# BOUNDS ON CODING THEORY FROM ALGEBRAIC GEOMETRY

GUILHERME ZEUS DANTAS E MOURA

ABSTRACT. Coding theory is concerned with finding efficient ways to encode a message so that one may correct errors in the message. In algebraic coding theory, we study efficient codes generated from algebraic geometric methods.

In this paper, I will construct the Reed–Solomon codes, generalize them using projective curves, and understand the results from [1] on finding a bound better than the well-known Gilbert–Varshamov one.

## 1. CODING THEORY

**Definition 1.1** (Code)**.** A *code $C$ over an alphabet $A$* is a subset of $A^n = A \times \cdots \times A$. We define $n$ as the *length of $C$*. A code $C$ over a field $A$ is a *linear code* if $C$ is a vector subspace of $A^n$. An element of a code $C$ is called a *code word*.

In this paper, $A$ is a finite field unless otherwise stated.

**Definition 1.2** (Hamming distance)**.** We define *Hamming distance* between $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in A^n$ as

$$\operatorname{dist}(\mathbf{x}, \mathbf{y}) = \# \left(x_i \neq y_i \mid i \in \{1, 2, \ldots, n\}\right), \tag{1.1}$$

in other words, the number of positions $\mathbf{x}$ and $\mathbf{y}$ differ.

**Proposition 1.3.** *Hamming distance is a metric over $A^n$, i.e., the following holds for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in A^n$:*

- $\operatorname{dist}(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$*;*
- $\operatorname{dist}(\mathbf{x}, \mathbf{y}) = \operatorname{dist}(\mathbf{x}, \mathbf{y})$*;*
- $\operatorname{dist}(\mathbf{x}, \mathbf{y}) \leq \operatorname{dist}(\mathbf{x}, \mathbf{z}) + \operatorname{dist}(\mathbf{z}, \mathbf{y})$*.*

**Definition 1.4** (Parameters of a code)**.** If $C$ is a linear code over $A$, we define *dimension of $C$* as $k = \dim_A(C)$ and *minimum distance of $C$* as $d = \min\{\operatorname{dist}(\mathbf{x}, \mathbf{y}) \mid \mathbf{x}, \mathbf{y} \in C\}$. (*If $C$ is a nonlinear code over an alphabet with size $q$, we can coherently define $k = \log_q |C|$.*) The length $n$, dimension $k$ and minimum distance $d$ are the *parameters* of $C$.

*Date*: last compiled April 5, 2021.

Suppose Alice wants to send a message to Bob through a noisy channel. They previously agree on a choice of code $C \subset A^n$, with parameters $n, k, d$. Alice will choose one of the $|A|^k$ code words and send it to Bob. Since the channel is not a perfect medium, some positions of the code may change; however, if less than $\frac{d}{2}$ of such changes occur, Bob can take the closest code word to the receiving message using Hamming distance and restore the original message.

Thus, a good code has two properties: it has large $d$ with respect to $n$, in order to correct as many errors as possible; but also has large $k$ with respect to $n$, so that Alice has a wider variety of possible messages to send and send more information.

**Definition 1.5.** If $C$ is a code, its code rate is $R = k/n$ and its relative minimum distance is $\delta = d/n$. Note that $R, \delta \in [0, 1]$.

Therefore, a good code is one with large $R$ — not much redundancy — and large $\delta$ — corrects many errors.

## 2. Singleton bound and a promising example

**Theorem 2.1** (Singleton Bound). *If $C$ is a code with parameters $n, k, d$, then*

$$k + d \leq n + 1, \tag{2.1}$$

*or equivalently,*

$$R + \delta \leq 1 + 1/n. \tag{2.2}$$

*Proof.* We will provide the proof for Theorem 2.1 when $C$ is a linear code.
WRITE PROOF. □

**Definition 2.2** (Reed–Solomon Codes). Let $q$ be a power of a prime, and $\mathbb{F}_q = \{\alpha_1, \alpha_2, \ldots, \alpha_q\}$ the field with $q$ elements. Let $k$ be an integer, and $\mathcal{L}_k$ the set of all polylomials over $\mathbb{F}_q$ with degree smaller than $k$. Let $k \leq n \leq q$ be an integer. The Reed–Solomon code $RS_q(n, k)$ over $\mathbb{F}_q$ is

$$RS_q(n, k) = \{(f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n)) \mid f \in \mathcal{L}_k\}. \tag{2.3}$$

**Proposition 2.3.** *The Reed–Solomon code $R_q(n, k)$ is a linear code with length $n$, dimension $k$ and minimum distance $n - k + 1$. Thus, any Reed–Solomon code meets the inequality of the Singleton Bound.*

*Proof.* $R_q(n, k)$ is a subset of $\mathbb{F}_q^n$, thus it has length $n$. Note that $\mathcal{L}_k$ is a vector space over $\mathbb{F}_q$. Note that $\{1, x, x^2, \ldots, x^{k-1}\}$ is a choice of basis for this vector space, thus it has dimension $k$. Consider the map $\phi : \mathcal{L}_k \to \mathbb{F}_q^n$ given by

$$f \mapsto (f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n)). \tag{2.4}$$

Note that the map $\phi$ is a linear transformation. Thus, its image $\operatorname{Im}\phi = RS_q(n,k)$ is also a vector space. Additionally, if $\phi(f) = \phi(g)$, then $f - g$ has at least $n$ roots, but has degree less than $n$; thus $f - g$ is the zero polylomial, which implies $f = g$. Therefore, $\phi$ is also injective. This implies that the dimension of the domain $\mathcal{L}_k$ is the same as the dimension of the image $RS_q(n,k)$, i.e., $\dim RS_q(n,k) = k$.

Finally, consider distinct $f, g \in \mathcal{L}_k$ and define $d = \operatorname{dist}(\phi(f), \phi(g))$, $f - g$ has at least $n - d$ roots. Futhermore, $f - g$ is a non-zero polynomial with degree less than $k$, thus has at most $k-1$ roots. Then, $k-1 \geq \#\,\text{roots} \geq n-d$. If we choose $f, g$ such that $d$ is the minimal distance, we get $k + d_{\min} \geq n+1$, which together with Singleton Bound implies

$$k + d_{\min} = n + 1. \tag{2.5}$$

□

The Reed–Solomon codes are good codes in the sense that key have the largest possible sum $k + d$ for their length $n$. However, Reed–Solomon codes are limited becuase their length is at most the the alphabet size.

## 3. Generalized Reed–Solomon codes

We shall redefine the Reed–Solomon codes using language related to a projective line. There is a way to replace the "projective line" with a "projective plane curve" and create other codes, called *Generalized Reed–Solomon codes* or simply *algebraic geometric codes*. We want large $R$ and $\delta$, and these codes yield

$$R + \delta \geq 1 + 1/n - g/n, \tag{3.1}$$

where $n$ is the number of rational points of a curve $X$, with genus $g$.

## 4. Final thoughts

On equation (3.1), we observe that good algebraic geometric codes are generated by curves with a large ratio between $n$ and $g$. On [1], the authors present a sequence of such curves, with $n/g$ large enough to create a better bound than the Gilbert–Varshamov one.

## References

[1] M. A. Tsfasman, S. G. Vlăduţ, and Th. Zink. "Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound". In: *Math. Nachr.* 109.1 (1982), pp. 21–28. DOI: 10.1002/mana.19821090103.