Error Correcting Codes

1 Codes correcting a single symmetric error

In this section we consider the case when only a single bit can be corrupted during transmission. In this case the bit value will be negated. We assume that the codewords are binary sequences $x_1x_2\ldots x_n$ of length n for some fixed n. Let l be a number such that

$$2^{l-1} \leq n < 2^l.$$

In other words, $l = \lfloor \log n \rfloor + 1$. Any integer i from the interval $[0, n)$ can be represented in the binary system by using l bits. Denote by $e_l(i)$ the binary word of length l which is the representation of i.

Furthermore, for a binary word $X = x_1x_2\ldots x_n$ denote

$$H(X) = \sum_{i=1}^{n} x_i \, e_l(i). \qquad (1)$$

Evidently, $H(X)$ is a binary word of length l obtained by summing up component-wise modulo 2 some binary words of length l that correspond to the 1's in X. Consider the code H_n defined by

$$H_n = \{X = x_1x_2\ldots x_n \mid H(X) = (\underbrace{00\ldots0}_{l})\}.$$

Example 1 *Let* $n = 6$*,* $X = 010101$ *and* $Y = 110100$*. Then* $l = 3$*,* $H(X) = 010 \oplus 100 \oplus 110 = 000$ *and* $H(Y) = 001 \oplus 010 \oplus 100 = 111$*. Hence,* $X \in H_n$ *and* $Y \notin H_n$*.*

For a binary word $X = x_1x_2\ldots x_n$ denote by $N(X)$ the decimal number, whose binary expansion is $x_1x_2\ldots x_n$. For example, $N(101) = 5$. Assume that a codeword $X \in H_n$ is sent and a word Y (of the same length) is received. If the j-th symbol was corrupted, $Y = x_1x_2\ldots x_{j-1}(x_j \oplus 1)x_{j+1}\ldots x_n$. One has

$$H(Y) = \sum_{i=1}^{n} x_i \, e_l(i) \oplus e_l(j) = H(X) + e_l(j) = e_l(j),$$

since $H(X) = 00\ldots0$. The sent word X can be restored by flipping the bit of Y with index $N(H(Y)) = N(e_l(j)) = j$.

Example 2 *Let* $X = 010101 \in H_6$ *is sent and* $Y = 010111$ *is received. Since* $H(Y) = 010 \oplus 100 \oplus 101 \oplus 110 = 101$*, the symbol with index 5 is corrupted.*

The code H_n was designed by Hamming. For the number of codewords, one has $|H_n| = 2^{n-l}$. Since $l = \lfloor \log n \rfloor + 1 = \lceil \log(n+1) \rceil$, one has

$$\frac{2^{n-1}}{n} \leq |H_n| = 2^{n-\lceil \log(n+1) \rceil} \leq \frac{2^n}{n+1}.$$

In particular, $|H_6| = 8$. The code H_6 is presented in Table 1.

H_6	W_6	$W_{6,12}$	N_6
000000	000000	000000	000000
111000	100001	100011	001101
110011	010010	010101	011010
001011	001100	001110	100111
101101	110100	111001	110100
010101	001011	110110	
011110	110011		
100110	101101		
	011110		
	111111		

Table 1: Some binary codes

Note that $|H_n| = \frac{2^n}{n+1}$ if and only if n is of the form $2^k - 1$ for some $k > 1$. In this case the set of all binary words can be partitioned into balls of radius 1 around the codewords of H_n.

2 Codes correcting a single substitution error

Assume that only a single zero in the transmitted word can be substituted with 1 during the transmission. For a binary word $X = x_1 x_2 \ldots x_n$ denote

$$W(X) = \sum_{i=1}^{n} x_i \cdot i.$$

Obviously, $W(X)$ is the sum of indices of non-zero bits of X. For a given k define the code $W_{n,k}$ as

$$W_{n,k} = \{X = x_1 x_2 \ldots x_n \mid W(X) \equiv 0 \pmod{k}\}, \tag{2}$$

and put $W_n = W_{n,n+1}$.

Example 3 *Let* $n = 6$*,* $X = 110100$ *and* $Y = 010101$*. Then* $W(X) = 1 + 2 + 4 = 7$*,* $W(Y) = 2 + 4 + 6 = 12$*. Hence,* $X \in W_6$ *and* $Y \notin W_6$*.*

We show that the code $W_{n,k}$ for $k \geq n+1$ (in particular, the code W_6) is a code correcting a single error of the type $0 \to 1$. Assume that a codeword X was sent, a word Y is received, and at most one error occurred during the transmission. Clearly, $W(Y) = W(X)$ in case of no error, and $W(Y) = W(X) + j$ if the j-th bit is corrupted. Since $W(X) \equiv 0 \pmod{k}$, in the last case one has $W(Y) \equiv j \pmod{k}$. This allows to figure out the index of the corrupted bit.

Example 4 *Let $X = 110100 \in W_6$ was sent and $Y = 110101$ is received. Since $W(Y) = 1 + 2 + 4 + 6 = 13$ and $13 \equiv 6 \pmod{7}$, the bit number 6 is corrupted.*

The code W_n is constructed by Varshamov and Tennenholz. One can show that

$$|W_n| = \frac{1}{2(n+1)} \sum_{\substack{d|n+1 \\ d \text{ is odd}}} \phi(d) \cdot 2^{(n+1)/d}$$

where $\phi(d)$ is the number of numbers i in the interval $[0, d]$ that are relatively prime with d, that is, $\gcd(i, d) = 1$ (Euler function). In particular, $|W_6| = \frac{2^7+6\cdot 2}{14} = 10$. The code W_6 is shown in Table 1.

3 Codes correcting a single deletion or insertion

Here we assume that at most one bit can be dropped from a codeword during the transmission. We show that the code $W_{n,k}$ with $k \geq n + 1$ can correct a single error of this type.

Assume that a single bit is dropped from a codeword $X \in W_{n,k}$ during the transmission and a word $Y = y_1 y_2 \dots y_{n-1}$ is received. Denote by n_1 (respectively, n_0) the number of ones (respectively, zeros) located to the right of the dropped bit in X and $W(Y) = \sum_{i=1}^{n-1} y_i \cdot i$.

Note, that is a zero is dropped at position j, then each of the ones to the right of position j (whose number is n_1) will contribute one less to the sum. Therefore, $W(X) - W(Y) = n_1$. If a one is dropped at position j, the entire sum will additionally decrease on j units, so $W(X) - W(Y) = j + n_1 = n - n_0$ (because $n_0 + n_1 = n - j$). Obviously, in either case $0 \leq W(X) - W(Y) \leq n < k$.

Denote $\Delta Y = k - W(Y)$. Since $W(X) \equiv 0 \pmod{k}$, one has $W(X) - W(Y) = \Delta Y$. Taking into account

$$n_1 \leq \|Y\| \leq (n-1) - n_0 < n - n_0,$$

comparing ΔY and $\|Y\|$ one can figure out what symbol (0 or 1) was dropped during the transmission. Namely, if $\Delta Y \leq \|Y\|$, then 0 was dropped, and to restore the sent codeword X one should insert a 0 in Y at position j so that there is ΔY ones to the right of j. Similarly, if $\Delta Y > \|Y\|$, one should insert a 1 at position l so that there is $n - \Delta Y$ zeros to the right of l.

Example 5 *Let* $X = 110100 \in W_6$ *was sent and* $Y = 10100$ *is received (after dropping the first symbol from* X*). One has* $\|Y\| = 2$, $W(Y) = 4$, *hence* $\Delta Y = 3$*. Since the condition* $\Delta Y > \|Y\|$ *is satisfied, count* $n - \Delta Y = 3$ *zeros from the right of* Y *and insert a 1 there. Note, that we insert a 1 at a different position (position 2 in this case), but the obtained this way word is equal to the one being sent.*

It turns out that any code correcting s or less deletions is at the same time a code correcting s or less insertions. It is leaved as an exercise to figure out how to restore the codeword of W_n after a single insertion.

4 Codes correcting a single arithmetic error

Arithmetic errors during the transmission lead to adding or subtracting a power of 2 to/from the codeword. Consider a code N_n consisting of all binary words $X = x_1 x_2 \dots x_n$ such that

$$N(X) = \sum_{i=1}^{n} x_i 2^{n-i} \equiv 0 \pmod{2n+1}. \tag{3}$$

If a codeword $X \in N_n$ was sent and a single arithmetic error of the type $\pm 2^i$ occurred, the received word satisfies the condition

$$N(Y) = N(X) \pm 2^i.$$

Hence, $N(Y) \equiv \pm 2^i \pmod{2n+1}$. Therefore, the code N_n can correct a single arithmetic error if the numbers

$$1, 2, \dots, 2^{n-1}, -1, -2, \dots, -2^{n-1} \tag{4}$$

are all distinct and nonzero mod $2n + 1$.

We show that this condition is satisfied in the following two cases:

a. The number $p = 2n + 1$ is prime and 2 is a primitive root modulo p. This means that all the numbers

$$1, 2, \dots, 2^{n-1}, 2^n, \dots, 2^{2n-1} \tag{5}$$

are pairwise distinct modulo p.

b. The number $p = 2n + 1$ is prime, 2 is not a primitive root modulo p, and -2 is a one. This means that all the numbers

$$1, -2, 2^2, -2^3, \dots, 2^{2n-2}, -2^{2n-1} \tag{6}$$

are all distinct modulo p.

Indeed, if $p = 2n + 1$ is prime then, by the little Fermat theorem, $2^{p-1} - 1 = 2^{2n} - 1 \equiv 0 \pmod{p}$. This implies $(2^n+1)(2^n-1) \equiv 0 \pmod{p}$. Therefore, either $2^n \equiv -1 \pmod{p}$ or $2^n \equiv 1 \pmod{p}$.

If 2 is a primitive root modulo p then $2^n \not\equiv 1 \pmod{p}$. Hence, $2^n \equiv -1 \pmod{p}$. Then the set of numbers (5) is equal to the set (4), and the required condition is satisfied.

On the other hand, if 2 is not a primitive root modulo p, but -2 is a one, then n is odd. Indeed, if n would be even, then the fact that -2 is a primitive root modulo p implies $2^n = (-2)^n \not\equiv 1 \pmod{p}$. Hence, $2^n \equiv -1 \pmod{p}$. But then the numbers set (6) is the same as

$$1, 2^{n+1}, 2^2, 2^{n+3}, \ldots, 2^{2n-1}, 2^n, 2, 2^{n+2}, 2^3, \ldots, 2^{2n-2}, 2^{n-1},$$

which implies 2 is a primitive root modulo p. Therefore, n is odd and $(-2)^n = -2^n \not\equiv 1 \pmod{p}$, hence $2^n \equiv 1 \pmod{p}$. But then the set of numbers (6) is the same as

$$1, -2, 2^2, -2^3, \ldots - 2^{n-2}, 2^{n-1}, -1, 2, -2^2, \ldots, 2^{n-2}, -2^{n-1}$$

and the required condition is also satisfied in case b).

Therefore, if the number n satisfies one of the conditions a) or b), then the code N_6 corrects a single arithmetic error.

Example 6 *The condition a) is satisfied for $n = 6$, and the positive residues modulo 13 of the numbers*

$$1, 2, 2^2, 2^3, 2^4, 2^5, -1, -2, -2^2, -2^3, -2^4, -2^5 \tag{7}$$

are equal, respectively, to

$$1, 2, 4, 8, 3, 6, 12, 11, 9, 5, 10, 7. \tag{8}$$

Note that $X = 110100 \in N_6$, because $N(X) = 2^5 + 2^4 + 2^2 = 52 \equiv 0 \pmod{13}$. Assume a single arithmetic error -2^3 occurred by transmission of the word X, so the received word is $Y = 101100$. Since $N(Y) = 2^5 + 2^3 + 2^2 = 44 \equiv 5 \pmod{13}$, take the number of the set (7) corresponding to the number 5 of the set (8). This number is -2^3, which equals the arithmetic error.

The code N_n was discovered by Brown. It is easily seen that $|N_n| = \left\lceil \frac{2^n}{2n+1} \right\rceil$.