LIST DECODING OF CONCATENATED CODES: IMPROVED PERFORMANCE ESTIMATES

ALEXANDER BARG AND ANDREW MCGREGOR

Abstract. An improved bound is proved on the list-decoding radius of a concatenated code relying upon a combination of (soft-decision) algebraic list decoding and generalized minimum distance (GMD) decoding in the outer level. This bound is further improved if the inner code is a random linear code.

1. Introduction

Until recently the best estimates of the number of errors corrected by concatenated codes were obtained under GMD decoding of the outer code (see G. D. Forney [4] for the main result and I. Dumer [2] for an overview of improvements). Recently V. Guruswami and M. Sudan [7] showed that Forney's bound on the number of correctable errors can be improved by using list decoding of the outer (Reed-Solomon) codes. Other results in this direction were obtained in [8, 9]. This paper continues the same line of research, combining the GMD and list decoding schemes for better error correction.

A q-ary $[Nn, Kk, Dd]$ linear concatenated code C is constructed from an inner q-ary code $A[n, k = rn, d = \delta n]$ and an outer Q-ary $[N, K = \kappa n, D = \Delta N]$ Reed-Solomon (RS) code, where $Q = q^k$. A typical codeword of the concatenated code C can be thought of as a q-ary $n \times N$ matrix in which the ith column, $1 \le i \le N$, represents an encoding with the code A of the q-ary representation of the ith symbol of the RS codeword. The rate of the code C is $R = r\kappa$. We assume that $q^k \le N^{O(1)}$, so that we can do full maximum likelihood decoding of the inner code in poly(Nn) time.

A GMD decoding procedure that corrects errors up to half the designed distance Dd was introduced in [4]. Consider the transmission of a codeword $z_1 z_2 \dots z_N$ from C where each $z_i \in A$ over a q-ary symmetric channel. We denote the received word by $r_1 r_2 \dots r_N$, where each column r_i represents a corrupted version of the codeword z_i. Under GMD decoding, each column r_i is decoded with the code A to obtain a codeword $y_i \in A$. The distance $h_i = \mathrm{d}(r_i, y_i)$ is used as a reliability estimate of the decoding decision in the ith column and passed to the outer decoder. Let $b_i \in \mathbb{F}_Q$ be the symbol of the RS code alphabet which corresponds to y_i under the encoding mapping of the code A[1]. The outer decoder performs several decoding attempts of the "received vector" $(b_1, \dots, b_N)$ correcting errors and erasures.

Recently a new algebraic list decoding algorithm of RS codes was introduced in [6]. We formulate their result in a way suitable for concatenated decoding. Suppose that in each column i, an integer non-negative weight $w_{i,j}$ is assigned to each codeword $x_j \in A$.

Theorem 1 (G-S Decoding [6])**.** *In time polynomial in Nn, the G-S decoding method can find all codewords $(z_1, \dots, z_N)$ such that*

$$\sum_{\substack{i=1 \\ x_j = z_i}}^{N} w_{i,j} > \sqrt{K \sum_{i,j} w_{i,j}^2}.$$

The G-S decoding algorithm corrects (in the list decoding sense) more errors than other algebraic decoding procedures of RS codes. Even more importantly, the above theorem enables one to use

Research supported in part by NSF under grant CCR0310961.

[1]By abuse of notation, in such cases below we write $b_i = y_i$.

reliability information of the symbols and provides a powerful and flexible decoding scheme useful in a wide range of applications such as concatenated decoding and digital fingerprinting.

In this paper we first effectively combine the G-S decoding algorithm and the GMD decoding algorithm such that we improve upon the decoding radius of both. Our result relies on the observation that if the most reliable columns are still not particularly reliable then the decoding radius of GMD can be improved. On the other hand, if the most reliable columns are very reliable, then the G-S decoding algorithm can be tweaked to yield a better bound on the decoding radius. Secondly, we further improve the decoding radius if the inner code is a random linear code by making use of known results about the coset distribution of random linear codes. Improvements when the inner code is random linear have already been studied by V. Guruswami and P. Indyk [5] but they focused on polynomial-time decoding of concatenated codes up to half the Gilbert-Varshamov radius proportion of errors[2]. In contrast, we are concerned with improvements over GMD decoding in a broad range of code rates.

2. Decoding Better than GMD and G-S Decoding

Consider transmission over a q-ary with concatenated codes. GMD decoding of the outer code enables one to correct up to $Dd/2$ errors in a received word with polynomial complexity. This estimate was recently improved in [7] relying on Theorem 1 and the decoding algorithm associated with it. The result of [7] is as follows.

Theorem 2. [7] *There exists a polynomial-time decoding algorithm that is capable of list decoding the code C up to*

$$Nn\Big[\Big(1-\frac{1}{q}\Big)\Big(1-\sqrt{1-\frac{\delta}{1-1/q}}\Big)-\sqrt{\delta\kappa}\Big]$$

errors.

This improves upon the $Dd/2$ bound for small values of the outer code rate κ. In this section we combine GMD decoding and Theorem 1 to further improve the (list decoding) correction radius of concatenated codes.

Let l_i be the number of errors in the ith column and $L=\sum_{1\le i\le N} l_i$ be the total number of errors. As above, for each column r_i let y_i be the nearest codeword of A. Let $h_i = d(y_i, r_i)$. Without loss of generality let $h_1 \le h_2 \le \dots h_N$. Let $H=\sum_{1\le i\le N-D} h_i$.

Theorem 3. *In time polynomial in Nn we can list-decode C up to the following fraction of errors,*

$$\frac{L}{Nn} \le \min_{0\le H/Nn\le \kappa(\delta-J)} \max\{A,B\}$$

where

$$A = J+\kappa(\delta-J)-\frac{H}{Nn}-\sqrt{\kappa\Big(\Big(1-\kappa+\frac{H/Nn}{\delta-J}\Big)\delta+\Big(\kappa-\frac{H/Nn}{\delta-J}\Big)\delta^2\Big)} \tag{1}$$

$$B = \frac{\delta(1-\kappa)}{2}+\frac{H}{Nn}. \tag{2}$$

We will use an improvement of the decoding radius of GMD decoding due to I. Dumer, [1] that takes into account that GMD decoding only focuses on the D most unreliable symbols.

Theorem 4 (GMD Decoding)**.** *In time polynomial in Nn, GMD decoding will list-decode C up to the number of errors,*

$$L < dD/2 + H.$$

[2]in the footsteps of [10], where the same result was attained with superpolynomial complexity for low code rates, $R(C)\le 0.02$. As could be expected, the range of rates in [5] in which it is possible to make the corresponding claim is even lower, namely $R(C)\le 10^{-4}$.

Note that the bigger H is, the more errors we can provably correct with GMD. To decode up to the radius claimed in Theorem 3 we combine the G-S decoding method with a different setting of weights with GMD decoding.

Proof of Theorem 3. In our decoding we will do both GMD decoding and G-S decoding and argue a sufficient condition for a codeword to be output by at least one of decodings. For the G-S decoding, in each column i we find all codewords in the inner code up to a distance $\max\{Jn, d - h_i\}$ from r_i. We assign the weight

$$w_{i,j} = \max\{\max\{Jn, d\} - d(x_j, J), 0\}$$

to codeword x_j in this column. We quote the following combinatorial result from [7].

Lemma 5. *Suppose that in the ith column $d - h_i \leq Jn$. Then*

$$\sum_j w_{i,j}^2 \leq \delta n^2. \tag{3}$$

Alternatively, if $Jn < d - h_i$, there is only one codeword that gets positive weight (by the triangle inequality) and therefore

$$Jn < d - h_i \Rightarrow \sum_j w_{i,j}^2 = (d - 2h_i)^2 \leq \delta^2 n^2$$

Let there be X columns such that $Jn > d - h_i$. Consider X', the number of columns among those with the $N - D$ smallest values of h_i that have $Jn > d - h_i$. In these columns we required $h_i > d - Jn$ but we also have $H = \sum_{1 \leq i \leq N-D} h_i$. Hence we get that $X'(d - Jn) \leq H$. Therefore

$$X \leq D + X' \leq D + \frac{H}{d - Jn}$$

We also will want a lower bound on $\sum_{1 \leq i \leq N} \max\{Jn, d - h_i\}$. First let

$$\begin{aligned} \sum_{1 \leq i \leq N} \max\{Jn, d - h_i\} &\geq DJn + \sum_{1 \leq i \leq N-D} \max\{Jn, d - h_i\} \\ &\geq NJn + (N - D)(d - Jn) - H \end{aligned}$$

From Theorem 1, G-S decoding will output a codeword z in our list if

$$\sum_{1 \leq i \leq N, x_j^i = z_i} w_{i,j} \geq \sqrt{K \sum_{i,j} w_{i,j}^2}$$

Now using the bounds we have established in terms of H we get

$$\sqrt{K \sum_{i,j} w_{i,j}^2} \leq \sqrt{K(X\delta n^2 + (N - X)\delta^2 n^2)} \leq Nn\sqrt{\kappa\Big(\Big(1 - \kappa + \frac{H/Nn}{\delta - J}\Big)\delta + \Big(\kappa - \frac{H/Nn}{\delta - J}\Big)\delta^2\Big)}$$

and

$$\begin{aligned} \sum_{1 \leq i \leq N, x_j^i = z_i} w_{i,j} &= \sum_{1 \leq i \leq N} \max\{0, \max\{Jn, d - h_i\} - l_i\} \\ &\geq \sum_{1 \leq i \leq N} \max\{Jn, d - h_i\} - l_i \\ &\geq NnJ + (N - D)(d - Jn) - H - \sum_{1 \leq i \leq N} l_i \end{aligned}$$

Hence G-S decoding will output a codeword z in our list if

$$\sum_{1\le i\le N} l_i \le Nn\Big[J+\kappa(\delta-J)-\frac{H}{Nn}-\sqrt{\kappa\Big(\Big(1-\kappa+\frac{H/Nn}{\delta-J}\Big)\delta+\Big(\kappa-\frac{H/Nn}{\delta-J}\Big)\delta^2\Big)}\Big], \tag{4}$$

Note that the smaller H is the more errors we can correct with our modified version of G-S.

Hence by doing both GMD and G-S decoding with our setting of weights we will decode up to the maximum of the decoding radii given in equations (1) and (2). Minimizing over H gives the theorem. □

Remark. The result of Theorem 3 improves on the standard G-S decoding bound and is above the GMD bound for a certain set of κ values for any given value of δ. Both the improvement of [7] and our result are better than the decoding radius of GMD decoding only for small values of the outer rate κ. Our main result here is a combination of GMD decoding and G-S decoding in one algorithm.

The nonexplicit nature of the bound in Theorem 3 is somewhat unfortunate. However if we relax the bound on the error correction radius in the G-S decoding given by (4) to the slightly weaker

$$\sum_{1\le i\le N} l_i \le Nn\Big(J+\kappa(\delta-J)-\frac{H}{Nn}-\sqrt{\kappa\delta}\Big)$$

then we get the following simpler error correction radius:

$$\frac{L}{Nn}\le\begin{cases} J-\sqrt{\delta\kappa} & T(\delta,\kappa)\ge\kappa(\delta-J)\\ \frac{\delta(1-\kappa)}{2}+T(\delta,\kappa) & 0\le T(\delta,\kappa)\le\kappa(\delta-J)\\ \frac{\delta(1-\kappa)}{2} & T(\delta,\kappa)\le 0,\end{cases}$$

where

$$T(\kappa,\delta)=\frac{1}{2}(J+\kappa(\delta-J)-\sqrt{\delta\kappa}-(1-\kappa)\delta/2)$$

3. Using Random Codes

In this section we consider concatenated codes with random linear codes in the inner level. In [7], a probabilistic construction of a concatenated code C whose outer code is Reed-Solomon and whose inner code is random is given that decodes up to a $J-\sqrt{\delta_{\rm GV}\kappa}$ fraction of errors. The construction works because a random code has a relative distance $\delta_{\rm GV}$ with high probability. However we also know that a random code has a certain coset distribution with high probability. The following theorem is due to V.V. Zyablov and M.S. Pinsker [10]. Let $h_q(x)=-x\log_q\frac{x}{q-1}-(1-x)\log_q(1-x)$.

Theorem 6. *For almost all $[n,rn]$ linear codes the number of codewords in a sphere of radius t, where $t=n(\delta_{\rm GV}-\varepsilon)$, is at most*

$$q^{(1-r)/\varepsilon h_q'(\delta_{\rm GV})}$$

and does not depend on n.

For $\delta_{\rm GV}$ not too large, i.e., r not too small, $h_q'(\delta_{\rm GV})\ge h_q(\delta_{\rm GV})$, and the number of codewords is at most $q^{1/\varepsilon}$. For $q=2$ this is true for $r\ge 0.345$. Hereafter we restrict our attention to $q=2$ and $r\ge 0.345$. Using the above result we can improve upon the bound of Lemma 5 as follows:

Lemma 7. *Let $C=\{x_1,x_2,\ldots\}$ be a linear code chosen with uniform probability from the ensemble of $[n,rn]$ linear codes. With probability $\to 1$ as $n\to\infty$ we get*

$$\sum_{j:x_j\in C} w_j^2(x)\le\delta_{\rm GV}^2E(c)$$

where

$$E(c):=\Big(1-\frac{1}{c}\Big)^2+7\Big(1-\frac{1}{c}-\frac{1}{2}\Big)^2+(\ln 2)\int_{1/2}^{1-1/c}\Big(1-\frac{1}{c(1-u)}\Big)^2 2^{\frac{1}{1-u}}du$$

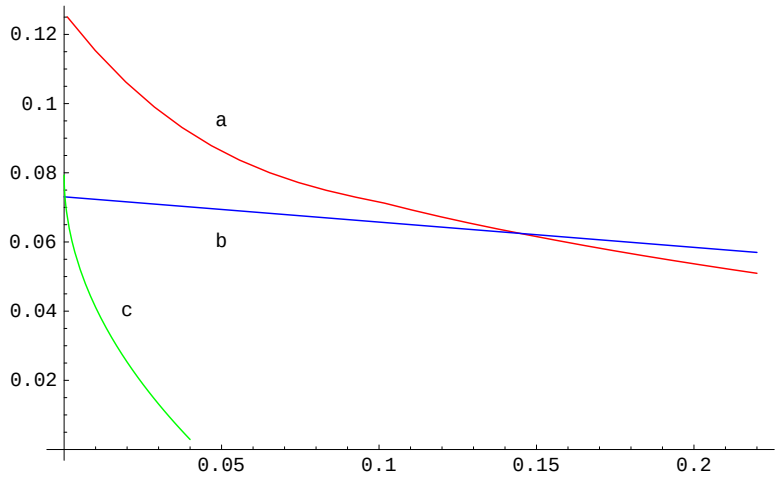

FIGURE 1. The graph depicts the decoding radius for various methods for varying rates of the outer code κ. a) Theorem 8, using a random inner code, b) The GMD decoding radius and c) The decoding radius of G-S. For all the bounds the inner code is binary of rate $r = 0.4$ and relative distance $\delta_{\mathrm{GV}}(0.4) = 0.1461$.

$$w_j(x) := \max\left\{0, \left(1 - \frac{1}{c}\right)\delta_{\mathrm{GV}}(r)n - d(x_j, x)\right\}$$

and c is an arbitrary constant ≥ 2.

Proof. With high probability the minimum distance of C is $d := \delta_{\mathrm{GV}} n$. We have

$$\begin{aligned}\frac{1}{d^2}\sum_j w_j^2 &= \frac{1}{d^2}\sum_{x_j : d(x_j,y) \leq d/2} w_j^2 + \frac{1}{d^2}\sum_{x_j : d/2 \leq d(x_j,y) \leq d(1-1/c)} w_j^2 \\ &\leq \left(1 - \frac{1}{c}\right)^2 + \sum_{x_j : d/2 \leq d(x_j,y) \leq d(1-1/c)} \left(\left(1 - \frac{1}{c}\right) - \frac{d(x_j,y)}{d}\right)^2\end{aligned}$$

since, by the triangle inequality, there exists at most one codeword at distance $\leq d/2$ from r_i and $w_j^2 \leq d^2(1 - 1/c)^2$. Using Theorem 6 and the fact that $w_j(x)$ is decreasing with $d(x_i, x)$ we bound the second term as follows.

$$\begin{aligned}&\sum_{x_j : \frac{d}{2} \leq d(x_j,r_i) \leq d(1-\frac{1}{c})} \left(\left(1 - \frac{1}{c}\right) - \frac{d(x_j,x)}{d}\right)^2 \\ &= \sum_{l=d/2}^{d(1-1/c)} \left(1 - \frac{1}{c} - \frac{l}{d}\right)^2 |\{x_j : x_j \in C \text{ and } d(x_j, x) = l\}| \\ &\leq 7\left(1 - \frac{1}{c} - \frac{1}{2}\right)^2 + \sum_{l=d/2+1}^{d(1-1/c)} \left(1 - \frac{1}{c} - \frac{l}{d}\right)^2 (2^{\frac{1}{1-l/(d(1-1/c))}} - 2^{\frac{1}{1-(l-1)/(d(1-1/c))}}) \\ &\leq 7\left(1 - \frac{1}{c} - \frac{1}{2}\right)^2 + \int_{1/2}^{1-1/c} \left(1 - \frac{1}{c} - u\right)^2 d\mu(u)\end{aligned}$$

where $\mu(u) = 2^{1/(1-u)}$. This gives the lemma. □

Theorem 8. *Let C be a concatenated code chosen from the ensemble of codes with an RS outer code and random inner code. With probability $\to 1$ as $n \to \infty, N \to \infty$, the code C can be list-decoded with polynomial complexity up to the following fraction of errors,*

$$\frac{L}{Nn} \leq \begin{cases} \delta_{\mathrm{GV}} \max_{c \geq 2} \left[(1 - \frac{1}{c}) - \sqrt{\kappa E(c)} \right] & \kappa \leq \kappa^* \\ \delta_{\mathrm{GV}} \frac{(1-\kappa)}{2} & \kappa \geq \kappa^* \end{cases}$$

where κ^* is the root of

$$\max_{c \geq 2} \left[\left(1 - \frac{1}{c}\right) - \sqrt{\kappa E(c)} \right] = \frac{1-\kappa}{2}.$$

Proof. As before we do a combination of G-S decoding with weights set as in Lemma 7 and GMD decoding. To get the improvement over GMD decoding for $\kappa > \kappa^*$ we assume that the random inner code has minimum distance $\delta_{\mathrm{GV}} n$ and that Lemma 7 holds true.

$$\sum_{1 \leq i \leq N, x_j^i = z_i} w_{i,j} \quad \geq \quad Nn\delta_{\mathrm{GV}}(r)\Big(1 - \frac{1}{c}\Big) - \sum_{1 \leq i \leq N} l_i$$

Then by Theorem 1 and Lemma 7 we find a codeword in the G-S decoding if

$$L \leq \delta_{\mathrm{GV}}(r) Nn \max_{c \geq 2} \left[\left(1 - \frac{1}{c}\right) - \sqrt{\kappa E(c)} \right]$$

□

Remark. The bound in Theorem 8 represents a significant improvement over the previous bound in Theorem 2. The new result also represents an improvement over GMD decoding. For instance for $\delta = 0.1461$ $(= \delta_{\mathrm{GV}}(0.4))$ and $q = 2$, the new result the error-correcting radius of GMD decoding for outer rates $\kappa \in [0, 0.1449]$ whereas the bound of Theorem 2 does the same only for $\kappa \in [0, 2.73 \times 10^{-4}]$. Note however that Theorem 2 is a more general result that does not assume a random inner code. The bounds are shown in Figure 1.

References

[1] I. Dumer, *A choice of cascaded decoding techniques according to the noise level*, Proceedings of The Eight All-Union Conference on Coding Theory and Inform. Transmission Pt. 2 (1981), 61–65 (in Russian)

[2] ———, *Concatenated codes and their multilevel generalizations*, in Handbook of Coding Theory, V. Pless and W. C. Huffman, Eds., Vol. 2, Amsterdam: Elsevier Science, 1998.

[3] G. Forney, *Concatenated Codes*, MIT Press, Cambridge, MA, 1966

[4] ———, *Generalized minimum distance decoding*, IEEE Trans. Inform. Theory **12** (1966): no. 2, 125–131

[5] V. Guruswami and P. Indyk, *Efficiently decodable low-rate codes meeting Gilbert Varshamov bound* presented at the 41st Annual Allerton Conference on Communication, Control and Computing, October 2003.

[6] V. Guruswami and M. Sudan, *Improved decoding of Reed-Solomon codes and algebraic-geometry codes*, IEEE Trans. Inform. Theory **45** (1999): no. 6, 1757–1767.

[7] ———, *Decoding concatenated codes using soft information*, Proceedings of the Seventeenth IEEE Conference on Computational Complexity, pages 148-157, Montreal, Canada, 21-24 May, 2000.

[8] R. Koetter and A. Vardy, *Algebraic soft-decision decoding of Reed-Solomon codes*, manuscript (2001).

[9] R. R. Nielsen, *Decoding concatenated codes with Sudan's algorithm*, maniscript (2001).

[10] V. V. Zyablov and M. S. Pinsker, *List cascade decoding*, Problems of Information Transmission **17** (1981): no. 4 29–34, (in Russian) (1982): 236–240 (in English)

Department of ECE, University of Maryland, College Park, MD 20742
E-mail address: abarg@ieee.org

Department of CIS, University of Pennsylvania, 3330 Walnut Street, Philadelphia, PA 19104
E-mail address: andrewm@cis.upenn.edu