Error Correcting Codes: Combinatorics, Algorithms and Applications **(Fall 2007)**

Lecture 25: Justesen Codes

October 26, 2007

Lecturer: Atri Rudra *Scribe: Kanke Gao*

In the last lecture, we introduced code concatenation, where we compose an outer code C_{out} with an inner code C_{in}. We derived the Zyablov bound by picking C_{out} on the Singleton bound and C_{in} on the GV bound. We also presented a polynomial time construction of a code that achieves the Zyablov bound (and hence, an asymptotically good code). A somewhat unsatisfactory aspect of this construction was the brute force search for a suitable inner code (which lead to the polynomial construction time). In today's lecture, we will study a strongly explicit construction of an asymptotically good code.

1 Strongly explicit construction

A polynomial time construction of an asymptotically good code was presented in the last lecture. A natural followup question is if we can have a strongly explicit construction. Technically speaking, by strongly explicit construction, we mean a log space construction. However, we will not formally define this notation. We will now consider the so called *Justesen code* [1]. Justesen code is concatenation code with *different* linear inner codes, which is composed of an $(N, K, D)_{q^k}$ outer code C_{out} and different $(n, k, d)_q$ inner codes $C_{in}^i : 1 \leq i \leq N$. Formally, the concatenation of these codes, denoted by $C_{out} \circ (C_{in}^1, \ldots, C_{in}^N)$, is defined as follows. Given a message $\mathbf{m} \in [q^k]^K$, let the outer codeword be denoted by $(c_1, \ldots, c_N) \overset{def}{=} C_{out}(\mathbf{m})$. Then $C_{out} \circ (C_{in}^1, \ldots, C_{in}^N)(\mathbf{m}) = (C_{in}^1(c_1), C_{in}^2(c_2), \ldots, C_{in}^n(c_N))$.

We will need the following result.

Theorem 1.1. *Let $\varepsilon > 0$. There exists an ensemble of inner codes $C_{in}^1, C_{in}^2, \ldots, C_{in}^N$ of rate $\frac{1}{2}$, where $N = q^k - 1$, such that for at least $(1-\varepsilon)N$ values of i, C_{in}^i has relative distance $\geq H_q^{-1}(\frac{1}{2} - \varepsilon)$.*

In fact, this ensemble is the following. For $\alpha \in \mathbb{F}_{q^k}^*$, the inner code $C_{in}^\alpha : \mathbb{F}_q^k \to \mathbb{F}_q^{2k}$, is defined as $C_{in}^\alpha(x) = (x, \alpha x)$. This ensemble is due to Wozencraft and is called the *Wozencraft ensemble*. It is easy to check that C_{in}^α for every $\alpha \in \mathbb{F}_{q^k}^*$ is linear.

2 Justesen code

For the Justesen code, the outer code C_{out} is Reed-Solomon code over $\mathbb{F}_{q^k}$ evaluated over $\mathbb{F}_{q^k}^*$ of rate R, $0 < R < 1$. The outer code C_{out} have relative distance $\delta_{out} = 1 - R$ and block length of $N = q^k - 1$. The set of inner codes is the Wozencraft ensemble $\{C_{in}^\alpha\}_{\alpha \in \mathbb{F}_{q^k}^*}$. So Justesen code is the concatenated code $C^* \overset{def}{=} C_{out} \circ (C_{in}^1, C_{in}^2, \ldots, C_{in}^N)$ with the rate $\frac{R}{2}$. The following proposition estimates the distance of C^*.

Proposition 2.1. *Let $\varepsilon > 0$. C^* has relative distance at least $(1-R-\varepsilon)H_q^{-1}(\frac{1}{2}-\varepsilon)$*

Proof. Consider $\mathbf{m}^1 \neq \mathbf{m}^2 \in (\mathbb{F}_{q^k})^K$. By the distance of outer codes $|\mathbf{S}| \geq (1-R)N$, where $\mathbf{S} = \{i | C_{out}(\mathbf{m}^1)_i \neq C_{out}(\mathbf{m}^2)_i\}$. Call the ith inner code "good", if C_{in}^i has distance at least $d \overset{def}{=} H_q^{-1}(\frac{1}{2}-\varepsilon) \cdot 2k$. Otherwise, the inner code is considered bad. Note that by Theorem 1.1, there are at most εN bad inner codes. Let $\mathbf{S}_g$ be the set of all good inner codes in $\mathbf{S}$, while $\mathbf{S}_b$ is the set of all bad inner codes in $\mathbf{S}$. Since $\mathbf{S}_b \leq \varepsilon N$,

$$|\mathbf{S}_g| = |\mathbf{S}| - |\mathbf{S}_b| \geq (1-R-\varepsilon)N. \tag{1}$$

For each good $i \in \mathbf{S}$

$$\Delta(C_{in}^i(C_{out}(\mathbf{m}^1)_i), C_{in}^i(C_{out}(\mathbf{m}^2)_i)) \geq d. \tag{2}$$

Finally, from (1) and (2), we obtain that the distance of C^* is at least

$$(1-R-\varepsilon)Nd = (1-R-\varepsilon)H_q^{-1}(\frac{1}{2}-\varepsilon)N \cdot 2k,$$

as desired. □

Since the Reed-Solomon codes as well as the Wozencraft ensemble are strongly explicit, the above result implies the following:

Corollary 2.2. *The concatenated code C^* is an asymptotically good code and has a strongly explicit construction.*

Thus, we have now satisfactorily answered the question of whether explicit asymptotically good (binary) codes exist modulo Theorem 1.1, which we prove next.

Proof of Theorem 1.1. Fix $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2) \in \mathbb{F}_q^{2k} \setminus \{\mathbf{0}\}$. Note that this implies that $\mathbf{y}_1 = \mathbf{0}$ and $\mathbf{y}_2 = \mathbf{0}$ are not possible. We claim that $\mathbf{y} \in C_{in}^{\alpha}$ for at most one $\alpha \in \mathbf{F}_{2^k}^*$. The proof is by a simple case analysis.

- Case 1: $\mathbf{y}_1 \neq \mathbf{0}$ and $\mathbf{y}_2 \neq \mathbf{0}$, then $\mathbf{y} \in C_{in}^{\alpha}$, where $\alpha = \frac{\mathbf{y}_2}{\mathbf{y}_1}$.
- Case 2: $\mathbf{y}_1 \neq \mathbf{0}$ and $\mathbf{y}_2 = \mathbf{0}$, then $\mathbf{y} \notin C_{in}^{\alpha}$ for every $\alpha \in \mathbb{F}_{2^k}^*$(as $\alpha\mathbf{y}_1 \neq \mathbf{0}$).
- Case 3: $\mathbf{y}_1 = \mathbf{0}$ and $\mathbf{y}_2 \neq \mathbf{0}$, then $\mathbf{y} \notin C_{in}^{\alpha}$ for every $\alpha \in \mathbb{F}_{2^k}^*$(as $\alpha\mathbf{y}_1 = \mathbf{0}$).

Now assume that $wt(\mathbf{y}) < H_q^{-1}(1-\varepsilon)n$. Note that if $\mathbf{y} \in C_{in}^{\alpha}$, then C_{in}^{α} is "bad"(i.e. has relative distance $< H_q^{-1}(\frac{1}{2}-\varepsilon)$). Since $\mathbf{y} \in C_{in}^{\alpha}$ for at most one value of α, the total number of bad codes is at most

$$\begin{aligned}|\{\mathbf{y} | wt(\mathbf{y}) < H_q^{-1}(\frac{1}{2}-\varepsilon) \cdot 2k\}| &\leq Vol_q(\mathbf{0}, H_q^{-1}(\frac{1}{2}-\varepsilon) \cdot 2k) \\ &\leq q^{H_q(H_q^{-1}(\frac{1}{2}-\varepsilon)) \cdot 2k} \end{aligned} \tag{3}$$

$$
\begin{aligned}
&= q^{(\frac{1}{2}-\varepsilon)\cdot 2k} \\
&= \frac{q^k}{q^{2\varepsilon k}} \\
&< \varepsilon(q^k - 1) && (4) \\
&= \varepsilon N. && (5)
\end{aligned}
$$

In the above, (3) follows from our good old upper bound on the volume of a Hamming ball while (4) is true for large enough k. Thus for at least $(1-\varepsilon)N$ values of α, C_{in}^{α} has relative distance at least $H_q^{-1}(\frac{1}{2}-\varepsilon)$, as desired. □

Concatenating an outer code of distance D and an inner code of distance d, we can obtain a code of distance at least $\geq Dd$ (Dd is called its *design distance*). For asymptotically good codes, we have obtained polynomial time construction of such codes, as well as strongly explicit construction of similar codes. Further, since these codes were linear, we also get polynomial time encoding. However, the following natural question about decoding still remains unanswered.

Question 2.3. *Can we decode concatenated codes up to half their design distance in polynomial time?*

We will study this question in the next lecture.

References

[1] J. Justesen. Class of constructive asymptotically good algebraic codes. *IEEE Trans. Inform. Theory*, pages 652–656, Sep 1972.