* Curves for codes

Master Course, Spring term 2025

Fernando Chamizo https://matematicas.uam.es/~fernando.chamizo/

Contents. Riemann-Roch. Algebraic geometry codes. Reed-Solomon codes

2.1 It is only Riemann-Roch

A *divisor* on a compact Riemann surface S is a formal finite sum

$$D=\sum n_P P \qquad \text{with} \quad n_p\in\mathbb{Z} \text{ and } P\in S.$$

Divisors can be added, with an abelian group structure (we write 0 for the identity element), and multiplied by integers in the obvious way. The *degree* is defined as $\deg D=\sum n_P$. A not identically zero meromorphic function f on S has an associate divisor

$$\operatorname{div}(f)=\sum_{P\in S} n(f,P)P$$

with $n(f,P)$ the order of f at P. This is said to be a *principal divisor*. When appending 0 to the principal divisors it is obtained a subgroup fo the divisors. The argument principle implies that $\deg\operatorname{div}(f)=0$ because S is compact. It is not true in general that a divisor with $\deg D=0$ is a principal divisor. In fact, *Abel-Jacobi theorem* shows that the quotient of both groups has, in some sense, the geometric structure of a torus.

A differential form ω on S (a 1-form) has a local expression $\omega=f\,dz$ and then it is possible to assign to it a divisor K. This is called a *canonical divisor*. Given one of them K, the rest are of the form $K+\operatorname{div}(f)$ because the "quotient" of two differential forms makes sense and it is a function in this 1-complex dimensional setting. In particular, $\deg K$ is uniquely determined.

The inequality $D_1\geq D_2$ for two divisors means that $D=D_1-D_2$ has all its n_P greater or equal than 0. The condition $\operatorname{div}(f)\geq 0$ implies that f is constant because f is a function from S to $\widehat{\mathbb{C}}=\mathbb{C}\cup\{\infty\}$ and nonconstant maps between Riemann surfaces are surjective.

Riemann-Roch theorem imposes some restrictions to choose meromorphic functions with $n(f,P)$ bounded. Consider for a fixed divisor D the set

$$\mathcal{L}(D)=\{0\}\cup\big\{f \text{ meromorphic } \not\equiv 0 \,:\, \operatorname{div}(f)+D\geq 0\big\}.$$

It is a linear space. Let $\ell(D)$ be its dimension. According to the previous comment, for a principal divisor $\mathcal{L}(D)=\mathbb{C}$ and $\ell(D)=1$.

Theorem 2.1 (Riemann-Roch)**.** *If S is a compact Riemann surface of* genus *g and K is a canonical divisor, then*

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

For a proof, see [1]. Geometrically the genus is the number of handles of the surface (e.g., $g = 0$ for the sphere and $g = 1$ for the torus) and analytically the number of linearly independent holomorphic differentials.

Corollary 2.2. *We have $\ell(K) = g$,* $\deg K = 2g-2$ *and $\ell(D) = \deg D+1-g$ if* $\deg D > 2g-2$.

Proof. Taking $D = 0$ and using $\ell(0) = 1$, we have $\ell(K) = g$. Substituting this in the theorem with $D = K$, it is deduced $\deg K = 2g - 2$. Finally, if $\deg D > 2g - 2$ then $\deg(K - D) < 0$ implies $\mathcal{L}(K - D) = \{0\}$, because $\deg \operatorname{div}(f) > 0$ cannot hold, and $\ell(D) = \deg D + 1 - g$ follows. □

If we review the proof of the formula for $\dim \mathcal{M}_k$, we will see that the only case that cannot deduced from previous arguments is $k = 2$. This case can be settled with Riemann-Roch theorem because if $f \in \mathcal{M}_2$ then $f\,dz$ is a holomorphic differential form on the Riemann surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}^*$, which has $g = 0$ because it is homeomorphic to a sphere. But Corollary 2.2 implies $\deg K = -2$ hence such holomorphic differential forms cannot exist.

Compact Riemann surfaces are the same as projective (nonsingular) algebraic curves over $\mathbb{C}$. It turns out that Theorem 2.1 is also true for any curve of this kind over an algebraically closed field [4], in particular over $\overline{\mathbb{F}}_p$, the algebraic closure of $\mathbb{F}_p$. The finite subextensions of $\overline{\mathbb{F}}_p/\mathbb{F}_p$ give all the *finite fields*. There is exactly one (up to isomorphisms) for each degree [5, §20]. The standard notation is $\mathbb{F}_{p^n}$ to indicate the field with $[\mathbb{F}_{p^n} : \mathbb{F}_p] = n$. As a linear space, $\mathbb{F}_{p^n}$ is isomorphic to $\mathbb{F}_p^n$. To carry out computations it is convenient to represent $\mathbb{F}_{p^n}$ as $\mathbb{F}_p[x]/(Q)$ with $Q \in \mathbb{F}_p[x]$ any irreducible polynomial of degree n. In cryptography, very often the messages are composed by blocks in finite fields.

2.2 Messages and vector spaces

Before worrying about the evil Eve, sometimes Alice and Bob have to deal with a more immediate (a perhaps more malicious) problem: an unreliable channel. This problem encourages to encode messages with some redundancy to detect errors and correct them, if possible. This is the topic of the area called *coding theory* ([2] and [7] are readable references).

A common setting is that of *linear codes*. With it, each message is encoded as a vector of a linear subspace $V \subset \mathbb{F}_q^n$ with $\dim V = k$. This linear space has some equations, in other words, there exists a linear endomorphism $H : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^n$ such that $V = \mathrm{Ker}(H)$. We test if $\vec{c} \in \mathbb{F}_q^n$ is a licit encoding of a message with $H(\vec{c}) = \vec{0}$. By this reason, the matrix of H is called the *check matrix*. If n is much larger than k, we are introducing a lot of redundancy and $H(\vec{c}) = \vec{0}$ is a very good test (albeit not 100% safe) to know if $\vec{c}$ is error free.

If $H(\vec{c}) \neq \vec{0}$ we can assure that the defective channel has caused one or several errors and we would like to correct them. In this context, a natural concept is the *Hamming distance*, when applied to $\vec{v}$ and $\vec{w}$ it counts the number of coordinates with $v_j \neq w_j$. For instance, the Hamming distance between $(0, 1, 0)$ and $(1, 0, 0)$ is 2.

The *code distance* of V is the minimum Hamming distance between different vectors in V. It is easy to see that the definition is equivalent to count the minimal number of nonzero coordinates for a nonzero vector in V (less obvious is that it coincides with the minimum number of linearly dependent columns of the check matrix). For any $\vec{x} \in \mathbb{F}_q^n$ there is at most one vector of V at Hamming distance less than $d/2$. In this way, choosing the nearest element of V,

$$\text{code distance} = d \quad \Longrightarrow \quad \text{It is possible to correct } \lfloor (d-1)/2 \rfloor \text{ errors.}$$

In the literature a linear code is identified with pair $(V, \mathbb{F}_q^n)$. The values of n and $k = \dim V$ are called the *code length* and the *code dimension* by obvious reasons. It is said to be an $[n, k, d]$ code (or an $[n, k, d]_q$ code).

A rather trivial example of linear code is $V = \{(0,0,0), (1,1,1)\} \subset \mathbb{F}_2^3$. In plain terms it consists in repeating each bit three times. This is a $[3, 1, 3]$ code, consequently it allows to correct one error. For instance, if Alice sends $(0,0,0)$ and Bob receives $(0,1,0)$ because the channel is defective, assuming that there is only an error, it can be corrected to $(0,0,0)$ using the simple rule "majority wins".

In general, one would like codes with small redundancy, represented by $n-k$, and allowing to correct several errors. There is a limitation in this goal given by the *Singleton bound* (see Exercise 1 for a proof)

$$d \leq n - k + 1 \qquad \text{for any } [n, k, d] \text{ code.} \tag{1}$$

There exist codes reaching the equality.

2.3 Messages on curves

During the 80's of the last century some linear codes related to algebraic curves were developed. They gained popularity among algebraic geometers and number theorists who witnessed the relevance of some abstract results in their areas. Although nowadays there is active research on the topic, it seems that this kind of codes is not very often employed in practice.

The starting point is a projective (nonsingular) algebraic curve X over $\mathbb{F}_q$, distinct points $P_1, \ldots, P_n \in X$ and a divisor D "disjoint" from these points (i.e., $n_{P_j} = 0$). We also impose $2g - 2 < \deg D < n$. The associated linear code is

$$V = \big\{ (f(P_1), f(P_2), \ldots, f(P_n)) \,:\, f \in \mathcal{L}(D) \big\} \subset \mathbb{F}_q^n.$$

Codes of this kind are called *AG codes* where AG stands for algebraic geometry.

Lemma 2.3. *The previous code is an* $[n, \deg D + 1 - g, d]$ *code with* $d \geq n - \deg(D)$.

Proof. Corollary 2.2 shows that the assumption $\deg D > 2g-2$ implies $\dim V = \deg D + 1 - g$. By the definition of d, there exists $f \in \mathcal{L}(D)$ such that $f(P_j) \neq 0$ for exactly d values of j that renaming the points we can assume to be $j = 1, \ldots, d$. Then $f \in \mathcal{L}\big(D - \sum_{j=d+1}^n P_j\big)$ and $\deg \operatorname{div}(f) = 0$ implies $\deg\big(D - \sum_{j=d+1}^n P_j\big) \geq 0$, hence $\deg D - (n-d) \geq 0$. □

Combining this result and the Singleton bound (1) with $k = \deg D + 1 - g$,

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

If we only think in terms of this inequality, we observe that the optimal choice is $g = 0$ when the algebraic geometric structure is trivial and the AG code coincide with other previously known codes. This is not the end of the story because because one may be interested in other things different from d, for instance the connection with the information rate given by k/n, that leads to interesting theoretical problems [6], but perhaps explain the persistence of the older methods in applications.

2.4 When the curve is not curved

Let us rephrase one of the aforementioned older methods with the language of AG codes. In its original formulation [3] the *Reed-Solomon codes* correspond to the linear subspace of $\mathbb{F}_q^n$

$$V = \big\{(P(x_1), P(x_2), \ldots, P(x_n)) \,:\, P \in \mathbb{F}_q[x],\ \deg P < k\big\}$$

where $x_1, \ldots, x_n \in \mathbb{F}_q$ are distinct. Two distinct polynomials cannot agree in more than $k-1$ points then the distance d is at least $n - (k-1)$ and recalling (1) it is deduced that Reed-Solomon codes are $[n, k, n-k+1]$ codes. They reach the highest capability of correcting errors for n and k given.

From the point of view of algebraic geometry, a polynomial of degree N is a rational function on the projective line having a pole of order N at infinity. Therefore, the Reed-Solomon codes corresponds to choosing as X the genus 0 curve $\mathbb{P}^1$, the points $P_j = (x_j : 1)$ and the divisor $D = (k-1)P_\infty$ with $P_\infty = (1 : 0)$ the point at infinity.

We consider a message as $\vec{m} = (m_0, \ldots, m_{k-1}) \in \mathbb{F}_q^k$, a sequence of k blocks in $\mathbb{F}_q$. A natural representation as a polynomial is

$$\vec{m} \in \mathbb{F}_q^k \longmapsto P_{\vec{m}} = \sum_{j=0}^{k-1} m_j x^j$$

It is easy to see that the block matrix

$$(-BA^{-1}|I) \qquad \text{with} \quad A = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k & x_k^2 & \ldots & x_k^{k-1} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & x_{k+1} & x_{k+1}^2 & \ldots & x_{k+1}^{k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{k-1} \end{pmatrix}$$

is a check matrix for the code, in other words, $\vec{v}$ is of the form $(P(x_1), P(x_2), \ldots, P(x_n))^t$ with $\deg P < k$ if and only if $(-BA^{-1}|I)\vec{v} = \vec{0}$.

There is a variant of the Reed-Solomon codes that makes them closer to the so-called *BCH codes* [7]. For α a fixed generator of $\mathbb{F}_q^*$, consider the map

$$P \in \mathbb{F}_q[x] \longmapsto QP \in \mathbb{F}_q[x] \quad \text{with} \quad Q = \prod_{j=1}^{n-k} (x - \alpha^j)$$

applies a polynomial of degree less than k into polynomials of degree less than n divisible by Q. The resulting polynomial can be identified with its n coefficients and hence considered as an element of $\mathbb{F}_q^n$. This map form $\{P \in\in \mathbb{F}_q[x] \,:\, \deg P < k\}$ to $\mathbb{F}_q^n$ is one to one, then the parameters are as in the original formulation of Reed-Solomon codes and we have an $[n, k, n-k+1]$ code. Summing up, each message $\vec{m} = (m_0, \ldots, m_{k-1}) \in \mathbb{F}_q^k$ is coded as $(c_0, \ldots c_n) \in \mathbb{F}_q^n$ following the procedure:

$$\vec{m} = (m_0, \ldots, c_{k-1}) \mapsto \sum_{j=0}^{k-1} m_j x^j \mapsto \sum_{j=0}^{k-1} m_j x^j \prod_{\ell=1}^{n-k} (x - \alpha^\ell) = \sum_{j=0}^{n-1} c_j x^j \mapsto \vec{c} = (c_0, \ldots, c_{n-1}).$$

The image of $\mathbb{F}_q^k$ constitutes the linear space V giving the code.

The advantage of this approach is that there are efficient algorithms for checking and correcting errors. Note that we can test $\vec{c} \in V$ verifying if $\sum c_j x^j$ vanishes at $x = \alpha^\ell$, $0 < \ell \leq n-k$. With respect to error correction, we only examine here for illustration the case of one or two errors.

Imagine that $\vec{c}$ is affected by a single error in the m-th coordinate, namely, we have received $\vec{r}$ that coincides with a valid code $\vec{c}$ except for $r_m = c_m + e_m$. The polynomial $R = \sum_{j=0}^{n-1} r_j x^j$ associated to $\vec{r}$ satisfies

$$\frac{R(\alpha^2)}{R(\alpha)} = \frac{e_m \alpha^{2m}}{e_m \alpha^m} = \alpha^m$$

and m is uniquely determined (by the discrete logarithm) if $n < q$ because α is a generator (and sometimes the condition $n < q$ is not necessary). Once we know that the error is in the m-th coordinate we can correct it using $e_m = \alpha^{-m} R(\alpha)$. Note that for these calculations we need $n - k \geq 2$ to assure $R(\alpha^2) = e_m \alpha^{2m}$ in concordance with (1) because the safe correction of one error requires $d \geq 3$.

If $n - k \geq 4$ then $d \geq 5$ and we can correct two errors. If they affect to the coordinates ℓ and m then $R(\alpha^j) = e_\ell \alpha^{jl} + e_m \alpha^{jm}$ and a calculation shows that

$$\begin{pmatrix} R(\alpha) & R(\alpha^2) \\ R(\alpha^2) & R(\alpha^3) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = - \begin{pmatrix} R(\alpha^3) \\ R(\alpha^4) \end{pmatrix} \quad \text{is solved by} \quad \begin{cases} x_1 = \alpha^\ell \alpha^m, \\ x_2 = -(\alpha^\ell + \alpha^m). \end{cases}$$

Once we have obtained x_1 and x_2 solving the linear system, α^ℓ and α^m are the roots of $X^2 + x_2 X + x_1 = 0$. They determine ℓ and m, with them e_ℓ and e_m can be eliminated from $R(\alpha) = e_\ell \alpha^l + e_m \alpha^m$ and $R(\alpha^2) = e_\ell \alpha^{2l} + e_m \alpha^{3m}$.

In general, for a higher number N of errors, a linear system allows to determine the *symmetric polynomials* [5] in N variables evaluated at α^ℓ with ℓ running over the location of the errors. Solving a polynomial equation of degree N the values of α^ℓ are obtained and the conditions imposed by the values of $R(\alpha^j)$ allow to get the magnitude e_ℓ of the errors.

Exercises

Exercise 1. For a linear $[n, k, d]$ code consider the linear map $f : V \longrightarrow \mathbb{F}_q^n$ assigning to $(v_1, v_2, \ldots, v_n)$ the vector $(0, \ldots, 0, v_d, v_{d+1}, \ldots, v_n)$. Show that $\dim V = \dim \mathrm{Im}(f)$ and deduce from it the Singleton bound (1).

EXERCISE 2. Prove that, with the above notation, $(-BA^{-1}|I)$ is actually a check matrix for the original formulation of the Reed-Solomon codes. Explain also why A^{-1} makes sense.

EXERCISE 3. Write a check matrix for the Reed-Solomon code $V = \{(P(0), P(1), P(2)) : \deg P < 2\} \subset \mathbb{F}_7^3$ and decide if $(2, 5, 1)$ and $(1, 6, 5)$ are valid codes.

EXERCISE 4. We encode $(m_0, m_1, m_2, m_3) \in \mathbb{F}_5^4$ as the vector in $\mathbb{F}_5^6$ determined by the coefficients of $\sum_{j=0}^{3} m_j x^j (x-2)(x-4) \in \mathbb{F}_5[x]$ in increasing degree order. If we receive $(3, 3, 3, 4, 4, 1)$, show that there is at least one error and correct it assuming that there is only one.

EXERCISE 5. Write a sagemath code automatizing and generalizing the process of the previous problem when $n < q$. For fixed q prime, n and α, a first function of the program must have $(m_0, \ldots, m_{n-3}) \in \mathbb{F}_q^{n-2}$ as input and offer as output a list given by the coefficients of the polynomial $\sum_{j=0}^{n-3} m_j x^j (x-\alpha)(x-\alpha^2)$ with a random error in one of them. A second function must take this output and correct the error to recover the original message.

References

[1] E. Girondo and G. González-Diez. *Introduction to compact Riemann surfaces and dessins d'enfants*, volume 79 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2012.

[2] R. Hill. *A first course in coding theory*. Oxford Applied Mathematics and Computing Science Series. The Clarendon Press, Oxford University Press, New York, 1986.

[3] I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *J. Soc. Indust. Appl. Math.*, 8:300–304, 1960.

[4] I. R. Shafarevich. *Basic algebraic geometry. 1*. Springer, Heidelberg, third edition, 2013. Varieties in projective space.

[5] I. Stewart. *Galois Theory*. Chapman & Hall/CRC Mathematics. Chapman & Hall/CRC, Boca Raton, FL, third edition, 2004.

[6] M. A. Tsfasman and S. G. Vlăduţ. *Algebraic-geometric codes*, volume 58 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1991. Translated from the Russian by the authors.

[7] S. A. Vanstone and P. C. van Oorschot. *An introduction to error correcting codes with applications*, volume 71 of *Kluwer Int. Ser. Eng. Comput. Sci.* Bosten, MA etc.: Kluwer Academic Publishers, 1989.