Linear Codes for an Effective Quantization of Data

Vladimir N. Potapov

Sobolev Institute of Mathematics

Problems of Redundancy in Information and Control Systems

Saint–Petersburg, Russia in 26–29 September 2016

Introduction

Quantization is the important stage for lossy compression of real data (image, speech). From the nature of the things a part of data values is on the edges of the quantization intervals. The last bit of such value is the least significant one for the quality of quantization. A special method for choosing this least significant bits is used for data hiding in image and video. It is possible to utilize this redundancy for data compression.

Consider an n-tuple consisted of the last bits of quantized values. Suppose that each n-tuple contains negligible bits. Let C be some code with cardinality 2^{n-k} and let for each n-tuple there exists a codeword such that this n-tuple and the codeword differ only in negligible bits. We will transmit the codeword (rather its number) instead of the initial n-tuple. So, we will truncate k bits of the n-bit message.

Definitions

A subset T of the hypercube is called a binary covering array $\mathrm{CA}(|T|, n, k)$ with strength k if for each $v \in F_2^n$ and for any k positions there exists $u \in T$ such that v and u can differ only in these k positions.

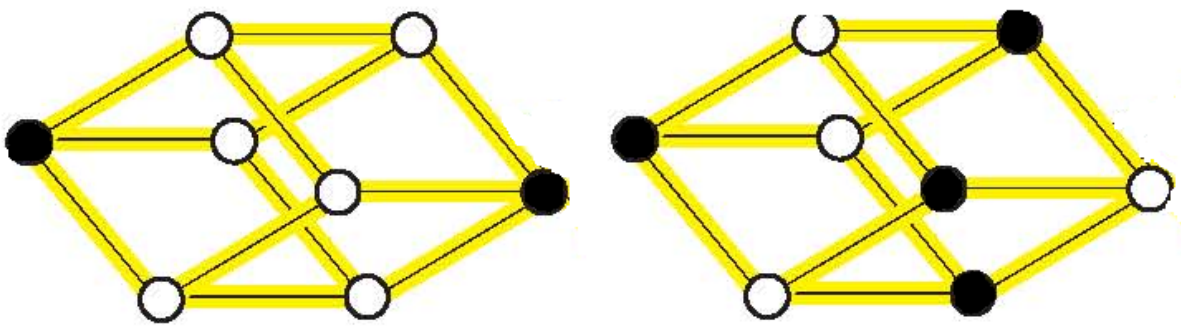
We will consider a bit different mathematical problem: to construct a partial covering array $S_k \subset F_2^n$, $|S_k| = 2^{n-k}$ with the following property: the number of k-faces containing elements of S_k is as large as possible.

Examples

A perfect solution for this problem would be a set of codewords containing only one element of each k-face of F_2^n. Such sets have cardinalities 2^{n-k} are called MDS codes.

In the Boolean hypercube, there exist only two nonequivalent MDS codes: the parity check code ($k = 1$) and the pair of antipodal vectors ($k = n - 1$).

$\{000, 111\}$ $\{000, 011, 101, 110\}$



Results

Denote by $\nu_k(S)$ the ratio between the number of k-faces that contain elements of S and the number $\binom{n}{k}2^{n-k}$ of all k-faces. Define $\nu_k(n) = \max \nu_k(S)$ where $S \subset F_2^n$, $|S| = 2^{n-k}$.

Proposition 1

$\lim\limits_{n\to\infty} \mathrm{E}\nu_k(T) = 1 - 1/e$, where $T \subset F_2^n$ is a random set, $|T| = 2^{n-k}$, and k is fixed.

Results

Theorem 1

Let $r \leq k$ be even. Then $\nu_k(n) \leq 1 - \frac{1+o(1)}{2^{k+1}}\binom{k}{r}$ as $n \to \infty$.

Let $r(k)$ be the nearest even number for k. Then $\frac{1}{2^{k+1}}\binom{k}{r(k)} = \frac{1+o(1)}{\sqrt{2\pi k}}$ as $k \to \infty$. For example, $\lim_{n\to\infty} \nu_3(n) \leq \frac{13}{16}$.

Results

Denote by $C_{k,m}$ a code with the check matrix H of size $k \times m(2^k - 1)$ consisting of m columns $b_k(j)$ for all $j = 1, \ldots, 2^k - 1$, where $b_k(j)$ is the binary representation of j. If $m = 1$ then $C_{k,1}$ is the Hamming code.

$(111\ldots1)$; $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$; $\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$;

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

are the check matrices for $C_{1,m}$, $C_{2,1}$, $C_{2,2}$ and $C_{3,1}$ respectively.

Results

Denote by $\mu_k(C)$ the number of k-faces containing only one element $\overline{0}$ of a linear code C. It is clear that $\nu_k(C) \geq \mu_k(C)/\binom{n}{k}$.

Theorem 2

For fixed k and $n = m(2^k - 1)$ the maximum value of $\mu_k(C_{k,m})$ corresponds to the code $C_{k,m}$.

Results

Proposition

Let $n = 2^k - 1$, $k \geq 2$ then
$\mu_k(C_{k,1}) = (2^k - 1)(2^k - 2)(2^k - 4) \cdots (2^k - 2^{k-1})/k!$ and
$\nu_k(C_{k,1}) =$
$$\frac{1}{k!2^k\binom{n}{k}} \sum_{t=1}^{k-1} (2^k - 1)(2^k - 2)(2^k - 4) \cdots (2^k - 2^t)\binom{2^{t+1}-t-2}{k-t-1}2^{t+1}.$$

It is possible to calculate that $\nu_2(C_{2,1}) = 1$, $\nu_3(C_{3,1}) = 9/10$, $\nu_4(C_{4,1}) = 10/13$.

Results

Define $\nu'_k(C) = \mu_k(C)/\binom{n}{k}$. Obviously, $\nu'_k(C)$ is a lower bound of $\nu_k(C)$. Then $\nu'_k(C_{k,1}) = (2^k-1)(2^k-2)(2^k-4)\cdots(2^k-2^{k-1})/(2^k-1)(2^k-2)(2^k-3)\cdots(2^k-k)$ and $\lim_{k\to\infty} \nu'_k(C_{k,1}) \approx 0.2888$.

Proposition

$\lim_{s\to\infty} \nu'_k(C_{k,s}) = (2^k-1)(2^k-2)(2^k-4)\cdots(2^k-2^{k-1})/(2^k-1)^k$.

Conclusion

Linear code is not better than a random set as k is large. But if k is a small integer than the best linear code is tight to the best unrestricted code.

Problems to find the best unrestricted code or to find an asymptotic of its cardinality are open.