ELSEVIER

Discrete Mathematics 264 (2003) 37–43

DISCRETE MATHEMATICS

www.elsevier.com/locate/disc

Greedy loop transversal codes for correcting error bursts

Dug-Hwan Choi[a,*], Jonathan D.H. Smith[b]

[a] *Authentication Technology Research Center, Sungkyunkwan University, Suwon, Kyunggi-do 440-746, South Korea*

[b] *Department of Mathematics, Iowa State University, Ames, IA 50011, USA*

Received 17 March 2001; received in revised form 6 September 2001; accepted 22 February 2002

Abstract

The greedy loop transversal algorithm is used to construct linear codes in a binary channel for the correction of error bursts. The dimensions of the codes constructed are compared with the dimensions of the corresponding white noise greedy loop transversal codes, and with the dimensions of the few previously best-known codes for the burst-error patterns. The dimensions of the greedy loop transversal codes for burst-error correction match or exceed the dimensions of these previously known codes.

MSC: 94B05

Keywords: Loop transversal code; Greedy algorithm; Burst error; Lexicode

1. Introduction

Loop transversal codes were introduced in [5]. They are also described in [6,7]. In the loop transversal approach to the construction of linear error-correcting codes, attention focusses on the set of errors to be corrected. This set of errors does not have to correspond to "white noise". The idea is to specify a loop structure (abstractly an abelian group) on the set of errors as a loop transversal to the linear code as a subgroup of the channel. A greedy algorithm for specifying this loop structure, and thus for the construction of loop transversal codes, was discussed in [3,4]. In [4], it is

* Corresponding author.

E-mail address: dchoi@dosan.skku.ac.kr (D.-H. Choi).

PII: S0012-365X(02)00548-4

shown that for binary channels, the codes constructed by the greedy loop transversal algorithm coincide with the "lexicodes" of Conway and Sloane [2]. However, for good channels, the set of errors to be corrected is much smaller than the set of codewords, so the greedy loop transversal algorithm is more efficient than the lexicode algorithm. For non-binary channels, the lexicodes become non-linear, while the loop transversal codes remain linear. In [3], the greedy loop transversal algorithm was used to construct linear codes in binary and ternary channels for the correction of "white noise" errors. The codes obtained, including the binary and ternary Golay codes, were always within a dimension or two of the best known linear codes. The current paper may be seen as a counterpart of [3], but dealing with burst-errors rather than "white noise". The greedy loop transversal algorithm is used to construct linear codes in a binary channel for the correction of single errors and bursts of length 2, alone, or in conjunction with a single error or another burst of length 2. The dimensions of the codes constructed are compared with the dimensions of the corresponding white noise greedy loop transversal codes from [3], and with the dimensions of the few previously best known codes for these error patterns from [9]. The dimensions of the greedy loop transversal codes for burst-error correction match or exceed the dimensions of these previously known codes.

2. Loop transversal codes

A *transversal* T to a subgroup C of a group $(V,+,0)$ is a subset of V with $V=\dot{\bigcup}_{(t\in T)}(C+t)$. Thus each element v can be expressed uniquely as $v=v\delta+v\varepsilon$ with $v\delta\in C$ and $v\varepsilon\in T$. A received word v is decoded to a codeword $v\delta$ with the error $v\varepsilon$. A binary operation $*$ is defined on T by

$$t*u=(t+u)\varepsilon. \tag{1}$$

For any t,u in T, the equation $x*t=u$ has a unique solution x. If the equation $t*y=u$ also has a unique solution, then T is called a *loop transversal*. Equivalently, the algebra $(T,*,0\varepsilon)$ is a loop. If V is abelian, then each transversal is a loop transversal, and the loop $(T,*,0\varepsilon)$ is an abelian group. For u_i in T, it is convenient to use the notation $\prod_{i=1}^{0}u_i=0\varepsilon$ and $\prod_{i=1}^{r}u_i=[\prod_{i=1}^{r-1}u_i]*u_r$ for $r>0$. In compound expressions, $*$ and $\prod$ will bind more strongly than $+$ and $\sum$.

Now specialize to the usual coding theory case that V is a finite-dimensional vector space over a field F. Define $\lambda\times t=(\lambda t)\varepsilon$ for λ in F and t in T. This makes $(T,*,F)$ a vector space over F. Induction on r extends Eq. (1) to

$$\left(\sum_{i=1}^{r}\lambda_i t_i\right)\varepsilon=\prod_{i=1}^{r}(\lambda_i\times t_i) \tag{2}$$

for t_i in T. Assume that T contains a basis $\{e_1,\ldots,e_n\}$ for V, e.g. $V=F^n$ and each e_i has 1 in the ith place as its only non-zero coordinate. Then the knowledge of the vector space $(T,*,F)$ is sufficient to determine the code C. Indeed, by Eq. (2), if

$v = \sum_{i=1}^{n} \lambda_i e_i$,

$$C = \{v\delta \mid v \in V\} = \{v - v\varepsilon \mid v \in V\} = \left\{ \sum_{i=1}^{r} \lambda_i e_i - \prod_{i=1}^{r} (\lambda_i \times e_i) \mid \lambda_i \in F \right\}.$$

As an abstract vector space, the transversal $(T, *, F)$ is isomorphic to the dual of the code C. Knowing small parts of the transversal $(T, *, F)$ is sufficient to identify specific codewords; conversely knowing specific codewords determines part of the structure of $(T, *, F)$. In particular, for t_i in T,

$$\sum t_i - \prod t_i \in C. \tag{3}$$

Duality (3) between C and T is such that small parts of C determine small parts of T, and vice versa. The relationship is described as *local duality*.

Here is another way of passing between the code and the transversal. By the fact that $(x + y)\varepsilon = x\varepsilon * y\varepsilon$ and $(\lambda x)\varepsilon = \lambda \times (x\varepsilon)$ for $\lambda \in F$ and $x, y \in V$, the *parity map* $(V, +, F) \to (T, *, F)$ is a linear transformation. The parity-check matrices can be given by matrices of ε with respect to appropriate bases. Any complete set of coset leaders for a given code C yields the underlying set T of a loop transversal.

3. The greedy loop transversal algorithm

Each natural number n (including 0) has a unique expansion $n = \sum_{i=0}^{\infty} n(i)2^i$ with $n(i) \in \mathrm{GF}(2)$ for each i, where GF(2) is the Galois field of order 2. Moreover, $n(i) = 0$ for $i > \lfloor \log_2 n \rfloor$. The set $\mathbb{N}$ of natural numbers is the nested union $\bigcup_{(d>0)} V_d$ of vector spaces V_d, where $V_d = \mathrm{GF}(2)^d$ is the d-dimensional vector space over the Galois field GF(2). The set $\{2^i \mid 0 \leqslant i \leqslant d-1\}$ is a basis for V_d. The operation $+_2$ or $\sum$ is the *nim sum* of [1, p. 51], which corresponds to the exclusive-or operation.

The set $\mathbb{N}$ of natural numbers is ordered by the lexicographic ordering $\subseteq_2$ on the binary expansions of its members. A subset X of a poset $(Y, \sqsubseteq)$ is said to be *self-subordinate* if $y \sqsubseteq x \in X$ implies $y \in X$. A self-subordinate subset E of $(\mathbb{N}, \subseteq_2)$ is called an *error pattern* if it contains the set $2^{\mathbb{N}} = \{2^i \mid i \in \mathbb{N}\}$. Error patterns model sets of possible errors to be corrected in the various channels V_d. For example,

$$B_2 = \{a2^i +_2 b2^j \mid a, b \in \mathrm{GF}(2); i, j \in \mathbb{N}, |i - j| \leqslant 1\}$$

is the error pattern describing burst errors of length at most 2 [5, (3.5)]. Also,

$$B_2S = \{a2^i +_2 b2^j +_2 c2^k \mid a, b, c \in \mathrm{GF}(2); i, j, k \in \mathbb{N}, i < j < k,$$
$$\text{either } |i - j| \leqslant 1 \text{ or } |j - k| \leqslant 1 \text{ with } |i - k| \geqslant 3\}$$

is one describing burst errors of length at most 2 together with at most one single error. The error pattern for two burst errors of length at most 2 [8] is described by

$$2B_2 = \{a2^i +_2 b2^j +_2 c2^k +_2 d2^\ell \mid a, b, c, d \in \mathbb{N}, i < j < k < \ell,$$
$$|i - j| \leqslant 1, |k - \ell| \leqslant 1, |j - k| \geqslant 2\}.$$

Error patterns form partial algebras under the operations of the vector space $(\mathbb{N}, +_2, \mathrm{GF}(2))$.

Suppose that an error pattern E is given. Then an *E-syndrome*, or just *syndrome*, is a partial function $s: E \to \mathbb{N}$ which

$$\begin{array}{ll} \text{(a) injects;} & \\ \text{(b) is a partial vector space homomorphism;} & \\ \text{(c) has domain self-subordinate in } (E, \leqslant), \text{and} & \\ \text{(d) satisfies: } \forall n \in \mathbb{N},\ \exists r \in \mathbb{N} \cdot 2^{\mathbb{N}} \cap s(V_n \cap E) \text{ spans } V_r. & \end{array} \tag{4}$$

The syndrome is said to be *proper* if s is a properly partial function. In view of (c), this is equivalent to finiteness of the domain of s. For a proper syndrome, the *length* is defined to be

$$n = \max\{1 + \lfloor \log_2 m \rfloor \mid m \in \mathrm{dom}\, s\}.$$

The *redundancy* is defined to be

$$r = \max\{1 + \lfloor \log_2(ms) \rfloor \mid m \in \mathrm{dom}\, s\}.$$

A proper syndrome s defines a *parity map*

$$\varepsilon_s : V_n \to V_r$$

by linearity and $2^i \varepsilon_s = 2^i s$ for $i < n$. By (c), these values $2^i s$ are defined. Condition (b) guarantees that s agrees with ε_s on $V_n \cap E$. Condition (d) yields that ε_s is surjective. Condition (a) guarantees that dom s embeds into V_r under ε_s. A code C_n in the channel V_n correcting the set $V_n \cap E$ of errors, and having dimension $n - r$, is then given as the kernel of ε_s.

The greedy loop transversal algorithm determines an E-syndrome s by the partial linearity (b) in (4) and the greedy choice of $2^n s$ given that $s: (V_n \cap E) \to \mathbb{N}$ has already been defined. The greedy algorithm picks $2^n s$ to be the least integer not in *anathema*, the set (cf. [3, (2.7)]):

$$\{es +_2 fs \mid e, f \in V_n \cap E; 2^n +_2 e \in E\}.$$

4. Binary burst-error-correcting syndrome functions

When E is some burst-error pattern, the improper syndrome function $s_E : 2^{\mathbb{N}} \to \mathbb{N}$ with $0 s_E = 0$ can be constructed by the greedy algorithm. To compare the syndrome functions $s_E : E \to \mathbb{N}$ for each error pattern E, they may be graphed with $\log_2 x$ on the ordinate and y on the abscissa. We define a *nodal point* of the syndrome function s_E to be a point on its graph of the form $(2^{n-1}, 2^k)$ for $n-1,\ k \in \mathbb{N}$. The proper syndrome given by the restriction of s_E to the channel V_n then yields this burst-error-correcting code C_n with redundancy $r = k + 1$ satisfying

$$f_E(n) \cdot 2^{n-r} \leqslant 2^n,$$

Table 1
Efficiencies

$r\backslash E$	B_2	B_2S	$2B_2$
2	100[2]	100[2]	100[2]
3	86[3]	93[3]	93[3]
4	90[6]	92[4]	92[4]
5	86[10]	89[5]	90[5]
6	94[25]	93[7]	87[6]
7	91[41]	92[9]	90[8]
8	97[108]	88[11]	88[10]
9	93[165]	87[14]	88[13]
10	95[372]	85[17]	86[16]
11	96[771]	85[23]	86[21]
12		84[29]	85[27]
13		83[37]	84[33]
14		82[47]	84[43]

Note: The numbers inside [ ] give the lengths of the channel achieving the corresponding efficiency.

since the channel must contain a disjoint union of 2^{n-r} "stars", each with $f_E(n)$ points (cf. [5, (5.2)]). When the burst-error pattern E has the form B_2 ($B_2S, 2B_2$), the error-correcting code for that error pattern E will also be denoted by B_2 (resp. B_2S, $2B_2$). The functions $f_E(n)$ for these burst-error patterns are given as follows.

Error pattern E	$f_E(n)$
B_2	$2n$
B_2S	$(3/2)n^2 - (9/2)n + 7$
$2B_2$	$2n^2 - 7n + 8$

The *efficiency* of the code C_n of redundancy r is the ratio of $\log_2 f_E(n)$ to r, usually expressed as a percentage. Table 1 lists the highest efficiencies for $1 < r < 15$ in each burst-error pattern. The numbers inside [ ] are the length of the channel which has the highest efficiency for the given redundancy. Tables 2 and 3 give the dimensions of the binary greedy loop transversal codes for each burst-error pattern considered. Table 2 gives the dimensions of the codes for the error pattern B_2. For comparison, the table also gives the dimensions, in ( ), for the white noise double error-correcting greedy loop transversal codes from [3], i.e. corresponding to minimum Hamming distance $d = 5$. Table 3 gives the dimensions of the codes for the error patterns B_2S and $2B_2$. Again for comparison, the table also gives the dimensions, in ( ), for the white noise triple and quadruple error correcting greedy loop transversal codes from [3], i.e. corresponding to minimum Hamming distances $d = 7, 9$. The comparisons enable one to observe the quantitative gain in information rate resulting from restriction of the higher-weight errors to burst patterns.

Table 2
Dimensions of codes for the error pattern B_2

n	B_2	n	B_2
7–10	$n-5$	109–120	$n-9(n-15)$
11	5	121–156	$n-9(n-16)$
12–17	$n-6(n-8)$	157–165	$n-9(n-17)$
18–21	$n-6(n-9)$	166–203	$n-10(n-17)$
22–25	$n-6(n-10)$	204–266	$n-10(n-18)$
26–29	$n-7(n-10)$	267–342	$n-10(n-19)$
30–38	$n-7(n-11)$	343–360	$n-10(n-20)$
39–41	$n-7(n-12)$	361–372	$n-10$
42–53	$n-8(n-12)$	373–771	$n-11$
54–69	$n-8(n-13)$		
70–92	$n-8(n-14)$		
93–108	$n-8(n-15)$		

Note: The numbers inside ( ) give dimensions of codes from [3] for $d=5$.

Table 3
Dimensions of codes for error patterns B_2S and $2B_2$

n	B_2S	$2B_2$	n	B_2S	$2B_2$	n	B_2S	$2B_2$
10	2	2[2]	27	15(13)	15(9)	44	30	30
11	3	2	28	16(13)	15(10)	45	31	30
12	3(2)	3(1)	29	17(14)	16(11)	46	32	
13	4(3)	4(1)	30	17(15)	17(12)	47	33	
14	5(4)	4(2)	31	18(16)	18(12)	48	33	
15	5(5)	5(2)[5]	32	19(16)	19(13)	49	34	
16	6(5)	6(2)	33	20(17)	20(14)	50	35	
17	7(6)	6(3)	34	21(18)	20	51	36	
18	7(7)	7(3)	35	22(19)	21	52	37	
19	8(8)	8(4)	36	23(20)	22	53	38	
20	9(9)	9(5)	37	24(21)	23	54	39	
21	10(10)	10(5)[8]	38	24(22)	24	55	40	
22	11(11)	10(6)	39	25(23)	25	56	41	
23	12(12)	11(6)[11]	40	26	26	57	42	
24	12(12)	12(7)	41	27	27	58	43	
25	13(12)	13(8)	42	28	28	59	44	
26	14(12)	14(9)	43	29	29			

Note: Numbers inside ( ) are data from [3] when $d=7$ and $d=9$ and numbers inside [ ] from [9].

References

[1] J.H. Conway, On Numbers and Games, Cambridge University Press, Cambridge, 1975.
[2] J.H. Conway, N.J.A. Sloane, Lexicographic codes: error-correcting codes from game theory, IEEE Trans. Inform. Theory 32 (1986) 337–348.
[3] F.-L. Hsu, F.A. Hummer, J.D.H. Smith, Logarithms, syndrome functions, and the information rate of greedy loop transversal codes, J. Combin. Math. Combin. Comput. 22 (1996) 33–49.

[4] F.A. Hummer, J.D.H. Smith, Greedy loop transversal codes, metrics, and lexicodes, J. Combin. Math. Combin. Comput. 22 (1996) 143–155.

[5] J.D.H. Smith, Loop transversals to linear codes, J. Combin. Inform. System Sci. 17 (1992) 1–8.

[6] J.D.H. Smith, Loop transversal codes, Cubo 3 (2001) 137–148.

[7] J.D.H. Smith, A.B. Romanowska, Post-Modern Algebra, Wiley, New York, 1999.

[8] W.M.C.J. van Overveld, Some constructions of new burst-error-correcting codes, IEEE Trans. Inform. Theory 33 (1) (1987) 153.

[9] H.C.A. van Tilborg, An overview of recent results in the theory of burst-correcting codes, Lecture Notes in Computer Science, Vol. 388, Springer, New York, Berlin, 1989, pp. 164–184.