A Combinatorial Bound on the List Size

Yuval Cassuto and Jehoshua Bruck
California Institute of Technology
Electrical Engineering Department
MC 136-93
Pasadena, CA 91125, U.S.A.
E-mail: {ycassuto,bruck}@paradise.caltech.edu

Abstract

We provide a simple, closed-form upper bound for the classical problem of worst case list-size of a general q-ary block code. This new bound improves upon the best known general bound when the alphabet of the code is large. We also show that with parameters of Reed-Solomon codes this bound is very close to the algebraic bound derived using the constructions of the Guruswami-Sudan decoder.

Index Terms

List decoding, List size bounds, Johnson bound

I. Introduction

Upper bounding the codeword list size output by a radius t decoder is a well studied problem with both theoretical and practical appeal. Codes that have short lists for relatively large decoding radii are termed *list-decodable* and their design is of prime interest. Though list-decodability of a code does not necessarily imply good minimum distance, the minimum distance does ensure a certain degree of list-decodability. The bound presented here, as those that predate it, uses the minimum distance of a code to bound the list size of a radius t decoder for that code. The reason we can bound the number of codewords in a radius t Hamming ball just based on the minimum distance is rather obvious; packing many codewords in a small sphere is impossible when every pair of codewords should be at least d apart. As mentioned in other works that deal with the worst case list-size, most notably [2] and [3, Ch.3], this problem is closely related to the problem of bounding the maximal size of a constant weight code. Accordingly, Johnson bound based arguments [7, pp.525], with necessary modifications, prove effective for the list-size problem. In [2], the Johnson bound is shown to provide a valid list-size bound in the binary case. The q-ary case was addressed in [9], though the main bound there can be extracted from the q-ary generalization of the Johnson bound, and a simple argument of its applicability to the list-size problem (see for example 2.3.1 of [6] for the q-ary Johnson bound). An improvement over [9] for short codes was reported in [5] using a geometric approach. The bound presented here is better than its predecessors when the alphabet size q is "large enough". The threshold alphabet size for the bound to be tighter depends solely on the asymptotic ratios $\frac{d}{n}$ and $\frac{t}{n}$. Therefore, for code families such that their alphabet grows with the length, this bound will be asymptotically tighter. An epitome of such codes is Reed-Solomon codes. For them, we show by examining sample codes, that an improvement is achieved even for relatively short RS codes. This encouraging behavior of the bound is further validated by showing that the bound is very close, at most a small constant away, to the algebraic bound of [8], despite being simpler and more general. We note that recently in [10], the authors proved a similar conclusion that the Guruswami-Sudan algebraic bound on the decoding radius of a list decoder applies to a general block code. Nevertheless, this result yields no closed form expression for the list size. Such a simple closed form expression is often required to analyze the behavior of the code. In [1], for example, the bound proved herein was used to prove a lower bound on the miscorrection probability of linear codes.

[1] This work was supported in part by the Lee Center for Advanced Networking at the California Institute of Technology.

II. A Combinatorial Bound on The List Size

For a code with length n and minimum distance d we want to bound the number of codewords that can reside in an arbitrary Hamming sphere of radius t. Similarly to Elias [2] and Goldreich et al [9], our analysis is combinatorial and will thus apply to a general code. In distinction from those known bounds, we will give a bound which is independent of the alphabet size of the code.

Theorem 1: Let $\mathcal{C}^*$ be an (n, K, d) code and $M(E)$ the number of codewords in distance at most t from a particular received word E, $M(E) = |\{C \in \mathcal{C}^* : d(C, E) \leq t\}|$. Then if $\lfloor \frac{d-1}{2} \rfloor < t < n\left(1 - \sqrt{1 - d/n}\right)$ we have

$$M(E) \leq \frac{n(d-t)}{t^2 - 2nt + dn} \tag{1}$$

Proof: We first consider the maximal number of codewords on the *surface* of the t sphere. Let $M'(E) = |\{C \in \mathcal{C}^* : d(C, E) = t\}|$. We fix E and define $M' = M'(E)$. For any pair of codewords C_i, C_j that are both in distance t from E, we have $d \leq D(C_i, C_j) \leq 2t$. We use $X(l)$ to denote the l^{th} coordinate of X. Then we define a pair of binary vectors K_i, K_j to be $K_i(l) = 1$ if $C_i(l) \neq E(l)$ and 0 otherwise, similarly for K_j. Then $|\{l : K_j(l) = 1\}| = |\{m : K_i(m) = 1\}| = t$. We define the *span* of two binary vectors as the number of coordinates that are 1 in at least one of the vectors

$$\text{span}(X_i, X_j) \equiv |\{l : X_i(l) = 1\} \cup \{m : X_j(m) = 1\}|$$

We claim that

$$\text{span}(K_i, K_j) \geq d$$

Otherwise there were more than $n - d$ coordinates in which $C_i(l) = C_j(l) = E(l)$, which would contradict the distance requirement $D(C_i, C_j) \geq d$. So a necessary condition to find M' codewords in distance t from E is the existance of M' binary vectors of weight t such that each pair (i, j), $i \neq j$, will have $\text{span}(K_i, K_j) \geq d$. Therefore, an upper bound on the latter will be an upper bound on the former for any alphabet size. The weight and span requirements together imply that the ones of K_i and K_j are allowed to overlap on at most $2t - d$ coordinates: $|\{l : K_i(l) = 1\} \cap \{m : K_j(m) = 1\}| \leq 2t - d$. Using the technique of the Johnson bound proof [7] with the above overlap we get

$$\frac{t^2 M'^2}{n} - tM' \leq (2t - d)M'(M' - 1)$$

$$\left(\frac{t^2}{n} - 2t + d\right) M' \left[M' - \frac{d - t}{\frac{t^2}{n} - 2t + d}\right] \leq 0$$

Solving for M', we get

$$M'(E) \leq \frac{n(d-t)}{t^2 - 2nt + dn}$$

under the condition

$$\frac{t^2}{n} - 2t + d > 0 \tag{2}$$

Solving for the condition (2)

$$t < n\left(1 - \sqrt{1 - d/n}\right) \tag{3}$$

To complete the proof we want to show that the upper bound on M' is also an upper bound on M. We define $W(X)$ to be the Hamming weight of X and claim the following. If we have M binary words such that every pair (K_i, K_j) taken from them satisfies
(1) $W(K_i), W(K_j) \leq t$
(2) $\text{span}(K_i, K_j) \geq d$
then there exist M binary words such that any pair satisfies
(1) $W(K_i), W(K_j) = t$
(2) $\text{span}(K_i, K_j) \geq d$
This can be shown by observing that increasing the weight of K_i or K_j cannot decrease the span.

□

From the first part of the proof of theorem 1 it is clear that the bound provided applies to the problem of constant weight codes as well. However, when considering this problem of placing codewords *on the surface* of the sphere, the same result can be proved using the q-ary Johnson bound by replacing the Cauchy-Schwartz argument with an integrality argument that is more efficient for large q. This simple shortcut does not seem to generalize to our problem in which some codewords are possibly placed strictly inside the t-Hamming ball.

III. Analysis of the List-Size Bound

We now turn to analyze the proposed bound. In subsection III-A we evaluate it in comparison to the best known combinatorial bounds and give exact threshold on the alphabet size, above which it is tighter than the best known. In subsection III-B we explore the strong link the bound has to the seemingly unrelated algebraic bound for Reed-Solomon codes.

A. Comparison with Known Combinatorial Bounds

A possible justification for a q-independent bound arises from the following. Ignoring the alphabet size in the proof of theorem 1 required us to count the overlapping coordinates towards d, which is less restrictive (and thus result in a looser bound) than the Johnson bound in the binary case. However, if the alphabet size is large "enough", it will not limit the overlap symbols and the span requirement will capture the limitation on the number of codewords in the sphere. As it turns out, this simplification proves advantageous for giving strictly tighter bounds for alphabets above some threshold.

To simplify the analysis we will fix the asymptotic distance by $\gamma = 1 - \frac{d}{n}$ and the decoding radius by $\delta = 1 - \frac{t}{n}$. Now the bound (1) is given by

$$M \leq \frac{\delta - \gamma}{\delta^2 - \gamma} \tag{4}$$

Henceforth we will denote the bound in (4) by L_C. For nontrivial codes we require $0 < \gamma < 1$ and for δ we require $\sqrt{\gamma} < \delta < \frac{1+\gamma}{2}$. The lower limit is to maintain positive denominator in (4) and the upper limit represents decoding beyond half the minimum distance. The main bound of [9, Thm4.2], which to the best of our knowledge is the tightest known, asserts

$$M \leq \frac{(1-\gamma)(1-\frac{1}{q})}{(\delta - \frac{1}{q})^2 - (1-\frac{1}{q})(\gamma - \frac{1}{q})} \tag{5}$$

which for large q tends to $\frac{1-\gamma}{\delta^2-\gamma}$, a value larger than (4) since $\delta < 1$. The exact alphabet size q_0, above which (4) is tighter than (5) can be recovered, as a function of γ, δ, by solving the following inequality for q

$$\frac{(1-\gamma)(1-\frac{1}{q})}{(\delta - \frac{1}{q})^2 - (1-\frac{1}{q})(\gamma - \frac{1}{q})} > \frac{\delta - \gamma}{\delta^2 - \gamma} \tag{6}$$

The above simplifies to a linear inequality and yields the threshold

$$q > q_0(\gamma, \delta) = \frac{\delta(1+\gamma) - 2\gamma}{\delta^2 - \gamma}$$

Table I shows a comparison of the bounds for sample codes. The rightmost column is the q_0 found above for the corresponding parameters n, d, t.

$(n, d, t), q$	(1)	[9]	q_0
$(31, 17, 9), 32$	4	10	2
$(31, 17, 10), 32$	31	51	11
$(255, 33, 17), 256$	120	239	9
$(18, 17, 13), 19$	10	18	8

TABLE I

Bound comparison for sample decoders

B. Connection to Algebraic Bound for Reed-Solomon Codes

The decoding radii for which the proposed bound applies are those that satisfy (3). For Reed-Solomon codes that implies

$$t < n - \sqrt{(k-1)n}$$

which equals exactly the famous Guruswami-Sudan bound for decoding Reed-Solomon codes efficiently using the GS algorithm [4]. This coincidence of domains between the bounds allows us to set forth a comparison between the general combinatorial list-size bound, and the Reed-Solomon specific algebraic bound.

Algebraic list-size Bound

In [8] McEliece provides a two step, closed form list-size bound, derived from arguments on maximal degrees of bivariate polynomials. The first step is determining the minimum multiplicity[2] required to achieve decoding radius of t

$$m > (k-1) \cdot \frac{t + \sqrt{n(2t+k-1-n)}}{2((n-t)^2-(k-1)n)} = \gamma \frac{1-\delta+\sqrt{\gamma-2\delta+1}}{2(\delta^2-\gamma)} \tag{7}$$

The second step uses a list-size bound L_A that is given as a function of the multiplicity m.

$$L_A \approx \left(m + \frac{1}{2}\right)\sqrt{\frac{n}{k-1}} \tag{8}$$

$\approx$ here means the right side is less than 1 greater than the true value of the bound. Substituting m from (7) into (8) we get

$$L_A \approx \frac{\sqrt{\gamma}}{2} \cdot \frac{1-\delta+\sqrt{1-2\delta+\gamma}+\frac{1}{\gamma}(\delta^2-\gamma)}{\delta^2-\gamma}$$

So far we have a combinatorial bound $L_C = \frac{\delta-\gamma}{\delta^2-\gamma}$ and an algebraic bound L_A above. We want to argue that L_C is close to L_A despite being more general. The following theorem shows that when approaching the strongest GS decoder (decoding radii that attain the GS bound) L_C and L_A converge to the same bound.

Theorem 2: $\lim_{\delta\to\sqrt{\gamma}} \frac{L_A}{L_C} = 1$

Proof: Elementary substitution $\delta = \sqrt{\gamma}$ into $\frac{L_A}{L_C}$.

□

It is also possible to show that the difference $L_C - L_A$ is small for general γ, δ.

Theorem 3: For every pair γ, δ the combinatorial and algebraic bounds on the list size satisfy

$$L_C - L_A < \frac{1}{4}\left[1 + \frac{2}{1-\sqrt{\gamma}}\right]$$

Proof: We first prove a simple lemma.

Lemma 1: If $\sqrt{\gamma} < \delta < \frac{1+\gamma}{2}$ then

$$\sqrt{1-2\delta+\gamma} > \frac{1-2\delta+\gamma}{1-\sqrt{\gamma}} \tag{9}$$

Proof:

$$\left(\frac{1-2\delta+\gamma}{1-\sqrt{\gamma}}\right)^2 - \left(\sqrt{1-2\delta+\gamma}\right)^2 =$$

$$= \frac{4\delta^2 - \delta(2\gamma+2+4\sqrt{\gamma}) + 2\sqrt{\gamma}(1+\gamma)}{(1-\sqrt{\gamma})^2} =$$

$$= \frac{4\overbrace{(\delta-\sqrt{\gamma})}^{>0}\overbrace{(\delta-\frac{1+\gamma}{2})}^{<0}}{(1-\sqrt{\gamma})^2} < 0$$

the lemma follows since both sides of (9) are positive so $x^2 - y^2 < 0 \Rightarrow x < y$.

We are now ready to prove the theorem

$$L_A + \frac{1}{4}\left[1 + \frac{2}{1-\sqrt{\gamma}}\right] - L_C =$$

[2] In [8] t is bounded given m so the expression here is the corresponding bound on m given t.

$$= \frac{\sqrt{\gamma}}{2} \cdot \frac{1-\delta+\sqrt{1-2\delta+\gamma}+\frac{1}{\gamma}(\delta^2-\gamma)}{\delta^2-\gamma} + \frac{3-\sqrt{\gamma}}{4(1-\sqrt{\gamma})} - \frac{\delta-\gamma}{\delta^2-\gamma} >$$

$$> \frac{\sqrt{\gamma}}{2} \cdot \frac{1-\delta+\frac{1-2\delta+\gamma}{1-\sqrt{\gamma}}+\frac{1}{\gamma}(\delta^2-\gamma)}{\delta^2-\gamma} + \frac{3-\sqrt{\gamma}}{4(1-\sqrt{\gamma})} - \frac{\delta-\gamma}{\delta^2-\gamma} =$$

$$= \frac{(\delta-\sqrt{\gamma})(2+\sqrt{\gamma}-\gamma)}{4\sqrt{\gamma}(1-\sqrt{\gamma})(\delta+\sqrt{\gamma})} > 0$$

The first inequality follows from lemma 1, the equality from straight forward manipulation and the last inequality from the positivity of both the numerator and denominator for $0 < \gamma < 1$, $\sqrt{\gamma} < \delta < \frac{1+\gamma}{2}$.

□

Substituting sample values of γ we get the following corollary.

Corollary 1:

(1) $L_C - L_A < 2$ for all δ when $\gamma \leq 0.5$

(2) $L_C - L_A < 5$ for all δ when $\gamma \leq 0.8$

(3) $L_C - L_A < 10$ for all δ when $\gamma \leq 0.9$

REFERENCES

[1] Y. Cassuto and J. Bruck. Miscorrection probability beyond the minimum distance. In *Proc. of the IEEE International Symposium on Info. Theory*, Chicago, Illinois, June 2004. IEEE.

[2] P. Elias. Error-correcting codes for list decoding. *IEEE-Trans-IT*, 37(1):5–12, 1991.

[3] V. Guruswami. *List decoding of error-correcting codes*. Ph. D. Dissertation, Massachusetts Institute of Technology, 2001.

[4] V. Guruswami and M. Sudan. Improved decoding of reed-solomon and algebraic-geometric codes. *IEEE-Trans-IT*, 45:1755–1764, 1999.

[5] V. Guruswami and M. Sudan. Extensions to the johnson bound. *Manuscript*, 2001.

[6] W.C Huffman and V. Pless. *Fundamentals of Error-Correcting Codes*. Cambridge university press.

[7] F.J MacWilliams and N.J.A Sloane. *The Theory of Error-Correcting Codes*. North Holland, Amsterdam, The Netherlands, 1977.

[8] R.J McEliece. The guruswami-sudan algorithm for decoding reed-solomon codes. Technical Report IPN progress report 42-153, JPL, http://www.ipnpr.jpl.nasa.gov/progress_report/42-153/, 2003.

[9] O. Goldreich R. Rubinfeld and M. Sudan. Learning polynomials with queries: The highly noisy case. *SIAM J. Discrete Math*, 13(4):535–570, November 2000.

[10] G. Ruckenstein and R.M Roth. Bounds on the list-decoding radius of reed-solomon codes. *SIAM J. Discrete Math*, 17(2):171–195, November 2003.