Mathematics of DS - Project report

Nischal Mainali

April 2018

1 Introduction

I worked on the so called 'Matrix Spencer Problem' for the final project.

This problem is a problem in Discrepancy Theory. These kind of problem arises when one tries to beat the union bound and give a tighter bound to discrepancies. A particularly famous example of this is what Spencer showed in his seminal "Six Standard Deviations Suffice" result (1985). The problem at hand, also named Matrix-Spencer is a generalization of the problem in which when the Matrices are diagonal, the problem corresponds to Spencer's Six Deviation Suffice.

The problem follows:

Prove or Disprove: There exists a universal constant C such that, for any choice of n symmetric matrices, $H_1, \dots H_n \in R^{nxn}$, satisfying $||H_k|| \leq 1$ for all K, there exists $\epsilon_1, \dots \epsilon_n \in [1, -1]$ s.t

$$\|\sum_{k=1}^{n} \epsilon_k H_k\| \leq C\sqrt{n}$$

I'll start with an exposition to the problem, and a particularly enlightening proof of Spencer Six Deviation Suffice using Convex Geometry, which is a more promising approach of extending the proof to the case of Matrices.

We can start by looking at a specific variant of the Matrix-Chernoff bound: Given n symmetric matrices, $H_1, \dots H_n \in R^{nxn}$,

$$Pr_{\epsilon \in [1,-1]} \left[\|\sum_{k=1}^{n} \epsilon_k H_k\| \geq t \|\sum_{k=1}^{n} \epsilon_k H_k^2\|^{1/2} \right] \leq 2ne^{-t^2/2}$$

Then, with high probability,

$$\left\|\sum_{k=1}^{n} \epsilon_k H_k\right\| = O(\sqrt{logn}) \left\|\sum_{k=1}^{n} \epsilon_k H_k^2\right\|^{1/2}$$

The Matrix-Spencer question then is whether we can improve in the $O(\sqrt{logn})$ by carefully choosing the ϵ's.

2 Literature Review

We will now look at an illuminating short Convex geometry proof for Spencer's six deviations suffice, and restate the Matrix-Spencer problem differently.
(This is heavily derived from Raghu Meka's blog post [1]: Discrepancy bounds from Convex Geometry)

Similarity between Spencer's and Gluskin's proof begins and end at the fact they start by relaxing the problem a bit and allowing $\epsilon \in [-1, 0, 1]$, but limiting the 0's and looking for $\Omega(n)$ non-zero coordinates, which we will call the support. We can then recurse on the set that had coordinate 0. So, proving the following lemma suffices:

Lemma: For vectors $A_1, \dots A_n \in R^{nxn}$, satisfying $\|A_k\| \leq 1$ for all K, there exists $\epsilon \in \{1, 0, -1\}^n$ s.t $\forall k, \langle A_k, \epsilon \rangle \leq C\sqrt{n}$ and $|support(\epsilon)| = \Omega(n)$.

Consider $K \subset R^n$ a symmetric convex set defined as follows for $\Delta = O(\sqrt{n})$:

$$K = \{x : |\langle A_k, x \rangle| \leq \Delta, \forall k\}$$

We start by using Sidak's lemma[2] for $g \sim N(0, 1)^n$:

$$Pr(g \in K) \geq \prod_{j=1}^{n} Pr[|\langle A_k, g \rangle| \leq \Delta] \geq \prod_{j=1}^{n} (1 - 2e^{-\Delta^2/2n})$$

Then, for sufficiently big $\Delta = O(\sqrt{n})$:

$$Pr(g \in K) \geq (3/4)^n$$

Now, let μ be a uniform distribution on [-1, 1]. And consider the corollary due to Kanter's Lemma[3]: $N(0, 1)^n \preceq \mu^n \Rightarrow \mu^n(K) \geq (3/4)^n$.
Then, for $K' = K \cap [-1, 1]^n$

$$Vol(K') \geq 2^n (3/4)^n = (3/2)^n$$

Then, for sufficiently small $\delta > 0$,

$$Vol((2 - \delta)K') \geq (2 - \delta)^n (3/2)^n = 2^n 2^{\Omega(n)}$$

Now, by Minkowski's theorem[4], $(2 - \delta)K'$ has at least $2^{\Omega(n)}$ lattice points.

But note that all the lattice points in $(2 - \delta)K'$ are elements of $\{-1, 0, 1\}^n$. So, there are $2^{\Omega(n)}$ from $\{-1, 0, 1\}^n$ in K.

Then, by counting argument, there is at least one point with $\Omega(n)$ non-zero points, meaning $|support(\epsilon)| = \Omega(n)$.

Similarly, the Matrix-Spencer can be restated as the following problem:

Prove or Disprove: There exists a universal constant C such that, for any choice of n symmetric matrices, $H_1, \dots H_n \in R^{nxn}$, satisfying $||H_k|| \leq 1$ for all K,

$$Pr_{g \sim N^n} \left[\|g_1 H_1 + g_2 H_2 + \dots + g_n H_n\| \leq C\sqrt{n} \right] \geq (3/4)^n$$

3 Attempts Numerics

Thinking about this problem was a wonderful journey albeit not very fruitful in terms of making any progress. I started out with Spencer's original proof and realized there was no obvious generalization to the case of Matrices. So, I started thiking about reducing the problem to a vector problem, perhaps by using the eigen value since that is waht we are controlling.

3.1 Weyl's Inequality

For the matrices s.t H = M + P, with eigen values $h_1 \dots h_n, m_1 \dots m_n, p_1 \dots p_n$, and for all $i = 1 \dots n$

$$m_i + p_n \leq h_i \leq m_i + p_1$$

So, I tried mapping each Matrix to eigen vectors and controlling the eigen vector the sum by iteratively using the weyl's inequality and using spencer's six deviation suffice on these vectors, since controlling eigen value means controlling the spectral norm. Problem: The bound becomes looser and looser with each iteration, and the only case where we get a meaningful bound is when all the eigen values are the same, which is the same as spencer's proof. So, this approach doesn't lead anywhere.

3.2 Added Restriction

Because I was belwildered and because SDP matrices have been particularly useful, I thought maybe adding further restriction of SDP in the matrices might be useful. It is aslo interesting because the truth for SDP case doesn't imply Spencer's Six-deviation suffice. Theoritically, I got nowhere. But I'll present SDP numerics later.

3.3 Generalized Gluskin's Proof

the most straightforward generalization of Gluskin's proof appeared to be regarding a Matrix M_{nxn} as a vector $M \in R^{nxn}$. But it appears to be that information on spectral norm is hardly preserved in this format and even if it is, I was at loss of extracting spectral norm from the vector representation. We do get the Forbenius norm from the vector which bound ths spectral norm, but it is a pretty useless bound in our context.

3.4 Numerics

The major concern of the numerics was to assume that the conjecture is true and find a lower bound for the Constant used in the conjecture. The following construction gives an immediate lower bound, but it was improved upon by the numeric. Consider:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} and \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then, we get an upper bound of 2 implying C can't be smaller than $\sqrt{2}$

The numerics lead to the following lower bound for C. So C must be greater than 1.63. The random matrix were constructed by means of random orthonormal matrix and diagonal matrix with 1 or -1 in the diagonal randomly to push closer to the limit.

n	$C \geq$
2	1.41
3	1.48
4	1.63
5	1.51
6	1.43
7	1.37

And in case of SDP matrix where the random matrix is constructed with orthonormal matrices and diagonal matrix with diagonal entries with 0 or 1.

n	bound
2	1
3	1.98
4	1.95
5	2.06
6	1.96
7	1.96
8	2.06
9	2.19

As one would imagine the bounds are smaller and appear to hover around a set value. Perhaps SDP condition will allow us to have a global bound independent of n, similar to Komlos conjecture?

All in all, I think I learned a great deal thinking about the problem, and know for sure that I completely enjoyed it.

References

[1] https://windowsontheory.org/2014/02/17/discrepancy-bounds-from-convex-geometry/

[2] Lemma 3 (Sidak's Lemma): Let $v_1, \dots, v_m \in R^n$ and let $g \sim \mathcal{N}(0,1)^n$ be a standard Gaussian vector. Then, for all $t_1, \dots, t_m \in R_+$,

$$\begin{aligned} &Pr\left[|\langle v_1, g\rangle| \le t_1 \ \wedge \ |\langle v_2, g\rangle| \le t_2 \ \wedge \ \cdots \ \wedge \ |\langle v_m, g\rangle| \le t_m\right] \ge \\ &\qquad Pr\left[|\langle v_1, g\rangle| \le t_1\right] \cdot Pr\left[|\langle v_2, g\rangle| \le t_2\right] \cdots Pr\left[|\langle v_m, g\rangle| \le t_m\right]. \end{aligned}$$

[3] Lemma 4 (Kanter's lemma): Let p, q be two symmetric distributions on R^n such that $p \preceq q$ and let μ be a unimodal distribution on R^m. Then, the product distributions $p \times \mu$, $q \times \mu$ on $R^{n \times m}$, satisfy $p \times \mu \preceq q \times \mu$.

[4] Theorem 6 (Minkowski's Theorem): Let $C \subseteq R^n$ be a symmetric convex set of Lebesgue volume more than $2^n \cdot \ell$ for an integer $\ell \ge 1$. Then, C contains at least ℓ points from the integer lattice Z^n.