Closed-form expression for the parameters of binary lexicodes

Proposed version: Ari Trachtenberg

I. Description

A minimum distance d binary lexicode is greedily constructed as follows: start with the all-zero vector; repeatedly add (until exhaustion) the lexicographically earliest (*i.e.*, appearing first in a dictionary) vector at Hamming distance $\geq d$ from the previously chosen codewords. The length of the code is determined as the support size of the code. As a simple example, constructing the minimum distance $d = 2$ binary lexicode would entail adding, in order, codewords 0, 11, 110, and 111 resulting in a code of length $n = 3$ with 4 codewords.

A. Open problem:

Provide an explicit, closed-form expression for the length n of an arbitrary, minimum distance d, binary lexicode with 2^k codewords. Even a solution for low dimensions (say $k \leq 30$) could be of interest.

II. Motivation

Lexicographic codes (nicknamed *lexicodes*) were introduced by Levenshtein [1] and by Conway and Sloane [2, 3]. Despite their greedy construction, they have surprisingly good code parameters. For a given length and distance, they are typically within one of the best-known minimum distance, and they often exhibit the smallest known covering radius. They also have a number of famous optimal codes as special cases, including the Hamming codes, the binary Golay code, and certain quadratic-residue codes. As such, they have an empirically "almost-optimal" character to them, and studying their properties could tighten bounds and provide automated generation for optimal error-correcting codes.

III. Prior work

In their initial work, Conway and Sloane [2, 3] provided a short table of various lexicode parameters. Thereafter, a number of researchers have provided improved algorithms for generating lexicodes and their parameters, based on game theoretic and coding theoretic considerations (see [4–9] and the references contained therein). The most extensive lists of parameters we know are based on the work of Bob Jenkins [10] and available through the On-Line Encyclopedia of Integer Sequences [11]. These tables get very short as Hamming distance increases, and this is consistent with the fact that existing algorithms for generating lexicodes are exponential in the Hamming distance d or the co-dimension $n - k$.

Some general properties of the lexicode parameters are also known. For one, the minimum distance 4 binary lexicodes are all extended Hamming codes or shortenings thereof [3], meaning that their dimension k and length n observe the relationship $k = n-2-\lfloor \log_2(n-1) \rfloor$. In addition, binary lexicodes with odd minimum distance are punctured versions of their even minimum distance counterparts [3]. More generally, lexicodes have some weak bounds [4]

$$k \geq \begin{cases} n - 2 - \lfloor \log_2(n-1) \rfloor & \text{if } d = 4, \\ \left\lfloor \frac{4n-d-12}{2d-4} \right\rfloor & \text{if } d \equiv 0 \pmod 4, d \neq 4, 8, \\ \left\lfloor \frac{n}{3} \right\rfloor & \text{if } d = 8,\ n > 18, \\ \left\lfloor \frac{4n-d-14}{2d-4} \right\rfloor & \text{if } d \equiv 2 \pmod 4). \end{cases} \tag{1}$$

and very slightly tighter bounds [7]

$$n \leq \left(k + \frac{1}{2}\right) \frac{d+4}{3} - \frac{2}{3}, \tag{2}$$

or asymptotically $n \leq \frac{kd}{3}$. Lexicodes, like most linear codes, also meet the Gilbert bound [12].

A. Proposed direction

Since binary lexicodes are linear, it is sufficient to generate their basis to determine the length of the code. Dimension $k = 1$ binary lexicodes have exactly two codewords, 0 and 1^d, where we use the notation a^b to denote b copies of a. Thus, dimension 1 binary lexicodes have length n equal to their minimum distance d.

By inspection, we can see that the next basis vector to add to our dimension 1 lexicode will be of the form $1^x 0^y 1^z$, where $x = y = d - \left\lfloor \frac{d}{2} \right\rfloor$ and $z = \left\lfloor \frac{d}{2} \right\rfloor$. Thus, the binary lexicodes of dimension $k = 2$ have length $2d - \left\lfloor \frac{d}{2} \right\rfloor$. This result can also be determined more systematically by using Theorem 3 in [7], which defines a relationship between the coset leaders of lexicodes of dimension k and $k+1$, for any given k. We can apply Theorem 3 to the coset leaders of the dimension $k = 1$ lexicode to get a general form for the coset leaders of the $k = 2$ lexicode, from which the code length is apparent.

Though Theorem 3 from [7] can be applied once again to determine the coset leaders (and hence length) of the $k \geq 3$ lexicode, the number of constraints quickly becomes too complicated for simple hand calculations. At a minimum, it should be possible to develop a symbolic algebra for manipulating these constraints by computer to generate explicit, closed-form expressions for low-dimensional lexicodes of arbitrary minimum distance.

References

[1] V.I. Levenšteĭn, "A class of systematic codes," *Soviet Math. Dokl.*, vol. 1, no. 1, pp. 368–371, 1960.

[2] J.H. Conway, "Integral lexicographic codes," *Discrete Math*, vol. 83, pp. 219 – 235, 1990.

[3] J.H. Conway and N.J.A. Sloane, "Lexicographic codes: Error-correcting codes from game theory," *IEEE Trans. on Inform. Theory*, vol. 32, pp. 337–348, 1986.

[4] R.A. Brualdi and Vera S. Pless, "Greedy codes," *Journal of Comb. Th. (A)*, vol. 64, pp. 10–30, September 1993.

[5] A. Trachtenberg, *Error-Correcting Codes on Graphs: Lexicodes, Trellises, and Factor Graphs*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 2000, available at http://ipsit.bu.edu/documents/phdthesis.ps.

[6] A. Trachtenberg and A. Vardy, "Lexicographic codes: Constructions, bounds, and trellis complexity," in *Proc. 31st Annual Conference on Information Sciences and Systems*, Baltimore, MD, March 1997, pp. 521–527.

[7] Ari Trachtenberg, "Designing lexicographic codes with a given trellis complexity," *IEEE Trans. on Inform. Theory*, vol. 48, no. 1, pp. 89, 2002.

[8] N. Suparta, *Counting Sequences, Gray Codes, and Lexicodes*, Ph.D. thesis, Delft University of Technology, 2006.

[9] A.S. Fraenkel and O. Rahat, "Complexity of error-correcting codes derived from combinatorial games," *Lecture Notes in Computer Science*, vol. 2883, pp. 201–212, 2003.

[10] Bob Jenkins, "Table of lexicodes," http://www.burtleburtle.net/bob/math/lexicode.html.

[11] N.J.A. Sloane, "The online encyclopedia of integer sequences," http://www.research.att.com/ njas/sequences/.

[12] W.C. Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*, Cambridge University Press, 2003.