BOUNDS ON CODING THEORY FROM ALGEBRAIC GEOMETRY

GUILHERME ZEUS DANTAS E MOURA

1. Coding theory

Suppose Alice wants to send a message to Bob though a noisy channel. When Bob receives the message, it is possible that some of the information is misinterpreted. For example, if Alice sends the message '`food`' through the noisy channel, one of its letters may be misinterpreted and Bob could actually receive '`mood`'. So, the question Coding Theory tries to answer is: How can Alice and Bob agree on a system beforehand so that, if Alice sends a message to Bob, even if some misinterpretations occur, Bob will be able to understand the correct meaning?

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*. An element of a code C is called a *code word*. A code C over a finite field $\mathbb{F}_q$ is a *linear code* if C is a vector subspace of $\mathbb{F}_q^n$.

Back to the analogy with Alice and Bob, the code C is the set of all the messages that Alice may send to Bob, according to their agreement.

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

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

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 $\mathbf{x}, \mathbf{y}, \mathbf{z} \in A^n$:*

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

The proof of Proposition 1.3 is ommited from this short paper.

Definition 1.4 (Parameters of a code)**.** If C is a code over A, we define *dimension of C* as $k = \log_{|A|} |C|$ and *minimum distance of C* as $d = \min \{\operatorname{dist}(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \neq \mathbf{y} \in C\}$. (*If C is a linear code, the definition above is*

Date: 2021-05-07.

equivalent to $k = \dim_A(C)$.) The length n, dimension k and minimum distance d are the *parameters* of C.

Alice and Bob agree on a choice of code $C \subset A^n$, with parameters n, k, d. To send a message, 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 δ — corrects many errors.

2. The 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,

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

We will provide the proof for Theorem 2.1 when C is a linear code over a field K.

Proof. Let $W := \{\mathbf{x} = (x_1, \dots, x_n) \mid x_d = \cdots = x_n = 0\}$, which is a vector subspace of K^n with dimension $d - 1$. Since C is a vector space, $\vec{0} \in C$. For any non-zero vector $\vec{v} \in C$, $\mathrm{dist}(\vec{0}, \vec{v}) \geq d$, thus $\vec{v}$ has at least d non-zero entries, and therefore $\vec{v} \notin W$. Thus, $W \cap C = \{\vec{0}\}$.

Let $\vec{w}_1, \dots, \vec{w}_{d-1}$ and $\vec{v}_1, \dots, \vec{v}_k$ be a choice of basis for W and C, respectively. Since $W \cap C = \{\vec{0}\}$, the vectors $\vec{w}_1, \dots, \vec{w}_{d-1}, \vec{v}_1, \dots, \vec{v}_k$ are linearly independent. They all are vectors in K^n, therefore, $k + d - 1 \leq n$. ☐

Definition 2.2 (Reed–Solomon codes, [1])**.** Let q be a power of a prime, and $\mathbb{F}_q = \{\alpha_1, \alpha_2, \dots, \alpha_q\}$ 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

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

Proposition 2.3. *The Reed–Solomon code $RS_q(n,k)$ 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. $RS_q(n,k)$ 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 $\{1, x, x^2, \dots, x^{k-1}\}$ 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

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

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 $\mathcal{L}_k$ 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. Furthermore, $f-g$ 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 the Singleton Bound, implies $k + d_{\min} = n+1$. □

The Reed–Solomon codes are very good codes in the sense that they 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 well-known Gilbert–Varshamov bound shows that there are codes with

$$1 - R \approx H_q(\delta), \text{ as } n \to \infty, \tag{2.5}$$

in which

$$H_q(x) = x\log_q(q-1) - x\log_q(x) - (1-x)\log_q(x). \tag{2.6}$$

The article [2] uses Algebraic Geometry to show that there are codes that give better bounds.

3. Curves, rational functions and divisors

Through this section, let K be a field and L its algebraic closure.

Definition 3.1 (Algebraic plane curves)**.** An affine algebraic plane curve $\mathfrak{C}_f$ is the zero set of a polynomial $f \in K[x,y]$. A projective algebraic plane curve $\mathfrak{C}_F$ is the zero set in a projective plane of a homogeneous polynomial $F \in K[X, Y, Z]$.*

Note that there exists a natural injection, called *homogenization*, from the set of polynomials in $K[x, y]$ and the set of homogeneous polynomials in $K[X, Y, Z]$ defined by $f(x, y) \mapsto Z^{\deg f} f(X/Z, Y/Z)$. Because of this, we'll denote by $\mathfrak{C}$ both $\mathfrak{C}_f$ and $\mathfrak{C}_F$, where the usage is clear by context.

Definition 3.2 (Rational points, [3])**.** Let $\mathfrak{C}$ be a projective curve defined by $F(X, Y, Z) = 0$, where $F \in K[X, Y, Z]$ is a homogeneous polynomial. A point $(X_0 : Y_0 : Z_0) \in \mathbb{P}^2(L)$ is called a *L-rational point* on $\mathfrak{C}$ if $F(X_0, Y_0, Z_0) = 0$. Additionally, if $(X_0 : Y_0 : Z_0)$ is also in $\mathbb{P}^2(K)$, we call it a *K-rational point* on $\mathfrak{C}$, or simply a *rational point* on $\mathfrak{C}$.

Example. Consider the curve $\mathfrak{C}$ defined by $X^2 + Y^2 + Z^2 = 0$ over $\mathbb{F}_7$. The point $(1 : 2 : 3)$ is a $\mathbb{F}_7$-rational point on $\mathfrak{C}$. Note that the polynomial $x^2 - 3$ is irreducible over $\mathbb{F}_7$, but there it has a root α in its algebraic closure, $\overline{\mathbb{F}_7}$. The projective point $(1 : \alpha : \alpha)$ is a $\overline{\mathbb{F}_7}$-rational point on $\mathfrak{C}$.

Theorem 3.3 (Bezout's theorem, [3])**.** *If $F, G \in K[X, Y, Z]$ are homogeneous polynomials and there is no homogeneous polynomial in $K[X, Y, Z]$ that divides both F and G, then their curves $\mathfrak{C}_F$ and $\mathfrak{C}_G$ intersect in exactly $\deg(F) \cdot \deg(G)$ L-rational points, counting multiplicity.*

Definition 3.4 (Rational function on a curve)**.** Let $F \in K[X, Y, Z]$ be a homogeneous polynomial and let $\mathfrak{C} = \mathfrak{C}_F$. If $G, H \in K[X, Y, Z]$ are homogeneous polynomials of equal degree, then the fraction G/H is called a *rational function* on $\mathfrak{C}$. Fractions G/H and G'/H' define the same rational function if $G'H - GH'$ is a multiple of F, i.e., if $G'H - GH'$ vanishes at all L-rational points on $\mathfrak{C}$.

Note that it makes sense to evaluate a rational function on a projective point, since

$$\frac{G(X_0, Y_0, Z_0)}{H(X_0, Y_0, Z_0)} = \frac{\lambda^d G(X_0, Y_0, Z_0)}{\lambda^d H(X_0, Y_0, Z_0)} = \frac{G(\lambda X_0, \lambda Y_0, \lambda Z_0)}{H(\lambda X_0, \lambda Y_0, \lambda Z_0)},$$

i.e., the result is the same regardless of the representation of a given projective point.

After cancelling common factors of G and H, a L-rational point P on $\mathfrak{C}$ is called a *zero* of G/H whenever G vanishes at P, and is called a *poles* of G/H whenever H vanishes at P. We say G/H has order n at P if P is a zero of multiplicity n; order $-n$ if P is a pole with multiplicity n; and order 0, otherwise.

*In more general terms, algebraic curves (as opposed to algebraic plane curves) are algebraic varieties of dimension 1. For this paper, we will not worry about this; our discussion will be based on algebraic plane curves, but the results also follow in higher dimensions.

Definition 3.5 (Divisor)**.** Let $\mathfrak{C}$ be a projective plane curve defined over K. A *divisor* D on $\mathfrak{C}$ is a formal finite sum of the form $D = \sum a_P P$, where P varies over the L-rational points on $\mathfrak{C}$, a_P is an integer and $a_P = 0$ for all but finitely many points P. The *degree* of D is $\sum a_P$. The *support* of D, denoted by $\operatorname{Supp} D$, is the set of L-rational points P on $\mathfrak{C}$ that satisfy $a_P \neq 0$.

Definition 3.6 ([1])**.** If $D = \sum n_p P$, then define $\mathcal{L}(D)$ as the vector space formed by all homogeneous rational functions f such that the order of f at each L-rational point P on $\mathfrak{C}$ is greater or equal to $-n_P$ (plus the zero polynomial).

Proposition 3.7. *If* $\deg D < 0$*, then* $\mathcal{L}(D)$ *only contains the zero polynomial.*

Proof. Bezout's theorem implies that $\mathfrak{C}$ and $\mathfrak{C}_G$ intersect in the same number of points, counting multiplicity, than $\mathfrak{C}$ and $\mathfrak{C}_H$. Therefore, the number of zeros of a function is the same as the number of poles, counting multiplicity. Thus, the sum of the orders of G/H at the L-rational points on $\mathfrak{C}$ must be zero. On the other hand, if $G/H \in \mathcal{L}(D)$, then the sum of the orders of G/H at the L-rational points on $\mathfrak{C}$ must be $\geq -\deg(D)$.

Thus, if $\deg(D) < 0$, then a rational function G/H cannot be in $\mathcal{L}(D)$. □

For our study, an important theorem from Algebraic Geometry that will be useful in Section 4 when creating new types of codes is the following one:

Theorem 3.8 (Riemann–Roch Theorem, [3])**.** *Let* $\mathfrak{C}$ *be a nonsingular projective curve of genus** g *defined over the field* $\mathbb{F}_q$ *and let* D *be a divisor on* $\mathfrak{C}$*. Then*

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

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

4. Generalized Reed–Solomon codes

Let $\mathbb{P}^1(\mathbb{F}_q)$ denote the projective line over $\mathbb{F}_q = \{\alpha_1, \dots, \alpha_q\}$. We will write $(a : b)$ to denote the projective point corresponding to the 1-dimensional vector space containing (a, b). The points on $\mathbb{P}^1(\mathbb{F}_q)$ are the points

$$P_i = (\alpha_i : 1), \quad 1 \leq i \leq q, \qquad \text{and} \qquad P_\infty = (1 : 0). \tag{4.1}$$

Consider the divisor $D = (k-1)P_\infty$ and the associated vector space $\mathcal{L}(D)$, which can be seen as the set of two-variable homogeneous rational functions which have a pole of order less than k in the point P_∞.

*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.

Note that, if a polynomial $f(x) \in \mathcal{L}_k$ has degree $d < k$, then the rational function

$$F(X,Y) = \frac{Y^d f(X/Y)}{Y^d}$$

has P_∞ as its only pole, with order at most $k-1$. Thus, $F(X,Y) \in \mathcal{L}(D)$ and $f(\alpha_i) = F(P_i)$ for any α_i.

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

$$RS_q(n,k) = \left\{ \big(f(P_1), f(P_2), \ldots, f(P_n)\big) \mid f \in \mathcal{L}(D) \right\}. \tag{4.2}$$

Here, we are evaluating the function f in the points of the projective line. We shall generalize the idea by changing the projective line to an arbitrary projective curve, and allowing other divisors.

Definition 4.1 (Algebraic geometric codes, [1, 3])**.** Let $\mathfrak{C}$ be an irreducible nonsingular* projective plane curve, let D be a divisor on $\mathfrak{C}$ and let $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ be a set of rational points on $\mathfrak{C}$. Assume $\mathcal{P}$ and $\operatorname{Supp} D$ have no points in common, thus, no P_i can be a pole of any $f \in \mathcal{L}(D)$, and $f(P_i)$ is well-defined for any $f \in \mathcal{L}(D)$ and any $P_i \in \mathcal{P}$. Then, the *algebraic geometric code* associated to $\mathfrak{C}, \mathcal{P}$ and D is

$$C(\mathfrak{C}, \mathcal{P}, D) := \{(f(P_1), \ldots, f(P_n)) \mid f \in L(D)\} \subset \mathbb{F}_q^n. \tag{4.3}$$

Theorem 4.2 ([3])**.** *Let $\mathfrak{C}, \mathcal{P}, D$ be as above. Let g denote the genus of $\mathfrak{C}$. Suppose D satisfies $2g-2 < \deg D < n$. Then the algebraic geometric code $C(\mathfrak{C}, \mathcal{P}, D)$ is linear of:*

- *length n,*
- *dimension $\deg D + 1 - g$,*
- *minimum distance $d \geq n - \deg D$.*

Proof. Since $C(\mathfrak{C}, \mathcal{P}, D)$ is a subset of $\mathbb{F}_q^n$, it is a code of length n.

Let $\psi : \mathcal{L}(D) \to \mathbb{F}_q^n$ be defined by $f \mapsto (f(P_1), \ldots, f(P_n))$. This is a linear transformation, and $\mathcal{L}(D)$ is a vector space, thus the image of this function, which is precisely $C(\mathfrak{C}, \mathcal{P}, D)$, is also a vector space; thus it is a linear code.

Futhermore, we shall prove that ψ is injective. If $\psi(f) = \psi(g)$, then $\psi(f-g) = \vec{0}$. Let $h := f - g$. We have $h(P_1) = h(P_2) = \cdots = h(P_n) = 0$, so every P_i is a zero of h, and finally $a_{P_i} \geq 1$, for all $1 \leq i \leq n$. Therefore, $h \in \mathcal{L}(D - P_1 - \cdots - P_n)$. Since $\deg D < n$, we have $\deg(D - P_1 - \cdots - P_n) < 0$ and, consequenly, h must be the zero polynomial by Proposition 3.7. Therefore, $f = g$, finishing the proof that ψ is injective. Since ψ is a injective linear transformation, the dimension of the domain and of the image are the same. Therefore, the dimension of $C(\mathfrak{C}, \mathcal{P}, D)$ is $k = \dim \mathcal{L}(D)$, which evaluates to $\deg D + 1 - g$ by the Riemann-Roch Theorem.

*These adjectives just make sure $\mathfrak{C}$ does not have some weird behavior. The reader should not worry much about them.

Finally, let $f \neq g$ be polynomials such that $\psi(f)$ and $\psi(g)$ yield the minimum distance d. Then, $\psi(f) - \psi(g) = \psi(f - g)$ and $\psi(g) - \psi(g) = \vec{0}$ also yield the same distance d. Therefore, if we define $h = f - g$, exactly d coordinates of $\psi(h)$ are nonzero. Without loss of generality, say that $h(P_{d+1}) = \cdots = h(P_n) = 0$. Similarly as in the previous paragraph, this implies $h \in \mathcal{L}(D - P_{d+1} - \cdots - P_n)$. Since h is not the zero polynomial, Proposition 3.7 implies $\deg(D - P_{d+1} - \cdots - P_n) = \deg D - (n - d) \geq 0$, i.e., $d \geq n - \deg D$. □

This theorem is an interesting result! We want large R and δ, and these codes yield $k + d \geq n + 1 - g$, i.e.,

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

where n is the number of rational points of a curve $\mathfrak{C}$, with genus g. Futhermore, all the work done can be applied for algebraic curves in higher dimensional projective spaces.

5. Final thoughts

On equation (4.4), we observe that good algebraic geometric codes are generated by curves with a small ratio between g and n. On [2], the authors present a sequence of such curves, with g/n large enough to create a better bound than the Gilbert–Varshamov one.

Write one or two more sentences about this article.

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.