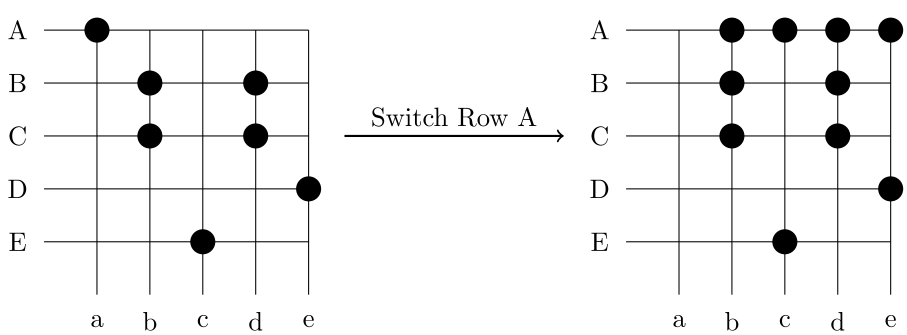
CS 466/666 Algorithm Design and Analysis **Last Updated:** July 8, 2019
Homework 4 **Due Date: July 19 (11pm), 2019**

You are allowed to discuss with others but not allowed to use any references other than the course notes. Please list your collaborators for each question. This will not affect your marks. In any case, you must write your own solutions.

There are totally 50 marks plus 20 bonus marks. The full mark is 50. This homework is counted 8% of the course.

---

1. **Unbalancing Lights**
   (10 marks + 10 bonus marks)
   (a) **Bonus:** Let $X = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n \in \{-1, +1\}$ are independent random variables with $\Pr[X_i = 1] = \Pr[X_i = -1] = 1/2$. Prove that $\mathbb{E}[|X|] = \Theta(\sqrt{n})$.
   (Hint: There are different ways to prove this. One simpler way is to compute moments of X and use the power mean inequality (i.e. $(\mathbb{E}[Y^p])^{1/p} \geq (\mathbb{E}[Y^q])^{1/q}$ for $p \geq q$ and a non-negative random variable Y) and the Cauchy-Schwarz inequality. You can assume these inequalities without providing proofs.)
   (b) Given a matrix $A \in \{-1, +1\}^{n \times n}$, use part (a) or otherwise to prove that there exist $x, y \in \{-1, +1\}^n$ such that $x^T A y = \Omega(n^{3/2})$.
   (c) Suppose we have an $n \times n$ array of lights in some initial state where each light is either on or off. We have $2n$ switches, one for each horizontal line and one for each vertical line that switches the whole line (i.e. on to off, off to on). Use part (b) or otherwise to prove that no matter what is the initial state of the lights, there is always a way to turn switches so that $n^2/2 + \Omega(n^{3/2})$ many lights are on.



2. **Graph Drawing**

   (10 marks) A graph is *planar* if it can be drawn on the plane such that the edges do not intersect with each other (except that they meet at the vertices). It is a well-known result that a simple planar graph with n vertices can have at most $3n-6$ edges. We say a graph G has intersecting number k if k is the maximum number such that any drawing of G on the plane has at least k pairs of edges intersecting. By the above result, a simple bound is that the intersecting number is at least t if the graph has at least $3n-6+t$ edges. Prove that the intersecting number is at least $m^3/(64n^2)$ for any simple graph with at least $m \geq 4n$ edges.

   (Hint: Do random sampling and apply the simple bound.)

3. **Isolated Vertices**

   (10 marks) Consider a random graph $G_{n,p}$ with $p = c \ln n/n$ for a constant c. Prove that if $c < 1$ then, for any constant $\epsilon > 0$ and for n sufficiently large, the graph has isolated vertices with probability at least $1-\epsilon$.

   (Remark: This implies that the uniform sampling algorithm for graph sparsification in L04 requires $\Omega(n \ln n)$ edges to form a good sparsifier if the input graph is a complete graph.)

4. **Random Bit Strings**

   (10 marks) Professor X from the Statistics department claims that she can easily distinguish a machine generated random string and a human generated random string. Her observation is simple: human generated random strings tend not to have a long run of consecutive ones or consecutive zeros, while machine generated random strings tend to have such substrings. We study this phenomenon in this problem.

   Let $s = (s_1, s_2, \ldots, s_n)$ be a random n-bit string, where we assume that n is sufficiently large. We say a substring $(s_i, s_{i+1}, \ldots, s_{i+k-1})$ is a k-consecutive substring if $s_i = s_{i+1} = \ldots = s_{i+k-1} = 0$ or $s_i = s_{i+1} = \ldots = s_{i+k-1} = 1$. Determine the largest k (that you can) so that s has a k-consecutive substring with probability at least 0.99.

5. **Discrepancy**

   (10 marks + 10 bonus marks) Let $H = (V, \mathcal{E})$ be a t-uniform hypergraph (i.e. each hyperedge is of size t) where $\deg_H(v) \leq t$ for all $v \in V$ (i.e. each vertex belongs to at most t hyperedges). In the discrepancy problem, we would like to assign $+1$ or -1 to each vertex so that the sum on each hyperedge is as close to zero as possible. Prove that there exists an assignment $\chi : V \to \{-1, +1\}$ so that for every hyperedge $e \in \mathcal{E}$ it holds that

   $$|\sum_{i \in e} \chi(i)| \leq O(\sqrt{t \log t}).$$

   Bonus: Provide a randomized polynomial time to find such an assignment. To prove the time complexity, you can assume that a random n-bit string can be compressed to $n-c$ bits for $c \geq 1$ with probability at most 2^{-c}.