Homework 2 Solutions

- Exercise 1 - Private Vectors

Alice has two unit vectors, \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \). Bob wants to know the value \( \mathbf{x} \cdot \mathbf{y} \). He is allowed to choose any vector \( \mathbf{v} \in \mathbb{R}^n \), send \( \mathbf{v} \) to Alice and she will send him \( \mathbf{v} \cdot \mathbf{x} \) and \( \mathbf{v} \cdot \mathbf{y} \).

a. Can Bob compute \( \mathbf{x} \cdot \mathbf{y} \) using fewer than \( n \) queries?

Solution: It seems impossible for Bob to compute \( \mathbf{x} \cdot \mathbf{y} \) using fewer than \( n \) queries. For example, if \( n = 2 \), and we make one query corresponding to the following equations,

\[
\begin{align*}
v_1 x_1 + v_2 x_2 &= a, \\
v_1 y_1 + v_2 y_2 &= b,
\end{align*}
\]

then even using the fact that \( \mathbf{x} \) and \( \mathbf{y} \) are unit vectors, we cannot uniquely determine \( \mathbf{x} \cdot \mathbf{y} \). This can be generalized for \( n > 2 \).

b. Suppose Bob wants to estimate\(^1\) the dot product \( \mathbf{x} \cdot \mathbf{y} \). Design an algorithm for Bob that uses the fewest possible queries to Alice to estimate \( \mathbf{x} \cdot \mathbf{y} \).

Solution: Bob can make \( k \) queries and each time, choose each entry in \( \mathbf{v} \) according to the normal distribution \( \mathcal{N}(0, \frac{1}{k}) \). Then the queries \( \mathbf{v} \cdot \mathbf{x} \) and \( \mathbf{v} \cdot \mathbf{y} \) are random projections of \( \mathbf{x} \) and \( \mathbf{y} \). Let \( \mathbf{x}' \in \mathbb{R}^k \) and \( \mathbf{y}' \in \mathbb{R}^k \) denote these \( k \) random projections. Then by the Johnson-Lindenstrauss Lemma, \( \mathbf{x}' \cdot \mathbf{y}' \) is a good estimate for \( \mathbf{x} \cdot \mathbf{y} \) when \( k = O\left(\frac{1}{\epsilon^2}\right) \) (since we only need to union bound over a constant number of dot products).

- Exercise 2 - Jaccard Distance

Let \( U \) denote a universe of \( N \) items. Let \( A \subseteq U \) and \( B \subseteq U \) denote two sets, each containing \( n \) unique items, where \( n < N \). The Jaccard distance between \( A \) and \( B \) is defined as

\[
\frac{|A \Delta B|}{|A \cup B|},
\]

where \( A \Delta B \) is the symmetric difference of the sets \( A \) and \( B \) (i.e., the number of elements in \( A \setminus B \) plus the number of elements in \( B \setminus A \)).

1. Suppose that \( A \) and \( B \) are each constructed by (independently) choosing an \( n \)-item subset of \( U \) uniformly at random among all such subsets. What is the expected value of the Jaccard distance between \( A \) and \( B \)?

2. Prove or disprove: The Jaccard distance obeys the triangle inequality. (The triangle inequality holds if \( d(A, B) \leq d(A, C) + d(B, C) \) for any three sets \( A, B, C \subseteq U \).)

- Exercise 3 - Small-Set Expansion on the Hypercube

A graph \( G = (V,E) \) is a \((\delta, \epsilon)\)-small-set expander if for all subsets of vertices \( S \subset V \) such that \(|S| = \delta |V|\),\(^2\)

\[
\frac{E(S, \bar{S})}{\sum_{v \in S} \deg(v)} \geq 1 - \epsilon.
\]

\(^1\)By estimate, we mean compute a multiplicative \( 1 \pm \epsilon \) or an additive \( \pm \epsilon \) approximation, whichever is larger.

\(^2\)If you prefer to use the constraint \(|S| \leq \delta |V|\) rather than \(|S| = \delta |V|\), you are free to do so.
A hypercube is a graph on $2^n$ vertices, where each vertex corresponds to a (distinct) length-$n$ binary string. Two vertices are connected by an edge if their respective binary strings differ in exactly one bit. For what values of $\epsilon, \delta$ is the hypercube a $(\delta, \epsilon)$-small-set expander?

- Exercise 4 - Unique Games and Cycle Covers

Suppose we are given a cubic graph $G = (V, E)$ and a constraint $x_i - x_j \equiv c_{ij} \mod p$ for each edge $ij \in E$. Our objective is to find an assignment from $[0, p)$ to the variables $\{x_i\}$ that maximizes the number of satisfied constraints.

a. Give a $\frac{2}{3}$-approximation algorithm for this problem.

b. Suppose that $G$ has a cycle cover containing at least $\frac{n}{10}$ cycles. Give an approximation algorithm for this problem with the best approximation factor you can find. (A cycle cover is a subgraph of $G$ with vertex set $V$ and edgeset $F \subset E$ such that each vertex $v \in V$ has degree exactly two.)

- Exercise 5 - Spectral Partitioning

a. Give an example of a $d$-regular graph $G$ in which the spectral partitioning algorithm (Lecture 8) has an output close to its upper bound.

b. Give an example of a $d$-regular graph $G$ in which the spectral partitioning algorithm has a output close to the lower bound (i.e., close to $\lambda_2(G)/2$).

- Exercise 6 - Local Connectivity

Assume that $G = (V, A)$ is an directed, unweighted graph such that the solution for the linear programming relaxation for ATSP on $G$ has value $|V|$. Give an algorithm for Local-Connectivity ATSP that is 2-light.