CONSTRUÇÃO DOS CÓDIGOS (2^M, M+1, 2^{M-1}), COM $M \geq 3$, USANDO BLOCOS LEXICOGRÁFICOS

Francisco Antonio de ALENCAR MENEZES[1]
Elvio César GIRAUDO[2]

- RESUMO: Conceitos relacionados à teoria da codificação linear são apresentados para fundamentar a nova construção de códigos lexicográficos e propor para o código (8, 4, 4), uma construção por matrizes ou por bloco lexicográfico. Na construção, a partir da matriz geradora, suas palavras código estão a uma mesma distância de Hamming e a uma mesma distância euclidiana, dois a dois. Como propriedade fundamental, a matriz de paridade do código é a transposta de sua matriz geradora, o que sugere um procedimento de decodificação diferenciado do código de Hamming. Nesse contexto é definida a família de códigos de parâmetros (2^M, M+1, 2^{M-1}), com $M \geq 3$. Curva de taxa de erro de bit é apresentada para este código num sistema de modulação digital PSK.

- PALAVRAS-CHAVE: Código lexicográfico; código linear; código corretor de erro.

1 Introdução

Os fundamentos matemáticos para a comunicação digital foram estabelecidos por Shannon (1948) que formulou o problema básico da transmissão de informação confiável, usando modelos probabilísticos para fontes e canais de comunicação. Baseado em tal formulação, adotou uma medida para o conteúdo de uma fonte de informação e estabeleceu limites básicos sobre uma taxa máxima, que uma informação digital pode ser transmitida confiavelmente, através de um canal de comunicação. Os limites teóricos deduzidos por pesquisadores como Levenshtein (1960), Van Lint (1962), McWilliams e Sloane (1977), Conway (1990) e Calderbank (1995), são referências na fundamentação de projetos e desenvolvimento de sistemas de comunicação digital mais eficientes. Os códigos corretores de erros são utilizados sempre que se deseja transmitir ou armazenar dados. São exemplos, todas as comunicações via satélite, as comunicações internas de um computador e o armazenamento de dados. A classe de códigos mais utilizada é a dos códigos lineares, bem fundamentada por Lin e Costello Jr (1983) e como exemplo pode-se citar, segundo Hefez e Vilela (2002), o código de Hamming, o código de Golay e o código de Reed-Muller.

1 Centro de Ciências Exatas e Tecnologia, Coordenação de Ciências-Matemática, Universidade Estadual Vale do Acaraú, CEP: 62.040-370, Sobral, Ceará, Brasil, E-mail: *alencarmenezes@gmail.com*
2 Departamento de Engenharia de Teleinformática, Centro de Tecnologia, Universidade Federal do Ceará – UFCE, Caixa Postal 6007, CEP: 60.455-760, Fortaleza, Ceará, Brasil, E-mail: *elvio@deti.ufc.br*

O estudo da construção dos códigos lexicográficos, comprovadamente lineares e de ótimo desempenho, com o conceito matemático de lexicografia e a definição da lexicografia de códigos, com Conway (1990), Herscovici (1991) e Trachtenberg (2002) é um ponto inicial na descoberta de um novo método de construção denominada construção por bloco lexicográfico, fundamentada a partir do código (8,4,4), objeto deste trabalho.

Na Seção 2, são apresentadas as seqüências lexicográficas de vetores segundo Brualdi e Pless (1993). Na Seção 3, abordam-se brevemente as características da matriz geradora. Já na Seção 4, o conceito de lexicografia de códigos segundo Conway (1990) e Herscovici (1991) e os códigos lexicográficos de Trachtenberg (2002) definem uma construção lexicográfica baseada na matriz geradora. Esta serve de fundamentação para uma nova construção denominada de construção por bloco lexicográfico que produz uma família de códigos de parâmetros $(2^M, M+1, 2^{M-1})$, tal que, $M \geq 3$. Para $M = 3$ é definido o código (8, 4, 4) cujo desempenho é analisado na Seção 5, comparando a taxa de erro de bit (*BER*) em função da relação em decibéis entre a energia de bit e a densidade espectral de potência do ruído E_b/N_0 para o sistema PSK codificado e não codificado, mediante simulação no ambiente computacional do software livre SCILAB, permitindo conclusões e perspectivas.

2 Sequência lexicográfica de vetores

Vetores ordenados por critérios convenientemente propostos podem sugerir diferentes lexicografias. No contexto do GF(2) considera-se dois vetores, em sequência, $(x_n) = (x_1, x_2, ..., x_n)$ e $(y_n) = (y_1, y_2, ..., y_n)$, tais que:

$$(y_n) > (x_n) \Leftrightarrow \exists (y_j) > (x_j) : y_k = x_k \forall\ k < j\,. \tag{1}$$

A Tabela 1 mostra duas seqüências lexicográficas. A seqüência 2 é a concepção lexicográfica de Brualdi e Pless (1993), que consideram um alfabeto infinito {0, 1, 2, …}, representado por sequências em ordem. Na sequência 1, tem-se a lexicografia considerada neste trabalho, definida na Expressão (1), que define também a sequência 2, considerando $y_k = x_k \forall\ k > j$.

Tabela 1 - Vetores em ordem lexicográfica

Sequência 1	Sequência 2
(0...000)	(000...0)
(0...001)	(100...0)
(0...010)	(010...0)
(0...011)	(110...0)
(0...100)	(001...0)
(0...101)	(101...0)
(0...110)	(011...0)
(0...111)	(111...0)
⋮	⋮

3 Matriz geradora de um código linear

Definição 1. *Sejam K um corpo finito com q elementos e* $C(n,k,d) \subset K^n$ *um código linear, onde k é a dimensão de C, d é a distância mínima de Hamming sobre K e* $\beta = (v_1, v_2, \cdots, v_k)$ *uma base ordenada de C. A matriz G, cujas linhas são os vetores* $v = (v_{i_1}, v_{i_2}, \cdots, v_{i_k})$, $i = 1:k$, *tal que,*

$$G = \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{pmatrix}, \tag{2}$$

é denominada matriz geradora do código C associada à base β.

Em particular, as linhas de uma matriz geradora são linearmente independentes e a menor distância de Hamming entre elas é denominada distância mínima do código gerado, que serve de parâmetro para medição da capacidade de correção de erro do código.

Os códigos apresentados no contexto corrigem (*d*-1)/2 erros e nosso objetivo é construir matrizes geradoras lexicograficamente, por blocos, para geração de códigos com distâncias mínimas cada vez maiores, o que possibilita a correção de uma quantidade maior de erros.

4 Construção lexicográfica do código (2^M, M+1, 2^{M-1}) com $M \geq 3$

O código de comprimento n = 8, dimensão k = 4 e distância mínima de Hamming $d_{\min}$ = 4, mencionado por Trachtenberg (2002), é exemplo de código ótimo e a sua construção lexicográfica é dada a partir de sua matriz geradora *T*, na Equação (3).

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3}$$

Esta construção escolhe vetores linearmente independentes em ordem lexicográfica obedecendo à distância mínima de Hamming requerida pelo código. Observa-se que nessa matriz, a distância máxima é d_{max}= 8, diferente da distância mínima de Hamming $d_{\min}$ = 4, do código. Sendo assim, pode-se dizer que os vetores escolhidos lexicograficamente, que compõem *T*, diferem na distância de Hamming dois a dois.

Definição 2. *Um bloco lexicográfico é uma matriz geradora de um código com distância mínima de Hamming d =* 2^{M-1}*, tal que, M* ≥ *3, construída lexicograficamente a partir do vetor* (0^d | 1^d) = (00...0011...11), *de d zeros e d uns, onde o peso de Hamming e a distância euclidiana entre os seus vetores linha, tomados dois a dois, são iguais a d.*

Para M = 3 tem-se o código (2^M, M+1, 2^{M-1}) = (8, 4, 4) donde sua matriz geradora *H*, exibida na Equação (4), é construída a partir do vetor (0^4 | 1^4) = (00001111), mediante a

escolha de vetores linearmente independentes, considerando uma lexicografia onde possuem a mesma distância de Hamming, dois a dois. Como propriedade fundamental tem-se que a distância máxima da matriz H é igual à sua distância mínima d. Este fato não acontece para a construção lexicográfica da matriz T na Equação (3). Além disso, essa nova construção lexicográfica permite que todos os vetores geradores da matriz H, se encontrem dois a dois, à mesma distância euclidiana.

$$H = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

Cada bloco lexicográfico que dá origem a uma matriz geradora, dá origem a outro bloco lexicográfico que é matriz geradora para outro código em sequência. A matriz B da Equação (5) é o bloco lexicográfico que dá origem a H, na Equação (6).

$$B = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{5}$$

Nesse caso tem-se que H, na Equação (6), é um bloco lexicográfico e dessa forma pode-se

$$H = \begin{pmatrix} 0^d & 1^d \\ B & B \end{pmatrix}. \tag{6}$$

obter uma nova matriz geradora N, como definida na Equação (7), do código $C(16, 5, 8)$,

$$N = \begin{pmatrix} 0^d & 1^d \\ H & H \end{pmatrix}, \tag{7}$$

exibida na Equação (8).

$$N = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{8}$$

As matrizes B, H e N formam uma sequência lexicográfica de matrizes. Como obedecem a um método de construção bem definido, possuem as mesmas características e propriedades fundamentais, tal que, quando multiplicadas pela sua transposta, retornam a matriz nula, ou seja, $BB^{\mathrm{T}} = 0$, $HH^{\mathrm{T}} = 0$ e $NN^{\mathrm{T}} = 0$. Por definição, tem-se que H^{T} é a matriz cheque de paridade do código $C(8, 4, 4)$. Essa propriedade é uma virtude dos blocos lexicográficos que, além disso, devido à sua construção, apresentam-se irredutíveis

à forma padrão, o que sugere um modelo de decodificação diferente do código de Hamming.

5 Simulação para a nova construção lexicográfica

Na Figura 1, mostra-se o diagrama em blocos de um sistema de comunicação digital, onde se destacam, entre outros, os blocos do canal de transmissão, codificador de canal e modulador, como em Lin e Costello Jr (1983). Nesse contexto, se compara o desempenho deste sistema de comunicação com modulação digital PSK, sem codificação e com codificação de canal pelo código (8, 4, 4). O canal de transmissão usado nesta simulação é com ruído aditivo branco e gaussiano. Todo o sistema foi programado no ambiente computacional SciLab.

Como resultado desta comparação, mostra-se na Figura 2, as curvas de *BER* em função da E_b/N_0[dB] para o sistema codificado e não codificado. Pode-se observar que tal sistema codificado com o código (8, 4, 4), apresenta, por exemplo, para potências aproximadamente E_b/N_0[dB] > 5dB, menores taxa de erro de bit que o sistema não codificado.

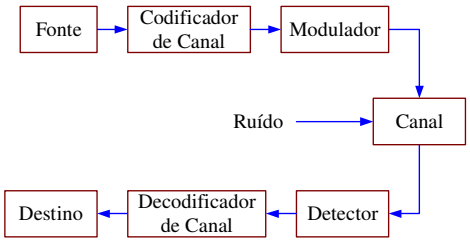

Figura 1 - Diagrama em Blocos de um Sistema de Comunicações Digitais.

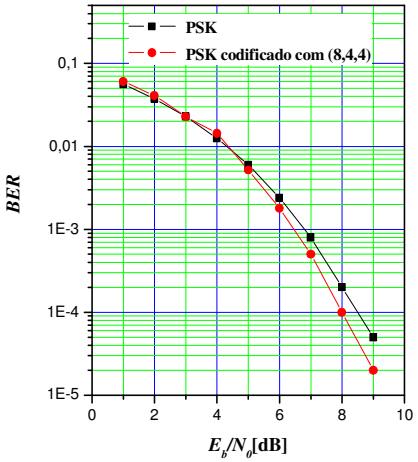

Figura 2 - Desempenho do sistema PSK codificado e não codificado usando o código (8, 4, 4).

Conclusões

Conceitos relacionados à teoria da codificação linear foram apresentados para fundamentar uma nova construção para o código (8, 4, 4), no contexto dos corpos de Galois de característica 2, $GF(2^m)$, denominada construção por bloco lexicográfico.

O código lexicográfico (8, 4, 4) aqui é definido a partir de sua matriz geradora *H*, construída a partir da matriz *B*, denominada bloco lexicográfico. Note-se que *H*, também

um bloco lexicográfico, é uma matriz cujos vetores estão a uma mesma distância de Hamming e a uma mesma distância euclidiana, dois a dois. Como propriedade fundamental, tem-se que $HH^t = 0$ e nesse caso, H^t é a matriz cheque de paridade do código. Deve-se ressaltar que o código apresentado não é o de Hamming. Nesse contexto apresenta-se, por simulação computacional, o desempenho do código (8, 4, 4) mostrando a taxa de erro de bit, *BER*, como função da energia de bit sobre a densidade espectral de potência do ruído E_b/N_0[dB] para o sistema PSK codificado e não codificado. Como resultado mostra-se na Fig.2, a diminuição da *BER* com o aumento da relação E_b/N_0 do sistema PSK codificado em relação ao não codificado. Observa-se que para uma dada probabilidade de erro, o sistema codificado com a nova construção lexicográfica, precisa menos potência de transmissão que o sistema não codificado.

Uma direção futura para a família de códigos lexicográficos (2^M, $M+1$, 2^{M-1}), tal que, $M \geq 3$, é a definição desses códigos sobre corpos de Galois com característica diferente de 2.

ALENCAR MENEZES, F. A. de; GIRAUDO, E. C. Construction of the codes (2^M, M+1, 2^{M-1}), with M ≥ 3, using lexicographical blocks. *Rev. Bras. Biom.*, São Paulo, v.25, n.4, p.157-163, 2007.

- *ABSTRACT: Concepts related to the theory of the linear codification are presented to the new construction of lexicographical codes, and to propose for the code (8, 4, 4), a lexicographic construction for matrices or for lexicographical blocks. The construction starts from the generating matrix, whose code words are at same Hamming distance and the same Euclidean distance, two by two. A fundamental property is that the code's parity check matrix is the transposed one of its generating matrix, what hints to the Hamming code differentiated decode procedure. In this context the family of codes of parameters is defined (2^M, $M+1$, 2^{M-1}), with $M \geq 3$. Curves of bit error rate are present for this code in a PSK digital modulation system.*

- *KEYWORDS: Lexicographic code; linear code; error correction code.*

Referências

BRUALDI, R. A.; PLESS, V. Greedy codes. *J. Comb. Theory Ser. A*, New York, v.64, p.10–30, 1993.

CALDERBANK, A. R. (Ed.) *Different aspects of coding theory*: Am. Math. Soc. (Short Course). In: SYMPOSIAN IN APPLIED MATHEMATICS, San Francisco, 1995. *Procidings...* San Francisco, 1995. 239p.

CONWAY, J. H. Integral lexicographic codes. *Discrete Math.*, Amsterdam, v.83, p.219–235, 1990.

HEFEZ, A.; VILELA, M. L. T. *Códigos corretores de erros*. Rio de Janeiro: IMPA, 2002. 206p.

HERSCOVICI, D. S. Minimal distance lexicographic codes over an infinite alphabet. *IEEE Trans. Inform. Theory*, New York, v.37, n.5, p.1366-1368, 1991.

LEVENSHTEIN, V. I. A class of systematic codes. *Soviet Math. Dokl.*, Providence, v.1, n.1, p.368-371, 1960.

LIN, S.; COSTELLO Jr., D. *Error control coding*. Prentice Hall, 1983. 1260p.

MACWILLIAMS, F. J.; SLOANE, N. J. A. *The theory of error correcting codes*. North-Holland, Amsterdam, 1978. 796p.

SCILAB. copyright © 1989-2007. INRIA. Disponível em: <http://www.scilab.org/>. Acesso em: jul. 2007.

SHANNON, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, New York, v.27, p.379–423, 623–656, 1948.

TRACHTENBERG, A. Designing lexicographic codes with a given trellis complexity. *IEEE Trans. Inform. Theory*, New York, v.48, n.1, p.89-100, 2002.

VAN LINT, J. H. *Introduction to coding theory*. Berlin: Springer, 1992. (Graduate Texts in Mathematics). 256p.

Recebido em 18.09.2007.

Aprovado após revisão em 15.02.2008.