Problem set III

Nati Linial

A little coding theory:
A *binary code* C of *length* n is simply a subset $C \subseteq \{0,1\}^n$. Members $x \in C$ are called *codewords.* Codes are used in order to communicate over noisy channels. A *transmitter* is sending messages to a receiver, using only words from C. When the received word y is in C the assumption is that indeed y is the word that was transmitted. However, if the received word z is not in C, we have to make an intelligent guess which word from C is the one that has actually been transmitted. One of the standard solutions is to find a word $x \in C$ which differs from z in the least number of coordinates and assume that x is the transmitted word. There are two critical parameters associated with a binary code of length n

- The cardinality $|C|$ which we want to maximize in order to better utilize the communication channel. The usual thing is to consider the *rate* of C that is defined as $R(C) := \frac{1}{n}\log_2|C|$. (This quantifies the rate at which information is transmitted when we communicate using C as our code book).
- The property that allows us to deal with noisy channels is that codewords differ substantially from each other. The metric that we use in the *Hamming metric* on $\{0,1\}^n$ that is defined via $d_H(x,y) := |\{i | x_i \neq y_i\}|$. The *distance of* C is defined as $d(C) := \min_{x \neq y \in C} d_H(x,y)$.

A major question in this area is how to find codes that have both high rate and large distance. A key function that quantifies this set of problems is

$$R(\delta) := \lim_{n\to\infty}\sup\{R(C) | C \text{ is a binary code of length } n \text{ and } d(C) \geq \delta n\}.$$

Here are a few problems on this function.

- Show the Gilbert-Varshamov bound $R(\delta) \geq 1 - H(\delta)$ where H is the binary entropy function. (This means that very good codes exist.) Hint: Try to construct a good code by picking words one by one greedily.
- Show that $R(\delta)$ vanishes for $\delta > 1/2$. This means that if $|C|$ is large as a function of n, then we can find two words $x \neq y \in C$ of distance $\leq \frac{n}{2}$. Actually more is true (and is easier to prove). Namely, if $|C|$ is large as a function of n, then the *average* distance between the words in C is $\leq \frac{n}{2}$.
- The *weight* of $x \in \{0,1\}^n$ is defined as $|x| := |\{i | x_i = 1\}|$. Show that if the average of $|x|$ over $x \in C$ is pn for some $1 \geq p \geq 0$, then the average distance of words in C is $\leq 2p(1-p)n$.

- **The Elias upper bound:** Show that $R(\delta) \leq 1 - H(\frac{1-\sqrt{1-2\delta}}{2})$. This is done in a way that resembles the proof of the Sperner Lemma shown in Sudakov's class and the proof of the Erdős-Ko-Rado Theorem from a previous problem sheet. Let C be a binary code of length n with distance $d(C) = \delta n$.

  - Pick a random Hamming sphere of radius pn, namely a set S of the form

    $$\{v \in \{0,1\}^n | d_H(v,z) = pn\}.$$

    The *center* z of S is chosen at random, and we discuss the parameter $1 \geq p \geq 0$ below. What is the average cardinality $|S \cap C|$? Now pick S so that $|S \cap C|$ is at least as large as the average.

  - Note that the distances in $S \cap C$ are the same as in $z \oplus (S \cap C)$ (where $\oplus$ stands for mod2 coordinate-wise addition and $w \oplus A$ stands for $\{w \oplus a | a \in A\}$.

  - Select p cleverly as a function of δ so you can apply a previous item concerning the average distances in large sets of words and deduce the Elias bound.