Foundations and Frontiers of Probabilistic Proofs (Summer 2021)
Worksheet B.0: Warm-up
Date: 2021.07.26

The goal of this worksheet is to gain familiarity with basic facts of codes and property testing.

Problem 1. (Basics of linear codes) Let $C\colon \mathbb{F}^k \to \mathbb{F}^n$ be a *linear* code (i.e., $C(x) + C(y) = C(x+y)$ and $C(\alpha x) = \alpha C(x)$ for every $x, y \in \mathbb{F}^k$ and $\alpha \in \mathbb{F}$).

1. Prove that the (relative) distance of C is $\delta = \frac{\min_{x \neq 0} |C(x)|}{n}$, where $|y| = |\{\, i \in [n] : y_i \neq 0 \,\}|$ is the *Hamming weight* of y.

   What can you say about the cardinality of (the image of) C if $\delta > 0$? What about when $\delta = 0$?

2. Show that there exists $G \in \mathbb{F}^{n \times k}$ such that $C(x) = G \cdot x$ for every $x \in \mathbb{F}^k$. (In other words, C is the image of the *generator matrix* G.)

3. Show that there exists $H \in \mathbb{F}^{(n-k) \times n}$ such that $C(x)^\intercal \cdot H = 0$ for every $x \in \mathbb{F}^k$. (In other words, C is the kernel of the *parity-check matrix* H.)

4. Give an example of a code with $k = 2$ and $n = 3$ (over the finite field of your choice). Compute its relative distance and show that the generator and parity-check matrices are not unique by exhibiting G_1, G_2, H_1, H_2 satisfying items 2 and 3 with $G_1 \neq G_2$ and $H_1 \neq H_2$.

Problem 2. (Identity testing) Fix an arbitrary string $s \in \Sigma^n$ and $\varepsilon = \varepsilon(n) \in (0,1)$. Show that the query complexity of detecting whether an unknown string $x \in \Sigma^n$ is equal to s or differs from s in at least an ε fraction of locations is $\Theta(1/\varepsilon)$. That is:

1. Construct an algorithm that makes $O(1/\varepsilon)$ queries to x, and always accepts if $x = s$ and rejects with probability at least $2/3$ if x is ε-far from s.
2. Argue that no algorithm making $o(1/\varepsilon)$ queries satisfies both conditions.

Problem 3. (Hadamard code) The code $\mathrm{Had}\colon \mathbb{F}^k \to \mathbb{F}^{|\mathbb{F}|^k}$ is defined as $\mathrm{Had}(x) := (\langle x, y \rangle)_{y \in \mathbb{F}^k}$ (i.e., the encoding of x is the linear function $\mathrm{Had}(x)\colon \mathbb{F}^k \to \mathbb{F}$ where $\mathrm{Had}(x)(y) = \langle x, y \rangle$). Show that Had has relative distance $1 - 1/|\mathbb{F}|$. (Despite its exponential block length, this code has important features that will be useful in this course: *local testability* and *local decodability*.)