ALGEBRAIC GEOMETRY CODES

RYAN CATULLO, KEVIN RIZK

Abstract. We give a brief overview of algebraic geometry codes following [Sti08]. These are codes that generalize Reed-Solomon codes by considering polynomials evaluated at rational points lying on more general curves. The important implication is that these codes were the first infinite family to beat the Gilbert-Varshamov bound in the thirty years since it was established.

Contents

1. Algebraic Geometry Preliminaries

The field of algebraic geometry is incredibly deep and complex, so the purpose of this section is to give a working background on the algebraic geometry necessary to define this family of codes. As such, we will have to black box a number of important theorems but will try to provide motivation whenever possible.

1.1. **Varieties**

We start with a motivating question: what is a circle? We can define it as the set of points (x, y) such that $x^2+y^2 = 1$. Equivalently, we say it is the *vanishing locus* $V(f(X, Y))$ of the polynomial $f(X, Y) = X^2 + Y^2 - 1$, i.e. the set of solutions to $f(X, Y) = 0$ in $\mathbb{A}^2$. In fact, this makes sense over other fields, so we can define $V(f(X, Y)) \subset \mathbb{A}^2_k$ for any field k.

More generally, we have the following definition.

Definition 1.1 An *algebraic variety over* k is the vanishing locus $V(f_1,\dots,f_r)\subseteq\mathbb{A}^n_k$ where $f_1,\dots,f_r\in k[X_1,\dots,X_n]$ are polynomials in n variables with coefficients in k.

Now consider $\mathbb{A}^n_{\mathbb{C}}$ whose points are complex $(z_1,\dots,z_n)$. We can describe this space as $V(0)$ where we consider $0\in\mathbb{C}[X_1,\dots,X_n]$. If you are familiar with the notion of ideals, observe that the maximal ideals of $\mathbb{C}[X_1,\dots,X_n]$ are exactly $(X_1-z_1,\dots,X_n-z_n)$. Thus we have a natural bijection

$$\left\{\begin{array}{c}\text{maximal ideals of }\mathbb{C}[X_1,\dots,X_n]\\ \mathfrak{m}=(X_1-z_1,\dots,X_n-z_n)\end{array}\right\}\longleftrightarrow\left\{\begin{array}{c}\text{points in }V(0)=\mathbb{A}^n_{\mathbb{C}}\\ (z_1,\dots,z_n)\end{array}\right\}$$

How can we generalize this idea to varieties? Just like we can evaluate polynomials at points, we can "evaluate them at maximal ideals". For a polynomial $f\in k[X_1,\dots,X_n]$, we say that "$f(\mathfrak{m})=0$" if $\mathfrak{m}$ contains (f), i.e. $f\in\mathfrak{m}$.

In the previous correlation, if $\mathfrak{m}=(X_1-z_1,\dots,X_n-z_n)$ note that $k[X_1,\dots,X_n]/\mathfrak{m}$ identifies $X_1\equiv z_1,\dots,X_n\equiv z_n$. We can equivalently say that f vanishes at $(z_1,\dots,z_n)$ if

$$f(X_1,\dots,X_n)\equiv f(z_1,\dots,z_n)\equiv 0 \text{ in } k[X_1,\dots,X_n]/\mathfrak{m}$$

That is, $f\equiv 0$ in this ring. Equivalently, $f\in\mathfrak{m}$ or $\mathfrak{m}$ contains (f), which is the intuition behind this definition.

In general, if $f_1,\dots,f_r\in k[X_1,\dots,X_n]$ then maximal ideals $\mathfrak{m}$ containing $(f_1,\dots,f_r)$ correspond bijectively to maximal ideals of $k[X_1,\dots,X_n]/(f_1,\dots,f_r)$.

Definition 1.2 For an algebraic variety $V=V(f_1,\dots,f_r)$ over k in $\mathbb{A}^n_k$, the *coordinate ring of* V is $\mathcal{O}_V=k[X_1,\dots,X_n]/(f_1,\dots,f_r)$.

As we saw with the motivating example, if k is algebraically closed then we have a bijection

$$\left\{\begin{array}{c}\text{maximal ideals }\mathfrak{m}\text{ of}\\ \mathcal{O}_V=k[X_1,\dots,X_n]/(f_1,\dots,f_r)\end{array}\right\}\longleftrightarrow\left\{\begin{array}{c}\text{(closed) points in }V\subseteq\mathbb{A}^n_k\\ (z_1,\dots,z_n)\in k^n\cong k[X_1,\dots,X_n]/\mathfrak{m}\end{array}\right\}$$

We have an intuitive notion of dimension as well. Namely, every polynomial f should "cut down" the dimension by 1. For example, if $f_1(x,y,z)=x^2+y^2+z^2-1$ and $V=V(f_1)\subset\mathbb{A}^3_k$, then $\mathbb{A}^3_k$ is 3-dimensional and we cut out one equation f_1, leaving V with 2 dimensions. Note that V is a sphere which is naturally a 2-dimensional space.

Note that we want to cut out by new equations, i.e. $V(x,y,(x+y)^2)=V(x,y)$ in $\mathbb{A}^3$ is 1-dimensional, not 0 dimensional since if $x=y=0$ then $(x+y)^2=0$ is trivially satisfied. In turns out that the following is the correct notion of dimension.

Definition 1.3 Let $V\subset\mathbb{A}^n_k$ be a variety such that $V=V(f_1,\dots,f_r)$ and f_i is not nilpotent in $k[X_1,\dots,X_n]/(f_1,\dots,f_{i-1})$. Then the *dimension of* V is defined as $\dim(V)=n-r$.

The only varieties we need to consider for AG codes are *curves*, i.e. varieties of dimension 1. We will fix all curves to be connected, irreducible, and projective to avoid complaints on some of the technicalities. There are a few special properties of curves that allow a lot of wiggle room for computation. We list some of them.

For one, there is a correspondence between curves C over k and algebraic field extensions $k(C)/k$ given by the *function field* $k(C)$ of C, which is roughly the set of rational functions $f/g \in k(X_1, \ldots, X_n)$ such that g does not vanish identically on C, i.e. $g \not\equiv 0 \in \mathcal{O}_C$.

We can also consider for a point $p \in C$ the field of functions $f/g \in k(X_1, \ldots, X_n)$ such that $g(p) \neq 0$. This field is denoted $k(p)/k$ and is called the *residue field at p over k*. As an aside, we always have a surjection

$$\bigsqcup_{p \in C} k(p) \twoheadrightarrow k(C)$$

The *degree* $\deg(p)$ of a point $p \in C$ is the degree of the extension $k(p)/k$ (which is a finite extension iff p is closed, but we are only considering closed points). If k is algebraically closed then naturally all points are degree 1, so we will often make this assumption.

1.2. Divisors

It will also be useful to have a notion of a *divisor*, which is not the most illuminating terminology but is standard. For our purposes, a divisor D on a curve C over k is a formal integral combination of regular points

$$D = \sum_{p \in C^{\text{reg}}} a_p[p]$$

where all but finitely many a_p are 0. By regular we intuitively mean just "smooth" but there is a technical difference between the two. Regular means the tangent space at p has the same dimension as C, i.e. dimension 1, and smooth means that the Jacobian doesn't vanish at p, but these are not equivalent for all curves. The *degree* of a divisor $\deg(D)$ is defined as $\sum_{p \in C^{\text{reg}}} a_p \deg(p) = \sum_{p \in C^{\text{reg}}} a_p$ (since we're assuming all points have degree 1).

As an example, consider the curve cut out by $y^2 = x^3 - x$ in $\mathbb{A}^2$ over $\mathbb{C}$. A divisor could be $D = 2[(1,0)] - 4[(-1,0)] + [(2, \sqrt{6})]$, which has $\deg(D) = -1$. Given some rational function $r = f/g \in k(C)$, we can construct a divisor

$$\operatorname{div}(r) = \sum_{p \in C^{\text{reg}}} \operatorname{ord}_p(r)[p]$$

where $\operatorname{ord}_p(r)$ is the *order of vanishing of r at p*. In other words, $\operatorname{ord}_p(r) = \operatorname{ord}_p(f/g) = \operatorname{ord}_p(f) - \operatorname{ord}_p(g)$ is the number of zeros minus the number of poles, counted with multiplicity, at p. For example, if C is the curve as in the previous example $y^2 = x^3 - x$ and $r = y^2/(x-2)^2$ one can check that

$$\operatorname{div}(r) = 2[(0,0)] + 2[(1,0)] + 2[(-1,0)] - 2[(2, \sqrt{6})] - 2[(2, -\sqrt{6})]$$

Note that $\deg(\operatorname{div}(r)) = 2$, but it is actually true that the degree of the divisor of any rational function is always 0, i.e. "any rational function has the same number of zeros and poles, counted with multiplicity". The reason our count is off is that technically r has another pole at ∞.

Transform our curve by adding a variable z and setting x, y to $x/z, y/z$. Then our curve becomes $(y/z)^2 = (x/z)^3 - (x/z)$. If we multiply both sides by z^3, we get $y^2z = x^3 - xz^2$. Note if we set $z = 1$ we get back our original curve, but now we can consider what happens when $z = 0$. Note that $x/z, y/z$ "go to infinity" as z approaches 0, so in a sense this will tell us what our curve looks like at infinity.

Thus our equation becomes $0 = x^3$, which is just a triple point with special notation $[x : y : z] = [0 : 1 : 0]$. Note that with this same trick, r becomes $(y/z)^2/((x/z)-2)^2 = y^2/(x-2z)^2$. Then it's apparent that if $x = z = 0$ we get another pole of order 2, bringing our degree count to 0.

Definition 1.4 A curve C with these extra added points at infinity is called a *projective curve*. Any rational divisor on a projective curve has degree 0.

Given a divisor D, we can construct $\mathcal{O}_C(D)$ as the $\mathcal{O}_C$-module of rational functions $r \in k(C)$ such that $\operatorname{div}(r) + D \geq 0$, i.e. $\operatorname{div}(r) + D$ is a divisor with all nonnegative coefficients. Intuitively, if D has an $n[p]$ term for $n \geq 0$ then r is allowed to have a pole at p of order at most n, and if it has an $-n[p]$ term then r must have a zero of order at least n at p.

It turns out that $\mathcal{O}_C(D)$ is something called a *line bundle*, and in fact every line bundle $\mathscr{L}$ on a curve C is given by $\mathcal{O}_C(D)$ for some divisor D. We will ignore this, and instead endow it with a natural structure as a vector space over k. If we have two rational functions $r, s \in \mathcal{O}_C(D)$ and $a \in k$ then $r + s$ and ar are both in $\mathcal{O}_C(D)$. It is obvious for ar since multiplying by a constant doesn't change zeros or poles. Note that if D has an $n[p]$ term, then $\operatorname{ord}_p(r + s) \geq \min\{\operatorname{ord}_p(r), \operatorname{ord}_p(s)\} \geq -n$ since $\operatorname{ord}_p(r) + n \geq 0$ and $\operatorname{ord}_p(s) + n \geq 0$ by assumption.

Thus, we can define the *dimension* of D to be $\ell(D) = \dim_k \mathcal{O}_C(D)$. This is geometrically the dimension of the space of global sections of $\mathcal{O}_C(D)$ if that is meaningful to you. These numbers will be the focus of most of our calculations, and Riemann-Roch gives a tool for computing this. One easy observation is the following.

Lemma 1.1 For a divisor Z on C, $\ell(Z) = 0$ if and only if $\deg(Z) < 0$.

Proof. By definition, $f \in \mathcal{O}_C(Z)$ if and only if $\operatorname{div}(f) + Z \geq 0$, and taking degrees shows $\deg(Z) \geq 0$ since $\deg \operatorname{div}(f) = 0$ on any projective curve C. □

Note if $D = 0$ then $\mathcal{O}_C(D) = \mathcal{O}_C$, and under our assumptions (connected, irreducible, projective) we have that $\ell(0) = 1$, i.e. $\mathcal{O}_C \cong k$. We won't have time to prove this, but roughly $\mathcal{O}_C = k^r$ where r is the number of connected pieces of the curve and since we assume there is only one piece, we have $r = 1$.

We also need a canonical choice of divisor, which we make as follows. Let ω be a rational differential 1-form, i.e. locally near a point $p \in U \subseteq C$, $\omega = r dz$ where $r \in k(U) \simeq k(C)$ is a rational function. We say ω has a zero (resp. pole) at p if and only if r has a zero (resp.

pole) at p. Thus we can define a divisor K_C as $\operatorname{div}(\omega)$, which is well-defined up to $\operatorname{div}(s)$ for some rational $s \in \mathcal{O}_C$. In short, if ω' is another 1-form then $\omega/\omega' = s$ is a rational function, hence $K_C = K'_C + \operatorname{div}(s)$.

We'll do a quick computation since this idea can be confusing. Let $C = \mathbb{P}^1$ which is covered by two open sets $U_x = \{[x:y] \colon x \neq 0\}, U_y = \{[x:y] \colon y \neq 0\}$. On U_x we have a local coordinate system $s = [x:y] = [1:y/x]$ and on U_y we have another coordinate $t = [x:y] = [x/y:1]$, which are related on the overlap $U_x \cap U_y$ by $t = 1/s$. Basically, t is defined everywhere except ∞ ($[1:0]$) and s is defined everywhere except 0 ($[0:1]$), and as $s \to 0$ we have $t \to \infty$ by the transition $t = 1/s$.

On U_y, define our differential 1-form $\omega = dt$. This obviously has no zeros or poles on U_y, but what about on U_x? The transition map tells us that $\omega = dt = d(1/s) = -ds/s^2$ (this is just standard differentiation of $1/s$), hence ω has a pole at $s = 0$ of order 2. Thus $\operatorname{div}(\omega) = K_C = -2[1:0]$, which you might notice has degree -2 even though everything is projective.

The reason is that this is not a rational divisor, since the differentiation can introduce extra poles or zeros like we just saw. In fact, we will see in the next section why this is.

This divisor is actually incredibly special and has a lot of really nice properties. We give $\mathcal{O}_C(K_C)$ the special name Ω_C and call it the canonical bundle, even though we treat it as a vector space as before throughout this paper. We summarize these informalisms into the following (important) definition.

Definition 1.5 Given a rational differential 1-form ω on C, we define the *canonical divisor* $K_C := \operatorname{div}(\omega)$. The vector space $\Omega_C := \mathcal{O}_C(K_C)$ is the *canonical or dualizing bundle.*

1.3. **Riemann-Roch**

We can now state the powerhouse theorem for computation of AG codes.

Theorem 1.2 (Riemann-Roch for Curves) Let C be a projective curve, and let $D = \sum_{p \in C^{\text{reg}}} a_p[p]$ be a divisor on C. Then

$$\ell(D) = \deg(D) + 1 - g + \ell(K_C - D)$$

Here g is a special invariant of the curve called its *genus*, which has a really nice interpretation. Note that a curve C over $\mathbb{C}$ geometry looks like a $2d$ surface, since 1 complex dimension = 2 real dimensions. A common example is elliptic curves which can be translated into a complex torus via the Weierstrass transformation. The genus g is the "number of holes" in the surface. For example, a sphere has genus 0, and a torus has genus 1. Formally, g is defined as $\ell(K_C)$.

The Euler characteristic is another invariant given by $\chi = 2 - 2g$. You might have seen that $\chi = V - E + F = 2$ for planar graphs: this is directly related to the fact that $\chi = 2$ for the sphere, i.e. $g = 0$ or a sphere is a 0-holed torus. In fact, Euler characteristic is a well-defined notion for line bundles and is defined as $\chi(\mathcal{O}_C(D)) = \ell(D) - \ell(K_C - D)$. When $D = 0$ we get $\chi(\mathcal{O}_C) = \ell(0) - \ell(K_C) = 1 - g$, and when $D = K_C$ we get $\chi(\Omega_C) = \ell(K_C) - \ell(0) = g - 1$.

Therefore, Riemann-Roch stated differently says

$$\chi(\mathcal{O}_C(D)) = \deg(D) + \chi(\mathcal{O}_C)$$

We can use this to compute $\deg(K_C)$ as a nice first application. If we let $D = K_C$ in the above then

$$\deg(K_C) = \chi(\Omega_C) - \chi(\mathcal{O}_C) = (g-1) - (1-g) = 2g-2 \tag{1}$$

This brings our example computation of $K_C = -2[1:0]$ for $\mathbb{P}^1$ in the previous subsection full circle, since $\mathbb{P}^1$ is a genus 0 curve and hence $\deg(K_C) = 2g-2 = -2$ as we saw.

With this theorem in hand, we can start a discussion on AG codes and begin working out their properties.

2. Algebraic Geometry Codes

In this section, we will define Algebraic Geometry Codes, their properties, and how they can be used. They were first introduced by V.D. Goppa in the 1970s [Gop77; Gop81; Gop82]. AG codes development proved very important in the coding theory field as they were the first time the GV bound was beaten after its introduction. We will also discuss briefly a popular class of AG codes, Hermitian codes.

One of the big advantages of AG codes is that they can give rise to long codes with good distance and rate while still keeping the alphabet size relatively small. We will first motivate the definition of AG by using the RS codes as a starting point.

2.1. **From RS to AG Codes**

Recall the definition of Reed-Solomon codes.

Definition 2.1 Let $n, k \geq 0$ and $q \geq n$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q$ and define the *Reed-Solomon Code* with parameters q, n, k, α as

$$RS_q(n,k) = \{(f(\alpha_1), \ldots, f(\alpha_n)) \mid f \in \mathbb{F}_q[x], \deg f < k\}$$

That is, $RS_q(n,k)$ codes are defined as evaluation maps of polynomials of degree at most k on n points $\alpha_1, \ldots, \alpha_n$ in $\mathbb{F}_q$. Now, by going into a more geometric view, we can see that $\mathbb{F}_q$ is an affine line which we can projectivize to $\mathbb{P}^1_{\mathbb{F}_q} =: \mathbb{P}^1$ (we use the latter notation when the base field $\mathbb{F}_q$ is clear from context). Recall that now our points are $[x : z]$ defined up to nonzero scaling, and we recover "normal" points by $x/z = \lambda x/\lambda z$. Thus our "point at infinity" is the point $[1 : 0]$, where x/z doesn't make sense on $\mathbb{F}_q$.

We can view $f(x) \in \mathbb{F}_q[x]$ equivalently as a rational function $F(x,z)$ on $\mathbb{P}^1$ defined by $F(x,z) = f(x/z)$. For example, if $f(x) = x^3+1$ then $F(x,z) = (x/z)^3 + 1 = (x^3+z^3)/z^3$. Note the numerator and denominator are always homogeneous. Then the condition $\deg f < k$ is restated equivalently as $F(x,z)$ having a pole of order at most $k-1$ at $[1:0]$, since if $f(x) = a_0 + \ldots + a_d x^d$ where $a_d \neq 0$, we have $F(x,z) = (a_0 z^d + \ldots + a_d x^d)/z^d$ which obviously has a pole of order d at $z = 0$.

As a slight abuse of notation, we write $F(x,z)$ just as f. Then the above in the language of Section 1 says $\operatorname{div}(f) + (k-1)[1:0] \geq 0$ as a divisor, or equivalently $f \in \mathcal{O}_{\mathbb{P}^1}((k-1)[1:0])$.

Thus what if instead we try curves C other than $\mathbb{P}^1$, and divisors G other than $(k-1)[1:0]$? This is exactly the intuition behind algebraic geometry codes.

Throughout this section, we will fix the following notation. C will denote a projective connected irreducible algebraic curve over $\overline{\mathbb{F}}_q$ of genus g. We will let $p_1, \ldots, p_n$ be pairwise distinct $\mathbb{F}_q$-rational points of C (which are degree 1 as our base field is algebraically closed). We will let $D = [p_1] + \ldots + [p_n]$ denote the corresponding divisor, and G will denote another divisor on C such that $\operatorname{Supp} D \cap \operatorname{Supp} G = \emptyset$ (they have no points in common).

First, let's define algebraic codes for general curves (function fields) and general divisors.

Definition 2.2 The algebraic geometry (AG) code associated with the curve C and divisors D, G above is denoted $\mathcal{C}_{\mathcal{O}}(D,G)$ and is given by

$$\mathcal{C}_{\mathcal{O}}(D,G) := \{(f(p_1), \cdots, f(p_n)) : f \in \mathcal{O}_C(G)\} \subseteq \mathbb{F}_q^n$$

Note that $f(p_i)$ is well-defined since $\operatorname{div}(f) + G \geq 0$ and $\operatorname{Supp} D \cap \operatorname{Supp} G = \emptyset$, i.e. if f has a pole at p_i then $\operatorname{ord}_{p_i}(f)[p_i] < 0$, which would imply $\operatorname{div}(f) + G \not\geq 0$ since G has no $[p_i]$ term.

So now having defined AG codes, we can ask ourselves about the properties of such codes, for example, the distance and the message length.

Theorem 2.1 The following hold.

(i) $\mathcal{C}_{\mathcal{O}}(D,G)$ is a linear $[n,k,d]_q$ code such that

$$d \geq n - \deg G \quad \textit{and} \quad k = \ell(G) - \ell(G-D)$$

(ii) If $\deg G < n$ then $k \geq \deg G + 1 - g$. If in addition $2g - 2 < \deg G < n$, we have equality. That is,

$$k = \deg G + 1 - g$$

(iii) If $\{f_1, \ldots, f_k\}$ is a basis of $\mathcal{O}_C(G)$ then the $n \times k$ matrix

$$M = \begin{pmatrix} f_1(p_1) & f_2(p_1) & \ldots & f_k(p_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(p_n) & f_2(p_n) & \ldots & f_k(p_n) \end{pmatrix}$$

is a generator matrix for $\mathcal{C}_{\mathcal{O}}(D,G)$.

The proof of this theorem is a nice application of Riemann-Roch from Section 1.

Proof. If we consider the evaluation map, $ev : \mathcal{O}_C(G) \to \mathbb{F}_q^n$, defined as:

$$ev(f) = (f(p_1), \cdots, f(p_n))$$

we see that f is in the kernel of ev if and only if $\operatorname{div}(f) + G - D \geq 0$, i.e. f has at a zero of order at least 1 at each p_i. Equivalently, $f \in \mathcal{O}_C(G-D)$ and so the kernel of ev has dimension $\dim_k \mathcal{O}_C(G-D) = \ell(G-D)$. By rank-nullity we indeed get

$$k = \dim(\mathcal{C}_{\mathcal{O}}(D,G)) = \ell(G) - \ell(G-D)$$

If $\deg G < n$ then $\deg(G-D) < 0$, so by Lemma 1.1 $\ell(G-D) = 0$ (that is, ev is injective). Then $k = \ell(G)$ which can be computed directly using Riemann-Roch (Theorem 1.2).

$$\ell(G) = \deg(G) + 1 - g + \ell(K_C - G)$$

In particular, $\ell(G) \geq \deg G + 1 - g$ and we have equality by Lemma 1.1 iff $\deg(K_C - G) < 0$, i.e. $2g - 2 = \deg(K_C) < \deg(G)$ where the first equality is by (1).

For the distance, let us consider a non-zero $f \in \mathcal{O}_C(G)$ such that $ev(f)$ is of weight d. This means f vanishes on $n-d$ regular points $p_{i_1}, \cdots, p_{i_{n-d}}$. Then $f \in \mathcal{O}_C(G - (p_{i_1} + \cdots + p_{i_{n-d}}))$ and so $\ell(G - (p_{i_1} + \cdots + p_{i_{n-d}})) > 0$, so in particular, we must have $\deg G - (n-d) \geq 0$ which implies that $d \geq n - \deg G$.

It's clear that if $f = c_1 f_1 + \ldots + c_k f_k$ then $f(p_i) = c_1 f_1(p_i) + \ldots + c_k f_k(p_i) = [Mf]_i$. □

Now note that if we add up our inequalities for d and k in Theorem 2.1 above, we get

$$d + k \geq n + 1 - g \tag{2}$$

In particular, we see that we get an MDS code if and only if $g = 0$,

We will not go deeply into the theory of $g = 0$ codes, which are called rational codes, as it can be shown that rational codes are, in a sense, ”equivalent” to generalized Reed-Solomon (GRS) codes. In fact, the genus $g = 0$ case imposes many restrictions on the code, limiting what can happen on the curve. The curve will be isomorphic to $\mathbb{P}\mathbb{F}_q$, and in particular, the divisor of degree 0 will have a special form (it will be generated by at most one element), making it possible to prove that we get an equivalence with GRS codes. Moreover, we can also express BCH and Goppa codes as subfields of rational AG codes, providing a new perspective on all these codes.

Let's now discuss the dual code, which is a natural concept to consider when studying a new type of code.

2.2. **Dual Codes**

We can construct another code using the two divisors G,D, using another point of view. The AG codes defined above used an evaluation point of view. We can also use the residual point of view to create some linear codes.

Fix the same notation for $C, D = [p_1] + \ldots + [p_n]$, and G such that $\operatorname{Supp} G \cap \operatorname{Supp} D = \emptyset$ as in the previous subsection, and recall the notions of the rational differential 1-form ω on C, canonical divisor $K_C = \operatorname{div}(\omega)$, and canonical bundle $\Omega_C = \mathcal{O}_C(K_C)$. We can analogously define the vector space $\Omega_C(Z)$ as rational differential 1-forms η such that $\operatorname{div}(\eta) - Z \geq 0$. Note that $\Omega_C(Z) = \mathcal{O}_C(K_C - Z)$ since giving a rational function f such that $\operatorname{div}(f) + K_C - Z = \operatorname{div}(f) + \operatorname{div}(\omega) + Z \geq 0$ is the same as giving a rational 1-form $\eta = f\omega$ such that $\operatorname{div}(\eta) - Z \geq 0$.

For a rational 1-form η, let η_{p_i} denote the local form of η at p_i. For example, if $C = \mathbb{P}^1$ and $\eta = dt = -ds/s^2$ as in Section 1, then $\eta_0 = dt$ and $\eta_\infty = -ds/s^2$. By convention, if $\eta_p = f(z)dz$ then the point $z = 0$ corresponds to p.

DEFINITION 2.3 If $\eta_p = f(z)dz$, the *residue of* η at p is defined as the coefficient of $1/z$ in the Laurent series expansion of $f(z)$ at $z = 0$, and is denoted $\mathrm{Res}_p(\eta)$.

For example, if $f(z) = (1+z)/(z^2-z^3)$ then the Laurent series expansion is $(1/z^2)(1+z)/(1-z) = (1/z^2)(1+z)(1+z+z^2+\ldots) = (1/z^2)(1+2z+2z^2+\ldots) = 1/z^2 + 2/z + 2 + 2z + \ldots$ and therefore the residue would be 2. Here we used $1/(1-z) = 1 + z + z^2 + \ldots$ near $z = 0$, which is essentially the geometric series formula. This is a very beautiful notion coming from complex analysis, in particular because of the following theorem which we state but do not prove.

THEOREM 2.2 (Residue Theorem) Let ω be a rational differentiable 1-form on a smooth projective curve C. Then

$$\sum_{p \in C^{\mathrm{reg}}} \mathrm{Res}_p(\omega) = 0$$

DEFINITION 2.4 Let C, D, G be as before with $\operatorname{Supp} D \cap \operatorname{Supp} G = \emptyset$. Define the following code.

$$\mathcal{C}_\Omega(D, G) = \{(\mathrm{Res}_{p_1}(\eta), \ldots, \mathrm{Res}_{p_n}(\eta)) \colon \eta \in \Omega_C(G - D)\} \subseteq \mathbb{F}_q^n$$

Then we have an analogue of Theorem 2.1 whose proof contains essentially the same ideas.

THEOREM 2.3 The following statements hold.

(i) $\mathcal{C}_\Omega(D, G)$ is a linear $[n, k', d']_q$ code such that

$$d' \geq \deg G - (2g-2) \quad \textit{and} \quad k' = \ell(K_C - G + D) - \ell(K_C - G)$$

(ii) If $2g - 2 < \deg G$ then $k' = \ell(K_C - G + D) \geq n - \deg G + g - 1$.

(iii) If $2g - 2 < \deg G < n$ we have equality, i.e.

$$k' = n - \deg G + g - 1$$

Proof. Consider the residue map $\mathrm{Res}\colon \Omega_C(G - D) \to \mathbb{F}_q^n$, defined as

$$\mathrm{Res}(\eta) = (\mathrm{Res}_{p_1}(\eta), \ldots, \mathrm{Res}_{p_n}(\eta))$$

The kernel consists of rational 1-forms η with 0 residue at each p_i. Since $\mathrm{div}(\eta) - G + D \geq 0$, each η has a pole of order at most 1 at every p_i and having residue 0 implies it has no pole, i.e. $\mathrm{div}(\eta) - G \geq 0$. Thus the kernel is $\Omega_C(G) = \mathcal{O}_C(K_C - G)$ which has dimension $\ell(K_C - G)$. Similarly $\dim_k \Omega_C(G - D) = \dim_k \mathcal{O}_C(K_C - G + D) = \ell(K_C - G + D)$. By rank-nullity,

$$k' = \ell(K_C - G + D) - \ell(K_C - G)$$

In particular, if $\deg(K_C - G) < 0$ or $2g - 2 = \deg K_C < \deg G$ then $\ell(K_C - G) = 0$ and by Riemann-Roch

$$k' = \ell(K_C - G + D) = \ell(G - D) - \deg(G - D) + g - 1 \geq n - \deg G + g - 1$$

If in addition $\deg(G-D) < 0$ or $\deg G < \deg D = n$ then $\ell(G-D) = 0$ and we have equality.

For distance, suppose $n - d'$ residues vanish $p_{i_1}, \ldots, p_{i_{n-d}}$ for some 1-form η. Then $\eta \in \Omega_C(G - D + p_{i_1} + \ldots + p_{i_{n-d'}})$ so $\ell(K_C - G + D - (p_{i_1} + \ldots + p_{i_{n-d'}})) > 0$. We must then have $\deg K_C - \deg G + n - (n - d') \geq 0$, i.e.

$$d' \geq \deg G - (2g - 2)$$

□

In the particular case $2g - 2 < \deg G < n$, the above theorem and Theorem 2.1 imply

$$k' + k = (n - \deg G + g - 1) + (\deg G + 1 - g) = n$$

but it's not too hard to check this holds in general, i.e.

$$k' + k = \ell(K_C - G + D) - \ell(K_C - G) + \ell(G) - \ell(G - D) = n$$

(*Hint: use Riemann-Roch*). There's good reason for this, as we see with the next theorem.

Theorem 2.4 (Duality) The codes $\mathcal{C}_{\mathcal{O}}(D, G)$ and $\mathcal{C}_\Omega(D, G)$ are dual to each other;

$$\mathcal{C}_{\mathcal{O}}(D, G) = \mathcal{C}_\Omega(D, G)^\perp$$

Proof. We showed that $k+k' = n$ so it suffices to show any codewords satisfy $\langle ev(f), \mathrm{Res}(\eta)\rangle = 0$. Note that since $\mathrm{div}(f) + G \geq 0$ and $\mathrm{div}(\eta) - G + D \geq 0$, we have $\mathrm{div}(f\eta) + D \geq 0$ so $f\eta$ has a pole of order at most 1 at each p_i. The residue of $f\eta$ at p_i is $f(p_i)\,\mathrm{Res}_{p_i}(\eta)$, hence

$$\langle ev(f), \mathrm{Res}(\eta)\rangle = \sum_{i=1}^{n} f(p_i)\,\mathrm{Res}_{p_i}(\eta) = \sum_{i=1}^{n} \mathrm{Res}_{p_i}(f\eta)$$

Note that as $\mathrm{div}(f\eta) + D \geq 0$ these are the only possible poles, i.e. the only points with nonzero residues, hence by the residue theorem (Theorem 2.2)

$$\sum_{i=1}^{n} \mathrm{Res}_{p_i}(f\eta) = \sum_{p \in C^{\mathrm{reg}}} \mathrm{Res}_p(f\eta) = 0$$

□

We've established that these codes are mathematically extremely well-behaved, but what can we say about their merits as error-correcting codes? We address this question in the next few sections.

2.3. **Block Length and the Hasse–Weil Bound**

Recall from (2) that for AG codes, we have:

$$k + d \geq n + 1 - g.$$

We see that the lower the genus, the better our distance and rate are, and the closer we are to being an MDS (meeting the singleton bound). However, a problem arises with low-genus curves: they cannot be very long codes. In fact, as we have seen in the definition of AG codes, n is upper-bounded by the number of $\mathbb{F}_q$-rational points on the curve. But we know that the number of such points is bounded for a fixed g (for example by $q + 1$ for $g = 0$). In fact, we have the following result:

Theorem 2.5 (Hasse–Weil Bound) For a curve C of genus g, the number of $\mathbb{F}_q$-points of C, denoted by $\#C(\mathbb{F}_q)$, satisfies:

$$|\#C(\mathbb{F}_q) - (q + 1)| \leq 2g\sqrt{q}.$$

One interpretation of the bound is to notice that $q + 1$ is the number of points on the projective line $\mathbb{P}_{\mathbb{F}_q}$, and thus $\leq 2g\sqrt{q}$ can be interpreted as an error term (a square-term error in terms of q). Therefore, for small g, there are not many more points compared to a genus zero curve.

The proof is quite involved. One of the ideas is to note that all $\mathbb{F}_q$-rational points of C are, in fact, $\overline{\mathbb{F}_q}$-points fixed by the Frobenius morphism $\mathrm{Frob}_q : (x_1, \cdots, x_n) \to (x_1^q, \cdots, x_n^q)$. Using intersection theory, we can prove this bound.

Remark (Small Parenthesis) The Hasse-Weil Bound is really analogous to the Riemann Hypothesis for curves over finite fields. In fact, we can also define $\zeta(s, C)$, a zeta function for curves, and define similar conjectures to the Riemann Hypothesis (see Weil conjectures). The Riemann Hypothesis for "finite fields" (one of Weil's conjectures) was proved by Deligne, but the proof could not be replicated for $\mathbb{Z}$, which is our usual Riemann Hypothesis that could give more information about primes.

Having established the Hasse-Weil bound, we see that for a fixed genus, n cannot be arbitrarily large. Therefore, to have a long code with a smaller alphabet size q, we really need to have a larger genus g. Thus, we see that AG codes, in particular, give us a tradeoff between the block length of the code and its distance, as we approach being an MDS and meeting the singleton bound.

So, the problem of finding long codes with a smaller alphabet size q, but still with good rate and distance, is transformed into a question of finding curves with a large number of rational points—specifically, curves that meet the Hasse-Weil bound.

2.4. **Hermitian Codes**

Let's find a curve of genus $g > 0$ curve that satisfies the Hasse-Weil bound. One such code would be the Hermitian curve $C = \mathcal{H}_q$ defined over $\mathbb{F}_{q^2}$, curve with its affine curve equation given by:

$$X^q + X = Y^{q+1}.$$

or $X^qZ + XZ^q = Y^{q+1}$ in projective coordinates. We can prove that C is smooth, irreducible, and has genus $g = q(q - 1)/2$.

The number of $\mathbb{F}_{q^2}$ -rational points is q^3+1 with P_∞ ($[0:0:1]$) the point at infinity and q^3 other points that we will denote $P_{\alpha,\beta}$. To see why there are q^3 other points, we will use nice properties of the trace and norm field which in a way motivate the construction of the curve. Using the field and trace norm we can see that $Y^{q+1} = \mathrm{N}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(Y)$ and $X^q+X = \mathrm{Tr}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(X)$.

So in particular, for any $\beta \in \mathbb{F}_{q^2}$, $\mathrm{N}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(\beta) \in \mathbb{F}_q$. Now for any $c \in \mathbb{F}_q$, there are exactly q solutions to $\mathrm{Tr}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(\alpha) = c$ in $\mathbb{F}_{q^2}$. So now we have q^2 choice for β and for each β, we have q possible α, so we indeed get what we want.

We directly see that $q^3+1 = (q^2+1) + \sqrt{q^2}q(q-1)/2$, so the Hermitian curve indeed meets the Hasse bound.

Now we can define the one-point Hermitian code.

Definition 2.5 Let $r \in \mathbb{N}$. Let

$$D = \sum_{\alpha,\beta:\alpha^q+\alpha=\beta^{q+1}} P_{\alpha,\beta} \text{ and } G = rP_\infty$$

Indeed $\mathrm{Supp}(D) \cap \mathrm{Supp}(G) = \emptyset$, and so we can define the one-point Hermitian code $H_{q,r}$ as:

$$H_{q,r} = \mathcal{C}_{\mathcal{O}}(D,G).$$

We are evaluating the $f \in \mathcal{O}_{\mathcal{H}_q}(rP_\infty)$ on the points $P_{\alpha,\beta}$. One useful thing is we can explicitly write down a basis for $\mathcal{O}_{\mathcal{H}_q}(rP_\infty)$, in fact:

$$\mathcal{O}_{\mathcal{H}_q}(rP_\infty) = \left\{X^iY^j \mid iq + j(q-1) \leq r\right\}.$$

This explicit description makes the code easier to work with.

Moreover, we have if $r < q^3$, we have that $d \geq n-r$ and $k \geq r - q(q-1)/2+1$. In particular, we have

$$d + k \geq n+1-g = n+1-g = q^3+1-q^2/2+q/2$$

So we see that the distance and rate are still good, and we do not lose a lot if q is of appropriate size. Moreover, we also have that the size of the code can be longer than that of RS codes if we want not a very big alphabet size. In particular, if we work with an alphabet size q^2, RS codes can only have $n \leq q^2$ while Hermitian codes can have size q^3, so we have a factor q difference.

There are other nice properties of Hermitian codes that we will not cover here. For example, the dual of $H_{q,r}$ is $H_{q,q^3+q^2-q-r-2}$ or the different efficient algorithms to decode them.

All these properties make the hermitian codes a useful code in practice in cryptography and quantum error correcting codes, for example.

Now we will see how AG codes can be used to beat the GV bound.

2.5. **Tsfasman-Vladut-Zink Bound (TVZ) Bound**

Let's recall that $R = k/n$ is the rate of a code and $\delta = d/n$ is the relative distance of a code.

In coding theory, one popular lower bound and the existence of codes with good distance and rate were given by the GV bound.

Theorem 2.6 (Gilbert–Varshamov bound theorem, [Gil52; Var57]) For any linear code over $\mathbb{F}_q$, for every $0 < \delta < 1-1/q$ and $\epsilon > 0$, there exists a linear code with rate $R > 1-H_q(\delta)-\epsilon$ and relative distance δ.

The proof of this bound used a random linear code. So, finding an explicit family of codes that achieves this bound was and remains a good question. AG codes were one of the first to give codes that not only achieve the bound but also beat it. We will try to motivate the ideas behind the proof.

Recall that we had $k + d \geq n + 1 - g$. By dividing by n, we have $R + \delta \geq 1 + 1/n - g/n$. To maximize this quantity, the goal is to find, for each genus g, the curve containing the maximal number of $\mathbb{F}_q$ rational points.

Let's define the natural quantity:

$$N_q(g) = \{\#C(\mathbb{F}_q) \mid C \text{ curve of genus } g\}.$$

So now, as we are interested in g/n, we can take $n \approx \#C(\mathbb{F}_q)$. Recall from the Hasse-Weil bound that the only way $n \to \infty$ is for $g \to \infty$. As we will be interested in a family of curves where n and thus $\#C(\mathbb{F}_q)$ go to infinity, we want to see what happens to $N_q(g)$ when $g \to \infty$. So now we can define the Ihara constant:

Definition 2.6 The Ihara constant $I(q)$ is defined as:

$$A(q) = \limsup_{g\to\infty} \frac{N_q(g)}{g}.$$

Now, if we take a family of curves that meets the Ihara constants, we get that:

$$R + \delta > 1 - \frac{1}{A(q)}.$$

Thus, the question becomes one of estimating $A(q)$. First, recall that by the Hasse-Weil bound, we can directly prove that $A(q) \leq 2\sqrt{q}$, but this inequality was not tight. By a more careful argument, [VD83] proved that:

$$A(q) \leq \sqrt{q} - 1.$$

The goal is to find an upper bound for $A(q)$, or in other words, to find a sequence of curves or function fields that achieve this bound.

In fact, [TVZ82; Iha82] were able to find, for $q = p^{2k}$, where p is a prime number, a series of curves that achieve the upper bound of $A(q)$, so $A(q) = \sqrt{q} - 1$ for q a square prime power. The curves used are reductions modulo p of modular curves. The proof and the constructions are quite involved and do not yield easy explicit codes.

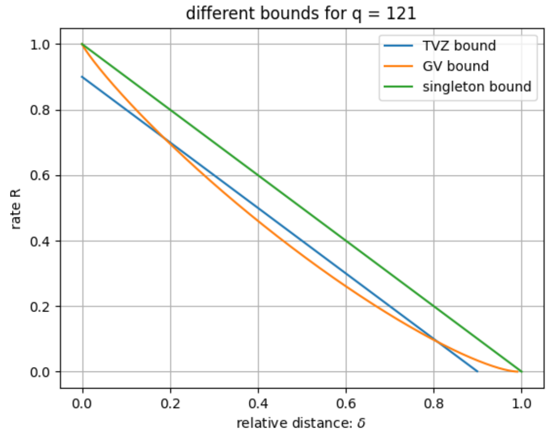

Figure 1: TVZ bound and GV bound and singleton bound plotted for $q = 121$

Theorem 2.7 (Tsfasman-Vladut-Zink Bound) Let q be a square number. Then, for every $0 \leq \delta < 1 - (q^{1/2} - 1)^{-1}$, there exists a sequence of linear codes over $\mathbb{F}_q$ such that their asymptotic relative distance δ and rate R satisfy:

$$R + \delta > \left(1 - \frac{1}{q^{1/2} - 1}\right).$$

In fact, we can prove that we beat the GV bound for $q \geq 49$ and q square numbers. Figure 1 shows a plot of the GV bound and TVZ bound for $q = 121$, and we can observe that the TVZ bound beats the GV bound.

The problem with the Ihara and TVZ proof was that the construction of the curves was not very explicit, making the AG codes complicated to construct. So, even if $A(q) = \sqrt{q} - 1$ was achieved, some works still attempted to find explicit examples of a sequence of curves that achieve the upper bound for q as a square number. Here, the language of function fields proved to be easier to work with, as the idea was to try to find an explicit tower of recursively defined function fields. This is what [GS95; GS96] did. One of the towers of function fields they proposed was defined as follows:

$$E_n := F(x_0, \cdots, x_n),$$

with x_0 being a transcendental element and x_{i+1} satisfying:

$$x_i^{q-1} x_{i+1}^q + x_{i+1} = x_i^q.$$

If needed, one can use these relationships to define curves explicitly. These towers of function fields indeed converged to the bound $A(q) \leq \sqrt{q} - 1$. This provides a more explicit construction than [Iha82; TVZ82].

Moreover, the method of towers of function fields also helped to obtain results for q that is not a square. For example, using a tower of function fields, [Bas+15] proved that for $m > 1$:

$$A(p^m) \leq 2\left(\frac{1}{p^{\lceil m/2 \rceil} - 1} + \frac{1}{p^{\lfloor m/2 \rfloor} - 1}\right)^{-1}.$$

This is tight for even values of m. For example, for $m = 3$, we get:

$$A(p^3) \geq \frac{2(p^2 - 1)}{p + 2},$$

which beats the GV bound for $q = p^3 > 7^3$, for instance.

However, for q that is not a square number, the exact value of $A(q)$ is still not known and remains an open problem.

2.6. **Decoding AG codes**

We have seen how to define AG codes, some of their most well-known examples, and how they can be used to beat the GV bound. However, what also makes a code interesting, aside from its rate and distance, is how easy it is to decode, correct errors, or list-decode it, and how many errors we can actually correct efficiently. This is a very important and extensive topic, and it is still an active area of research. We did not go into detail about the algorithms and different techniques, but it would be interesting as a continuation of our topic.

If we denote $d^* = \deg G - (2g - 2)$ (recall that $d' \geq d^*$ for $\mathcal{C}_\Omega(D, G)$), there are several general algorithms for decoding AG codes, most of which use the dual version or residual AG codes. For example, [SV90] developed a general algorithm that can correct up to $\lfloor (d^* - 1 - g)/2 \rfloor$ errors in $O(n^3)$ time for any G and D. Notably, this algorithm is still not perfect as it still can correct more errros.

[FR93], for instance, used ideas similar to BCH decoding and the Peterson method to develop a simpler algorithm that can correct $\lfloor (d^* - 1)/2 \rfloor$ errors in $O(n^2)$ time for $G = kQ$ for some point Q.

There are many different algorithms that correct errors with varying complexities and error-correction capabilities for AG codes.

Finally, it is worth noting that the Guruswami-Sudan method can be used to list-decode AG codes, but the complexities depend on other parameter properties related to the curve, aside from the genus.

In conclusion, there are various methods for decoding and list-decoding AG codes efficiently, each with its advantages and disadvantages, and research in this area is ongoing.

3. Conclusion

In this report, we have provided a brief overview of some of the theory and intuition behind the algebraic geometry used for AG codes. We introduced the codes, their properties, their dual perspective, and their advantages. In particular, we discussed Hermitian codes and explored the intuition behind how AG codes can be used to beat the GV bound.

It would be interesting to further explore the decoding and list-decoding algorithms for AG codes, and investigate how other AG and algebraic techniques—perhaps involving higher-dimensional objects can be used to better understand these codes or define new ones. Additionally, it would be fascinating to explore whether coding theory insights can, conversely, help us gain a deeper understanding of the geometry of curves over finite fields.

References

[Bas+15] A. Bassa et al. "Towers of Function Fields over Non-prime Finite Fields". In: *Moscow Mathematical Journal* 15 (2015), pp. 1–29.

[FR93] G.-L. Feng and T.R.N. Rao. "Decoding algebraic-geometric codes up to the designed minimum distance". In: *IEEE Transactions on Information Theory* 39.1 (1993), pp. 37–45. DOI: 10.1109/18.179340.

[Gil52] E. N. Gilbert. "A comparison of signalling alphabets". In: *Bell System Technical Journal* 31.3 (1952), pp. 504–522. DOI: 10.1002/j.1538-7305.1952.tb01393.x.

[Gop77] V. D. Goppa. "Codes associated with divisors". In: *Probl. Peredachi Inform.* 13.1 (1977). Translation: Probl. Inform. Transmission, vol. 13, pp. 22–26, 1977, pp. 33–39.

[Gop81] V. D. Goppa. "Codes on algebraic curves". In: *Dokl. Akad. Nauk SSSR* 259 (1981). Translation: Soviet Math. Dokl., vol. 24, pp. 170–172, 1981, pp. 1289–1290.

[Gop82] V. D. Goppa. "Algebraico-geometric codes". In: *Izv. Akad. Nauk SSSR* 46 (1982). Translation: Math. USSR Izvestija, vol. 21, pp. 75–91, 1983.

[GS95] A. Garcia and H. Stichtenoth. "A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vlăduț bound". In: *Inventiones Mathematicae* 121 (1995), pp. 211–222.

[GS96] A. Garcia and H. Stichtenoth. "On the asymptotic behaviour of some towers of function fields over finite fields". In: *Journal of Number Theory* 61 (1996), pp. 248–273.

[Iha82] Y. Ihara. "Some remarks on the number of rational points of algebraic curves over finite fields". In: *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* 28 (1982). 1981, pp. 721–724.

[Sti08] Henning Stichtenoth. *Algebraic Function Fields and Codes*. 2nd. Springer Publishing Company, Incorporated, 2008. ISBN: 3540768777.

[SV90] A.N. Skorobogatov and S.G. Vladut. "On the decoding of algebraic-geometric codes". In: *IEEE Transactions on Information Theory* 36.5 (1990), pp. 1051–1060. DOI: 10.1109/18.57204.

[TVZ82] M. A. Tsfasman, S. G. Vlădutx, and Th. Zink. "Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound". In: *Mathematische Nachrichten* 109 (1982), p. 21. DOI: 10.1002/mana.19821090104.

[Var57] R. R. Varshamov. "Estimate of the number of signals in error correcting codes". In: *Dokl. Akad. Nauk SSSR* 117 (1957), pp. 739–741.

[VD83] Serge Vlăduţ and V. Drinfel'd. "Number of Points of an Algebraic Curve". In: *Functional Analysis and Its Applications - FUNCT ANAL APPL-ENGL TR* 17 (Jan. 1983), pp. 53–54. DOI: 10.1007/BF01083182.