Exploring Komlós Conjecture

V. Mullachery

May 1, 2018

Abstract

Komlós conjecture in discrepancy theory states that for some K and for any $n \times n$ matrix whose columns lie in the unit ball there exists a vector $\mathbf{x} \in \{-1, 1\}^n$ such that $\|\mathbf{Ax}\|_\infty \leq K$. In this work we explore the problem to arrive at an intuition and a few numerical and analytical attempts at arriving at an approximation to this upper bound.

1 Introduction

Our work here attempts to build an intuition and narrates the trail of struggle with Komlós conjecture. We also explore various techniques numerical and analytical, to find an upper bound. The progression of these techniques coincides our exploration of this topic and our own learning trajectory, and in itself may be entertaining and perhaps educational. The only redeeming aspect of our attempt is our sincerity of intent and our honesty in laying bare our failures and flaws.

In his famous *six standard deviations* theorem, Spencer [Spe94] gives an upper bound $\leq 6\sqrt{n}$, to a closely related problem. Towards the end of his paper, Spencer notes that the actual constant is 5.32.

Our main empirical results are in section 6. In sections 7 and 9 we advance a few ideas that are closely related and appear promising in analyzing this conjecture.

2 Definitions

A few definitions might make the discussion simpler and less repetitive to the reader and easier for the author(s).

$$A = \text{Matrix in } \mathbb{R}^{n \times n}, \text{ where each } \|A_i\| \leq 1$$
$$\mathbf{x} = \text{Vector of } \pm 1 \text{ assignments to each } A_i$$
$$\|A\mathbf{x}\|_\infty = \text{Infinity norm or maximum valued coordinate, of } A\mathbf{x}$$
$$K(n) = \text{Komlós constant for Matrices up to size } n \times n, \text{ dependent on } n$$
$$K = \text{Universal Komlós constant (conjectured)}$$
$$\begin{bmatrix} cos\theta \\ sin\theta \end{bmatrix} \text{ polar coordinates of a point } x, y \text{ on a unit circle in } \mathbb{R}^2$$
$$\begin{bmatrix} cos\phi \\ sin\phi * cos\theta \\ sin\phi * sin\theta \end{bmatrix} \text{ spherical coordinates } z, x, y \text{ of a point on the sphere in } \mathbb{R}^3$$
$$A^T A = \text{the symmetric covariance matrix of } A$$
$$A_i = \text{the } i\text{th vector of } A$$

3 Intuition

We could assume that the matrix A is a collection of n vectors $\in \mathbb{R}^n$. Now we assign ± 1 to each of the vectors, which is the same as $A\mathbf{x}$, where $\mathbf{x} \in \pm 1^n$. Which can be written as $\sum_{i=1}^n A_i x_i$. Now notice that if $A_i = \mathbf{e}_i$, then we have unit orthogonal vectors in $\mathbb{R}^n$. Each one of these would contribute a unit length in each of the orthogonal directions and there will be no coinciding coordinate. In such a case, $A\mathbf{x}$ would yield a 1 along each coordinate, i.e. $\begin{bmatrix} \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix}$. That is the $\|A\mathbf{x}\|_\infty = 1$. Each vector of A had a non-competing contribution in one coordinate direction. That was easy to see.

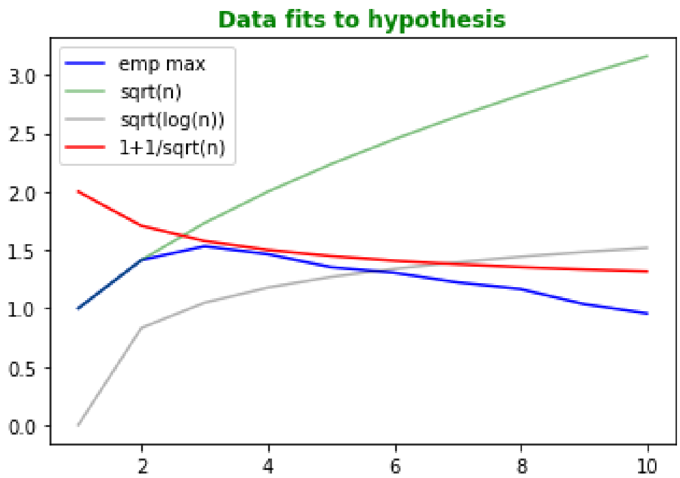
If the constituents of A are not the unit vectors, but vectors of unit L2 norm (i.e. $\|A_i\|_2 = 1$, then what we notice is that the assignment of ± 1 to vector A_i causes it's non-zero components along a coordinate direction to influence the outcome of $A\mathbf{x}$. Further if we used a different $\mathbf{x} \in \pm 1^n$ then the outcome changes. The number of choices of $\mathbf{x}$ is exponential $(= 2^n)$ in n. Very quickly one sees the combinatorial nature of the problem. The choices of A_i are in the continuous space and is infinite (only considering the $\|A\| = 1$, all points on a unit ball in $\mathbb{R}^n$). Now if we also consider $\|A_i\|_2 < 1$ (as in the conjecture), the space of A becomes even larger - it is the entire unit solid ball in $\mathbb{R}^n$.

The next item that we notice is that the conjecture states that there exists an upper bound on the minimum infinity norm given a free choice of $\mathbf{x} \in \pm 1^n$. That means the minimum infinity norm (which is the minimum of the maximum accrued value on any of the coordinates) cannot exceed a certain universal constant. If one were to consider a particular assignment of $\mathbf{x}$ that gives rise to a large infinity norm, then there exists a rearrangement of ± 1 among the elements of $\mathbf{x}$ such that the infinity norm is reduced below $K(n)$ and the as yet unknown (unproven) K.

In order to find or approximate $K(n)$ for a particular n it is necessary to compute $\|A\mathbf{x}\|_\infty$ for all $\mathbf{x}$ for a particular A. The lowest infinity norm for a particular A is now a possible candidate at an approximation to $K(n)$. And then repeat this process for *smart* choices of A. Now the essential question becomes what are some *smart* choices of A in a particular dimension n. We require that A yield high values for infinity norm, even in the worst possible assignment of $\pm 1^n$. What properties would such a A need to have? One could see this as the central question that needs to be answered. Now, these are some foundational intuitions. Equipped with these, we now run a few numerical experiments.

4 A first numerical approach

What would happen if we generated n random unit vectors $A_i \in \mathbb{R}^n$ of unit norm, and then exhaustively computed the max norm against all possible assignment vectors $\mathbf{x} \in \{+1, -1\}^n$? This experiment yielded a result that appeared promising, as witnessed by the plot below. We hypothesized that the upper bound $K(n)$ should reduce with n, since in higher dimensions any two random column vectors that compose A are nearly orthogonal and that could mean the upper limit should vary as $1 + 1/\sqrt{n}$. Our numerical experiments had a tendency to follow our deep wishful thinking that our hypothesis was correct.



What is wrong with this picture?

– The blue curve plots the empirical (observed) minimum infinity norm

– Since every $\mathbb{R}^n$ contains all other $\mathbb{R}^k$, $k < n$, the discrepancy should not reduce as dimensionality increases; it should be non-decreasing at the least

Why did this happen?

– For one, we proved that we were no better than a common bat, or rat for that matter

– As the dimensionality increases, the number of $\mathbf{x}$ increases exponentially, 2^n. So, given a fixed amount of computation, we ended up taking fewer observations (samples) of A.

– And interestingly, the higher the dimension, the lower the chances of sampling a large discrepancy causing A

– So, in expectation the chances of a discrepancy are lower, the larger the dimension

What do we learn?

– This problem will need very careful construction of A

– This will mean a very careful study of the properties of such a high discrepancy A

– Net utility to the world by our ruminations on this problem $- > 0 \in \mathbb{R}^n$

Well, for one, as the dimensions increase, the random vectors that we generate are indeed more likely to be orthogonal to each other and so their contributions are unlikely to be of the coordinate mixture that could yield a large discrepancy. However, the plot appears promising to someone studying expected deviations and not to our study here of minimum infinity norm.

5 Analytical Inspection

5.1 $\mathbb{R}^1$

Trivially, one can see that the discrepancy in 1-D case is 1, simply use $A = [1]$

5.2 $\mathbb{R}^2$ and Hadamard Matrix

Here we take an algorithmic and analytic approach in attempting to construct a large discrepancy yielding matrix. Two unit norm vectors $\in \mathbb{R}^2$ can be represented in polar coordinates as:

$$\begin{bmatrix} \cos\theta & \cos\alpha \\ \sin\theta & \sin\alpha \end{bmatrix} \tag{1}$$

In order for them to yield the max norm on either of the coordinates independent of their assignments to $\{+1, -1\}$, they must add up to equal maximum values $\cos\theta + \cos\alpha$ or $\sin\theta - \sin\alpha$. The other cases where the signs could be reversed are symmetric to this argument. Further, for those values to be a maximum, their first order partial differentials with respect to θ and α must be zero. So,

$$\frac{\partial\alpha}{\partial\theta} = -\frac{\sin\theta}{\sin\alpha}$$
$$\frac{\partial\theta}{\partial\alpha} = \frac{\cos\alpha}{\cos\theta}$$

Combining these with the earlier statements (of equal sized max values):

$$\cos\theta + \cos\alpha = \sin\theta - \sin\alpha$$

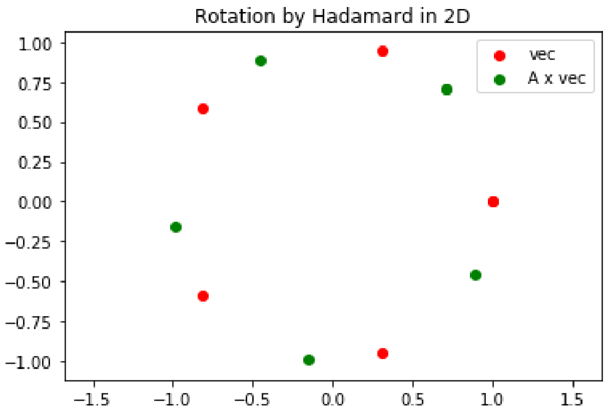
$$\sin(\theta + \alpha) = 0$$
$$\cos^2\theta = 1/2$$

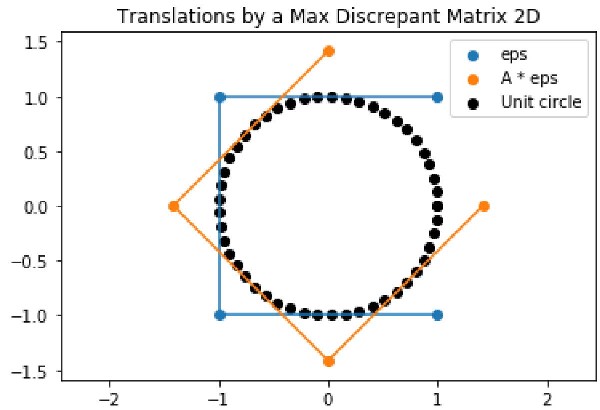
Thus, we arrive at $\theta = -\alpha = \pi/4$. That means the best possible discrepancy value in $\mathbb{R}^2$ is no larger than $\sqrt{2} = 1/\sqrt{2} + 1/\sqrt{2}$. So, an example of minimum infinity norm accruing vectors $\in \mathbb{R}^2$ are then:

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \tag{2}$$

This should immediately remind the reader of a normalized Hadamard matrix. In $\mathbb{R}^2$, a normalized Hadamard matrix is the most discrepant, and $K(2) = \sqrt{2}$

As can be seen in the figures, the most discrepant matrix has one role to perform and that is to translate (rotate) the vertexes of a centered unit cube onto the coordinate axes. This is so because the corners of the unit cube correspond to possible choices of $\mathbf{x}$. And the infinity norm is the maximal value along any coordinate:





5.3 $\mathbb{R}^3$

In $\mathbb{R}^3$, using spherical coordinates (as above) and 3 vectors, one should have 6 variables. However, here things break down, since there does not appear to be a principled method of symmetrical assignment of $\pm 1^3$ to 3 vectors, and so an *equal* sum hypothesis becomes untenable. So, we now switch over to numerical simulations and outcomes. And since we are already past page 3, and we haven't said anything worthwhile and we have lost all our readers except perhaps the most tortured souls, we might as well continue on, at the cost of assignment grades, and academic reputation. But then, if grades and reputation were our motivations to begin with, we might as well not have commenced on this journey at all.

6 Numerical Simulations

We undertook repeated simulations to push the discrepancy numbers higher. Though we did perform simulations for dimensions up to 13, the outcomes were poor for higher dimensions. This is so because we exhaust computational and memory resources as n grows larger. So, we focused on R^3 and R^4, with occasional detours to R^5:

Algorithm 1 $K(n)$ estimation algorithm

Input : $n > 0$, r (number of repetitions)
Generate $A = n$ normal random vectors $\in \mathbb{R}^n$
Normalize columns of A to be unit vectors, $\|.\|_2 = 1$
Generate $S =$ set of all permutation of $\epsilon \in \pm 1^n$
for $t = 1$ to r **do**
  Compute $g_t = \textbf{MinInfNorm}(A_t)$
end for
Return $\max g_{1:t}$

Function **MinInfNorm**(B)
  for $p = 1$ to $|S|$ **do**
    Compute $x_p = \|B\epsilon_p\|_\infty$
  end for
  Return $\min x_p$
EndFunction

For a given n, we notice that the results slowly grow larger as we increase r, the number of repetitions, as expected, and eventually should clue us on $K(n)$

n	$K(n)$ estimation observations	Repetitions
3	1.5559	10 million
4	1.6206	400 million
5	1.4799	12 million
6	1.3885	12 million
7	n/a	n/a
8	1.1649	120 thousand
9	1.0828	120 thousand
10	0.9564	1200
11	0.9850	1200

$K(n)$ estimations decrease with n because the search space grows exponentially with n and each successive row searches an exponentially smaller region. Notwithstanding that, these simulations are only marginally satisfying, since they do not provide us any intuition about the dynamics involved.

7 A few interesting ideas

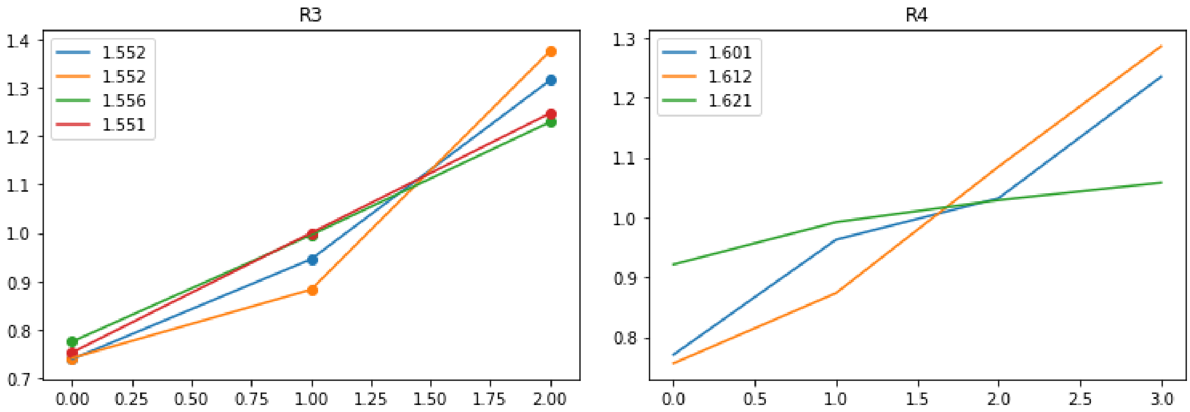
In this section we enumerate a few lines of attack that appeared appealing to us.

7.1 Gradient Free Optimization techniques: Nelder Mead

An idea that was suggested to us was to use gradient free optimization techniques. However, this yielded $K(n) = 1$ for $\mathbb{R}^3$, and summarily appeared unpromising for the discovery of $K(n)$ in an entire $\mathbb{R}^n$ space. From the outward appearance of it, these techniques could be helpful once we have identified a region of interest through initial random sampling. We surmise that these techniques could assist our future numerical explorations, to perform local searches.

7.2 Covariance Matrix and Spectral Range

An interesting line of pursuit is to compare the spectral values of the empirically most discrepant matrix to some that weren't. As a proxy, we could consider the symmetric equivalent, the covariance matrix, $P = A^T A$. These matrices have a 1 along the diagonal. In the dimensions that admit a Hadamard matrix (typically powers of 2), we get $P = A^T A = \frac{1}{\sqrt{n}} H^T \frac{1}{\sqrt{n}} H = I_n$. Thus in these dimensions, P will have all unit spectral value, and a range of 0. There appears to be a tendency for the spectral ranges of P in $\mathbb{R}^2$, and $\mathbb{R}^4$ tend to span a small region. However, not so with other dimensions:

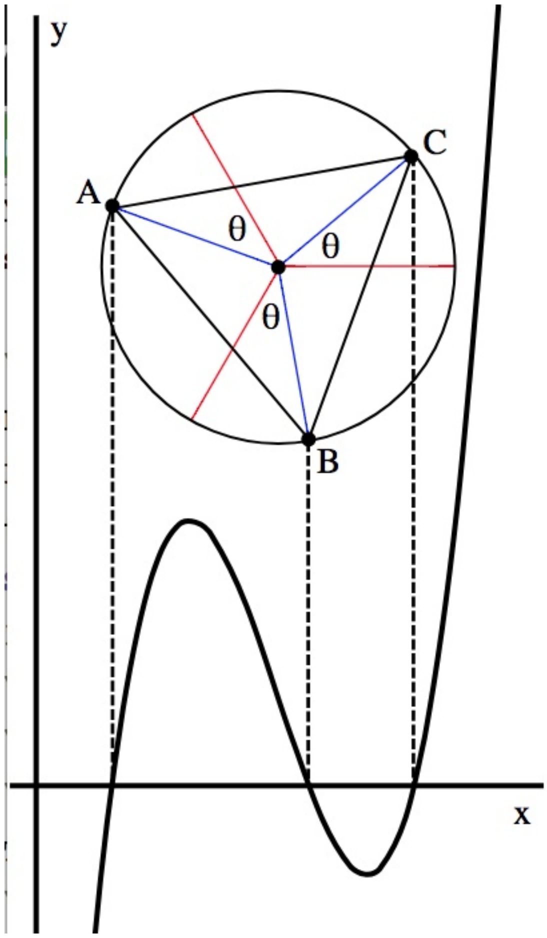


n	Eigenvalues (max, min)	Spectral range
2	1, 1	0
3	0.77553741, 1.2288312	0.453
4	0.92147169, 1.05785574	0.136

In general the covariance matrix will be: $\begin{bmatrix} 1 & A_1^T A_2 & \cdots & A_1^T A_n \\ A_2^T A_1 & 1 & \cdots & A_2^T A_n \\ A_n^T A_1 & A_n^T A_2 & \cdots & 1 \end{bmatrix}$. Notice that this matrix is described by $n(n-1)/2$ scalar values $\in [-1, 1]$. So, this prompts a faster exploration technique, which is to sample fewer scalar values randomly as opposed to $n \times n$ matrix. This should yield us a computational speed up for numerical simulations.

7.3 Covariance Matrix and Characteristic Polynomial

If we attempted to solve the characteristic polynomial for the covariance matrix in the $\mathbb{R}^3$ case, we get: $\begin{bmatrix} 1-\lambda & \alpha & \beta \\ \alpha & 1-\lambda & \gamma \\ \beta & \gamma & 1-\lambda \end{bmatrix}$. The characteristic polynomial is given by : $(1-\lambda)^3 - (\alpha^2+\beta^2)(1-\lambda) + 2\alpha\beta\gamma - \gamma^2$. The roots of this depressed form cubic polynomial are given by an expression: $t_k = 2v\cos(\frac{w}{3} - \frac{2\pi k}{3})$, for $k = 0, 1, 2$, where $v = 2\sqrt{\frac{\alpha^2+\beta^2}{3}}$, $w = \arccos\frac{3(\gamma^2-\alpha\beta\gamma)}{2(\alpha^2+\beta^2)}\sqrt{\frac{3}{\alpha^2+\beta^2}}$. These roots of the polynomial yield the $t_k = 1 - \lambda_k$ values. In a geometric sense they form vertexes of an equilateral triangle as shown here:



We surmise that such an interpretation and comparison of the best discrepant covariance matrix with the others could uncover properties of interest. For instance, what would happen to the discrepancy number as we rotate this triangle (that is specify different eigenvalues) - is there a set of geometries that correspond to maximum discrepancy.

8 Conclusion

We attempted numerous numerical simulations and witness that the discrepancy numbers that we observe are of the range 1.62, much smaller than the tightest analytical proofs indicate. We believe that this should be an area of great interest to practitioners at the intersection of numerous fields: Statistics, Machine Learning, Applied Mathematics etc.

9 Future Work

We would like to advance these areas for future work in analyzing Komlós conjecture:

1. Analysis of the spectral range of $A^T A$ and the discrepancy value
2. Convex surrogate to the infinity norm using $\log \sum \exp(.)$. Combining this with a two player game theoretic formulation could yield us an unconstrained convex dual problem
3. Explore ideas from the Lévy-Steinitz Theorem and Polygonal Confinement Theorem [Ros87], to attempt alternative intuitions to Komlós
4. Parallel ideas between Margins for Support Vector Machines using L^∞ norm regularization and discrepancy values A
5. Generating a two community graph so as to yield a maximal discrepancy
6. Relax $\mathbf{x}$ to be a vector in $\mathbb{R}^n$, along the lines of work by A. Nikolov [Nik13] on vector colorings

References

[Nik13] Aleksandar Nikolov. The Komlós Conjecture holds for Vector Colorings, 2013.

[Ros87] Peter Rosenthal. The Remarkable Theorem of Lévy and Steinitz, 1987.

[Spe94] Joel Spencer. Ten Lectures on the Probabilistic Method: Second edition. 1994.