Coursework, Advanced Algorithms 2017

Due: April 24, 11:59am

- You should write up your own solutions independently. You need to indicate the names of the people you discussed a problem with; ideally you should discuss with no more than two other people.
- Typeset with LaTex is highly recommended. Please submit the PDF file through SAFE before the deadline.
- Do not share written solutions, and remember to cite all collaborators and sources of ideas. Sharing written solutions, and getting solutions from outside sources such as the Web or students not enrolled in the class is strictly forbidden.

Problem 1: We are given two independent streams of elements from $\{1, \ldots, n\}$, and we only consider the cash register model. Let $A[1, \ldots, n]$ and $B[1, \ldots, n]$ be the number of occurrences of item i in two streams, respectively.

1. Design a randomised streaming algorithm to estimate $X = \sum_{i=1}^{n} A[i]B[i]$ with additive error $\varepsilon \cdot \|A\|_1 \cdot \|B\|_1$.

   (20 points)

2. Analyse the space complexity of your proposed algorithm, and prove why your algorithm returns the correct answer. (20 points)

Problem 2: Let $\mathcal{H}$ be a finite set of functions such that $h : \{0,1\}^m \to \{0,1\}^n$ for every $h \in H$. We say that H is universal if

$$\mathbb{P}_{h\in\mathcal{H}}[h(x) = h(y)] \leq 2^{-n}$$

for any $x \neq y \in \{0,1\}^m$. Moreover, we say that $\mathcal{H}$ is pairwise independent if

$$\mathbb{P}_{h\in\mathcal{H}}[h(x) = u \wedge h(y) = v] = 2^{-2n}$$

for any different $x, y \in \{0,1\}^m$ and any $u, v \in \{0,1\}^n$.

In this problem we consider the constructions of universal (or pairwise independent) hash function families of small size.

1. Consider a family $\mathcal{H}$ of size 2^{m+n-1} where each $h \in \mathcal{H}$ is identified with a string $h \in \{0,1\}^{m+n-1}$. Given such a string, we create an $n \times m$ matrix $a(h) \in \mathbb{R}^{n\times m}$ where

   $$a(h)_{i,j} = h_{i-j+m},$$

   where $h_k, k \in \{1, \ldots, m+n-1\}$, is the kth bit of h. Then we define

   $$h(x) = a(h)x$$

   for all $x \in \{0,1\}^m$, where $a(h)$ is the matrix product of $a(h)$ with the column vector x, and where all arithmetic is taking place over the field GF(2) $= \{0,1\}$. (So 1+1=0, etc.) Prove that $\mathcal{H}$ is universal, and explain why $\mathcal{H}$ is not pairwise independent. (20 points)

2. Now we enlarge $\mathcal{H}$ to a family of size 2^{m+2n-1}, with each $h \in \mathcal{H}$ being identified with a string $h \in \{0,1\}^{m+2n-1}$. We again define an $n \times m$ matrix $a(h)$ by

$$a(h)_{i,j} = h_{i-j+m}$$

and we also define a vector $b(h)$ of length n by

$$b(h)_i = h_{n+m-1+i}$$

for $i = 1, \ldots, n$. (Note that $a(h)$ and $b(h)$ use disjoint bits of h.) Then we define

$$h(x) = a(h)x + b(h)$$

for all $x \in \{0,1\}^m$. Prove that $\mathcal{H}$ is pairwise independent. (20 points)

Problem 3: For any function $\alpha : \mathbb{N} \to [0,1]$, we say that a boolean formula ϕ is a $2/3\mathsf{CNF}(\alpha)$ if it is of the form

$$\phi(x_1, \ldots, x_n) = C_1 \wedge \cdots \wedge C_m,$$

where each of $\alpha(m) \cdot m$ clauses C_i is an $\vee$ of exactly 2 literals, and each of the remaining $(1-\alpha(m)) \cdot m$ clauses is an $\vee$ of exactly 3 literals. We further write $2/3\mathsf{SAT}(\alpha)$ as the set of satisfiable $2/3\mathsf{CNF}(\alpha)$ formulas. We know the following two results:

- When $\alpha(m) = 1$, $2/3\mathsf{SAT}(\alpha)$ is polynomial-time solvable;
- When $\alpha(m) = 0$, $2/3\mathsf{SAT}(\alpha)$ is NP-hard.

Now you need to generalise these two results above. Formally, you need to prove that under which conditions of α the $2/3\mathsf{SAT}(\alpha)$ problem is polynomial-time solvable, and under which conditions of α this problem becomes NP-hard. Please notice that the tighter your results are, the higher points you will get. (Hint: As a start, you can show the case where $\alpha(m)$ is constant for any m.) (20 points)