Reflections on the Gray code

Kamil Łopuszański

September 6, 2015

Abstract

The work aims at gathering theorems about the Gray code. In this paper there are presented a generation algorithm of the Gray code, as well as its relationship with A001511 sequence and flowing from that conclusions. Thanks to a presented theorem, one will be able to find numbers occurring in a subset on a certain position. Nevertheless, a formula for position of a full set $I_n = \{1, 2, \ldots, n\}$ is proved.

Then there is a discussion on non-binary Gray order and extension of previous theorems.

Introduction

Gray code is a binary code, in which 2 consecutive elements differ in exactly one bit. Gray code can be defined in multiple ways. The following, from [2], is one of them:

$$\begin{aligned} \Gamma_0 &= \epsilon \; (\textit{empty string}) \\ \Gamma_{n+1} &= 0\Gamma_n, 1\mathcal{R}(\Gamma_n) \end{aligned}$$

where $\mathcal{R}(\cdot)$ is a reversion operator.

Furthermore, Gray code is equivalent with an order (a sequence) of subsets of $\{1, 2, ...n\}$ where n is the number of bits. Assumed is that 1 encodes an element, and 0 its absence. Decoding a binary string into a subset, we go backwards from the end of the string, i.e. from the least to the most significant bit. Such a sequence of subsets is called **Gray n-order** or **binary Gray n-order**. 0^{th} subset is always an empty set and in this work enumeration always starts from 0.

For example, for $n = 4$

0:	
1:	1
2:	1, 2
3:	2
4:	2, 3
5:	1, 2, 3
6:	1, 3
7:	3
8:	3, 4
9:	1, 3, 4
10:	**1, 2, 3, 4**
11:	2, 3, 4
12:	2, 4
13:	1, 2, 4
14:	1, 4
15:	4

Figure 1: Γ_4 - Gray 4-order

1 Gray order generation algorithm

Gray order generation algorithm has been known for over 30 years and can be found in [1]. Let us define negation of en element e as follows: if $e \in S$ then $S := S \backslash \{e\}$, else $S := S \cup \{e\}$. Suppose z is the first element in the subset S of Gray n-order, $1 \leq z \leq n$. Then, to reach a next subset in an order, one has to perform the following steps:

1. **if** $S = \{n\}$ **then** exit;

2. **if** $|S|$ is odd, **then** negate $z + 1$;

3. **else** $|S|$ is even **then** negate 1;

In every single step we negate a number, i.e. if it is present in the subset, we remove it, and if it is not, we add it. The sequence of positions of negated elements in the successive steps $(n \to \infty)$ give A001511 sequence. It is also called the ruler function $\rho(k)$, defined as follows: $j = \rho(k)$ when j is the highest integral that $2k$ is a multiple of 2^j.

The complexity depends on a structure type that is used to store subsets. The main operation is counting the length of subset, so if a list is used, then the algorithm works in $O(|S|)$, thus $O(n)$. Using an array is more efficient and guaranties the complexity $O(1)$ with respect to n, compare [1].

The whole Gray order can be generated in this fashion, using a *while* loop with a condition: if $S \neq \{n\}$ then compute the next subset. Except from this, there exists also a more efficient, loopless algorithm, described i.a. in [2].

2 Conclusions resulting from a Gray code relationship with A001511 sequence

From the relationship one can deduce the position of any subset in Gray order.
Since 0^{th} subset is an empty string, if p is included in the subset $S \subset \Gamma_n$, $p \leqslant n$, the position i of

S in Γ_{n}, p must appear in A001511 sequence on positions $1 \dots i$ certain odd number of times. See [2] for explanation. It is a base for using Equation 3.

Theorem 1. *Number $p \in \{1, 2, \dots, n\}$, $1 \leqslant p \leqslant n$, belongs to a subset $S \subset \{1, 2, \dots, n\}$ if and only if the corresponding to the subset position belongs to a following generalized sum:*

$$\bigcup_{k \in \mathbb{N}_0} [2^{p-1} + 2k \cdot 2^p, 2^{p-1} + (2k+1)\dot{2}^p) \tag{1}$$

Proof. Let X_p be a set determined with (1).
A001511 sequence can be expressed in recurrence form:

$$\rho(n) = \begin{cases} 1 & \text{for } n \text{ odd} \\ \rho(\frac{n}{2}) + 1 & \text{for } n \text{ even} \end{cases} \tag{2}$$

For $p = 1$: $X_1 = [1, 3) \cup [5, 7) \cup [9, 11) \cup \dots$
We see from (2) that ones are on odd positions and that theorem is true.

Lemma 1. *For $p \geqslant 1$, if theorem is true for p, is also true for $p + 1$*

Proof.

$$\begin{aligned} X_{p+1} &= \bigcup_{k \in \mathbb{N}_0} [2^p + 2k \cdot 2^{p+1}, 2^p + (2k+1)\dot{2}^{p+1}) \\ &= \bigcup_{k \in \mathbb{N}_0} [2 \cdot (2^{p-1} + 2k \cdot 2^p), 2 \cdot (2^{p-1} + (2k+1) \cdot 2^p)) = \{2x : x \in X_p\} \end{aligned}$$

Let $f(p, n)$ be the function of the form:

$$f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$

$f(p, n)$ is a number of elements in the A001511 sequence on such positions $1 \leqslant i \leqslant n$ that $\rho(i) = p$.

$$X_p = \{n : f(p, n) \text{ } is \text{ } odd\} \tag{3}$$

There is a bijection between every two sets A, B of the form:

$$A = \{i : \rho(i) = p\}, B = \{i : \rho(i) = p + 1\}$$

for $p \in \mathbb{N}$:

$$A \ni x \to 2x \in B \tag{4}$$

It is obviously implicated by (2).
It means that if only X_p is the right set for p, X_{p+1} is the right set for $p+1$. So Lemma 1 is proved hereafter. □

By the induction principle, from *(1)* and *(2)* theorem is true for all $p \leqslant n$. □

Example 1. The set $1, 2, 3, 4$ for $n = 4$ is at position 10, since the product of the sets:
For 1: $[1, 3) \cup [5, 7) \cup [9, 11) \cup [13, 17)$
For 2: $[2, 6) \cup [10, 14)$
For 3: $[4, 12)$
For 4: $[8, 24)$
gives the number 10 (cmp. Figure 1).

Generally:

Theorem 2. *The set* $\{1, 2, 3, ..., n\}$ *is located at the position*

$$2^n - 2^{n-1} + \ldots + (-1)^n 2^0 - 1 \text{ for } n \text{ even} \tag{5}$$

or

$$2^n - 2^{n-1} + \ldots + (-1)^n 2^0 \text{ for } n \text{ odd} \tag{6}$$

It can be presented in an easier form. Let us examine a recurrence relation:

$$c_n = -c_{n-1} + 2^n, c_0 = 1; n \geqslant 1$$

The solution of its homogeneous part is

$$c_n^{(h)} = A \cdot (-1)^n$$

And solution of the particular one is:

$$c_n^{(p)} = B \cdot 2^n$$

Afterwards one determines the values of A and B and obtains: $A = \frac{1}{3}$ and $B = \frac{2}{3}$. This leads to the solution of the recurrence relation:

$$c_n = \frac{2^{n+1} + (-1)^n}{3} \tag{7}$$

So we can create shared formula knowing the fact that equations (5) and (6) differ only by -1 at the end:

$$\frac{2^{n+1} + (-1)^n}{3} - \frac{(-1)^n + 1}{2} = \frac{2^{n+2} + 2 \cdot (-1)^n - 3 \cdot (-1)^n - 3}{6} = \frac{2^{n+2} - 3 - (-1)^n}{6}$$

The following formula comes from Jon Stadler. The formula generates A000975 sequence.

Theorem 3. *The set* $I_n = \{1, 2, 3, ..., n\}$ *is located at the position*

$$\psi = \frac{2^{n+2} - 3 - (-1)^n}{6} \tag{8}$$

Proof (of theorems 2, 3). (1) implicates that a remainder r of modulo division $\psi = r \ mod \ 2^{p+1}$, where ψ, $0 \leqslant \psi < 2^n$, is the position of the subset I_n and every $p \in I_n$, satisfies the inequality:

$$2^{p-1} \leqslant r < 2^p + 2^{p-1} \tag{9}$$

In order to prove theorems 2 and 3, it's enough to prove the above inequality for all $p \in I_n$.

The remainder ψ is a sign-alternating sum of exponents with base equal 2. Number of terms in the sum which equals p or $p + 1$ (which depends on a parity of n). Exponents higher than 2^p are omitted since we divide by $mod \ 2^{p+1}$.

And now 2 cases:

I. Suppose p and n are both even or both odd.

$$\begin{array}{rcll} r & = & 2^p - 2^{p-1} + 2^{p-2} + \ldots + (-1)^p - 1 & \text{for } n \text{ even} \\ r & = & 2^p - 2^{p-1} + 2^{p-2} + \ldots + (-1)^p & \text{for } n \text{ odd} \end{array} \tag{10}$$

Furthermore,

$$r = 2^{p-1} + 2^{p-3} + \ldots + (-1)^p \text{ (for } p > 1 \text{ we can omit } -1 \text{ term)}$$

so (9) is obviously satisfied for $p > 1$. What about $p = 1$? p is odd, so n is odd either and formula takes the form $r = 2^1 + (-1)^1 = 2 - 1 = 1$ and $2^0 \leqslant 1 < 2^1 + 2^0$.

II. Suppose either p or n is even (the parity is different)

In order to compute the remainder, we must subtract the alternating sum from 2^{p+1} since the alternating sum itself is now negative (the term with the highest exponent equals $(-1)^{n-1}2^p$).

$$\begin{array}{rcll} r & = & 2^{p+1} - 2^p + 2^{p-1} - 2^{p-2} + \ldots + (-1)^p + 1 & \text{for } n \text{ even} \\ r & = & 2^{p+1} - 2^p + 2^{p-1} - 2^{p-2} + \ldots + (-1)^p & \text{for } n \text{ odd} \end{array} \tag{11}$$

Again, for $p > 1$ we can omit $+1$ term and write simply:

$$r = 2^{p+1} - 2^{p-1} - 2^{p-3} - \ldots - (-1)^{p-1}$$

One can see that:

$$2^{p+1} - 2^{p-1} = 2^p + 2^p - 2^{p-1} = 2^p + 2^{p-1}$$

$$r = 2^p + 2^{p-1} - 2^{p-3} - \ldots - (-1)^{p-1}$$

so (9) is true for $p > 1$. Let us examine $p = 1$ case. As p is odd, n is even. Formula says: $r = 2^2 - 2^1 + (-1)^1 + 1 = 4 - 2 - 1 + 1 = 2$ and of course $1 \leqslant 2 < 3$.

We see that in both cases (9) is satisfied. □

Simple explanation of Theorem 2 is based on the fact that the searched subset is in the second half of first half of second half ... of second half of Gray order.

See also [4].

3 Non-binary mirror Gray code

There won't be presented more than one sort of Gray code, although there exist another non-binary codes like modular Gray code. Nevertheless, one can extend the definition of the mirror Gray code on non-binary strings.

$$\begin{aligned} &\textit{Let } k > 2 \\ &\Gamma_0^k = \varepsilon \ (\textit{empty string}) \\ \Gamma_{n+1}^k = 0\Gamma_n^k, &\underbrace{\mathcal{R}(\mathcal{R}(\ldots(\mathcal{R}}_{k-1}(k\Gamma_n^k), (k-1)\Gamma_n^k), \ldots, 1\Gamma_n^k) \end{aligned}$$

where the symbol $\mathcal{R}(\cdot)$ means reversion of a sequence and Γ_n^k is called **k-digits mirrored Gray n-order**. Where there is no sign between symbols, this should be thought as concatenation of sequences and where is a comma, consecutive subsets are separated.

Again we create an order of subsets and assume that n encodes presence of a digit in the subset and if appropriate bit is less than n, it encodes absence an element.

For the mirrored Gray code theorem Equation 1 might be extensible:

Hypothesis 1. *Number $p \in \{1, 2, \ldots, n\}$ belongs to a subset $S \subset \{1, 2, \ldots, n\}$ if and only if the corresponding to the subset ordinal number belongs to a following generalized sum:*

$$\bigcup_{j\in\mathbb{N}_0} [(k-1)k^p + k(k-1)j \cdot k^p, (k-1)k^p + (kj+1)(k-1) \cdot k^p) \tag{12}$$

We do not give any proof of the above hypothesis. Despite that, it would be probably analogous to proof pertaining to binary code. What makes the only difference is the ruler function which would be modified and defined in following fashion: $j = \rho_k(m)$ when j is the highest integer that $k \cdot m$ is a multiple of k^j.

Example for $k = 3$, $n = 3$

0:	0 0 0
1:	1 0 0
2:	2 0 0
3:	2 1 0
4:	1 1 0
5:	0 1 0
6:	0 2 0
7:	1 2 0
8:	2 2 0
9:	2 2 1
10:	1 2 1
11:	0 2 1
12:	0 1 1
13:	1 1 1
14:	2 1 1
15:	2 0 1
16:	1 0 1
17:	0 0 1
18:	0 0 2
19:	1 0 2
20:	2 0 2
21:	2 1 2
22:	1 1 2
23:	0 1 2
24:	0 2 2
25:	1 2 2
26:	2 2 2

Figure 2: Γ_3^3 - 3-digits Gray 3-order

For example, for $p = 2$: $X_1^3 = [2 \cdot 3, 2 \cdot 3 + 2 \cdot 3^1) = [6, 12) \cup [24, 30) \cup \cdots$

References

[1] Albert Nijenhuis Herbert S.Wilf. *Combinatorial Algorithms for Computers and Calculators.* Academic Press, 1978.

[2] Donald E. Knuth. *Generowanie wszystkich krotek i permutacji.* Warszawa, 2007.

[3] Gonzalo Navarro. *A Guided Tour to Approximate String Matching.* Dept. of Computer Science, University of Chile, Blanco Encalada 2012 - Santiago - Chile.

[4] Robert L. Lamphere. *A Recurrence Relation in the Spinout Puzzle.* "The College Mathematics Journal", 1996