

CSE 203A Homework 2

Spring 2026

This homework is due in class Friday May 1st 11:59pm on gradescope. Make sure to justify all of your answers with a mathematical proof. For algorithms you should both prove correctness (with appropriately small probability of error) as well as appropriate bound on runtime.

Question 1 (Hashing Large Universes, 35 points). *One awkward feature of k -wise independent hash families when applied to very large universes (for example if the universe consists of strings with at most a million characters, encoding decent sized text documents) is that with standard constructions, a member of a k -wise independent family would require $k \log_2(N)$ bits to store and would take $O(k \log(N))$ time to compute, where N is the universe size.*

One solution to this is to first apply a relatively weak hash function $H : U \rightarrow [M]$ where $M = O(n^3)$ so that:

1. *For any $S \subset U$ of size n , with reasonably high probability H causes no collisions on S .*
 2. *A member of the family of H can be stored in $O(\log \log(N) + \log(n))$ bits and can be computed in $O(\log(N))$ time.*
- (a) *Give an example of such a construction. Hint: It might be useful to operate in two stages to first map to a set of size $O(\log(N))$ and then to one of size n^3 . [25 points]*
- (b) *Show that this construction is optimal in the sense that for $N > m > n > 1$ that no hash family $H : [N] \rightarrow [m]$ consisting of fewer than $\log(N)/\log(m)$ functions can have the property that for any set $S \subset [N]$ of size n the probability that H causes a collision among the elements of S is less than $1/2$. [10 points]*

Question 2 (Cuckoo Hashing with 3-wise independence, 35 points). *Consider a universe $U = \mathbb{F}_2^T$ and identify $[m]$ with \mathbb{F}_2^t . Consider a family of hash functions h by letting h be a uniform random affine linear transformation from $\mathbb{F}_2^T \rightarrow \mathbb{F}_2^t$. In particular, $h(x) = Ax + b$ where A is a $t \times T$ matrix over \mathbb{F}_2 , and b is an \mathbb{F}_2 vector.*

- (a) *Show that this defines a 3-wise independent hash family. [10 points]*
- (b) *Show that if this family is used to implement Cuckoo Hashing for a carefully chosen set $S \subset U$ with $|S| = O(m^{5/6})$ that after picking a random hash function h from this family, one will be forced to rehash with at least constant probability. Hint: Let S be a random subset of a subspace of dimension $t + 2$. [25 points]*

Question 3 (Fingerprinting and Communication Complexity, 30 points). *In communication complexity, two players (traditionally named Alice and Bob) are each given half of the input to some function $F(X, Y)$. They take turns sending one bit of information to the other, and eventually want one to declare the correct value of the function F . The goal is to minimize the number of bits sent regardless of the amount of computation required.*

In particular, we will look at the equality problem. Alice is given an n bit string X and Bob is given an n bit string Y and they want to compute the function $F(X, Y)$, which is 1 if $X = Y$ and 0 otherwise.

- (a) *Prove that for deterministic protocols, n bits of communication are required. Hint: Consider the full communication transcripts if $X = Y = s$ for various values of s . [15 points]*

(b) Show that only 1 bit of communication is required if Alice and Bob have a source of shared randomness (i.e. there is a random string that Alice and Bob can both read without this counting as communication). In particular, you should design what one might call a coRP algorithm where if $X = Y$, they definitely return 1, but if $X \neq Y$ there is at least a 50% chance that they return 0. [15 points]

Question 4 (Extra credit, 1 point). Approximately how much time did you spend working on this homework?