Constructive discrepancy minimization for convex sets

Thomas Rothvoß
Department of Mathematics
University of Washington
Seattle, USA
Email: rothvoss@uw.edu

***Abstract*—A classical theorem of Spencer shows that any set system with n sets and n elements admits a coloring of discrepancy $O(\sqrt{n})$. Recent exciting work of Bansal, Lovett and Meka shows that such colorings can be found in polynomial time. In fact, the Lovett-Meka algorithm finds a half integral point in any "large enough" polytope. However, their algorithm crucially relies on the facet structure and does not apply to general convex sets.**

We show that for any symmetric convex set K with Gaussian measure at least $e^{-n/500}$, the following algorithm finds a point $y \in K \cap [-1,1]^n$ with $\Omega(n)$ coordinates in ± 1: (1) take a random Gaussian vector x; (2) compute the point y in $K \cap [-1,1]^n$ that is closest to x. (3) return y.

This provides another truly constructive proof of Spencer's theorem and the first constructive proof of a Theorem of Gluskin and Giannopoulos.

***Keywords*-Discrepancy theory; combinatorics; convex optimization**

I. Introduction

Discrepancy theory deals with finding a *bi-coloring* $\chi : \{1,\dots,n\} \to \{\pm 1\}$ of a set system $S_1,\dots,S_m \subseteq \{1,\dots,n\}$ so that the worst inbalance $\max_{i=1,\dots,m} |\chi(S_i)|$ of a set is minimized, where we denote $\chi(S_i) := \sum_{j\in S_i} \chi(j)$. A seminal result of Spencer [1] says that there is always a coloring χ so that $|\chi(S_i)| \leq O(\sqrt{n})$ if $m = n$. The result is in particular interesting since it beats the random coloring which has discrepancy $\Theta(\sqrt{n \log n})$. Spencer's technique, which was first used by Beck in 1981 [2] is usually called the *partial coloring method* and is based on the argument that due to the pigeonhole principle many of the 2^n many colorings χ, χ' must satisfy $|\chi(S_i)-\chi'(S_i)| \leq O(\sqrt{n})$ for all sets S_i. Then one can take the *difference* between such a pair of colorings with $|\{j \mid \chi(j) \neq \chi'(j)\}| \geq \frac{n}{2}$ to obtain a *partial coloring* of low discrepancy. Iterating the argument $\log n$ times provides a full coloring.

Few years later and on the other side of the iron curtain, Gluskin [3] obtained the same result using convex geometry arguments. In a paraphrased form, Gluskin's result showed the following:

Theorem 1 (Gluskin [3], Giannopoulos [4]). *For a small constant $\delta > 0$, let $K \subseteq \mathbb{R}^n$ be a symmetric convex set with Gaussian measure $\gamma_n(K) \geq e^{-\delta n}$ and $v_1,\dots,v_m \in \mathbb{R}^n$ vectors of length $\|v_i\|_2 \leq \delta$. Then there are partial signs $y_1,\dots,y_m \in \{-1,0,1\}$ with $|supp(y)| \geq \frac{m}{2}$ so that $\sum_{i=1}^m y_i v_i \in 2K$.*

For the proof, consider all 2^m many translates $\sum_{i=1}^m y_i v_i + K$ with $y \in \{\pm 1\}^m$. Then one can estimate that the total measure of the translates must be much bigger than 1, so there must be many pairs $y', y'' \in \{\pm 1\}^m$ so that the translates overlap. Then take a pair that differs in at least half of the entries and $y := \frac{1}{2}(y' - y'')$ gives the vector that we are looking for. For more details, we refer to the very readable exposition of Giannopoulos [4].

In both, Spencer's original result and the convex geometry approach of Gluskin and Giannopoulos, the argument goes via the pigeonhole principle with exponentially many "pigeons" and "pigeonholes" which makes both type of proofs non-constructive. In a more recent breakthrough, Bansal [5] showed that a random walk, guided by the solution of an SDP can find the coloring for Spencer's Theorem in polynomial time. However, the approach needs a very careful choice of parameters and the feasibility of the SDP still relies on the non-constructive argument. A simpler and truly constructive approach was provided by Lovett and Meka [6] who showed that a "large enough" polytope of the form $P = \{x \in \mathbb{R}^n : |\langle v_i, x\rangle| \leq \lambda_i \ \forall i \in [m]\}$ has a point $y \in P \cap [-1,1]^n$ that can be found in polynomial time and satisfies $y_i \in \{-1,1\}^n$ for at least half of the coordinates. If the v_i's are scaled to unit length, then the "largeness" condition requires that

$$\sum_{i=1}^m e^{-\lambda_i^2/16} \leq \frac{n}{16}. \tag{1}$$

The approach of Lovett and Meka is surprisingly simple: start a random walk at the origin and each time you hit one of the constraints $\langle v_i, x\rangle = \pm\lambda_i$ or $x_i = \pm 1$, continue the random walk in the subspace of the tight constraint. The end point of this random walk is the desired point y.

Still, the algorithm of Lovett and Meka does not seem to generalize to arbitrary convex sets and the condition in (1) might not be satisfied for convex sets even if they have a large measure.

A. Related work

If we have a set system $S_1,\dots,S_m$ where each element lies in at most t sets, then the partial coloring technique

0272-5428/14 $31.00 © 2014 IEEE
DOI 10.1109/FOCS.2014.23

from above can be used to find a coloring of discrepancy $O(\sqrt{t} \cdot \log n)$ [7]. A linear programming approach of Beck and Fiala [8] shows that the discrepancy is bounded by $2t-1$, independent of the size of the set system. On the other hand, there is a non-constructive approach of Banaszczyk [9] that provides a bound of $O(\sqrt{t \log n})$ using a different type of convex geometry arguments. A conjecture of Beck and Fiala says that the correct bound should be $O(\sqrt{t})$. This bound can be achieved for the vector coloring version, see Nikolov [10].

More generally, the theorem of Banaszczyk [9] shows that for any convex set K with Gaussian measure at least $\frac{1}{2}$ and any set of vectors $v_1, \dots, v_m$ of length $\|v_i\|_2 \leq \frac{1}{5}$, there exist signs $\varepsilon_i \in \{\pm 1\}$ so that $\sum_{i=1}^m \varepsilon_i v_i \in K$.

A set of k permutations on n symbols induces a set system with kn sets given by the prefix intervals. One can use the partial coloring method to find a $O(\sqrt{k} \log n)$ discrepancy coloring [11], while a linear programming approach gives a $O(k \log n)$ discrepancy [12]. On the other hand, for $k = 3$, the discrepancy was recently shown to be $\Theta(\log n)$ [13] which disproved a conjecture of Beck. Also the recent proof of the Kadison-Singer conjecture by Marcus, Spielman and Srivastava [14] can be seen as a discrepancy result. They show that a set of vectors $v_1, \dots, v_m \in \mathbb{R}^n$ with $\sum_{i=1}^m v_i v_i^T = I$ can be partitioned into two halfs S_1, S_2 so that $\sum_{i \in S_j} v_i v_i^T \preceq (\frac{1}{2} + O(\sqrt{\varepsilon}))I$ for $j \in \{1, 2\}$ where $\varepsilon = \max_{i=1,\dots,m}\{\|v_i\|_2^2\}$ and I is the $n \times n$ identity matrix. Their method is based on interlacing polynomials and no polynomial time algorithm is known to find the desired partition.

For a very readable introduction into discrepancy theory, we recommend Chapter 4 in the book of Matoušek [15] or the book of Chazelle [16].

B. Our contribution

Our main contribution is the following:

Theorem 2. *There is a polynomial time algorithm, which for any symmetric convex set $K \subseteq \mathbb{R}^n$ with Gaussian measure at least $e^{-n/500}$ finds a point $y \in K \cap [-1, 1]^n$ with $y_i \in \{-1, 1\}$ for at least $\frac{n}{9000}$ many coordinates. Here it suffices if a polynomial time separation oracle for the set K exists.*

In fact, our method is extremely simple (see Figure 1 for a visualization):

Algorithm:

(1) take a random Gaussian vector $x^* \sim N^n(0, 1)$
(2) compute the point
$y^* = \text{argmin}\{\|x^* - y\|_2 \mid y \in K \cap [-1, 1]^n\}$
(3) return y^*

II. Preliminaries

In the following, we write $x \sim N(0, 1)$ if x is a *Gaussian random variable* with expectation $\mathbb{E}[x] = 0$ and variance

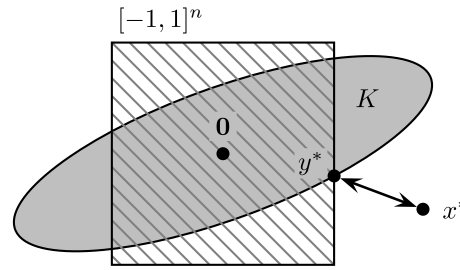

Figure 1. Visualization of the algorithm.

$\mathbb{E}[x^2] = 1$. By $N^n(0, 1)$ we denote the *n-dimensional Gauss distribution* and γ_n denotes the corresponding measure with density $\frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}$ for $x \in \mathbb{R}^n$. In other words, $\gamma_n(K) = \Pr_{x \sim N^n(0,1)}[x \in K]$ whenever K is a measurable set. In fact, all sets K that we deal with will be closed and convex and thus trivially measurable.

For a convex set K, let $d(x, K) := \min\{\|x - y\|_2 \mid y \in K\}$ be the *distance* of x to K and for $\delta \geq 0$, let $K_\delta := \{x \in \mathbb{R}^n \mid d(x, K) \leq \delta\}$ be the set of points that have at most distance δ to K (in particular $K \subseteq K_\delta$). A *half-space* is a set of the form $H := \{x \in \mathbb{R}^n \mid \langle v, x \rangle \leq \lambda\}$ for some $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. The key theorem on Gaussian measure that we need is the *Gaussian Isoperimetric inequality* (see e.g. [17] for a proof):

Theorem 3. *Let $K \subseteq \mathbb{R}^n$ be a measurable set and H be a halfspace so that $\gamma_n(K) = \gamma_n(H)$. Then for any $\delta \geq 0$, $\gamma_n(K_\delta) \geq \gamma_n(H_\delta)$.*

A simple consequence is that any set K that is not too small, is close to almost all the measure[1].

Lemma 4. *Let $\varepsilon > 0$. Then for any measurable set K with $\gamma_n(K) \geq e^{-\varepsilon n}$ one has $\gamma_n(K_{3\sqrt{\varepsilon n}}) \geq 1 - e^{-\varepsilon n}$.*

Proof: We assume that indeed $\gamma_n(K) = e^{-\varepsilon n} \leq \frac{1}{2}$. Choose $\lambda \in \mathbb{R}$ so that the halfspace $H = \{x \in \mathbb{R}^n \mid x_1 \leq \lambda\}$ has measure $\gamma_n(H) = \gamma_n(K)$ (note that $\lambda \leq 0$). First, we claim that $|\lambda| \leq \frac{3}{2}\sqrt{\varepsilon n}$. This follows from

$$\int_{-\infty}^{-\frac{3}{2}\sqrt{\varepsilon n}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \leq e^{-\frac{9}{8}\varepsilon n} \leq e^{-\varepsilon n}$$

using the estimate $\int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \leq e^{-t^2/2}$ for all $t \geq 0$. By symmetry, we get $\gamma_n(K_{3\sqrt{\varepsilon n}}) \geq 1 - e^{-\varepsilon n}$. ■

For a vector $v \in \mathbb{R}^n$ and $\lambda \geq 0$, the set $S = \{x \in \mathbb{R}^n : |\langle v, x \rangle| \leq \lambda\}$ is called a *strip*. If v is a unit vector, then

[1]Instead of using the Gaussian isoperimetric inequality, one can prove Lemma 4 also using the well-known *measure concentration* inequality for Gaussian space: given a 1-Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$ (i.e. $|F(x) - F(y)| \leq \|x - y\|_2$) one has $\Pr_{x \sim N^n(0,1)}[|F(x) - \mu| > \lambda] \leq 2e^{-\lambda^2/2}$ with $\mu = \mathbb{E}_{x \sim N^n(0,1)}[F(x)]$. One can then choose $F(x) := d(x, K)$ with $\lambda := \frac{3}{2}\sqrt{\varepsilon n}$ and one obtains $\Pr[|d(x, K) - \mu| > \frac{3}{2}\sqrt{\varepsilon n}] \leq 2e^{-\frac{9}{8}\varepsilon n} < e^{-\varepsilon n}$ for n large enough. Since $\gamma_n(K) \geq e^{-\varepsilon n}$, we know that $\mu \leq \frac{3}{2}\sqrt{\varepsilon n}$ and thus $\Pr[d(x, K) > 2 \cdot \frac{3}{2}\sqrt{\varepsilon n}] \leq e^{-\varepsilon n}$ as claimed.

the strip has width 2λ and $\gamma_n(S) = \Phi(\lambda)$ where we define $\Phi(\lambda) := \int_{-\lambda}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. Useful estimates are $\Phi(1) \geq e^{-1/2}$ and $\Phi(\lambda) \geq 1 - e^{-\lambda^2/2}$ for all $\lambda \geq 0$.

A convex body is called *symmetric* if $x \in K \Leftrightarrow -x \in K$. It is a convenient fact, that if we intersect a symmetric convex body with a strip, the measure decreases only slightly.

Lemma 5 (Šidák [18], Khatri [19])**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and $S \subseteq \mathbb{R}^n$ be a strip. Then $\gamma_n(K \cap S) \geq \gamma_n(K) \cdot \gamma_n(S)$.*

The still unproven *correlation conjecture* suggests that this claim is true for any pair K, S of symmetric convex sets. For more details on Gaussian measures, see the book of Ledoux and Talagrand [17].

For $0 \leq \varepsilon \leq 1$, let $H(\varepsilon) = \varepsilon \log_2(\frac{1}{\varepsilon}) + (1-\varepsilon)\log_2(\frac{1}{1-\varepsilon})$ be the *binary entropy function*. Recall that for $0 \leq \varepsilon \leq \frac{1}{2}$, the number of subsets $I \subseteq \{1, \ldots, n\}$ of size $|I| \leq \varepsilon n$ is bounded by $2^{H(\varepsilon)n}$. One can easily estimate that $2^{H(\varepsilon)} \leq e^{\frac{3}{2}\varepsilon \log_2(\frac{1}{\varepsilon})}$ which provides us with a bound for later.

A simple fact about convexity is that the optimum solution to a convex optimization problem does not change if we discard constraints that are not tight for the optimum.

Lemma 6. *Let $P, Q \subseteq \mathbb{R}^n$ be convex sets and let $g : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Suppose that x^* is an optimum solution to $\min\{g(x) \mid x \in P \cap Q\}$ and x^* lies in the interior of Q. Then x^* is also an optimum solution to $\min\{g(x) \mid x \in P\}$.*

Proof: Suppose for the sake of contradiction that there is a $y^* \in P$ with $g(y^*) < g(x^*)$, then some convex combination $(1-\lambda)y^* + \lambda x^*$ with $0 < \lambda < 1$ lies also in Q and has a better objective function than x^*, which is a contradiction. ■

III. Proof of the Main Theorem

Now we have everything to analyze the algorithm.

Theorem 7. *Let $0 < \varepsilon \leq \frac{1}{9000}$ be a constant and $\delta := \frac{3}{2}\varepsilon \log_2(\frac{1}{\varepsilon})$. Suppose that $K \subseteq \mathbb{R}^n$ is a symmetric, convex body with $\gamma_n(K) \geq e^{-\delta n}$. Choose a random Gaussian $x^* \sim N^n(0,1)$ and let y^* be the point in $K \cap [-1,1]^n$ that minimizes $\|x^* - y^*\|_2$. Then with probability $1 - e^{-\Omega(n)}$, y^* has at least εn many coordinates i with $y^*_i \in \{-1, 1\}$.*

Proof: First, we want to argue that x^* has at least a distance of $\Omega(\sqrt{n})$ to the hypercube. A simple calculation shows that

$$\Pr_{x \sim N^n(0,1)}[|x_i| \geq 2] = 2\int_2^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt > \frac{1}{25}.$$

Then with probability $1 - e^{-\Omega(n)}$ we have

$$d(x^*, [-1,1]^n) \geq \sqrt{\frac{n}{25} \cdot (2-1)^2} = \frac{1}{5} \cdot \sqrt{n}.$$

The crucial idea is that by the Gaussian isoperimetric inequality, x^* will not be far from any body that has a large enough Gaussian measure. The set $K \cap [-1,1]^n$ itself has only a tiny Gaussian measure, but we can instead consider the super-set $K(I^*) := K \cap \{x \in \mathbb{R}^n : |x_i| \leq 1 \ \forall i \in I^*\}$ where $I^* := \{i \in [n] \mid y^*_i \in \{\pm 1\}\}$ are the tight cube constraints for y^*. We claim that $d(x^*, K \cap [-1,1]^n) = d(x^*, K(I^*))$ since the distance is already defined by the tight constraints for y^*! More formally, this claim follows from an application of Lemma 6 with $P := K(I^*)$, $Q := \{x \in \mathbb{R}^n \mid |x_i| \leq 1 \ \forall i \notin I^*\}$ and $g(y) := \|x^* - y\|_2$ which is a strictly convex function.

Now, let us see what happens if $|I^*| \leq \varepsilon n$. We can apply the Lemma of Šidák and Khatri (Lemma 5) to lower bound the measure of $K(I^*)$ as

$$\begin{aligned} \gamma_n(K(I^*)) &\geq \gamma_n(K) \cdot \prod_{i \in I^*} \gamma_n(\{x \in \mathbb{R}^n : |x_i| \leq 1\}) \\ &\geq \gamma_n(K) \cdot e^{-|I^*|/2} \\ &\geq e^{-\delta n} \cdot e^{-(\varepsilon/2)n} \geq e^{-2\delta n} \end{aligned}$$

using that strips of width 2 have measure at least $e^{-1/2}$ and that $\varepsilon \leq \delta$. Now we know that the measure of $K(I^*)$ is not too small and hence almost all Gaussian measure is close to it. Formally we obtain $\gamma_n(K(I^*)_{3\sqrt{2\delta n}}) \geq 1 - e^{-2\delta n}$ by Lemma 4. It seems we are almost done since we derived that with high probability, a random Gaussian vector has a distance of at most $3\sqrt{2\delta n}$ to $K(I^*)$ and one can easily check that $3\sqrt{2\delta n} < \frac{1}{5}\sqrt{n}$ for all $\varepsilon \leq \frac{1}{9000}$. But we need to be a bit careful since I^* did depend on x^*. So, let us define $B := \bigcap_{|I| \leq \varepsilon n} (K(I)_{3\sqrt{2\delta n}})$. Observe that we have defined δ so that there are at most $e^{\delta n}$ many sets $I \subseteq [n]$ with $|I| \leq \varepsilon n$. Then by the union bound $\gamma_n(B) \geq 1 - e^{\delta n} \cdot e^{-2\delta n} \geq 1 - e^{-\delta n}$. Now we can conclude that with probability $1 - e^{-\Omega(n)}$, a random Gaussian will have distance at least $\frac{1}{5}\sqrt{n}$ to the hypercube while at the same time it has distance at most $3\sqrt{2\delta n} < \frac{1}{5}\sqrt{n}$ to all sets $K(I)$ with $|I| \leq \varepsilon n$. This shows that with high probability $|I^*| > \varepsilon n$. ■

We get the constants as claimed in Theorem 2 if we choose $\varepsilon = \frac{1}{9000}$ and observe that in this case $\delta \geq \frac{1}{500}$. We should spend few words on the computational aspects of our algorithm: the problem of finding the point $y^* \in K \cap [-1,1]^n$ that is closest to x^* is a convex optimization problem that can be solved in polynomial time using the Ellipsoid method [20]. Typically one has to be aware of numerical issues when dealing with arbitrary convex sets. But K as above is guaranteed to be full-dimensional and it must even contain a ball of radius $\Omega(e^{-\delta n})$, otherwise the measure could not be $e^{-\delta n}$. Moreover, we had a small slack in the above arguments, which means we can solve the optimization problem and find a y^* in a slightly scaled body $(1 - \Theta(\frac{1}{n}))K$; then even if we made a small numerical error, the resulting y^* would be in K. This concludes the proof of the main result, Theorem 2.

IV. Extension to intersection with subspaces

As already mentioned, our algorithm includes the result of Lovett and Meka in the following sense: Suppose our convex set is a polytope of the form $K = \{x \in \mathbb{R}^n : |\langle v_i, x\rangle| \leq \lambda_i \; \forall i \in [m]\}$ where all the v_i's are unit vectors and $\lambda_i \geq 1$. In this case, the strip $S = \{x \in \mathbb{R}^n : |\langle v_i, x\rangle| \leq \lambda_i\}$ of length $2\lambda_i$ has measure $\gamma_n(S) = \Phi(\lambda_i) \geq 1 - e^{-\lambda_i^2/2} \geq \exp(-2e^{-\lambda_i^2/2})$ using that $\lambda_i \geq 1$. By the Lemma of Šidák-Khatri this means that

$$\begin{aligned}\gamma_n(K) &\geq \prod_{i=1}^m \exp(-2e^{-\lambda_i^2/2}) \\ &= \exp\Big(-2\sum_{i=1}^m e^{-\lambda_i^2/2}\Big) \overset{!}{\geq} e^{-n/500}\end{aligned}$$

as long as $\sum_{i=1}^m e^{-\lambda_i^2/2} \leq \frac{n}{1000}$, exactly as in Lovett-Meka (apart from different constants). Please note that this line of arguments appeared already in the paper of Giannopoulos [4]. In the following we want to argue how $\Omega(n)$ many constraints with $\lambda_i = 0$ can be incorporated in the analysis.

For a subspace H we denote $N_H(0,1)$ as the $\dim(H)$-dimensional Gauss distribution restricted to the subspace H and we denote γ_H as the corresponding measure. For example one can generate a random $z \sim N_H(0,1)$ by selecting any orthonormal basis $u_1, \ldots, u_{\dim(H)}$ of H and letting $z = \sum_{i=1}^{\dim(H)} g_i u_i$ where $g_1, \ldots, g_{\dim(H)} \sim N(0,1)$ are independent 1-dim. Gaussians. Note that $\gamma_H(H) = 1$ and $\gamma_H(\mathbb{R}^n \backslash H) = 0$. We want to remind the reader that for any symmetric convex set K and any subspace H, one has $\gamma_H(K) \geq \gamma_n(K)$.

We want to argue that the following variation of our main claim still holds:

Theorem 8. *Fix $0 < \varepsilon \leq \frac{1}{60000}$ and $\delta := \frac{3}{2}\varepsilon \log_2(\frac{1}{\varepsilon})$. Let $K \subseteq \mathbb{R}^n$ be a symmetric, convex body with $K \subseteq H$ and $\gamma_H(K) \geq e^{-\delta n}$ where $H = \{x \in \mathbb{R}^n \mid \langle v_i, x\rangle = 0 \; \forall i \in [m]\}$ is a subspace defined by $m \leq 2\delta n$ equations. Choose a random Gaussian $x^* \sim N^n(0,1)$ and let y^* be the point in $K \cap [-1,1]^n$ that minimizes $\|x^* - y^*\|_2$. Then with probability $1 - e^{-\Omega(n)}$, y^* has at least εn many coordinates i with $y_i^* \in \{-1, 1\}$.*

Proof: Reinspecting the proof of Theorem 7, we see that it suffices to argue that most of the measure is still close to the sets $K(I)$. Formally, we will argue that for all $|I| \leq \varepsilon n$ one has $\gamma_n(K(I)_{7\sqrt{2\delta n}}) \geq 1 - 2e^{-2\delta n}$. Then $7\sqrt{2\delta n} < \frac{1}{5}\sqrt{n}$ for $\varepsilon \leq \frac{1}{60000}$ and the claim follows.

Hence, take a random point $x^* \sim N^n(0,1)$ and let $z^* \in H$ be the projection of x^* onto H (that means z^* is the point in H closest to x^*). We may assume w.l.o.g. that $v_1, \ldots, v_m$ are orthonormal. First, at least some part of the measure is close to H, since $\gamma_n(H_{\sqrt{2\delta n}}) \geq \gamma_n(\{x \in \mathbb{R}^n : |\langle v_i, x\rangle| \leq 1 \; \forall i \in [m]\}) \geq e^{-2\delta n}$ by Lemma 5. By Lemma 4 this implies that

$$\gamma_n(H_{4\sqrt{2\delta n}}) = \gamma_n((H_{\sqrt{2\delta n}})_{3\sqrt{2\delta n}}) \geq 1 - e^{-2\delta n}$$

and hence with the latter probability $\|x^* - z^*\|_2 \leq 4\sqrt{2\delta n}$.

In a second step, observe that we need to argue that z^* is close to $K(I)$. We know that $\gamma_H(K(I)) \geq \gamma_H(K) \cdot e^{-(\varepsilon/2)n} \geq e^{-2\delta n}$ as before. Since z^* is an orthogonal projection of a Gaussian, we know that $z^* \sim N_H(0,1)$ and we obtain that $d(z^*, K(I)) \leq 3\sqrt{2\delta n}$ with probability $1 - e^{-2\delta n}$. The claim then follows. ■

For being able to use the algorithm iteratively to find a full coloring, it is important that we admit centers that are not the origin. But this is very straightforward to obtain. In the following, for $x_0 \in \mathbb{R}^n$ and $K \subseteq \mathbb{R}^n$ we define $x_0 + K = \{x_0 + x : x \in K\}$ as the *translate* of K by x_0.

Lemma 9. *Let $\varepsilon \leq \frac{1}{60000}$ and $\delta := \frac{3}{2}\varepsilon \log_2(\frac{1}{\varepsilon})$. Given a subspace $H \subseteq \mathbb{R}^n$ of dimension at least $(1-\delta)n$, a symmetric convex set $K \subseteq H$ with $\gamma_H(K) \geq e^{-\delta n}$ and a point $x_0 \in [-1,1]^n$. There exists a polynomial time algorithm to find a point $y \in (x_0 + K) \cap [-1,1]^n$ so that at least $\frac{\varepsilon}{2}n$ many indices i have $y_i \in \{-1,1\}$.*

Proof: After translating by x_0 our goal is to find a $y \in K \cap \{x : -a_i \leq x_i \leq b_i \; \forall i\}$ with $a_i + b_i = 2$. Assume without loss of generality that $b_i \leq a_i$. Imagine that we take all the boundaries $x_i \geq -a_i$ and replace them with $x_i \geq -b_i$, which geometrically means that we push them closer to the origin. Then we aim to find a $y \in K \cap \{x : |x_i| \leq b_i \; \forall i\}$ with εn many coordinates satisfying $|y_i| = b_i$. Obviously we might have the problem that many coordinates i are tight at "fake" boundaries, i.e. $y_i = -b_i$. But either y or the mirrored point $-y$ will have $\frac{1}{2}\varepsilon n$ many coordinates i having a value of $+b_i$.

So it suffices to find points $y \in K \cap \{x : |x_i| \leq b_i \; \forall i\}$ with many coordinates i satisfying $|y_i| = b_i$. The easy solution is to rescale $K \cap \{x : |x_i| \leq b_i \; \forall i\}$ along the coordinate axes to $\tilde{K} \cap \{x : |x_i| \leq 1 \; \forall i\}$, which only increases the Gaussian measure (we will see formal arguments for $\gamma_H(\tilde{K}) \geq \gamma_H(K)$ in Cor. 13). Then we can apply Theorem 8 to find the desired vector y. ■

We want to briefly outline how one can iteratively apply Lemma 9 in order to find a full coloring (similar arguments can be found in [4]). Intuitively, whenever we induce on a subset of coordinates, the convex set needs to be still large enough. For a subset $J \subseteq [n]$ of indices, we call $U = \{x \in \mathbb{R}^n : x_i = 0 \; \forall i \in J\}$ an *axis-parallel subspace*.

Lemma 10. *Suppose that $K \subseteq \mathbb{R}^n$ is a symmetric convex body so that for all axis-parallel subspaces $U \subseteq \mathbb{R}^n$ one has that $\gamma_U(K) \geq e^{-\dim(U)/500}$. Then there is a polynomial time algorithm to compute a $y \in \{\pm 1\}^n \cap O(\log n) \cdot K$.*

Proof: In iteration $t = 1, \ldots, T$ we compute $y^{(t)} \in (y^{(t-1)} + K) \cap [-1,1]^n$ using Lemma 9 starting with $y^{(0)} :=$

$\mathbf{0}$ and ending with $y := y^{(T)}$, each time restricting to the subspace $U := \{x : x_i = 0 \text{ for } |y_i^{(t-1)}| < 1\}$. In each iteration a constant fraction of coordinates becomes integral and after $T = O(\log n)$ iterations we have $y^{(T)} \in \{\pm 1\}^n$. If $\|x\|_K := \min\{\lambda \geq 0 : x \in \lambda K\}$ denotes the *Minkowski norm* of x, then $\|y^{(t)} - y^{(t-1)}\|_K \leq 1$ and hence $\|y\|_K \leq T$. This settles the claim. ■

For Spencer's theorem it turns out that the $O(\log n)$-term can be replaced by $O(1)$ since the incurred discrepancy bounds decrease from iteration to iteration. A general way to state this is as follows:

Lemma 11. *Suppose that $K \subseteq \mathbb{R}^n$ is a symmetric convex body so that for all axis parallel subspaces $U \subseteq \mathbb{R}^n$ one has $\gamma_U((\frac{\dim(U)}{n})^\varepsilon K) \geq e^{-\dim(U)/500}$ for some constant $\varepsilon > 0$. Then one can compute a vector $y \in \{\pm 1\}^n \cap (c_\varepsilon K)$ in polynomial time.*

Proof: Now we can apply the procedure from Lemma 10 even with a body $\tilde{K} := (\frac{\dim(U)}{n})^\varepsilon \cdot K$ that shrinks over the course of the iterations. For some constant $0 < c < 1$ we have $\dim(U) \leq c^{t-1} \cdot n$ in iteration t, hence

$$\|y\|_K \leq \sum_{t=1}^{T} \|y^{(t)} - y^{(t-1)}\|_K \leq \sum_{t=1}^{\infty} \Big(\frac{c^{t-1} n}{n}\Big)^\varepsilon = \frac{1}{1 - c^\varepsilon}.$$

■

Let us illustrate how to apply Lemma 11 in Spencer's setting. Consider a set system $S_1, \ldots, S_n \subseteq [n]$ with n sets over n elements and define a convex body $K := \{x \in \mathbb{R}^n : |\sum_{j \in S_i} x_j| \leq 100\sqrt{n} \ \forall i \in [n]\}$. If at some point we have already all elements except of m many colored, then this means that we have a subspace U of dimension $\dim(U) = m$ left. For such a set system with m elements (but still $n \geq m$ sets), we can reduce the right hand side from $100\sqrt{n}$ to a value of $100\sqrt{m \cdot \log \frac{2n}{m}}$ and the Gaussian measure is still large enough. More formally, if we want $\gamma_U(\lambda \cdot K) \geq e^{-m/500}$, then a scalar of size $\lambda = 100\sqrt{m \cdot \log \frac{2n}{m}}/(100\sqrt{n}) \geq (\frac{m}{n})^{1/3}$ suffices. Then Lemma 11 finds a full coloring of discrepancy $O(\sqrt{n})$.

V. Extension to vector balancing

The attentive reader might have realized that we have essentially proven Giannopolous' Theorem only in the variant in which the vectors v_i correspond to the unit basis vectors. But we want to argue here that the algorithm from above can also handle Giannopoulos' general claim (apart from the fact that our partial signs x_i will be in $[-1,1]$ and not in $\{-1,0,1\}$).

For this sake, consider $Q = \{x \in \mathbb{R}^m \mid \sum_{i=1}^m x_i v_i \in K\}$. Then Q is again a symmetric convex set and all we need to do is to find a vector $y \in Q \cap [-1,1]^m$ that has $\Omega(m)$ many entries in ± 1. We know that it suffices to show that $\gamma_m(Q)$ is not too small — and this is what we are going to do now.

First, let us discuss how the Gaussian measure of a body can change if we scale it in some direction:

Lemma 12. *Let $K \in \mathbb{R}^n$ be symmetric and convex and for some $\lambda \geq 0$ define $Q := \{(x_1, x_2, \ldots, x_n) \mid (\lambda x_1, x_2, \ldots, x_n) \in K\}$. Then Q is symmetric and convex and $\gamma_n(Q) \geq \frac{1}{\max\{1,\lambda\}} \cdot \gamma_n(K)$.*

Proof: Define $f(x_1) := \Pr_{x_2,\ldots,x_n \sim N(0,1)}[x \in K]$. Note that f is a symmetric function and it is monotone in the sense that $0 \leq x_1 \leq y_1 \Rightarrow f(x_1) \geq f(y_1)$. Then we can express both measures as

$$\begin{aligned}\gamma_n(Q) &= 2\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdot f(\lambda x_1)\, dx_1 \\ &= 2\int_0^\infty \underbrace{\frac{1}{\sqrt{2\pi}\lambda} e^{-(x_1/\lambda)^2/2}}_{(*)} \cdot f(x_1)\, dx_1\end{aligned}$$

and

$$\gamma_n(K) = 2\int_0^\infty \underbrace{\frac{1}{\sqrt{2\pi}} e^{-x_1^2/2}}_{(**)} \cdot f(x_1)\, dx_1$$

For $\lambda \leq 1$, we see that $f(\lambda x_1) \geq f(x_1)$ and hence $\gamma_n(Q) \geq \gamma_n(K)$. For $\lambda \geq 1$, we can estimate that $\frac{(*)}{(**)} = \frac{1}{\lambda}\exp(\frac{1}{2}x_1^2(1 - \frac{1}{\lambda^2})) \geq \frac{1}{\lambda}$ and hence $\gamma_n(Q) \geq \frac{1}{\lambda}\gamma_n(K)$. ■

Since also the scaled set Q is symmetric, iteratively applying Lemma 12 gives:

Corollary 13. *Let $K \subseteq \mathbb{R}^n$ be symmetric and convex and $\lambda \in \mathbb{R}^n$. Then*

$$\begin{aligned}&\Pr_{x \sim N^n(0,1)}[(\lambda_1 x_1, \ldots, \lambda_n x_n) \in K] \\ \geq\ & \frac{1}{\prod_{i=1}^n \max\{1, |\lambda_i|\}} \Pr_{x \sim N_n(0,1)}[x \in K]\end{aligned}$$

Lemma 14. *Let $v_1, \ldots, v_m \in \mathbb{R}^n$ vectors with $\|v_i\|_2^2 \leq \varepsilon$ for $i = 1, \ldots, m$ and let $K \subseteq \mathbb{R}^n$ be a symmetric convex set. For $Q = \{x \in \mathbb{R}^m \mid \sum_{i=1}^m x_i v_i \in K\}$ one has $\gamma_m(Q) \geq \gamma_n(K) \cdot e^{-\varepsilon m}$.*

Proof: We consider the random vector $X = \sum_{i=1}^m x_i v_i$ with independent Gaussians $x_i \sim N(0,1)$. It is a well known fact in probability theory (see e.g. page 84 in [21]), that there is an orthonormal basis $b_1, \ldots, b_n \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$ so that one can write $X = \sum_{i=1}^n y_i u_i b_i$ with $y_1, \ldots, y_n \sim N(0,1)$ being independent Gaussians and the total variance of X is preserved, that means $\|u\|_2^2 = \sum_{i=1}^m \|v_i\|_2^2$. If we abbreviate $\Lambda := \prod_{i=1}^n \max\{1, |u_i|\}$, then we can apply Corollary 13 to lower bound

$$\begin{aligned}\gamma_m(Q) &= \Pr[X \in K] = \Pr_{y \sim N^n(0,1)}\Big[\sum_{i=1}^n y_i u_i b_i \in K\Big] \\ &\geq \frac{1}{\Lambda}\Pr_{y \sim N^n(0,1)}\Big[\sum_{i=1}^n y_i b_i \in K\Big] = \frac{1}{\Lambda}\gamma_n(K)\end{aligned}$$

using the rotational symmetry of γ_n. It remains to provide a (fairly crude) upper bound on Λ, which is

$$\begin{aligned}\Lambda &= \prod_{i=1}^{n} \max\{1,|u_i|\} \leq \prod_{i=1}^{n}(1+u_i^2) \\ &\stackrel{1+x\leq e^x}{\leq} \exp\Big(\sum_{i=1}^{n} u_i^2\Big) = \exp\Big(\sum_{i=1}^{m} \|v_i\|_2^2\Big) \leq e^{\varepsilon m}\end{aligned}$$

■

For example, if $\gamma_n(K) \geq e^{-m/1000}$ and $\|v_i\|_2^2 \leq \frac{1}{1000}$, then $\gamma_m(Q) \geq e^{-m/500}$ and we can apply Theorem 2 to obtain:

Theorem 15. *Given a symmetric convex set $K \subseteq \mathbb{R}^n$ with $\gamma_n(K) \geq e^{-m/1000}$ and vectors $v_1, \ldots, v_m \in \mathbb{R}^n$, with $\|v_i\|_2 \leq \frac{1}{40}$ for all $i = 1, \ldots, m$, there is a polynomial time algorithm to find a $y \in [-1,1]^m$ with $\sum_{i=1}^{m} v_i y_i \in K$ and at least $\frac{m}{9000}$ many indices i that have $y_i \in \{\pm 1\}$. Here it suffices to have access to a polynomial time separation oracle for K.*

Acknoweledgements: The author is very grateful to Daniel Dadush and to the anonymous referees for their helpful comments.

References

[1] J. Spencer, "Six standard deviations suffice," *Transactions of the American Mathematical Society*, vol. 289, no. 2, pp. 679–706, 1985.

[2] J. Beck, "Roth's estimate of the discrepancy of integer sequences is nearly sharp," *Combinatorica*, vol. 1, no. 4, pp. 319–325, 1981.

[3] E. D. Gluskin, "Extremal properties of orthogonal parallelepipeds and their applications to the geometry of banach spaces," *Mathematics of the USSR-Sbornik*, vol. 64, no. 1, p. 85, 1989.

[4] A. Giannopoulos, "On some vector balancing problems," *Studia Mathematica*, vol. 122, no. 3, pp. 225–234, 1997.

[5] N. Bansal, "Constructive algorithms for discrepancy minimization," in *FOCS*, pp. 3–10, 2010.

[6] S. Lovett and R. Meka, "Constructive discrepancy minimization by walking on the edges," in *FOCS*, pp. 61–67, 2012.

[7] A. Srinivasan, "Improving the discrepancy bound for sparse matrices: Better approximations for sparse lattice approximation problems," in *SODA'97*, (Philadelphia, PA), pp. 692–701, ACM SIGACT, SIAM, 1997.

[8] J. Beck and T. Fiala, ""Integer-making" theorems," *Discrete Appl. Math.*, vol. 3, no. 1, pp. 1–8, 1981.

[9] W. Banaszczyk, "Balancing vectors and Gaussian measures of n-dimensional convex bodies," *Random Structures Algorithms*, vol. 12, no. 4, pp. 351–360, 1998.

[10] A. Nikolov, "The Komlos Conjecture Holds for Vector Colorings," *ArXiv e-prints*, Jan. 2013.

[11] J. H. Spencer, A. Srinivasan, and P. Tetali, "The discrepancy of permutation families." Unpublished manuscript.

[12] G. Bohus, "On the discrepancy of 3 permutations," *Random Structures Algorithms*, vol. 1, no. 2, pp. 215–220, 1990.

[13] A. Newman, O. Neiman, and A. Nikolov, "Beck's three permutations conjecture: A counterexample and some consequences," in *FOCS*, pp. 253–262, 2012.

[14] A. Marcus, D. A. Spielman, and N. Srivastava, "Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem," *ArXiv e-prints*, June 2013.

[15] J. Matoušek, *Geometric discrepancy*, vol. 18 of *Algorithms and Combinatorics*. Berlin: Springer-Verlag, 1999. An illustrated guide.

[16] B. Chazelle, *The discrepancy method - randomness and complexity*. Cambridge University Press, 2001.

[17] M. Ledoux and M. Talagrand, *Probability in Banach spaces*. Classics in Mathematics, Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[18] Z. Šidák, "Rectangular confidence regions for the means of multivariate normal distributions," *J. Amer. Statist. Assoc.*, vol. 62, pp. 626–633, 1967.

[19] C. G. Khatri, "On certain inequalities for normal distributions and their applications to simultaneous confidence bounds," *Ann. Math. Statist.*, vol. 38, pp. 1853–1867, 1967.

[20] M. Grötschel, L. Lovász, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *Combinatorica*, vol. 1, no. 2, pp. 169–197, 1981.

[21] W. Feller, *An introduction to probability theory and its applications. Vol. II.* Second edition, John Wiley & Sons, Inc., New York-London-Sydney, 1971.