Lecture 9

AG Codes I

In this lecture we will give a brief introduction to codes from algebraic geometry.

9.1. Introduction

We continue with our general discussion of evaluation codes. We saw that Reed-Solomon codes are evaluation codes that have the best possible minimum distance given their block length and their dimension. The drawback, however, is that they are restricted in their length: the block length is at most equal to the size of the field. In particular, it is not possible to get sequences of Reed-Solomon codes for which the length goes to infinity, while the field size is fixed.

One of the ways to overcome this difficulty was to look at Reed-Muller codes. For these codes the block length can be as large as desired. However, we saw that as we increase the block length, either the relative minimum distance, or the rate, converge to 0. In particular, we cannot obtain *asymptotically good* codes, i.e., sequences of codes for which the relative distance and the rate converge to nonzero values.

The new idea that helps overcome this difficulty is due to V.D. Goppa, from 1981, though our presentation differs in part substantially with the one provided in the established literature on this subject.

Looking back at Reed-Solomon codes, the main reason why they yielded good codes was our control on the number of zeros of the code. The reason why Reed-Muller codes are not as good is that the evaluation of an m-variate polynomial on $\mathbb{F}_q^m$ will lead to a large number of zeros. Would it be possible, though, to evaluate the polynomial on a subset of $\mathbb{F}_q^m$ as to gain better control on the number of zeros? For this we need a tool that limits the number of zeros of such an evaluation. We are going to describe the tool first in the most familiar setting, that of the 2-dimensional plane.

Suppose that $m = 2$, i.e., we want to evaluate a bivariate polynomial $h(x, y) \in \mathbb{F}_q[x, y]$ on a subset of the plane $\mathbb{F}_q^2$. Suppose that this subset is the set of zeros of a bivariate polynomial $f(x, y) \in \mathbb{F}_q[x, y]$. Do we have any chance to obtain an upper bound on the number of zeros of $h(x, y)$ on the set of zeros of f?

We can phrase this question in yet a different way: the zeroset of $f(x, y)$ is a *curve* in $\mathbb{F}_q^2$, as is the zeroset of $h(x, y)$. So, we are asking how many intersection points these two curves will have at most. Figure 9.1 gives an example of such a setting in the plane $\mathbb{R}^2$, where $h(x, y) = x^3 - 3xy + y^3 - 1.8$ and $f(x, y) = y^2 - x^3 + 2x - 2$. In this case, the number of intersection points is 5.

9.2. Irreducible Curves and Bézout's Theorem

In general, nothing much can be said about the number of intersection points of $h(x, y)$ and $f(x, y)$. In fact, over an infinite field, this number can be infinity. For example, suppose that $f(x, y) = (x - y)(x + y)$ and $h(x, y) = (x - y)(x^2 + y^2 - 1)$. In this case, both $f(x, y) = 0$ and $h(x, y) = 0$ have the *component* $x - y = 0$ in common, and over an infinite field, this component has infinitely many points (and over the field $\mathbb{F}_q$ it has q points).

In our applications, $f(x, y)$ will be a fixed curve, and we will evaluate polynomials $h(x, y)$ of various degrees on it. There is no guarantee, a priori, that $h(x, y)$ and $f(x, y)$ have no components in common, because we have little

Lecture and scribes by Amin Shokrollahi

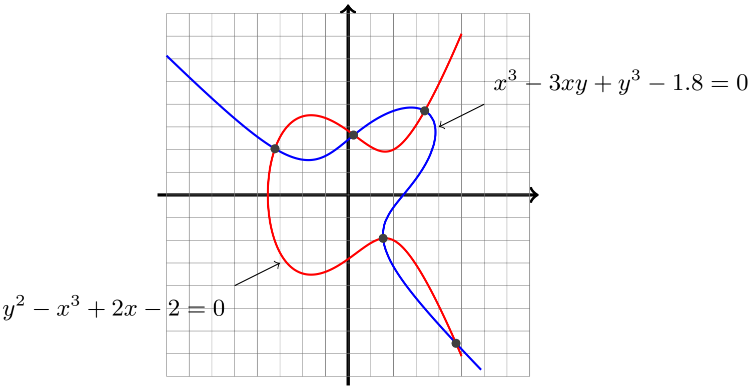

Figure 9.1: The curves given by $x^3 - 3xy + y^3 - 1.8 = 0$ and $y^2 - x^3 + 2x - 2 = 0$ intersect at 5 points in the real plane. By Bézout's theorem, the number of intersection points is at most 9.

control over $h(x,y)$. However, if we assume that $f(x,y)$ is *irreducible*, i.e., that it cannot be factored in a nontrivial way, then we can say a little bit more. Namely, in this case, $h(x,y)$ and $f(x,y)$ have a common factor iff h is a multiple of f. If this is not the case, i.e., if f is irreducible and h is not a multiple of f, then we can bound the number of intersection points of f and h.

For this, we need the concept of *degree*. The degree $\deg(g)$ of a polynomial $g \in \mathbb{F}_q[x,y]$ is the largest degree of a monomial in g. For example, the degree of $x^3y + xy + 3x - y$ is 4, since the degree of the monomial x^3y is 4. We have the following theorem.

Theorem 9.1 (Bézout's Theorem in the plane). *Suppose that $f(x,y), h(x,y) \in \mathbb{F}_q[x,y]$, that $f(x,y)$ is irreducible, and that $h(x,y)$ is not a multiple of $f(x,y)$. Then the number of points $(a,b) \in \mathbb{F}_q^2$ such that $f(a,b) = h(a,b) = 0$ is at most* $\deg(f)\deg(h)$.

For a proof of this theorem, we refer the reader to the book *Algebraic Geometry* by Joe Harris, or to the book *Algebraic Complexity Theory* by Bürgisser et al., or to the book *Ideals, Varieties, and Algorithms* by Cox et al.

Let us check the theorem in the situation of Figure 9.1: here, the degree of $f(x,y) = y^2 - x^3 + 2x - 2$ is 3, and the degree of $h(x,y) = x^3 - 3xy + y^3 - 1.8$ is also 3. The polynomial $f(x,y)$ is irreducible (without proof), and $h(x,y)$ is not a multiple of $f(x,y)$. Hence, Bézout's theorem implies that the number of intersection points is at most 9. (In fact, the theorem bounds the number even over the field $\mathbb{C}$ of complex numbers. The reader is invited to show that over this field the number of intersection points is exactly 9.)

9.3. The Space of Functions

Plane AG codes are evaluation codes in which we evaluate certain bivariate polynomials on a subset of $\mathbb{F}_q^2$. Based on the discussions in the last sections, the subset we will be considering will be the zeroset of an irreducible bivariate polynomial $f(x,y) \in \mathbb{F}_q[x,y]$. The polynomials we want to evaluate are bivariate polynomials that are not multiples of $f(x,y)$. Our first goal is to find a basis for the vector space of such polynomials.

Let $R := \mathbb{F}_q[x,y]/(f(x,y)\mathbb{F}_q[x,y])$ be the residue class ring modulo $f(x,y)$. The polynomials we want to evaluate are nonzero elements of R. To find a basis for R over $\mathbb{F}_q$, we make the following simplifying assumption. If d denotes the degree of f, we assume that $f(x,y) = ax^d + f_1(x,y)$, wherein the x-degree of f_1 is smaller than d, and a is nonzero. We can always assume w.l.o.g. that f is of this form. Otherwise, if $f(x,y)$ has a monomial of the form $x^{d-k}y^k$, then we replace y by $y + x$, and obtain the desired form for f. This only results in a change of basis in $\mathbb{F}_q^2$ and does not change anything else. In particular, after this transformation, the new f is irreducible as well.

Theorem 9.2. *Suppose that $f \in \mathbb{F}_q[x,y]$ is of degree d, and that $f = ax^d + f_1(x,y)$, where the degree in x of $f_1(x,y)$ is smaller than d and a is nonzero. Then for any polynomial $h \in \mathbb{F}_q[x,y]$ there is a unique element h_1 in*

$$\Gamma := \mathbb{F}_q[y] \oplus x\mathbb{F}_q[y] \oplus \cdots \oplus x^{d-1}\mathbb{F}_q[y],$$

such that $h - h_1$ is a multiple of f.

Proof. The assertion is very easily seen, when we consider all the polynomials as polynomials in x over $\mathbb{F} := \mathbb{F}_q(y)$.

First, we show that if for $h, g \in \Gamma$ we have that $h - g$ is divisible by f, then $h = g$. Considering these polynomials and f as polynomial in $\mathbb{F}[x]$, we see that their x-degree of $g - h$ is smaller than that of f, and hence f divides $g - h$ iff $g = h$.

Next, take any $h \in \mathbb{F}_q[x, y]$, interpret it as a polynomial in $\mathbb{F}[x]$, and perform polynomial division with remainder by f to obtain a remainder h_1. Since the "leading coefficient" of f (in x) is in $\mathbb{F}_q$, h_1 is an element of Γ (i.e., we never divide by a polynomial in $\mathbb{F}_q[y]$).

Putting everything together, any h has an h_1 in Γ, and this h_1 is unique since otherwise the difference of two non-equal elements in Γ would be divisible by f. □

The polynomials we would like to evaluate on the zeroset of $f(x, y)$ will have to come from Γ. We need to have a control on their degree, which, by Bézout's theorem, gives us control on the number of zeros of nonzero codewords, and hence on the minimum distance of the code. For this, we need to calculate the dimension of $\Gamma_{<m}$, the space of polynomials in Γ of degree less than m.

Theorem 9.3. *For $m \geq d - 1$, the dimension of the space of polynomials of degree $< m$ in Γ, called $\Gamma_{<m}$, is*

$$(m-1)d - \frac{(d-1)(d-2)}{2} + 1.$$

Proof. Note that we have

$$\Gamma_{<m} = \mathbb{F}_q[y]_{<m} \oplus x\mathbb{F}_q[y]_{<m-1} \oplus \cdots \oplus x^{d-1}\mathbb{F}_q[y]_{<m-d+1}.$$

This shows that if $m \geq d$, then the dimension of $\Gamma_{<m}$ is

$$\begin{aligned} \sum_{i=0}^{d-1}(m-i) &= md - \frac{d(d-1)}{2} \\ &= (m-1)d - \frac{(d-1)(d-2)}{2} + 1, \end{aligned}$$

and proves our assertion. □

9.4. AG Codes

Now we are ready to define plane AG codes. These are evaluation codes via $(M, S; \mathbb{F}_q)$.

We choose an irreducible polynomial $f(x, y) \in \mathbb{F}_q[x, y]$ of degree d, $f(x, y) = ax^d + f_1(x, y)$ with x-degree of f_1 smaller than d, and $a \neq 0$. The evaluation set S is equal to $\{(a, b) \in \mathbb{F}_q^2 \mid f(a, b) = 0\}$.

The space of functions M is equal to $\Gamma_{<m}$, and the evaluation is simply the evaluation of a polynomial at a point. The code obtained this way is called a *geometric Goppa code* or an *AG code*. We denote it by $C(f, m)$.

Theorem 9.4. *Let n be the number of $(a, b) \in \mathbb{F}_q^2$ such that $f(a, b) = 0$. If the degree of f is d, $m \geq d$, and $n > (m - 1)d$, then*

$$\dim C(f, m) = (m-1)d - \frac{(d-1)(d-2)}{2} + 1,$$

and the minimum distance of $C(f, m)$ is at least $n - (m - 1)d$.

Proof. If $h \in \Gamma_{<m}$ is nonzero, then by Bézout's Theorem the number of (a, b) such that $h(a, b) = f(a, b) = 0$ is at most $(m - 1)d$. Therefore, the image of the evaluation map applied to f has at most $(m - 1)d$ zeros, and thus at least $n - (m - 1)d$ nonzeros. If $n > (m - 1)d$, then the number of nonzeros is positive, and hence the evaluation map is injective. As a result, the dimension of $C(f, m)$ equals that of $\Gamma_{<m}$, which was determined in Theorem 9.3. By the same token, the minimum distance is at least $n - (m - 1)d$. □

The AG codes we obtain are $[n, k, d]_q$-codes with

$$k + d \geq n - \frac{(d-1)(d-2)}{2} + 1.$$

On the other hand, the Singleton bound asserts that $k + d \leq n + 1$, so that we can think of the term $(d - 1)(d - 2)/2$ as some kind of a *defect* of the code.

In order to obtain good AG codes, what we have to do is to find polynomials $f(x, y)$ that

1. are irreducible;
2. have a lot of zeros in $\mathbb{F}_q^2$, and
3. do not have a large degree.

Unfortunately, the second and the third goal above compete with one another: a deep theorem in algebraic geometry, called the *Hasse-Weil Bound*, asserts that the number of zeros of f is at most $q + (d-1)(d-2)\sqrt{q}$.

Even though we will be able to exhibit polynomials that achieve this bound, one should note that this approach will not yield asymptotically good codes, since the number of zeros of $f(x, y)$ will never exceed q^2. In the next lecture, we will show how to overcome this difficulty. For now, we will be content with some examples of plane AG codes.

9.5. Eisenstein's Irreducibility Criterion

In order to apply our results to specific examples, we need to test whether a given polynomial $f(x, y)$ is irreducible. There are a variety of sufficient conditions for irreducibility, of which we highlight one, called the *Eisenstein Criterion*.

First, let us discuss a trivial case in which $f(x, y)$ is reducible. We write $f(x, y) = \sum_{i=0}^{m} f_i(x)y^i$. Suppose that $f_0(x), \ldots, f_m(x)$ have a nontrivial gcd, say $h(x)$ of degree larger than 0. Then $f(x, y) = h(x)f_1(x, y)$ for some other polynomial $f_1(x, y)$, and $f(x, y)$ becomes reducible.

We say that a polynomial $f(x, y)$ is primitive (with respect to x), if the gcd $h(x)$ is trivial, i.e., if the $f_0(x), \ldots, f_m(x)$ are co-prime. Clearly, if we want $f(x, y)$ to be irreducible, we definitely want it to be primitive.

Theorem 9.5. *Suppose that $f(x, y) = \sum_{i=0}^{m} f_i(x)y^i$ is primitive with respect to x, and that there is an irreducible polynomial $p(x) \in \mathbb{F}_q[x]$ such that $p(x)$ divides $f_0(x), f_1(x), \ldots, f_{m-1}(x)$, but $p(x)$ does not divide $f_m(x)$, and $p^2(x)$ does not divide $f_0(x)$. Then $f(x, y)$ is irreducible.*

Proof. Suppose that $f(x, y)$ is reducible. Then

$$f(x, y) = \left(\sum_{i=0}^{\ell} h_i(x)y^i\right) \cdot \left(\sum_{j=0}^{m-\ell} g_j(x)y^j\right),$$

with $h_\ell(x)$ and $g_{m-\ell}(x)$ nonzero. Since $f(x, y)$ is supposed to be primitive with respect to x, both ℓ and $m - \ell$ are larger than 0. We have:

$$f_0(x) = h_0(x)g_0(x)$$

Since $p(x)$ is irreducible, it divides $f_0(x)$, and $p(x)^2$ does not divide $f_0(x)$, this polynomial either divides $h_0(x)$ or $g_0(x)$, but not both. W.l.o.g. assume that it divides $h_0(x)$. Then, since

$$f_1(x) = h_0(x)g_1(x) + h_1(x)g_0(x)$$

$p(x)$ also divides $h_1(x)$, since it divides $f_1(x)$ by assumption, and it does not divide $g_0(x)$. Continuing this way, we see that since

$$f_\ell(x) = h_0(x)g_\ell(x) + \cdots + h_\ell(x)g_0(x),$$

and $\ell < m$, $p(x)$ divides $f_\ell(x)$, and it divides $h_0(x), \ldots, h_{\ell-1}(x)$, by induction. Since it does not divide $g_0(x)$, it must divide $h_\ell(x)$. But

$$f_m(x) = h_\ell(x)g_{m-\ell}(x),$$

which shows that $p(x)$ divides $f_m(x)$, a contradiction. This shows that $f(x)$ is irreducible. □

9.6. Examples

9.6.1. Elliptic Curves

Consider the polynomial $x^2 + x - y^3 \in \mathbb{F}_2[x, y]$. Eisenstein's criterion shows that it is irreducible: consider the polynomial $p(x) = x$. It divides $x^2 + x$, but x^2 does not divide $x^2 + x$. It divides the coefficients of y and y^2 of the polynomial (both these coefficients are zero), but it does not divide the coefficient of y^3 which is 1. Moreover, the polynomial is primitive.

What kind of a code do we obtain from this? The number of zeros of this polynomial is 2: the zeros consist of the points $(0, 0)$, $(1, 0)$, so this will hardly give us a reasonable code over $\mathbb{F}_2$.

However, if we change the field to $\mathbb{F}_4$, something remarkable happens. Suppose that $\mathbb{F}_4 = \{0, 1, \omega, \omega^2\}$, where $\omega^2 + \omega + 1 = 0$. Then the zeros of f consist of the 8 points

$$\begin{array}{ccc} (0,0) & (1,0) & \\ (\omega,1) & (\omega,\omega) & (\omega,\omega^2) \\ (\omega^2,1) & (\omega^2,\omega) & (\omega^2,\omega^2) \end{array}$$

This gives us $[8, 3(m-1), 8-3(m-1)]_4$-codes for $m \geq 2$. In other words, we obtain an $[8, 3, 5]_4$-code and an $[8, 6, 2]_4$-code.

9.6.2. Hermitian Curves

This example is a generalization of the last one. Consider the field $\mathbb{F}_{q^2}$, and the polynomial

$$f(x, y) = x^q + x - y^{q+1}.$$

The zeroset of this polynomial is called a *Hermitian Curve*. In the same way as above, we can show that $f(x, y)$ is irreducible.

How many zeros does f have? We use the fact that the mapping $x \mapsto x^q + x$ is the *trace* of $\mathbb{F}_{q^2}$ over $\mathbb{F}_q$, and that the mapping $y \mapsto y^{q+1}$ is the *norm*, which means that it maps $\mathbb{F}_{q^2}$ to $\mathbb{F}_q$. This implies that for any $y \in \mathbb{F}_{q^2}$, there are exactly q values of x such that $x^q + x = y^{q+1}$: given y, the value of y^{q+1}, which is the norm of y, is in $\mathbb{F}_q$, and hence, since the trace map is surjective, there are exactly q values for x such that $x^q + x = y^{q+1}$.

The degree of f is $q + 1$, so putting things together, we see that for any $m \geq q$ and $(m-1)(q+1) < q^3$ we obtain a code with parameters

$$\left[q^3, (m-1)(q+1) - \frac{q(q-1)}{2} + 1, \geq q^3 - (m-1)(q+1)\right]_{q^2}.$$

Note that a Reed-Solomon code would yield codes of length at most q^2, so the length of the codes we have constructed is considerably larger than that of Reed-Solomon codes, while for large q the relative minimum distance and the rate of these codes behaves almost like those of an MDS-code. (The rate is roughly $m/q^2 + O(1/q)$ and the relative minimum distance is roughly $1 - m/q^2$.)