Códigos Corretores de Erros

Quando armazenamos ou transmitimos dados por um canal de comunicação, ocorrem erros que não podemos prever. Isso ocorre pela própria natureza do canal de comunicação. A *Teoria dos Códigos* se preocupa em detectar e até corrigir esses erros. E onde a Matemática entra nessa teoria?

A resposta é bastante óbvia: a construção de códigos que possam corrigir erros aleatórios, os chamados *códigos corretores de erros*, usa estruturas de natureza algébrica, combinatória e – pasmem! – geométrica.

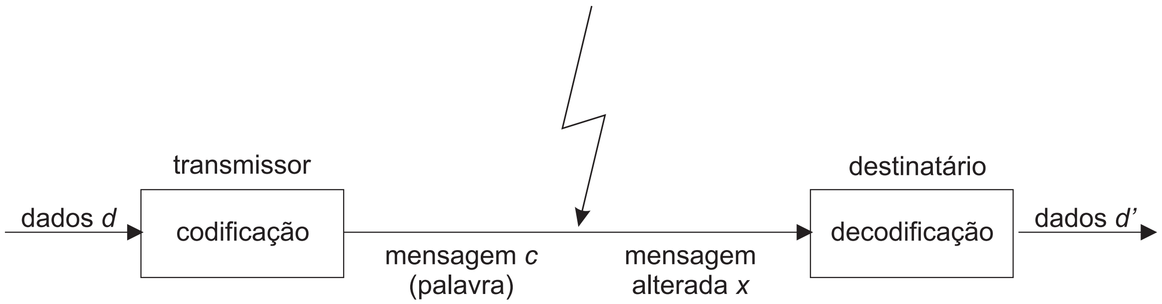
Veremos aqui um pouco de Teoria dos Códigos e alguns códigos baseados em geometrias finitas como a projetiva e a afim.

1. Afinal, o que é Teoria dos Códigos?

Você já deve ter enviado um email ou, pelo menos, assistido à televisão ou ouvido rádio (você ainda não fez isso? você existe?). Pois bem, de vez em quando pode ocorrer de o email aparecer corrompido, ou a televisão aparecer com problemas de recepção (sua TV nunca teve "chuvisco"?), ou a estação de rádio que você está ouvindo ficar com chiado (se não aconteceu nenhuma dessas coisas com você, considere-se uma pessoa de sorte!!). Pois bem, esses pequenos incômodos acontecem porque ocorreram alguns erros durante a transmissão. Esses erros não são intencionais (não, ninguém quis sabotar seu email ou sua TV ou seu rádio!) e geralmente não podem ser controlados.

Todas essas experiências do dia-a-dia podem ser modelados da seguinte forma: um *transmissor* (nos nossos exemplos, seria você, o canal de TV e a estação de rádio, respectivamente) quer enviar dados (o email, o programa de TV e a música do rádio) para um *destinatário* (a pessoa para quem você estava escrevendo email ou você assistindo à TV ou ouvindo rádio). Esses dados são transmitidos via um *canal de comunicação* que, por maiores cuidados que se tomem, pode não transmitir os dados sem alterá-los.

Falando de maneira mais precisa: o transmissor *codifica* os dados d em uma *mensagem* c. A mensagem c é transmitida (e não os dados *per se*). O destinatário *decodifica* a mensagem recebida x (que não é necessariamente igual a c pois erros podem ter ocorrido durante a transmissão) em dados d' e tenta descobrir se ocorreram erros durante a transmissão ou não. Se o código utilizado para transformar d em c e x em d' é um *código corretor de erros* é possível reconstruir os dados originais, mesmo que x seja diferente de c (nesse caso, teríamos $d' = d$).



Os erros que a Teoria dos Códigos estudam são de natureza aleatória, o que excluem erros gerados intencionalmente (esses são estudados pela *Criptologia*). E os erros estudados são exclusivamente alterações de símbolos. Ou seja, não consideraremos aqui inclusão ou exclusão de símbolos.

2. Um pouquinho de Álgebra Linear

Para entender (e apreciar) melhor a modelagem e os teoremas que vamos provar nesse artigo, é importante saber um pouquinho sobre espaços vetoriais. Procure por mais em [5].

2.1. Espaços vetoriais

Definição 2.1. *Um espaço vetorial sobre um corpo F é um conjunto V que satisfaz*

- *Para $u, v, w \in V$, valem as propriedades usuais da adição: associativa ($u + (v + w) = (u + v) + w$), existência de elemento neutro ($u + 0 = 0 + u = u$), existência de oposto ($u + (-u) = (-u) + u = 0$) e comutativa ($u + v = v + u$).*
- *Existe uma operação entre elementos de F e V, que resulta em elementos de V, denominada multiplicação por escalar, que satisfaz suas propriedades usuais (você pode se sentir mais confortável se pensar no exemplo $F = R$ e V sendo o conjunto das matrizes de ordem 2 com entradas reais): para $\alpha, \beta \in F$ e $u, v \in V$, valem as distributivas ($\alpha(u+v) = \alpha u + \alpha v$ e $(\alpha+\beta)v = \alpha v + \beta v$), associativa ($\alpha(\beta v) = (\alpha\beta)v$) e multiplicação por unidade ($1v = v$).*

Os elementos de V e F são chamados *vetores* e *escalares*, respectivamente.

Nas próximas definições, V é um espaço vetorial sobre F.

Definição 2.2. *Um subespaço vetorial de V é um subconjunto de V que é um espaço vetorial.*

Teorema 2.1. *$S \subset V$ é um subespaço vetorial se, e somente se, o vetor nulo 0 pertence a S e, para todos $v, w \in S$ e $\lambda \in F$, $v + \lambda w \in S$.*

Demonstração

Basta provar que as propriedades usuais da adição e multiplicação por escalar, que já valem no espaço vetorial V, são válidas em S também.

Primeiro, tomando $\lambda = 1$ (a unidade do corpo), vemos que a adição é fechada em S. É claro que valem as propriedades associativa e comutativa. O elemento neutro 0 já pertence a S. Dado $w \in S$, tomando $v = 0$ e $\lambda = -1$ vemos que $-w$ pertence a S, ou seja, todo elemento de S admite oposto.

Sobre a multiplicação por escalar, basta notar que se $v \in S$ então $\lambda v \in S$ para todo $\lambda \in F$. Assim, claramente valem a associativa, multiplicação por unidade e a distributiva $(\alpha + \beta)v = \alpha v + \beta v$. E, tomando λv no lugar de v, temos $\lambda v + \lambda w \in S$, de modo que podemos colocar λ em evidência. ∎

Definição 2.3. *Uma combinação linear de vetores $v_1, v_2, \ldots, v_n \in V$ é o vetor $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$, sendo $\alpha_1, \alpha_2, \ldots, \alpha_n$ escalares.*

Definição 2.4. *Sendo $S \subset V$, o subespaço gerado por S é o conjunto das combinações lineares de elementos de S.*

O leitor não deve encontrar dificuldades em provar que o conjunto das combinações lineares de S é um subespaço vetorial de V.

Definição 2.5. *Dizemos que um conjunto $\{u_1, u_2, \ldots, u_n\} \subset V$ é linearmente independente quando, sendo $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$ tais que $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = 0$ implica $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.*

Definição 2.6. *Um conjunto linearmente independente que gera V é uma base de V e o seu número de elementos é a dimensão de V sobre F, que é denotada por $\dim(V)$. Observe que B pode ser infinito; nesse caso, a dimensão também é infinita.*

Feitas as apresentações (ou melhor, as definições!), podemos provar alguns fatos interessantes. Daqui em diante, só trabalharemos com espaço vetoriais finitamente gerados.

Teorema 2.2. *Dada uma base B de V, todo elemento de V pode ser representado unicamente como uma combinação linear dos vetores de B.*

Demonstração

Como B gera V, cada elemento de V pode ser representado como uma combinação linear dos vetores de B. Basta provarmos que tal representação é única.

Suponha que um vetor admita duas representações distintas como combinação linear dos vetores de $B = \{u_1, u_2, \ldots, u_n\}$, ou seja, que $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n$ para alguma escolha de α_i's e β_i's. Então:

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n \iff (\alpha_1 - \beta_1)u_1 + (\alpha_2 - \beta_2)u_2 + \cdots + (\alpha_n - \beta_n)u_n = 0$$

Da definição de conjunto linearmente independente,

$$\alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \cdots = \alpha_n - \beta_n = 0 \iff \alpha_i = \beta_i \text{ para } i = 1, 2, \ldots, n$$

∎

Por que a dimensão de um espaço está bem definida? Um espaço vetorial pode ter bases diferentes, de modo que é possível, a princípio, que duas bases tenham cardinalidades diferentes. Veremos que isso não pode ocorrer.

Teorema 2.3. *Todas as bases de um espaço vetorial finitamente gerado tem o mesmo número de elementos.*

Demonstração

Sejam B e C bases do espaço vetorial V. Suponha, por absurdo, que o número m de elementos de $C = \{c_1, c_2, \ldots, c_m\}$ é maior que o número n de elementos de $B = \{b_1, b_2, \ldots, b_n\}$. Podemos escrever cada elemento de C de forma única como combinação linear de elementos de B, ou seja,

$$\left|\begin{array}{l} c_1 = \alpha_{11} b_1 + \alpha_{12} b_2 + \cdots + \alpha_{1n} b_n \\ c_2 = \alpha_{21} b_1 + \alpha_{22} b_2 + \cdots + \alpha_{2n} b_n \\ \cdots \\ c_m = \alpha_{m1} b_1 + \alpha_{m2} b_2 + \cdots + \alpha_{mn} b_n \end{array}\right. \iff \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Se virmos a última equação como um sistema linear nos b_i's, com n incógnitas e m equações, ou seja, mais equações que variáveis, percebemos que ele é possível e, portanto, somente as n primeiras equações já determinam os b_i's (já que, da hipótese, os c_i's são linearmente independentes). Mas, se substituirmos os valores dos b_i's nas demais $m - n$ equações, podemos escrever c_j, $n < j \leq m$, como combinação linear dos c_i's, $1 \leq i \leq n$. Absurdo. ∎

Isso tem uma conseqüência importante: para caracterizar completamente um espaço vetorial, basta encontrar uma base desse espaço. Podemos, inclusive, modelar todo espaço vetorial de dimensão n em função de uma base $B = \{u_1, u_2, \ldots, u_n\}$ como n-uplas ordenadas: se $v = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$, podemos representar v por $(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

Teorema 2.4. *Se $S \subset V$ é um subespaço vetorial e* $\dim S = \dim V$ *então $S = V$.*

Demonstração

Seja B uma base de V e C uma base de S. Então B e C têm o mesmo número de elementos. Escrevendo os vetores de C como combinação linear de vetores de B, obtemos um sistema linear nos vetores de B, que é possível (claro!) e determinado, pois o número de equações é igual ao número de incógnitas e os vetores de C são linearmente independentes. Logo todo vetor de B e, portanto, todo vetor de V, pode ser escrito como combinação linear de vetores de C (é só resolver o sistema!). Logo $V \subset S$ e concluímos que $S = V$. ∎

3. O problema principal da Teoria dos Códigos

Para nós, a partir de agora, uma mensagem é um vetor do espaço vetorial $V = Z/2Z^n$ (ou seja, uma mensagem é uma n-upla ordenada de zeros e uns, e somamos e multiplicamos $\bmod 2$). Você pode pensar em bits, se achar melhor. A partir de agora, V denotará esse espaço vetorial.

O problema que vamos estudar agora é o seguinte: ao transmitir uma mensagem, o canal de comunicação soma ao vetor transmitido c um *vetor erro* e, de modo que o destinatário recebe o vetor $x = c + e$. A meta do destinatário é decodificar x, ou seja, determinar o vetor erro para reconstruir c a partir de x.

O nosso medidor de erros é a importantíssima *distância de Hamming*.

4. Distância de Hamming

Definição 4.1. *Seja $v = (v_1, v_2, \ldots, v_n)$ e $w = (w_1, w_2, \ldots, w_n)$ vetores de V. A distância $d(v, w)$ é definida como o número de posições nas quais v e w diferem, ou seja,*

$$d(v, w) = |\{i \mid v_i \neq w_i\}|$$

(aqui, $|X|$ indica o número de elementos do conjunto X)

O conceito de distância vem da Topologia. Dado um conjunto A, uma *métrica em A* é uma função $d: A \times A \to R$ deve ter as seguintes propriedades: para $u, v, w \in A$,

(1) $d(v, w) \geq 0$ com igualdade se, e somente se, $v = w$.

(2) $d(v, w) = d(w, v)$

(3) Vale a *desigualdade triangular*, ou seja,

$$d(u, w) \leq d(u, v) + d(v, w)$$

A distância de Hamming é uma métrica em V. De fato:

(1) $d(v, w) \geq 0$ pois é um número natural e $d(v, w) = 0$ significa que v e w não diferem em nenhuma posição, ou seja, $v = w$.

(2) Claramente $d(v, w) = d(w, v)$.

(3) Podemos supor, sem perda de generalidade, que u e w diferem exatamente nas primeiras $a = d(u, w)$ posições. Dentre essas a posições, suponha que v e w diferem em b posições. Além disso, suponha que v e w diferem em c posições fora das a primeiras. Temos, então, $d(v, w) = b + c$.

$$\begin{array}{ll}
u & \overbrace{\mathrm{x\ x\ x\ x\ x}}^{a}\ \mathrm{x\ x\ x\ x\ x\ x\ x} \\
w & \mathrm{o\ o\ o\ o\ o\ x\ x\ x\ x\ x\ x\ x} \\
v & \underbrace{\mathrm{x\ x\ x}}_{b}\ \mathrm{o\ o}\ \underbrace{\mathrm{o\ o\ o}}_{c}\ \mathrm{x\ x\ x\ x}
\end{array}$$

Olhando a figura acima, podemos concluir que $d(u, v) = a - b + c$. Assim,

$$d(u, v) + d(v, w) = a - b + c + b + c = a + 2c \geq a = d(u, w)$$

Em Topologia, logo depois de definir métrica definimos esferas. Na geometria, uma esfera de raio r e centro P é o conjunto de pontos cuja distância a P é menor ou igual a r. Da mesma forma:

Definição 4.2. *Seja $v \in V$ e r um inteiro não negativo. Então*

$$S_r(v) = \{x \in V \mid d(x, v) \leq r\}$$

é uma esfera de raio r e centro v.

Enfim, chegamos em:

5. Códigos corretores de erros

Definição 5.1. *Seja t um inteiro positivo. Um subconjunto C de V é um t-código corretor de erro se quaisquer dois elementos distintos de $v, w \in C$ satisfazem*

$$d(v, w) \geq 2t + 1$$

Uma outra maneira de dizer isso é que a *distância mínima* de C, definida por

$$d(C) = \min\{d(c, c') \mid c, c' \in C, c \neq c'\},$$

é pelo menos $2t + 1$.

Por que o nome t-código corretor de erro? Considere as esferas com centro em vetores de C e raio t. Essas esferas não se interceptam, já que a distância entre dois centros quaisquer é pelo menos $2t+1$ (de fato, se x pertence à interseção de duas esferas $S_t(u)$ e $S_t(v)$, então $d(u, x) \leq t$ e $d(x, v) \leq t \implies d(u, x)+d(x, v) \leq 2t < 2t + 1 \leq d(u, v)$, o que contradiz a desigualdade triangular). Assim, se um vetor $c \in C$ é transmitido com no máximo t erros, obtendo x, basta encontrar a esfera a que x pertence. Em outras palavras: dado x, o vetor correto c é o vetor de C mais próximo de x.

Assim, temos (?) como resolver o problema: primeiro medimos ou estimamos o número máximo t de erros que o canal de comunicação pode gerar; em seguida, construímos um t-código corretor de erro.

Nesse ponto, podemos dizer de modo mais preciso qual é a meta da Teoria dos Códigos: construir códigos que

- tenham distância mínima grande (e, portanto, possam corrigir um número grande de erros);
- permitam um algoritmo de decodificação eficiente.

Exercícios

01. Um subconjunto C de $V = Z/2Z^n$ é um *t-código detector de erro* quando, sendo $e \in V$ um vetor com no máximo t entradas iguais a 1, para todo $c \in C$ o vetor $c + e$ não pertence a C.

Prove que C é um t-código detector de erro se, e somente se, $d(C) \geq t + 1$.

6. Um exemplo de código corretor de erro

O (?) na seção anterior é um aviso: não sabemos ainda se o problema de construir um código corretor de erro é fácil ou não; e o pior, não sabemos nem se existem t-códigos corretores de erro para valores arbitrários de t (embora não pareça muito difícil provar isso).

Por isso, exibimos o seguinte exemplo, que vai servir para esse artigo.

Um 1-código corretor de erro

0000000	1111111
1110000	0001111
1001100	0110011
1000011	0111100
0101010	1010101
0100101	1011010
0011001	1100110
0010110	1101001

Verifique que, de fato, a distância mínima entre dois vetores quaisquer acima é 3.

7. Códigos lineares

Até agora, não temos um modo sistemático de criar códigos corretores de erros. Tudo o que temos é só um exemplo. Mas, antes de construir mais códigos, vamos pensar um pouco em como facilitar nosso trabalho. Veja que dado um código, precisamos saber determinar seus elementos, sua distância mínima (um problema de ordem $|C|^2$) e decodificar todos os possíveis vetores que o destinatário pode receber.

Para melhorar um pouquinho esses problemas e começar a ter códigos que possam realmente ser aplicados na prática, usamos *códigos lineares*. O termo "linear" não é coincidência: faremos bastante uso de Álgebra Linear.

Definição 7.1. *Um código $C \subset V$ é dito linear quando C é um subespaço vetorial de V. Sendo k a dimensão de C (k existe porque C é finito), dizemos que C é um $[n,k]$-código linear.*

7.1. Por que usar códigos lineares, parte 1: determinando seus elementos

A vantagem em utilizar códigos lineares é evidente: para caracterizar todos os elementos de um código linear basta obter uma de suas bases.

Definição 7.2. *Seja $\{c_1, c_2, \ldots, c_k\}$ uma base do $[n,k]$-código linear C. A matriz $k \times n$ cuja i-ésima linha são as entradas do vetor c_i é uma matriz geradora de C.*

Exemplo 7.1.

Uma matriz geradora do nosso código exemplo é

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

Note que para guardar uma matriz geradora precisamos de somente k vetores, o que é uma fantástica economia, face aos 2^k vetores que compõem o código.

7.2. Por que usar códigos lineares, parte 2: determinando sua distância mínima

Precisamos de algumas definições.

Definição 7.3. *O peso $w(x)$ de um vetor $x \in V$ é o número de entradas não nulas de x, ou seja, $d(x, 0)$.*

Definição 7.4. *O peso mínimo $w(C)$ de um código C é dado por*

$$w(C) = \text{mín}\{w(c) \mid c \in C, c \neq 0\}$$

Teorema 7.1. *Seja C um código linear. Então*

$$d(C) = w(C)$$

Demonstração

Para todo código linear, $d(C) \leq w(C)$, já que $d(C)$ é a distância mínima entre dois vetores quaisquer de C e $w(C)$ é uma distância entre dois vetores de C (um deles é o vetor nulo 0).

Para terminar, basta provar que existe um vetor com peso $d(C)$. Sejam c e c' dois vetores cuja distância é mínima, ou seja, $d(c, c') = d(C)$. Assim,

$$w(c - c') = d(c, c') = d(C)$$

Como C é linear, sendo c e c' elementos de C então $c - c' \in C$ (basta subtrair as representações de c e c' como combinações lineares dos vetores da base). Logo $c_0 = c - c' \in C$ tem peso $d(C)$. ∎

Assim, para achar a distância mínima, só precisamos calcular todos os pesos dos vetores, o que é um problema que precisa de no máximo $|C|$ passos, no lugar de $|C|^2$.

7.3. Por que usar códigos lineares, parte 3: decodificando

Essa é a parte em que utilizaremos mais Álgebra Linear e também a mais longa (e interessante!) das três.

Definição 7.5. *Seja $C \subset V$ um código. O código dual de C, denotado por $C^\perp$, é definido como o conjunto dos vetores que são ortogonais a todos os vetores de C, ou seja,*

$$C^\perp = \{v \in V \mid v \cdot c = 0 \text{ para todo } c \in C\}$$

sendo que $v \cdot c$ é o produto interno usual: se $v = (v_1, v_2, \ldots, v_n)$ e $c = (c_1, c_2, \ldots, c_n)$ então

$$v \cdot c = v_1c_1 + v_2c_2 + \cdots + v_nc_n$$

(lembre que, no caso, a aritmética é feita $\bmod 2$)

Infelizmente, o nome é dual mesmo no lugar de ortogonal.

Teorema 7.2. *Se C é um $[n, k]$-código linear, então $C^\perp$ é um espaço vetorial de dimensão $n - k$.*

Demonstração

Primeiro, provemos que $C^\perp$ é um subespaço de V. Como $0 \cdot c = 0$ para qualquer vetor $c \in V$, $0 \in C^\perp$.

Vamos provar que sendo $v, w \in C^\perp$ e $\lambda \in Z/2Z$, $v + \lambda w \in C^\perp$. Aqui, $v = (v_1, v_2, \ldots, v_n)$, $w = (w_1, w_2, \ldots, w_n)$ e $c = (c_1, c_2, \ldots, c_n)$.

$$\begin{aligned}
\left|\begin{array}{l} v \in C^\perp \\ w \in C^\perp \end{array}\right. &\iff \left|\begin{array}{l} v \cdot c = 0, c \in C \\ w \cdot c = 0, c \in C \end{array}\right. \\
&\iff \left|\begin{array}{l} v_1c_1 + v_2c_2 + \cdots + v_nc_n = 0, c \in C \\ w_1c_1 + w_2c_2 + \cdots + w_nc_n = 0, c \in C \end{array}\right. \\
&\implies v_1c_1 + v_2c_2 + \cdots + v_nc_n + \lambda(w_1c_1 + w_2c_2 + \cdots + w_nc_n = 0), c \in C, \lambda \in Z/2Z \\
&\iff (v_1 + \lambda w_1)c_1 + (v_2 + \lambda w_2)c_2 + \cdots + (v_n + \lambda w_n)c_n = 0, c \in C, \lambda \in Z/2Z \\
&\iff (v + \lambda w) \cdot c = 0, c \in C, \lambda \in Z/2Z
\end{aligned}$$

Agora, encontremos a dimensão de $C^\perp$. Observando que

$$v_1c_1 + v_2c_2 + \cdots + v_nc_n = 0 \iff \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = 0$$

consideremos a matriz geradora G de C. Note que para que um vetor v seja ortogonal a todos os vetores de C é necessário e suficiente que v seja ortogonal a todos os vetores da base de C. Assim, v deve ser solução do sistema linear, $n \times k$, $G \cdot v = 0$. O posto de G é k, pois as linhas de G são linearmente independentes. Assim, o sistema $G \cdot v = 0$ tem grau de indeterminação $n - k$, o que gera um subespaço de dimensão $n - k$ (você pode pensar no seguinte: cada variável arbitrária aumenta a dimensão do subespaço em 1). ∎

Como será o código dual do código dual de C, ou seja, o que podemos esperar de $C^{\perp\perp}$? Isso é o que diz o

Teorema 7.3. *Se C é um código linear, $C^{\perp\perp} = C$.*

Demonstração

Primeiro provaremos que $C \subset C^{\perp\perp}$. O conjunto $C^{\perp\perp}$ consiste de todos os vetores ortogonais a todos os vetores de $C^{\perp}$. Eles incluem os vetores de C, pois $C^{\perp}$ é o conjunto de todos os vetores que são ortogonais a cada vetor de C. (dica, segundo a própria referência [1]: leia esse parágrafo de novo, mas bem devagar. Aí você vai entender. Para mim, funcionou bem.)

A dimensão de $C^{\perp\perp}$ é $n - (n-k) = k$. Logo $\dim C^{\perp\perp} = \dim C$, e portanto $C^{\perp\perp} = C$. ∎

Mais algumas definições. Calma, que falta pouco para aprendermos a decodificar códigos lineares!

Definição 7.6. *Seja $C \subset V$ um $[n,k]$-código linear. Uma matriz H cujas linhas formam uma base de $C^{\perp}$ é uma matriz de checagem de paridade de C.*

Observe que H, como definido anteriormente, é uma matriz $(n-k) \times n$.

Definição 7.7. *Seja H um matriz de checagem de paridade de um código linear C. Para cada vetor $v \in V$, a sua síndrome $s(v)$ é definida por*

$$s(v) = v \cdot H^t,$$

sendo H^t a transposta de H. Note que a síndrome de um vetor é um vetor de dimensão $n-k$.

Notemos que se $v \in C$ então $s(v) = 0$, já que cada entrada de $s(v)$ é o produto escalar de um vetor de $C^{\perp}$ e v. De fato, usando essa propriedade podemos descrever um código linear muito facilmente usando síndromes.

Teorema 7.4. *Se C é um código linear com matriz de checagem de paridade H então*

$$C = \{v \in V \mid s(v) = 0\}$$

Demonstração

Seja $v \in V$. Então

$$\begin{aligned} s(v) = 0 &\iff v \cdot H^t = 0 \\ &\iff v \text{ é ortogonal a todos os vetores de uma base de } C^{\perp} \\ &\iff v \in C^{\perp\perp} \\ &\iff v \in C \end{aligned}$$

∎

Uma *coclasse* de um espaço vetorial $S \subset V$ é definida como

$$v + S = \{v + w \mid w \in S\},$$

sendo $v \in V$.

O teorema a seguir nos diz que a síndrome $s(v)$ depende somente da coclasse que contém v.

Teorema 7.5. *Seja H uma matriz de checagem de paridade de um código linear $C \subset V$. Então, sendo $v, w \in V$*

$$s(v) = s(w) \iff v + C = w + C$$

Demonstração

$$s(v) = s(w) \iff v \cdot H^t = w \cdot H^t \iff (v-w) \cdot H^t = 0 \iff v - w \in C \iff v + C = w + C$$

Definição 7.8. *Seja $C \subset V$ um código linear. Um vetor é dito líder de uma coclasse de C se tem peso mínimo entre todos os vetores da coclasse.*

Em geral, o líder de uma coclasse não é único. Mas o resultado a seguir mostra como os líderes podem facilitar nossa vida.

Teorema 7.6. *Seja $C \in V$ um t-código corretor de erro linear. Então*

(a) cada vetor de V cujo peso é no máximo t é líder de alguma coclasse;

(b) o líder de uma coclasse que contém um vetor de peso no máximo t é unicamente determinado.

Demonstração

Seja $v \in V$ um vetor com peso menor ou igual a t. Devemos provar que se $v' \in v + C$ é diferente de v então o peso de v' é pelo menos $t + 1$.

Observe que $v' - v \in C$. Logo, sendo C um t-código corretor de erro linear, $w(C) = d(C) \geq 2t + 1$. Assim, o peso de $v' - v$ é pelo menos $2t + 1$. Assim, pela desigualdade triangular,

$$2t + 1 \leq d(v - v', 0) = d(v, v') \leq d(v, 0) + d(v', 0) \leq t + w(v')$$

e, portanto, $w(v') \geq t + 1$. ∎

Note que os vetores de peso no máximo t são os possíveis erros, ou seja, os líderes considerados são as possíveis diferenças. Deste modo, o último teorema nos dá um algoritmo de decodificação: primeiro, determina-se a coclasse de C que contém a mensagem recebida x. O vetor de erro de x é o líder da coclasse. Assim, é só somar x ao líder para obter a mensagem correta.

Veja que, pelo teorema 7.6., o líder e determina a coclasse; como x pertence à coclasse, $x \in e + C$, ou seja, $x = e + c$ para algum $c \in C$. Observe que $d(x, c) = w(e) \leq t$, logo $x \in S_t(c)$. Pelo que foi discutido na seção 5, a mensagem transmitida só pode ser c! E resolvemos o problema da decodificação.

Decodificação de síndromes. *Seja $C \subset V$ um $[n, k]$-código linear que é t-corretor de erro. Dada uma lista com os líderes de coclasses e suas respectivas síndromes, decodificamos um vetor x recebido da seguinte forma:*

(i) calculamos sua síndrome $s(x)$;

(ii) procuramos $s(x)$ na lista das síndromes;

(iii) encontramos o líder correspondente e;

(iv) decodificamos x para $c = x + e$.

Exemplo 7.2.

Considere o nosso código exemplo. Uma matriz de checagem de paridade desse código (mais tarde provaremos isso) é

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

Os líderes de coclasse são 0000000, 0000001, 0000010, etc. Assim, nossa lista de líderes de coclasses e síndromes é (verifique!!)

líder de coclasse	**síndrome**
0000000	000
0000001	111
0000010	011
0000100	101
0001000	110
0010000	001
0100000	010
1000000	100

Se, por exemplo, recebemos o vetor $x = 0010001$, calculamos sua síndrome $s(x) = 110$. Da lista, obtemos o líder 0001000. Logo a mensagem é $0010001 + 0001000 = 0011001$, que realmente consta no nosso código.

Exercícios

02. Decodifique o vetor 1100011 no código acima.

03. Seja $C \subset Z/2Z^n$ definido por $C = \{(a_1, a_2, \ldots, a_n) \mid a_1 + a_2 + \cdots + a_n = 0\}$. Prove que C é um código linear e 1-detector de erros.

8. Códigos de Hamming

Definiremos os códigos de Hamming a partir de sua matriz de checagem de paridade.

Definição 8.1. *Sejam r um inteiro positivo e $n = 2^r - 1$. Considere uma matriz H, $r \times (2^r - 1)$, com entradas binárias cujas colunas são as r-uplas de entradas binárias diferentes do vetor nulo 0. Definimos o código de Hamming* $\mathrm{Ham}(r)$ *por*

$$\mathrm{Ham}(r) = \{c \in Z/2Z^n \mid c \cdot H^t = 0\}$$

O nosso código exemplo é $\mathrm{Ham}(3)$ (verifique e aproveite para verificar que a matriz checadora de paridade dada anteriormente está correta!!).

Veja que o posto da matriz H é r (é só tomar as colunas cujas entradas são todas nulas, exceto uma, ou seja, $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 1)$). Assim $\mathrm{Ham}(r)$ é um $[2^r - 1, 2^r - 1 - r]$-código linear.

Primeiro vamos descobrir quantos erros esse código pode corrigir.

Teorema 8.1. *Os códigos de Hamming são 1-corretores de erro.*

Demonstração

Como $d(C) = w(C)$ para todo código linear C, basta provar que o peso de códigos de Hamming é pelo menos 3.

Suponha que o peso de um código de Hamming seja 1. Assim, esse código contém um vetor c cujas entradas são todas nulas, com exceção de uma, digamos, na i-ésima posição. Da definição de $\mathrm{Ham}(r)$, $c \cdot H^t = 0$. Isso implica que todas as entradas da i-ésima coluna de H são nulas, absurdo.

Agora, suponha que o peso de um vetor c de um código de Hamming seja 2, ou seja, que todas as entradas de c são nulas, com exceção de duas, digamos, nas posições i e j. De $c \cdot H^t = 0$, concluímos que a soma das colunas i e j de H é o vetor nulo. Isso implica que tais colunas (note que estamos trabalhando mód 2) são iguais, outro absurdo.

Logo o peso de códigos de Hamming é pelo menos 3. ∎

Os códigos de Hamming podem corrigir poucos erros, mas são os mais densos, no seguinte sentido:

Definição 8.2. *Um t-código corretor de erro $C \subset V$ é dito perfeito se qualquer vetor de V tem distância no máximo t de um elemento de C (que, claro, é único).*

Pensando topologicamente, um código C é perfeito quando as esferas de raio t e centro em vetores de C cobrem todo o espaço vetorial V.

Considerando que todas as esferas citadas são disjuntas, parece difícil acreditar na existência de códigos perfeitos. Provaremos que os códigos de Hamming são perfeitos.

Mas, antes, provemos o

Lema 8.1. *Seja $V = Z/2Z^n$ e seja $C \subset V$ um 1-código corretor de erro. Então*

$$|C| \leq \frac{2^n}{n+1}$$

A igualdade ocorre se, e somente se, C é perfeito.

Demonstração

Contemos o número de vetores em cada esfera. O número de vetores à distância zero do centro é 1: o próprio centro. A quantidade de vetores à distância 1 do centro é n: basta trocar cada entrada da n-upla que representa o centro. Logo cada esfera contém $n+1$ vetores.

Como há $|C|$ esferas disjuntas e 2^n vetores em V,

$$|C|(n+1) \leq 2^n \iff |C| \leq \frac{2^n}{n+1}$$

A igualdade ocorre se, e somente se, as esferas cobrem V, ou seja, se, e somente se, C é perfeito. ∎

Como corolário, temos que um 1-código corretor de erro perfeito $|C|$ deve ter dimensão da forma $n = 2^r - 1$, já que $2^n/(n+1)$ deve ser inteiro.

Exercícios

04. Prove que se existe um t-código corretor de erro contido em $Z/2Z^n$, linear e perfeito, então

$$\sum_{0 \leq i \leq t} \binom{n}{i}$$

é uma potência de 2. Mostre que se $t = 3$, então $n = 7$ ou $n = 23$.

Observação: O caso $n = 7$ é trivial (tente descobrir por quê). No caso $n = 23$, só existe essencialmente um código G_{23}, denominado *código de Golay*. Uma matriz de checagem de paridade de G_{23}, que tem 4096 vetores, é

$$\begin{pmatrix}
1&0&0&0&0&0&0&0&0&0&0&1&0&0&1&1&1&0&0&0&1&1&1\\
0&1&0&0&0&0&0&0&0&0&0&1&0&1&0&1&1&0&1&1&0&0&1\\
0&0&1&0&0&0&0&0&0&0&0&1&0&1&1&0&1&1&0&1&0&1&0\\
0&0&0&1&0&0&0&0&0&0&0&1&0&1&1&1&0&1&1&0&1&0&0\\
0&0&0&0&1&0&0&0&0&0&0&1&1&0&0&1&1&1&0&1&1&0&0\\
0&0&0&0&0&1&0&0&0&0&0&1&1&0&1&0&1&1&1&0&0&0&1\\
0&0&0&0&0&0&1&0&0&0&0&1&1&0&1&1&0&0&1&1&0&1&0\\
0&0&0&0&0&0&0&1&0&0&0&1&1&1&0&0&1&0&1&0&1&1&0\\
0&0&0&0&0&0&0&0&1&0&0&1&1&1&0&1&0&1&0&0&0&1&1\\
0&0&0&0&0&0&0&0&0&1&0&1&1&1&1&0&0&0&0&1&1&0&1\\
0&0&0&0&0&0&0&0&0&0&1&0&1&1&1&1&1&1&1&1&1&1&1
\end{pmatrix}$$

Teorema 8.2. *Os códigos de Hamming são perfeitos.*

Demonstração

Como $\dim(\mathrm{Ham}(r)) = 2^r - 1 - r$, $|\mathrm{Ham}(r)| = 2^{2^r-1-r}$. Sendo $n = 2^r - 1 \iff 2^r = n+1$, $|\mathrm{Ham}(r)| = 2^n/2^r = 2^n/(n+1)$. Do lema 8.1., o resultado segue. ∎

Mas o que é interessante sobre códigos de Hamming é a simplicidade de sua decodificação. Para isso, vamos ordenar as colunas da matriz de checagem de paridade H, na ordem numérica. Ou seja, a i-ésima coluna s_i é a representação binária do número inteiro i. Veja que a última coluna de H só tem uns.

Teorema 8.3. *Seja* $\mathrm{Ham}(r)$ *o código cuja matriz* H *está ordenada como descrito acima. Então, para cada vetor* $v \in V - C$ *sua síndrome* $s(v)$ *é o valor* s_i *para o qual* $v - e_i \in C$*, onde* e_i *é o vetor com todas as entradas nulas, exceto a* i*-ésima. Portanto a síndrome indica a posição onde ocorreu o erro.*

Demonstração

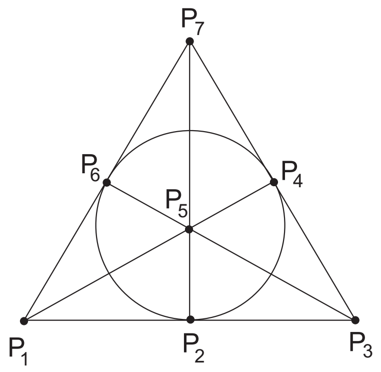
Sendo os e_i's os únicos vetores erro possíveis, todo vetor de $V - C$ é da forma $v = c + e_i$, $c \in C$, i inteiro. Visto que

$$s(v) = (c + e_i) \cdot H^t = c \cdot H^t + e_i \cdot H^t = e_i \cdot H^t = s_i,$$

a síndrome de v é s_i. ∎

Assim, o algoritmo de decodificação de um vetor x é extremamente simples: primeiro, calcula-se a síndrome $s(x)$, a lemos como um inteiro i na base binária, e somamos o vetor e_i a x, obtendo o vetor correto.

Na verdade, os códigos de Hamming têm a ver com geometria projetiva. Considere o nosso exemplo $\mathrm{Ham}(3)$. Além disso, considere o plano projetivo $P_2(Z/2Z)$ (olha o plano heptaprojetivo aí de novo, gente!).



Como cada elemento de $\mathrm{Ham}(3)$ tem comprimento 7, podemos associar a cada elemento $c \in \mathrm{Ham}(3)$ um conjunto $\mathcal{S}(c)$ de pontos de $P_2(Z/2Z)$, onde o ponto P_i pertence a $P_2(Z/2Z)$ se, e somente se, a i-ésima entrada de c é 1. Dizemos que c é um *vetor característico* de $\mathcal{S}(c)$. Por exemplo, se $c = 1110000$ então $\mathcal{S}(c) = \{P_1, P_2, P_3\}$.

Quais são os conjuntos de pontos de $P_2(Z/2Z)$ que correspondem aos elementos de $\mathrm{Ham}(3)$? Você pode verificar que, exceto o conjunto vazio e o conjunto de todos os pontos, esses conjuntos são as *retas* de $P_2(Z/2Z)$ e os complementos das mesmas retas!

Podemos verificar geometricamente que $\mathrm{Ham}(3)$ é 1-corretor de erro. No caso, um erro em $c \in \mathrm{Ham}(3)$, que é trocar um por zero ou vice-versa, corresponde a tirar ou colocar um ponto do conjunto $\mathcal{S}(c)$. Assim, basta provar que todo conjunto de pontos de $P_2(Z/2Z)$ pode ser transformado em um conjunto correspondente a um elemento de $\mathrm{Ham}(3)$ adicionando ou tirando um ponto.

Vamos tentar decodificar os vetores, mas agora pensando geometricamente.

Vetores de peso 1, 2 e 6 são fáceis de decodificar (basta tirar o ponto, considerar a reta que passa pelos dois pontos e colocar o ponto que falta, respectivamente).

Vetores de peso 3 que correspondem a retas já correspondem a elementos de $\mathrm{Ham}(3)$, assim não há o que decodificar. Vamos pensar num conjunto de 3 pontos não colineares, A, B e C. Tirar um ponto não ajuda, assim devemos adicionar um ponto para formar um complemento de reta. Há um único ponto D pertencente à reta que passa por A e B e um outro único ponto E pertencente à reta que passa por A e C (pois a interseção dessas duas retas é A). A reta que passa por D e E contém mais um único ponto F, distinto de A, B e C. Logo devemos considerar o complemento da reta $\{D, E, F\}$.

Um vetor de peso 4 que não pertence a $\mathrm{Ham}(3)$ é uma reta mais um ponto (veja o raciocínio do parágrafo anterior). Assim, basta tirar o ponto.

Um vetor de peso 5 só não contém dois pontos. Esses dois pontos determinam uma reta que contém um terceiro ponto P pertencente ao conjunto correspondente a esse vetor. Deste modo, basta tirar o ponto P para obter o complemento de uma reta.

Exercícios

05. Um *quadrilátero* de um plano projetivo é um conjunto de quatro pontos do plano que não contém três pontos colineares. Prove que todo quadrilátero de $P_2(Z/2Z)$ é o complemento de uma reta, ou seja, que os conjuntos correspondentes a vetores de Ham(3) são o vazio, todos os pontos, as retas e os quadriláteros de $P_2(Z/2Z)$.

06. Prove que todo conjunto de 5 pontos de $P_2(Z/2Z)$ é a união de duas retas de $P_2(Z/2Z)$.

O nosso exemplo Ham(3) pode ser comparado com um plano projetivo baseado em $Z/2Z$. E o código Ham(r), em geral? Antes de continuarmos precisamos falar de

8.1. Espaços projetivos

Relembrando: um *plano projetivo* é um conjunto de pontos sendo que existem subconjuntos chamados retas tais que dois pontos estão contidos em uma única reta e duas retas interceptam-se em um único ponto. Além disso, existem quatro pontos, três a três não colineares.

Essa última condição caracteriza o conjunto como plano. Porém, podemos definir *espaços projetivos* de dimensões maiores. Só que as propriedades que queremos nesses espaços são um pouquinho diferentes.

Definição 8.3. *Um espaço projetivo é um conjunto de pontos sendo que existem subconjuntos especiais, denominados retas que satisfazem as seguintes condições:*

(i) Dois pontos estão contidos em uma única reta.

(ii) Sejam A, B, C e D pontos tais que a reta AB intercepta CD. Então AC intercepta BD. Em outras palavras: se uma reta intercepta dois lados de um triângulo então intercepta o terceiro.

(iii) Cada reta contém pelo menos três pontos.

(iv) Existem pelo menos duas retas.

Você pode se perguntar: "ei! e como achamos a dimensão de um espaço projetivo?" Não vamos nos preocupar em definir dimensão sinteticamente (embora isso possa ser feito, com o auxílio de idéias semelhantes às da Álgebra Linear). Na verdade, para dimensões maiores que 2, vale sempre o

Teorema 8.4. *(Teorema de Desargues) Sejam A_1, A_2, A_3, B_1, B_2 e B_3 pontos tais que A_i e B_i são colineares com um ponto C, $i = 1, 2, 3$ e não haja pontos colineares entre C, A_1, A_2, A_3 nem entre C, B_1, B_2, B_3. Então os três pontos definidos por P_{ij} = interseção das retas A_iA_j e B_iB_j, $i, j = 1, 2, 3$, $i \neq j$ são colineares.*

E para todos os espaços nos quais vale esse teorema (sim, existem planos projetivos nos quais esse teorema não vale!), podemos encontrar um anel de divisão (que é quase um corpo; só não vale necessariamente a comutativa da multiplicação; um exemplo são os *quatérnions*, que podem ser encontrados em [5]) que coordenatizam o espaço. Aí definiremos dimensão com as coordenadas.

Exercícios

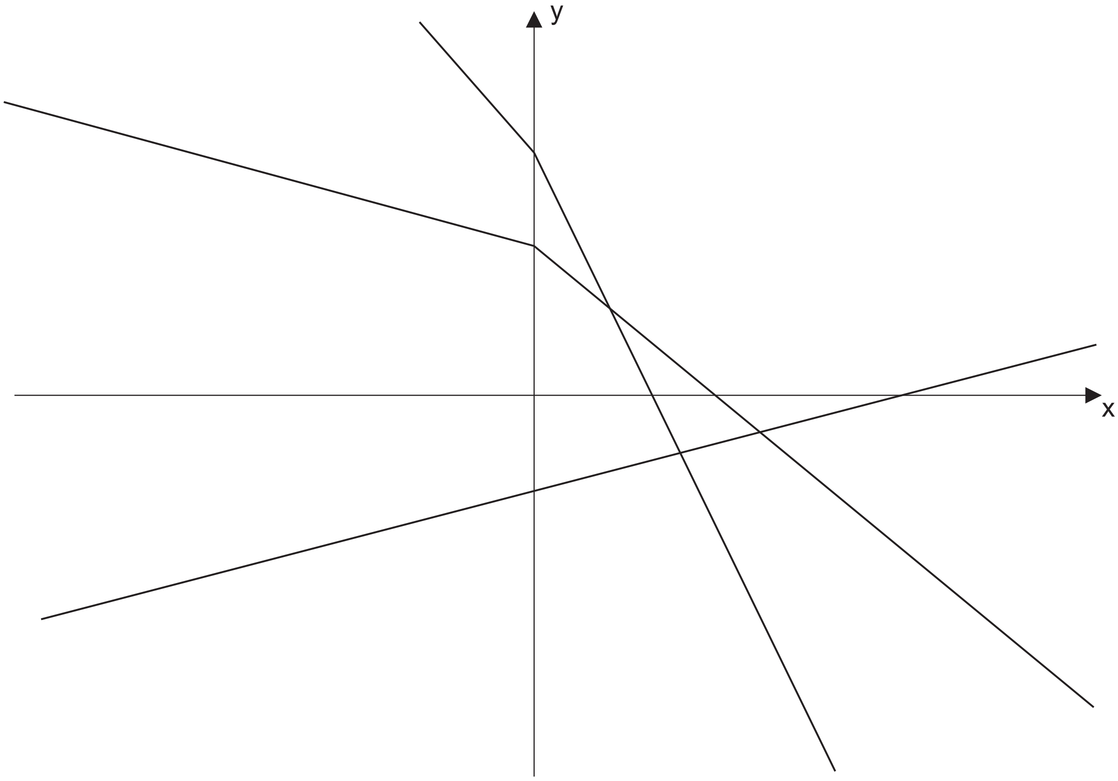
07. Prove que podemos definir planos projetivos usando as condições (i) a (iv) acima, mas trocando (ii) por

(ii') Quaisquer duas retas se interceptam.

Ou seja, que as duas definições dadas aqui coincidem.

08. O *plano de Moulton* pode ser definido como o conjunto de pontos de R^2 no qual as retas são descritas ou por uma equação da forma $x = c$, c constante real, ou da forma $y = mx + b$, m, b constantes reais. Só

que o critério para um ponto pertencer a uma reta é diferente: um ponto (x_0, y_0) pertence a uma reta da forma $x = c$ quando $x_0 = c$. Se $x_0 \leq 0$ ou $m \geq 0$, então (x_0, y_0) pertence a $y = mx + b$ se, e somente se, $y_0 = mx_0 + b$ (até aqui, normal). Agora, se $x_0 > 0$ e $m < 0$, então (x_0, y_0) pertence a $y = mx + b$ se, e somente se, $y_0 = \mathbf{2}mx_0 + b$.



Adicione a esse plano as retas do infinito. Mostre que o plano de Moulton unida às retas do infinito formam um plano projetivo no qual *não vale* o teorema de Desargues.

Mas, voltemos à nossa generalização. Relembramos que podemos coordenatizar um plano projetivo a partir de um corpo (ou anel de divisão) K definindo-o como um conjunto $P_2(K)$ de ternas $(x, y, z) \neq (0, 0, 0)$, $x, y, z \in K$, onde ternas da forma (x, y, z) e (kx, ky, kz) devem ser consideradas iguais, e na qual a reta dual do ponto (a, b, c) é o conjunto de pontos (x, y, z) que satisfazem

$$ax + by + cz = 0$$

Para espaços projetivos, não poderia ser muito diferente. Um espaço projetivo de dimensão d, baseado no corpo K, pode ser definido como um conjunto de $d+1$-uplas ordenadas $(a_0, a_1, \ldots, a_d) \neq (0, 0, \ldots, 0)$, onde $d + 1$-uplas da forma $(a_0, a_1, \ldots, a_d)$ e $(ka_0, ka_1, \ldots, ka_d)$, $k \in K$, devem ser consideradas iquais. Notamos esses espaços como $P_d(K)$.

Um *subespaço de dimensão t de* $P_d(K)$ é descrito por uma matriz H, $(d-t) \times (d+1)$, e de posto $d-t$ (se $t = 1$, o subespaço é chamado de *reta*; se $t = 2$, o subespaço é um *plano*; se $t = d-1$, temos um *hiperplano*). Os pontos desse subespaço são as soluções do sistema homogêneo

$$c \cdot H^t = 0$$

Nesse caso, como a equação $a_0x_0 + a_1x_1 + \cdots + a_dx_d = 0$ representa um hiperplano, ocorre dualidade entre pontos e hiperplanos em vez de retas.

Também podemos descrever subespaços t-dimensionais da seguinte forma: é um conjunto $\{\alpha_0 P_0+\alpha_1 P_1+\cdots+\alpha_t P_t\}$, onde $\alpha_0, \alpha_1, \ldots, \alpha_t$ são elementos de $Z/2Z$ e $P_0, P_1, \ldots, P_t$ são pontos linearmente independentes. Observe que esse conjunto é o subespaço que passa pelos pontos $P_0, P_1, \ldots, P_t$. A descrição sintética é próxima disso, mas é recursiva: tome dois pontos P_0 e P_1. Todos os pontos da reta P_0P_1 caracterizam o subespaço que passa por esses dois pontos e tem dimensão 1. Se houver um ponto P_2 fora de P_0P_1, ligue P_2 a todos os pontos de P_0P_1 e junte à reta. Obtemos o plano $P_0P_1P_2$. Para obter o subespaço de dimensão 3 que passa por P_0, P_1, P_2 e P_3, basta considerar o plano $P_0P_1P_2$ e as retas que ligam $P_3 \notin P_0P_1P_2$ a cada ponto do plano. E assim por diante.

Exercícios

09. Prove que um subespaço de dimensão t de $P_d(\mathrm{GF}(q))$, tem $1+q+\cdots+q^t$ pontos. *Dica: indução sobre t.*

10. Prove que o teorema de Desargues vale para todo espaço projetivo de dimensão maior ou igual a 3. *Dica: prove o teorema para R^3 e tnte imitar a demonstração no caso geral.*

11. (IMC 2003) Encontre todos os inteiros positivos n para os quais existe uma família F de subconjuntos de três elementos de $S=\{1,2,\ldots,n\}$ que satisfaz as seguintes condições:

(i) Para quaisquer elementos distintos $a, b \in S$ existe exatamente um $A \in F$ tal que $a, b \in A$.

(ii) Se a, b, c, x, y, z são tais que $\{a, b, x\}$, $\{a, c, y\}$, $\{b, c, z\} \in F$ então $\{x, y, z\} \in F$.

Dica: reveja os axiomas de espaços projetivos!

8.2. Voltando aos códigos de Hamming...

Considere a matriz de checagem de paridade H de $\mathrm{Ham}(r)$. Como H tem $2^r-1=1+2+2^2+\cdots+2^{r-1}$ colunas, vamos pensar no espaço projetivo de dimensão $r-1$. O que representam as linhas de H, em termos de pontos de $P_{r-1}(Z/2Z)$? Lembrando que as colunas são os números de 1 a 2^r-1 em sua representação binária, observamos que a linha i representa os pontos cuja representação binária tem um 1 na sua i-ésima casa, da esquerda para a direita. Ou seja, os pontos que **não** pertencem ao hiperplano $\pi_i\colon x_{r-i}=0$. Veja o exemplo $r=4$:

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

Note que, ao fazermos o produto escalar de dois vetores característicos u e v de dois subconjuntos de $P_{r-1}(Z/2Z)$, estamos calculando a paridade da interseção desses subconjuntos, já que estamos operando mód 2 e

$$u_i \cdot v_i = 1 \iff u_i = v_i = 1 \iff i \in \mathcal{S}(u) \text{ e } i \in \mathcal{S}(v) \iff i \in \mathcal{S}(u) \cap \mathcal{S}(v)$$

Assim, estamos procurando pelos conjuntos de pontos $\mathcal{P}$ tais que $\mathcal{P} \cap \overline{\pi_i} = \mathcal{P} - \pi_i$ tem um número par de pontos para $i=1,2,\ldots,r$.

Vamos considerar subespaços de $P_{r-1}(Z/2Z)$.

Lema 8.2. *Todo vetor característico de um subespaço de $P_{r-1}(Z/2Z)$ pertence a* $\mathrm{Ham}(r)$.

Demonstração

A interseção de dois subespaços é um subespaço (para ver isso, basta notar que para encontrar a interseção de dois subespaços precisamos resolver o sistema obtido juntando os sistemas correspondentes a cada subespaço). Logo, pelo exercício anterior, a interseção de dois subespaços tem cardinalidade da forma $1+2+\cdots+2^t$, que é ímpar.

Logo, se $\mathcal{P}$ é um subespaço de $P_{r-1}(Z/2Z)$,

$$|\mathcal{P} - \pi_i| = |\mathcal{P}| - |\mathcal{P} \cap \pi_i| \equiv 1 - 1 = 0 \pmod{2}$$

Portanto o vetor característico de $\mathcal{P}$ pertence a $\mathrm{Ham}(r)$. ∎

Lema 8.3. *Sejam $\mathcal{S}(u)$ e $\mathcal{S}(v)$ subespaços de $P_{r-1}(Z/2Z)$. Então $\mathcal{S}(u)\,\Delta\,\mathcal{S}(v) = \mathcal{S}(u+v)$.*

Demonstração

Decorre diretamente da definição de diferença simétrica. ∎

Note que isso quer dizer que podemos trocar a soma de vetores binários pela diferença simétrica entre os conjuntos que são representados pelos vetores.

Agora estamos prontos para caracterizar geometricamente $\mathrm{Ham}(r)$.

Lema 8.4. *Todo subconjunto de $P_{r-1}(Z/2Z)$ tem interseção não vazia com algum dos conjuntos $\overline{\pi_i}$.*

Demonstração

Como uma equação de π_i é $x_i = 0$, uma equação de $\overline{\pi_i}$ é $x_i = 1$. Logo, como cada ponto P de $P_{r-1}(Z/2Z)$ tem alguma coordenada igual a 1, P pertence a algum dos conjuntos $\overline{\pi_i}$. Na verdade, $P \in \overline{\pi_i}$ se, e somente se, a $r-i$-ésima coordenada de P é igual a 1. ∎

Teorema 8.5. $\mathrm{Ham}(r)$ *consiste nos vetores característicos de todos os subconjuntos de $P_{r-1}(Z/2Z)$ obtidos a partir da diferença simétrica de retas de $P_{r-1}(Z/2Z)$.*

Demonstração

Seja $\mathcal{P}$ um subconjunto qualquer de $P_{r-1}(Z/2Z)$. Suponha que $|\mathcal{P}| \geq 2$ e considere dois pontos distintos de $\mathcal{P}$. Seja r a reta que passa por esses dois pontos. Então, como r tem 3 pontos, $\mathcal{P}' = \mathcal{P}\,\Delta\, r$ tem no máximo $|\mathcal{P}|-1$ pontos (afinal, estamos tirando pelo menos dois pontos de $\mathcal{P}$ e colocando no máximo um). Note que $\mathcal{P} = \mathcal{P}'\,\Delta r$. Repetindo o procedimento um número finito de vezes, obtemos um conjunto $\mathcal{Q}$ vazio ou unitário. Se $\mathcal{Q} = \emptyset$, acabou. Se $\mathcal{Q}$ consiste de um ponto, do lema 8.4., esse ponto pertence a algum conjunto $\overline{\pi_i}$, de modo que o produto escalar dos vetores característicos de $\mathcal{Q}$ e $\overline{\pi_i}$ é 1. Logo, nesse caso, o vetor característico de $\mathcal{Q}$ não pertence a $\mathrm{Ham}(r)$. ∎

Note que esse teorema prova geometricamente que $\mathrm{Ham}(r)$ é 1-corretor de erro (e não melhor do que isso) e que todo subespaço de $P_d(Z/2Z)$ pode ser escrito como diferença simétrica de retas.

8.3. Códigos de Hamming estendidos

Podemos melhorar um pouquinho os códigos de Hamming, aumentando um pouquinho a sua distância mínima.

Definição 8.4. *Considere o código* $\mathrm{Ham}(r)$*. O código de Hamming estendido* $\mathrm{Ham}(r)^*$ *é obtido estendendo cada vetor de* $\mathrm{Ham}(r)$ *em uma posição; atribuímos a essa nova entrada o valor* 0 *ou* 1 *de modo que cada vetor tenha um número par de uns.*

Exemplo 8.1.

O código a seguir é $\mathrm{Ham}(3)^*$.

00000000	11111111
11100001	00011110
10011001	01100110
10000111	01111000
01010101	10101010
01001011	10110100
00110011	11001100
00101101	11010010

Teorema 8.6. $\mathrm{Ham}(r)^*$ *é um* $[2^r, 2^r - 1 - r]$*-código linear com distância mínima* 4.

Demonstração

Primeiro temos que provar que $\mathrm{Ham}(r)^*$ é linear, isto é, é um espaço vetorial. Para isso, basta mostrarmos que, para $u^*, v^* \in \mathrm{Ham}(r)^*$ temos $u^* + v^* \in \mathrm{Ham}(r)^*$.

Sejam u e v os vetores correspondentes em $\mathrm{Ham}(r)$. Sabemos que $u + v \in \mathrm{Ham}(r)$, assim só basta provarmos que a última entrada de $u^* + v^*$ é 1 se $w(u+v)$ é ímpar e 0 caso $w(u+v)$ é par. Mas não é difícil ver que $w(u) + w(v) \equiv w(u+v) \pmod{2}$. Se $w(u+v)$ é ímpar, $w(u)$ e $w(v)$ têm paridades diferentes, ou seja, a última entrada de um dos vetores u^*, v^* é 1 e do outro vetor é 0. Logo $u^* + v^*$ tem como última entrada 1. Da mesma forma, se $w(u+v)$ é par, $w(u)$ e $w(v)$ têm a mesma paridade e, portanto, $u^* + v^*$ têm números iguais na última entrada. Conseqüentemente, $u^* + v^*$ tem como última entrada 1. Logo $\mathrm{Ham}(r)^*$ é um espaço vetorial.

Não é muito difícil ver que $\mathrm{Ham}(r)^*$ e $\mathrm{Ham}(r)$ têm a mesma dimensão (basta pensar em combinações lineares!!).

Agora, sendo $\mathrm{Ham}(r)^*$ linear, basta provar que $w(\mathrm{Ham}(r)^*)$ é 4. Mas isso é fácil: $w(\mathrm{Ham}(r)^*)$ é par e é no mínimo 3 e no máximo 4. Logo a distância mínima de $\mathrm{Ham}(r)^*$ é 4. ∎

Vamos construir agora uma matriz de checagem de paridade de $\mathrm{Ham}(r)^*$. É natural utilizarmos a matriz de $\mathrm{Ham}(r)$.

Teorema 8.7. *Seja H uma matriz de checagem de paridade de* $\mathrm{Ham}(r)$*. Então podemos obter uma matriz de checagem de paridade de* $\mathrm{Ham}(r)^*$ *da seguinte forma: acrescente uma coluna de zeros a H; em seguida, adicione à matriz obtida uma linha de uns.*

Exemplo 8.2.

Uma matriz de checagem de paridade de $\mathrm{Ham}(3)^*$ é

$$H^* = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Demonstração

Como $\mathrm{Ham}(r)^*$ é $[2^r, 2^r - r - 1]$-linear, suas matrizes de checagem de paridade devem ser $(r+1) \times 2^r$.

Provemos que a matriz construída H^* tem posto $r+1$. As r primeiras linhas são linearmente independentes, pois parte delas são as linhas de uma matriz de checagem de paridade de $\mathrm{Ham}(r)$. E como a última entrada da última linha de H^* é 1, enquanto a das outras linhas é 0, a última linha de H^* não pode ser escrita como combinação linear das outras linhas. Logo H^* tem posto $r+1$.

Para terminar, basta provar que todas as linhas são ortogonais a todos os vetores de $\mathrm{Ham}(r)^*$. As r primeiras linhas são ortogonais a todos os vetores de $\mathrm{Ham}(r)^*$ pois são ortogonais aos vetores de $\mathrm{Ham}(r)$ e a última entrada não interessa, pois a última entrada das linhas correspondentes é zero. A última linha é ortogonal a todos os vetores de $\mathrm{Ham}(r)^*$ pois o produto escalar dessa linha com um vetor de $\mathrm{Ham}(r)^*$ é a soma de suas entradas vista $\bmod 2$, que é sempre zero, já que todos os vetores de $\mathrm{Ham}(r)^*$ têm um número par de uns. ∎

Exercícios

12. (OBM 2002) Arnaldo e Beatriz se comunicam durante um acampamento usando sinais de fumaça, às vezes usando uma nuvem grande, às vezes uma pequena.

No tempo disponível antes do café da manhã, Arnaldo consegue enviar uma seqüência de 24 nuvens. Como Beatriz nem sempre consegue distinguir uma nuvem pequena de uma grande, ela e Arnaldo fizeram um dicionário antes de ir para o acampamento. No dicionário aparecem N seqüências de 24 tamanhos de nuvem (como por exemplo a seqüência $PGPGPGPGPGPGGPGPGPGPGPGP$, onde G significa nuvem grande e P significa nuvem pequena). Para cada uma das N seqüências, o dicionário indica seu significado. Para evitar interpretações erradas, Arnaldo e Beatriz evitaram incluir no dicionário seqüências parecidas. Mais precisamente, duas seqüências no dicionário sempre diferem em pelo menos 8 das 24 posições.

Demonstre que $N \leq 4096$.

Observação: o código de Golay G_{23} estendido (notado por G_{24}) tem 4096 vetores com 24 entradas e com distância mínima 8. Familiar? Se você quiser entender melhor a construção de G_{24}, veja, por sua conta e risco (a construção é um pouco enrolada!!), [6].

8.4. Generalizações e outras aplicações de códigos de Hamming

Podemos generalizar códigos de Hamming para $\mathrm{GF}(q)$, q potência de primo. Os códigos de Hamming continuam sendo códigos perfeitos e são construídos de modo análogo, assim como sua versão estendida.

Códigos de Hamming, além de corrigirem erros, podem ser utilizados para compactar dados com pouca perda de informação. Isso vem do fato de $\mathrm{Ham}(r)$ ser perfeito e 1-corretor de erro: em vez de guardar uma n-upla binária x, $n = 2^r$, encontramos o vetor $c \in \mathrm{Ham}(r)$ mais próximo a x e guardamos a $(n-r)$-upla menor m, que é a única solução de $m \cdot G = c$, onde G é uma matriz geradora de $\mathrm{Ham}(r)$. Para decodificar m, basta tomarmos $c = m \cdot G$.

Como $\mathrm{Ham}(r)$ é perfeito e 1-corretor de erro, para todo x existe um vetor c nesse código tal que a distância entre x e c é no máximo 1 e c é único. Assim, o erro na descompactação é no máximo de uma posição.

Todavia, o problema de compactar dados é, de certo modo, dual ao problema de corrigir erros. Bons códigos corretores de erro têm distância mínima grande e bastantes vetores. Em contraste, bons códigos de compactação de dados têm distância mínima pequena (nesse caso, pensamos em *raio de cobertura*, o menor d tal que as esferas de raio d e centro nos vetores do código cobrem todo o espaço de todos os possíveis vetores) e poucos vetores.

9. Um pouco mais sobre códigos e geometria projetiva

Ainda não respondemos a seguinte pergunta: existem códigos que corrijam um número arbitrário de erros?

A resposta a essa pergunta vem, por incrível que pareça, da geometria projetiva. Mas antes de falarmos de retas e pontos, um lema.

Lema 9.1. *Seja C um código linear de vetores de n entradas com matriz de paridade H. Então*

$$d(C) \geq d \iff \text{todo conjunto de } d-1 \text{ colunas de } H \text{ é linearmente independente}$$

Demonstração

Como C é um código linear, $d(C) = w(C)$. Seja $w = w(C)$.

Demonstremos a volta primeiro. Seja $v = (v_1, v_2, \ldots, v_n)$ um vetor de C de peso mínimo w. Sendo $h_1, h_2, \ldots, h_n$ as colunas de H, temos

$$v \cdot H^t = 0 \iff v_1 h_1 + v_2 h_2 + \cdots + v_n h_n = 0$$

Logo a soma de w colunas é nula, ou seja, essas w colunas não são linearmente independentes. Logo $w > d - 1 \iff d(C) \geq d$.

Reciprocamente, suponha, por absurdo, que existam $s \leq w - 1$ colunas que não são linearmente independentes, ou seja, que

$$h_{i_1} + h_{i_2} + \cdots + h_{i_s} = 0$$

Considere o vetor x cujas entradas não nulas são as de posição $i_1, i_2, \ldots, i_s$. Temos $x \cdot H^t = 0 \iff x \in C$, o que é uma contradição, já que $w(x) = s < w$. ∎

Vamos refrasear o problema de encontrar t-códigos corretores de erro em termos de geometria projetiva.

Definição 9.1. *Um conjunto de pontos de um espaço projetivo coordenatizado é independente quando os vetores correspondentes aos pontos são linearmente independentes.*

Veja que é muito fácil construir conjuntos de t pontos independentes em um espaço de dimensão d, $t \leq d+1$. Basta tomar os t pontos $(1,0,0,\ldots,0)$, $(0,1,0,\ldots,0)$, $\ldots$, $(0,0,0,\ldots,1,\ldots,0)$, em que o i-ésimo ponto tem todas as coordenadas nulas, exceto a i-ésima.

Definição 9.2. *Um conjunto de n pontos de um espaço projetivo é um (n,s)-conjunto se s é o maior inteiro positivo para o qual todos os seus subconjuntos de s elementos são independentes.*

Teorema 9.1. *Seja n e r inteiros positivos. Um $[n, n-r]$-código linear de distância mínima d (e, portanto, $\lfloor \frac{d-1}{2} \rfloor$-corretor de erro) existe se, e somente se, existe um $(n, d-1)$-conjunto de n pontos no espaço projetivo $P_{r-1}(Z/2Z)$.*

Demonstração

Seja C um $[n, n-r]$-código com distância mínima d e H uma de suas matrizes de checagem de paridade.

Pelo lema anterior, as colunas de H são n r-uplas binárias, sendo que quaisquer $d-1$ delas são linearmente independentes. Consideremos esses vetores como coordenadas de pontos de $P_{r-1}(Z/2Z)$. O conjunto $\mathcal{M}$ formado tem n pontos, sendo que quaisquer $d-1$ deles compõem um conjunto independente. Além disso, $d-1$ é o valor máximo para o qual isso acontece, pois se no lugar dele fosse $D > d-1$ a distância mínima seria pelo menos $D + 1 > d$.

Reciprocamente, considere um $(n, d-1)$-conjunto $\mathcal{M}$ de $P_{r-1}(Z/2Z)$. Considere as coordenadas dos pontos de $\mathcal{M}$ como colunas de uma matriz H, $r \times n$. Escalonando completamente a matriz, obtemos uma matriz H' de posto p, sendo que as p primeiras linhas contêm como submatriz a matriz identidade I_p e as demais $r-p$ linhas são nulas. Troque a $p+1$-ésima linha de zeros por 1. Isso aumenta o posto em 1 e não altera a $(d-1)$-independência das colunas. Repita o procedimento até obter uma matriz de posto r. Considere o código definido por

$$C = \{x \in Z/2Z^n \mid x \cdot H^t = 0\}$$

Como quaisquer $d-1$ colunas continuam independentes, pelo lema 9.1., $d(C) \geq d$. ∎

Agora que sabemos da existência de códigos que corrigem um número arbitrário de erros, queremos agora os códigos mais amplos e que corrigem o maior número de erros.

Primeiro vejamos qual é a maior distância mínima possível de um $[n, n-r]$-código linear.

Teorema 9.2. *(limitante de Singleton) Seja d a distância mínima de um $[n, n-r]$-código linear. Então*

$$d \leq r + 1$$

Demonstração

Seja C um $[n, n-r]$-código linear. Sendo C linear, basta provarmos que $w(C) \le r+1$. Isso é simples: considere a matriz geradora G de C. Essa matriz é $(n-r)\times n$ e tem posto $n-r$. Assim, escalonando G obtemos uma matriz G', que também é geradora de C (lembre que C é um espaço vetorial):

$$G' = \left(\begin{array}{ccccc|c} 1 & 0 & 0 & \dots & 0 & \\ 0 & 1 & 0 & \dots & 0 & \\ 0 & 0 & 1 & \dots & 0 & G^* \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \dots & 1 & \end{array}\right)$$

Como cada linha de G^* tem no máximo r entradas não nulas, $w(C) \le r+1$. ∎

Definição 9.3. *Um $[n, n-r]$-código linear C que satisfaz $w(C) = r+1$ é chamado código MDS (do inglês "maximum distance separable").*

Do teorema anterior, a existência de códigos lineares depende da existência de conjuntos independentes em espaços projetivos.

Definição 9.4. *O maior número n tal que existe um $(n, d-1)$-conjunto em $P_{r-1}(q)$ é denotado por $\text{máx}_{d-1}(n, q)$.*

Assim, $\text{máx}_r(r, 2)$ é o maior comprimento de um $[n, n-r]$-código MDS linear. Pode-se provar, por exemplo, que $\text{máx}_3(r, 2) = 2^{r-1}$. Mas ainda não foram calculados os valores de $\text{máx}_3(r, q)$ para todos os valores de q.

Exercícios

13. Prove que $\text{máx}_2(r, q) = q^{r-1} + q^{r-2} + \cdots + q + 1$.
14. Prove que $\text{máx}_3(r, 2) = 2^{r-1}$.

Vamos agora falar de códigos de Reed-Muller. Porém, para definirmos códigos de Reed-Muller, vamos usar geometria afim.

10. Geometria afim

Definimos geometria afim a partir da geometria projetiva (embora o mais usual seja o inverso!). E, por incrível que pareça, é mais fácil trabalhar com geometria projetiva do que com afim (você vai entender por que logo).

Definição 10.1. *Seja P um espaço projetivo com dimensão $d \ge 2$ e seja H_∞ um hiperplano de P. O conjunto $A = P - H_\infty$, no qual as retas (ou mesmo os subespaços) de A são as mesmas de P, tirando os pontos de H_∞, é um espaço afim de dimensão d.*

O hiperplano H_∞ costuma ser chamado *hiperplano no infinito* e seus pontos, *pontos no infinito*. E dizemos que dois subespaços U e W são *paralelos* e denotamos $U \parallel W$ quando $U \cap H_\infty = W \cap H_\infty$.

Do mesmo modo que existem espaços projetivos finitos, é claro que existem espaços afins finitos e podemos coordenatizá-los.

Considere o espaço projetivo coordenatizado $P_d(K)$, sendo K um corpo. O espaço afim baseado em K, denotado por $A_d(K)$, é obtido tomando como hiperplano no infinito $H_\infty: x_0 = 0$. Nesse caso, $A_d(K)$ consiste nas $(d+1)$-uplas $(x_0, x_1, \dots, x_d)$ nas quais $x_0 \ne 0$ e $(x_0, x_1, \dots, x_d)$ e $(kx_0, kx_1, \dots, kx_d)$ devem ser considerados iguais. Porém, podemos escolher k tal que $kx_0 = 1$. Assim, reduzimos $A_d(K)$ às $(d+1)$-uplas $(1, x_1, \dots, x_d)$. Por fim, podemos eliminar o 1 da primeira coordenada, que é redundante, e obtemos d-uplas $(x_1, \dots, x_d)$. Desta vez, $(x_1, \dots, x_d)$ e $(kx_1, \dots, kx_d)$ são pontos diferentes!

A reta que passa pelos pontos $P = (p_1, \dots, p_d)$ e $Q = (q_1, \dots, q_d)$ é dada por $\{P + \lambda Q = (p_1 + \lambda q_1, \dots, p_d + \lambda q_d) \mid \lambda \in K\}$.

Contemos o número de pontos em subespaços de espaços afins finitos.

Teorema 10.1. *Seja $A_d(K)$ um espaço afim de dimensão d baseado num corpo K com q elementos. Então cada reta de $A_d(K)$ contém q pontos e, de modo mais geral, os subespaços de dimensão t de $A_d(K)$ contêm q^t pontos.*

Demonstração

Dados os pontos P e Q, e sendo $PQ = \{P + \lambda Q \mid \lambda \in K\}$, temos q escolhas para λ, logo cada reta tem q pontos.

O número de pontos de um subespaço S de dimensão t do espaço $P_d(K)$ é $1 + q + \cdots + q^t$. Devemos subtrair desse total os pontos que pertencem ao hiperplano $H_\infty: x_0 = 0$. Essa equação, junto com as que geraram S em $P_d(K)$, definem um subespaço de dimensão $t-1$, e que tem $1 + q + \cdots + q^{t-1}$ pontos. Logo o total de pontos do subespaço correspondente a S em $A_d(K)$ é $1 + q + \cdots + q^t - (1 + q + \cdots + q^{t-1}) = q^t$. ∎

Teorema 10.2. *Seja U um subespaço, de dimensão t, de $P_d(K)$ ou $A_d(K)$, sendo que o corpo K contém q elementos. Há $q^{d-t-1} + \cdots + q + 1$ subespaços de dimensão $t+1$ que contêm U.*

Demonstração

Primeiro vamos demonstrar o resultado para espaços projetivos.

Há $q^d + \cdots + q + 1 - (q^t + \cdots + q + 1) = q^d + \cdots + q^{t+1}$ pontos de $P_d(K)$ fora de U, logo basta escolher um desses pontos e ligar todas as retas que ligam esse ponto a cada ponto de U. Assim, há, a princípio, $q^d + \cdots + q^{t+1}$ subespaços. Todavia, um subespaço de dimensão $t+1$ tem $q^{t+1} + \cdots + q + 1$ pontos, sendo $q^t + \cdots + q + 1$ em U. Assim, cada subespaço contém $q^{t+1} + \cdots + q + 1 - (q^t + \cdots + q + 1) = q^{t+1}$ pontos fora de U, de modo que está sendo contado q^{t+1} vezes.

Logo há $(q^d + \cdots + q^{t+1})/q^{t+1} = q^{d-t-1} + \cdots + q + 1$ subespaços de dimensão $t+1$ que contêm U.

Para espaços afins, a demonstração é a mesma. Só que temos que excluir de U e dos subespaços que contêm U os pontos no infinito, o que poderia excluir um subespaço inteiro. Mas como U não está contido no hiperplano no infinito, nenhum subespaço está. ∎

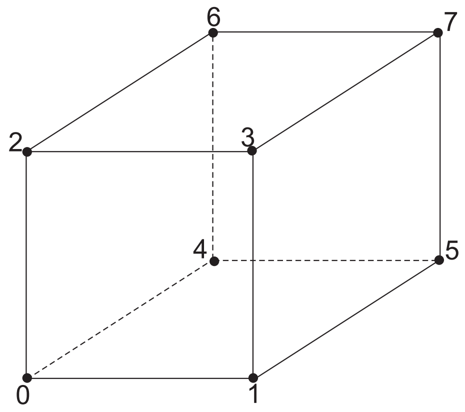
11. Códigos de Reed-Muller

Considere $A_d(Z/2Z)$. Numere os pontos de $A_d(Z/2Z)$ em qualquer ordem (você pode ler as d-uplas na base 2 e associar a esse número, se quiser), $P_1, P_2, \ldots, P_{2^d}$. Definimos o vetor característico de um subconjunto $\mathcal{M}$ de $A_d(Z/2Z)$ como $\chi(\mathcal{M}) = (a_1, a_2, \ldots, a_{2^d})$ com $a_i = 0$ se $P_i \notin \mathcal{M}$ e $a_i = 1$ se $P_i \in \mathcal{M}$. A partir de agora, não distinguiremos $\chi(\mathcal{M})$ e $\mathcal{M}$, ou seja, os subconjuntos de pontos de $A_d(Z/2Z)$ serão, para nós, vetores de $Z/2Z^{2^d}$.

Definição 11.1. *O código de Reed-Muller de ordem r é o código $C \subset Z/2Z^{2^d}$ gerado por todos os subespaços de dimensão $d-r$ de $A_d(Z/2Z)$.*

Exemplo 11.1.

Considere $d = 3$. Representamos $A_3(Z/2Z)$ a seguir.



Se $r = 1$, os vetores característicos dos 14 planos (por que são só 14, e não 15?) são

11110000	00001111
11001100	00110011
11000011	00111100
10101010	01010101
10100101	01011010
10011001	01100110
10010110	01101001

O subespaço gerado por esses vetores é

00000000	11111111
11110000	00001111
11001100	00110011
11000011	00111100
10101010	01010101
10100101	01011010
10011001	01100110
10010110	01101001

Se $r = 2$, basta tomar todas as retas, ou seja, todos os vetores característicos dos conjuntos de dois pontos como geradores.

Da mesma forma que fizemos com os códigos de Hamming, podemos descrever os códigos de Reed-Muller como diferenças simétricas de subespaços de dimensão $d - r$.

Veja que os códigos de Reed-Muller são claramente lineares. Assim, para achar a distância mínima de um código de Reed-Muller, basta calcular o peso mínimo desse código.

Teorema 11.1. *Seja C o código de Reed-Muller de ordem r. Então*

$$d(C) = 2^{d-r}$$

Um vetor de C tem peso 2^{d-r} se, e somente se, é o vetor característico de um subespaço de dimensão $d-r$ de $A_d(Z/2Z)$.

Demonstração

Observe que a caracterização geométrica de códigos de Hamming foi feita com uma indução. Usamos indução quando estamos estudando coisas sem muita estrutura ou difícil de caracterizar de modo simples, como retas em espaços projetivos. Nesse caso, temos subespaços de um espaço afim, que também não são muito estruturados e de caracterização simples, de modo que faremos indução sobre $d-r$.

Comecemos pela base de indução. Se $d-r=1 \iff 2^{d-r}=2$, basta provar que nenhum vetor de peso 1 pertence a C. Mas isso é óbvio, visto que os vetores característicos dos subespaços de dimensão 1 (ou seja, retas) têm peso 2 e a paridade de pesos é invariante $\mod 2$ quando somamos vetores. Isto é, o peso de todos os vetores de C é par. Como todo par de pontos de $A_d(Z/2Z)$ forma uma reta, a igualdade ocorre se, e somente se, consideramos uma reta.

Vamos ao passo de indução. Suponha que $d-r>1$ e que o teorema vale para todos os códigos de Reed-Muller de com $d'-r'<d-r$. Seja $\mathcal{M}$ um conjunto arbitrário de subespaços de dimensão $d-r$ de $A_d(Z/2Z)$ e $\mathcal{H}$ o conjunto definido pela diferença simétrica de todos os subespaços de $\mathcal{M}$. Temos que provar que $\mathcal{H}$ tem pelo menos 2^{d-r} pontos, com igualdade se, e somente se, é um subespaço de dimensão 2^{d-r}.

Suponha que $|\mathcal{H}| \le 2^{d-r}$. Mostremos que $\mathcal{H}$ é um subespaço de dimensão $d-r$. Faremos isso em quatro passos.

Como usar a hipótese de indução? Podemos pensar num espaço de dimensão menor que d, e nada mais natural que um subespaço de $A_d(Z/2Z)$.

Passo 1. *Seja U um subespaço de dimensão s de $A_d(Z/2Z)$, com $r+1 \le s \le d-1$. Então $\mathcal{H} \cap U$ é um vetor do código de Reed-Muller de ordem r de U. Pela hipótese de indução, sendo $s-r<d-r$, $\mathcal{H} \cap U$ é vazio ou contém pelo menos 2^{s-r} pontos, sendo a igualdade ocorre se, e somente se, $\mathcal{H} \cap U$ é um subespaço de dimensão $s-r$ de U.*

Observando que $(A \Delta B) \cap C = (A \cap C) \Delta (B \cap C)$, vemos que $\mathcal{H} \cap U$ é a diferença simétrica das interseções de U e dos subespaços de $\mathcal{M}$. Seja então $M \in \mathcal{M}$ e consideremos $M \cap U$. Se $M \cap U = \emptyset$, não há problemas, já que $A \, \Delta \, \emptyset = A$. Suponha, então, $M \cap U \neq \emptyset$. Assim, $M \cap U$ é um subespaço de $A_d(Z/2Z)$.

Para achar a dimensão de $M \cap U$, basta ver que M é o conjunto verdade de um sistema de r equações homogêneas, tirando os pontos no infinito e U é o conjunto verdade de um sistema de $d-s$ equações homogêneas, tirando os pontos no infinito. Assim, $M \cap U$ é o conjunto verdade de um sistema de $r+d-s = d-(s-r)$ equações homogêneas, tirando os pontos do infinito. Como a matriz desse sistema tem posto no máximo $d-(s-r)$, o número de variáveis arbitrárias é pelo menos $d+1-(d-(s-r)) = s-r+1$. Assim, a dimensão de $M \cap U$ é pelo menos $s-r$.

Como podemos escrever um subespaço de dimensão maior que $s-r$ como união disjunta de subespaços paralelos de dimensão $s-r$ (tente provar isso, tendo em vista a construção de espaços projetivos a partir de seus pontos!), o vetor característico de $M \cap U$ pertence ao código de Reed-Muller de ordem r em U. Sendo esse código linear, $\mathcal{H} \cap U$ é um vetor desse código.

Observemos que, do teorema 10.2., se considerarmos um subespaço U_r de dimensão r, há $2^{d-r-1}+\cdots+2+1 = 2^{d-r}-1$ subespaços de dimensão $r+1$ que contêm U_r, um número bem próximo de 2^{d-r}. Vamos usar isso a nosso favor.

Primeiro, vamos conseguir um subespaço U_r interessante. Mas não parece ser fácil encontrar um subespaço assim diretamente, de modo que vamos usar indução.

Passo 2. *Seja P um ponto de $\mathcal{H}$. Então, para $i \in \{0, 1, \ldots, r\}$ existe um subespaço U_i, de dimensão i, cuja interseção com $\mathcal{H}$ consiste apenas de P.*

Isso é claro para $i=0$ (U_0 é um ponto!). Suponha que $i \ge 1$ e que a afirmação é verdadeira para $i-1$. Considere o subespaço U_{i-1}. $\mathcal{H}$ corta esse subespaço em apenas um ponto. Vamos mostrar que existe um subespaço de dimensão i que corta $\mathcal{H}$ em só um ponto. Seja $\mathcal{S}_i$ a família de subespaços de dimensão i que contém U_{i-1}. Temos $|\mathcal{S}_i| = 2^{d-i+1} - 1$. Se todos os subespaços de $\mathcal{S}_i$ cortam $\mathcal{H}$ em mais de um ponto (veja que $P \in U_{i-1}$ pertence a todos tais subespaços), $\mathcal{H}$ contém um ponto de cada subespaço fora de U_{i-1}. Como dois subespaços de $\mathcal{S}_i$ têm como interseção somente U_{i-1}, $\mathcal{H}$ contém pelo menos $2^{d-i+1} - 1 + 1 = 2^{d-i+1}$ pontos (um de cada subespaço de $\mathcal{S}_i$ e P). Mas $2^{d-i+1} > 2^{d-r}$, absurdo. Logo existe U_i.

Agora, usaremos U_r a nosso favor.

Passo 3. $|\mathcal{H}| = 2^{d-r}$.

Considere $U = U_r$. Aplicando o passo 1 a um subespaço V de dimensão $r+1$ que contém U_r, temos que $\mathcal{H} \cap V$ é uma palavra de um código de Reed-Muller, que tem peso pelo menos 2. Assim, $\mathcal{H} \cap V$ tem pelo menos dois elementos, um de U_r e outro fora de U_r. Assim, utilizando um argumento análogo ao passo anterior e sabendo que há $2^{d-r} - 1$ subespaços de dimensão $r+1$ que contêm U_r, concluímos que $|\mathcal{H}| \geq 2^{d-r} - 1 + 1 = 2^{d-r}$. Logo, como supomos $|\mathcal{H}| \leq 2^{d-r}$, o resultado segue.

Agora que conseguimos uma parte do teorema, precisamos provar a outra.

Passo 4. *Os pontos de $\mathcal{H}$ formam um espaço afim de dimensão $d-r$.*

Observe que, do passo 3, todo subespaço de dimensão $r+1$ que contém U_r, um subespaço que corta $\mathcal{H}$ em um único ponto, corta $\mathcal{H}$ em dois pontos, que formam por si só um espaço afim de dimensão 1. Além disso, o próprio U_r corta $\mathcal{H}$ em um único ponto, que é um subespaço de dimensão zero. Assim, parece razoável tentarmos provar o seguinte fato:

Vamos dizer que um subespaço de dimensão r é bom quando corta $\mathcal{H}$ em exatamente um ponto. Então se W_s é um subespaço de dimensão s, $r \leq s \leq d$, que contém um subespaço bom então $W_s \cap \mathcal{H}$ é um subespaço de dimensão $s-r$.

O caso d prova o passo 4, pois $W_d \cap \mathcal{H}$ seria um subespaço de dimensão $d-r$, com 2^{d-r} pontos. Como $\mathcal{H}$ tem essa mesma quantidade de pontos, $W_d \cap \mathcal{H} = \mathcal{H}$.

Já provamos acima os casos $s = r$ e $s = r+1$. Vamos tentar encaixar (de novo!) uma indução.

Suponha que $r+1 \leq s \leq d$ e assuma que o resultado é válido para $s-1$ e s. Provaremos que vale também para $s+1$.

Considere um subespaço W que contém um subespaço bom G. Seja W_{s-1} um subespaço, de dimensão $s-1$, de W que contém G. Pela hipótese de indução, $X = \mathcal{H} \cap W_{s-1}$ é um subespaço de dimensão $s-1-r$. Há exatamente três subespaços s-dimensionais U_0, U_1, U_2 que passam por W_{s-1}. Logo, de novo pela hipótese de indução, $X_i = \mathcal{H} \cap U_i$, $i = 0, 1, 2$, são três subespaços de dimensão $s-r$. Veja que como $U_i \supset W_{s-1} \implies \mathcal{H} \cap U_i \supset \mathcal{H} \cap W_{s-1}$, os X_i's contêm X.

Suponha que $X_0 = X_1$. Considere as bases de X_0 e X_1. Temos $s-r+1$ geradores tanto para X_0 como para X_1. Para completar para U_0 e U_1, faltam r geradores. Porém, considere $X \subset W_{s-1}$. X tem dimensão $s-1-r$ e W_{s-1} tem dimensão $s-1$. Assim, para completar uma base de X para uma base de W_{s-1} faltam r geradores, que são os mesmos que completam X_0 e X_1 para U_0 e U_1, respectivamente. Logo se $X_0 = X_1$ então $U_0 = U_1$, absurdo. Portanto os três subespaços X_0, X_1 e X_2 são distintos, de modo que $X_0 \cap X_1 = X_0 \cap X_2 = X_1 \cap X_2 = X_0 \cap X_1 \cap X_2 = X$.

Considere o subespaço de dimensão $s-r+1$ que é gerado por X_0 e X_1. Como X_0 e X_1 estão contidos em W, esse subespaço também está contido em W. Esse subespaço contém exatamente três subespaços de dimensão $s-r$ que passam por X, que só podem ser X_0, X_1 e X_2. Logo X_2 está contido nesse subespaço e portanto X_0, X_1 e X_2 estão contidos num espaço de dimensão $s+1-r$. Seja Y esse subespaço. Podemos concluir que $Y \subset \mathcal{H}$, pois de $|X_0 \cup X_1 \cup X_2| = |X_0| + |X_1| + |X_2| - (|X_0 \cap X_1| + |X_0 \cap X_2| + |X_1 \cap X_2|) + |X_0 \cap X_1 \cap X_2| = 3 \cdot 2^{s-r} - 3 \cdot 2^{s-r-1} + 2^{s-r-1} = 2^{s+1-r}$ e $|Y| = 2^{s+1-r}$ concluímos que $X_0 \cup X_1 \cup X_2 = Y$. Assim, Y seria nosso tão sonhado subespaço de dimensão $s+1-r$ contido em $\mathcal{H}$ e conseguimos nosso passo de indução e o teorema. ∎

11.1. Aplicações dos códigos de Reed-Muller

Os códigos de Reed-Muller foram protagonistas de uma das aplicações mais importantes da Teoria dos Códigos: eles foram utilizados para codificar fotos enviadas de satélites para a Terra.

A meta da missão *Mariner 9* em 1971 era sobrevoar Marte e fotografar sua superfície inteira. Essas fotos seriam enviadas para a Terra e é claro que ao transmitirmos esses dados por essa gigantesca distância muitos erros devem ocorrer. Assim, esses dados deveriam ser codificados por um código muito bom, pois caso contrário todas as imagens tiradas pelos sofisticados equipamentos no satélite continuariam invisíveis para nós na Terra.

Cada foto tinha uma resolução de 700×832 pixels, sendo que cada pixel era uma 8-upla que representava um tom de cinza (sim, as fotos era em preto, branco e muito cinza!).

Esses dados foram divididos em blocos de 6 bits cada. Cada bloco foi codificado por um vetor de peso 32 (isto é, há 26 bits redundantes para corrigir erros). Para isso, foi utilizado um código de Reed-Muller de ordem 1 e cujos vetores tinham tamanho 64 (gerado pelos hiperplanos de $A_6(Z/2Z)$). Esse código é 7-corretor de erros.

12. Referências Bibliográficas

[1] Albrecht Beutelspacher, Ute Rosenbaum. *Projective Geometry*, Cambridge Press. Um livro muito bom sobre geometria projetiva, todavia com um contexto mais combinatório que propriamente geométrico (de fato, os autores atuam em Computação e Segurança de Informação, respectivamente, e o livro mostra aplicações em códigos e criptologia). Todavia, lá estão demonstrações de alguns dos principais teoremas, como Desargues e Pappus (que vale em espaços projetivos baseados em anéis de divisão se, e somente se, o anel de divisão é um corpo). Lá conheci o conceito de *Geometria Abstrata*. Quase todo o artigo está baseado no capítulo 5 deste livro.

[2] Carlos Shine. *Aplicações de Planos Projetivos em Teoria dos Números e Combinatória*, in: Revista Eureka! 15. Eu escrevi esse artigo para dar como aula na Semana Olímpica de 2003, em Goiânia. Foram duas aulas de 100 minutos, uma sobre teoria dos números e outra sobre combinatória. Eu gostei bastante de ambas as aulas, sobretudo a que trata de combinatória. O engraçado é que, na época que escrevi o artigo, ainda não tinha ganhado o livro [1] (veja as referências desse artigo!). Aliás, agradeço ao prof. Edmilson Motta pelas sugestões para aquele artigo e pelo livro. Considere também como referências bibliográficas as referências desse artigo, sobretudo o artigo do prof. Luciano Castro.

[3] J. I. Hall. *Notes on Coding Theory*. Este livro está on-line, na página

`http://www.mth.msu.edu/~jhall/classes/codenotes/coding-notes.html`

A discussão sobre generalizações e aplicações de códigos de Hamming foi retirada de lá.

[4] Peter J. Cameron. *Projective and Polar Spaces*. Alguns exercícios e as citações referentes aos códigos de Golay foram retiradas de lá. Também está disponível na Internet:

`http://www.maths.qmol.ac.uk/~pjc/pps`

[5] I. N. Herstein. *Topics in Algebra*, Wiley. Um livro de Álgebra Abstrata para quem quer aprender o básico (e saber o que são quatérnions!) e mais um pouco dessa fascinante e importantíssima área da Matemática.

[6] R. Chapman. *Constructions of the Golay Codes: A Survey*. Lá há várias construções do código de Golay estendido G_{24}. Só que para entender direitinho precisa saber um pouco de Álgebra Abstrata e bastante paciência. Veja isso em

`http://www.maths.ex.ac.uk/~rjc/etc/golay.pdf`

e mais em

`http://www.maths.ex.ac.uk/~rjc/rjc.html`