Loop Transversal Codes for Error Detection

Ruben Aydinyan and Jonathan D. H. Smith
Department of Mathematics, Iowa State University,
Ames, Iowa 50011-2064
raydinya@iastate.edu, jdhsmith@iastate.edu

Abstract

In a loop transversal code, the set of errors is given the structure of a loop transversal to the linear code as a subgroup of the channel. A greedy algorithm for specifying the loop structure, and thus for the construction of loop transversal codes, was discussed by Hummer et al. Apart from some theoretical considerations, the focus was mainly on error correction, in the white noise case constructing codes with odd minimum distance. In this paper an algorithm to compute loop transversal codes with even minimum distance is given. Some record breaking codes over a 7-ary alphabet are presented.

Introduction

In 1992 the second author introduced the idea of constructing codes by giving an algebraic structure to the set of errors [10]. Then the algebra of errors, called a loop transversal, is isomorphic to the dual of the code that will correct this set of errors. Consider a set $E \subset V$ of errors, where $(V, +, 0)$ is a (not necessarily abelian) group, the so-called *channel*. A linear map $\varepsilon : V \to V$ is defined such that $\varepsilon|_E$ is injective. The map ε is called the *syndrome* function. The kernel C of ε is realized as the code correcting E. Then V may be represented as the disjoint union of cosets of C,

$$V = \bigcup_{t \in T} (C + t)$$

where T is the set of coset representatives, called a *transversal* to C. Thus each element $v \in V$ can be expressed uniquely as

$$v = \delta(v) + \varepsilon(v) \tag{1}$$

where $\delta(v) \in C$ and $\varepsilon(v) \in T$. (In the coding context, a received word v has been exposed to error $\varepsilon(v)$, and hence has to be decoded as $\delta(v)$).

A binary operation $*$ is defined on T by

$$t_1 * t_2 = \varepsilon(t_1 + t_2) \tag{2}$$

For any $t_1, t_2 \in T$, the equation $x * t_1 = t_2$ has a unique solution x. If the equation $t_1 * y = t_2$ also has a unique solution, then T is said to be a *loop transversal.* The algebra $(T, *, \varepsilon(0))$ is a loop [10]. In traditional coding theory the channel V is an abelian group, and thus each transversal is a loop transversal, and the loop $(T, *, \varepsilon(0))$ is an abelian group. For t_i in T the product $\prod_{i=1}^{r} t_i$ is defined inductively by

$$\prod_{i=1}^{0} t_i = \varepsilon(0)$$

and

$$\prod_{i=1}^{r} t_i = \left[\prod_{i=1}^{r-1} t_i\right] * t_r.$$

Now, if V is a finite dimensional vector space over a field $\mathbb{F}$, define $\lambda \times t = \varepsilon(\lambda t)$ for λ in $\mathbb{F}$ and t in T. Then the algebra $(T, *, \mathbb{F})$ becomes a vector space over $\mathbb{F}$. Induction on r extends (2) to

$$\varepsilon\left(\sum_{i=1}^{r} \lambda_i t_i\right) = \prod_{i=1}^{r} (\lambda_i \times t_i) \tag{3}$$

for t_i in T. It is reasonable to require $T = E$ to contain a basis $\{e_1, \ldots, e_n\}$ for $V = \mathbb{F}^n$, where e_i has 1 in the i-th position[1] and 0's everywhere else (the set of single errors). Then knowledge of a portion of the vector space $(T, *, \mathbb{F})$ is sufficient to determine the corresponding portion of the code C. Indeed, for $k \leq n$,

[1] The i-th position is counted from the right, i.e. $e_1 = 0\ldots001$, $e_2 = 0\ldots010$, $e_3 = 0\ldots100$, etc.

$$\begin{aligned} C = \{\delta(v) |\, v \in V\} &= \{v - \varepsilon(v) |\, v \in V\} \\ &= \Big\{ \sum_{i=1}^{k} \lambda_i e_i - \varepsilon\Big(\sum_{i=1}^{k} \lambda_i e_i\Big) \Big|\, \lambda_i \in \mathbb{F} \Big\} \\ &= \Big\{ \sum_{i=1}^{k} \lambda_i e_i - \prod_{i=1}^{k} (\lambda_i \times e_i) |\, \lambda_i \in \mathbb{F} \Big\} \end{aligned}$$

the so-called *principal of local duality* [10].

Greedy algorithm

The syndrome function ε is constructed so that it is linear and $\varepsilon|_E$ is injective. First, the algorithm sorts the set E of errors into a lexicographic order [6]-[8]. Then for each error $e \in E$ a syndrome value

$$\varepsilon(e) = \begin{cases} 0 & \text{if } e = 0 \\ \min\{v \in V |\, v \neq \varepsilon(e'),\ e' < e_{i+1}\} & \text{if } e = e_i \\ \sum_{i=1}^{t} \alpha_i \varepsilon(e_i) & \text{if } e = \sum_{i=1}^{t} \alpha_i e_i \end{cases} \tag{4}$$

is assigned successively. The latter guarantees the linearity condition, and the former is the greedy choice. Then the kernel of ε defined in (4) is a linear loop transversal code correcting E.

The error set E may match any kind of channel statistics, for example burst errors [4]. In the classical "white noise" case, where E is the set of all vectors $e \in \mathbb{F}^n$ of weight up to t, the LT code produced will have a minimum distance $d = 2t + 1$.

Error Detection

From now on, we will restrict our attention to the case of "white noise" statistics. Let $E = \{e \in \mathbb{F}^n |\, wt(e) \leq t\}$ be the set of errors to be corrected, and let $D = \{v \in \mathbb{F}^n |\, wt(v) = t + 1\}$ be the set of errors to be detected. Here $wt(v)$ is the (Hamming or Lee) weight of vector v. We will say that ε *avoids* the set S if $\varepsilon(v) \neq 0$ for all $v \in S$.

Theorem 1 *Let ε be a syndrome function of a greedy loop transversal code correcting E. If ε avoids the set*

$$S = \{(v - e) | v \in D,\ e \in E,\ wt(v - e) = 2t + 1\},$$

then the code defined by ε has a minimum distance $2(t+1)$.

Proof. If we wish to detect $v \in D$, we want $v \neq c+e$ for any $c \in C$ and $e \in E$, i.e. $v-e$ must not be a codeword. Furthermore, an E-correcting LT code is guaranteed to have a minimal distance, and hence minimal weight $d = 2t+1$. Therefore, no difference vector $v-e$ with $wt(v-e) < 2t+1$ can be a codeword. So, if in the construction of the greedy LT code we require ε to avoid the set S, then no vector of weight $2t+1$ will be in the kernel of ε. Hence the resulting code will detect $(t+1)$-errors and have a minimum distance $2(t+1)$. □

Theorem 1 shows that adding an extra check to the greedy loop transversal algorithm given in [7] will enable one to construct LT codes with even minimal distances and error detection. Note that extending the LT codes this way is independent of whether we choose the Hamming metric or the Lee metric on V. The algorithm simplifies a great deal when we restrict our attention to the Hamming metric: Corollary 1 below. The chain of ideas leading through Theorem 2 to the corollaries are well-known (compare [12] for example), but it is useful to work out their implications in the context of LT codes for the Hamming metric.

Definition 1 *A set S of vectors is called* m-independent *if each set of m vectors from S is linearly independent.*

Definition 2 *An m-independent set $S \subset V$ is called* maximal *if for any other m-independent set $D \subset V$, $|D| \leq |S|$.*

Our next theorem states that $(d-1)$-independent sets are exactly the ones that are the images of the set $\{e_1, e_2, \ldots, e_n\}$ under the syndrome map.

Theorem 2 *A linear map $\varepsilon : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^n$ is a syndrome function defined by a linear $[n,k,d]$ code if and only if the set $S = \{\varepsilon(e_1), \varepsilon(e_2), \ldots, \varepsilon(e_n)\}$ is $(d-1)$-independent.*

Proof. If ε is the syndrome function of an $[n,k,d]$ linear code, then each codeword has a Hamming weight at least d, and thus there is no vector $v \in \mathbb{F}_q^n$ of Hamming weight less than d such that $\varepsilon(v) = 0$, i.e., the equation

$$0 \neq \alpha_1\varepsilon(e_{i_1}) + \alpha_2\varepsilon(e_{i_2}) + \ldots + \alpha_{d-1}\varepsilon(e_{i_{d-1}})$$

holds for all $\alpha_i \in \mathbb{F}_q$. Hence every subset $\{\varepsilon(e_{i_1}), \varepsilon(e_{i_2}), \ldots, \varepsilon(e_{i_{d-1}})\}$ of S is linearly independent, and therefore S is a $(d-1)$-independent set.

Conversely, if $S = \{v_1, \dots, v_n\}$ is a $(d-1)$-independent set, define a linear function $\varepsilon : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^n;\ e_i \mapsto v_i$. Then if $\varepsilon(u_1) = \varepsilon(u_2)$ for some $u_1 \neq u_2 \in \mathbb{F}_q^n$, then by linearity of ε the vector $u_1 - u_2$ is in the kernel of ε. Using the standard representation we can write:

$$u_1 = \alpha_1 e_1 + \ldots + \alpha_n e_n\ ;$$

$$u_2 = \beta_1 e_1 + \ldots + \beta_n e_n.$$

Then

$$0 = \varepsilon(u_1 - u_2) = (\alpha_1 - \beta_1)\varepsilon(e_1) + (\alpha_2 - \beta_2)\varepsilon(e_2) + \ldots + (\alpha_n - \beta_n)\varepsilon(e_n)$$

and

$$0 = \gamma_1 v_1 + \gamma_2 v_2 + \ldots + \gamma_n v_n\ , \tag{5}$$

where $\gamma_i = \alpha_i - \beta_i$. Since $u_1 \neq u_2$, not all of the γ_i are zero. Then the $(d-1)$-independence of S guarantees that there must be at least d nonzero summands on the right hand side of equation (5). Thus $\alpha_i \neq \beta_i$ for at least d values of $i \in \{1, 2, \dots, n\}$, i.e. $wt_H(u_1 - u_2) \geq d$. Hence the kernel C of ε defines a linear code with minimum distance d. □

Thus the knowledge of a $(d-1)$-independent set S completely determines a corresponding $[n, k, d]$ linear code, where $n = |S|$ and $k = n - dim(S)$.

Corollary 1 *The $(d-1)$-independent set of vectors chosen greedily (with lexicographic order) defines the syndrome function for an LT code of minimum distance d.*

Corollary 2 *A loop transversal [n,k,d] code is optimal if and only if the corresponding $(d-1)$-independent set is maximal.*

Results

The implementation of the greedy loop transversal algorithm has produced many astonishing results in both the binary and nonbinary cases. A vast number of optimal and best-known codes are produced using the algorithm. Along with these codes the extended Hamming codes, the binary Golay $[24, 12, 8]$, Reed-Muller $[16, 5, 8]$, the quadratic residue $[18, 9, 6]$, and the ternary Golay $[12, 6, 6]$ codes are obtained. Our previous theorem and the corollaries give a new way of computing the syndrome function. It makes it easier and faster to compute a $(d-1)$-independent set. Among the quaternary codes a code equivalent to the octacode described in [5] is obtained, and consequently the Nordstrom-Robinson code as its binary image under the Gray map [3], [5], [9], [11]. Record breaking codes of minimum distance six with parameters $[32, 24, 6]$, $[33, 25, 6]$, $[34, 26, 6]$, $[35, 27, 6]$, $[36, 28, 6]$, $[37, 29, 6]$, and $[38, 30, 6]$ over $\mathbb{Z}_7$ are obtained. The procedure for computing a generating matrix for greedy LT codes was given in [1].

The records were compared with codes in A. Brouwer's online catalog [2].

References

[1] R. H. Aydinyan, *Loop transversal codes over finite rings*, Ph.D. Dissertation, Math. Dept., Iowa State University, 2004.

[2] A. E. Brouwer, *Linear code bounds* (on-line server), Eindhoven University of Technology, The Netherlands, http://www.win.tue.nl/ aeb/voorlincod.html (Retrieved: 19 December 2004).

[3] A. R. Calderbank, A. R. Hammons, Jr., P. V. Kumar, N. J. A. Sloane, and P. Solé, *The $\mathbb{Z}_4$-linearity of Kerdock, Preparata, Goethals and related codes*, IEEE Trans. Inform. Theory, **40** (1994), pp. 301-319.

[4] D.-H. Choi and J. D. H. Smith, *Greedy loop transversal codes for correcting error bursts*, Discrete Math. **264** (2003), pp. 37-43.

[5] G. D. Forney Jr., N. J. A. Sloane, and M. D. Trott, *The Nordstrom-Robinson code is the binary image of the octacode*, Coding and Quantization: DIMACS/IEEE Workshop, October 19-21, 1992, R.Calderbank, G.D.Forney Jr., and N.Moayeri, Eds., Amer. Math. Soc., 1993, pp. 19-26.

[6] F. A. Hummer, *Loop transversal codes*, Ph.D. Dissertation, Math. Dept, Iowa State University, 1992.

[7] F. A. Hummer and J. D. H. Smith, *Greedy loop transversal codes, metrics, and lexicodes*, J. Comb. Math. Comb. Comp. **22** (1996), pp. 143–155.

[8] F. -L. Hsu, F. A. Hummer, and J. D. H. Smith *Logarithms, syndrome functions, and the information rates of greedy loop transversal codes*, J. Comb. Math. Comb. Comp. **22** (1996), pp. 33-49.

[9] N. J. A. Sloane, *Algebraic coding theory: recent developments related to $\mathbb{Z}_4$*, Study of Algebraic Combinatorics (Proceedings Conference on Algebraic Combinatorics, Kyoto 1993), Research Institute for Mathematical Sciences, Kyoto, 1995, pp. 38-52.

[10] J. D. H. Smith, *Loop transversals to linear codes.* Journal of Combinatorics, Information & System Sciences, **17** (1992), Nos. 1–2, 1–8.

[11] S. L. Snover, *The uniqueness of the Nordstrom-Robinson and the Golay binary codes*, Ph.D. Dissertation, Math. Dept., Michigan State Univ., 1973.

[12] R. R. Varshamov, *Estimate of the number of signals in error correcting codes*, Dokl. Acad. Nauk SSSR, **117** (1957), pp. 739 – 741.