BOUNDS ON CODING THEORY FROM ALGEBRAIC GEOMETRY

GUILHERME ZEUS DANTAS E MOURA

1. Coding theory

WRITE INTRODUCTION WITH ALICE AND BOB.

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

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

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.

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.

Date: last compiled April 6, 2021.

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 δ — 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 L_k the set of all polynomials 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

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

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 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 $\phi : L_k \to \mathbb{F}_q^n$ given by

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

Note that the map ϕ 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 polynomial, which implies $f = g$. Therefore, ϕ is also injective. This implies that the dimension of the domain 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 $f, g \in L_k$ and define $d = \operatorname{dist}(\phi(f), \phi(g))$, $f - g$ has at least $n - d$ roots. Furthermore, $f - g$ is a non-zero polynomial with degree less than k, thus has at most $k-1$ roots. Then, $k-1 \geq \#$ 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 very good codes in the sense that key have the largest possible sum $k+d$ for their length n. However, Reed–Solomon codes are limited because their length is at most the alphabet size. So, a question naturally arises: Given fixed $\mathbb{F}_q$, are there codes over $\mathbb{F}_q$ with arbitrarily large n and $R + \delta = 1 + 1/n$? If not, how large can R and δ be when n gets larger? The Gilbert–Varshamov bound shows that there are codes with

$$1 - R \approx q(\delta), \text{as } n \to \infty, \tag{2.6}$$

in which

$$q(x) = x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(x). \tag{2.7}$$

Are there any better codes?

3. Rational Functions and Divisors

Definition 3.1 (Rational Function)**.** A rational function f is a function which is the ratio g/h of two polynomials. It is homogeneous if g, h are homogeneous. After cancelling common roots of g, h, the roots of g are called *zeros* of f and the roots of h are called the *poles* of f.

We say f has order n in P if P is a zero of muliplicity n; order $-n$ if P is a pole with multiplicity n; order 0, otherwise.

Definition 3.2 (Divisor)**.** Let F be the algebraic closure of $\mathbb{F}_q$. Let X be an irreducible nonsingular projective curve in N-dimensional projective space over F. A *divisor* on X is a formal finite sum of the form $D = \sum a_P P$, where P are points of X, a_P are integers and $a_P = 0$ for all but finitely many points P. The *degree* of D is $\sum n_P$. The *support* $\operatorname{Supp} D$ is the set $\{P \in X : a_P \neq 0\}$

If $D = \sum n_p P$, then define the vector space $\mathcal{L}(D)$ as the set of all homogeneous rational functions f such that the order of f at each point P of X is greater or equal to n_P. For our study, an important theorem is the following:

Theorem 3.3 (Riemman–Roch Theorem, [3])**.** *Let X be a nonsingular projective curve of genus*[a] *g defined over the field $\mathbb{F}_q$ and let D be a divisor on X. Then*

$$\dim \mathcal{L}(D) \geq \deg D + 1 - g, \tag{3.1}$$

with equality holding if $\deg D > 2g - 2$.

[a]For the reader that is not familiar with genus, it is enough to know that it is an integer that can be calculated for any given curve.

4. Generalized Reed–Solomon codes

Let $\mathbb{P}^1(\mathbb{F}_q)$ denote the projective line over $\mathbb{F}$. We will write $(a : b)$ to denote the projective point corresponding to the 1-dimensional vector space through (a, b). The points on $\mathbb{P}^1(\mathbb{F}_q)$ are the points

$$P_i = (\alpha_i : 1), \qquad 1 \leq i \leq q, \tag{4.1}$$

and

$$P_\infty = (1 : 0). \tag{4.2}$$

Following [1], let $\mathcal{L}_k$ be the set of two-variable homogeneous rational functions which have a pole of order less than k in the point Q.

Proposition 4.1. *The sets $\mathcal{L}_k$ and L_k are mapped with a bijection $\phi : f(x) \mapsto f(x/y)$.*

Proof. WRITE PROOF. □

Then, we can rewrite the Reed–Solomon code from 2.2 as

$$RS_q(n, k) = \{f(P_1), f(P_2), \dots, f(P_n) \mid f \in \mathcal{L}_k\}. \tag{4.3}$$

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 δ, and these codes yield

$$R + \delta \geq 1 + 1/n - g/n, \tag{4.4}$$

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

5. Final thoughts

On equation (4.4), we observe that good algebraic geometric codes are generated by curves with a large ratio between n and g. On [2], 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] J. H. van Lint and T. A. Springer. "Generalized Reed-Solomon codes from algebraic geometry". In: *IEEE Trans. Inform. Theory* 33.3 (1987), pp. 305–309. DOI: 10.1109/TIT.1987.1057320.

[2] 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.

[3] Judy L. Walker. *Codes and curves*. Vol. 7. Student Mathematical Library. American Mathematical Society, 2000. DOI: 10.1090/stml/007.