. By Chernoff bounds, it is easy to see that with high probability, the additional discrepancy will be $O(\sqrt{n})$ and the result follows. □

1.3.2 The Algorithm

Theorem 1.3 was proven by Bansal [3], who gave the following algorithm. $A(t)$ stands for the floating variables at step t.

Repeat the following subroutine $\log(2n)^3$ times (abort if the floating points are less than $a/2$ where $a \equiv \#$ of floating points at the beginning)
Bansal's routine
1. For each set A_i, let d_i be the discrepancy of A_i until now. 2. Let S_k be the set of k-dangerous sets, namely sets A_i s.t $d_i \in [b(k), b(k+1))$ with $b(k) = C\sqrt{a}\log(4n/a)(2-\frac{1}{k})$. 3. Find feasible solution u of SDP1 4. Construct c_t from u_i, by setting $c_t(i) = s\langle g, u_i\rangle$ (where $g(i) \sim \mathcal{N}(0,1)$) 5. Update $\chi_t = \chi_{t-1} + c_t$ (abort in case $\lvert\chi_t(i)\rvert > 1$ for some i 6. For each i, if $\chi_t(i) \geq 1 - \frac{1}{\log 2n}$ then set $\chi_t(i) = 1$ with probability $p = \frac{1+\chi_t(i)}{2}$ and -1 with $1-p$. 7. Update $A(t)$.

The SDP1 is the following:

$$1. \sum_{i\in[n]} ||u_i||_2^2 \geq A(t-1)/2$$

$$2. ||\sum_{i\in A_j} u_i||_2^2 \leq \frac{C' a\log(2n/a)}{(k+1)^5} \qquad \forall k, \forall A_j \in S_k$$

$$3. ||u_i||_2^2 \geq 1 \qquad \forall i \in A(t-1)$$

$$4. ||u_i||_2^2 = 0 \qquad \forall i \notin A(t-1)$$

1.3.3 Discussion

Let us first give some observations for SDP1. Clearly, the first constraint exists because we want at a current step, at least half of the floating points to become fixed and hence in a sense we want to make large progress. The constraint 4 is trivial since we want our fixed variables to continue to be fixed, thus we don't want to increase them. Constraint 3 is trivial too, because we want a tiny increase in the floating variables. Finally, constraint 2 is useful to certify that the k-dangerous sets (for some k) will not become bigger than discrepancy $O(\sqrt{n})$.

The idea of the algorithm, is to seperate the sets A_i in groups with respect to the discrepancy (k-dangerous) and find a vector that changes χ_t so as to maintain discrepancy of each set low. Additionally, using theorem 1.1, it can be proved that the number of k-dangerous sets decreases exponentially with respect to k with high probability w.r.t k (namely exponential). This fact implies that SDP1 is feasible. These arguments suffice to prove theorem 1.3 (the rest are just technical).

What is important to mention is the choice of some parameters in the algorithm. Even if Spencer's result is not constructive since he uses entropy's property *good chance*, it is clear that except of the partial coloring method, mostly Bansal's algorithm was influenced by Entropy method. Namely, the non-constructive entropy method is a major component for the algorithm above. The parameters $b(k)$ and $C\sqrt{a}\log(4n/a)(2-\frac{1}{k})$ which guide the semi-definite program were given by Entropy method.

Chapter 2

Lower Bounds, Eigenvalues and Other cases

2.1 Lower Bound

2.1.1 Definitions

In this section, we mention the necessary definitions we need to prove the lower bound of combinatorial discrepancy (proof follows in section 2.1.2).

Definition 2.1. Hadamard Matrix H is a square matrix with entries $-1, +1$ and with rows that are pairwise orthogonal. From the observation $HH^T = nI_n$, it follows that the columns of H are also othogonal and also it holds that

$$|det(H)| = n^{n/2} \tag{2.1}$$

It's an open question whether or not, there exists an Hadamard matrix with dimension $4k \times 4k$ for every $k \in \mathbb{N}^*$. By Sylvester's construction, for every $k \in \mathbb{N}$, there exists a Hadamard matrix with size $2^k \times 2^k$ (*). The construction follows from the recursion below

$$H_1 = [1]$$

$$H_{2^{k+1}} = \begin{bmatrix} H_{2^k} & H_{2^k} \\ H_{2^k} & -H_{2^k} \end{bmatrix} \tag{2.2}$$

It is clear that the first column of H_{2^k} has entries 1 and also that the number of 1 is equal to the number of -1 for all the other columns (because of the orthogonality).

2.1.2 The Proof

In this subsection we will construct a family of sets $\mathcal{A}$ such that $disc(\mathcal{A}) \geq \sqrt{n}/2$ using some Linear Algebra and also (*) from previous section. We can represent $\mathcal{A}$ as a square matrix M which we will call incident matrix, $M_{ij} = 1$ if element $j \in A_i$ and $M_{ij} = 0$ otherwise. Therefore to prove that Spencer's result is tight, it suffices to find a matrix M such that $\min_{\chi \in \{-1,+1\}^n} ||M\chi||_\infty$ is $\Omega(\sqrt{n})$. This is true because

$$disc(\mathcal{A}) = \min_{\chi \in \{-1,+1\}^n} ||M\chi||_\infty \tag{2.3}$$

(observe that a coloring χ can be seen as a $1 \times n$ vector with entries $-1, +1$). The following theorem certifies the tightness of Spencer's result.

Theorem 2.1. *For every $k \geq 2$, there exists a family $\mathcal{A}$ of sets $A_1, ..., A_n \subseteq \{1, 2, ..., n\}$ such that $disc(\mathcal{A}) \geq \sqrt{n}/2$, where $n = 2^k$*

Proof. Let H_n be a hadamard matrix as defined in section 2.1.1 and let $M = (H_n + J)/2$ where J is a square matrix with entries all ones. It is clear that each entry of M is either 0 or 1 (plays the role of an incident matrix as mentioned above). For an arbitrary $\chi \in \{-1, +1\}^n$ we get that

$$||H_n\chi||_2^2 = \sum_{i=1}^{n} \chi_i^2 ||H_{2^k}^i||_2^2 = n^2 \tag{2.4}$$

where H_n^i denotes the i-th column of the hadamard matrix H_n and $||u||_2$ denotes the euclidean norm of vector u.

Moreover, using the fact that the first column of H_n has entries 1 and also that the number of 1 is equal to the number of -1 for all the other columns (mentioned above), it holds that

$$|\sum_{i=1}^{n}(H_n\chi)_i| = n|(\chi)_1| = n \tag{2.5}$$

Finally, setting $s = \sum_{i=1}^{n} \chi_i$, observe that $J\chi = (s, s, ..., s)$, so

$$||(H_n + J)\chi||_2^2 = ||H_n\chi + (s, s, ..., s)||_2^2 \tag{2.6}$$

$$= ||H_n\chi||_2^2 + ns^2 + 2s\sum_{i=1}^{n}(H_n\chi)_i \tag{2.7}$$

$$\overbrace{=}^{2.4,2.5} n^2 + ns^2 \pm 2ns \tag{2.8}$$

Since $s \in Z$, (2.8) is minimized for $s = 0, \mp 2$ (observe that since n is even, s is even too, thus $sneq \pm 1$). Namely, we conclude that

$$||(H_n + J)\chi||_2^2 \geq n^2 \tag{2.9}$$

From Pingeonhole Principle and (2.9), it follows that there exists a coordinate j such that $|((H_n + J)\chi)_j| \geq \sqrt{n}$ and hence $||(H_n + J)\chi||_\infty \geq \sqrt{n}$ or equivalently $||M||_\infty \geq \sqrt{n}/2$.

Therefore, M is the matrix we are looking for and the proof finishes (as stated at the beginning of the section and 2.3) □

2.2 Eigenvalues

The proof above with the Hadamard matrices gives some interesting observations and tools to deal with combinatorial discrepancy. The idea is to consider the coloring χ, as a vector that belongs to $\{-1, +1\}^n$ and observe that $disc(\mathcal{A}) = \min_{\chi \in \{-1,+1\}^n} ||M\chi||_\infty$ where M is the incident matrix of $\mathcal{A}$. For purposes of completeness, we mention some important theorems that concern discrepancy and which are nice and powerful tools to deal with this problem and helps in obtaining lower bounds. In many cases and it can also be seen in the proof of the lower bound, it is more convenient to use Euclidean norm instead of infinity norm. Please notice that we use $||.||$ and $||.||_2$ for euclidean norm interchangeably. So using the same notation as before (M to be the incident matrix, etc), we may define $disc_2(\mathcal{A})$ as follows:

$$disc_2(\mathcal{A}) = \frac{1}{\sqrt{n}} \min_{\chi \in \{-1,+1\}^n} ||M\chi||_2 \tag{2.10}$$

It is straightforward that $disc(\mathcal{A}) \geq disc_2(\mathcal{A})$.

Theorem 2.2. *Let $\mathcal{A}$ be a family of sets $A_i \subset \Omega$ as previously and M be the incident matrix. It holds that:*

$$disc_2(\mathcal{A}) \geq \sqrt{\lambda_{min}} \tag{2.11}$$

where λ_{min} denotes the minimum eigenvalue of matrix $M^T M$.

In case the number of sets is n and $\Omega = \{1, ..., m\}$ then the inequality becomes $disc_2(\mathcal{A}) \geq \sqrt{\frac{m}{n}\lambda_{min}}$. Before continuing with the proof, we will give some intuition and discuss about theorem 2.2. First of all, matrix $M^T M$ is symmetric and hence all its eigenvalues are reals. Additionally, for a given vector $y \neq 0$ we have that $y^T M^T M y = (y^T M^T)^T M y = ||My||_2^2 \geq 0$ and hence $M^T M$ is a semidefinite matrix, or equivalently all its eigenvalues are nonnegative. Finally, it follows that $\min_{||\chi||=1} \frac{||\chi^T M^T M \chi||}{||\chi||} = \lambda_{min}$ (similar to the spectral norm).

Proof. Using the arguments mentioned above we get that

$$\begin{aligned}
\sqrt{\lambda_{min}} &= \min_{||\chi||=1} \sqrt{\frac{||\chi^T M^T M \chi||}{||\chi||}} \\
&= \min_{||\chi||=1} \frac{||M\chi||}{||\chi||} \\
&\overbrace{=}^{\text{scaling}} \min_{||\chi||=\sqrt{n}} \frac{||M\chi||}{||\chi||} \\
&\leq \min_{\chi \in \{-1,+1\}^n} \frac{||M\chi||}{||\chi||} \\
&= disc_2(\mathcal{A})
\end{aligned}$$

□

2.3 Beck-Fiala's theorem

2.3.1 Discussion

One of the most common and interesting settings of combinatorial discrepancy which has plenty of applications, is the case where each element of Ω is allowed to be used at most k times in $\mathcal{A}$ (bounded degree). In the section that follows, we will prove that if the degree is bounded from k, then the discrepancy is at most $2k-1$ using techniques from Linear Algebra. Bednarchak and Helm [5] improved the result to $2k-3$ (which is the best result in terms of the maximum degree k, namely considering that k is constant and also $k \ll n$). It was conjectured that discrepancy is of order $O(\sqrt{k})$ (still an open question). Finally, if the conjecture is true, it will be tight since it has been proven that discrepancy is $\Omega(\sqrt{k})$.

2.3.2 Proof

The proof we mention below, uses some Linear algebra without any randomness. However, it has some similarities (at least the beginning) with Bansal's.

Theorem 2.3. *Let $\mathcal{A}$ be a family of sets on an arbitrary finite set Ω such that $|\{A_i \in \mathcal{A} : x \in A_i\}| \leq t$ for all $x \in \Omega$. Then $disc(\mathcal{A}) \leq 2t-1$.*

Proof. The proof is algorithmic. Let $x_j \in [-1, 1]$ be a variable that corresponds to the value of element j. Initially, $x_j = 0$ for all j. A variable x_j will be called *floating* if $x_j \in (-1, 1)$ at an arbitrary step of the algorithm. Once $x_j = -1$ or $x_j = 1$, then it is fixed. A set A_i will be called *dangerous* at an arbitrary step if more than t elements of A_i are still floating. The algorithm finishes when there are no dangerous sets, and then we give arbitrary values (± 1) to the remaining floating variables. Finally, at each step, we assume that the discepancy of each dangerous set is equal to 0, namely

$$\sum_{j \in A_i} x_j = 0 \qquad \text{for all dangerous sets } A_i \tag{2.12}$$

(each element has a real value in $[-1, 1]$, at the beginning every set has discrepancy 0). The inductive step as described below, ensures that at each step of the algorithm, at least one floating variable gets a fixed value ± 1.

Inductive Step: The main idea which ensures that we can decrease the number of fixed variables at each step, maintaining 2.10, is the following claim:

Claim: At each step, the number of the floating variables is larger that the number of the dangerous sets.

Proof of claim: It comes from a simple double counting. Let l be the number of dangerous sets in an arbitrary step. Then the number of floating variables is at least $l \cdot (k+1)$, where each variable is counted at most k times, hence we have at least $\frac{(k+1) \cdot l}{k} > l$ floating variables.

Hence if we write for each dangerous set A_i, the constraint that the discrepancy is equal to 0, we get a system of l linear equations (where l is the number of dangerous sets) and more unknowns, so we can force a floating variable to aquire value ± 1, by adapting and all the other floating variables.

In the end, since each set is not dangerous, it has at most t floating variables, whose values belong to $(-1, 1)$. Finally, each floating variable takes an arbitrary value and hence it may increase the discrepancy of a set A_i which is currently 0 (at the end of the algorithm) by less than 2, hence for a total discrepancy of less than $2k$ for each set $A_i \in \mathcal{A}$. Thus $disc(\mathcal{A}) < 2k$, so $disc(\mathcal{A}) \leq 2k - 1$. □

2.4 2,3 - Permutation problem

We have already defined the modified version of combinatorial discrepancy called k-permutation problem, where the family of sets $\mathcal{A}$ contains sets A_{ij} where $A_{ij} = \{\pi_i(1), ..., \pi_i(j)\}$ for given permutations $p_1, ..., p_k$ of ground set $\Omega = \{1, 2, ..., n\}$. In this last section, we mention the cases $k = 2, 3$ (notice that for $k = 1$ trivially $disc(\mathcal{A}) = 1$ if we consider the coloring χ where $\chi(\pi(i)) = -\chi(\pi(i-1))$ and $\chi(\pi(1)) = 1$ for the given permutation π). Surprisingly for $k = 2$ we also have that discrepancy equals to 1. Beck conjectured that the discrepancy of the 3-permutation problem is $O(1)$ and Spencer offered 100$ for the resolution of this conjecture. However, on April 2011, Alantha Newman and Aleksandar Nikolov [9] gave a counterexample to this conjecture. They actually showed that discrepancy is $\Omega(\log n)$ for $k = 3$.

2.4.1 Case $k = 2$

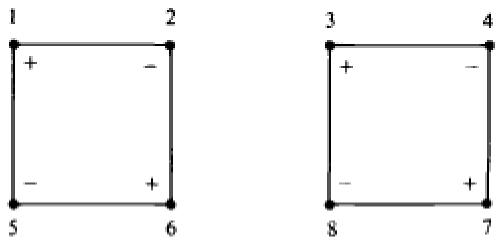
In this subsection, we present an elegant and short proof that for the case $k = 2$, discrepancy is at most 1 (see also [12]).

Theorem 2.4. *Let p_1, p_2 be two permutations of $\{1, ..., n\}$ and $A_{ij} = \{p_i(1), ..., p_i(j)\}$ for*

$i \in \{1, 2\}$ and $1 \leq j \leq n$. Then it holds that

$$disc(\mathcal{A}) \leq 1 \tag{2.13}$$

Proof. We may assume that p_1 maps every element to itself, so we have the permutations $1, 2, ..., n$ and $q(1), ..., q(n)$. We may also assume that n is even. In case n is odd, we add a dummy element and the argument still holds. Consider the graph G with $V(G) = \{1, ..., n\}$. For the edges, $(2i-1, 2i) \in E(G)$ and $q(2i-1), q(2i) \in E(G)$ for every $1 \leq i \leq n/2$. It is clear that each vertex has degree two, hence G is a union of disjoint cycles. Also observe that the cycles have even length (the edges of the cycles alternate). So what we do, is to color each cycle alternately $+1$ and -1. Clearly each permutation breaks into $+, -$ or $-, +$ pairs, thus no partial sum is more than one (see the figure below, q is permutation $3, 8, 2, 6, 1, 5, 4, 7$).



+	−	+	−	−	+	+	−
1	2	3	4	5	6	7	8
3	8	2	6	1	5	4	7
+	−	−	+	+	−	−	+

Figure 1

$\square$

2.4.2 Case $k = 3$

It was an open question for more than two decades whether or not discrepancy of 3-permutation problem is $O(1)$. Few months ago, Newman et al. [9] gave a recursive construction (family of sets $\mathcal{A}_k$ for $n = 3^k$ (k is a parameter) and they proved by induction the following theorem:

Theorem 2.5. $disc(\mathcal{A}_k) \geq \lceil \frac{k}{3} + 1 \rceil = \lceil \frac{\log_3 n}{3} + 1 \rceil$

The construction can be discribed with the following matrix:

$$\begin{matrix} A & B & C \\ C & A & B \\ B & C & A \end{matrix}$$

A corresponds to the numbers on the interval $[1, n/3]$, B corresponds to $[n/3+1, 2n/3]$ and C to the $[2n/3+1, n]$. Each element of the matrix, recursively has the form of the row that it appears, namely the A at the second row has the form $C\ A\ B$, where C corresponds to the numbers on the interval $[2n/9, 3n/9]$, B to the numbers on the interval $[n/9, 2n/9]$ and A to the numbers on the interval $[1, n/9]$ (etc). An example can be shown below for $n = 3^3$.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
27	25	26	21	19	20	24	22	23	9	7	8	3	1	2	6	4	5	18	16	17	12	10	11	15	13	14
14	15	13	17	18	16	11	12	10	23	24	22	26	27	25	20	21	19	5	6	4	8	9	7	2	3	1.

Figure 2

The main idea of the proof that is proven by induction is the following observation:
The sum of the discrepancies family of sets, each corresponding to one of the permutations, increases by 1 as k increases by 1, so by pigeonhole principle the discrepancy of at least one set is at least $\frac{k}{3} = \Theta(\log_3 n)$.

Bibliography

[1] N. Alon. Non-constructive proofs in combinatorics. *Proceedings of the International Congress on Mathematics*, 2:1421–1429, 1990.

[2] N. Alon and J.H Spencer. *The Probabilistic Method.* Wiley Series, 2008.

[3] N. Bansal. Constructive algorithms for discrepancy minimization. *FOCS*, pages 3–10, 2010.

[4] J. Beck and T. Fiala. Integer-making theorems. *Discrete Appl. Math.*, 3:1–8, 1981.

[5] D. Bednarchak and M. Helm. A note on the beck-fiala theorem. *Combinatorica*, 17:147–149, 1997.

[6] S. Boyd and L. Vandenberghe. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[7] J. Matoušek. An lp version of the beck-fiala conjecture. *European Journal of Combinatorics*, 19(2):175–182, 1998.

[8] J. Matoušek. *Geometric Discrepancy: An Illustrated Guide.* Springer, 1999.

[9] A. Newman and A. Nikolov. A counterexample to beck's conjecture on the discrepancy of three permutations. *arxiv*, 2011.

[10] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[11] J.H Spencer. Six standard deviations suffice. *Trans. Am. Math. Soc*, (289):679–706, 1985.

[12] J.H Spencer. *Ten Lectures on the Probabilistic Method.* Regional Conference Series in Applied Mathematics, 1994.

[13] A. Srinivasan. Improving the discrepancy bound for sparse matrices: better approximations for sparse lattice approximation problems. *In Proc. 8th ACM-SIAM Symposium on Discrete Algorithms*, pages 692–701, 1997.

[14] J.H Spencer. A. Srinivasan and P. Tetali. The discrepancy of permutation families. *arxiv.*

Appendices

Theorem .6. *(Markov's Inequality) Let X be a (discrete) random variable that takes positive values. Then for any number $a > 0$ we get that*

$$\Pr[X \geq a] \leq \frac{E[X]}{a}$$

Proof.

$$\begin{aligned} E[X] &= \sum_x x \Pr[X = x] \\ &\geq \sum_{x \geq t} x \Pr[X = x] \\ &\geq \sum_{x \geq t} t \Pr[X = x] \\ &\geq t \Pr[X \geq t] \end{aligned}$$

□

Theorem .7. *(Chernoff Bounds) Let X_i, $1 \leq i \leq n$ be mutually independent random variables with*

$$\Pr[X_i = 1] = \Pr[X_i = -1] = \frac{1}{2}$$

and let $S_n = X_1 + ... + X_n$. For every $a > 0$ we get that

$$\Pr[|S_n| > a] < 2e^{-a^2/2n}$$

Proof. Let $\lambda > 0$ then $E[e^{\lambda X_i}] = \frac{e^{\lambda}+e^{-\lambda}}{2}$. Since X_i are mutually independent it follows that

$$E[e^{S_n}] = \prod_{i=1}^{n} E[e^{X_i}] = \left(\frac{e^{\lambda} + e^{-\lambda}}{2}\right)^n < e^{n\lambda^2/2}$$

Therefore we get that

$$\begin{aligned} \Pr[S_n > a] &= \Pr[e^{\lambda S_n} > e^{\lambda a}] \\ &\overbrace{\leq}^{\text{Markov}} \frac{E[e^{\lambda S_n}]}{e^{\lambda a}} \\ &< e^{n\lambda^2/2 - \lambda a} \end{aligned}$$

Thus choosing $\lambda = \frac{a}{n}$ the result follows. □

Lemma .1. *(Useful Inequality) For every natural $k < n$ it holds that*

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

Proof. First of all $\binom{n}{k} = \frac{n(n-1)...(n-k+1)}{k!} < \frac{n^k}{k!}$. Additionally since $e^k = \sum_{j=0}^{\infty} \frac{k^j}{j!}$, we get that $e^k > \frac{k^k}{k!}$ and thus $\binom{n}{k} < \frac{n^k e^k}{k^k}$. □

Remarks on conditional expectation: Two properties that we used in this thesis are the following: Let X, Y be random variables and $g(x)$ real function. It holds that

$$E[E[X|Y]] = E[X]$$

and

$$E[Xg(Y)|Y] = g(Y)E[X|Y]$$