Construction of q-ary Constant Weight Sequences using a Knuth-like Approach

Elie N. Mambou & Theo G. Swart

Department of Electrical and Electronic Engineering Science,
University of Johannesburg (UJ)
South Africa

2017 IEEE International Symposium on Information Theory
RWTH University, Aachen, Germany

June 25-30, 2017

About this work

- Extension on “Encoding and decoding of balanced q-ary sequences using a Gray code prefix,” ISIT 2016.
- ”A Construction for Balancing Non-Binary Sequences Based on Gray Code Prefixes”, arXiv:1706.00852v1.
- Received the chancellor’s medal award for best master dissertation.
- Thanks to my advisor and external examiners.

Overview

Definitions

- Consider a q-ary information sequence $\boldsymbol{x} = x_0 x_1 x_2 \ldots x_{k-1}$, $x_i \in \{0, 1, \ldots, q-1\}$, of length k.
- Let the prefix that will be appended to $\boldsymbol{x}$ be of length r; and let the information and the prefix together be denoted by $\boldsymbol{c} = c_0 c_1 c_2 \ldots c_{k-1}$, $c_i \in \{0, 1, \ldots, q-1\}$, of length $n = k + r$.
- The weight of $\boldsymbol{c}$, $w(\boldsymbol{c})$ is defined as

$$w(\boldsymbol{c}) = \sum_{i=0}^{k-1} c_i.$$

- $\boldsymbol{c}$ is called constant weight (CW) sequence with weight, $w(\boldsymbol{c})$ and it is said to be balanced if $w(\boldsymbol{c}) = \beta_{n,q} = \frac{n(q-1)}{2}$.

Balancing of q-ary sequences

- It has been proven [1], that $\boldsymbol{x}$, can always be balanced by adding modulo q one sequence from a set of balancing sequences $\boldsymbol{b}(s,p) = b_1 b_2 \dots b_k$ generated as follows:

$$b_i = \begin{cases} s, & i-1 \geq p, \\ s+1 \pmod{q}, & i-1 < p, \end{cases} \quad \text{where} \quad \begin{cases} 0 \leq s \leq q-1, \\ 0 \leq p \leq k-1. \end{cases}$$

- Let z be the iterator through these balancing sequences, with $z = sk + p$, $0 \leq z \leq kq - 1$. $\boldsymbol{b}(s,p)$ and $\boldsymbol{b}(z)$ refers to the same.
- Let $\boldsymbol{y}$ denote the sequence after a balancing sequence is added, $\boldsymbol{y} = \boldsymbol{x} \oplus_q \boldsymbol{b}(z)$. At least one $\boldsymbol{b}(z)$ will lead to a balanced output $\boldsymbol{y}$.

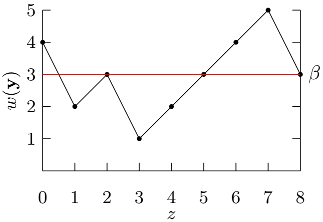
[1] T. G. Swart and J. H. Weber, "Efficient balancing of q-ary sequences with parallel decoding," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009.

Balancing of q-ary sequences (Cont'd)

Example 1

For $q = 3$, $k = 3$, consider the sequence $\boldsymbol{x} = 202$. The balancing value is $\beta_{k,q} = 3$.

z	$\boldsymbol{b}(z)$	$\boldsymbol{x} \oplus_q \boldsymbol{b}(z) = \boldsymbol{y}$	$w(\boldsymbol{y})$
0	000	$202 \oplus_3 000 = 202$	4
1	100	$202 \oplus_3 100 = 002$	2
2	110	$202 \oplus_3 110 = 012$	**3**
3	111	$202 \oplus_3 111 = 010$	1
4	211	$202 \oplus_3 211 = 110$	2
5	221	$202 \oplus_3 221 = 120$	**3**
6	222	$202 \oplus_3 222 = 121$	4
7	022	$202 \oplus_3 022 = 221$	5
8	002	$202 \oplus_3 002 = 201$	**3**



q-ary Gray Codes

- Invented by Frank Gray [2]; originally used to solve problems in pulse code communication; and extended to several other fields.
- $\boldsymbol{d} = d_1 d_2 \ldots d_{r'}$ denotes a sequence amongst the set of q-ary sequences of length r' listed in lexicographic order. They are mapped to Gray code sequences, $\boldsymbol{g} = g_1 g_2 \ldots g_{r'}$. Any two adjacent sequences differ in only one symbol position, with weight difference of either -1 or $+1$.
- *4-ary Gray code of length 2*

z	$\boldsymbol{d}$	$\boldsymbol{g}$	z	$\boldsymbol{d}$	$\boldsymbol{g}$	z	$\boldsymbol{d}$	$\boldsymbol{g}$	z	$\boldsymbol{d}$	$\boldsymbol{g}$
0	00	00	4	10	13	8	20	20	12	30	33
1	01	01	5	11	12	9	21	21	13	31	32
2	02	02	6	12	11	10	22	22	14	32	31
3	03	03	7	13	10	11	23	23	15	33	30

[2] F. Gray, "Pulse code communication," *U. S. Patent 2632058*, 1953.

Encoding and Decoding of q-ary Gray codes [3]

Gray code encoding algorithm The parity of the sum S_i of the first $i-1$ digits of $\boldsymbol{g}$ determines the Gray code symbols, where $2 \leq i \leq r'$ and $g_1 = d_1$, then

$$S_i = \sum_{j=1}^{i-1} g_j, \quad \text{and} \quad g_i = \begin{cases} d_i, & \text{if } S_i \text{ is even,} \\ q-1-d_i, & \text{if } S_i \text{ is odd.} \end{cases}$$

Gray code decoding algorithm

$$S_i = \sum_{j=1}^{i-1} g_j, \quad \text{and} \quad d_i = \begin{cases} g_i, & \text{if } S_i \text{ is even,} \\ q-1-g_i, & \text{if } S_i \text{ is odd.} \end{cases}$$

[3] D.-J. Guan, "Generalized Gray codes with applications," in *Proc. National Science Council, Republic of China, Part A*, 1998.

Applications of CW sequences

- They play an important role in communication system where high security and confidentiality are needed, because of various properties such as correlations, balanced value distributions and strong linear complexity.
- Frequency hopping in GSM networks.
- Detection of unidirectional errors and threshold setting in barcode implementations.
- DNA sequences (Biology field).
- In VLC (visible light communication), to eliminate flickering in CSK and performing dimming in FSK, OOK.

Research goal

- Construction of CW sequences through an efficient encoding and decoding scheme.

Encoding

Generating q-ary CW Sequences

- We want to construct an (n, k, W, q) CW sequence of length n, weight W with k information symbols.
- The length of Gray code prefix is, $r' = \log_q(kq) = \log_q(k) + 1$; such that cardinalities of the set of Gray code prefix and that of weighting sequences are equal.
- Lemma 1. For any q-ary information sequence $\boldsymbol{x}$ of length k, where parameters k and q are not coprime, we can find a $\boldsymbol{b}(z)$ such that the weight of $\boldsymbol{y} = \boldsymbol{x} \oplus_q \boldsymbol{b}(z)$ is $\omega_1 \leq w(\boldsymbol{y}) \leq \omega_2$, where $\omega_1 = \beta_{k,q} - (q-1)$ and $\omega_2 = \beta_{k,q} + (q-1)$.
- Theorem 1. An (n, k, W, q) CW sequence can be constructed from any q-ary information sequence $\boldsymbol{x}$ of length k where

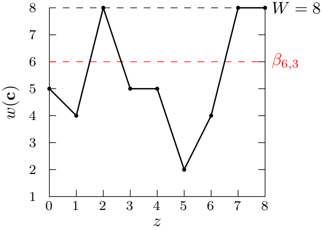
$$\frac{(k-2)(q-1)}{2} \leq W \leq \frac{(k+2r'+4)(q-1)}{2}. \quad (1)$$

Encoding (Cont'd)

Example 2

Encoding the ternary sequence $\boldsymbol{x} = 212$ into a CW sequence of weight $W = 8$. The condition $k = q^t$, is imposed and the Gray code prefix length is $r' = log_3 3 + 1 = 2$.

z	$\boldsymbol{x} \oplus_q \boldsymbol{b}(z) = \boldsymbol{y}$	$\boldsymbol{c} = [u \vert \boldsymbol{g} \vert \boldsymbol{y}]$	$w(\boldsymbol{c})$
0	$212 \oplus_3 000 = 212$	$\underline{\mathbf{0}00}212$	5
1	$212 \oplus_3 100 = 012$	$\underline{\mathbf{0}01}012$	4
2	$212 \oplus_3 110 = 022$	$\underline{\mathbf{2}02}022$	**8**
3	$212 \oplus_3 111 = 020$	$\underline{\mathbf{0}12}020$	5
4	$212 \oplus_3 211 = 120$	$\underline{\mathbf{0}11}120$	5
5	$212 \oplus_3 221 = 100$	$\underline{\mathbf{0}10}100$	2
6	$212 \oplus_3 222 = 101$	$\underline{\mathbf{0}20}101$	4
7	$212 \oplus_3 022 = 201$	$\underline{\mathbf{2}21}201$	**8**
8	$212 \oplus_3 002 = 211$	$\underline{\mathbf{0}22}211$	**8**



Encoding (Cont'd)

Generating q-ary CW Sequences with extended weight range

- Appending a redundant vector $\boldsymbol{u}$ of length e to $\boldsymbol{c}' = [\boldsymbol{g}|\boldsymbol{y}]$, then the output sequence becomes $\boldsymbol{c} = [\boldsymbol{u}|\boldsymbol{g}|\boldsymbol{y}]$. This leads to (n, k, W, q) CW sequences where $n = k + r' + e$.
- This will lead to an increase of weight range as $w(\boldsymbol{u}) \in [0, e(q-1)]$.
- Theorem 2. Any q-ary information sequence of length k can generate an (n, k, W, q) CW sequence where

$$\frac{(k-2)(q-1)}{2} < W < \frac{(k+2r'+2e+1)(q-1)}{2}. \tag{2}$$

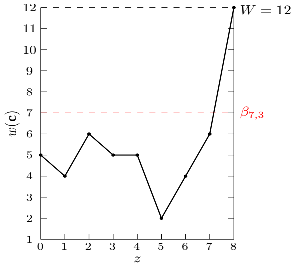
- The redundant vector $\boldsymbol{u} = u_1 u_2 \ldots u_e$ is such that $u_i \in \{0, 1, \ldots, q-1\}$ and $w(\boldsymbol{u}) = W - w(\boldsymbol{c}')$ if and only if $W \geq w(\boldsymbol{c}')$, otherwise $\boldsymbol{u} = \boldsymbol{0}$.

Encoding (Cont'd)

Example 3

Consider the same ternary information sequence $\boldsymbol{x} = 212$ of length 3 as in Example 2. We would like to generate a $(7, 3, 12, 3)$ CW sequence of weight $W = 12$ and $n = 7$.

z	$\boldsymbol{x} \oplus_q \boldsymbol{b}(z) = \boldsymbol{y}$	$\boldsymbol{c} = [\boldsymbol{u} \vert \boldsymbol{g} \vert \boldsymbol{y}]$	$w(\boldsymbol{c})$
0	$212 \oplus_3 000 = 212$	0000212	5
1	$212 \oplus_3 100 = 012$	0001012	4
2	$212 \oplus_3 110 = 022$	0002022	6
3	$212 \oplus_3 111 = 020$	0012020	5
4	$212 \oplus_3 211 = 120$	0011120	5
5	$212 \oplus_3 221 = 100$	0010100	2
6	$212 \oplus_3 222 = 101$	0020101	4
7	$212 \oplus_3 022 = 201$	0021201	6
8	$212 \oplus_3 002 = 211$	2222211	**12**



Range has been extended from [2, 10] to [2, 12].

Encoding (Cont'd)

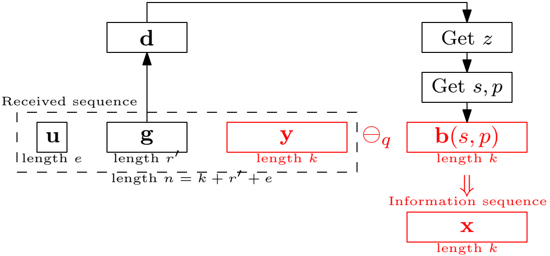
Parameters evaluation

- W was calculated according to equation (2).

	t	$k = q^t$	W	n	r'	e
$q = 2$	2	4	$[2, 8]$	10	3	3
	3	8	$[4, 12]$	16	4	4
	4	16	$[8, 16]$	24	5	3
$q = 3$	1	3	$[2, 11]$	7	2	2
	2	9	$[8, 21]$	15	3	3
	3	27	$[26, 41]$	34	4	3
$q = 4$	1	4	$[4, 18]$	8	2	2
	2	16	$[22, 42]$	22	3	3
	3	64	$[94, 120]$	72	4	4

Decoding

- The redundant vector $\boldsymbol{u}$ is dropped, then the r' symbols are extracted as the Gray code prefix and converted to corresponding iterator z.
- z is used to determine the parameters s and p, then $\boldsymbol{b}(s, p)$ can be derived.
- Finally, the original sequence is recovered through $\boldsymbol{x} = \boldsymbol{y} \ominus_q \boldsymbol{b}(s, p)$.



Decoding (Cont'd)

Decoding of $(2, 4)$-Gray code

Gray code ($\boldsymbol{g}$)	Sequence ($\boldsymbol{d}$)	z	s, p	$\boldsymbol{b}(s, p)$
00	00	0	$0, 0$	0000
01	01	1	$0, 1$	1000
02	02	2	$0, 2$	1100
03	03	3	$0, 3$	1110
13	10	4	$1, 0$	1111
12	11	5	$1, 1$	2111
11	12	6	$1, 2$	2211
10	13	7	$1, 3$	2221
20	20	8	$2, 0$	2222
21	21	9	$2, 1$	3222
22	22	10	$2, 2$	3322
23	23	11	$2, 3$	3332
33	30	12	$3, 0$	3333
32	31	13	$3, 1$	0333
31	**32**	**14**	$\mathbf{3, 2}$	**0033**
30	33	15	$3, 3$	0003

Decoding (Cont'd)

Example 4

Consider the decoding of the $(7, 4, 14, 4)$ CW sequence, $\underline{231}3113$.

- The redundant symbol $u = 2$ is dropped. Then the Gray code sequence of length 2, is extracted as 31.
- The Gray code $\boldsymbol{g} = 31$ corresponds to $\boldsymbol{d} = 32$, and index $z = 14$. This implies that $s = 3$ and $p = 2$, therefore $\boldsymbol{b}(3, 2) = 0033$ (presented in the previous table).
- Finally, the information sequence is recovered as

$$\boldsymbol{x} = \boldsymbol{y} \ominus_q \boldsymbol{b}(s, p) = 3113 \ominus_3 0033 = 3120.$$

Cardinality study

- $\mathcal{N}_1$ is the cardinality of q-ary CW sequences for specific W of length n and $\mathcal{N}_2$, the cardinality of q-ary information sequences of length k.
- To construct an (n, k, W, q) CW sequence, one clearly requires enough parity bits r such that $\mathcal{N}_1 \geq \mathcal{N}_2 = q^k$, where $n = k + r$.

W		q	n	k	$\mathcal{N}_1$	$\mathcal{N}_2$
$\beta_{n,q} - q + 1$	3	2	7	4	35	16
	5	2	12	8	792	256
	10	2	21	16	352716	65536
	3	3	5	3	30	27
	10	3	12	9	58278	19683
	6	4	6	4	336	256
$\beta_{n,q}$	4	2	7	4	35	16
	6	2	12	8	924	256
	11	2	21	16	352716	65536
	5	3	5	3	51	27
	12	3	12	9	737789	19683
	9	4	4	6	580	256
$\beta_{n,q} + q$	6	2	8	4	28	16
	9	2	13	8	715	256
	13	2	22	16	497420	65536
	9	3	6	3	50	27
	16	3	13	9	129844	19683
	15	4	7	4	728	256

Redundancy and complexity

- The redundancy is $r = \log_q k + e + 1 \Rightarrow k = q^{r-1-e}$.
- For $e = 1$, the redundancy becomes $r = \log_q k + 2$, which is similar as that presented in [4].
- The addition of the vector $\boldsymbol{u}$ does not change the complexity of this construction compared to the one in [4].
- This method requires $\mathcal{O}(qk \log_q k)$ digit operations for the encoding and $\mathcal{O}(k)$ digit operations for the decoding process.
- Comparison of our scheme with other constructions based on redundancy and complexity can be found in [4].

[4] E. N. Mambou and T. G. Swart, "Encoding and decoding of balanced q-ary sequences using a Gray code prefix," in *Proc. IEEE Int. Symp. Inform. Theory*, Barcelona, Spain, 2016.

Conclusion

- An efficient algorithm was proposed for encoding and decoding (n, k, W, q) CW sequences based on Gray code prefixes with a method to extend the achievable CW range.
- The construction does not make use of memory-consuming lookup tables, and only simple operations such as addition and subtraction are needed.
- The decoding process can be performed mostly in parallel.
- As the proposed method is only applicable to information sequences of length k where $k = q^t$, the improvement would be to extend this algorithm to the case where $k \neq q^t$.

Thanks for your attention!

"We cannot solve our problems with the same thinking we used when we created them." Albert Einstein

QUESTIONS AND COMMENTS!!

This work is based on the research supported in part by the National Research Foundation of South Africa.

NRF National Research Foundation