A Quick Introduction to Coding Theory

Unit 1

Gil Cohen

October 29, 2024

Overview

What is a code?

Definition 1 (Codes)

An error correcting code is a map $c : \Sigma^k \to \Sigma^n$.

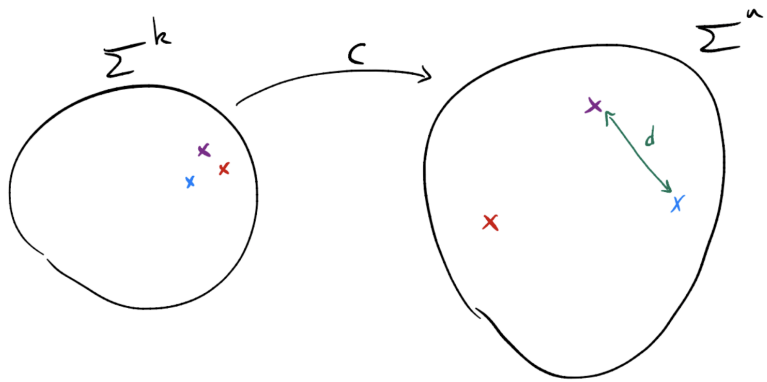
The distance of the code is

$$d = \min_{x \neq y \in \Sigma^k} \text{dist}(c(x), c(y)),$$

where $\text{dist}(w, z) = \{i \in [n] \mid w_i \neq z_i\}$ is the Hamming distance between w and z. The relative distance is defined by $\delta = \frac{d}{n}$.

The rate of the code is given by $\rho = \frac{k}{n}$.

The elements in the image of c are called codewords.



Introduction to coding theory

Codes were introduced by Shannon (1948) and Hamming (1950) as a primitive so to allow for communication over imperfect channels.

The larger δ, ρ - the better. But, of course, $\delta, \rho \in [0,1]$ are competing parameters.

Lemma 2 (Singleton bound (1964); Joshi (1958); Komamiya (1953))

For any code,

$$\rho + \delta \leq 1 + \frac{1}{n}.$$

The proof is via the pigeonhole principle. Consider, say, the first $n - d + 1$ coordinates. If $k > n - d + 1$ then two distinct codewords must agree on these coordinates, rendering the distance $< d$.

Is the Singleton bound attainable?

Linear codes

When the code $c : \Sigma^k \to \Sigma^n$ is over a (finite) field $\Sigma = \mathrm{F}$ and it is F-linear, we say that c is a linear code.

In such case, $c(\mathrm{F}^k)$ is a subspace of F^n. Sometimes we abuse definition, ignore the map itself, and consider the image $c(\mathrm{F}^k)$ as the code.

Note that the distance of linear codes is given by

$$d = \min_{x \neq y \in \Sigma^k} \mathrm{dist}(c(x), c(y)) = \min_{0 \neq z \in \Sigma^k} |c(z)|,$$

where

$$|w| = \mathrm{dist}(w, 0) = \{i \in [n] \mid w_i \neq 0\}$$

is the Hamming weight of w.

Many constructions of codes in the literature are linear.

Overview

The Reed-Solomon code (1960)

Let q be a prime power and $q, k \leq n$.

Let $\mathbb{F}_q$ be the finite field of size q and let $\mathfrak{p}_1, \ldots, \mathfrak{p}_n \in \mathbb{F}_q$ distinct. Define

$$\mathrm{RS} : \mathbb{F}_q^k \to \mathbb{F}_q^n$$
$$a = (a_0, a_1, \ldots, a_{k-1}) \mapsto (f_a(\mathfrak{p}_1), \ldots, f_a(\mathfrak{p}_n)),$$

where

$$f_a(x) = \sum_{i=0}^{k-1} a_i x^i \in \mathbb{F}_q[x].$$

Note that RS is a linear code with rate $\rho = \frac{k}{n}$.

Analyzing the distance of the Reed-Solomon code

As RS is linear, we fix $0 \neq a \in \mathbb{F}_q^k$ and bound $|\mathrm{RS}(a)|$ from below. But,

$$|\mathrm{RS}(a)| = |\{i \in [n] \mid f_a(\mathfrak{p}_i) \neq 0\}|.$$

As $a \neq 0$ we have that $f_a(x) = \sum_{i=0}^{k-1} a_i x^i \neq 0$.

By the fundamental theorem of algebra, $f_a(x)$ has at most

$$\deg f_a(x) \leq k - 1$$

distinct roots, and so

$$|\mathrm{RS}(a)| \geq n - (k-1) = n - k + 1$$

Thus, $d \geq n - k + 1$, and so

$$\rho + \delta \geq 1 + \frac{1}{n}.$$

Therefore, RS attains the Singleton bound!

Overview

The Reed-Muller code (1954)

RS is great in that it attains the Singleton bound (namely, it is a maximum distance separable (MDS) code).

The disadvantage of RS is in its large alphabet size $q \geq n$. This can be partially addressed using code concatenation.

A natural, more algebraic, idea is to move to bivariate polynomials.

The Reed-Muller Code (1954)

Let q be a prime power, and $k < n = q^2$.

Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ be distinct elements in $\mathbb{F}_q \times \mathbb{F}_q$. Define

$$\begin{aligned} \mathrm{RM} : \mathbb{F}_q^{\binom{r+1}{2}} &\to \mathbb{F}_q^n \\ a = (a_{i,j})_{0 \le i+j < r} &\mapsto (f_a(\mathfrak{p}_1), \ldots, f_a(\mathfrak{p}_n)), \end{aligned}$$

where $f_a(x, y) \in \mathbb{F}_q[x, y]$ is given by

$$f_a(x, y) = \sum_{i,j} a_{i,j} x^i y^j.$$

We have that $k = \binom{r+1}{2} \ge \frac{r^2}{2}$, and so $\rho \ge \frac{r^2}{2n}$.

The Reed-Muller Code (1954)

In the recitation you will invoke Schwartz-Zippel to bound the relative distance of RM by

$$\delta \geq 1 - \frac{r}{q} = 1 - \frac{r}{\sqrt{n}},$$

and so

$$r \geq (1-\delta)\sqrt{n}.$$

Thus,

$$\rho \geq \frac{r^2}{2n} \geq \frac{(1-\delta)^2}{2} \geq \frac{1}{2} - \delta.$$

In any case, unfortunately, $\rho < \frac{1}{2}$.

Interestingly, one can improve upon the rate by using multiplicity codes in which one outputs not only $f_a(\mathfrak{p}_i)$ but also the derivatives

$$\frac{\partial f_a}{\partial x}(\mathfrak{p}_i), \frac{\partial f_a}{\partial y}(\mathfrak{p}_i).$$

Overview

The Gilbert-Varshamov Bound (1952, 1957)

The following question begs an answer:

What is the best possible trade off between ρ and δ for a given alphabet size q?

The Gilbert-Varshamov bound is a non-explicit proof for the existence of codes giving a great tradeoff.

Theorem 3 (The Gilbert-Varshamov Bound; Informal, inaccurate version)

For every ρ, δ satisfying

$$\rho + \delta \geq 1 - \frac{1}{\log q}$$

there exists a code with relative distance δ and rate ρ.

The Gilbert-Varshamov Bound (1952, 1957)

The formal result is as follows.

Theorem 4 (The Gilbert-Varshamov Bound)

There exists a universal constant c s.t. for every q, every n large enough, and every ρ and $\delta < 1 - \frac{1}{q}$ satisfying

$$\rho + h_q(\delta) \geq 1 - \frac{c}{\log q},$$

there exists a code of of block length n on alphabet size q with relative distance δ and rate ρ. Here,

$$h_q(x) = x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x).$$

Is the Gilbert-Varshamov bound tight?

It is a common theme that the probabilistic method yields tight, or near-tight, results, and it is only the explicitness question that remains challenging.

Two famous counterexamples, rooting for structure, include:

1. Ramanujan graphs, and
2. Error correcting codes.

Indeed, the main result we prove in this course is

Theorem 5 (Ihara (1981); Tsfasman-Vladut-Zink (1982); Garcia-Stichtenoth (1995))

For every q which is an even power of a prime and all $\delta < 1 - \frac{1}{\sqrt{q}-1}$ there exist explicit arbitrary long codes satisfying

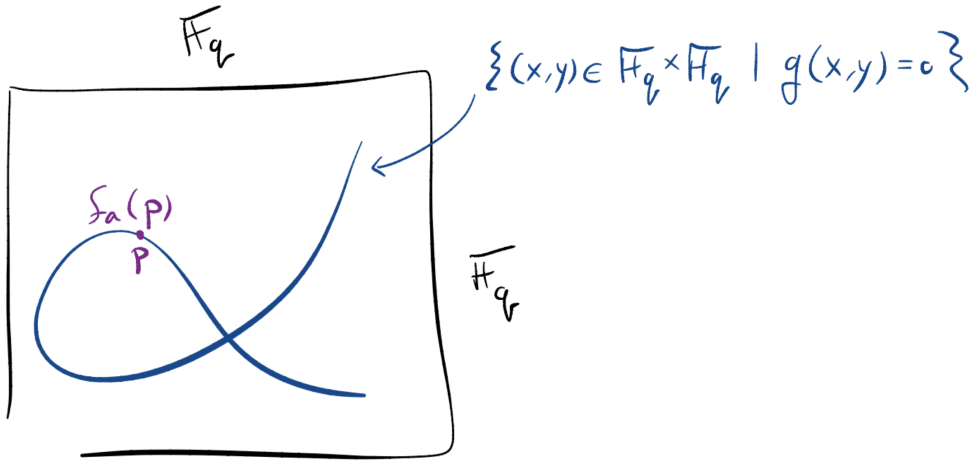
$$\rho + \delta \geq 1 - \frac{1}{\sqrt{q}-1}.$$

Overview

Algebraic-Geometric Codes

How do these codes look like?

Much like RM, we interpret the message a as a polynomial $f_a(x, y)$. However, we evaluate f_a not on the entire plane $\mathbb{F}_q \times \mathbb{F}_q$ but rather on an algebraic curve, namely, the zero set of a fixed polynomial $g(x, y) \in \mathbb{F}_q[x, y]$.



This suggestion is due to Goppa (1981), hence the name Goppa Codes.

Algebraic-Geometric Codes

A nice choice of a curve is given by

$$g(x,y) = y^p + y - x^{p+1},$$

where $q = p^2$ and p a prime power. This is called the Hermitian curve.

A "better" curve is given was suggested by Garcia and Stichtenoth

$$g(x,y) = y^p - y - \frac{x^p}{1 - x^{p-1}}.$$

A third example is

$$g(x,y) = y^2 - x^3 - x$$

which is an example for an elliptic curve.

Algebraic-Geometric Codes via Towers

We will in fact consider curves that are embedded in a higher dimensional ambient space, e.g.,

$$\{(x, y, z) \in \mathbb{F}_q^3 \mid g(x, y) = g(y, z) = 0\}.$$

The major questions we will ask in order to analyze the Goppa code associated with the curve are:

1. How many $\mathbb{F}_q$-points lie on the curve?
2. What is the genus of the curve?

The answer to the first question is very much related to the Riemann Hypothesis for curves over finite fields.

The path to proving the Riemann Hypothesis for curves over finite fields is lengthy and passes through the fundamental Riemann-Roch Theorem, which itself is based on the concept of genus.