Introduction to Coding Theory **CMU: Fall 2018**

Lecture 6: Algebraic Geometric Codes

October 5, 2018

Lecturer: Venkatesan Guruswami *Scribe: Taisuke Yasuda*

1 Introduction

1.1 Approaching the Singleton Bound

So far, we have studied two cases very nicely – large alphabet (Reed-Solomon) codes and binary (concatenations, expander-based) codes. These are good to start with, but what can we say about a general fixed alphabet size q? Focusing on the rate vs distance trade off, can we get codes like Reed-Solomon codes that are close to the Singleton bound with $R \approx 1 - \delta$ over fixed alphabets?

First note that the Gilbert-Varshamov (GV) bound gives us that $R \geq 1 - h_q(\delta) \approx 1 - \delta - O(1/\log q)$. We are off from the Singleton bound by about $1/\log q$, so we can get within ε of this bound with code constructions achieving the GV bound by taking $q = q(\varepsilon) \leq \exp(1/\varepsilon)$. Of course, this code construction will not be explicit, which is a problem as usual. Explicitly, we can get for instance from expander-based codes a construction with $q(\varepsilon) \leq (1/\varepsilon)^{O(1/\varepsilon^2)}$. Although this is not great, it at least matches the random construction.

In terms of the lower bound on $q(\varepsilon)$, we have from the Plotkin bound that for a q-ary code, we have $R \leq 1 - \frac{q}{q-1}\delta + o(1)$. Thus, to be within ε of the Singleton bound, we must have at least $q = \Omega(1/\varepsilon)$. However, this lower bound is still very far from the exponential upper bounds that we have seen.

Today, we will see that algebraic geometric (AG) codes will allow us to achieve $q(\varepsilon) \leq O(1/\varepsilon^2)$. Thus, it not only beats the random construction bound, it crushes it. Note however that for a tight characterization of $q(\varepsilon)$, we are still a factor of $1/\varepsilon$ off, and closing this gap is an open problem to this day.

1.2 History

The general framework of algebraic gemoetric codes was introduced by Goppa [Gop81], but he did not have constructions. Later, Tsfasman, Vladut, and Zink [TVZ82] showed the existence of AG codes attaining a rate of $R \geq 1 - \delta - 1/(\sqrt{q} - 1)$ with an alphabet size of $q = p^{2\ell}$ for p prime. Note that for $\delta = 1/2$ and $q \geq 44$, we have that $\delta + 1/(\sqrt{q} - 1) \leq h_q(\delta)$. Thus, at $q \geq 49$, we get codes beating the GV bound. Garcia and Stichtenoth then obtained concrete constructions in the mid 90s [GS95, GS96], which [SAK+01] later improved to an $\tilde{O}(n^3)$ time generator matrix construction.

Although AG codes may be intimidating at first, they are useful as black boxes and achieve remarkable bound not known in any other way. It is good to at least know how they look, so we give a simple presentation today.

2 Algebraic geometric codes

2.1 Motivation: bivariate polynomials

Given the success of Reed-Solomon (RS) codes, it is reasonable to attempt at better codes using algebraic objects. Recall that in RS codes, we constructed codes by mapping polynomials to their evaluations, i.e. $f \mapsto (f(p_1), \ldots, f(p_n))$. The problem here was that the length of this code was limited to the alphabet size q. A simple idea for increasing the number of evaluation points is to increase the dimension by considering bivariate polynomials, so that we may evaluate the polynomial at q^2 of $\mathbb{F}^q \times \mathbb{F}^q$. Let us evaluate the parameters of this new code in comparison to RS codes. RS codes were $[q, k, q-k+1]_q$ codes, and it is easy to see that our bivariate polynomial code is the tensor concatenation of this RS code with itself, so we have a $[q^2, k^2, (q-k+1)^2]_q$ code. Note that we've succeeded in lengthening the code, but in order to approach the Singleton bound, we wanted to achieve a distance of $q^2 - k^2$, and what we got was

$$(q-k+1)^2 \approx (q^2 - k^2) - 2k(q-k).$$

So, this is already off from the Singleton bound: we obtained a longer code, but that distance is already too bad.

The good idea here was to go to higher dimensions for longer codes, but we run into trouble if we implement this idea too simply. As an intuition for why RS codes met the Singleton bound but bivariate polynomials did not, consider what happens when we know that some symbols of an RS codeword are 0. Then as long as we set k symbols to 0, no other symbols were forced to be 0. In other words, each symbol set to 0 lead to an independent constraint, up to k symbols, and this is what lead to an optimal Singleton bound-achieving code. However, this property does not carry over to bivariate polynomials: when we take a particular evaluation of a row and know that some of its symbols are 0, then we in fact know that other entries of the codeword are also 0.

2.2 Preliminaries

The above example motivates us to use multivariate polynomials, but choosing all q^2 points available to us to evaluate the polynomial was a bad idea. The idea then is to choose a special set of evaluation points $S \subset \mathbb{F}_q^2$ such that no message polynomial f intersects S too much. This is where the idea of using algebraic geometry comes in. To execute the above idea, we will pick our evaluation point set S itself to be a curve, i.e. the set of roots of a polynomial $g(X, Y) \in \mathbb{F}_q[X, Y]$. The idea then is that we will construct g so that it does not intersect any other curve f too much, using techniques of algebraic geometry. Of course, one could choose a polynomial with high enough degree so that every point in $\mathbb{F}_q^2$ is a root. Thus, the idea here is to use a *low degree* polynomial g to construct the evaluation points S.

We now introduce a theorem of algebraic geometry.

Theorem 2.1 (Bézout's theorem (in the plane))**.** *Let $f, g \in \mathbb{F}_q[X, Y]$ have no common factors. Then, the size of the set $\{(\alpha, \beta) \in \mathbb{F}_q^2 : f(\alpha, \beta) = g(\alpha, \beta) = 0\}$ is at most the product of degrees. That is, if f has degree D_1 and g has degree D_2 and they share no common factors, then they have at most $D_1 D_2$ common zeros.*

It is easy to see that the theorem is tight, since for any $T, S \subseteq \mathbb{F}_q$ the polynomials

$$\prod_{\alpha \in T}(X - \alpha), \qquad \prod_{\beta \in S}(Y - \beta)$$

share no common factors, have degree $|T|$ and $|S|$ each, and have $|T||S|$ common zeros at all $(\alpha, \beta) \in T \times S$.

We will crucially use Bézout's theorem to obtain a code with good distance. To design our good code, we roughly first set our messages to be $M = \{f \in \mathbb{F}_q[X, Y] : \deg f < k\}$. We then want to choose g, the polynomial specifying our evaluation set S, such that g shares no factors with any $f \in M$. If we take g to be irreducible, then we may accomplish this by requiring $g \nmid f$, which in turn can be guaranteed by requiring $\deg(g) > \deg(f)$. Once this is done, we may say by Bézout's theorem that if we restrict our attention to the set S of the roots of g, f may only have at most $\deg(f)\deg(g)$ zeros and thus the minimum weight is at least $n - \deg(f)\deg(g)$. Note that this introduces a tension in the choice of our parameters, since we now want $\deg(g)$ to be both large and small.

2.3 Hermitian codes

We will see the above ideas in action in Hermitian codes. We set $n = q^{3/2}$ over the alphabet $\mathbb{F}_q$ with $q = r^2$, a degree 2 extension of a prime field. This will result in a $\left[n = r^3, k, n - k - \frac{r(r-1)}{2} + 1\right]_{q=r^2}$ code. Note that this backs off from the Singleton bound by $r(r-1)/2$, but this is asymptotically small since $r(r-1)/2 = \Theta(n/\sqrt{q})$. Thus, we will achieve a rate of $R \geq 1 - \delta - \Theta(1/\sqrt{q}) = 1 - \delta - o(1)$.

2.3.1 Field theory preliminaries

Recall the field norm and field trace. For the field $\mathbb{F}_r$ and its finite extension $\mathbb{F}_q = \mathbb{F}_{r^2}$, we have that the norm and trace, respectively, are

$$N(X) = N_{\mathbb{F}_q/\mathbb{F}_r}(X) = X^{r+1}$$
$$\mathrm{Tr}(X) = \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_r}(X) = X^r + X.$$

We now introduce the Hermitian curve.

Definition 2.2 (Hermitian curve)**.** *Consider the bivariate polynomial* $g(X, Y) = N(Y) - \mathrm{Tr}(X)$*. Then, we define the* Hermitian curve *to be*

$$S = \left\{(\alpha, \beta) \in \mathbb{F}_r^2 : g(\alpha, \beta) = \mathrm{Tr}(\beta) - N(\alpha) = 0\right\}.$$

As the notation suggests, this will be our choice for the evaluation set. Furthermore, the g above is irreducible, as we wanted previously. We will not prove this and take the fact for granted.

Fact 2.3. *The bivariate polynomial* $g(X, Y) = \mathrm{Tr}(Y) - N(X)$ *is irreducible.*

Finally, to set up for our analysis of this code, we recall some facts about the norm and the trace.

Fact 2.4. *The following hold for the trace and norm:*

- *We have* $\mathrm{Tr}(\alpha+\beta) = \mathrm{Tr}(\alpha) + \mathrm{Tr}(\beta)$ *and* $N(\alpha\beta) = N(\alpha)N(\beta)$.
- *A priori,* Tr, N *are maps from* $\mathbb{F}_q \to \mathbb{F}_q$. *In fact, it can be shown that they actually only take values in* $\mathbb{F}_r$.

  Proof. We have that

  $$\mathrm{Tr}(X)^r = (X^r + X)^r = X^q + X^r = X + X^r = \mathrm{Tr}(X)$$

  and

  $$N(X)^r = (X^{r+1})^r = X^{q+r} = X^{1+r} = N(X)$$

  so $\mathrm{Tr}(X), N(X) \in \mathbb{F}_r$. □

- *For all* $\gamma \in \mathbb{F}_r$, *there are exactly* r *values of* $\alpha \in \mathbb{F}_q$ *such that* $\mathrm{Tr}(\alpha) = \gamma$. *Similarly, for all* $\gamma \in \mathbb{F}_r^*$, *there are exactly* $r+1$ *values of* α *such that* $N(\alpha) = \gamma$.

2.3.2 The construction

We finally introduce the construction of the Hermitian code. Let S be the Hermitian curve, i.e.

$$S = \left\{(\alpha, \beta) \in \mathbb{F}_q^2 : \beta^r + \beta = \alpha^{r+1}\right\}.$$

Note that for each $\alpha \in \mathbb{F}_q$, there are exactly r values of $\beta \in \mathbb{F}_q$ such that $\beta^r + \beta = \alpha^{r+1}$. Since there are r^2 choices of α, we have that $|S| = r^3$.

Define also the message space $M \subseteq \mathbb{F}_q[X, Y]$ to be all polynomials of X degree at most r and total degree at most D for D to be specified, i.e.

$$M = \left\{ f \in \mathbb{F}_q[X,Y] : f(X,Y) = \sum_{j=0}^{r} X^j h_j(Y), \deg(h_j) \le D - j \right\}$$

We now have a complete specification of the Hermitian code.

2.3.3 Analysis

We now compute the parameters of the Hermitian code.

Proposition 2.5. *The Hermitian code has distance at least* $n - D(r+1)$.

Proof. Note that we have set the degree of X in $f \in M$ to be strictly less than $r+1$. Furthermore, $g(X,Y) = \mathrm{Tr}(Y) - N(X)$ is irreducible, so $g \nmid f$ and thus g and f do not share factors. Thus, by Bézout's theorem, g and f share at most $D(r+1)$ common zeros and thus the codeword $\langle f(\alpha,\beta)\rangle_{(\alpha,\beta)\in S}$ has weight at least $n - D(r+1)$. □

Proposition 2.6. *The Hermitian code is a* $\left[n = r^3, k, n - k - \frac{r(r-1)}{2} + 1\right]_{q=r^2}$ *code.*

Proof. We first must compute the dimesion k. This is the number of monomials we are allowed in the message space. Counting by the degree of h_j, we have

$$k = \sum_{j=0}^{r} (D - j + 1) = (D+1)(r+1) - \frac{r(r+1)}{2}.$$

Solving for D gives

$$D = \frac{k + r(r+1)/2}{r+1} - 1$$

so distance is

$$n - D(r+1) = n - k - \frac{r(r+1)}{2} + r + 1 = n - k - \frac{r(r-1)}{2} + 1 \qquad \square$$

We have thus given the full construction and analysis for the Hermitian code, modulo the proof of Bézout's theorem and the irreducibility of g.

2.4 Extensions of the Hermitian code

Hermitian codes are far from the end of the story of AG codes. We will briefly discuss how one obtains longer codes from generalizations of the above construction. The way one obtains longer codes is to recursively execute the above scheme, producing what is known as the Hermitian tower. In this direction, one may consider a multivariate message space $M \subset \mathbb{F}_q[X_1, X_2, \ldots, X_m]$ and an evaluation set $S_m \subseteq \mathbb{F}_q^m$ where

$$\mathrm{Tr}(X_2) = N(X_1), \mathrm{Tr}(X_3) = N(X_2), \ldots, \mathrm{Tr}(X_m) = N(X_{m-1})$$

of size $|S_m| = r^{m+1}$.

In fact, the above does not quite work, and what one actually uses is the Garcia-Stichtenoth tower [GS95, GS96]. This construction is essentially the same thing, where one instead considers the polynomial equations

$$\mathrm{Tr}(X_{i+1}) = \frac{N(X_i)}{\mathrm{Tr}(X_i)}$$

for $i \in [m]$ with a message space of rational functions. This construction indeed achieves the desired rate of $R \geq 1 - \delta - 1/(\sqrt{q} - 1)$ with much longer codes.

Note that our degree arguments no longer work with this new message space, since we are working with rational functions instead of polynomials. In order to carry out analgous analyses, one introduces the notion of the *order* $\mathrm{ord}(f)$ of a rational function f. Roughly speaking, $\mathrm{ord}(f)$ is the number of poles of f at some point of a curve, and behaves much like the degree of a polynomial. For instance, it satisfies the following properties:

- $\mathrm{ord}(fh) = \mathrm{ord}(f) + \mathrm{ord}(h)$
- $\mathrm{ord}(\alpha f + \beta h) = \max\{\mathrm{ord}(f), \mathrm{ord}(h)\}$
- if f is zero on $\mathrm{ord}(f) + 1$ points of S, then $f \equiv 0$

Looking at the last property, if we take our message space to be $M = \{f : \mathrm{ord}(f) < \ell\}$, then this already gives us a code with distance at least $n - \ell$. Furthermore, when one introduces the notion of a *genus* g, one then finds that the dimension of the code is $\dim(\mathcal{C}) \geq (\ell+1) - g$. For some vague idea of the genus is, the genus for the Hermitian curve is $r(r-1)/2$.

References

[Gop81] Valerii Denisovich Goppa. Codes on algebraic curves. In *Soviet Math. Dokl.*, volume 24, pages 170–172, 1981.

[GS95] Arnaldo Garcia and Henning Stichtenoth. A tower of artin-schreier extensions of function fields attaining the drinfeld-vladut bound. *Inventiones mathematicae*, 121(1):211–222, 1995.

[GS96] Arnaldo Garcia and Henning Stichtenoth. On the asymptotic behaviour of some towers of function fields over finite fields. *Journal of number theory*, 61(2):248–273, 1996.

[SAK+01] Kenneth W Shum, Ilia Aleshnikov, P Vijay Kumar, Henning Stichtenoth, and Vinay Deolalikar. A low-complexity algorithm for the construction of algebraic-geometric codes better than the gilbert-varshamov bound. *IEEE Transactions on Information Theory*, 47(6):2225–2241, 2001.

[TVZ82] Michael A Tsfasman, SG Vlădutx, and Th Zink. Modular curves, shimura curves, and goppa codes, better than varshamov-gilbert bound. *Mathematische Nachrichten*, 109(1):21–28, 1982.