A New Analysis of Rateless IBLTs

Lei Yang

Aug. 1, 2024

The paper Intermediate Performance of Rateless Codes mentioned a new result (c.f. references 3, 4 there) that allows analyzing the peeling algorithm. Importantly, this method does not assume local tree-like structure in the Tanner graph. As a result, the method is well suited for producing a rigorous analysis of Rateless IBLTs.

First, look at Theorem 1, but ignore $s(r, P)$—we can view it simply as a parameter q which is the fraction of source symbols we expect to recover. Then, Theorem 1 says something that is quite similar to what we do in Density Evolution: the inequality in Eq. 3 must hold for all t smaller than q. Recall that in Density Evolution, there is also an inequality that must hold for all t smaller than q. If that condition is true, and if the number of coded symbols follows Poisson(rk), we can decode q-fraction of the k source symbols. Notice that the mean of the distribution is rk, so the overhead is r.

Now, the main task is to compute $P'(t)$, i.e., the derivative of the probability generating function of the coded symbol degree distribution in Rateless IBLTs.

Now that we have introduced the result in the Intermediate Performance paper, we switch to the notation in my paper, where $P(t)$ is $\Psi(x)$.

My previous analysis already gives a closed-form expression of $\Psi'(x)$.

$$\Psi'(x) = \frac{n}{m} \sum_{i=0}^{m-1} \rho(i)(\rho(i)(x-1)+1)^{n-1}. \tag{1}$$

However, it's the sum of $m-1$ terms where m, the number of coded symbols, goes to infinity in the analysis. Our goal is to find the limit of the said series.

Fortunately, this is done as we compute $f(q)$ in the original analysis. Based on results in the paper,

$$\lim_{n\to\infty} f(q) = e^{\frac{1}{\alpha}\mathrm{Ei}(-\frac{q}{\alpha\eta})}, \tag{2}$$

$$f(q) = e^{-\sum_{i=0}^{m-1} \rho(i)(1-q\rho(i))^{n-1}}. \tag{3}$$

Let $q = 1 - x$, take the limit on both sides in Eq. 3, and combine it with Eq. 2, we get

$$e^{\frac{1}{\alpha}\mathrm{Ei}(-\frac{1-x}{\alpha\eta})} = e^{-\lim_{n\to\infty}\sum_{i=0}^{\eta n-1}\rho(i)((x-1)\rho(i)+1)^{n-1}}. \tag{4}$$

Take log on both sides, we get

$$\lim_{n\to\infty}\sum_{i=0}^{\eta n-1}\rho(i)((x-1)\rho(i)+1)^{n-1} = -\frac{1}{\alpha}\mathrm{Ei}(-\frac{1-x}{\alpha\eta}) \tag{5}$$

Plugging Eq. 5 into Eq. 1, we get

$$\lim_{n\to\infty}\Psi'(x) = -\frac{1}{\alpha\eta}\mathrm{Ei}\left(\frac{x-1}{\alpha\eta}\right). \tag{6}$$

Theorem 1 of the Intermediate Performance paper tells us that for $\eta > 0$ such that

$$\forall 0 \le t < z : \eta\Psi'(t) + \log(1-t) > 0, \tag{7}$$

the fraction of recovered source symbols goes to z.

Isn't Eq. 7 familiar? Of course. Let $q = 1 - t$. (Pardon the overloaded symbol q. I have to use q here since it is the notation used in my original analysis.) Then, take exp on both sides, plug in Eq. 6, and rearrange it a bit, we get

$$\forall q \in (1-z, 1] : q > e^{\frac{1}{\alpha}\mathrm{Ei}\left(-\frac{q}{\alpha\eta}\right)}. \tag{8}$$

Let $z = 1$, i.e., we want to recover all source symbols. Eq. 8 is identical to the inequality in my original analysis. In other words, we have arrived at the exactly same conclusion through a different path. One thing that is worth remarking is that this new analysis requires that the number of received coded symbols to follow Poisson distribution with mean ηn, where the original analysis says the success probability at *exactly* ηn coded symbols tends to 1 as the length of the code goes to infinity. We do not get as strong a result. However, the new analysis does tell us the average overhead, which is what we are after. Another remark is that the original analysis is not technically incorrect. As mentioned in New model for rigorous analysis of LT-codes, the tree-based analysis has been used as a heuristic prior to that work.