Greg Kuperberg's Lectures on

Introduction to Algebraic Geometry Codes

Sanchayan Dutta
dutta@ucdavis.edu

Contents

1 Lecture 1 (20 November 2022)

1.1 Introduction

The point in the affine case is to look at polynomial functions on algebraic curves. Just the field itself is an example of an algebraic curve. If you look at polynomial functions on the field itself, that's a polynomial in one variable, because say a straight line is an example of a curve. This is easier to visualize when using $\mathbb{R}$ or $\mathcal{C}$, except such fields are not usually used for error correction.

An algebraic curve over $\mathbb{R}$ is a curve in the sense of high school math – for instance, a straight line. Even in the case of finite fields you can still conceptualize them as curves. What is the next example of an algebraic curve, for ordinary geometry over the reals – it would be a conic section, which is a solution set to a quadratic equation. Over a finite field, $\mathbb{F}_q$, you can look at solutions in $\mathbb{F}_{q^2}$ and that is a finite field analogue of a conic. For a small enough field, like F_4, every subset is a conic – it's a degenerate case. However, in a large enough field, say F_{23} then the plane over F_{23} has 23^2 points – there only certain finite subsets are conics.

After quadratic curves, you can look at cubic curves. For one thing, cubics can have inflection points unlike conics. A cubic over the reals can be of two types – one connected and one disconnected. Still, cubic curves are much more restrictive than all curves. Cubic curves on the plane are called elliptic curves.

1.2 Lagrange Interpolation

Take the simplest curve – an affine line. To make codes, we're studying polynomials upto some degree on $\mathbf{F}_q$ itself. The domain only has q points. It follows from Lagrange interpolation that every possible function is represented by a polynomial of degree $q-1$ or less. In such a case, we consider polynomials upto some lesser degree. That's exactly the definition of complete unextended Reed-Solomon code. (Reed-Solomon codes were discovered in the 1960s and Goppa codes were discovered in 1980s – around 20 years later.)

1.3 Constructing AG Codes

Start with a finite field $\mathbb{F}_7$. A function from $\mathbb{F}_7 \to \mathbb{F}_7$ has 7 values. If it a preferred function, then you just list the 7 values and that's the codeword. Which functions should be favoured? What is the dimension of the space of quadratic functions from F_7 to itself? It's 3 because a quadratic function has 3 coefficients. The codewords are indexed by F_7. The message space consists of 7 lettered words from F_7 and this message space contains a code subspace that's 3-dimensional over F_7. You need to plug in the values in some order. So you get a 3-dimensional code in 7-dimensional wordspace.

One thing that's remarkable is that it's remarkably easy to calculate the minimum distance. What is the minimum distance? It's the same as the minimum weight as it's a linear weight. It's governed by the number of zeroes it can have. So the minimum weight in the quadratic case is $7-2=5$. The codewords are simply the quadratic functions (due to the Lagrange interpolation theorem – the evaluation of each quadratic can be uniquely identified with the codeword list).

Getting a minimum distance of 5 for a 7-dimensional wordspace is miraculous – in fact, it's the best that you can do.

1.4 The simplest algebraic (curve) codes

Take the functions $f\colon F_q \to F_q$ given by polynomials of degree d. List all of the values of each f (using a fixed ordering of F_q). Call each list a codeword, and the set of all them a code $C_{q,d}$.

1) What is the length of $C_{q,d}$? q.

2) What is the dimension of $C_{q,d}$? $d+1$

3) What is the minimum distance? $q-d$

Theorem: Each such code is optimal!

1.5 Permuting letters of the codewords

Does the indexing order of the list by the elements of F_q matter? Does permuting the letters of the codewords matter?

Warmup question: Let A be a finite alphabet and let C in A^n be some code of length n. $|C|$ is something, mindist$C$ is something. C may or may not be linear if A is a finite field.

Define C' from C by swapping the jth and kth letters simultaneously for all w in C, for some $1 \leq i, j \leq n$. How it C' (probably) different and how is it necessarily the same?

For a start, same number of codewords and same minimum distance. The codewords themselves are usually different but it's still an equivalent code. Swapping the jth and kth factors is both an isometry of A^n (in Hamming distance) and a linear automorphism if A is a field. So C' is an equivalent code but it doesn't equal C as a set. (So even saying we need an ordering of $\mathbb{F}_q$ is a red herring. What's sneaky is that $\mathbb{F}_q$ is being used for the alphabet as well as the indexing.)

Theorem: Each such code is optimal! (Due to the singleton bound.)

1.6 Generalization

Let S be a subset of $\mathbb{F}_q$ of some size $|S| = k \leq q$. Then $C_{k,q,d}$ can be constructed the same way, as lists of values of polynomials $f\colon S \to \mathbb{F}_q$ of degree $\leq q$, and (theorem) it is still optimal.

1) Length of $C_{k,q,d}$? k

2) Dimension? $d+1$

3) Minimum distance? $k-d$

These are all possibly somewhat generalized Reed-Solomon. More careful name: $\mathrm{RS}_{q,S,d}$.

Something goes very wrong in the first version, even for just questions (1) and (2), when $d \geq q$.

Theorem on evaluation of polynomials:

Let $\mathbb{F}$ be a field. $\mathbb{F}[x]$ is the algebra of all polynomials (all degrees). $\mathrm{Fun}(\mathbb{F}, \mathbb{F})$ is the algebra of all functions from $\mathbb{F}$ to $\mathbb{F}$. If p in $\mathbb{F}[x]$ is a polynomial, then $p(x)$ is the associated function, or element of $\mathrm{Fun}(\mathbb{F}, \mathbb{F})$. Call this map $\mathrm{ev}\colon \mathbb{F}[x] \to \mathrm{Fun}(\mathbb{F}, \mathbb{F})$.

Theorem:

(1) If $\mathbb{F}$ is infinite, then ev is injective and not surjective.

(2) If $\mathbb{F}$ if finite, then ev is surjective and not injective.

Taking $\mathbb{F} = \mathbb{F}_q$ (it's a PID), there's a single polynomial generator. $\ker \mathrm{ev} = (x^q - x)$. So the polynomials of degree $\leq d = q-1$ represents all functions in $\mathrm{Fun}(\mathbb{F}_q, \mathbb{F}_q)$, and any higher degree is redundant.

Actually you can take $\mathrm{ev}_{S,\mathbb{F}}$ from $\mathbb{F}[x]$ to $\mathrm{Fun}(S, \mathbb{F})$. This is injective and not surjective whenever S is infinite. It is surjective and not injective whenever S is finite. Howework problem: What is $\ker \mathrm{ev}_{S,F}$?

1.7 Code equivalences

(a) Given A^n (Hamming space), permuting the index set $\{1, \dots, n\}$ is an isometry and thus a source of code equivalences.

(b) Permuting A *separately for each position* is also an isometry and thus a code equivalence.

Theorem: The two together are all of $\mathrm{Isom}(A^n)$.

Start of proof. When $A = \mathbb{Z}/2$, let $f\colon (\mathbb{Z}/2)^n \to (\mathbb{Z}/2)^n$ be an isometry. Using (b), WLOG, $f(0 \dots 0) = 0 \dots 0$. Since f is an isometry, the n neighbours of $0 \dots 0$, namely the standard basis of $(\mathbb{Z}/2)^n$ are permuted in some way. Using (a), we can simplify f further and assume WLOG $f(\text{any standard basis vector}) = \text{itself}$.

Now (lemma) If f is of this restricted WLOG form, then actually f is the identity. Proof is by induction on weight. E.g. $f(1100 \dots 0) = \text{itself}$, because it is the only point at distance 2 from 0 which is a neighbour of both $1000 \dots 0$ and $0100 \dots 0$, all three of which are fixed.

In the linear case, if $A = \mathbb{F}_q$, we are interested in the linear isometries of $\mathbb{F}_q^n$ rather than all of them. (a) is linear. (b) is linear when each of the n permutations of $\mathbb{F}_q$ is multiplication by a scalar. The group of linear isometries and thus the group of monomial matrices in $M(n, \mathbb{F}_q)$, the matrices with one non-zero entry in each row and column.

Theorem: The dual of $\mathrm{RS}_{q,S,d}$ is equivalent to Reed-Solomon. This is a type of self-dual.

Q: The topic is Reed-Solomon. Jeez, I don't remember. Are we using polynomials for the codewords or the parity checks? The answer is yes.

1.8 The one limitation of Reed-Solomon

The only limitation is code length, which is at most q when $A = \mathbb{F}_q$. Otherwise, this is optimal because it matches the Singleton bound (which is a non-existence bound). It is a relatively crude upper bound on the size of an arbitary code block C with block length n, size M and minimum distance d (d stands for dimension here). The proof is similar to the RS construction, which uses the pigenhole principle.

However, there is a way to increase the max length by 1, which is the extended Reed-Solomon. This is a very important but desperate moment. We'll run out after this ...

1.9 Extended Reed-Solomon

FP^1 is by definition the set of slopes of lines in F^2, for any field F. As a set, $FP^1 = F \cup \{\infty\}$. $\mathbb{R}P^1$ has the topology of a circle. $\mathbb{C}P^1$ has the topology of a sphere, namely, the Riemann sphere.

$|\mathbb{F}_q P^1| = q + 1$.

The elegant construction of extended Reed-Solomon: $P^1\mathbb{F}$ is $\mathbb{F}^2 \setminus \{0, 0\}$, up to the equivalence $(x, y) \sim (ax, ay)$. Now choose a section S of this equivalence, i.e., one non-zero representative (x_k, y_k) of each line. We want a favourable class of functions $f\colon S \to \mathbb{F}_q$, and we use *homogeneous* polynomials $f(x, y)$ of degree d. Get a code $\mathrm{ERS}_{q,S,d}$.

1) Length of $\mathrm{ERS}_{q,S,d}$: $|S| = q + 1$

2) Dimension: $d + 1$

3) Minimum distance: $q + 1 - d$

It looks like this construction depends on S. However, theorem, all choices of S yield equivalent codes. (All linearly (b) equivalent.)

1.10 The Goppa Construction

Same as ERS, but use a projective curve X in $\mathbb{F}_qP^t$ for some $t \geq 1$, often $t = 2$. You still need a section but only for points in X. So X replaces $\mathbb{F}_qP^1$. Actually, there is a subtlety, given that the ambient home F_qP^t of X is negotiable, can always (for a fixed X) assume $d = 1$. X is a "curve" means that it is the set of all solutions to a system of homogeneous equations in $t + 1$ variables. The legit way to do this is to take homogenous polynomials the entire time.

We're taking polynomials of degree d and there's the subtlety, you can take what the same curve and put it in the same or larger projective space, so that you only ever really have to look at polynomials of degree 1. There's a more abstract language for that called *sections of line bundles*.

All that an AG code could ever be is that X may not be a curve, but some other variety. That idea is due to Goppa. However, Goppa knew more than that – counterintuitive result that 1-dimensional algebraic varieties are often at least as powerful as higher dimensional varieties. AG often turns out to be useful in CS, but 1-dimensional most often suffices. Like in cryptography, elliptic curves is the 1-dimensional case, called Diffie-Hellman.