This is a summary of the talk I gave on July 26th of 2011 in University of Waterloo. The aim is to describe two of the three main results of the paper

> Constructive Algorithms for Discrepancy Minimization, by Nikhil Bansal (FOCS'10).

The results presented are indicated as (2) and (3) in the abstract of the paper.
Abbas Mehrabian

The Setup

$V = \{1, 2, \ldots, n\}$ is a set of *elements*, also called *points*.

$\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ is a family of m subsets of V.

The *maximum degree*, d, is the maximum number of times an element appears in sets.

A *complete colouring* is a function $\chi : V \to \{-1, +1\}$.

A *partial colouring* is a function $\chi : V \to [-1, +1]$.

The *discrepancy* of a set S is $\text{disc}(S) = |\sum_{i \in S} \chi(i)|$.

The *discrepancy* of a colouring χ is $\max\{\text{disc}(S) : S \in \mathcal{S}\}$.

The *discrepancy* of a set system $(V, \mathcal{S})$ is the minimum possible discrepancy of a complete colouring of it.

The *hereditary discrepancy* of a set system $(V, \mathcal{S})$, λ, is the maximum, over all $W \subseteq V$, of the discrepancy of the set system

$$(W, \{S \cap W : S \in \mathcal{S}\}).$$

A Martingale Tail Bound. Let $0 = X_0, X_1, \ldots, X_n$ be a martingale with increments $Y_i = X_i - X_{i-1}$. That is, for $1 \le i \le n$,

$$\mathbb{E}[Y_i | X_0, X_1, X_2, \ldots, X_{i-1}] = 0.$$

Suppose that for $1 \le i \le n$, $Y_i | (X_0, X_1, \ldots, X_{i-1})$ is distributed as $N(0, \kappa_i^2)$, where κ_i is determined by $(X_0, X_1, \ldots, X_{i-1})$ and has absolute value at most σ. Then for any $\alpha > 0$,

$$\mathbf{Pr}\left[|X_n| \ge \alpha\sigma\sqrt{n}\right] \le 2\exp(-\alpha^2/2).$$

Additive Property of Gaussians. Let $g \in \mathbb{R}^n$ be a random Gaussian vector, i.e. each coordinate is chosen independently according to distribution $N(0, 1)$. Then for any $v \in \mathbb{R}^n$, the random variable $< g, v >$ is distributed as $N(0, \| v \|_2^2)$.

Problem. Find upper bounds for discrepancy.

Results:

	existential	algorithmic	
random colouring	$O\left(\sqrt{n \log m}\right)$	$O\left(\sqrt{n \log m}\right)$	random colouring
Spencer'85	$O\left(\sqrt{n \log \frac{2m}{n}}\right)$	$O\left(\sqrt{n}\log \frac{2m}{n}\right)$	this paper
Banaszczyk'98	$O\left(\sqrt{d \log n}\right)$	$O\left(\sqrt{d} \log n\right)$	this paper
by definition	λ	$O\left(\lambda \log(mn)\right)$	this paper

Inapproximability of Discrepancy: Charikar, Newman and Nikolov proved that there are systems with $m = O(n)$ sets, such that no polynomial algorithm can distinguish whether the discrepancy is 0 or $\Omega(\sqrt{n})$, unless $P = NP$.

Theorem 1. There is a randomized polynomial algorithm that with probability at least $1/2n$ constructs a complete colouring with discrepancy $O(\lambda \log(mn))$.

Proof. The algorithm starts with a partial colouring, with all points coloured 0. The colour of each point starts from 0, and changes over time, and becomes -1 or 1 at some moment during the execution of the algorithm. When this happens, we say that the point is *dead* and its colour does not change. Otherwise, the point is *alive*. Let A_t denote the set of alive points at time t. We denote the colouring vector at time t by $x_t \in \mathbb{R}^n$, so $x_0 = (0, 0, \dots, 0)$ and $x_t(i)$ is the colour of point i at time t. In this language, point i is alive at time t if $x_t(i) \notin \{-1, 1\}$, otherwise it is dead.

The algorithm is given in the next page (taken from the original paper). $\square$

4.1 Algorithm

Initialize, $x_0(i) = 0$ for all $i \in [n]$. Let $s = 1/(4n(\log(mn))^{1/2})$. Let $\ell = 8 \log n/s^2$.
For each time step $t = 1, 2, \ldots, \ell$ repeat the following:

1. Find a feasible solution to the following semidefinite program:

$$||\sum_{i \in S_j} v_i||_2^2 \quad \leq \quad \lambda^2 \qquad \text{for each set } S_j \tag{4}$$

$$||v_i||_2^2 \quad = \quad 1 \qquad \forall i \in A(t-1) \tag{5}$$

$$||v_i||_2^2 \quad = \quad 0 \qquad \forall i \notin A(t-1) \tag{6}$$

   This SDP is feasible as setting $v_i \cdot v_j = \mathcal{X}(i)\mathcal{X}(j)$, where $\mathcal{X}$ is the minimum discrepancy coloring of the set system restricted to $A(t-1)$ is a valid solution. Let $v_i \in \mathbb{R}^n$, $i \in [n]$ denote some arbitrary feasible solution to the SDP above.

2. Construct $\gamma_t \in \mathbb{R}^n$ as follows: Let $g \in \mathbb{R}^n$ be obtained by choosing each coordinate $g(i)$ independently from the distribution $\mathcal{N}(0, 1)$. For each $i \in [n]$, let $\gamma_t(i) = s\langle g, v_i \rangle$.
   Update $x_t = x_{t-1} + \gamma_t$.
   If $|x_t(i)| > 1$, for any i, abort the algorithm.

3. For each i, set $x_t(i) = 1$ if $x_t(i) \geq 1 - 1/n$ or set $x_t(i) = -1$ if $x_t(i) < -1 + 1/n$.
   Update $A(t)$ accordingly.

Return the final coloring x_ℓ.

4.2 Analysis

We begin with some simple observations.

1. At each time step t, we have $||v_i||_2^2 = 1$ for each $i \in A(t-1)$ and $||v_i||_0^2 = 0$ for $i \notin A(t-1)$. Thus, by lemma 2.1, conditioned on $i \in A(t-1)$, we have $\gamma_t(i) \sim N(0, s^2)$ for $i \in A(t-1)$ and $\gamma_t(i) = 0$ otherwise. Similarly, conditioned on the evolution of the algorithm until $t-1$, the increment $\gamma_t(S_j)$ for S_j at time t is an unbiased Gaussian with variance at most $s^2\lambda^2$ (the precise value of the variance will depend on $v(S_j) = \sum_{i \in S_j : i \in A(t-1)} v_i$, which depends on the SDP solution at time t, which in turn depends on the evolution of the algorithm until time $t-1$, in particular on the set of alive variables $A(t-1)$).

2. The rounding in step 3 of the algorithm can effect the overall discrepancy by at most $n \cdot (1/n) = 1$, as each variable is rounded up or down at most once and is never modified thereafter. Note $\lambda \geq 1$, unless the set system is empty, so we will ignore the effect of this rounding step henceforth.

3. For the algorithm to abort in step 2 at time t, it is necessary that $\gamma_t(i) > 1/n = 4s(\log n)^{1/2}$, as step 3 ensures that $|x_{t-1}(i)| < 1 - 1/n$. Since $\gamma_t(i)$ is distributed as $N(0, s^2)$, this probability is at most $\exp(-8 \ln mn) = (mn)^{-8}$. Since there at most n variables and only $\ell = O(n^2 \log^2(mn))$ time steps, by union bound the probability that the algorithm ever aborts due to this step is at most $1/(mn)^4$.

The following key lemma shows that the number of alive variables halves in $O(1/s^2)$ steps with reasonable probability. The proof below follows a simpler presentation due to Joel Spencer.

Observations:

1. The SDP is feasible in all iterations (follows from the definition of λ).

2. For an active point i, the increment in $x(i)$ has distribution $N(0, s^2)$ (follows from additive property of Gaussians and constraint (5) of the SDP).

3. For a set S, the increment in $x(S)$, which is defined as

$$\gamma_t(S) = \sum_{i \in S} \gamma_t(i),$$

   has distribution $N(0, \kappa^2)$ for some $\kappa^2 \leq s^2 \lambda^2$ (follows from additive property of Gaussians and constraint (4) of the SDP).

4. The rounding in step 3 effects the overall discrepancy by at most $n \times 1/n = 1$, so we ignore it in the analysis.

5. The probability that the algorithm aborts in some iteration in step 2 is at most $(mn)^{-4}$: If the algorithm aborts in step 2, we must have

$$\gamma_t(i) > 1/n = 4s\sqrt{\log mn}.$$

   $\gamma_t(i)$ is distributed as $N(0, s^2)$ so the probability of this event is at most $\exp(-8\log(mn)) = (mn)^{-8}$. There are n variables and l iterations, so by the union bound the probability that the algorithm aborts in some iteration in step 2 is at most $(mn)^{-4}$.

Lemma. Let y be a partial colouring with at most k alive variables. After running the loop for $8/s^2$ times, with probability at least $3/4$, the number of alive variables is at most $k/2$.

Proof. Let $u = 8/s^2$, y_t be the colouring after t iterations, k_t be the number of alive variables after t iterations. Define

$$r_t = \begin{cases} \sum_i y_t(i)^2 & if\ k_{t-1} \geq k/2 \\ r_{t-1} + s^2 k/2 & otherwise. \end{cases}$$

Claim. For any y_{t-1}, $\mathbb{E}[r_t - r_{t-1} | y_{t-1}] \geq s^2 k/2$.

Proof of Claim. Clearly true for $k_{t-1} < k/2$. Otherwise

$$\begin{aligned} \mathbb{E}[r_t - r_{t-1}|y_{t-1}] &= \mathbb{E}[r_t|y_{t-1}] - r_{t-1} \\ &= \mathbb{E}\left[\sum_i (y_{t-1}(i) + \gamma_t(i))^2 | y_{t-1}\right] - \sum_i y_{t-1}(i)^2 \\ &= \sum_i \left(2y_{t-1}\mathbb{E}[\gamma_t(i)|y_{t-1}] + \mathbb{E}[\gamma_t(i)^2|y_{t-1}]\right) \geq s^2 k_{t-1} \geq s^2 k/2, \end{aligned}$$

because $\gamma_t(i)$ has mean 0 and variance s^2, and $k_{t-1} \geq k/2$. □

The claim implies $\mathbb{E}[r_u] \geq us^2 k/2$. For any t with $k_t \geq k/2$, we have $r_t = \sum_i y_t(i)^2 \leq k$. So $r_u \leq k + us^2 k/2$. Thus

$$us^2 k/2 \leq \mathbb{E}[r_u] \leq \mathbf{Pr}[k_u \geq k/2]k + (1 - \mathbf{Pr}[k_u \geq k/2])(k + us^2 k/2),$$

hence $\mathbf{Pr}[k_u \geq k/2] \leq \frac{k}{us^2 k/2} = 1/4$. □

Corollary. After $l = 8 \log n / s^2$ iterations, with probability at least $(3/4)^{\log n} \geq 1/n$, all variables are dead.

Proof of Theorem 1. By the Corollary, with probability at least $1/n$ when the algorithm finishes, x is a complete colouring. Let B denote the event that there is a set with discrepancy more than $2\sqrt{l \log mn}\lambda s$, and let B_j denote the event that set S_j has discrepancy more than this. Note that $x_t(S_j)$ forms a martingale, where each increment (conditional upon the history until time $t-1$) is a Gaussian with mean 0 and variance at most $\lambda^2 s^2$, so by the martingale tail bound we have

$$\mathbf{Pr}[B_j] = \mathbf{Pr}\left[|x_l(S_j)| \geq 2\sqrt{\log mn} \times \lambda s \sqrt{l}\right] \leq 2\exp(-2\log mn) = 2(mn)^{-2}.$$

Taking union bound over the m sets gives

$$\mathbf{Pr}[B] \leq \frac{2}{mn^2} \leq \frac{1}{2n}. \quad \square$$

Theorem 2. Assume that every element is contained in at most d sets. There is a randomized polynomial time algorithm that with probability at least $1/2n$, constructs a complete coloring with discrepancy $O(\sqrt{d}\log n)$.

Lemma[Srinivasan'97]. Assume that every element is contained in at most d sets. There is a partial colouring $\chi : V \to \{-1, 0, +1\}$ that assigns $\{-1, +1\}$ to at least half of the variables, whose discrepancy is at most $c\sqrt{d}$, for some absolute constant c.

The proof of the lemma is based on a Lemma by Beck'81.

Proof of Theorem 2. The algorithm is pretty much the same as the one for Theorem 1. However, the SDP is changed (see the next page, taken from the original paper). The feasibility of the SDP follows from Srinivasan's lemma. □

4.3 Constructive version of Srinivasan's result

We prove theorem 1.2. Let n denote the number of elements, and let m denote the number of sets. Since, each element lies in at most t sets, we can assume that $m \leq nt$. The algorithm is essentially identical to that in section 4. The only difference is that, at any step t in the algorithm, the entropy method, as applied in [19], only guarantees us a partial coloring (instead of a complete coloring) of the alive variables $A(t-1)$ with discrepancy $ct^{1/2}$. So we modify the first step of the algorithm above as follows:

Find a feasible solution to the following semidefinite program:

$$||\sum_{i \in S_j} v_i||_2^2 \leq c^2 t \qquad \text{for each set } S_j \tag{7}$$

$$\sum_{i \in A(t-1)} ||v_i||_2^2 \geq |A(t-1)|/2 \tag{8}$$

$$||v_i||_2^2 \leq 1 \qquad \forall i \in A(t-1) \tag{9}$$

$$||v_i||_2^2 = 0 \qquad \forall i \notin A(t-1) \tag{10}$$

The constant c is not stated explicitly in [19], but it can be calculated (in fact our algorithm can do a binary search on c do determine the smallest value c for which the SDP has a feasible solution). This program is feasible, as $v_i(1) = \mathcal{X}(i)$, where $\mathcal{X}$ is the partial coloring of $A(t-1)$ with discrepancy $ct^{1/2}$, is a feasible solution.

The analysis is essentially identical to that in section 5. As in lemma 4.1, during $16/s^2$ steps, the number of alive variables reduces by a factor of 2, with probability at least $1/2$ (note that we have $16/s^2$ steps above instead of $8/s^2$ steps in Lemma 4.1, because of the partial coloring instead of complete coloring of $A(t-1)$). Thus, there is a proper coloring with probability at least $1/n$ at end of $(16/s^2) \cdot \log n$ steps. The expected discrepancy of each set S in this coloring is at most $t^{1/2}(\log n)^{1/2}$. As there at most nt sets, arguing as at the end of section 4.2, conditioned on obtaining a proper coloring at the end, each set has discrepancy at most $O((t \log n)^{1/2}(\log(nt))^{1/2}) = O(t^{1/2} \log n)$.

5 Constructive version of Spencer's result

In this section we prove theorem 1.1. In fact, we will prove the more general guarantee for $O(n^{1/2}\log(2m/n))$ for set systems with n elements and m sets, where $m \geq n$.

To show this, we will design an algorithmic subroutine with the following property.

Theorem 5.1. *Let $x \in [-1,1]^n$ be some fractional coloring with at most a alive variables (i.e. i with $x(i) \notin \{-1,+1\}$). Then, there is an algorithm that with probability at least $1/2$, produces a fractional coloring $y \in [-1,1]^n$ with at most $a/2$ alive variables, and the discrepancy of any set increases by at most $O(a^{1/2}\log(2m/a))$.*

Given theorem 5.1, the main result follows easily.

Lemma 5.2. *The procedure in theorem 5.1 implies an algorithm to find a proper $\{-1,+1\}$ coloring with discrepancy $O(n^{1/2}\log(2m/n))$. Moreover, the algorithm succeeds with probability at least $1/(2\log m)$.*

Proof. We start with the coloring $x = (0,0,\ldots,0)$, and apply theorem 5.1 for $\ell = \log\log m$ steps. With probability at least $2^{-\ell} = 1/\log m$, this gives a fractional coloring y with at most $n/2^\ell = n/\log m$ alive