Universidade Federal do Rio Grande do Sul

Instituto de Matemática

Programa de Pós-Graduação em Matemática

CÓDIGOS CORRETORES ALGÉBRICOS

por

Ricardo Queiroz de Araújo Fernandes

Porto Alegre, 11 de Dezembro de 2007

Dissertação submetida por Ricardo Queiroz de Araújo Fernandes como requisito parcial para a obtenção do grau de Mestre em Matemática pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática da Universidade Federal do Rio Grande do Sul.

Professor Orientador:

Dr. Luis Gustavo Doninelli Mendes

Banca Examinadora:

Dr. Ivan Edgardo Pan Perez

Dr. Luis Gustavo Doninelli Mendes

Dra. Luisa Rodriguez Doering

Dra. Miriam del Milagro Abdón (UFF-Niterói- RJ)

Data de Defesa: 11 de Dezembro de 2007.

dedico
à minha família e
aos meus amigos.

Agradecimentos

Agradeço a oportunidade de viver.

Agradeço ao meu orientador, que desde o primeiro encontro é muito mais meu amigo do que o doutor com pós-doutorado na França.

Agradeço à minha Mãe por ter se empenhado para que eu estudasse apesar das dificuldades.

Agradeço à minha esposa pela dedicação e suporte para que eu tenha tranqüilidade para estudar.

Agradeço aos meus amigos e colegas que tornam a caminhada mais agradável.

Agradeço o amor ao estudo.

Resumo

A Teoria de Códigos lida com um problema crítico do processo de comunicação: o controle eficiente do ruído em um canal de comunicação. Este trabalho tem por objetivo descrever algumas relações entre as necessidades da Teoria de Códigos e sua a modelagem matemática, utilizando-se, para isso, conceitos de álgebra e geometria sobre corpos finitos. Apresentaremos motivações para diferentes códigos, dando ênfase a aspectos geométrico-projetivos do código Hamming e conceitos geométrico-algébricos subjacentes aos de Reed-Solomon e Goppa.

Abstract

The Coding Theory deals with a critical communication problem: the eficient noise control. This work aims to describe some relationships between the needs of the Coding Theory and its Mathematical Models throught concepts of algebra and geometry over finite fields. We present some motivation for different codes, searching for projective geometric aspects of Hamming codes and algebraic geometric aspects of Reed-Solomon and Goppa codes.

Sumário

1 Modelo de Comunicação **1**
1.1 Codificação 2

2 Códigos Corretores **5**
2.1 Definição abrangente e seus algoritmos 7

3 Códigos Lineares **12**
3.1 Definições e algoritmos 13
3.2 Códigos MDS 16
3.2.1 Relações com a Geometria Projetiva 16
3.3 Código de Hamming 18
3.3.1 Relações com a Geometria Projetiva 20

4 Códigos Algébricos **23**
4.1 Divisores 26
4.1.1 Espaço vetorial de funções 27
4.2 Códigos de Reed-Solomon 29
4.3 Diferenciais sobre uma curva 31
4.3.1 Espaços vetoriais sobre diferenciais 34
4.4 Códigos de Goppa 35
4.5 Teorema de Riemann-Roch 36
4.5.1 Lacunas de Weierstrass 37
4.5.2 Parâmetros dos códigos de Reed-Solomon 38
4.5.3 Parâmetros dos códigos de Goppa 39
4.5.4 Dualidade entre Reed-Solomon e Goppa 40
4.6 Um código de Reed-Solomon em detalhe 42
4.7 Bons códigos 45

5 Apêndice: Geometria Projetiva **46**

1 Modelo de Comunicação

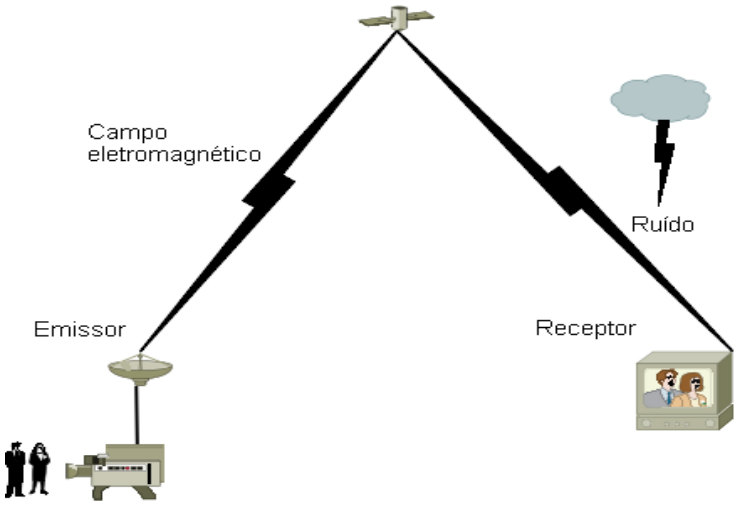

Figura 1: Exemplo de comunicação

Um *emissor* possui uma *informação* a ser transmitida a um *receptor* através de um *meio de comunicação*.

Faremos algumas hipóteses sobre este modelo.

O emissor transmite informações por um meio de comunicação que seja compatível, ou seja, o receptor é capaz de receber informações por este meio escolhido. Em outras palavras, não adianta enviar informações eletromagnéticas para um receptor humano.

O transporte da informação através do meio será chamado de *sinal*. Como o sinal depende do meio, faz sentido falarmos de sinal eletromagnético, sinal elétrico, sinal mecânico, dentre outros.

O meio produz *ruído* na transmissão de maneira independente do emissor e do receptor e, portanto, será considerado aleatório. Entenderemos o ruído como um sinal que o meio introduz ao sinal originado pelo emissor. Desta maneira, como o ruído é um sinal, ele terá amplitude mensurável. Chamaremos o efeito do ruído de *erro*.

Chamaremos de *canal* a alocação ou reserva de recursos do meio para a comunicação. Como exemplo podemos pensar na atribuição de um intervalo de freqüência para uma determinada estação de rádio.

É interessante observar que o ruído não altera a informação, mas sim a mensagem que a representa. Assim o ruído não ameaça a integridade da informação, mas compromete de maneira crítica a capacidade do receptor conhecer a informação. O ruído, então, pode impossibilitar toda a comunicação. Um bom exemplo é tentar conversar em um ambiente com música alta.

1.1 Codificação

Para que a informação seja transmitida, é necessário que ela seja codificada para utilizar este meio e passa a se chamar *mensagem*. Uma vez recebida a mensagem, deverá esta ser decodificada em uma informação compreensível. Assim fazemos nós ao expressarmos nossas palavras em fonemas audíveis.

Existem diversos meios para se estabelecer uma comunicação. É de se esperar, portanto, que existam diferentes formas de se codificar a informação para que possa ser transmitida. Mesmo um determinado meio possui diferentes codificações.

A classificação mais abrangente das codificações as separa em codificações analógicas e codificações digitais. Deixando de lado questões técnicas e econômicas, vejamos o que as diferencia.

A codificação analógica realiza transformações contínuas (em geral transformadas integrais) sobre parâmetros contínuos da informação. Assim, a informação é representada de forma contínua. Esta codificação oferece, então, possibilidade infinita de representações.

Para tornar mais claro, digamos que alguém deseja monitorar o nível da maré em alguma localidade. E, para isto, estabelece um marco e coloca um sensor que transmita a informação de forma analógica. Logo, cada ponto do intervalo que ele escolher será uma mensagem aceitável.

Concluímos assim que, neste modelo e nesta codificação, a interferência do ruído é irreversível, pois o efeito do ruído sempre produzirá outra mensagem aceitável. Não significa que a comunicação seja inviável, mas sim que as distorções nas mensagens não são passíveis de correção. A codificação analógica oferece representação infinita e precisão nula.

A codificação digital, por sua vez, realiza a amostragem (quantização) dos parâmetros da informação sobre um conjunto finito de possibilidades de representação. Como exemplo, o padrão de vídeo chamado RGB oferece 256 níveis de vermelho, verde e azul.

Isto pode parecer uma limitação inaceitável; mas, se levarmos em conta que a sensibilidade humana também possui limites, compreenderemos porque o CD substituiu o disco de vinil. O importante é que o número de níveis possíveis seja tão grande que nos pareça um intervalo contínuo ou que atenda a algum padrão de qualidade estipulado.

Chamaremos de *símbolos* os elementos da codificação que produzem todos os níveis de representação. No nosso último exemplo, os símbolos são todos os números de 0 a 255. Chamaremos de *alfabeto* o conjunto de todos os símbolos de uma determinada codificação.

Mas nossa atenção se volta para outra propriedade característica da codificação digital: a possibilidade de corrigir erros provocados pelo ruído. Um exemplo teórico pode ajudar.

Digamos que nossa codificação digital represente todas as informações em números inteiros não negativos múltiplos de 8 e menores que 1024 e os sinais transportam somente números inteiros. Se o receptor recebe um número que não é múltiplo de 8, ele poderá corrigir o número recebido para o múltiplo de 8 mais próximo do número recebido.

Concluímos que esta codificação oferece representação finita e precisão não nula. Nos interessa aumentar a precisão, controlando os níveis de qualidade, o que só é possível na codificação digital. Por esta razão tomaremos a codificação digital também como mais uma hipótese do modelo.

Sabendo que, em nosso modelo, a codificação é digital; podemos nos perguntar qual o possível efeito do ruído sobre a comunicação. Se levarmos em conta a experiência do cotidiano, em que falta de energia elétrica, curto-circuito e mau-contato são reais; é compreensível que o ruído possa alterar o fluxo das mensagens, suprimindo ou acrescentando símbolos.

Este problema é resolvido pelo *controle de fluxo*. Este mecanismo, que pode ser implementado em hardware ou software, evita alterações no tamanho das mensagens. Como desejamos estudar outro efeito do ruído, iremos supor que algum mecanismo de controle de fluxo é aplicado em nosso modelo, de tal maneira que teremos por hipótese que o ruído somente troca símbolos.

Colocaremos agora a síntese de nosso modelo de comunicação.

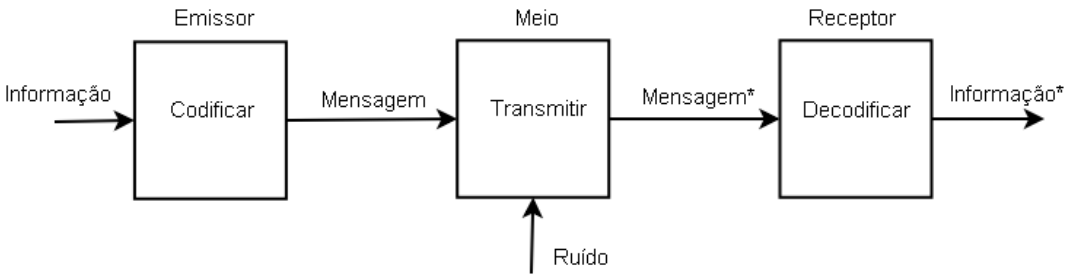

Figura 2: Modelo de comunicação

Hipóteses:

- O ruído é aleatório;
- A codificação é digital;
- O ruído somente troca símbolos.

2 Códigos Corretores

Tendo estabelecido nosso modelo de comunicação e suas hipóteses aplicáveis, desejamos agora construir códigos que nos permitam corrigir o efeito do ruído. Partiremos de uma situação simples como motivação. Apresentaremos depois uma definição ampla de o que é um código corretor.

Precisamos antes advertir que os códigos que iremos introduzir não correspondem à codificação do sinal abordada anteriormente. A codificação do sinal trata especificamente da transmissão do sinal. A codificação que iremos abordar usa a codificação digital para realizar o controle eficiente do ruído.

Nos cumpre também ressaltar que os termos e conceitos apresentados na parte inicial da seção, apesar de serem relevantes para a Teoria das Comunicações, não serão alvo do estudo deste trabalho, contribuindo somente para motivação.

Digamos que nosso alfabeto $\mathcal{A}$ seja o conjunto das letras. Se o receptor, após o ruído, recebe a letra L; como poderá ele saber se esta letra está correta ou trocada?

Para responder esta pergunta é necessário antes conhecer a *qualidade* do canal. Apesar de termos visto o ruído como um vilão incontrolável, podemos estudar o comportamento do canal na transmissão de nossos símbolos e atribuir a ele uma *taxa de erro esperada*.

Assim, se o receptor recebe por um canal com taxa de erro muito pequena, ele poderá aceitar a letra L como a mais provável, tolerando o risco de estar errado. O problema se torna mais interessante quando não dispomos de um canal com taxa de erro que atenda nosso nível de tolerância a falhas. O que fazer por exemplo quando a taxa de erros for alta, como acontece nas transmissões via satélite?

Um algoritmo bem simples consiste em obrigar o emissor a enviar redundâncias, por exemplo 5 cópias da mesma letra para cada letra a enviar. Assim, ao receber a seqüência $LLLLL$, o receptor ficará convencido pela maioria dos símbolos, pois sabe que a probabilidade do ruído produzir o mesmo efeito 5 vezes é ínfima. Se, no entanto, recebe $LPLLS$; saberá ainda que L é a letra mais provável.

Alguém poderia, então, pensar que bastaria aumentar indefinidamente o número de redundâncias. Isto, porém, resulta em um algoritmo que desperdiça recursos do canal, ou seja, ineficiente.

Por outro lado, se a taxa de erro for muito alta, pode acontecer de obtermos seqüências sem símbolo com maioria de representantes para uma porcentagem inaceitável de tentativas de comunicação. Em outras palavras, se a taxa de erro estiver próxima ou acima de 50%, nosso mecanismo será ineficaz.

Definição 2.1. *Chamaremos* $\mathcal{A}^n := \underbrace{\mathcal{A} \times \mathcal{A} \times ... \times \mathcal{A}}_{n}$ *de espaço de seqüências.*

É interessante observar, mesmo neste exemplo simples, que:

Lema 2.2. *A aplicação* $h : \mathcal{A}^n \times \mathcal{A}^n \to \mathbb{N}_0$*, definida por* $h(\alpha, \beta) :=$ *número de posições com símbolos diferentes entre* α *e* β*, é uma métrica em* $\mathcal{A}^n$*. Chamamos* h *de métrica de Hamming.*

Prova:

1. $h(\alpha, \beta) = 0 \iff \alpha = \beta$ por definição;
2. $h(\alpha, \beta) = h(\beta, \alpha)$ por definição;
3. $h(\alpha, \beta) \leq h(\alpha, \gamma) + h(\gamma, \beta)$:
   Suponhamos que $h(\alpha, \beta) > h(\alpha, \gamma) + h(\gamma, \beta)$. A partir de α, poderemos produzir γ, realizando alterações em $h(\alpha, \gamma)$ posições. A partir de γ poderemos produzir β, realizando alterações em $h(\gamma, \beta)$ posições. Considerando a composição destas alterações, podemos produzir β a partir de α realizando alterações em, no máximo, $h(\alpha, \gamma) + h(\gamma, \beta)$ posições. Contradição com a definição de $h(\alpha, \beta)$. $\diamond$

Pensando em eficiência, tentaremos fazer uma analogia com o processo lingüístico. Em nosso idioma, o alfabeto é conhecido. Logo todas as letras deste alfabeto são aceitas. No entanto, nem todas as seqüências são aceitas. Ao lermos, em português, a letra ***z***, deveremos aceitá-la. Ao lermos, porém, a sílaba ***bz***; saberemos que houve erro. O que isto sugere? Não somos capazes de corrigir símbolo por símbolo de um texto em nosso próprio alfabeto, mas sim, seqüências de símbolos.

Existirá, contudo, alguma vantagem em se corrigir seqüências de símbolos? A resposta quem dá é nosso cérebro; que, ao se deparar com um texto, corrige palavras e não letras. É de se esperar, portanto, que este processo seja mais eficiente que nosso algoritmo inicial.

Temos a imprsesão de ainda poedrmos entender um texto com pocuos erros de ortografia.

O que acontece, na verdade, é que o cérebro faz uma nova significação e assume como símbolos as palavras e não mais as letras isoladas, que ficam vazias de significado. As letras estão muito mais próximas do sinal de comunicação, pois representam fonemas, enquanto a informação está nas palavras.

Observamos, assim, que a seqüência $\boldsymbol{IMPRESSÃO} \in \mathcal{A}^8$ é uma seqüência aceitável. A seqüência $\boldsymbol{IMPRSESÃO} \in \mathcal{A}^8$ não é aceitável. E assim se constrói um dicionário, tomando-se a união de subconjuntos de $\mathcal{A}^1, \mathcal{A}^2, ..., \mathcal{A}^n$.

Contudo, a seqüência $\boldsymbol{IMPRSESÃO}$ está próxima o suficiente de $\boldsymbol{IMPRESSÃO}$ para que o cérebro corrija. Fica claro, neste contexto, que a distância entre seqüências nos oferece a oportunidade de recuperar a informação em um canal ruidoso.

Como, em nosso modelo de comunicação, o ruído somente troca símbolos, utilizaremos a métrica de Hamming, uma vez que esta identifica somente alterações de símbolos.

2.1 Definição abrangente e seus algoritmos

Introduziremos uma definição extremamente ampla dos objetos de nosso interesse: os códigos. Somente depois estabeleceremos estruturas adicionais, que permitirão uma definição mais restrita, útil e precisa de códigos.

Definição 2.3. *$\forall\ \mathcal{C} \subset \mathcal{A}^n$, $\mathcal{C}$ é um código corretor. $\forall\ c \in \mathcal{C}$, chamaremos c de palavra de código.*

Dada esta definição genérica, como utilizar algum código $\mathcal{C}$ para efetivamente corrigir erros?

Digamos que nosso interesse seja corrigir seqüências de tamanho $n \in \mathbb{N}$ em uma comunicação. Faremos então como o cérebro e escolheremos um conjunto $\mathcal{C} \subset \mathcal{A}^n$ tal que nossas informações serão representadas por elementos $c \in \mathcal{C}$.

- O algoritmo:

  1. O emissor envia $c \in \mathcal{C}$;
  2. A mensagem pode ser alterada pelo efeito do ruído;
  3. O receptor recebe uma mensagem $m \in \mathcal{A}^n$ possivelmente com erro;
  4. O receptor verifica se $m \in \mathcal{C}$;
  5. caso afirmativo: aceitar m;
  6. caso negativo: calcular $d_m := \min\{h(m,c)|c \in \mathcal{C}\}$;
  7. calcular $V_m := \{c \in \mathcal{C}|h(m,c) = d_m\}$;
  8. verificar se $\sharp V_m = 1$
  9. caso afirmativo: corrigir m para $c \in V_m$;
  10. caso negativo: rejeitar m.

O que este algoritmo faz é definir vizinhanças disjuntas em torno de cada elemento de $c \in \mathcal{C}$. Qualquer mensagem recebida em uma destas vizinhanças é corrigível, as demais não são corrigíveis. Fica, então, patente que a escolha dos elementos de $\mathcal{C}$ tem influência direta na construção destas vizinhanças e, conseqüentemente, na capacidade do código corrigir erros produzidos sobre cada palavra de código.

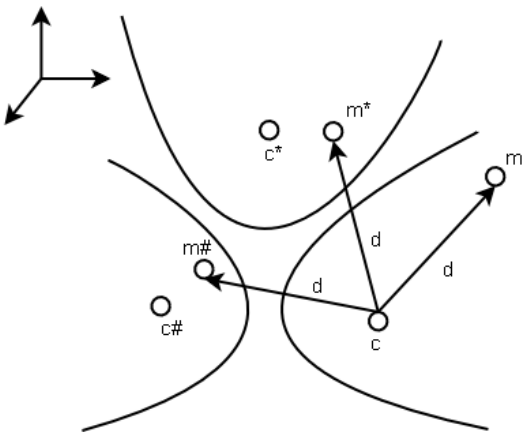

Figura 3: Vizinhanças

Este algoritmo tem a desvantagem de produzir vizinhanças que corrigem os erros de maneira dependente do valor específico do ruído e de sua amplitude. Quer dizer, o efeito do ruído sobre $c \in C$ pode produzir $m \in {\mathbb{F}_n}^m$ tal que $h(c, m) = d$, mas m pode ser corrigida para c ou $c' \in \mathcal{C}, c \neq c'$; dependendo do valor específico do erro.

Esta propriedade não nos interessa. Uma vez que o ruído é aleatório por hipótese, concluímos que nosso algoritmo também tem comportamento aleatório na correção de erros. Para eliminarmos o efeito do valor específico do ruído, adotaremos bolas fechadas disjuntas centradas em cada elemento de $\mathcal{C}$.

- O algoritmo :
  $B(c_i, \delta_i)$:= bola fechada centrada em $c_i \in \mathcal{C}$ e raio δ_i. As bolas são disjuntas e os raios são escolhidos de modo a fazer com que o conjunto das bolas contenha o maior número de elementos de $\mathcal{A}^n$.

  1. O emissor envia $c \in \mathcal{C}$;
  2. A mensagem pode ser alterada pelo ruído;
  3. O receptor recebe $m \in \mathcal{A}^n$;
  4. O receptor verifica se $m \in B(c_i, \delta_i)$, para algum $c_i \in \mathcal{C}$;
  5. caso afirmativo: corrigir m para c_i;
  6. caso negativo: rejeitar m.

Conseguimos, assim, um algoritmo que não depende do valor específico do ruído e, portanto, não aleatório. Obtemos, no entanto, uma desvantagem: reduzimos o número de palavras corrigíveis. Aquelas que não pertencem a qualquer $B(c_i, \delta_i)$ simplesmente são ignoradas. Por outro lado, poderíamos aceitar esta perda sob o argumento de que estas palavras seriam produzidas por ruído de grande amplitude. Deveremos, então, escolher o diâmetro das bolas para corrigir a amplitude esperada do erro.

E, se vamos falar na capacidade do código como um todo de corrigir erros, estaremos procurando necessariamente $R_{\mathcal{C}} = \min\{\delta_i\}$. Não é difícil de entender que $R_{\mathcal{C}}$ ficará definido quando encontrarmos a *distância mínima* $d_{\mathcal{C}}$ entre todos os pares de palavras de código. A relação entre estes dois parâmetros do código é dada por

$$d_{\mathcal{C}} = 2R_{\mathcal{C}} + 1.$$

A distância mínima é um parâmetro fundamental na descrição dos códigos: ao conhecermos $d_{\mathcal{C}}$, obtemos $R_{\mathcal{C}}$ e, conseqüentemente, a capacidade do código corrigir erros. Desta maneira, poderemos classificar os códigos em códigos corretores de 1 erro, 2 erros, ..., n erros, conforme o valor de $R_{\mathcal{C}}$.

Não há, portanto, ganho na capacidade global de correção em se tentar aumentar os diâmetros das bolas se $R_{\mathcal{C}}$ permanece igual. Ademais, se admitimos raios diferentes, a capacidade de correção, que já foi aleatória, dependerá da palavra de código enviada. Esta assimetria é indesejada, pois estamos construindo um método independente da representação da informação e, conseqüentemente, da freqüência de ocorrência das seqüências. Se fossem conhecidas as freqüências, poderíamos pensar em uma distribuição assimétrica.

- O algoritmo :
  $B_{\mathcal{C}}(c_i) :=$ bola fechada centrada em $c_i \in \mathcal{C}$ e raio $R_{\mathcal{C}}$.

  1. O emissor envia $c \in \mathcal{C}$;
  2. A mensagem pode ser alterada pelo ruído;
  3. O receptor recebe $m \in \mathcal{A}^n$;
  4. O receptor verifica se $m \in B_{\mathcal{C}}(c_i)$, para algum $c_i \in \mathcal{C}$;
  5. caso afirmativo: corrigir m para c_i;
  6. caso negativo: rejeitar m.

Conseguimos agora um algoritmo cujo comportamento corretor não é aleatório e não depende das palavras de código. Contudo, devemos observar que restringimos ainda mais o conjunto das palavras $p \in \mathcal{A}^n$ que são corrigíveis.

Mais uma vez, se conhecemos bem o comportamento médio esperado da amplitude do ruído e se $R_{\mathcal{C}}$ é suficiente para corrigir o erro médio esperado, desprezar $p \in \mathcal{A}^n \setminus \bigcup_{c_i \in \mathcal{C}} B_{\mathcal{C}}(c_i)$ não representa uma perda significativa para o processo de comunicação.

Neste ponto queremos lembrar que $\mathcal{C} \subset \mathcal{A}^n$ é o conjunto de símbolos utilizados para representar as informações que o emissor deseja enviar. Desta maneira o número de elementos de $\mathcal{C}$ está intimamente ligado à capacidade do código representar as informações, ou seja, este número significa quanto de informação o código consegue transmitir por mensagem.

Se, por exemplo, tomamos $p \in \mathcal{A}^n$ e definimos $\mathcal{C} := \{p\}$, certamente poderemos corrigir sempre de forma eficaz e eficiente todo e qualquer erro. Mas estaremos fadados a sempre enviar a mesma informação

Por esta razão procuraremos aumentar $\sharp\mathcal{C}$ para tornar o código mais eficiente na representação de informações. Isto significa tentar incluir cada vez mais bolas $B_{\mathcal{C}} \subset \mathcal{A}^n$ de maneira a diminuir o valor

$$O_{\mathcal{C}} := \sharp\mathcal{A}^n \setminus \bigcup_{c_i \in \mathcal{C}} B_{\mathcal{C}}(c_i).$$

Definição 2.4. *Dado o código $\mathcal{C} \subset \mathcal{A}^n$. Se $O_{\mathcal{C}} = 0$, então $\mathcal{C}$ será chamado de código perfeito.*

3 Códigos Lineares

Apresentamos até agora algoritmos sobre códigos que pressupõe a capacidade de armazenar e manipular todos os elementos de $\mathcal{C} \subset \mathcal{A}^n$. No entanto, se os valores de n e de $\sharp\mathcal{C}$ são muito grandes, este pressuposto interfere diretamente na eficiência e computabilidade do algoritmo. Isto acontece, pois a estrutura de conjuntos e de espaços métricos para os códigos é ineficiente para a realização computacional dos algoritmos.

Nosso alfabeto $\mathcal{A}$ até o momento só possui a estrutura de conjunto. Nosso interesse é fazer com que o efeito do ruído seja reversível. A expressão algébrica deste desejo nos leva a pensar na estrutura de grupo, e através dela corrigir o erro aplicando-se o elemento inverso do erro sobre a mensagem.

Entretanto, nosso interesse em utilizar as ferramentas da Álgebra Linear nos impõe a necessidade de duas operações definidas sobre os símbolos. Logo, a menor estrutura que atribuiremos a nossos alfabetos será a estrutura de *corpo finito*. Assim sendo, nossos alfabetos dependerão somente do número de elementos, uma vez que $\mathcal{A} := \mathbb{F}_q$ para algum $q \in \mathbb{N}$.

Nosso espaço de seqüências ${\mathbb{F}_q}^n$ será dotado da estrutura linear mais conhecida, a de espaço vetorial sobre $\mathbb{F}_q$ com produto interno usual. Desta maneira conseguiremos realizar operações algébricas sobre seqüências (vetores), definir aplicações inversíveis, bem como estabelecer relações geométricas.

Com relação aos códigos, qual o proveito que poderemos tirar desta estrutura? Se o código possui uma estrutura linear, será possível armazenarmos somente um subconjunto de elementos, a base do código. Mais do que isto, poderemos fazer cálculos e manipulações sobre matrizes, o que torna a modelagem computacional muito mais simples.

Definição 3.1. *Um código $\mathcal{C} \subset {\mathbb{F}_q}^n$ será chamado código linear se $\mathcal{C}$ for um subespaço linear. Chamaremos de $k_{\mathcal{C}}$ a dimensão do código.*

Quando não houver ambiguidade, poderemos escrever simplesmente d e k para a distância mínima e a dimensão do código $\mathcal{C}$.

3.1 Definições e algoritmos

Estabeleceremos agora diversas definições e lemas que serão úteis para a abordagem dos códigos lineares.

Definição 3.2. *A aplicação* $p : {\mathbb{F}_q}^n \to \mathbb{N}$, *definida por* $p(v) := h(v, 0)$ *é chamada de peso. Chamaremos* $p(v)$ *de peso do vetor* v.

O peso será, na verdade, uma maneira mais simples, sob certos aspectos, de observar distâncias.

Lema 3.3. *Seja* $\mathcal{C}$ *é um código linear. Então a distância mínima é dada por*

$$d_{\mathcal{C}} = \min\{\ p(c) \mid c \in \mathcal{C}\ \}$$

Prova:
Existem vetores c e $c' \in \mathcal{C}$ tais que $h(c, c') = d_{\mathcal{C}}$. Logo, $c - c' \in \mathcal{C}$ e $h(c - c', 0) = d_{\mathcal{C}} = p(c - c')$. $\diamond$

A ortogonalidade nos permite introduzir uma noção de dualidade.

Definição 3.4. *Seja* $\mathcal{C}$ *um código linear sobre* ${\mathbb{F}_q}^n$. *Definimos o código dual*

$$\mathcal{C}^{\perp} := \{p \in {\mathbb{F}_q}^n \mid \forall\ c \in \mathcal{C},\ p \cdot c = 0\}.$$

Sabemos que existe correlação linear entre os elementos de um código linear $\mathcal{C}$. Desta maneira, podemos escolher uma *base* de $\mathcal{C}$ e fazer manipulações a partir desta base.

Definição 3.5. *Seja* $b_1, ..., b_k$ *uma enumeração de todos os elementos de uma base de um código* $\mathcal{C}$. *A matriz*

$$\mathcal{G}_{\mathcal{C}} = \begin{bmatrix} b_1 \\ ... \\ b_k \end{bmatrix}$$

é chamada de matriz de geradores de $\mathcal{C}$. *Ela será uma matriz* $k \times n$.

Também será muito útil:

Definição 3.6. *Seja* $p_1, ..., p_{n-k}$ *uma enumeração de todos os elementos de uma base do código* $\mathcal{C}^{\perp}$. *A matriz*

$$\mathcal{P}_{\mathcal{C}} := \begin{bmatrix} p_1 \\ ... \\ p_{n-k} \end{bmatrix}$$

é chamada de matriz de paridade de $\mathcal{C}$. *Ela será uma matriz* $(n - k) \times n$.

Como modelar o erro neste contexto linear? Considerando que o emissor envia $c \in \mathcal{C}$ e o receptor recebe $m \in {\mathbb{F}_q}^n$, o erro e será o vetor diferença.

$$m = c + e \tag{1}$$

Sabemos que c pertence ao subespaço linear $\mathcal{C}$. O erro e, contudo, pode possuir uma componente em $\mathcal{C}$ e outra em $\mathcal{C}^\perp$. Se o erro possuir uma componente não nula em $\mathcal{C}$, a norma do erro será maior ou igual do que a norma mínima de $\mathcal{C}$ que é igual a $d_\mathcal{C}$. Neste caso, o erro ultrapassa a capacidade deste código corrigir erros. Desta maneira, só nos interessa erro cuja componente em $\mathcal{C}$ seja nula, ou seja, e pertencente a $\mathcal{C}^\perp$.

E como eliminar o erro dentro da mensagem? A idéia principal consiste em utilizar a matriz de paridade para eliminar a componente do código.

Definição 3.7. *Dado o código linear $\mathcal{C}$ e $\mathcal{P}_\mathcal{C}$ uma matriz de paridade, definimos a síndrome de um vetor por*

$$s(v) := v.\mathcal{P}_\mathcal{C}^T.$$

A síndrome é a projeção sobre o espaço dual $\mathcal{C}^\perp$. Fica claro, então, que o núcleo desta aplicação corresponde exatamente ao código $\mathcal{C}$ por definição. A síndrome não nula será, por sua vez, o sintoma de que ocorreu algum erro na transmissão. Ao aplicarmos a síndrome em (1):

$$s(m) = s(c + e) = s(c) + s(e) = s(e)$$

Assim, possuímos um algoritmo para perceber a presença do erro. Entretanto, nos interessa corrigir o erro e não somente percebê-lo. Como resolver o problema se diferentes vetores podem produzir $s(m)$?

Lema 3.8. *Dado o código linear $\mathcal{C}$, qualquer que seja $m \in {\mathbb{F}_q}^n$, existe somente um $e \in \mathcal{C}^\perp$, tal que $s(m) = s(e)$.*

Prova:
A existência é dada pela componente de m em $\mathcal{C}^\perp$. Vamos supor que existem e, $e' \in \mathcal{C}^\perp$ tais que $s(e) = s(e') = s(m)$. Pela linearidade, $s(e - e') = 0$. Logo, $e - e' \in \mathcal{C} \cap \mathcal{C}^\perp \Rightarrow e = e'.\diamond$

Entendemos, assim, o erro como um elemento de $\mathcal{C}^\perp$. Uma vez encontrado o vetor $e \in \mathcal{C}^\perp$, tal que $s(e) = s(m)$, basta fazer $c = m - e \in \mathcal{C}$ para corrigirmos a mensagem. A questão se torna, portanto, encontrar um método para determinar $e \in \mathcal{C}^\perp$ a partir de $s(m)$.

- O algoritmo:

    1. o emissor envia $c \in \mathcal{C}$;
    2. a mensagem pode ser alterada pelo ruído;
    3. o receptor recebe $m \in {\mathbb{F}_q}^n$;
    4. o receptor calcula $s(m)$;
    5. o receptor encontra o erro e associado a $s(m)$;
    6. o receptor corrige m para $m - e \in \mathcal{C}$.

Lembremos que $s(m)$ é uma projeção em $\mathcal{C}^{\perp}$. Será, no entanto, necessário nos preocuparmos em determinar esta projeção? Não seria possível aproveitarmos a matriz de geradores e estabelecermos diretamente a projeção da mensagem sobre $\mathcal{C}$?

Definição 3.9. *Sejam $\mathcal{C}$ um código linear e $\mathcal{G}_{\mathcal{C}}$ uma matriz de geradores. Definimos a aplicação*

$$f(v) := v.\mathcal{G}_{\mathcal{C}}^{T} \in \mathcal{C}$$

Esta aplicação é a projeção sobre o código $\mathcal{C}$. Basta aplicá-la à equação(1):

- O algoritmo:

    1. o emissor envia $c \in \mathcal{C}$;
    2. a mensagem pode ser alterada pelo ruído;
    3. o receptor recebe $m \in {\mathbb{F}_q}^n$;
    4. o receptor calcula $f(m)$;
    5. o receptor aceita $f(m)$ como vetor de ${\mathbb{F}_q}^n$ com componente $\mathcal{C}^{\perp}$ nula;

Esta abordagem, no entanto, não foi observada nas referências utilizadas.

3.2 Códigos MDS

Se desejarmos aumentar a capacidade de um determinado código $\mathcal{C}$ de corrigir erros, saberemos dizer que basta aumentar $d_{\mathcal{C}}$. Se, por outro lado, desejarmos aumentar a capacidade deste mesmo código de representar informações, bastaria aumentar $k_{\mathcal{C}}$. É fácil observarmos que existem limites. Mas qual seria a relação entre os parâmetros que envolvem um código linear?

Lema 3.10. *Seja $\mathcal{C}$ um código linear $\subset {\mathbb{F}_q}^n$. Então*

$$d_{\mathcal{C}} \leq n - k_{\mathcal{C}} + 1.$$

Prova:
$\mathcal{G}_{\mathcal{C}}$ é uma matriz de posto $k_{\mathcal{C}}$. Utilizando operações de eliminação gaussiana, podemos transformar $\mathcal{G}_{\mathcal{C}}$ em uma matriz da forma

$$\mathcal{G}'_{\mathcal{C}} = [I_{k_{\mathcal{C}}} | G^*]$$

A matriz $I_{k_{\mathcal{C}}}$ é matriz identidade de tamanho $k_{\mathcal{C}}$. Cada linha de G^* possui no máximo $n - k_{\mathcal{C}}$ elementos não nulos. Assim,

$$d_{\mathcal{C}} = \min\{\ p(c) \mid c \in \mathcal{C}\ \} \leq n - k_{\mathcal{C}} + 1.$$

Se a igualdade for obtida, estaremos em um caso muito especial de códigos.

Definição 3.11. *Um código linear $\mathcal{C}$ será chamado de código MDS (maximum distance separable) se $d_{\mathcal{C}} = n - k_{\mathcal{C}} + 1$.*

3.2.1 Relações com a Geometria Projetiva

Será possível obtermos algum código que alcance este limite? Esta pergunta, embora não pareça, é um problema geométrico.

Definição 3.12. *Diremos que um conjunto Θ com t elementos é um conjunto $LI(t, s)$, quando qualquer subconjunto $H \subset \Theta$, com s ou menos elementos, for LI.*

Definição 3.13. *Dados um código linear $\mathcal{C}$ e $\mathcal{P}_{\mathcal{C}}$ uma matriz de paridade. Definimos $\mathcal{H}_{\mathcal{C}} := \{h_1, h_2, ..., h_n\}$ como a enumeração de colunas de $\mathcal{P}_{\mathcal{C}}$ na ordem em que ocorrem.*

Vamos agora estabelecer a relação entre a distância mínima e os subconjuntos de $\mathcal{H}_{\mathcal{C}}$.

Lema 3.14. *Sejam $d \in \mathbb{N}$, $\mathcal{C}$ um código linear e $\mathcal{H}_\mathcal{C}$.*

$$d_\mathcal{C} \geq d \Leftrightarrow \mathcal{H}_\mathcal{C} \text{ é um conjunto } LI(n, d-1).$$

Prova:
$\Rightarrow$) Vamos supor que exista um conjunto $H \subset \mathcal{H}_\mathcal{C}, \sharp H = d' < d, H$ L.D. Seja $h_{i_1}, h_{i_2}, ..., h_{i_{d'}}$ a subseqüência H. Logo existe uma combinação linear não nula tal que

$$\sum_{h_{i_j} \in H} v_{i_j}.h_{i_j} = 0. \tag{2}$$

Seja o vetor v formado pelos valores v_{i_j} nas coordenadas i_j e valor zero nas demais coordenadas. Pela equação (2), $v \in \mathcal{C}$. Contradição, pois $p(v) \leq d' < d \Rightarrow d_\mathcal{C} < d$.
$\Leftarrow$) Seja $v = (v_1, v_2, ..., v_n) \in \mathcal{C} \setminus \{0\}$. Como $s(v) = 0$

$$v_1.h_1 + v_2.h_2 + ... + v_n.h_n = 0 \tag{3}$$

Se $p(v) < d$, chegamos a uma contradição, pois (3) ofereceria uma subseqüência H L.D., $\sharp H < d$, H formada pelas colunas que possuem coeficientes não nulos. Logo $\forall\, v \in \mathcal{C}$, $p(v) \geq d \Rightarrow d_\mathcal{C} \geq d$. $\diamond$

Esta relação ganha uma expressão geométrica, quando consideramos os elementos de $\mathcal{H}_\mathcal{C}$ como coordenadas homogêneas em um espaço projetivo finito. O Apêndice trata especificamente de espaços projetivos.

Definição 3.15. *Chamamos $PG(n, q)$ o espaço projetivo finito de dimensão n e ordem q.*

Lema 3.16. *Existirá um código $\mathcal{C}$ com distância mínima d e dimensão k no espaço $\mathbb{F}_q^n$ se e somente se existir um conjunto $LI(n, d-1) \subset PG(k-1, q)$.*

Prova:
$\Rightarrow$) Basta considerarmos as colunas h_i como coordenadas homogêneas de pontos de $PG(k-1, q)$. Este conjunto de pontos é $LI(n, d-1)$ pelo lema anterior. $\Leftarrow$) Análogo. $\diamond$

A busca por códigos MDS, portanto, se torna o problema geométrico de encontrar conjuntos $LI(n, s) \subset PG(n, q)$. Muito relacionados com estes conjuntos estão as quádricas e curvas normais.

Qualquer cônica em um plano projetivo ou cúbica normal em um espaço tridimensional é um conjunto $LI(a,3)$, em que a é o número de pontos na curva. Qualquer curva normal é um conjunto $LI(a,n)$, em que a é o número de pontos da curva e n é a dimensão do espaço projetivo em que a curva está.

A teoria em torno destas curvas pode ser encontrada em [4] e [7]. No Apêndice apresentamos de forma resumida a abordagem de [4].

3.3 Código de Hamming

Até este ponto não apresentamos qualquer exemplo concreto de código e sequer fizemos menção a algum código perfeito. Para dirimir esta lacuna, apresentaremos um dos códigos fundamentais no desenvolvimento da Teoria de Códigos: o código de Hamming.

O código de Hamming é um código linear. Para obtê-los usaremos o fato de a Álgebra Linear oferecer duas formas bem simples para construirmos subespaços: uma como Imagem e outra como Núcleo de uma transformação linear.

Definição 3.17. *Sejam $r \in \mathbb{N}$ e $n = 2^r - 1$. Seja H_r uma matriz cujas colunas sejam todos os elementos de $\mathbb{F}_2^r \setminus \{0\}$. Chamamos de código de Hamming*

$$Ham(r) := Ker(H_r : \mathbb{F}_2^n \to \mathbb{F}_2^r)$$

A capacidade de correção deste código tem propriedades conhecidas.

Lema 3.18. *$\forall\, r \in \mathbb{N}$, $R_{Ham(r)} \geq 1$.*

Prova:
Sabemos que $\forall\, c \in Ham(r)$

$$c.H_r^T = 0. \tag{4}$$

Vamos supor, por absurdo, que $\exists\, c \in Ham(r)$, tal que $p(c) = 1$. A equação (3) implica que existe uma coluna nula em H_r. Contradição. Supondo que $p(c) = 2$, temos que duas colunas de H_r devem ser iguais. Contradição. Assim $\forall\, c \in Ham(r)$, $p(c) \geq 3 \Rightarrow R_{Ham(r)} \geq 1$. $\diamond$

Lema 3.19. $\forall\, r \in \mathbb{N},\ R_{Ham(r)} = 1.$

Prova:
Seja i_1 o número da coluna de H_r que é a representação binária do número 1. De forma análoga temos i_2 e i_3. Tomemos o vetor v tal que possua valor 1 nas coordenadas i_1 i_2 e i_3 e que possua o valor zero nas demais coordenadas. Temos claramente que $p(v) = 3$ e $v \in Ham(r)$. Pelo lema anterior, $R_{\mathcal{C}} = 1$. $\diamond$

Os códigos corretores de 1 erro em $\mathbb{F}_2^n$ possuem um limite para seus números de elementos.

Lema 3.20. *Seja $\mathcal{C} \subset \mathbb{F}_2^n$, tal que $R_{\mathcal{C}} = 1$, então*

$$\sharp\mathcal{C} \leq \frac{2^n}{n+1}.$$

A igualdade só acontece se $\mathcal{C}$ for perfeito.

Prova:
$\forall\, c \in \mathcal{C}$, $B(c)$ será uma bola de raio igual a $R_{\mathcal{C}} = 1$ centrada em c.
$\sharp B(c) = n + 1$. A união de todas as bolas está contida no espaço $\mathbb{F}_2^n$. Como as bolas são disjuntas, $\sharp\mathcal{C}.\sharp B(0) \leq \sharp\mathbb{F}_2^n = 2^n$. Logo

$$\mathcal{C} \text{ é perfeito } \Leftrightarrow \sharp\mathcal{C} = \frac{2^n}{n+1} \diamond$$

Os código de Hamming não são quaisquer códigos corretores de 1 erro, mas sim os melhores destes códigos.

Lema 3.21. $\forall\, r \in \mathbb{N},\ Ham(r)$ *é perfeito.*

Prova:
Lembrando que $n = 2^r - 1$, temos que $\dim(H_r) = r \Rightarrow \dim(Ham(r)) = n - r \Rightarrow \sharp Ham(r) = 2^{n-r} \Rightarrow \sharp Ham(r).(n+1) = 2^{n-r}.2^r = 2^n = \sharp\mathbb{F}_2^n$. Pelo lema anterior, $Ham(r)$ é perfeito. $\diamond$

Como exemplo da elegância destes códigos, apresentamos:

Teorema 3.22. *Seja $Ham(r)$ construído através de uma matriz H_r^* cuja coluna s_i é a representação binária do inteiro i. Seja e_i o vetor que tem valor 1 na coordenada i e zero nas demais. Então, para qualquer $v \in \mathbb{F}_2^n$, $s(v)$ é a representação binária da coordenada onde ocorreu o erro.*

Prova:
Como $Ham(r)$ é perfeito, para qualquer $v \in \mathbb{F}_2^n \setminus Ham(r)$, existem $i \in \mathbb{N}$ e $c \in Ham(r)$ tais que $c = v - e_i \in Ham(r)$. Assim

$$s(v) = v.H^T = (c + e_i).H^T = e_i.H^T = s_i$$

Logo $s(v)$ é a coluna s_i que representa o número i. $\diamond$

3.3.1 Relações com a Geometria Projetiva

Os códigos de Hamming guardam profunda relação com a Geometria Projetiva, em especial com as retas de um espaço projetivo. A partir deste ponto seguimos a abordagem proposta em [1].

Definição 3.23. *Dado um conjunto U, chamaremos 2^U como o conjunto das partes de U.*

Definição 3.24. *Dado um espaço projetivo $PG(n,2)$, definimos uma enumeração $E_{PG(n,2)}$ de todos os seus pontos de acordo com a seqüência de valores das coordenadas homogêneas interpretadas na base binária.*

Assim podemos entender que $(0:1:0:1) \in PG(4,2)$ é representado na enumeração $E_{PG(4,2)}$ pelo ponto P_5, pois 0101 na base binária equivale a 5 na base decimal.

Definição 3.25. *Dada uma enumeração $E_{PG(n,2)} = \{P_1, P_2, ..., P_{2^n+...+2+1}\}$ de $PG(n,2)$. Definimos uma aplicação*

$$S : \mathbb{F}_2^{2^n+...+2+1} \longrightarrow 2^{E_{PG(n,2)}}$$

tal que $\forall\, v \in \mathbb{F}_2^{2^n+...+2+1}, P_i \in S(v) \Leftrightarrow e_i.v = 1$. Diremos, neste caso, que v é o vetor característico de $S(v)$.

Desta maneira, o vetor $v = (0,0,0,0,0,1,0,0,0,1,1,0,0,0,0)$ é levado, pela aplicação S, ao conjunto $S(v) = \{P_5, P_6, P_{10}\}$.

Sabemos que a matriz de paridade H_r^* de $Ham(r)$ tem $2^r - 1$ colunas. Como $2^r - 1 = 2^{r-1} + 2^{r-2} + ... + 2 + 1 = \sharp PG(r-1,2)$, consideraremos o espaço projetivo $PG(r-1,2)$ para estabelecermos as associações geométricas dos códigos de Hamming.

Definição 3.26. *Chamaremos de l_i a i-ésima linha de H_r^*.*

Definição 3.27. *Chamaremos de π_i o hiperplano $x_i = 0 \subset PG(r-1,2)$ e $\overline{\pi_i}$ seu complemento.*

Pela definições sabemos que

$$S(l_i) = \overline{\pi_i} \tag{5}$$

Lema 3.28. *$\sharp S(u) \cap S(v)$ é número par $\Leftrightarrow$ $u.v = 0$.*

Prova:
Para que o produto escalar seja nulo, o número de termos iguais a 1 na soma deve ser par. Sabemos que

$$u_i.v_i = 1 \Leftrightarrow u_i = v_i = 1 \Leftrightarrow P_i \in S(u) \wedge P_i \in S(v) \Leftrightarrow P_i \in S(u) \cap S(v). \diamond$$

Para qualquer $v \in Ham(r)$,

$$v.H_r^T = 0.$$

Logo $v \in Ham(r) \Leftrightarrow v.l_i = 0, \forall\ i \in \{1, ..., r\}$. Usando o lema anterior, $v \in Ham(r) \Leftrightarrow S(v) \cap \overline{\pi_i}$ tem número par de elementos $\forall\ i \in \{1, ..., r\}$.

Por esta razão estaremos interessados em subconjuntos $\mathcal{T} \in Q(\tilde{P})$ tais que

$$\mathcal{T} \cap \overline{\pi_i} = \mathcal{T} \setminus \pi_i$$

tem um número par de pontos $\forall\ i \in \{1, ..., r\}$.

Lema 3.29. *Todo vetor característico de um subespaço de $PG(r-1, 2)$ pertence a $Ham(r)$.*

Prova:
Se $\mathcal{T} \in Q(\tilde{P})$ é subespaço de $PG(r-1, 2)$, $\dim(\mathcal{T}) = t$ então
$\sharp\mathcal{T} = 2^t + 2^{t-1} + ... + 2 + 1$ que é ímpar. Logo

$$\sharp\mathcal{T} \setminus \pi_i = \sharp\mathcal{T} - \sharp(\mathcal{T} \cap \pi_i) \equiv 1 - 1 = 0.$$

Desta maneira $\sharp\mathcal{T} \setminus \pi_i$ é par $\forall\ i \in \{1, ...r\}$ e o vetor característico de $\mathcal{T}$ pertence a $Ham(r)$. $\diamond$

Lema 3.30. *Seja $\mathcal{T}$ um subconjunto qualquer não vazio de $PG(r-1, 2)$. Então existe $i \in \{1, ..., r\}$ tal que*

$$\mathcal{T} \cap \overline{\pi_i} \neq \emptyset$$

Prova:
Seja $P \in \mathcal{T}$. P possui uma coordenada i não nula. Assim $\mathcal{T} \cap \overline{\pi_i} \neq \emptyset$. $\diamond$

Definição 3.31. *Dados dois conjuntos A e B, definimos a diferença simétrica entre A e B,*

$$A \Delta B := (A \cup B) \setminus (A \cap B)$$

Lema 3.32. $(A\Delta B)\Delta B = A$

Lema 3.33. *Se $S(u)$ e $S(v)$ são subespaços, então $S(u)\Delta S(v) = S(u+v)$.*

Prova:
$\Rightarrow$) Seja $P_i \in (S(u) \cup S(v)) \setminus (S(u) \cap S(v))$. Sem perda de generalidade, vamos supor que $P_i \in S(u)$ e $P_i \notin S(v)$. Logo, $u_i = 1$ e $v_i = 0$, portanto, $u_i + v_i = 1$. Desta maneira $P_i \in S(u+v)$ e
$(S(u) \cup S(v)) \setminus (S(u) \cap S(v)) \subset S(u+v)$. $\Leftarrow$) Análogo. $\diamond$

Sabemos que $Ham(r)$ é um código linear, portanto, u, $v \in Ham(r) \Rightarrow u + v \in Ham(r)$. Assim, se dois conjuntos A e B possuem vetores característicos em $Ham(r)$, $A\Delta B$ também possuirá vetor característico em $Ham(r)$.

O próximo teorema apresenta a relação entre os elementos de $Ham(r)$ e subconjuntos especiais de $PG(r-1,2)$, a saber: as retas.

Teorema 3.34. *As palavras de código de $Ham(r)$ são os vetores característicos de todos os subconjuntos de $PG(r-1,2)$ obtidos a partir de diferenças simétricas de retas de $PG(r-1,2)$.*

Prova:
Sabemos, pelo lema 3.28, que o vetor característico de qualquer reta de $PG(r-1,2)$ pertence a $Ham(r)$. Sabemos também que a diferença simétrica de conjuntos com vetor característico em $Ham(r)$ resulta em conjunto com vetor característico também em $Ham(r)$. Falta mostrarmos que qualquer vetor de $Ham(r)$ é vetor característico de algum conjunto obtido por diferenças simétricas de retas. Seja $v \in Ham(r)$. Pelo lema 3.18, sabemos que $\sharp S(v) \geq 3$. Vamos considerar dois pontos distintos de $S(v)$. Seja r_1 a reta que passa por esses dois pontos. Logo o conjunto $S(v)_1 = S(v)\Delta r_1$ tem no máximo $\sharp S(v) - 1$ pontos, pois tiramos ao menos dois pontos e acrescentamos no máximo um. Temos que $S(v) = S(v)_1\Delta r_1$. Se $\sharp S(v)_1 \geq 2$ repetimos o processo um número finito de vezes até que

$$S(v) = S(v)_j\Delta r_j\Delta ...\Delta r_2\Delta r_1$$

em que $S(v)_j$ é vazio ou unitário. Se $S(v)_j$ é vazio, resolvido. Se $S(v)_j$ é unitário, então o vetor característico de $S(v)$ não pertence a $Ham(r)$, pela prova do lema 3.27 e pela observação após o mesmo lema, pois $S(v) \cap \overline{\pi_i}$ tem número ímpar de pontos. Contradição. $\diamond$

4 Códigos Algébricos

Do ponto de vista que abordamos, a questão essencial da Teoria de Códigos é encontrar subespaços vetoriais de dimensão finita com propriedades conhecidas, a saber, distância mínima e dimensão. Esta busca não se restringiu ao domínio das matrizes, se estendendo, inclusive, sobre espaços de dimensão infinita, especialmente sobre os anéis de polinômios sobre corpos finitos.

Contudo, existe uma íntima ligação entre os anéis de polinômios e o estudo de variedades pela Geometria Algébrica. Esta é a razão do interesse por parte da Teoria de Códigos pelo estudo dos espaços vetoriais sobre variedades, em especial sobre curvas algébricas projetivas.

Do ponto de vista histórico a passagem dos códigos de Hamming para os códigos corretores algébricos corresponde à passagem da Geometria Projetiva para a Geometria Algébrica. Para realizar esta passagem, seguiremos muitas vezes a seqüência proposta em [10].

Para a construção de um código corretor algébrico, utilizaremos, como nos de Hamming, transformações lineares. No entanto, em vez de tomarmos o Núcleo, faremos uso da Imagem da transformação. As transformações serão obtidas a partir de espaços vetoriais de funções definidas sobre uma variedade com imagem em um corpo. Fica claro que, para conhecermos melhor os parâmetros fundamentais dos códigos, precisaremos utilizar a teoria que envolve as variedades.

Inicialmente nossos objetos geométricos serão variedades de dimensão qualquer. Posteriormente faremos a restrição para tratarmos somente curvas.

As variedades são obtidas como raízes de polinômios. Portanto, o espaço onde estão definidos estes objetos possui dimensão infinita. Nosso objetivo, entretanto, é construir espaços vetoriais de dimensão finita.

Precisamos, então, de algum mecanismo que estabeleça restrições quanto às dimensões dos subespaços que iremos construir. A maneira mais simples de se fazer isto é restringir-se diretamente o grau dos polinômios envolvidos na construção do subespaço. No que se segue $\mathbb{F}$ será sempre o fecho algébrico de $\mathbb{F}_q$.

Definição 4.1. *Seja $G \in \mathbb{F}_q[x_1, ..., x_s]$. Definimos a variedade*

$$\mathcal{V}(G) = \{P \mid G(P) = 0\}$$

O grau da variedade é o grau de G.

Definição 4.2. *Seja $\mathcal{X}$ uma variedade definida sobre $\mathbb{F}_q$. Diremos que um ponto $P \in \mathcal{X}$ é um ponto racional de $\mathcal{X}$ se cada coordenada de P pertencer a $\mathbb{F}_q$.*

Definição 4.3. *Seja $V_l \subset \mathbb{F}_q[x_1, ..., x_s]$ o subespaço vetorial de polinômios de grau menor ou igual a l.*

Seja G um polinômio de grau m em $\mathbb{F}_q[x_1, ..., x_s]$ irredutível em $\mathbb{F}[x_1, ..., x_s]$. Sejam $P_1, ..., P_n$ pontos racionais da variedade $\mathcal{V}(G)$. Podemos definir um código sobre esta variedade:

$$\mathcal{C} := \{(F(P_1), ..., F(P_n)) \mid F \in V_l\}$$

Os parâmetros deste código são dados por

Teorema 4.4. *Seja $n > lm$. Com as definições acima temos*

$$d_{\mathcal{C}} \geq n - lm,$$

$$k_{\mathcal{C}} = \begin{cases} \binom{l+n}{n} \text{ se } l < m \\ \binom{l+n}{n} - \binom{l-m+n}{n} \text{ se } l \geq m \end{cases}$$

Prova:
O espaço V_l tem por base os monômios da forma $x_1^{\alpha_1} x_2^{\alpha_2} ... x_n^{\alpha_n}$, $\sum \alpha_i \leq l$. Logo $\dim V_l = \binom{l+n}{n} = \binom{l+n}{l}$. Seja $F \in V_l$. Se G divide F, então F gera o vetor $0 \in \mathcal{C}$. Por outro lado, se o polinômio F gera $0 \in \mathcal{C}$, então as equações $F = 0$ e $G = 0$ têm os pontos $P_1, ..., P_n$ em sua intersecção. Lembrando que F e G têm graus $l' \leq l$ e m respectivamente, pelo teorema de Bézout e pela hipótese $n > lm$, temos que F e G têm fator comum. Como G é irredutível, F é divisível por G. Então o conjunto das funções que geram $0 \in \mathcal{C}$ corresponde ao conjunto GV_{l-m}. Portanto, se $l < m$, $k_{\mathcal{C}} = \binom{l+n}{n}$. De outra maneira, se $l \geq m$, então $k_{\mathcal{C}} = \binom{l+n}{n} - \binom{l-m+n}{n}$. O mesmo argumento, através do teorema de Bézout, mostra que $\forall\, c \in \mathcal{C}$, $c \neq 0$, c tem no máximo lm coordenadas nulas, e, portanto, $p(c) \geq n - lm$. Temos assim que $d_{\mathcal{C}} \geq n - lm$. $\diamond$

Uma matriz de geradores deste código pode facilmente ser obtida a partir de uma base $F_1, ..., F_k$ de V_l/GV_{l-m}:

$$\{F_i(P_j) \mid 1 \leq i \leq k, 1 \leq j \leq n\}$$

Um observação, no entanto, deve ser feita. Este exemplo é apresentado em um espaço afim. Por esta razão, é possível trabalharmos com funções regulares sobre toda a variedade, simplesmente controlando a soma dos graus das funções. Isto deixa de ser verdade se temos interesse em trabalhar em espaços projetivos.

Sobre espaços projetivos nossos corpos de funções sobre uma variedade, para serem interessantes, devem admitir funções com singularidades. Neste contexto, somente as frações de polinômios homogêneos de mesmo grau fazem sentido. E, por esta razão, não poderíamos utilizar o mesmo raciocínio de controle da dimensão a partir do grau da função.

O controle do grau, no caso afim, equivale a restringir a soma das multiplicidade dos zeros do polinômio. Poderíamos, então, pensar em controlar, no caso projetivo, as somas das multiplicidades de zeros e pólos de uma função. Mas esta soma é sempre zero e a dimensão, assim, não se torna finita.

Se, em vez de controlarmos simplesmente a soma, impusermos restrições sobre as multiplicidades de zeros e pólos das funções em cada ponto, obteremos um espaço de funções mais interessante. Poderemos fazer isto através de uma definição muito importante para a Geometria Algébrica: a definição de divisor.

É uma opção muito útil, pois a teoria dos divisores nos oferece o grupo dos divisores com uma ordem parcial e uma relação de equivalência bem definidas, ferramentas que nos auxiliaram na construção de espaços vetoriais. Ademais, sobre os divisores, já temos disponível um espaço vetorial de dimensão finita já definido.

4.1 Divisores

Neste momento nos restringimos às curvas algébricas projetivas. No que segue $\mathcal{X}$ será sempre uma curva projetiva regular irredutível sobre um corpo algebricamente fechado $\mathbb{F}$.

Definição 4.5. *Um divisor é uma soma $D = \sum_{P \in \mathcal{X}} n_P P$, em que $n_P \in \mathbb{Z}$ e grau$\{D\} := \sum n_P < \infty$. O suporte de um divisor é o conjunto de pontos tais que $n_P \neq 0$. Diremos também que um divisor é efetivo se todos os seus coeficientes forem não negativos.*

É interessante observar que o conjunto de todos os divisores sobre uma determinada curva forma um grupo com respeito à soma. Denotaremos este grupo por $Div(\mathcal{X})$.

Como estamos em um espaço projetivo, nossas funções sobre curvas são todas racionais. A relação entre as funções e os divisores é dada por

Definição 4.6. *Seja ϕ uma função racional sobre $\mathcal{X}$, não identicamente nula, definimos o divisor de ϕ como*

$$(\phi) = \sum_{P \in \mathcal{X}} v_P(\phi)\ P,$$

em que $v_P(\phi)$ é a multiplicidade de zero ou de pólo da função ϕ em P. Chamaremos os divisores de funções racionais sobre $\mathcal{X}$ de divisores principais.

Sabemos que as funções racionais que fazem sentido em espaços projetivos devem possuir numerador e denominador com mesmo grau. Assim, é natural entendermos que o somatório das multiplicidades de zeros e pólos de um divisor de uma função racional deve respeitar

Lema 4.7. *O grau de um divisor principal é igual a zero.*

Estes divisores, dentro do grupo dos divisores, são muito importantes, pois definem uma relação de equivalência em $Div(\mathcal{X})$ que será a base para a construção de espaço vetoriais de dimensão finita definidos sobre variedades projetivas. Esta relação de equivalência fará a separação dos divisores pelo grau de cada divisor. Desta maneira, as classes de equivalência serão fechadas com respeito à soma com divisores de grau zero.

Definição 4.8. *Dados dois divisores D e D' sobre $\mathcal{X}$, diremos que são linearmente equivalentes se $D - D'$ for um divisor principal. Denotaremos esta equivalência por $D \equiv D'$.*

Contudo, apesar de o divisor descrever de maneira precisa as multiplicidades, ainda não seremos capazes de controlar a dimensão do espaço de funções pelo controle de multiplicidades. A razão disto é a carência de uma ordem ao menos parcial que nos permita estabelecer limites à dimensão dos espaços a serem construídos.

Uma ordem parcial nos permite construir conjuntos de elementos que respeitam uma relação de ordem com um limite estabelecido e isto nos é suficiente. Adotar uma ordem total acrescenta somente a informação de ordem entre os elementos do conjunto definido. Por esta razão não precisamos desta definição mais restritiva.

A definição de ordem parcial que iremos utilizar já foi apresentada na definição de divisor. Quando dizemos que um divisor D é efetivo, garantimos que, para cada ponto P de $\mathcal{X}$, o coeficientes de P é maior ou igual ao coeficiente do ponto P no divisor 0. Este fato denotamos por $D \succeq 0$. Desta maneira, a ordem parcial exige que comparemos os coeficientes dos divisores em cada ponto da curva.

4.1.1 Espaço vetorial de funções

De posse das definições de divisor, grupo de divisores e ordem parcial; chegamos à construção de um espaço vetorial de dimensão finita de funções racionais sobre uma curva projetiva regular irredutível sobre um corpo algebricamente fechado.

Definição 4.9. *Dado um divisor D sobre $\mathcal{X}$, definimos o espaço vetorial*

$$\mathcal{L}(D) = \{\phi \in \mathbb{F}(\mathcal{X})^* \mid (\phi) + D \succeq 0\} \cup \{0\}.$$

A dimensão de $\mathcal{L}(D)$ será denotada por $l(D)$.

É simples observar que $\mathcal{L}(D)$ é um espaço vetorial. Mas em que consiste este conjunto ? E mais: a dimensão $l(D)$ será sempre finita ?

Sabemos que qualquer divisor pode ser escrito com diferença de divisores efetivos de tal forma que $D = D' - D'' = \sum_{i=1}^{r} n_i P_i - \sum_{j=1}^{s} m_j Q_j$, D' e D'' efetivos. Logo a definição de $\mathcal{L}(D)$ exige que as funções possuam zeros de multiplicidade no mínimo m_j no pontos Q_j e pólos de multiplicidade no máximo n_i no pontos P_i.

Para responder à segunda questão, temos o teorema:

Teorema 4.10. *Seja D um divisor sobre a curva $\mathcal{X}$:*
i) $l(D) = 0$, se $grau(D) < 0$;
ii) $l(D) \leq 1 + grau(D)$, se $grau(D) \geq 0$.

A prova deste teorema exige lema e definições.

Lema 4.11. *Sejam D e D' divisores sobre $\mathcal{X}$, tais que $D \equiv D'$, então $\mathcal{L}(D)$ e $\mathcal{L}(D')$ são isomorfos.*

Prova:
$D \equiv D' \Rightarrow \exists\, \psi \in \mathbb{F}(\mathcal{X})^*$ tal que $(\psi) = D - D'$. Tomemos a aplicação $i : \phi \mapsto \phi\psi$ entre $\mathcal{L}(D)$ e $\mathcal{L}(D')$. O núcleo desta aplicação linear é $\{0\}$. Logo, i é injetora. Tomemos $\nu \in \mathcal{L}(D')$. A seqüência abaixo mostra que ν pertence à imagem de i. A álgebra em torno dos divisores pode ser encontrada em [7].

$$\begin{aligned}(\nu) + D' &\succeq 0\\ (\nu) + D' - D &\succeq -D\\ (\nu) - (\psi) &\succeq -D\\ (\nu) + (\psi^{-1}) &\succeq -D\\ (\nu\psi^{-1}) + D &\succeq 0\end{aligned}$$

i, portanto, é sobrejetora. Assim i é isomorfismo. $\diamond$

Definição 4.12. *Chamamos de anel local $\mathcal{O}_P(\mathcal{X})$ o conjunto das funções racionais sobre a curva $\mathcal{X}$ que são regulares no ponto P.*

Este anel é de fato um anel local, pois só tem um ideal maximal: o conjunto $\mathcal{M}_P(\mathcal{X})$ das funções racionais sobre a variedade que possuem o valor zero no ponto P. Sabemos que $\mathcal{M}_P(\mathcal{X})$ é também um ideal principal.

Definição 4.13. *Chamamos de parâmetro local qualquer gerador de $\mathcal{M}_P(\mathcal{X})$.*

Como um parâmetro local t é um gerador do anel maximal $\mathcal{M}_P(\mathcal{X})$, podemos escrever qualquer função $z \in \mathcal{O}_P(\mathcal{X})$ de forma única como $z = ut^m$, u não-nula e $m \in \mathbb{N}_0$. Se $m > 0$, dizemos que P é um zero de ordem m de z. Se $m < 0$, P será um pólo de ordem m de z. Escrevemos $m = v_P(z)$ para definirmos a valoração de z no ponto P. Por convenção dizemos que $v_P(0) = \infty$.

Prova do teorema 4.10:

1. $\forall\ \phi \in \mathbb{F}(\mathcal{X})^*$, $grau((\phi)) = 0 \Rightarrow grau((\phi) + D) < 0 \Rightarrow$ $L(D) = \{0\} \Rightarrow l(D) = 0$.

2. Pelo lema acima, se D não é efetivo e $\phi \in \mathcal{L}(D)^*$, $D' = D + (\phi)$ é efetivo. Sabemos que $\mathcal{L}(D)$ e $\mathcal{L}(D')$ são isomorfos e, portanto, tem mesma dimensão. Logo podemos supor que $D = \sum_{i=1}^{r} n_i P_i$ é um divisor efetivo. Vamos supor que $\phi \in \mathcal{L}(D)^*$. No ponto P_i, ϕ pode ter pólo de multiplicidade no máximo n_i. Seja t_i um parâmetro local no ponto P_i. Logo ϕ pertence ao espaço vetorial n_i-dimensional $(t^{-n_i}\mathcal{O}_{P_i})/\mathcal{O}_{P_i}$. Em cada ponto P_i, mapearemos ϕ para a função correspondente no espaço $(t^{-n_i}\mathcal{O}_{P_i})/\mathcal{O}_{P_i}$. Esta é uma aplicação linear de ϕ na soma direta destes espaços vetoriais. Se ϕ está no núcleo da aplicação, ϕ não tem pólos, logo é uma função constante. Assim obtemos
$$l(D) \leq 1 + \sum_{i=1}^{r} n_i = 1 + grau(D).\diamond$$

4.2 Códigos de Reed-Solomon

Podemos agora apresentar uma definição inicial de uma classe de códigos corretores algébricos:

Definição 4.14. *Sejam $D = P_1 + P_2 + ... + P_n$ e G divisores sobre $\mathcal{X}$ com suportes disjuntos. O código linear C(D,G) de tamanho n sobre $\mathbb{F}$ é a imagem da aplicação linear $\alpha : \mathcal{L}(G) \to \mathbb{F}^n$ definida por $\alpha(\phi) = (\phi(P_1), \phi(P_2), ..., \phi(P_n))$. Códigos deste tipo são chamados de códigos de Reed-Solomon.*

Esta definição corresponde de fato a um sub-espaço em $\mathbb{F}^n$. Contudo, precisaremos fazer adaptações para tornar estes códigos aplicáveis ao uso em computadores.

Precisamos, em primeiro lugar, que nosso alfabeto seja finito. Tomemos como alfabeto $\mathbb{F}_q$ com fecho algébrico $\mathbb{F}$. Nossas funções somente poderão ter coeficientes em nosso alfabeto.

Definição 4.15. *Uma curva irredutível é absolutamente irredutível se for irredutível também sobre o fecho algébrico do corpo.*

Tomaremos somente curvas que sejam absolutamente irredutíveis.

Definição 4.16. *Seja a curva $\mathcal{X}$ definida sobre o corpo $\mathbb{F}_q$, a aplicação $Fr : \mathbb{F} \to \mathbb{F}$ definida por $Fr(x) = x^q$ é chamada de aplicação de Frobenius. Para vetores, aplicamos coordenada a coordenada.*

Definição 4.17. *Um divisor D é chamado racional se os coeficientes de P e $Fr(P)$ forem os mesmos para todos os pontos da curva $\mathcal{X}$.*

Sabemos que para qualquer elemento $x \in \mathbb{F}_q$, $Fr(x) = x$. A partir da aritmética finita, temos que $G(x_1, ..., x_n)^q = G(x_1^q, ..., x_n^q)$ para qualquer polinômio $G \in \mathbb{F}_q[x_1, ..., x_n]$. Logo, se $\mathcal{X}$ é uma curva sobre $\mathbb{F}_q$,

$$P \in \mathcal{X} \Leftrightarrow Fr(P) \in \mathcal{X}.$$

Por outro lado, se P é um ponto racional de $\mathcal{X}$, cada coordenada pertence a $\mathbb{F}_q$. Logo, para qualquer ponto racional, $Fr(P) = P$. Por esta razão, se um divisor tem suporte composto somente por pontos racionais, ele será racional.

O espaço $\mathcal{L}(D)$ somente será considerado sobre divisores racionais com a restrição das funções com coeficientes em $\mathbb{F}_q$. Segundo Høholdt[10], estas restrições garantem que o teorema de Riemann-Roch continua sendo válido. Este teorema será fundamental para determinarmos os parâmetros dos códigos algébricos.

Podemos agora apresentar um nova definição dos códigos de Reed-Solomon.

Definição 4.18. *Sejam $P_1, P_2, ..., P_n$ pontos racionais de $\mathcal{X}$. Sejam $D = P_1 + P_2 + ... + P_n$ e G divisores sobre esta curva com suportes disjuntos. O código linear $C(D, G)$ de tamanho n sobre $\mathbb{F}_q$ é a imagem da aplicação linear $\alpha : \mathcal{L}(G) \to \mathbb{F}_q^n$ definida por $\alpha(\phi) = (\phi(P_1), \phi(P_2), ..., \phi(P_n))$.*

Exemplo 4.19. Seja $\mathcal{X}$ a reta projetiva sobre $\mathbb{F}_{q^m}$. Seja $n = q^m - 1$. Definimos $P_0 = (0 : 1)$, $P_\infty = (1 : 0)$ e $P_j = (\beta^j : 1)$, β é raiz n-ésima da unidade, $1 \leq j \leq n$. Tomemos os divisores $D = \sum_{j=1}^{n} P_j$ e $G = aP_0 + bP_\infty$, $a \geq 0$, $b \geq 0$. Podemos observar que as funções $(x/y)^i$, $-a \leq i \leq b$, formam uma base de $\mathcal{L}(G)$. Uma matriz de geradores de $C(D, G)$ é dada por uma matriz formada por linhas $(\beta^i, \beta^{2i}, ..., \beta^{ni})$, $-a \leq i \leq b$.

Que tenhamos obtido um sub-espaço vetorial não existe dúvida. Porém, ainda não conhecemos os parâmetros fundamentais destes códigos: a distância mínima $d_\mathcal{C}$ e a dimensão do código $k_\mathcal{C}$. Para resolver esta questão, utilizaremos o teorema de Riemann-Roch, que relaciona propriedades da curva $\mathcal{X}$ com a dimensão $l(D)$ de divisores sobre esta curva.

Neste ponto não fizemos qualquer restrição sobre os graus dos divisores D e G. Mais adiante abriremos mão desta liberdade com o propósito de estabelecer limites que nos facilitem conhecer com precisão os parâmetros dos códigos. Para alcançar este objetivo, deveremos passar pelas definições de diferenciais sobre uma curva.

4.3 Diferenciais sobre uma curva

O estudo das diferenciais nos permite conhecer as propriedades intrínsecas à curva. Ao conhecermos as propriedades de uma curva poderemos determinar os parâmetros dos códigos que construirmos sobre ela. Esta será nossa motivação para seguirmos os passos até a apresentação do teorema de Riemann-Roch.

Definição 4.20. *Seja $\mathcal{V}$ um espaço vetorial sobre $\mathbb{F}(\mathcal{X})$. Um aplicação $\mathbb{F}$-linear $\mathcal{D} : \mathbb{F}(\mathcal{X}) \to \mathcal{V}$ é chamada uma derivação se satisfaz a regra do produto:*

$$\mathcal{D}(\phi\psi) = \phi\mathcal{D}(\psi) + \psi\mathcal{D}(\phi).$$

Definição 4.21. *O conjunto de todas as derivações será denotado por $Der(\mathcal{X}, \mathcal{V})$. Se $\mathcal{V} = \mathbb{F}(\mathcal{X})$, denotaremos por $Der(\mathcal{X})$.*

Lema 4.22. *O conjunto de derivações é um espaço vetorial sobre $\mathbb{F}$.*

Prova:
Basta apresentarmos a soma de vetores e a multiplicação por escalares. Dadas duas derivações $\mathcal{D}_1$ e $\mathcal{D}_2 \in Der(\mathcal{X}, \mathcal{V})$, definimos a soma de duas derivações $(\mathcal{D}_1 + \mathcal{D}_2)(\phi) = \mathcal{D}_1(\phi) + \mathcal{D}_2(\phi)$. Definimos a multiplicação de uma derivação $\mathcal{D}$ por um escalar $\phi \in \mathbb{F}(\mathcal{X})$ como $(\phi\mathcal{D})(\psi) = \phi\mathcal{D}(\psi)$. $\diamond$

A primeira relação entre as propriedades da curva e as derivações é dada por

Lema 4.23. *Seja t um parâmetro local de um ponto $P \in \mathcal{X}$. Então existe uma única derivação $D_t \in Der(\mathcal{X})$ tal que $D_t(t) = 1$. Além disto $Der(\mathcal{X})$ é unidimensional sobre $\mathbb{F}(\mathcal{X})$ e, conseqüentemente, D_t é uma base de $Der(\mathcal{X})$ para cada parâmetro local t.*

Prova:
Construiremos uma transformação $D_t : \mathbb{F}(\mathcal{X}) \to \mathbb{F}(\mathcal{X})$, atribuindo os valores $D_t(t) = 1$ e $D_t(\phi) = 0$ para qualquer $\phi \notin \mathcal{M}_P$. Impomos que ela respeita a regra do produto. Usando o fato que $\mathcal{M}_P$ é ideal principal, obtemos que $D_t \in Der(\mathcal{X})$. O teorema do núcleo e da imagem de uma transformação linear impõe a $Der(\mathcal{X})$ ser unidimensional sobre $\mathbb{F}(\mathcal{X})$. $\diamond$

Definição 4.24. *Uma forma diferencial racional ou simplesmente diferencial sobre $\mathcal{X}$ é uma aplicação $\mathbb{F}(\mathcal{X})$-linear de $Der(\mathcal{X})$ em $\mathbb{F}(\mathcal{X})$. O conjunto de todas as diferenciais em $\mathcal{X}$ é denotado por $\Omega(\mathcal{X})$.*

Assim definida, entenderemos a diferencial como um funcional linear que atua sobre o espaço vetorial das derivações. Ou seja, as diferenciais são elementos do espaço dual ao espaço de derivações. Esta dualidade oferecerá mais adiante uma visão unificadora dos códigos e, de maneira muito elegante, se relacionará com a definição de dualidade entre códigos.

Como $\Omega(\mathcal{X})$ é um espaço vetorial sobre $\mathbb{F}(\mathcal{X})$, podemos definir uma derivação

$$d : \mathbb{F}(\mathcal{X}) \to \Omega(\mathcal{X})$$

tal que $d(\phi) := d\phi : Der(\mathcal{X}) \to \mathbb{F}(\mathcal{X})$ é definida por $d\phi(\mathcal{D}) = \mathcal{D}(\phi)$ para toda derivação $\mathcal{D} \in Der(\mathcal{X})$. De acordo com a notação $d \in Der(\mathcal{X}, \Omega(\mathcal{X}))$.

Teorema 4.25. *$\Omega(\mathcal{X})$ é um espaço vetorial de dimensão 1 sobre $\mathbb{F}(\mathcal{X})$ e dt é uma base em cada ponto P com parâmetro local t.*

A prova deste teorema é semelhante à prova do teorema anterior.

A partir deste teorema podemos concluir que, para cada ponto $P \in \mathcal{X}$ com parâmetro local t_P, uma diferencial ω pode ser representada de forma única por $\omega = \phi_P dt_P$, $\phi_P \in \mathbb{F}(\mathcal{X})$.

Como dt_P é uma base, ω acompanha a multiplicidade de zero ou de pólo da função racional ϕ_P. Por esta razão

Definição 4.26. *Seja ω uma diferencial sobre $\mathcal{X}$, representada em um ponto $P \in \mathcal{X}$ por $\omega = \phi_P dt_P$. A ordem ou valoração de ω em P é definida por $ord_P(\omega) := v_P(\omega) := v_P(\phi_P)$. A diferencial é dita regular se não apresenta pólos. O conjunto das diferenciais regulares sobre $\mathcal{X}$ formam um $\mathbb{F}[\mathcal{X}]$-módulo denotado por $\Omega[\mathcal{X}]$.*

Esta informação sobre zeros e pólos nos permite

Definição 4.27. *O divisor (ω) de uma diferencial é definido por*

$$(\omega) = \sum_{P \in \mathcal{X}} v_P(\omega) P.$$

Interessante observar que o número de pontos com coeficiente não-nulo deste divisor é finito, pois as funções racionais têm número finito de zeros e pólos.

Seja ω uma diferencial e $W = (\omega)$. Chamaremos W de *divisor canônico*, pois para qualquer outra diferencial ω' sobre $\mathcal{X}$, existe $\phi \in \mathbb{F}(\mathcal{X})$ tal que $\omega = \phi\omega'$ o que implica $(\omega') = W' = (\phi\omega) = (\phi) + (\omega) \equiv W$. Desta maneira, os divisores canônicos formam uma única classe de equivalência. Chamaremos esta classe de W.

Como as diferenciais formam um divisor, podemos pensar no espaço $\mathcal{L}(W)$. Este espaço vetorial de funções racionais pode ser mapeado em um espaço de diferenciais pela aplicação $\phi \to \phi\omega$. Pela definição de $\mathcal{L}(W)$, a imagem de ϕ por esta aplicação resulta em uma diferencial regular. Ou seja, $\mathcal{L}(W)$ é isomorfo a $\Omega[\mathcal{X}]$.

$\mathcal{L}(W)$ guarda muitas informações sobre a curva. Por esta razão destacamos

Definição 4.28. *O gênero de um curva $\mathcal{X}$ é dado por $g = l(W)$.*

Assim o gênero de uma curva regular é dado pelo número de diferenciais regulares que sejam linearmente independentes.

Esta definição de gênero é coerente com a definição de gênero utilizada no estudo topológico das superfícies de Riemann sobre o corpo dos complexos. Na topologia, o gênero é a contagem de alças presentes na superfície. A presença destas alças corresponde, de fato, à dimensão do espaço das diferenciais, ou 1-formas holomorfas. Maiores detalhes em [8].

Para curvas planas, dispomos de uma fórmula para o cálculo do gênero que depende somente do grau da curva.

Teorema 4.29. *Se $\mathcal{X}$ é uma curva projetiva não singular de grau m em $\mathbb{P}^2$, então*

$$g = \frac{1}{2}(m-1)(m-2)$$

Esta é uma das várias fórmulas atribuídas a Plücker.

4.3.1 Espaços vetoriais sobre diferenciais

Para a definição de código $C(D, G)$, utilizamos o espaço vetorial $\mathcal{L}(G)$ e, para cada função, associamos em cada ponto racional escolhido um elemento do corpo. As diferenciais, por sua vez, também oferecem a possibilidade de construções semelhantes para definirmos outra classe de códigos. Precisamos, contudo, adaptar algumas definições que foram aplicadas a divisores para o contexto das diferenciais.

Definição 4.30. *Seja D um divisor sobre uma curva $\mathcal{X}$. Definimos*

$$\Omega(D) = \{\omega \in \Omega(\mathcal{X}) \mid (\omega) - D \succeq 0\} \cup \{0\}$$

A dimensão de $\Omega(D)$ sobre $\mathbb{F}$ será denotada por $\delta(D)$ e é chamada de índice de especialidade de D.

A conexão com as funções é dada por

Lema 4.31. *$\delta(D) = l(W - D)$.*

Prova:

- Se $W = (\omega)$ é trivial, então a definição de $\mathcal{L}(W - D)$ corresponde à definição de $\Omega(D)$.
- Se $W = (\omega)$ é não trivial, construímos a aplicação linear $\phi : \mathcal{L}(W - D) \to \Omega(D)$ definida por $\phi(f) = f\omega$.
  - $\phi(f) = 0 \Rightarrow f = 0$. Logo ϕ é injetora.
  - Seja $\gamma \in \Omega(D)$. Como $\gamma = f\omega$, para alguma função racional f, a seqüência abaixo mostra que γ é a imagem de ϕ para a função f.

$$\begin{aligned} (\gamma) - D &\succeq 0 \\ (f\omega) - D &\succeq 0 \\ (f) + (\omega) - D &\succeq 0 \\ (f) + (W - D) &\succeq 0 \\ f &\in \mathcal{L}(W - D). \diamond \end{aligned}$$

4.4 Códigos de Goppa

Na definição 4.18 utilizamos a aplicação linear α, que calcula o valor da função em cada ponto. Em outras palavras, para cada coordenada do vetor $C(D, G)$, a aplicação α realiza uma transformação linear que retorna o valor da função no ponto correspondente à coordenada.

No contexto das diferenciais, não faz sentido perguntar pelo valor da diferencial em um ponto. Contudo, podemos aplicar outra transformação linear.

Definição 4.32. *Seja $P \in \mathcal{X}$, t um parâmetro local em P e $\omega = \phi dt$ a representação da diferencial ω em P. A função ϕ pode ser escrita como uma série de Laurent $\sum_i a_i t^i$. Definimos o resíduo $Res_P(\omega)$ de ω no ponto P como sendo a_{-1}.*

No que se segue, tomaremos os divisores $D = P_1 + P_2 + ... + P_n$ e G como na definição 4.18 com a restrição $2g - 2 < grau(G) < n$. Definimos agora a classe de códigos de Goppa.

Definição 4.33. *O código linear $C^*(D, G)$ de tamanho n sobre $\mathbb{F}_q$ é a imagem da aplicação linear $\alpha^* : \Omega(G - D) \to {\mathbb{F}_q}^n$ definida por*

$$\alpha^*(\eta) = (Res_{P_1}(\eta), Res_{P_2}(\eta), ..., Res_{P_n}(\eta)).$$

Assim como para os códigos de Reed-Solomon, para conhecermos os parâmetros dos códigos de Goppa, dependemos do teorema de Riemann-Roch.

4.5 Teorema de Riemann-Roch

O teorema de Riemann-Roch é o resultado que estabelece a ligação entre as propriedades algébricas e topológicas de uma curva. Este teorema tem profunda influência em diversas área da matemática: de Geometria Algébrica à Teoria dos Números. Veremos adiante como utilizá-lo na Teoria de Códigos.

A primeira apresentação do teorema segue a abordagem proposta por Hirschfeld[6], seção 2.14.

Seja F um polinômio e $\mathcal{X} = \mathcal{V}(F)$ uma curva sobre $\mathbb{F}_q$. A cada função $f \in \mathbb{F}_q(\mathcal{X})$ definida sobre a curva está associado um divisor $div(f)$.

Teorema 4.34. *(Riemann-Roch) Para qualquer divisor D definido sobre $\mathcal{X}$*

1. *$\mathcal{L}(D)$ é um espaço vetorial de dimensão $l(D)$ sobre $\mathbb{F}_q$;*
2. *Existe um inteiro não negativo g tal que*

$$l(D) \geq grau(D) + 1 - g$$

   para todos os divisores $D \in Div(\mathcal{X})$, e o menor inteiro que satisfaz esta condição é o gênero de $\mathcal{X}$.

3. *Se $grau(D) > 2g - 2$, então $l(D) = grau(D) + 1 - g$*

A abordagem de Høholdt[10] apresenta o teorema de Riemann-Roch com a restrição a curvas projetivas regulares de gênero g.

Teorema 4.35. *(Riemann-Roch) Seja D um divisor definido sobre uma curva projetiva regular de gênero g. Então, para qualquer divisor canônico W:*

$$l(D) - l(W - D) = grau(D) - g + 1.$$

Corolário 4.36. *Para um divisor canônico W, $grau(W) = 2g - 2$.*

Prova:
Basta substituir D por W no teorema e lembrar que $\mathcal{L}(0) = \mathbb{F}$ e, portanto, $l(0) = 1$. $\diamond$

O teorema de Riemann–Roch nos permite refinar o teorema 4.10.

Corolário 4.37. *Seja D um divisor sobre uma curva projetiva regular de gênero g e $grau(D) > 2g - 2$. Então*

$$l(D) = grau(D) - g + 1.$$

Prova:
Pelo corolário anterior $grau(W - D) < 0$, logo $l(W - D) = 0$. $\diamond$

Este corolário nos permite, dado um divisor D, construir bases para $\mathcal{L}(D)$ através do estudo das lacunas de Weierstrass.

4.5.1 Lacunas de Weierstrass

Dado um ponto $P \in \mathcal{X}$ e $m \in \mathbb{N}_0$, se existe uma função em $\mathbb{F}(\mathcal{X})$ com pólo de multiplicidade m em P, esta função é linearmente independente de todas as funções de $\mathcal{L}((m-1)P)$. Caso exista tal função podemos afirmar que $l(mP) = l((m-1)P) + 1$. Caso não exista tal função, concluimos que $l(mP) = l((m-1)P)$.

Definição 4.38. *Nesta condições, se $l(mP) = l((m-1)P)$, então m será uma lacuna de Weierstrass no ponto P.*

Lema 4.39. *O número de lacunas de um ponto P é igual ao gênero da curva.*

Prova:
Seja $m \in \mathbb{N}$. Se $m > 2g - 2$, então, pelo corolário anterior, $l(mP) = m - g + 1$. Ou seja, para inteiro maiores que $2g - 2$ a dimensão sempre cresce com m. Isto significa que não existe lacuna para inteiros maiores que $2g - 2$. A cadeia a seguir nos permite contar as lacunas

$$1 = l(0) \leq l(P) \leq l(2P) \leq ... \leq l((2g-1)P) = g.$$

Como temos $2g$ inteiros e só chegamos até dimensão g, temos g lacunas. $\diamond$

Dado um ponto P, se escolhermos $m \in \mathbb{N}_0$, tal que m não é lacuna de P, então existe uma função em $\mathbb{F}(\mathcal{X})$ que possui pólo de ordem m em P e nenhum outro pólo. Se realizarmos a multiplicação desta função com ela mesma, poderemos concluir que qualquer múltiplo positivo de m não será lacuna no ponto P.

Podemos concluir que se dois inteiros não negativos, m_1 e m_2, não são lacunas de P, então $m_1 + m_2$ também não é lacuna de P. Desta forma, os inteiros que não são lacunas formam o semigrupo de Weierstrass em $\mathbb{N}_0$.

Seja $(\rho_i | i \in \mathbb{N}_0)$ uma enumeração em ordem crescente dos inteiros que não são lacunas de P. Seja $f_i \in \mathcal{L}(\rho_i P)$ tal que $v_P(f_i) = -\rho_i$ para $i \in \mathbb{N}_0$. Então $\{f_j | 1 \leq j \leq i\}$ é uma base de $\mathcal{L}(\rho_i P)$.

4.5.2 Parâmetros dos códigos de Reed-Solomon

Estamos agora em condições de conhecer os parâmetros do código de Reed-Solomon. Para isto, daqui em diante, exigimos que D e G sejam divisores tais que

$$2g - 2 < grau(G) < n.$$

Teorema 4.40. *O código linear $C(D, G)$ de tamanho n sobre o corpo $\mathbb{F}_q$ tem dimensão $k = grau(G) - g + 1$ e distância mínima $d \geq n - grau(G)$.*

Prova:

- Se uma função f pertence ao núcleo da aplicação α da definição 4.18, então f possui zeros em $P_1, ..., P_n$ e com isto $f \in \mathcal{L}(G - D)$. Como $grau(G) < grau(D)$, o teorema 4.10 (i) nos garante que $f = 0$. Logo α é injetora e a dimensão de $C(D, G)$ é dada pela dimensão de $\mathcal{L}(G)$. Como $grau(G) > 2g - 2$, o corolário 4.32 nos garante que

$$k = l(G) = grau(G) - g + 1.$$

- Se o vetor $\alpha(f)$ tem peso $d > 0$, então existem $n - d$ pontos $P_{i_1}, ..., P_{i_{n-d}}$no suporte de D, tais que $f(P_{i_j}) = 0$, $1 \leq j \leq n - d$. Tomemos o divisor $E = P_{i_1} + ... + P_{i_{n-d}}$. Então $f \in \mathcal{L}(G - E)$. O teorema 4.10 exige que $grau(G) - grau(E) = grau(G) - (n - d) \geq 0$. Assim

$$d \geq n - grau(G).\diamond$$

4.5.3 Parâmetros dos códigos de Goppa

Sobre os parâmetros dos códigos de Goppa, podemos afirmar

Teorema 4.41. *O código $C^*(D,G)$ tem dimensão $k = n - grau(G) + g - 1$ e distância mínima $d \geq grau(G) - 2g + 2$*

Prova:

- Seja γ pertencente ao núcleo da aplicação α^*. Sabemos que $(\gamma) - G + D \succeq 0$, mas como ela não possui pólos nos pontos $P_1, ..., P_n$, temos que $(\gamma) - G \succeq 0$. Logo, $\gamma \in \Omega(G)$. Mas $\delta(G) = l(W - G)$ e $grau(G) > 2g - 2$. Desta forma $\gamma = 0$. Logo $k = \delta(G - D) = l(W - G + D)$. Pelo teorema de Riemann-Roch

$$\begin{aligned} l(G-D) - l(W-G+D) &= grau(G-D) - g + 1 \\ -l(W-G+D) &= grau(G) - n - g + 1 \\ l(W-G+D) &= n - grau(G) + g - 1. \end{aligned}$$

- Se $\alpha^*(\gamma)$ tem peso $d > 0$, então existem $n - d$ pontos $P_{i_1}, ..., P_{i_{n-d}}$ no suporte de D, tais que $Res_{P_{i_j}}(\gamma) = 0$, $1 \leq j \leq n - d$. Tomemos o divisor $E = P_{i_1} + ... + P_{i_{n-d}}$. Então $\gamma \in \Omega(G - D + E)$. O teorema 4.10 exige que $\delta(G - D + E) = l(W - G + D - E) \geq 0$. Logo,

$$\begin{aligned} grau(W - G + D - E) &\geq 0 \\ grau(W) - grau(G) + n - (n - d) &\geq 0 \\ 2g - 2 - grau(G) + d &\geq 0 \\ d &\geq grau(G) - 2g + 2. \diamond \end{aligned}$$

4.5.4 Dualidade entre Reed-Solomon e Goppa

As definições dos códigos de Reed–Solomon e de Goppa são semelhantes, mas utilizam funções e diferenciais respectivamente. Poderemos observar, a seguir, uma dualidade subjacente entre estas classes de código.

Para alcançar este resultado, dependemos do teorema de resíduos:

Teorema 4.42. *Se ω é uma diferencial sobre uma curva projetiva regular $\mathcal{X}$, então*

$$\sum_{P \in \mathcal{X}} Res_P(\omega) = 0$$

Tate[9] provou este resultado para qualquer corpo k através da correspondência entre pontos racionais e valorações discretas de $\mathcal{O}_P(\mathcal{X})$. Ele tomou o completamento do anel local e dos corpos de frações ([5] pág. 33-35) para definir séries de Laurent para funções e, com elas, demonstrar que sua definição abstrata de resíduo corresponde à escolha do coeficiente a_{-1}.

Sobre o corpo dos complexos a prova deste teorema pode utilizar o teorema de Stokes ou Cauchy.

Teorema 4.43. *Os códigos $C(D,G)$ e $C^*(D,G)$ são códigos lineares duais entre si.*

Prova:
Pelos dois teoremas anteriores, sabemos que $k_{C(D,G)} + k_{C^*(D,G)} = n$. Só nos resta mostrar que as palavras de códigos são ortogonais. Sejam $\phi \in \mathcal{L}(G)$ e $\eta \in \Omega(G - D)$. O produto escalar dos vetores é dado por

$$\alpha(\phi).\alpha^*(\eta) = \sum_{i=1}^{n} \phi(P_i) Res_{P_i}(\eta) = \sum_{i=1}^{n} Res_{P_i}(\phi\eta).$$

A diferencial $\phi\eta$ só admite pólos de ordem 1 nos pontos do suporte do divisor D, de acordo com a seqüência

$$\begin{aligned} (\eta) - G + D &\succeq 0 \\ (\phi) + (\eta) - G + D &\succeq (\phi) \succeq -G \\ (\phi\eta) - G + D &\succeq -G \\ (\phi\eta) + D &\succeq 0 \end{aligned}$$

Logo a soma $\sum_{i=1}^{n} Res_{P_i}(\phi\eta)$ ocorre sobre todos os pólos da diferencial e, assim, pelo teorema do resíduo,

$$\alpha(\phi).\alpha^*(\eta) = 0. \diamond$$

Concluímos as comparações entre as duas classes de códigos com um teorema que diz que esta divisão entre os códigos de Reed–Solomon e de Goppa é somente histórica, pois[10]:

Teorema 4.44. *Seja $\mathcal{X}$ um curva definida sobre $\mathbb{F}_q$. Sejam $P_1, P_2, ..., P_n$ n pontos racionais sobre $\mathcal{X}$ e $D = P_1 + P_2 + ... + P_n$. Então existe uma diferencial ω com pólos de de ordem 1 em cada ponto P_i tal que $Res_{P_i} = 1$ para todo i. Além disto*

$$C^*(D, G) = C(D, W + D - G)$$

para todos os divisores G que tenham suporte disjunto do suporte de D. W é o divisor de ω.

4.6 Um código de Reed-Solomon em detalhe

Seja $\mathcal{X}$ a curva plana definida pela equação $X^3 + Y^3 + Z^3 = 0$ sobre $\mathbb{F}_4 = \{0, 1, \alpha, \bar{\alpha} = 1 + \alpha = \alpha^2\}$. A curva $\mathcal{X}$ possui grau 3 e contém 9 pontos racionais:

	Q	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
x	0	0	0	1	α	$\bar{\alpha}$	1	α	$\bar{\alpha}$
y	1	α	$\bar{\alpha}$	0	0	0	1	1	1
z	1	1	1	1	1	1	0	0	0

Através da fórmula de Plücker temos que o gênero desta curva é dado por

$$g = \frac{1}{2}(3-1)(3-2) = 1$$

Pelo Corolário 4.31 o grau do divisor canônico é dado por

$$grau(W) = 2g - 2 = 0$$

logo o divisor canônico é igual a $W = 0$.

Tomemos os divisores $G = 3Q$ e $D = \sum P_i$, $1 \leq i \leq 8$. Desejamos, então, estudar o espaço de funções $\mathcal{L}(3Q)$ e encontrar uma base. Como o suporte só contém o ponto Q, funções neste espaço são linearmente independentes se possuirem ordem de pólo diferentes em Q.

Propomos as funções 1, $\frac{x}{y+z}$ e $\frac{y}{y+z}$.

Devemos verificar se cada uma delas pertence a $\mathcal{L}(G)$. Para isto devemos determinar os divisores gerados por elas. Para determinar zeros e pólos, usaremos a intersecção da curva com retas. Pelo teorema de Bézout estas intersecções devem conter 3 pontos contando-se as multiplicidades.

- $(1) = 0$;
- Seja a reta $\mathcal{R}$ com equação $X = 0$. O divisor intersecção $\mathcal{X}.\mathcal{R} = P_1 + P_2 + Q$. Seja a reta $\mathcal{S}$ com equação $Y + Z = 0$. O divisor intersecção $\mathcal{X}.\mathcal{S} = 3Q$. Logo $(\frac{x}{y+z}) = P_1 + P_2 - 2Q$.
- Seja a reta $\mathcal{T}$ com equação $Y = 0$. O divisor intersecção $\mathcal{X}.\mathcal{T} = P_3 + P_4 + P_5$. Logo $(\frac{x}{y+z}) = P_3 + P_4 + P_5 - 3Q$.

A verificação é feita diretamente a partir da definição de $\mathcal{L}(G)$:

$$(1) + G = 3Q \succeq 0$$
$$(\frac{x}{y+z}) + G = P_1 + P_2 - 2Q + 3Q = P_1 + P_2 + Q \succeq 0$$
$$(\frac{y}{y+z}) + G = P_3 + P_4 + P_5 - 3Q + 3Q = P_3 + P_4 + P_5 \succeq 0$$

Para determinar a ordem de pólo escolher um parâmetro local pode ser muito útil. Tomemos em Q o parâmetro local $t = \frac{x}{z}$.

- $1 = t^0$;
- $\frac{x}{y+z}$ não faz sentido em Q. Podemos reescrever esta função, lembrando que estamos em característica 2:

$$\frac{x}{y+z} = \frac{x(y^2+yz+z^2)}{(y+z)(y^2+yz+z^2)} =$$
$$\frac{x(y^2+yz+z^2)}{(y^3+y^2z+yz^2+y^2z+yz^2+z^3)} = \frac{x(y^2+yz+z^2)}{(y^3+z^3)} =$$
$$\frac{x(y^2+yz+z^2)}{x^3} = \frac{(y^2+yz+z^2)}{x^2} =$$
$$\frac{(y^2+yz+z^2)z^2}{x^2z^2} = t^{-2}\frac{(y^2+yz+z^2)}{z^2}$$

  E assim $\frac{x}{y+z}$ tem pólo de ordem 2 em Q.

- Analogamente:

$$\frac{y}{y+z} = \frac{y(y^2+yz+z^2)}{(y+z)(y^2+yz+z^2)} =$$
$$\frac{y(y^2+yz+z^2)}{(y^3+y^2z+yz^2+y^2z+yz^2+z^3)} = \frac{y(y^2+yz+z^2)}{(y^3+z^3)} =$$
$$\frac{y(y^2+yz+z^2)}{x^3} = \frac{y(y^2+yz+z^2)z^3}{x^3z^3} =$$
$$t^{-3}\frac{y(y^2+yz+z^2)}{z^3}$$

  E assim $\frac{y}{y+z}$ tem pólo de ordem 3 em Q.

Encontramos 3 funções linearmente independentes. Para sabermos se poderemos encontrar mais uma recorremos ao teorema de Riemann-Roch:

$$l(3Q) - l(0-3Q) = l(3Q) = grau(3Q) - 1 + 1 = 3$$

Concluímos, então, que as funções 1, $\frac{x}{y+z}$ e $\frac{y}{y+z}$ formam uma base de $\mathcal{L}(3Q)$. Podemos construir a matriz de geradores aplicando as funções em cada um dos pontos do suporte do divisor D. Apresentamos o cálculo somente para as situações menos triviais.

$$\frac{y}{y+z}(P_1) = \frac{\alpha}{\alpha+1} = \frac{\alpha}{\bar{\alpha}} = \frac{\alpha\alpha}{\bar{\alpha}\alpha} = \frac{\bar{\alpha}}{(\alpha+1)\alpha} = \frac{\bar{\alpha}}{\alpha^2+\alpha} = \frac{\bar{\alpha}}{\alpha+1+\alpha} = \bar{\alpha}$$

$$\frac{y}{y+z}(P_2) = \frac{\bar{\alpha}}{\bar{\alpha}+1} = \frac{\bar{\alpha}}{\alpha} = \frac{\bar{\alpha}\bar{\alpha}}{\alpha\bar{\alpha}} = \frac{(1+\alpha)(1+\alpha)}{1} = 1+\alpha+\alpha+\alpha^2 = 1+\bar{\alpha} = \alpha$$

A matriz de geradores de $C(D, G)$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & \alpha & \bar{\alpha} & 1 & \alpha & \bar{\alpha} \\ \bar{\alpha} & \alpha & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Estamos diante de um espaço de dimensão igual a 3. Além disto, podemos observar que a partir desta matriz de geradores que a distância mínima do código é 5, pois o menor peso é igual a 5. Estes fatos estão de acordo com o teorema 4.35:

$$\begin{aligned} k &= grau(G) - g + 1 = 3 \\ d &\geq n - grau(G) = 5. \end{aligned}$$

4.7 Bons códigos

Sabemos agora, por meio do teorema de Riemann–Roch, identificar algumas características das variedades que melhoram os parâmetros de códigos algébricos. Se desejamos aumentar a dimensão dos códigos, precisamos encontrar variedades com muitos pontos racionais. Se temos interesse em aumentarmos a distância mínima, procuraremos variedades de gênero próximo de zero. O teorema 4.3 nos mostra que, se $g = 0$, possuimos um código MDS, pois $d_{C(D,G)} \geq n - k_{C(D,G)} + 1 - g$.

O número de pontos racionais de uma variedade é um problema importante para Geometria Algébrica. Ele estabelece limites para a dimensão dos códigos. Por esta razão, esta é uma das áreas de maior interesse atualmente para a Teoria de Códigos.

5 Apêndice: Geometria Projetiva

Desejamos fazer a apresentação resumida dos pontos que consideramos mais relevantes na abordagem dos espaços projetivos proposta por Beutelspacher e Rosenbaum [4].

A geometria projetiva, assim como a geometria euclidiana, é construída a partir de um sistema axiomático. Não é de se estranhar que existam conexões entre o estudo de lógica e de geometria.

Neste contexto, dizemos que uma geometria é uma tripla $G = (\mathcal{P}, \mathcal{L}, \mathcal{I})$, em que chamamos $\mathcal{P}$ de conjunto de pontos, $\mathcal{L}$ o conjunto de retas e $\mathcal{I}$ uma relação de incidência entre retas e pontos.

Os axiomas do espaço projetivo:

1. Para quaisquer dois pontos distintos existe exatamente uma reta que é incidente nestes dois pontos;
2. (Veblen-Young) Sejam A, B, C e D pontos tais que AB possui intersecção com CD, então AC também possui intersecção com BD;
3. Qualquer reta é incidente em 3 pontos no mínimo;
4. Existem 2 retas no mínimo.

Definição 5.1. *Um subconjunto $\boldsymbol{U} \subset \boldsymbol{P}$ é chamado linear se, dados quaisquer dois pontos $P, Q \in \boldsymbol{U}$, o conjunto (PQ) de todos os pontos incidentes em PQ está contido em $\boldsymbol{U}$.*

Entenderemos os planos como sendo o menor conjunto linear que contenha 3 pontos não colineares. O que representamos por $\langle A, B, C\rangle$ ou $\langle\{A, B, C\}\rangle$.

Lema 5.2. *Seja $\boldsymbol{P} = \langle\{A, B, C\}\rangle$ um plano projetivo. Então qualquer reta g diferente de AB possui intersecção com AB. O mesmo vale para AC e BC.*

Prova:
P é formado pela reta AB e pelas retas CX, em que $X \in (AB)$. Tomemos uma reta g diferente de AB em **P**. Podemos tomar dois pontos $E, F \in (g)$ pelo axioma 3. Se E ou F incidem em AB, concluimos. Caso contrário, Sabemos que E é incidente em uma reta CX, em que $X \in (AB)$. O mesmo vale para F, para alguma reta CY. Concluimos o lema aplicando o axioma 2 aos pontos X, E, Y e F; pois as retas XE e YF se intersectam em C. O mesmo argumento é aplicável para as demais retas. $\diamond$

Lema 5.3. *Quaisquer duas retas retas em um plano projetivo se intersectam.*

Prova:
Seja $\mathbf{P} = \langle\{A, B, C\}\rangle$ um plano projetivo. Sejam g e h duas retas em $\mathbf{P}$. Se uma dela for BC, AC ou AB, concluimos pelo lema anterior. Caso contrário, pelo lema anterior, podemos tomar os pontos das intersecções $E = AB.g$, $F = AB.h$, $G = AC.g$ e $H = AC.h$. Concluimos aplicando o axioma 2 aos pontos E, F, G e H, pois as retas EF e GH se intersectam em A. $\diamond$

Todos os espaços projetivos nesta apresentação são finitamente gerados, o que equivale a dizer que existe um conjunto linear que pode ser construído a partir de um número finito de pontos.

Se estes pontos forem linearmente independentes, eles são uma base do espaço. Para qualquer espaço linear $\mathbf{P}$, se a cardinalidade de qualquer uma de suas bases é $d + 1$, diremos que o espaço $\mathbf{P}$ tem dimensão d.

Teorema 5.4. *(Fórmula da dimensão) Sejam $\boldsymbol{U}$ e $\boldsymbol{W}$ subespaços lineares do espaço projetivo $\boldsymbol{P}$. Então*

$$\dim(\langle \boldsymbol{U}, \boldsymbol{W}\rangle) = \dim(\boldsymbol{U}) + \dim(\boldsymbol{W}) - \dim(\boldsymbol{U} \cap \boldsymbol{W})$$

Definição 5.5. *Seja Q um ponto no espaço projetivo $\boldsymbol{P}$. Definimos a geometria quociente $\boldsymbol{P}/Q$ cujos pontos são as retas de $\boldsymbol{P}$ incidentes em Q, cujas retas são os planos de $\boldsymbol{P}$ e cuja relação de incidência seja a incidência induzida por $\boldsymbol{P}$.*

Teorema 5.6. *Seja um espaço projetivo $\boldsymbol{P}$ de dimensão d e ordem q e seja um ponto $Q \in \boldsymbol{P}$. Então a geometria quociente $\boldsymbol{P}/Q$ é um espaço projetivo de dimensão $d - 1$ e ordem q.*

Lema 5.7. *Sejam duas retas g_1 e g_2 no espaço projetivo $\boldsymbol{P}$. Então existe uma bijeção entre o conjunto de pontos (g_1) no conjunto (g_2).*

Quando o espaço $\mathbf{P}$ for finito, sabemos pelo lema anterior que cada reta possui o mesmo número $q + 1$ de pontos. Ao número q damos o nome de ordem do espaço $\mathbf{P}$.

Definição 5.8. *Um espaço projetivo é dito finito se seu conjunto de pontos for finito. Representamos um espaço projetivo finito por $PG(d, q)$ em que d é a dimensão e q é a ordem do espaço projetivo.*

Para contarmos pontos em $\mathbf{P}$, dispomos de

Teorema 5.9. *Seja* $\boldsymbol{P}$ *um espaço projetivo finito de dimensão* d *e ordem* q *e seja* $\boldsymbol{U}$ *um subespaço linear t-dimensional de* $\boldsymbol{P}$*. Então*

- *O número de pontos de* $\boldsymbol{U}$ *é dado por*

$$q^t + q^{t-1} + ... + q + 1 = \frac{q^{t+1} - 1}{q - 1};$$

- *O número de retas de* $\boldsymbol{U}$ *por um ponto de* $\boldsymbol{U}$ *é dado por*

$$q^{t-1} + ... + q + 1;$$

- *O número de retas de* $\boldsymbol{U}$ *é dado por*

$$\frac{(q^t + q^{t-1} + ... + q + 1)(q^{t-1} + ... + q + 1)}{q + 1}.$$

Prova:
A prova é feita por indução na dimensão do espaço. Para o caso $d = 1$, o teorema vale. Através de uma geometria quociente, conseguimos fazer valer a hipótese de indução em uma subestrutura do espaço $\mathbf{P}$. A partir deste resultado, contamos as retas e pontos em $\mathbf{P}$.$\diamond$

A questão de quais inteiros podem ser ordem de algum plano projetivo é uma das mais discutidas em geometria finita. Uma das conjectura mais conhecidas diz que a ordem deve ser potência de algum número primo. Na verdade, todos os planos projetivos finitos conhecidos tem ordem potência de primos conforme o quadro abaixo

ordem	2	3	4	5	6	7	8	9	10	11	12
existência	sim	sim	sim	sim	não	sim	sim	sim	não	sim	?

O plano projetivo carrega em si uma noção de minimalidade que pode ser observado no teorema de Ërdos e de Bruijn:

Teorema 5.10. *Seja* $\boldsymbol{L}$ *um espaço linear finito com* v *pontos e* b *retas. Então temos* $b \geq v$ *com igualdade se e somente se* $\boldsymbol{L}$ *for um plano projetivo ou uma estrutura em que todos os pontos menos* 1 *são colineares (near-pencil), como na figura 4.*

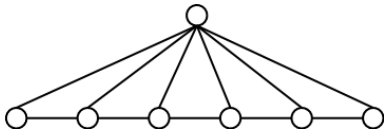

Figura 4: Near-pencil

Os problemas sobre configurações de retas e pontos são muito importantes para revelar características e propriedades de espaços projetivos. Os mais conhecidos são os teoremas de Desargues e de Pappus.

Teorema 5.11. *(Desargues) Dados quaisquer dois triângulos em perspectiva, as intersecções dos lados correspondentes são colineares.*

Teorema 5.12. *(Pappus) Sejam* $\boldsymbol{g}$ *e* $\boldsymbol{h}$ *quaisquer duas retas coplanares. Tomemos* $A, B, C \in \boldsymbol{g}$ *e* $A', B', C' \in \boldsymbol{h}$*, todos distintos entre si e diferentes de* $\boldsymbol{g}.\boldsymbol{h}$*. Então os pontos*

$$D = BC'.B'C, E = CA'.C'A, F = AB'.A'B$$

são colineares.

Sobre estes teoremas podemos afirmar que se um espaço projetivo $\mathbf{P}$ é pappusiano, então ele é desarguesiano e se $\mathbf{P}$ tem dimensão maior que 2, então $\mathbf{P}$ é desarguesiano.

Até o momento só possuimos a estrutura linear. Desejamos associar esta estrutura a uma estrutura vetorial para que possamos dispor de ferramentas mais poderosas da Álgebra Linear. Para fazermos esta associação, o teorema de Desargues será imprescindível. Primeiro definimos o que é um espaço projetivo construído a partir de um espaço vetorial.

Definição 5.13. *Seja V um espaço vetorial de dimensão $d+1 \geq 3$ sobre um anel de divisão F. Definimos a geometria $\boldsymbol{P}(V)$:*

- *Os pontos de $\boldsymbol{P}(V)$ são os subespaços de dimensão 1 de V;*
- *As retas de $\boldsymbol{P}(V)$ são os subespaços de dimensão 2 de V;*
- *A incidência é definida pela inclusão de conjuntos.*

Teorema 5.14. *Seja um espaço vetorial V de dimensão $d+1$ sobre o anel de divisão F. Então $\boldsymbol{P}(V)$ é desarguesiano.*

Teorema 5.15. *Seja um espaço vetorial V sobre o anel de divisão F. Então $\boldsymbol{P}(V)$ é pappusiano se e somente se F é comutativo, e, portanto, um corpo.*

Nos perguntamos se existiria uma contrapartida, de tal forma que Desargues implicasse Pappus. No caso de anéis de divisão finitos isto vale, pois sabemos que corpos finitos são comutativos através do teorema de Wedderburn.

Contudo, uma prova que só utilizasse a finitude do anel de divisão e o teorema de Desargues seria uma alternativa geométrica para provar que corpos finitos são comutativos[1].

Desejamos saber quando um determinado espaço projetivo $\mathbf{P}$ possui uma estrutura vetorial tal que $\mathbf{P} = \mathbf{P}(V)$ para algum espaço vetorial V.

Teorema 5.16. *Dado um espaço projetivo $\boldsymbol{P}$ de dimensão maior ou igual a 2. Se $\boldsymbol{P}$ é desarguesiano, então existe uma espaço vetorial V sobre um domínio de integridade F tal que $\boldsymbol{P}$ é isomorfo a $\boldsymbol{P}(V)$.*

Este resultado depende do estudo das colineações centrais. Através delas conseguimos construir um domínio de integridade e estabelecer as relações geométricas que precisamos.

Definição 5.17. *Uma colineação em um espaço projetivo $\boldsymbol{P}$ é uma união de uma bijeção no conjunto de pontos de $\boldsymbol{P}$ e de uma bijeção no conjunto de retas de $\boldsymbol{P}$ construída de tal forma que a incidência é preservada. Uma colineação é dita central quando ela deixa invariantes todas as retas por um ponto, o centro da colineação, e todos os pontos em um hiperplano, o eixo da colineação.*

[1]Agradecemos ao professor Marcos Sebastiani a apresentação do problema em aberto de uma prova geométrica para o teorema de Wedderburn.

Sobre as colineações centrais podemos afirmar:

Teorema 5.18. *Seja $\boldsymbol{P}$ um espaço projetivo desarguesiano. Para qualquer hiperplano $\boldsymbol{H}$ e quaisquer pontos C, P, P' colineares, tais que P e P' não pertencem a $\boldsymbol{H}$, existem uma e somente uma colineação com centro C e eixo $\boldsymbol{H}$ que leva P em P'.*

Este resultado significa a existência de todas as colineações possíveis. Fica evidente, desta maneira, a simetria inerente aos espaços projetivos. Para maiores detalhes sobre colineações, veja [4] capítulo 3.

Depois do estudo dos subespaços lineares, nos interessam os conjuntos quadráticos, principalmente para o problema de encontrarmos códigos MDS.

Para abordá-los de maneira sintética, precisamos de algumas definições. Tomaremos $\mathcal{Q}$ como um conjunto de pontos de um espaço projetivo $\mathbf{P}$.

Definição 5.19. *Um reta g será chamada tangente de $\mathcal{Q}$ se $(g).\mathcal{Q}$ contém um ou todos os pontos de (g).*

Esta definição parece artificial, mas é apropriada somente para o estudo de conjuntos quadráticos. Se fossemos, por exemplo, trabalhar com quárticas, esta definição não seria adequada para bitangentes.

Definição 5.20. *Para cada ponto $P \in \mathcal{Q}$, seja $\mathcal{Q}_P$ o conjunto formado pelos pontos de todas as retas tangentes a $\mathcal{Q}$ em P. Diremos que $\mathcal{Q}_P$ é o espaço tangente de $\mathcal{Q}$ em P.*

Definição 5.21. *Diremos que um conjunto $\mathcal{Q}$ é um conjunto quadrático de $\boldsymbol{P}$ se atende às condições:*

1. *Se uma reta g tem 3 pontos em comum com $\mathcal{Q}$, então $(g) \subset \mathcal{Q}$;*
2. *Para cada ponto $P \in \mathcal{Q}$, $\mathcal{Q}_P$ é um hiperplano ou $\boldsymbol{P}$.*

Os conjuntos quadráticos são os objetos geométricos que representam as quádricas. Eles oferecem relações geométricas muito ricas e elegantes além de oferecerem conjuntos $LI(n, 3)$ no plano. Para conjuntos $LI(a, n)$, as curvas normais racionais são de grande valia, pois seus pontos se encontram em posição geral.

Definição 5.22. *Seja $\boldsymbol{P}$ um espaço projetivo de dimensão d. Dizemos que um conjunto $\mathcal{S}$ de, no mínimo, $d+1$ pontos está em posição geral se qualquer subconjunto com $d+1$ pontos de $\mathcal{S}$ formam uma base de $\boldsymbol{P}$.*

Definição 5.23. *Seja* $\boldsymbol{P} = \boldsymbol{P}(V)$ *um espaço projetivo de dimensão* d *coordenado pelo corpo* F*. Sejam os pontos de* $\boldsymbol{P}$ *descrito por coordenadas homogêneas. Então o conjunto*

$$\mathcal{N} = \{(1 : t : t^2 : ... : t^d \mid t \in F)\} \cup \{(0 : 0 : ... : 0 : 1)\}$$

é chamado de curva normal racional.

Referências

[1] Códigos corretores de erros. http://www.obm.org.br/semana/codigos.ps.

[2] R. Lidl e H. Niederreiter. *Introduction to finite fields and their applications*. Cambridge University Press, 1994.

[3] Abramo Hefez e Maria Lúcia T. Villela. *Códigos corretores de erros*. Instituto Nacional de Matemática Pura e Aplicada, 2002.

[4] Albrecht Beutelspacher e Ute Rosenbaum. *Projective Geometry*. Cambridge University Press, 1998.

[5] Robin Hartshorne. *Algebraic Geometry*. Springer, 1997.

[6] James Hirschfeld. *Projective Geometries Over Finite Fields*. Oxford University Press, 1998.

[7] Klaus Hulek. *Elementary algebraic geometry*. American Mathematical Society, 2000.

[8] D. Mumford. *Algebraic Geometry I: Complex Projective Varieties*. Springer-Verlag, 1996.

[9] John Tate. Residues of differentials on curves. In *Annales scientifiques de l'École Normale Supérieure*, pages 149–159. 1968. Sér. 4, 1 no. 1.

[10] Tom Høholdt; Jacobus H. van Lint e Ruud Pellikaan. Algebraic geometry codes. In W.C. Huffman e R.A. Brualdi. V.S. Pless, editor, *Handbook of Coding Theory*, pages 871–961. Elsevier, 1998.

[11] Robert J. Walker. *Algebraic curves*. Princeton University Press, 1950.