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

Lecture 18: Johnson Bound

October 3, 2007

Lecturer: Atri Rudra *Scribe: Yang Wang & Atri Rudra*

In the last lecture, we started with some definitions related to the so called Johnson bound. Today, we will state and prove the Johnson bound.

1 Johnson bound

Let $J_q(n, d, e)$ be the maximum number of codewords in a Hamming ball of radius e for any code $C \subseteq [q]^n$ of distance d. Then obviously,

$$J_q\left(n, d, \left\lfloor \frac{d-1}{2} \right\rfloor\right) = 1.$$

Notice that if we can show $J_q(n, d, e) = n^{O(1)}$ for some $e > \lfloor \frac{d-1}{2} \rfloor$, list decoding is possible for any code of distance d up to e errors. We will show how to prove the Johnson bound for binary case. Before the proof, let's set up some notations.

Definition 1.1. $J_q(\delta) = (1 - \frac{1}{q})\left(1 - \sqrt{1 - \frac{q}{q-1}\frac{d}{n}}\right)$

As we will see later in these notes, $J_q(\delta) > \delta/2$ for every $0 < \delta < 1 - 1/q$.

Theorem 1.2 (Johnson bound)**.** *If* $\frac{e}{n} < \frac{q-1}{q}\left(1 - \sqrt{1 - \frac{q}{q-1}\frac{d}{n}}\right)$*, then* $J_q(n, d, e) \leq qnd$*.*

Proof. : We will only prove the Johnson bound for $q = 2$. The proof technique that we will use has a name: double counting.

We have to show that for every binary code $C \subseteq \{0, 1\}^n$ with distance d (i.e. for every $\mathbf{c_1} \neq \mathbf{c_2} \in C$, $\Delta(\mathbf{c_1}, \mathbf{c_2}) \geq d$) and every $\mathbf{y} \in \{0, 1\}^n$, $|B_2(\mathbf{y}, e) \bigcap C| \leq 2nd$.

Fix arbitrary C and $\mathbf{y}$. Let $\mathbf{c_1}, \ldots, \mathbf{c_M} \in B_2(\mathbf{y}, e)$. We need to show that $M \leq 2nd$. Define $\mathbf{c_i}' = \mathbf{c_i} - y$ for $1 \leq i \leq M$, then

(i) $wt(\mathbf{c_i}') \leq e$ for $1 \leq i \leq M$ because $\mathbf{c_i} \in B_2(\mathbf{y}, e)$.

(ii) $\Delta(\mathbf{c_i}', \mathbf{c_j}') \geq d$ for every $i \neq j$, because $\Delta(\mathbf{c_i}, \mathbf{c_j}) \geq d$.

Define

$$S = \sum_{i<j} \Delta(\mathbf{c_i}', \mathbf{c_j}').$$

Then from (ii), we have

$$S \geq \binom{M}{2} d \tag{1}$$

Take each $\mathbf{c_i}'$ as a row vector and we get the $n \times M$ matrix $(\mathbf{c_1}'^T, \cdots, \mathbf{c_M}'^T)$. Define m_i as the number of 1's in the i-th row for $1 \leq i \leq n$. Then the i-th row of the matrix contributes the value $m_i(M - m_i)$ to S because we can find this number of 0-1 pairs in that row.

Set $\overline{e} = \sum_i m_i/M$. From (i) above, we have $\overline{e} \leq e$. Using the Cauchy-Schwartz inequality (i.e., $\langle \mathbf{x}, \mathbf{y} \rangle \leq ||\mathbf{x}|| \cdot ||\mathbf{y}||$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$) by taking $\mathbf{x} = (m_1, \cdots, m_n)$, $\mathbf{y} = (1/n, \cdots, 1/n)$,

$$\left(\frac{\sum m_i}{n}\right)^2 \leq \left(\sum m_i^2\right)\frac{1}{n}. \tag{2}$$

Thus,

$$S = \sum_{i=1}^{n} m_i(M - m_i) = M^2\overline{e} - \sum_{i=1}^{n} m_i^2 \leq M^2\overline{e} - \frac{(M\overline{e})^2}{n} = M^2(\overline{e} - \frac{\overline{e}^2}{n}), \tag{3}$$

where the inequality follows from (2). By (1) and (3),

$$M^2\left(\overline{e} - \frac{\overline{e}^2}{n}\right) \geq \frac{M(M-1)}{2}d,$$

which implies that

$$\begin{aligned} M \leq \frac{dn}{dn - 2n\overline{e} + 2\overline{e}^2} = \frac{2dn}{2dn - n^2 + n^2 - 4n\overline{e} + 4\overline{e}^2} &= \frac{2dn}{(n - 2\overline{e})^2 - n(n-2d)} \\ &\leq \frac{2dn}{(n-2e)^2 - n(n-2d)}, \end{aligned}$$

where the last inequality follows from the fact that $\overline{e} \leq e$. Then from

$$\frac{e}{n} < \frac{1}{2}\left(1 - \sqrt{1 - \frac{2d}{n}}\right),$$

We get

$$n - 2e > \sqrt{n(n-d)}.$$

In other words

$$(n-2e)^2 > n(n-d).$$

Thus, $(n-2e)^2 - n(n-2d) \geq 1$ because n, e are all integers and we have $M \leq 2dn$. □

We now digress a bit to prove the following property of the function $J_q(\cdot)$:

Lemma 1.3. *Let $q \geq 2$ be an integer and let $0 \leq x \leq 1 - \frac{1}{q}$. Then the following inequalities hold:*

$$J_q(x) \geq 1 - \sqrt{1-x} \geq \frac{x}{2}.$$

Proof. We start with by proving the inequality

$$J_q(x) = \left(1 - \frac{1}{q}\right)\left(1 - \sqrt{1 - \frac{xq}{q-1}}\right) \geq 1 - \sqrt{1-x}.$$

Indeed, both the LHS and RHS of the inequality are zero at $x = 0$. Further, it is easy to check that the derivatives of the LHS and RHS are $\frac{1}{\sqrt{1-\frac{xq}{q-1}}}$ and $\frac{1}{\sqrt{1-x}}$ respectively. Finally, the latter is always larger than the former quantity, which implies that the LHS increases more rapidly than the LHS, which in turn proves the required inequality.

The second inequality follows from the subsequent relations. As $x \geq 0$,

$$1 - x + \frac{x^2}{4} \geq 1 - x,$$

which implies that

$$\left(1 - \frac{x}{2}\right)^2 \geq 1 - x,$$

which in turn implies the required inequality. □

Remark 1.4. *Theorem 1.2 and Lemma 1.3 imply that for* any *code, list decoding can potentially correct strictly more errors than unique decoding in polynomial time, as long as q is at most some polynomial in n (which will be true of all the codes that we discuss in this course).*

Theorem 1.2 and Lemma 1.3 also implies the following "alphabet free" version of the Johnson bound.

Theorem 1.5 (Alphabet free version)**.** *If $e \leq n - \sqrt{n(n-d)}$ then $J_q(n, d, e) \leq qnd$ for all the q.*

A natural question to ask is the following:

Question 1.1. *Is the Johnson bound tight?*

The answer is yes in the sense that there exist linear codes with relative distance δ such that we can find Hamming ball of radius larger than $J_q(\delta)$ with super-polynomially many codewords [1, 2].

On the other hand, it is not tight in the following sense. Note that by the Singleton bound, the Johnson bound implies that for any code one can hope to list decode from about $p \leq 1 - \sqrt{R}$ fraction of errors. However, this tradeoff between p and R is not tight: recall that for large q, the list decoding capacity is $1 - R > 1 - \sqrt{R}$. Figure 1 plots and compares the relevant trade-offs.

References

[1] Venkatesan Guruswami. Limits to list decodability of linear codes. *STOC 2002*, 802 - 811.

[2] Venkatesan Guruswami, Igor Shparlinski. Unconditional Proof of Tightness of Johnson Bound. *SODA 2003*, 754 - 755.

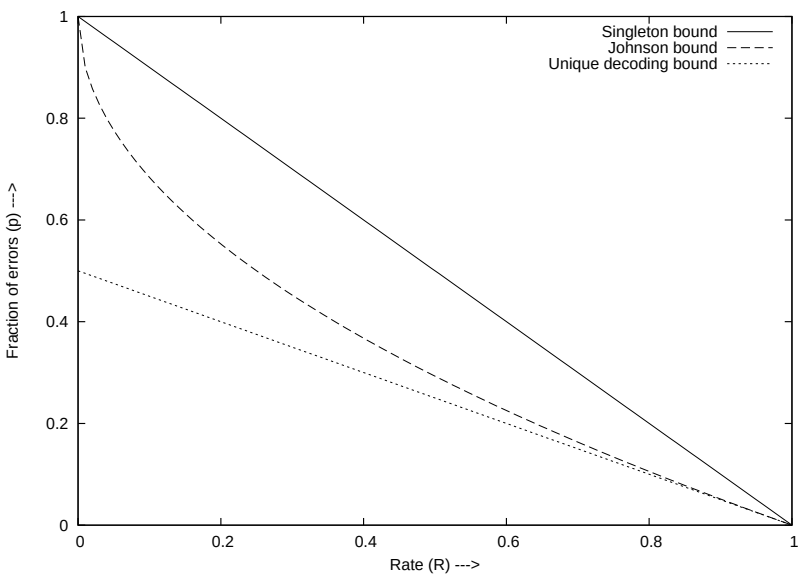

Figure 1: The tradeoff between rate R and the fraction of errors that can be corrected. $1 - \sqrt{R}$ is the tradeoff implied by the Johnson bound. The bound for unique decoding is $(1 - R)/2$ while $1 - R$ is the Singleton bound (and the list decoding capacity for codes over large alphabets).