ERROR-CORRECTING CODES AND FINITE FIELDS

HAORU LIU

Abstract. We investigate the properties of modern error-correcting codes from an algebraic perspective. First, using techniques of linear algebra over finite fields, we develop the basic concepts of linear codes such as minimum distance, dimension, and error-correcting capabilities. We then use the structure of polynomial rings to define an example of cyclic codes, the Reed-Solomon code, and derive some of its properties. Finally, we introduce algebraic function fields and reinterpret Reed-Solomon codes from that perspective, then introduce the BCH code from the perspective of both cyclic codes and function fields.

Contents

1. Introduction - Linear codes

1.1. **Introduction and motivation.** The motivation for algebraic codes is the correction of errors in electronic communication that may arise due to imperfections in the physical medium of transmission. Consider the scenario in which Alice sends the eight bits 01000001 (representing the character 'A' in the ASCII character scheme) as a part of a message to Bob via a noisy transmission line. Due to random physical fluctuations, the fifth bit is flipped and Bob instead receives the string 01001001 (representing the character 'I'). If this is the extent of the communication between Alice and Bob, then there is a high likelyhood that Bob will not receive what Alice sent, as a single error changes the meaning of the whole communication. Algebraic codes seek to remedy this problem by encoding redundant data into the transmission, providing a means of ensuring correctness.

Date: September 29, 2012.

The simplest kind of redundant data is repetition. To implement this scheme, Alice and Bob agree beforehand that each 16 bit string sent over the transmission line will be the concatenation of two copies of the 8 bit message that Alice wishes to send. In the example above, Alice would transmit the string 0100000101000001. Then, if the fifth bit is flipped during transmission again, Bob will receive the string 0100100101000001. Now, since Bob knows that the first 8 bits and the second 8 bits are supposed to be identical, he is able to *detect* the error in transmission and thus request that the message be retransmitted. In fact, Bob will be able to detect any single-bit error in the transmission.

Due to the simplicity of this agreement for redundancy, multiple bit errors may slip by its notice. For example, if simultaneous errors occurred at the 5th and 13th bits, then Bob will receive the string 0100100101001001. Now, he is unable to tell if Alice transmitted the character 'I' or if she transmitted an 'A' after two well-placed errors occurred in transmission. In addition, this agreement is only able to detect errors – if Bob detected an error but wanted to know what Alice really wanted to say, he would have to request the transmission of an additional 16 bits from her, greatly increasing the time required especially given the latency of real-world transmission lines. An agreement between Alice and Bob with which Bob could not only detect but also *correct* errors in transmission would be helpful when retransmission is slow or impractical. To do this, we could repeat the pattern 3 times. In this case, Bob can now correct a 1-bit error by taking the two copies which agree, and he can detect a 2-bit error. Additional correction or detection capability can be added by simply increasing the number of repeats. However, this is a massively inefficiently use of transmission resources, and there are ways to build more efficient agreements.

In order to do so, we turn to algebraic structures. The discrete nature of digital communication naturally lends itself to manipulation via the theory of finite fields, the simplest example being the correspondence of 1s and 0s in binary to the elements of $\mathbb{F}_2$. Further, we can view binary strings of length n as elements of the vector space $\mathbb{F}_2^n$, so the characters discussed above would have been considered elements of $\mathbb{F}_2^8$. Given these connections to algebra, we can then devise methods of redundancy using the language of linear algebra. Note that while we began our discussion with binary data, the methods we shall use are equally applicable over all finite fields, so the rest of the paper will consider $\mathbb{F}_q$ instead of $\mathbb{F}_2$.

1.2. **Linear codes.**

Definition 1.1. A linear code C over the vector space $\mathbb{F}_q^n$ is the image of an injective linear map G from $\mathbb{F}_q^k$ to $\mathbb{F}_q^n$. The map G is called the generator map. We call n the block length (or simply length) of C and k the dimension of C.

Notation 1.2. Since we will be considering codes as finite-dimensional vector spaces over finite fields, all the sets we consider here are finite. Let $|C|$ denote the cardinality of C.

Since we will work exclusively with the standard basis on the various vector spaces over $\mathbb{F}_q$, we may write the generator map G as a matrix and refer to it as the *generator matrix*.

In the example given in subsection 1.1, the agreement between Alice and Bob is a linear code over $\mathbb{F}_2^{16}$ with dimension 8. Its generator matrix is the top-to-bottom concatenation of two 8×8 identity matrices.

The generator map lets us encode strings x in $\mathbb{F}_q^k$ simply by applying G to x and decode error-free codes in C by applying the inverse of G. However, in order for codes to be useful in practice, we need a way to decode error-containing strings in $\mathbb{F}_q^n \backslash C$ and a way to evaluate their error-correcting capabilities.

Definition 1.3. The *Hamming distance* d between x and y in $\mathbb{F}_q^n$ is defined as the number of coordinates that differ between x and y. The *Hamming weight* w of an element $x \in C$ is defined as $w(x) = d(x, 0)$, or the number of nonzero coordinates in x

The Hamming distance is a metric on $\mathbb{F}_q^n$, so we may apply the usual terminology of metric spaces to it. It also allows us to define a property of linear codes that is closely associated to their error-correcting capabilities.

Definition 1.4. The *minimum distance* of a linear code C is defined to be $\min\{d(x, y) : x, y \in C\}$. Using this definition, we write that C is an $[n, k, d]$ code, where n and k are as in definition 1.1 and d is the minimum distance of C.

Proposition 1.5. *The minimum distance of C is the same as $min\{w(x) : x \neq 0, x \in C\}$.*

Proof. Suppose the minimum Hamming weight of elements in C is d, and let $x \in C$ such that $w(x) = d$. Then, the minimum distance of C is at least d, since $d(x, 0) = d$. Assume that there exist $a, b \in C$ such that $d(a, b) < d$. Then, we have $w(a - b) = d(a, b) < w(x)$, a contradiction. □

The above proposition gives a useful way to prove bounds on the minimum distance of a linear code.

We may decode an element x in $\mathbb{F}_q^n$ by writing $x = c + e$, where $c \in C$ and e has minimum weight among all possible choices of c. We call e the *error vector* of x, and the error-correcting capabilities of a code C are measured by the weight of the heaviest error vector that the code is capable of correcting and/or identifying. An element $x \in \mathbb{F}_q^n \backslash C$ has a correctible error if there is a unique element c with $d(x, c)$ minimal, and x has a detectable error if it is possible to determine that an error occurred in transmission.

Next, we state how the minimum distance of C relates to the error-correcting ability of C.

Proposition 1.6. *A linear code of minimum distance d can detect all errors of weight less than $d - 1$ and correct all errors of weight t, where $d \geq 2t + 1$.*

Proof. Let x be the transmitted string in C, and let x' be the received string in $\mathbb{F}_q^n$.

Detection: Suppose that an error e of weight $s \leq d - 1$ occurred during transmission. Then we have $x' = x + e$, or $d(x, x') = s < d$. Thus, x' cannot be in C, since that would contradict the minimality of d.

Correction: Suppose that an error e of weight t occurred during transmission, with $2t + 1 \leq d$. Suppose then that there is another element $y \in C$ with $d(x', y) \leq d(x', x) = t$. We then have $d(x, y) \leq d(x, x') + d(x', y) \leq 2t < d$, again contradicting the minimality of d. □

It is important to note that while an error of weight less than $d-1$ is guaranteed to be detectable, any error which results in a received string not in C is also detectable. For example, the example given in section 1.1 has a minimum distance of 2 (given

by the string 1000000010000000), and it can correct all errors of weight 1 but it fails on only certain errors of weight 2. We shall see an example of a code with correctional capabilities when we disucss Hamming codes.

The quality of a code is determined by its three parameters, n, k, and d. Ideally, we want to have a high value of d in order to improve error-correction capabilities, while keeping the value of n small in order to reduce the amount of data that needs to be transmitted. We can prove some results on the possible values of n, k, and d.

Lemma 1.7. *Let $C \subset \mathbb{F}_q^n$, and let d be the minimum distance of C. Assume that for all $x \in \mathbb{F}_q^n \backslash C$, there exists $c \in C$ such that $d(x, c) < d$. Then, we have $b \cdot |C| \geq q^n$, where b is the cardinality of the closed ball of radius $d-1$ about any point of $x \in \mathbb{F}_q^n$.*

Proof. Suppose that $b \cdot |C| < q^n$. Then, the total number of points within distance $d-1$ of a point in C is less than $|\mathbb{F}_q^n|$, so there exists some element $x \in \mathbb{F}_q^n$ such that $d(x, c) \geq d$ for all $c \in C$. The negation of this statement is our desired result. □

Proposition 1.8. *With d, n fixed and $b = |B_{d-1}(x)|$ as above, an $[n, k, d]$ code exists when $b < q^{n-k+1}$.*

Proof. We induct on k. For $k = 1$, the map $x \mapsto (x, x, \ldots, x, 0, 0, \ldots, 0)$ from $\mathbb{F}_q$ to $\mathbb{F}_q^n$ defines a code with minimum distance d, where d is the number of nonzero components in each element of the image.
Suppose now that an $[n, k-1, d]$ code C exists. Since $|C| = q^{k-1}$, we have $b \cdot |C| < q^{n-k+1} \cdot q^{k-1} = q^n$. By the above lemma, there exists some $x \notin C$ such that $d(x, c) \geq d$ for all $c \in C$. Consider the code $D = \text{span}\{C, z\}$. Suppose that there were an element l of weight $w < d$ in D. Then, we have $l = c + nz$, with $c \in C$ and $n \in \mathbb{F}_q$. Multiplying by $-n^{-1}$, we have $l \cdot -n^{-1} = c' - z$, where $c' \in C$. This means that $d(c', z) = w(l \cdot -n^{-1}) = w(l) < d$, a contradiction. Thus, the minimum weight of an element of D is the same as the minimum weight of an element of C, so they have the same minimum distance d. □

Proposition 1.9 (Singleton bound)**.** *For any linear $[n, k, d]$ code, $d \leq n - k + 1$.*

Proof. Consider the map $C \to \mathbb{F}_q^{n-d+1}$ defined by removing $d-1$ components of an element of C. This is an injective linear map, as every nonzero element has at least d nonzero components, ensuring that only 0 maps to 0. Thus, we have $n - d + 1 \geq k$, since C can be embedded in $\mathbb{F}_q^{n-d+1}$ as a subspace. Rearranging this expression give the desired bound. □

The Singleton bound represents an upper bound on how good codes can be, as it sets an upper bound for the error-correcting capabilities for a code of length n and dimension k. Codes for which $d = n - k + 1$ exist, and they are called maximum distance separated codes, or MDS codes. We will see examples of such codes later.

1.3. **Perfect codes and Hamming codes.** Now, we examine a class of linear codes which have the property that they achieve the best possible information density for their error-correcting capabilities. We saw before that for a $[n, k, d]$ code of minimum distance $d \geq 2t + 1$, the balls of radius t about each point in C are pairwise disjoint. If these balls also cover $\mathbb{F}_q^n$, then the code makes the most efficient use of the data contained in $\mathbb{F}_q^n$, as any increase in the dimension of C would render it incapable of correcting errors of length t. Such codes are deemed to be *perfect*.

For the rest of this paper, the inner product of a and b on $\mathbb{F}_q^n$ will be the one defined by $\sum_{i=1}^{n} a_i b_i$, where a_i and b_i are the i-th components of a and b, respectively.

Definition 1.10. Let the code $C \subset \mathbb{F}_q^n$ have parameters $[n, k, 2t+1]$. C is said to be *perfect* if there exists some $c \in C$ for every $x \in \mathbb{F}_q^n$ such that $d(x, c) \leq t$.

Definition 1.11. Fix a maximal set A of pairwise linearly independent vectors in $\mathbb{F}_q^m$, and let $n = |A|$. Let H be a $m \times n$ matrix with its columns consisting of all vectors in A. A *Hamming code* over $\mathbb{F}_q^n$ is defined as the set of vectors x which satisfy $Hx = 0$.

Lemma 1.12. *A maximal set of pairwise linearly independent vectors in $\mathbb{F}_q^m$ has size $(q^m - 1)/(q - 1)$.*

Proof. Let A be the set of equivalence classes of $\mathbb{F}_q^m \backslash \{0\}$ under scalar multiplication. Any maximal set B of linearly independent vectors is in bijection with A, since no two vectors in B can be in the same equivalence class, and a set which does not contain representatives from every class in A can be extended by appending a representative from the omitted class. Since each element has $q-1$ scalar multiples and there are $q^m - 1$ nonzero elements in $\mathbb{F}_q^n$, $|A| = (q^m - 1)/(q - 1)$. □

The above proposition is important, as it associates a Hamming code with each m independently of the choice of vectors.

Next, we determine the dimension and minimum distance of Hamming codes.

Proposition 1.13. *A Hamming code with its maximal pairwise linearly independent set originating from $\mathbb{F}_q^m$ has dimension $n - m$, where $n = (q^m - 1)/(q - 1)$.*

Proof. The value of n is given by 1.12. First, since the columns of H span $\mathbb{F}_q^m$, the rank of H is m, which implies that the m rows of H are linearly independent, Then, the Hamming code is simply the orthogonal complement of the space spanned by the rows of H, which has dimension $n - m$. □

Notation 1.14. A Hamming code of length n and dimension k is often denoted as $Ham(n, k)$.

Proposition 1.15. *The minimum distance of a Hamming code is* 3.

Proof. Suppose that less than 3 of the components of some nonzero vector x in the Hamming code are nonzero. If only one is nonzero, that implies that one of the columns of H is zero, which contradicts the definition of H. Likewise, if only two are nonzero, then a nontrivial linear combination of two columns of H is zero, again a contradiction. Suppose the minimum distance is greater than 3. Then, no nontrivial linear combination of three columns of H are zero, which contradicts the maximality of the set of columns of H. To see this, we add a column that is the sum of any two columns of H. The columns of H remain pairwise independent, as any linear combination of the new column with any other column is a linear combination of three columns of the original H. □

In terms of error-correcting capability, the Hamming code is relatively poor for large values of m, as it can only correct one-bit errors or detect two-bit errors regardless of code length. However, its attractiveness comes from the fact that it makes the most efficient possible use of the information in an element of $\mathbb{F}_q^n$ given the minimum distance 3.

Proposition 1.16. *Hamming codes are perfect codes.*

Proof. Fix an element $x \in \mathbb{F}_q^n$, and examine the value of Hx. If x is not in the Hamming code, then Hx is a nonzero linear combination of the columns of H. By the maximality of the set of columns of H, Hx forms a nontrivial linear combination with the ith column v_i of H evaluating to zero for some i. We then have $av_i + bHx = 0$, or $\frac{a}{b}v_i + Hx = 0$. Thus, if we add a/b to the ith index of x, we obtain a string in the Hamming code, so any element of $\mathbb{F}_q^n$ is at most distance 1 from an element of the Hamming code. □

Example 1.17. The Hamming code as originally discovered was the $Ham(7, 4)$ code. To construct it, first take a maximal set of pairwise linearly independent vectors in $\mathbb{F}_2^3$. As given by proposition 1.12, such a set contains 7 vectors. Since $\mathbb{F}_2^3$ only has 8 elements and one of them is zero, the only such set is the set of nonzero elements in $\mathbb{F}_2^3$. We then take these vectors and use them as the columns of a matrix H, whose rows span the orthogonal complement of $Ham(7, 4)$. We have

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}, \tag{1.18}$$

and we obtain $\{(1, 0, 0, 0, 1, 0, 1), (0, 1, 0, 0, 0, 1, 1), (0, 0, 1, 0, 1, 1, 1), (0, 0, 0, 1, 1, 1, 0)\}$ as a basis of $Ham(7, 4)$. We can then write the generator matrix of $Ham(7, 4)$ as

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}. \tag{1.19}$$

We now demonstrate the coding and decoding processes with correction using this code. Suppose Alice wishes to transmit the string 0100 to Bob. She first applies G to encode this string, resulting in the 7-bit string $x = 0100011$. Note that due to the structure of G, the encoded string contains the message verbatim in its first 4 bits. If Bob receives this string, he applies H to it and receives the zero vector. Thus, he decodes the message by applying the inverse of G. If Bob receives the message with a one-bit error (say $x' = 1100011$), he calculates $Hx' = (1, 0, 0)$. Then, since Bob knows the matrix that Alice used to encode her message, he can determine that 1100011 is distance 1 from 0100011, which is an element of $Ham(7, 4)$. Thus, Bob is able to correct the one bit error due to the minimum distance property of $Ham(7, 4)$.

2. Cyclic codes and Reed-Solomon codes

In this section, we describe a class of codes that have more structure than simply being a vector space over $\mathbb{F}_q$. This additional structure enables us to obtain a class of codes which have particularly strong error-correcting capabilities.

Definition 2.1. A *cyclic code* over $\mathbb{F}_q^n$ is a linear code with the additional condition that the code is closed under cyclic permutations of components.

The primary advantage of this condition is that it embues the code with additional algebraic structure. Consider the map $\phi : \mathbb{F}_q^n \to \mathbb{F}_q[x]$ that takes an element $(a_0, a_1, \ldots, a_{n-1})$ to $a_0 + a_1 x + \ldots + a_{n-1}x^{n-1}$. We can regard this as an injective map to $\mathbb{F}_q[x]/(x^n - 1)$. Then, we have that a cyclic shift in the coordinates of the codeword corresponds to multiplication by the equivalence class of x, easily verified by taking the remainder of $a_0 x + a_1 x^2 + \ldots + a_{n-1}x^n$ with respect to $x^n - 1$.

Let the ring R denote $\mathbb{F}_q[x]/(x^n - 1)$. The image of C under ϕ in R then has the additional property of being closed under multiplication by x and scalars, which in fact makes it into an ideal. Let $h(x) = b_0 + b_1 x + \ldots + b_{n-1}x^{n-1}$ be in R. Then, for any $c(x)$ in the ideal C of R, we have that $h(x) \cdot c(x) = b_0 \cdot c(x) + b_1 x \cdot c(x) + \cdots + b_{n-1}x^{n-1} \cdot c(x) \in C$. Using the Lattice isomorphism theorem, the ideals of R correspond to the ideals of $\mathbb{F}_q[x]$ which contain $(x^n - 1)$ through the projection map. Then, since the ring $\mathbb{F}_q[x]$ is a principal ideal domain, all the ideals of R are principally generated. Thus, the codewords in C are all multiples of the coset of some polynomial $g(x)$, called the *generator polynomial* for C, unique up to multiplication by some unit of R. Since the proper ideals of R are generated by the factors of $x^q - 1$, the units of R are the polynomials relatively prime to $x^n - 1$, which means that they are not contained in any proper ideal.

We summarize these results in the following proposition.

Proposition 2.2. *A cyclic code C is a principally generated ideal in the ring $R = \mathbb{F}_q[x]/(x^n - 1)$*

Since each element of R corresponds to a polynomial of degree less than n in $\mathbb{F}_q[x]$, we may perform operations in R and C as if they were performed on polynomials, as long as the degree does not exceed $n - 1$.

A particularly interesting example of cyclic codes is the class of codes known as the *Reed-Solomon codes*. They are defined as follows:

Definition 2.3. Fix some field $\mathbb{F}_q$, and choose some $k < q$. Let L_{k-1} be the subspace of $\mathbb{F}_q[x]$ with elements of degree less than k. The Reed-Solomon code $R(k, q)$ is defined as the subspace of $\mathbb{F}_q^{q-1}$ obtained by evaluating elements of L_{k-1} at each nonzero element of $\mathbb{F}_q$, or $\{(f(1), f(\alpha), f(\alpha^2), \cdots, f(\alpha^{q-2})) : \ f \in L_{k-1}\}$, where α is a primitive element of $\mathbb{F}_q^\times$.

Reed-Solomon codes have particularly nice values for their minimum distances. Since polynomials in L_{k-1} have at most $k-1$ zeros, $(f(1), f(\alpha), f(\alpha^2), \cdots, f(\alpha^{q-2}))$ has at most $k - 1$ zero components in it for any $f \in L_{k-1}$, which means that the minimum weight of any element of $R(k, q)$ is $q - k = (q - 1) - k + 1$. Since the dimension of $R(k, q)$ is $q - 1$, this makes it a MDS code.

It is not immediately obvious that $R(k, q)$ is a cyclic code. To show this, we use the following lemma characterizing cyclic codes.

Lemma 2.4. *A k-dimensional linear code in $\mathbb{F}_q^{q-1}$ viewed as a subring of $R = \mathbb{F}_q[x]/(x^{q-1} - 1)$ is cyclic if and only if the set of common roots in $\mathbb{F}_q^\times$ of the polynomials corresponding to the codewords has size $q - k - 1$.*

Proof. Every element in $\mathbb{F}_q^\times$ is a root of $x^{q-1} - 1$, since it is a group of order $q - 1$. Thus, we may write $x^{q-1} - 1$ as

$$\prod_{\theta \in \mathbb{F}_q^\times} (x - \theta). \tag{2.5}$$

($\Rightarrow$): Suppose a code $C \in R$ is cyclic. Then, it is an ideal in R, and its preimage under the projection map is an ideal I of $\mathbb{F}_q[x]$. I is generated principally by some g with minimum degree in I. This means that $\deg g < q-1$, since otherwise we would have that every element of I has degree greater than or equal to $q-1$, which is impossible unless C were empty. In addition, g divides $x^{q-1}-1$ by the Lattice isomorphism theorem, so all the roots of g are in $\mathbb{F}_q^\times$.

Now, note that the elements of I of degree less than $q-1$ is a vector space over $\mathbb{F}_q$. Call this space I_{q-1}. Each element of C may be identified with a unique element of I of degree less than $q-1$, and vice versa. This identification is a bijective linear map, so the dimension of I_{q-1} is k. Now divide every element of I_{q-1} by g. The result is a space of polynomials of degree less than or equal to $(q-2)-\deg g$. Also note that if we multiply any polynomial of degree less than or equal to $(q-2)-\deg g$ by g, we get an element of I_{q-1}. Then, the image under division by g has dimension $(q-2)-\deg g+1$. Since division by g is an injective linear map, we have the equivalence $(q-2)-\deg(g)+1=k$, or $\deg(g)=q-k-1$. The roots of g are then the $q-k-1$ common roots.

($\Leftarrow$): Suppose the elements of C have $q-k-1$ common zeros. Then, viewed as elements of $\mathbb{F}_q[x]$ as above, they all divide some g of degree $q-k-1$ and are thus in the ideal $(g(x)) \subset \mathbb{F}_q[x]$. Further, they are also in the space I_{q-1} defined by the elements of $(g(x))$ with degree less than $q-1$, which by the above argument has dimension k. Then, if we project down to R, all the elements we are dealing with have degree less than $q-1$, so projection is linear and injective. Thus, we have that C is a subspace of the projection of I_{q-1} with the same dimension, so they are equal. Now, we have that $\overline{I_{q-1}}$ is an ideal in R generated by $\bar{g}$, as any element of $(g(x)) \subset \mathbb{F}_q[x]$ is equivalent to an element of I_{q-1}. This shows that C is an ideal. □

Theorem 2.6. *The Reed-Solomon code $R(k,q)$ is cyclic with generator polynomial*

$$g(x) = \prod_{i=1}^{q-k-1} (x-\alpha^i), \tag{2.7}$$

where α is a primitive element of $\mathbb{F}_q^\times$.

Proof. Consider an element $f = a_0 + a_1 x + \cdots + a_{k-1}x^{k-1} \in L_{k-1}$. The codeword c corresponding to this polynomial is $(f(1), f(\alpha), \cdots, f(\alpha^{q-2}))$, but it may also be viewed as a polynomial in R in the manner of Lemma 2.4. If we evaluate this polynomial at some element α^r of $\mathbb{F}_q^\times$, we obtain

$$c(\alpha^r) = \sum_{i=1}^{q-2} \Big(\sum_{j=0}^{k-1} a_j \alpha^{ij} \Big) (\alpha^r)^i \tag{2.8}$$

$$= \sum_{j=0}^{k-1} a_j \sum_{i=1}^{q-2} \alpha^{i(j+r)} \tag{2.9}$$

Consider the case when $1 \le r \le q-k-1$. Then, $1 \le j+r \le q-2$, and $\alpha^{(j+r)}$ is a nonidentity element of $\mathbb{F}_q^\times$. Then, the inner sum in 2.9 is zero, as it is equal to $\frac{\alpha^{(j+r)(q-1)}-1}{\alpha^{(j+r)}-1} = \frac{1-1}{\alpha^{(j+r)-1}-1} = 0$. Thus, $\{\alpha, \alpha^2, \cdots, \alpha^{q-k-1}\}$ is a set of common roots for all elements c of $R(k,q)$, and the code is cyclic by Lemma 2.4. □

Example 2.10. As an example, we construct a Reed-Solomon code with block lenth 7 and show how it can be used in traditional binary communication. First, we let the dimension of the code be 3, so the minimum distance of this code is 5 due to the MDS property. Let α be a primitive element of $\mathbb{F}_8$ with minimal polynomial $x^3 + x^2 + 1$. Due to the large number of elements in a 3-dimensional vector space over this field, we will only construct a generator matrix for this code instead of listing all its elements.

From definition 2.3, we know that $R(3,7) = \{(f(1), f(\alpha), \ldots, f(\alpha^6)) : \ f \in L_2\}$. A basis for this space over $\mathbb{F}_8$ may be obtained by simply plugging in 1, x, and x^2 for f, as these form a basis for L_2. Thus, listing the elements of this basis as the columns of the generator matrix, we have that the generator for $R(3,7)$ is

$$G = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 \\ 1 & \alpha^2 & \alpha^4 \\ 1 & \alpha^3 & \alpha^6 \\ 1 & \alpha^4 & \alpha \\ 1 & \alpha^5 & \alpha^3 \\ 1 & \alpha^6 & \alpha^5 \end{pmatrix} \tag{2.11}$$

A difficulty with this code is that it is not amenable to transmission over digital lines, which are generally limited to transmitting binary data. In order to make it practical for use, we must encode the elements of $\mathbb{F}_8$ as binary strings. Fortunately, there is an easy way to do this – view the elements as octal digits, with $\alpha^i = (i+1)_8$, and write them in binary (for example, $1 = 001$ and $\alpha^3 = 100$). Representing the elements this way, the code then becomes a way to encode 9 bits into 21 bits.

As an example. we can view the bit-string 101110011 as the vector $(\alpha^4, \alpha^5, \alpha^2) \in \mathbb{F}_8^3$, which becomes the vector $(0, \alpha^6, 0, \alpha^4, \alpha^6, 1, 1) \in \mathbb{F}_8^7$ upon encoding, which translates to the bit-string 000111000101111001001 of length 21.

The minimum distance of a code guarantees correction of t arbitrarily placed errors anywhere within the codeword, but performance may be different for errors distributed in a particular fashion. Reed-Solomon codes have the advantage of high *burst-error* correction rates. Burst errors are occurrences where all errors are located in a contiguous region of the codeword. These errors are common in practice, as damage to a storage medium or interruptions to a transmission line are often localized in space and time, respectively. Thus, we would expect codes with higher burst-error correction capabilities to be more effective in these situations.

Consider a Reed-Solomon code over the field $\mathbb{F}_{2^r}$. These codes can be used to encode bit-streams by interpreting every block of r bits as an element of $\mathbb{F}_{2^r}$. Then, the resulting codeword has length $2^r - 1$ with elements in $\mathbb{F}_{2^r}$ and a length of $(2^r - 1)r$ when viewed as bits. This is similar to what we did in 2.10, where we took a codeword of length $2^3 - 1$ with elements in $\mathbb{F}_8$ and viewed it as a string of $(2^3 - 1) * 3$ bits. Suppose that a burst error of rt occurs somewhere within the codeword. Then, since a contiguous block of $t + 1$ components in $\mathbb{F}_{2^r}$ has length greater than rt bits, at most $t+1$ components of the codeword will be changed. By Proposition 1.6, a minimum distance of $2t + 3$ will be needed to correct the burst error of length rt, as opposed to the $2rt + 1$ required for rt random errors.

Example 2.12. To illustrate burst error, we go back to our example in 2.10, where we used the code $R(3,7)$. Suppose that when transmitting the 21-bit codeword, a

burst error of length 4 occurs. If we view the codeword as the concatenation of 7 segments each of 3 bits, then the burst is only capable of changing two segments. Since each of the segments corresponds to one element in the length 7 codeword over $\mathbb{F}_8$, the code is guarenteed to be able to correct a burst error of length 4. This is a significant improvement compared to the correctional potential for randomly distributed errors.

The above discussion also exposes a weakness of Reed-Solomon codes. If we wish to transmit coded information in $\mathbb{F}_2$, our only choice for a Reed-Solomon code has block length 1, which is fairly useless. We can get around this by viewing strings in $\mathbb{F}_2$ as strings in $\mathbb{F}_{2^r}$, but this may be disadvantageous in certain circumstances. There exist MDS codes with less constrained block length, which we shall examine from the perspective of algebraic function fields in the next section.

3. Rational Algebraic-Geometric codes

The rational algebraic-geometric codes, or rational AG codes, for short, are a class of linear codes that are defined in terms of the function field $\mathbb{F}_q(x)$. The "algebraic-geometric" part of the name originates from the relationship between function fields and algebraic curves over some field k, arising from the function field $k(x, y)$ with x and y satisfying the algebraic relation that defines the curve.

Throughout this section, we take $\alpha + P_i$ for $\alpha \in \mathcal{O}_i$ to be the projection of α under the quotient map (see section 4.1 in the appendix).

Definition 3.1. Let $P_1, \ldots, P_n$ be places of degree 1 in $\mathbb{F}_q(x)$, and let $D = P_1 + \cdots + P_n$. Let G be a divisor whose support is disjoint from that of D (see the definition of support from definition 4.15). Then, the rational AG code $C_{\mathcal{L}}(D, G)$ is the subspace of $\mathbb{F}_q^n$ defined by $\{(t + P_1, \ldots, t + P_n) : \ t \in \mathcal{L}(G)\}$.

This definition relies on the fact that the $t + P_i$ are in $\mathbb{F}_q$. Note that for all i, $v_{P_i}(t) \geq 0$, as $t \in \mathcal{L}(G)$ implies $v_{P_i}(t) \geq v_{P_i}(G) = 0$. Thus, the $t + P_i$ are well-defined, and they lie in the residue field of P_i which is isomorphic to $\mathbb{F}_q$, as $\deg P_i = 1$.

Using the machinery we developed in the appendix, we can quickly derive some of the parameters of rational AG codes in general.

Proposition 3.2. *The rational AG code $C_{\mathcal{L}}(D, G)$ is an $[n, k, d]$ code with $d \geq n - \deg G$ and $k = l(G) - l(G - D)$.*

Proof. First, we determine the dimension of this space over $\mathbb{F}_q$. Note that the code is defined as the image of a linear map from $\mathcal{L}(G)$, with dimension $l(G)$ (see definition 4.19). The kernel of this linear map consists of the elements $x \in \mathcal{L}(G)$ for which $v_{P_i}(x) \geq 1$ for all P_i. This is precisely the set $\mathcal{L}(G - D)$, so we have $k = l(G) - l(G - D)$.

Since the minimum distance is equivalent to the minimum number of nonzero components in any element of the code, we pick some $c \in C_{\mathcal{L}}(D, G)$ with exactly d of its components nonzero. Then, we have $n - d$ places $P_{i_1}, \ldots, P_{i_{n-d}}$ for which $v_{P_{i_j}}(x) \geq 1$. As above, this implies that any $x \in \mathcal{L}(G)$ which corresponds to c is also an element of $\mathcal{L}(G - (P_{i_1}, \ldots, P_{i_{n-d}}))$. This implies that $l(G - (P_{i_1}, \ldots, P_{i_{n-d}})) \geq 0$, or $\deg(G - (P_{i_1}, \ldots, P_{i_{n-d}})) \geq 0$ by proposition 4.21b. Since the degree is additive, we have $0 \leq \deg G - n + d$, or $d \geq n - \deg G$. □

Notation 3.3. The lower bound on the minimum distance is called the *designed distance*.

Note that the tradeoff between efficiency of the code and its error correcting capabilities still exists here. If we set $\deg G$ to be a low value, then $l(G)$ is correspondingly low, and $l(G-D)=0$ for $\deg G < n$. Thus, a high designed distance implies a low efficiency, as the block length is fixed at n.

Suppose now that $0 \leq \deg G \leq n-2$. We then have $k = l(G) = \deg G + 1 + l(W-G)$ for some canonical divisor W. However, since W has negative degree and G has positive degree, $l(W-G)=0$, and we have $k = 1+\deg G$. By the Singleton bound(1.9), we have $d \leq n-k+1 = n-1-\deg G+1 = n-\deg G$, but we also have $d \geq n-\deg G$ by the above proposition. Thus, $d = n-\deg G$, and $C_{\mathcal{L}}(D,G)$ is a MDS code for $0 \leq \deg G \leq n-2$.

One may ask what happens when $\deg G$ falls outside the range. If $\deg G < 0$, then $l(G)$ and $l(G-D)$ are both 0, so $k=0$, making the code empty. If $\deg G = n-1$, then $k=n$ by the argument in the paragraph above. Finally, if $\deg G \geq n$, then by the definition of genus, we have $0 = g \geq \deg(G-D) - l(G-D)+1$ or $l(G-D) \geq 1$. Since $k = l(G)-l(G-D)$ and $l(G) = 1+\deg G$, we have $k \geq \deg G$, or $k=n$ since $k \leq n$. We then see that rational AG codes are only interesting for $\deg G$ within the range $[0, n-2]$, as otherwise we end up with trivial codes.

Since rational AG codes are ultimately derived from elements of the field of rational functions over $\mathbb{F}_q$, there is a simpler representation for them in terms of polynomials in $\mathbb{F}_q[x]$, but the language of function fields is still useful in proving properties about them.

Theorem 3.4. *Let $C = C_{\mathcal{L}}(D,G)$ be a rational AG code over $\mathbb{F}_q$ with parameters $[n,k,d]$, and suppose $n \leq q$. Then, the code C can be described as*

$$\{v_1 f(a_1), \ldots, v_n f(a_n) : \ f \in \mathbb{F}_q[x], \deg f \leq k-1\}, \tag{3.5}$$

with the v_i being distinct nonzero elements of $\mathbb{F}_q$ and the a_i being distinct elements from $\mathbb{F}_q$, possibly zero.

Proof. Let $D = P_1 + \cdots + P_n$, with the P_i having degree 1. Since there are $q+1$ places of $\mathbb{F}_q(x)$ corresponding to the q irreducible polynomials of degree 1 and the place at infinity (see theorem 4.12), there is some place Q of degree 1 that is not equal to any of the P_i. Using the Riemann-Roch equation, we have that $l(P_1-Q) = \deg(P_1-Q)+1-g+l(W-Q+P_1)$. Since $\deg W = -2$ and $\deg(P_1-Q)=0$, we have $l(P_1-Q)=1$. By proposition 4.21c, P_1-Q is a principal divisor. Let the element of $\mathbb{F}_q(x)$ associated to P_1-Q be α. α is in every valuation ring except $\mathcal{O}_Q$, and it is only in the place P_1. If neither P_1 nor Q is the place at infinity, then we have that the denominator of α divides exactly one irreducible and the numerator divides exactly one irreducible. Thus, α is of the form $c(x-a)/(x-b)$, where $a,b,c \in \mathbb{F}_q$, so $\mathbb{F}_q(\alpha) = \mathbb{F}_q(x)$. If $Q = P_\infty$, then the numerator of α is $x-a$, and the denominator must be constant, so $\mathbb{F}_q(\alpha) = \mathbb{F}_q(x)$ again. Similarly, if $P_1 = P_\infty$, $\alpha = 1/(x-a)$. Thus, we conclude that $\mathbb{F}_q(\alpha) = \mathbb{F}_q(x)$ in all cases, so α is transcendental over $\mathbb{F}_q$.

If we assume that this code is non-trivial, then we let $k = \deg G + 1$ and examine the divisor $(k-1)Q - G$. This has degree 0, so $l((k-1)Q-G) = \deg((k-1)Q-G)+1-g+l(W-(k-1)Q-G) = 1$ by Riemann-Roch. Thus, by proposition 4.21c, this is another principal divisor. Let its element be u. We claim that $a_i = \alpha + P_i$ and $v_i = u + P_i$.

Consider the elements $\alpha^i u$ for $0 \leq i \leq k-1$. They are linearly independent over $\mathbb{F}_q$, since a linear dependence among them gives rise to a polynomial with α as a root, contradicting its transcendence. Look at the valuations of these elements. For Q, $v_Q(\alpha^i u) = iv_Q(\alpha) + v_Q(u) = -i + k - 1 - v_Q(G) \geq -v_Q(G)$. For P_1, $v_{P_1}(\alpha^i u) = iv_{P_1}(\alpha) + v_{P_1}(u) = i + k - 1 \geq 0 = -v_{P_1}(G)$. For any other place, we have $v_P(\alpha^i u) = v_P(G)$, so $\alpha^i u \in \mathcal{L}(G)$. Further, since the dimension of $\mathcal{L}(G)$ is k by Riemann-Roch, these elements form a basis of $\mathcal{L}(G)$, so any element of $\mathcal{L}(G)$ may be written as $uf(\alpha)$, where $f \in \mathbb{F}_q[\alpha]$ and $\deg f \leq k-1$. Reducing this modulo places P_i, we obtain elements of C as $((u + P_1)f(\alpha + P_1), \ldots, (u + P_n)f(\alpha + P_n))$, recovering the form in the theorem. □

Definition 3.6. The codes defined by the equation 3.5 are known as *generalized Reed-Solomon codes.*

Example 3.7. In order to illuminate the construction used in deriving the generalized Reed-Solomon form from the language of function fields, we present an example using the code $C_{\mathcal{L}}(P_1 + P_\gamma + P_{\gamma^2}, P_0)$ over the field $\mathbb{F}_4 = \{0, 1, \gamma, \gamma^2\}$ (we let P_t, with $t \in \mathbb{F}_4$, denote P_{x-t}).

Let $D = P_1 + P_\gamma + P_{\gamma^2}$, $G = P_0$. We begin by selecting a place not in the support of D, so let $Q = P_\infty$. We consider the principal divisor $P_1 - Q$. From inspection, this is the divisor of $x - 1 \in \mathbb{F}_4(x)$. We take $\alpha = x - 1$.

Next, we let $k = \deg G + 1 = 2$. Following the proof, we end up with the principal divisor $(2-1)P_\infty - G = P_\infty - P_0$. The element corresponding to this is $x^{-1} \in \mathbb{F}_4(x)$. We then have that the a_i are the projections of $x - 1$ with respect to P_1, P_γ, and P_{γ^2}, and the v_i are the projections of x^{-1} with respect to P_1, P_γ, and P_{γ^2}. Thus, we have $a_1 = 0, a_2 = \gamma^2, a_3 = \gamma$ and $v_1 = 1, v_2 = \gamma^2, v_3 = \gamma$, and we conclude that $C_{\mathcal{L}}(D, G) = \{(f(0), \gamma^2 f(\gamma^2), \gamma f(\gamma)) : f \in \mathbb{F}_4[x], \deg f \leq 1\}$.

We now examine a particular example of generalized Reed-Solomon codes. To do so, we first need a proposition about the orthogonal complements of rational AG codes. For the proof of the following result, see [1].

Proposition 3.8. *Let $\alpha_1, \ldots, \alpha_n$ be elements of $\mathbb{F}_q(x)$, and let $P_1, \ldots, P_n$ be the places corresponding to $(x - \alpha_1), \ldots, (x - \alpha_n)$. Let y be an element of $\mathbb{F}_q(x)$ such that $y + P_i = 1$ for all i, and define $h(x)$ to be the product of $(x - \alpha_i)$ over all i. Then, the orthogonal complement of a rational AG code, $C_{\mathcal{L}}(D, G)^\perp$, is equal to $C_{\mathcal{L}}(D, D - G + (y) + (h'(x)) - (h(x)) - 2P_\infty)$, where $h'(x)$ is the derivative of h, $D = P_1 + \cdots + P_n$, and P_∞ is as in theorem 4.12.*

Definition 3.9. Fix a field $\mathbb{F}_{q^m}$, and let $\beta \in \mathbb{F}_{q^m}^\times$ be an element of multiplicative order n. Let l be an integer and $\delta \geq 2$ also an integer. Then, a *BCH code* is defined as $BCH(\beta, l, \delta) = \{c \in \mathbb{F}_q^n : \ Hc = 0\}$, where

$$H = \begin{pmatrix} 1 & \beta^l & \beta^{2l} & \cdots & \beta^{(n-1)l} \\ 1 & \beta^{l+1} & \beta^{2(l+1)} & \cdots & \beta^{(n-1)(l+1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \beta^{l+\delta-2} & \beta^{2(l+\delta-2)} & \cdots & \beta^{(n-1)(l+\delta-2)} \end{pmatrix} \tag{3.10}$$

Note that the matrix in 3.10 defines a code C' with block length n on $\mathbb{F}_{q^m}^n$. The BCH code is in fact a subset of the orthogonal complement of C': the rows of H form a basis of C', and the inner product of each row with any element of the

BCH code is 0. Further, the BCH code is actually the space $C'^{\perp} \cap \mathbb{F}_q^n$, called the restriction of $C'^{\perp}$ to $\mathbb{F}_q^n$.

We will first relate BCH codes back to our discussion about cyclic codes before we examine them from the perspective of function fields

Proposition 3.11. *BCH codes $BCH(\beta, l, \delta)$ are cyclic codes with generator polynomial $g = lcm\ \{m_{\beta^l}, m_{\beta^{l+1}}, \ldots, m_{\beta^{l+\delta-2}}\}$, where m_{β^i} denotes the minimal polynomial of β^i over $\mathbb{F}_q$.*

Proof. First, note that if $\deg g$ exceeds n, we can safely rewrite g as a polynomial of degree less than n, as powers of β above n are all equal to some power of β below n because $\beta^n = 1$. Now, look at the elements of $C = BCH(\beta, l, \delta)$ as polynomials of degree less than n. Since the i-th row of H are the powers of β^{i+l-1} up to $n-1$, we have that β^j is a root of all $c \in C$ for $l \leq j \leq l + \delta - 2$. This means that g divides each $c \in C$, as g is the minimal (with respect to divisibility) polynomial with all the β^j as roots. This implies that $C \subset (g)$. To show the reverse inclusion, we note that every multiple fg of g has the β^j as roots, implying that $H(fg) = 0$. □

Example 3.12. The code constructed in example 2.10 is also a BCH code. To see this, let β be an element of $\mathbb{F}_8$ with minimal polynomial $x^3 + x^2 + 1$, like in example 2.10. This element has multiplicative order 7, as it is a primitive element, so set $n = 7$. Next, set $l = 1$ and $\delta = 5$. The matrix produced from the definition of a BCH code is then

$$H = \begin{pmatrix} 1 & \beta & \beta^2 & \beta^3 & \beta^4 & \beta^5 & \beta^6 \\ 1 & \beta^2 & \beta^4 & \beta^6 & \beta & \beta^3 & \beta^5 \\ 1 & \beta^3 & \beta^6 & \beta^2 & \beta^5 & \beta & \beta^4 \\ 1 & \beta^4 & \beta & \beta^5 & \beta^2 & \beta^6 & \beta^3 \end{pmatrix} \tag{3.13}$$

It is fairly easy to see that the code C from example 2.10 is a subset of $BCH(\beta, 1, 5)$ by applying H to the basis elements of C. Equality then follows from the equivalence of dimensions – the rows of H are linearly independent, so the kernel of H has dimension 3 over $\mathbb{F}_8$.

Now, we examine the BCH code from the perspective of function fields. To do so, we first consider its parent, the code generated by H in $\mathbb{F}_{q^m}^n$.

Proposition 3.14. *Consider the function field $\mathbb{F}_{q^m}(x)$, and fix some β and n as in definition 3.9. Let P_i for $1 \leq i \leq n$ be the place corresponding to $x - \beta^{i-1}$, and let $D_\beta = P_1 + \cdots + P_n$. Suppose further that a and b are integers with $0 \leq a + b \leq n - 2$. Then, we have that the code generated by H in 3.10 is equal to the rational AG code $C_{\mathcal{L}}(D_\beta, aP_0 + bP_\infty)$, where P_0 and P_∞ are as in theorem 4.12.*

Proof. We follow the proof of theorem 3.4 to write $C_{\mathcal{L}}(D_\beta, aP_0 + bP_\infty)$ in the form of 3.5. From the definition of D_β, we know that $P_\infty \notin \operatorname{supp} D_\beta$. Thus, we let (y) be the divisor equal to $P_1 - P_\infty$, so $y = x - 1$. Further, we have the $\deg G = \deg(aP_0 + bP_\infty) = a + b$, so set $(z) = (a+b)P_\infty - aP_0 - bP_\infty = aP_\infty - aP_0$, or $z = x^{-a}$.

We then have that $\mathcal{L}(aP_0 + bP_\infty) = \{zf(y) : \ f \in \mathbb{F}_q[\alpha], \deg f \leq a + b\}$. However, we can just as easily set $y = x$ to obtain another basis of this space, as adding a constant to the variable does not change the polynomial ring. Then, we have $y + P_i = \beta^{i-1}$ and $z + P_i = \beta^{-(i-1)a}$ (note that $x^{-a} - \beta^{-(i-1)a}$ factors as $(x - \beta^{i-1})(x^{(-a-1)} + x^a\beta^a + \cdots + \beta^{-ia})$). Then, by theorem 3.4, we have that

$\{(\beta^{-(1-1)a}\beta^{(1-1)j}, \beta^{-(2-1)a}\beta^{(2-1)j}, \ldots, \beta^{-(n-1)a}\beta^{(n-1)j})\}$ for $0 \leq j \leq \deg G = a+b$ is a basis for the code $C_{\mathcal{L}}(D_\beta, aP_0 + bP_\infty)$. Writing this basis into a generator matrix and substituting $l = -a$ and $\delta = a + b + 2$, we get the generator matrix in 3.10 back out. □

Next, we use proposition 3.8 to obtain an expression for the orthogonal complement of the above code in terms of rational AG codes.

Proposition 3.15. *The orthogonal complement of the code $C_{\mathcal{L}}(D_\beta, aP_0 + bP_\infty)$ with parameters as above is $C_{\mathcal{L}}(D_\beta, cP_0 + dP_\infty)$, where $c = -a-1$ and $d = n-b-1$.*

Proof. Set $y = x^{-n}$. This satisfies the condition in proposition 3.8, as $x^{-n} - 1$ has all powers of β as roots. We also note that multiplying up all the $x - \beta^{i-1}$ gives $h(x) = x^n - 1$ as β is a primitive n-th root of unity.

Now, use proposition 3.8. We have

$$\begin{aligned}
&D - (aP_0 + bP_\infty) + (t^n) + (h'(x)) - (h(x)) - 2P_\infty \\
&= D_\beta - (aP_0 + bP_\infty) + n(P_\infty - P_0) + (x^{n-1}) - (x^n - 1) - 2P_\infty \\
&= D_\beta - (aP_0 + bP_\infty) + (P_\infty - P_0) - (\prod_{i=1}^{n}(x - \beta^{i-1})) - 2P_\infty \\
&= D_\beta - (aP_0 + bP_\infty) + (P_\infty - P_0) - \sum_{i=1}^{n}(P_i - P_\infty) - 2P_\infty \\
&= D_\beta - (aP_0 + bP_\infty) + (P_\infty - P_0) - D_\beta + nP_\infty - 2P_\infty \\
&= -(a+1)P_0 + (n-b-1)P_\infty,
\end{aligned}$$

proving the proposition. □

From here, obtaining the BCH code is simple – restrict the code from the above proposition to $\mathbb{F}_q^n$. We can also quickly obtain a lower bound on its minimum distance – the minimum distance of the BCH code must be at least the minimum distance of the code $C_{\mathcal{L}}(D_\beta, cP_0 + dP_\infty)$, which is itself bounded below by $n - \deg(cP_0 + dP_\infty) = n + a + 1 - n + b + 1 = a + b + 2 = \delta$. Thus, δ is the designed distance of the BCH code.

The rational AG codes are also subject to a limitation on block length – the length of a rational AG code over a base field $\mathbb{F}_q$ is limited by the number of places of degree 1 in the function field. In the case of the BCH code we constructed above, we were able to circumvent that by taking an extension of the field we were interested in coding over, building a code over the extension, then restricting it to the field we were originally interested in. This lets us build a code whose length is constrained by the size of the field extension. However, this still has some undesirable results. The dimension of the restricted code is difficult to derive in general, and this has negative implications for the efficiency of the code. A better way to build a code with long block length would be to move to a different function field, one where there are more places of degree 1. For example, consider the function field $\mathbb{F}_2(x, y)$ where x and y satisfy $y^2 + y = x^3 + x$. This function field has 5 places of degree 1, an improvement over the 3 from the rational case (see chapter 6 of [1]).

4. Appendix: Function fields

Algebraic function fields are objects which will help us in defining and understanding generalizations of the Reed-Solomon codes.

4.1. **Places and valuations.**

Definition 4.1. A function field over a base field k is a finite extension of $k(x)$ for some x transcendental over k. The *rational function field* over k is the field $k(x)$.

Definition 4.2. A *valuation ring* $\mathcal{O}$ of a function field F/k is a proper subring of F containing k such that for all $y \in F$, $y \in \mathcal{O}$ or $y^{-1} \in \mathcal{O}$.

Proposition 4.3. *The set of non-units in a valuation ring $\mathcal{O}$, $P = \mathcal{O} \backslash \mathcal{O}^\times$, is an ideal of $\mathcal{O}$.*

Proof. Let $x \in P$, $y \in \mathcal{O}$. Since x is not a unit, xy also cannot be a unit, so $xy \in P$.

Let $x, y \in P$. One of x/y or y/x is in $\mathcal{O}$, so assume that $x/y \in \mathcal{O}$. Then, $y(1 + x/y) = y + x \in P$. □

$P \in \mathcal{O}$ is clearly a maximal ideal – any ideal strictly containing P must contain a unit and therefore be equal to $\mathcal{O}$. Also, any other proper ideal of $\mathcal{O}$ must be contained in P, since no proper ideal may contain a unit. Thus, we may use the following definition of a place of F/k.

Definition 4.4. A *place* of a function field F/k is a subring of F that is the unique maximal ideal of some valuation ring $\mathcal{O}$. There is then a unique place corresponding to each valuation ring, and vice versa. Denote the set of places of a function field F as $\mathbb{P}_F$.

We now introduce discrete valuations of F/k, which are closely related to places.

Definition 4.5. A *discrete valuation* of F/k is a function v from F to $\mathbb{Z} \cup \{\infty\}$ satisfying the following 5 properties, with $n + \infty = \infty + \infty = \infty$ and $\infty > n$ for all $n \in \mathbb{Z}$.

a: $v(xy) = v(x) + v(y)$ for all $x, y \in F$

b: $v(x + y) \geq \min(v(x), v(y))$ for all $x, y \in F$

c: $v(x) = \infty$ if and only if $x = 0$

d: $v(x) = 0$ for all $x \in k$

e: There exists some $z \in F$ for which $v(z) = 1$

We now illustrate the connection between places and discrete valuations. To do this, we make use of the following lemma from [4].

Lemma 4.6. *Places of F are principal ideals. If t is the generator of some place, then each nonzero element of F can be written as $t^n u$ for some unit u in the valuation ring $\mathcal{O}_P$ associated with P.*

With this lemma in mind, we can associate a discrete valuation to each place P. Let $v_P(x) = n$ for $x \in F^\times$, where n is the integer such that $x = t^n u$ with t a generator of P and u a unit of the valuation ring associated to P. Define $v_P(0) = \infty$ for all P.

Theorem 4.7. *The function v_P satisfies the properties of a discrete valuation, and we have*

$$\begin{aligned} P &= \{z \in F : v_P(z) > 0\} \\ \mathcal{O}_P^\times &= \{z \in F : v_P(z) = 0\} \\ \mathcal{O}_P &= \{z \in F : v_P(z) \geq 0\}. \end{aligned} \tag{4.8}$$

In addition, for any discrete valuation v, the sets given in 4.8 are valuation rings and places of F.

Proof. Properties a, d, and e of a discrete valuation rings are verified by simple manipulations of the expression $t^n u$, and property c results directly from the definition of $v_P(0)$. To prove the inequality in b, let $x = t^{n_1}u_1$, $y = t^{n_2}u_2$, and assume $n_1 \leq n_2$. Then $x + y = t^{n_1}(u_1 + t^{n_2-n_1}u_2)$. The valuation of $x + y$ is at least t^{n_1} by the proof of Theorem 4.6, as $u_1 + t^{n_2-n_1}u_2 \in \mathcal{O}_P$.

Any element of valuation greater than zero is of the form $t^n u$, with $t \in P$ and $u \in \mathcal{O}_P$, so the element is in P. If the valuation of an element is 0, then it is a unit of $\mathcal{O}_P$. Conversely, any element of $\mathcal{O}_P^\times$ has valuation zero, and any element of P is a multiple of its generator by some element of $\mathcal{O}_P$, establishing two of the above equivalences. We then see that $\mathcal{O}_P = \mathcal{O}_P^\times \cup P$.

For any discrete valuation v, the set $\mathcal{O} = \{z \in F : v(z) \geq 0\}$ is a valuation ring of F, as $v(z^{-1}) = -v(z)$, so one of them has to be in the set. The closure of multiplication and addition are given by properties a and b of valuations, as two elements of nonnegative valuation cannot multiply or add together to give a negative valuation. The set $\mathcal{O}^\times = \{z \in F : v(z) = 0\}$ is the group of units, as $v(z) = 0$ implies $v(z^{-1}) = 0$, and any element of positive valuation has an inverse with negative valuation, which is not in $\mathcal{O}$. Thus, $\{z \in F : v(z) \geq 0\} = \mathcal{O}\backslash\mathcal{O}^\times$ is a place of F. □

Due to this last theorem, we can now talk about discrete valuations, places, and valuation rings interchangably, though we will mostly be working with places and their associated valuations.

We can now define another property of a place P, its degree. Since P is a maximal ideal of the ring $\mathcal{O}_P$, the quotient ring $\mathcal{O}_P/P$ is a field. In addition, since $k \cap P$ contains only zero, every element of k lies in a different coset of P in $\mathcal{O}_P$. We may then regard k as a subfield of $\mathcal{O}_P/P$.

Definition 4.9. The *residue field* of a place P is the field $F_P = \mathcal{O}_P/P$, containing k as a subfield. The *degree* of P is the degree of the extension F_P/k.

Proposition 4.10. *For any $t \in P$, $\deg P \leq [F : k(t)] < \infty$.*

Proof. We wish to relate the extensions $F/k(t)$ and F_P/k as vector spaces. Consider a set of elements $a_1, \ldots, a_n$ of $\mathcal{O}_P \subset F$ whose images under the projection map $\mathcal{O}_P \to F_P$ are linearly independent. We wish to show that these elements themselves are linearly independent. Suppose that there is a nontrivial linear combination $\sum_i f_i(t)a_i = 0$. We can clear denominators and multiply both sides by a power of t so that the f_i are polynomials in t with at least one f_i having nonzero constant term. Apply the projection map to this linear combination. We then have that

$$0 + P = \sum_i (f_i(t) + P)(a_i + P) \tag{4.11}$$

Since $t \in P$, the nonconstant terms of $f_i(t)$ are sent to 0 under the projection. At least one of the f_i has a nonzero constant term, meaning that $f_i(t) + P \neq 0 + P$ for some i. However, this contradicts the linear independence of the $a_i + P$ in F_P. □

4.2. **The rational function field.** Now, we consider the case of the rational function field, from which we will build new algebraic codes. The places of the rational function field can be easily classified, which makes it easier to work with the rational function field to construct codes.

Theorem 4.12. *f and g be relatively prime polynomials over k. Then, the valuation rings of $k(x)$ are all described by one of the following two cases:*

$$\mathcal{O}_{p(x)} = \left\{ \frac{f(x)}{g(x)} : p(x) \nmid g(x) \right\}, \text{ where } p \text{ is an irreducible, or}$$
$$\mathcal{O}_{\infty} = \left\{ \frac{f(x)}{g(x)} : \deg f \leq \deg g \right\}.$$

The corresponding places are

$$P_{p(x)} = \left\{ \frac{f(x)}{g(x)} : p(x) \nmid g(x), p(x) | f(x) \right\} \text{ and}$$
$$P_{\infty} = \left\{ \frac{f(x)}{g(x)} : \deg f < \deg g \right\}$$

These valuation rings, excluding $\mathcal{O}_{\infty}$, are actually localizations of $k[x]$ at the maximal ideal generated by the corresponding irreducibles.

Proof. It is fairly easy to see that the rings $\mathcal{O}_{p(x)}$ and $\mathcal{O}_{\infty}$ are valuation rings – in the case of $\mathcal{O}_{p(x)}$, if $\alpha = f(x)/g(x) \in k(x)$ is not in $\mathcal{O}_{p(x)}$, then p divides g. If α^{-1} is also not in $\mathcal{O}_{p(x)}$, then p also divides f, which contradicts the assumption that f and g are relatively prime. For $\mathcal{O}_{\infty}$, we have either $\deg f \geq \deg g$ or $\deg g \geq \deg f$, so one of α or α^{-1} is in $\mathcal{O}_{\infty}$. The proofs of $P_{p(x)}$ and P_{∞} being the non-units of their respective valuation rings is similar.

We now introduce a lemma that will help us prove that these are the only places of $k(x)$.

Lemma 4.13. *Valuation rings are maximal proper subrings of $k(x)$.*

Proof. Let $\mathcal{O}$ be a valuation ring, and let $\alpha \in k(x) \backslash \mathcal{O}$. We will show that any ring which contains α and $\mathcal{O}$ must be $k(x)$. Choose some other $\beta \in k(x) \backslash \mathcal{O}$, and let $v(\beta) = -n$. Since $v(\alpha^{-1}) = -v(\alpha) \geq 1$, we have $v(\beta\alpha^{-n}) \geq 0$ implying $\beta\alpha^{-n} \in \mathcal{O}$. Thus, any element $\beta \in k(x) \backslash \mathcal{O}$ can be written as a power of α times an element of $\mathcal{O}$. We then have that any ring which strictly contains $\mathcal{O}$ must also contain $k(x)$. □

Fix a place P of $k(x)$, and let $\mathcal{O}_P$ be its associated valuation ring. Assume first that $x \in \mathcal{O}_P$, so then $k[x] \subset \mathcal{O}_P$. Consider $k[x] \cap P$ as an ideal of $k[x]$. This ideal cannot only contain 0, for if it were, then all nonzero elements of $k[x]$ would be invertible in $\mathcal{O}_P$, implying that $\mathcal{O}_P = k(x)$. We wish to show that $\mathcal{O}_{p(x)} \subset \mathcal{O}_P$, where $p(x)$ is the generator of the ideal $k[x] \cap P$, from which it follows that $\mathcal{O}_P = \mathcal{O}_{p(x)}$ by the maximality of $\mathcal{O}_{p(x)}$. Let $f(x)/g(x) \in \mathcal{O}_{p(x)}$. We have that $p(x) \nmid g(x)$, or $g(x) \notin k[x] \cap P$. Since $g(x) \in k[x]$, this means that $g(x) \notin P$, implying that $g(x)^{-1} \in \mathcal{O}_P$. Since $f(x) \in k[x] \subset \mathcal{O}_P$, we have $f(x)/g(x) \in \mathcal{O}_P$.

We now want to show that $\mathcal{O}_P = \mathcal{O}_\infty$ when $x \notin \mathcal{O}_P$. When this condition holds, $x^{-1} \in \mathcal{O}_P$, so we can repeat the above construction with x^{-1}. In this case, we know that the ideal $k[x^{-1}] \cap P$ is generated by x^{-1}, as x^{-1} is in both $k[x^{-1}]$ and P, so it is the generator by virtue of being an irreducible element in a principal ideal. Suppose $f(x)/g(x) \in \mathcal{O}_\infty$. Then, divide each term of f and g by $x^{\deg g}$ and denote the result as f' and g'. Then, $g' \in k[x^{-1}]$ has a constant term. The constant term precludes g' from being in $k[x^{-1}] \cap P$, so it is not in P because it is in $k[x^{-1}]$. Following the above argument, we have $g'(x)^{-1} \in P$ and $f' \in k[x^{-1}] \subset \mathcal{O}_P$, so $f(x)/g(x) = f'(x)/g'(x) \in \mathcal{O}_P$. This means that $\mathcal{O}_\infty \subset \mathcal{O}_P$, and $\mathcal{O}_\infty = \mathcal{O}_P$ follows from the maximality of $\mathcal{O}_\infty$. □

For the purposes of calculating parameters of the codes we will construct using these places, it is useful to know the degrees of these places.

Proposition 4.14. *In $k(x)$, the degree of the place $P_{p(x)}$ is $\deg p$, and the degree of P_∞ is 1.*

Proof. In the case of $P_{p(x)}$, we wish to show that the residue field $k(x)_P$ is isomorphic to the extension of k given by $k[x]/(p(x))$, as this extension has degree $\deg p$. Consider the map ϕ from $k[x]$ to $k(x)_P$ given by $f \mapsto f + P$. The kernel of this map is the ideal $(p(x))$, as $f \in P$ if and only if $p(x)|f(x)$. In addition, the map is surjective. Take some element $\alpha = g(x)/h(x) \in \mathcal{O}_{p(x)}$ with $p \nmid h$. Since $k[x]$ is an Euclidean domain, we can find a and b in $k[x]$ such that $ap + bh = 1$. Multiply both sides of this by α and we have that $\alpha = agp/h + bg$. Reducing both sides modulo P, we have $\alpha + P = bg + P$, as $agp/h \in P$. We then conclude ϕ is surjective, as $\alpha + P$ is equal to the image of the polynomial bg under ϕ Thus, the two fields are isomorphic by the First isomorphism theorem.

In the case of P_∞, consider some $\alpha = f(x)/g(x) \in \mathcal{O}_\infty$. Suppose that $\alpha \notin P_\infty$, so that $\deg f = \deg g = n$. We can write α as $a_n x^n/g(x) + f'(x)/g(x)$, where $f'(x)/g(x) \in P_\infty$. Thus, every element of $k(x)_{P_\infty}$ can be described by one element of k, namely the leading coefficient of f, so the two spaces are isomorphic. □

Note that the places of $k(x)$ of degree 1 correspond to the points on a line in the projective plane over k. This is not a coincidence – the correspondence holds between places of any function field F and the points on the projective curve which is defined by the algebraic relation between the generators of F. In fact, we can obtain places and valuation rings of a general function field by similarly localizing the coordinate ring of the curve corresponding to the function field at its maximal ideals. Since we will be primarily considering rational function fields, see [3] and section 1.3 of [4] for more details.

4.3. **Divisors and Riemann-Roch.** We now give some definitions of objects derived from places and valuation rings and state some of their properties.

Definition 4.15. A *divisor* D of a function field F/k is a formal sum of places $\sum_{P \in \mathbb{P}_F} n_P P$, with $n_P \in \mathbb{Z}$ and all but a finite number of the n_P zero. The divisors are thus elements of the free abelian group generated by $\mathbb{P}_F$. The zero divisor is the identity element of this group.

The *support* of a divisor D is the set of places P for which n_P is nonzero. This is a finite set denoted as supp D.

Define the valuation of a divisor D at a place P to be $v_P(D) = n_P$, the P-th component of D. We can then define the degree of a divisor to be $\sum_{P \in \text{supp } D} v_P(D) \cdot \deg P$.

Finally, we define a partial order on the set of divisors as follows: Let $D_1 \geq D_2$ if for all places P, $v_P(D_1) \geq v_P(D_2)$.

Example 4.16. Consider the function field $\mathbb{F}_5(x)$. From our discussion of places over rational function fields in the previous section, we have that the formal sum $P_x + P_{x+1} + P_{x+4} - P_\infty$ is a divisor. The degree of this divisor is $1+1+1-1=3$, again due to our discussion of rational function fields.

We now define divisors associated to elements of F.

Definition 4.17. Fit $t \in F$. Define the set of zeros of t to be the set of places P such that $v_P(t) > 0$, and the set of poles to be the set of places P such that $v_P(t) < 0$. Then, we define the following divisors:

$$\text{the zero divisor: } (t)_0 = \sum_{P \text{ a zero of t}} v_P(t)P$$
$$\text{the pole divisor: } (t)_\infty = \sum_{P \text{ a pole of t}} -v_P(t)P$$
$$\text{the principal divisor: } (t) = (t)_0 - (t)_\infty.$$

This definition only makes sense if the set of places for which $v_P(t) \neq 0$ is finite. The proof of this for general function fields is quite complex, but there is a simple explanation in the case of $k(x)$. If $v_P(t) \neq 0$, then either $t \in P$ or $t^{-1} \in P$. In $k(x)$, there can only be a finite number of places that x can be in, for the numerator of t has only a finite number of irreducible factors. Similarly, the denominator can also only have a finite number of irreducible factors.

The set of principal divisors forms a subgroup of the divisor group, since $v_P(x) + v_P(y) = v_P(xy)$ implies $(x)+(y) = (xy)$. We then can make the following definition.

Definition 4.18. Two divisors are said to be equivalent (i.e. $D \sim D'$) if they are equivalent modulo the subgroup of principal divisors.

Definition 4.19. The *Riemann-Roch* space associated to a divisor D is the space $\mathcal{L}(D) = \{x \in F : \ (x) + D \geq 0\}$.

This space is a vector space over k, as we have for each place P that $v_P(x+y) - v_P(D) \geq \min(v_P(x), v_P(y)) - v_P(D) \geq 0$ for $x, y \in \mathcal{L}(D)$ and $v_P(ax) - v_P(D) = v_P(x) - v_P(D) \geq 0$ for $x \in \mathcal{L}(D)$ and $a \in k$. We can then define the dimension of D, $l(D)$, to be the dimension of this space.

In the case of the rational function field, the Riemann-Roch space has the following interpretation. A divisor can be seen as a specification of how badly-behaved a rational function is allowed to be – if $v_P(D) = n \geq 0$, then any function in $\mathcal{L}(D)$ is only allowed to have at most n factors of the irreducible associated with P. If $P = P_\infty$, the P-th component of the divisor determines how fast the function is allowed to diverge at infinity. In addition, if P is associated with an irreducible of degree 1, $v_P(D)$ dictates how sharp of a singularity a function in $\mathcal{L}(D)$ is allowed to have at the point associated with P.

Example 4.20. As an example, we compute the Riemann-Roch space associated to the divisor $D = P_x + P_{x+1} + P_{x+4} - P_\infty$ from example 4.16.

We first rearrange the equation in definition 4.19 to obtain $(x) \geq -D$. That is, we want all elements t of $\mathbb{F}_5(x)$ such that $v_x(t) \geq -1$, $v_{x+1}(t) \geq -1$, $v_{x+4}(t) \geq -1$, and $v_\infty(t) \geq 1$. In addition, we require that $v_{p(x)}(t) \geq 0$ for all other irreducibles p. Over the rational function field, $v_{p(x)}(t)$ is the number of times that $p(x)$ appears in the numerator of t (with negative values if p appears only in the denominator), and $v_\infty(t)$ is the degree of the denominator minus the degree of the numerator.

Given these properties of valuations, we can now describe t. First, since $v_\infty(t) \geq 1$, the degree of the denominator must be at least 1 more than the degree of the numerator. Second, since $v_{p(x)}(t) \geq 0$ for all p besides x, $x+1$, and $x+4$, the degree of the denominator can be no larger than 3. Then, this implies that the degree of the numerator is no more than 2. In fact, we can write all t satisfying this condition in the form

$$\frac{f(x)}{x(x+1)(x+4)},$$

where $\deg f \leq 2$, since factors of x, $x+1$, or $x+4$ in f will result in a lower-degree denominator. Thus, we have that

$$\mathcal{L}(D) = \left\{ \frac{f(x)}{x(x+1)(x+4)} : f \in \mathbb{F}_5[x], \deg f \leq 2 \right\} \text{ and}$$
$$l(D) = 3.$$

The following proposition is a collection of facts about $\mathcal{L}(D)$ and $l(D)$. The proof of these facts are either trivial or technical and beyond the scope of this paper, and they may be found in section 1.4 of [1]

Proposition 4.21. *(a): Let $D \sim D'$. Then $l(D) = l(D')$ and $\deg D = \deg D'$*
(b): If $\deg D < 0$, then $l(D) = 0$
(c): The following 3 conditions are equivalent for a divisor of degree 0.

(1) D is a principal divisor
(2) $l(D) \geq 1$
(3) $l(D) = 1$

(d): For all divisors D over a fixed function field, $\deg D - l(D)$ is bounded above.

By part d of the previous proposition, we can define an important invariant of a function field.

Definition 4.22. The *genus* of a function field F is defined as the maximum of $degD - l(D) + 1$ taken over all divisors D of F.

Now, we state the Riemann-Roch theorem, which is a powerful result on the relation between the dimension of the Riemann-Roch space of some divisor and the degree of the divisor.

Theorem 4.23 (Riemann-Roch)**.** *There exists an equivalence class of divisors, called* canonical divisors*, for which the following equation holds for any canonical divisor W and any divisor D.*

$$l(D) = \deg D + 1 - g + l(W - D) \tag{4.24}$$

Corollary 4.25. *The degree of any canonical divisor W is $2g-2$, and the dimension of its Riemann-Roch space is g. Furthermore, for any divisor D, if $\deg D = 2g-2$ and $l(D) \geq g$, then D is canonical.*

Proof. Let $D = 0$. Then, we have $l(D) = 1$, $\deg D = 0$, and $l(W - D) = l(W)$. By 4.24, we have that $l(W) = g$
Let $D = W$. Then, we have $l(W) = \deg W + 1 - g + l(0)$, or $g = \deg W + 1 - g + 1$ implying that $\deg W = 2g - 2$.
Consider a divisor D with the properties given above. Choose any canonical divisor W, and substitute the values into the Riemann-Roch equation. We have $2g - 2 + 1 - g + l(W - D) \geq g$, or $l(W - D) \geq 1$. Since $\deg W = \deg D = 2g - 2$, we have $\deg(W - D) = 0$, so W and D are equivalent by 4.21c. Then, D is a canonical divisor, as it is in the same class as W. □

Since we will be frequently applying the Riemann-Roch theorem over the rational function field, let us derive its genus now.

Proposition 4.26. *The rational function field has genus* 0*.*

Proof. Consider the space $\mathcal{L}(nP_\infty)$, where n ranges over the natural numbers. We have for $0 \leq r \leq n$ that $(x^r) = rP_x - rP_\infty$, as the only irreducible that divides x^r is x, and the difference between the degree of the denominator and the degree of the numerator of x^r is $-r$. Then, we have $nP_\infty + rP_x - rP_\infty \geq 0$, so $x^r \in \mathcal{L}(nP_\infty)$. Since these elements are linearly independent over k, $l(nP_\infty) \geq n + 1$.

Let g be the genus of the rational function field, and take $n \geq 2g - 1$. By Proposition 4.21b and the fact we proved about the degree of a canonical divisor above, we have that $l(W - nP_\infty) = 0$. Thus, by Riemann-Roch, $\deg nP_\infty + 1 - g = l(nP_\infty) \geq n + 1$, or $-g \geq 0$. However, since $\deg 0 - l(0) + 1 = 0$, we have that $g \geq 0$, implying that $g = 0$. □

Example 4.27. Recall the divisor $D = P_x + P_{x+1} + P_{x+4} - P_\infty$ from example 4.20. In this example, we used an explicit calculation to calculate $l(D)$, but we can do this much quicker by using Riemann-Roch. Note first that over a rational function field like $\mathbb{F}_5(x)$, $\deg(W - D) \leq 0$ since $\deg D \geq 0$ and $\deg W = 0$. Thus, $l(W - D) = 0$. As we calculated in example 4.16, $\deg D = 2$. Thus, we have $l(D) = \deg D + 1 - g + l(W - D) = 2 + 1 - 0 + 0 = 3$, agreeing with our result from example 4.20.

Acknowledgements. I would like to thank my mentor, Daniel Le, for the substantial effort of editing this paper and explaining quite a bit of the geometric motivation behind function fields.

References

[1] Stichenoth, Henning. Algebraic function fields and codes. Berlin: Springer, 1993.
[2] Cox, David A., and John B. Little. Using algebraic geometry. 2. ed. New York, NY: Springer, 2005.
[3] Fulton, William. Algebraic curves, an introduction to algebraic geometry;. New York: Benjamin, 1969.
[4] Serre, Jean. Local fields. New York: Springer-Verlag, 1979.