Probabilistic and Smoothed Analysis of Algorithms
Assignment # 3

April 27, 2023

Roy, Akash | CS22M007
M.Tech CS | Indian Institute of Technology, Madras

Answering any 5.

1. This exercise is to demonstrate the limitations of considering expected running time of an algorithm as a useful measure. For any $n > 0$, describe a function $f : \{0,1\}^n \to \mathbb{N}$ such that
   - $\mathbb{E}[f] = \mathbb{E}_{x \in \{0,1\}^n}[f(x)] = n^c$ for some constant c and
   - $\mathsf{var}[f] = E[f^2] - (E[f])^2 = \Omega(2^n)$.

   I.e., there can be performance measures f which is polynomial in expectation, but variance being exponential. Give formal justification for your answer (i.e., computation of expectation and variance for the function f constructed). Further, write your views on why the function f that you have constructed may/maynot be a good performance measure for any algorithm in practice.

2. Compute the Fourier expansion of the following Boolean functions. (Note the functions are described in the Boolean/Fourier domain, you need to convert the function to Fourier domain first.)
   (a) Majority function $MAJ_3\{-1,1\}^n \to \{-1,1\}$, $MAJ_3(x_1, x_2, x_3) = \mathsf{sgn}(x_1 + x_2 + x_3)$.
   (b) $NAE_n : \{-1,1\}^n \to \mathbb{R}$, $NAE_n(x) = 1$ if not all of the input bits are equal, and 0 otherwise. Compute the Fourier coefficients for $n = 4$.
   (c) The inner product function: $IP_{2n} : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$, given by $IP_{2n}(x,y) = \mathsf{sgn}(\sum_{i=1}^{n} x_i y_i)$. First write down the Fourier coefficients for the case of $n = 2$ (3 points) and then obtain a general formula (2 points). No proof needed for the general formula.

3. Exercise 1.8 in the book " Analysis of Boolean Functions" by Ryan O'Donnel. Point split: 4+2+4.

4. (a) Exercise 1.15 in the book Analysis of Boolean Functions by Ryan O'Donnell. [5] Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. For $k \in [1,n]$, let $f^{=k}$ be the function given by $f^{=k}(x) = \sum_{S \subset [n], |S|=k} \widehat{f}(S)\chi_S(x)$. Prove the formula for $\langle f^{=k}, f^{=\ell} \rangle$ given in Exercise 1.18 of the book " Analysis of Boolean Functions" by Ryan O'Donnel.

5. We had stated in the class that Knapsack problem admits a pseudo linear time algorithm. This question is on developing such an algorithm and using it for a simple approximation algorithm for Knapsack.

   (a) Using the standard dynamic programing approach, develop a $O(\mathsf{poly}(n)W)$ time algorithm for the Knapsack problem with n items. Clearly mention each step and give a complexity analysis of the algorithm.

   Ans: Collaborators:

   (b) Using the above algorithm, devlop an algorithm that given ϵ, obtains an $1-\epsilon$ approximate solution(i.e., outputs a solution with profit at least (1-ϵ) times the maximum) in time $\mathsf{poly}(n, 1/epsilon)$.

   Ans: Collaborators:

6. Let A be an algorithm for a Euclidean optimization problem $\mathcal{P}$. in the plane. Consider an input instance x and the simple perturbation model with a given parameter $\delta \in (0, 1/2)$. Let $x \in ([0,1]^2)^n$. A δ perturbation of x is given by $(x_1 + y_1, \ldots, x_n + y_n)$, where $y_1, \ldots, y_n$ are chosen uniformly and independent on $[-\delta, \delta]$. The algorithm A is said to be (ϵ, δ) stable with respect to x, if $Pr[|A(x) - A(y)| \leq \epsilon] = 1 - o(1)$, where y is a δ perturbation of x.

   (a) Consider the Christofide's algorithm for ETSP (Euclidean TSP) in the plane. Construct an infinite family of inputs $X = (X_n)_{n>0}$ (i.e., for every $n > 0$, construct an instance X_n with $O(n)$ many points) such that $CHR(X_n) = (1.5 - o(n))TSP(X))$.

   Ans: Collaborators:

   (b) Are the instance you have constructed above (ϵ, δ) stable for suitable values of δ and ϵ? Give a detailed justification to your answer.

   Ans: Collaborators:

7. This question uses the same notion of perturbation as in the above question. Let $\mathcal{P}$ be a Euclidean optimization problem in the plane and $x \in (\mathbb{R})^2)^n$ be an input instance and y be a δ perturbation of x. We say that $\mathcal{P}$ is (ϵ, δ) stable at x, if $Pr[|\mathcal{P}(x) - \mathcal{P}(y)| \leq \epsilon] = 1 - o(1)$.

   (a) Suppose $\mathcal{P}$ be an NP hard Euclidean optimization problem such that $\forall x \in (\mathbb{R}^2)^n$, $\mathcal{P}$ is (ϵ, δ) stable at x, for some ϵ and δ. Can we have a smoothed polynomial time algorithm that solves $\mathcal{P}$ exactly? Justify your answer.

   (b) Suppose A is approximation algorithm for $\mathcal{P}$ with some approximation performance γ. Assume that $X = (X_n)$ is an infinite family of worst case instances for A (i.e., approximation ratio of A on X_n is $\gamma - o(1)$). Further, suppose that

X_n is (ϵ, δ) stable for $\mathcal{P}$ for every n. What will be the best possible smoothed approximation ratio of A under δ perturbations? Prove your answers.

8. Consider the pathTSP problem: Given a set of n points $x_1, \ldots, x_n \in [0,1]^2$, in the Euclidean plane, compute the minimum weight of a Hamiltonian path through the points. Define a Euclidean functional corresponding to pathTSP. Decide if the newly defined function satisfy any of the properties among subadditivity, superadditivity, smoothness, growth bound. Give detailed arguments justifying your answer.

Bonus

1. Let X be an instance of the k-means clusturing problem, with $|X| = n$. Let $B_1, \ldots, B_r$ be a sequence of clusturings obtained by the k-means algorithm. Show that there is some constant c such that if $r > c$, then at least one of the clusters among the clusterings $B_1, \ldots, B_r$, has seen at least three different configurations. (**Note:** in the class we had shown this for $r > 2^k$, using the pigeonhole principle.)