Homework 8
High-Dimensional Probability for Data Science, Fall 2023

Hints are in the back of the homework set.

The *Gram matrix* of a system of vectors $v_1, \ldots, v_n \in \mathbb{R}^d$ is defined as the $n \times n$ matrix G whose entries are the inner products between the vectors, i.e. $G_{ij} = \langle v_i, v_j \rangle$.

1. Gram matrices

(a) Check that the Gram matrix G of any system of vectors is symmetric and positive semidefinite.

(b) Conversely, prove that any $n \times n$ symmetric and positive semidefinite matrix G is a Gram matrix of some system of vectors $v_1, \ldots, v_n$ in $\mathbb{R}^n$.

Goemans-Williamson's semidefinite relaxation algorithm (Lecture 13) yields a 0.878-approximation of the max cut of a graph. A weaker result – a 0.5-approximation – can be achieved by a trivial algorithm: a random cut.

Consider a graph $G = (V, E)$. For each vertex $v \in V$, flip a coin independently. If it comes up heads, include the vertex in the subset V_1, otherwise include it in the subset V_2. Denote by $E(V_1, V_2)$ the resulting cut – the set of edges that cross from V_1 to V_2.

2. A 0.5-approximation of max cut

Prove that the random cut described above is at least 1/2 times the maximal cut:

$$\mathbb{E}|E(V_1, V_2)| \geq \frac{1}{2} \max|E(U_1, U_2)|$$

where the maximum is over all partitions $V = U_1 \cup U_2$.

Here we will check a few versions of Grothendieck's identity (Lecture 13).

3. Some versions of Grothendieck's identity

Let u, v be unit vectors in $\mathbb{R}^d$, and g be a standard normal random vector in $\mathbb{R}^d$, i.e.$g \sim N(0, I_d)$. Prove the following identities.

(a) $\mathbb{E}\langle u, g\rangle\langle v, g\rangle = \langle u, v\rangle$.

(b) $\mathbb{E}\langle u, g\rangle \operatorname{sign}(\langle v, g\rangle) = \sqrt{\frac{2}{\pi}}\langle u, v\rangle$.

(c) Consider the random variable $X_u = \langle u, g\rangle - \sqrt{\frac{\pi}{2}} \operatorname{sign}(\langle u, g\rangle)$, and similarly for X_v. Deduce from (a) and (b) that

$$\frac{\pi}{2}\,\mathbb{E} \operatorname{sign}(\langle u, g\rangle) \operatorname{sign}(\langle v, g\rangle) = \langle u, v\rangle + \mathbb{E}\left[X_u X_v\right]. \tag{1}$$

Here we analyze a benchmark combinatorial problem (Lectures 12, 13): given numbers a_{ij}, find signs $x_1, \ldots, x_n \in \{\pm 1\}$ that maximize the quadratic form $\sum_{i,j=1}^n a_{ij} x_i x_j$. This problem is NP-hard. We will find a semidefinite relaxation (thus efficiently computable) which gives an approximate solution to this problem.

Our solution is based on a "semidefinite" version of Grothendieck's inequality. The general Grothendieck's inequality (Lecture 14) holds with constant 1.781. You will improve it to $\pi/2 \approx 1.571$ assuming the matrix (a_{ij}) is symmetric and positive semi-definite:

$$\max_{u_i \in \mathbb{R}^d \text{ unit}} \sum_{i,j=1}^n a_{ij}\langle u_i, u_j\rangle \le \frac{\pi}{2} \cdot \max_{x_i \in \{\pm 1\}} \sum_{i,j=1}^n a_{ij} x_i x_j. \tag{2}$$

More importantly, a solution (u_i) to the semidefinite program (left hand side of (2)) can be converted to an approximate solution (x_i) of the combinatorial problem (right-hand side of (2)) by *randomized rounding*. Previously, we only achieved this for Goemans-Williamson's max-cut relaxation (October 10, Lecture 17) but not for the original problem (2).

The result you will prove was first established in the paper [1] of Alon and Naor from 2004. In just a month and a half of our course, you made it almost to the forefront of contemporary research! Congratulations, and keep doing the great work.

4. A semidefinite Grothendieck inequality

Let $A = [a_{ij}]_{i,j=1}^n$ be a symmetric and positive semidefinite matrix, and let $u_1, \ldots, u_n$ be unit vectors in $\mathbb{R}^d$. Perform the randomized rounding of the vectors u_i, i.e. let

$x_i = \text{sign}(\langle g, u_i \rangle)$, where $g \sim N(0, I_d)$. Using identity (1), show that

$$\mathbb{E}\left[\frac{\pi}{2} \cdot \sum_{i,j=1}^{n} a_{ij} x_i x_j\right] \geq \sum_{i,j=1}^{n} a_{ij} \langle u_i, u_j \rangle.$$

This immediately implies (2), since if expectation is large, it must be large for some realization of the random labels x_i.

TURN OVER FOR HINTS

Hints

Hints for Problem 1. There are several ways to prove (b) using linear algebra. For example, consider the spectral decomposition $G = \sum_{i=1}^n \lambda_i u_i u_i^\mathsf{T}$ and define the matrix $V = \sum_{i=1}^n \sqrt{\lambda_i} u_i u_i^\mathsf{T}$. (Why can we take the square root?) Check that $G = V^2$. Deduce from this that G is the Gram matrix of the rows of V.

Hint for Problem 2. For a given pair of vertices $u, v \in V$, compute the probability of the event $D(u, v) = \{u, v \text{ land in different subsets}\}$. Argue that the cut equals $\sum_{(u,v)\in E} \mathbf{1}_{D(u,v)}$, where $\mathbf{1}_{D(u,v)}$ denotes the indicator random variable (it equals 1 is $D(u, v)$ holds and 0 otherwise). Then take expectation of the sum.

Hint for Problem 3. Identities in (a) and (b) are proved in the paper [1, Section 5.1], see the reference below. Don't copy the computations from that paper verbatim; look at them but then write down your argument yourself. You can use without proof the value of the first absolute moment of the standard normal distribution: $\mathbb{E}|g| = \sqrt{\frac{2}{\pi}}$.

Hint for Problem 4. The argument is sketched in the paper [1, Section 5.1] (see the reference below). But I think it would be easier for you to prove it yourself, as follows. Substitute $u = u_i$, $v = u_j$ into equality (1), multiply both sides by a_{ij}, and sum over all i, j. The sum in the right hand side breaks into two sums; check that the second one is nonnegative because the matrix (a_{ij}) is positive semidefinite.

References

[1] N. Alon, A. Naor, *Approximating the cut-norm via Grothendieck's inequality.* In Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, pp. 72–80. 2004. Click to download the paper.