1 Recap - Distance amplification of Tanner codes

Last time, we saw an algorithm that decodes Tanner codes with errors up to 1/4 the distance of the code. However, $R(\delta) = 1 - 2h(\sqrt{\delta})$, which is only positive for $\delta < 0.11^2$. Hence, we cannot correct too many errors.
We also saw a construction [ABN+92] that amplified the distance of such codes, shown below.

Consider a binary code $C \subseteq \mathbb{F}_2^n$ (these constructions generalise to higher fields as well) and let $G = (L, R, E)$ be a $n \times n$ ϵ-pseudorandom r-regular bipartite graph. Then, define $G(C)$ in the following way:

1. For any codeword $c \in C$, let it "sit" on the left vertices L of G.
2. Define the codeword $G(c) \in G(C)$ as a vector of r-tuples where each coordinate is just the collection of the codeword bits along the edges incident on the vertex $G(C)_i$ (each coordinate corresponds to a vertex in R).

If $\delta(C) = \delta_0 > 0$, say 1/1000, then we can show that the resulting code $G(C)$ satisfies

- $R(G(C)) = \Omega(\gamma)$ (a factor r worse than $R(C)$).
- $\delta(G(C)) \geq 1 - \gamma$.
- Alphabet size is $2^{O(1/\gamma)}$.

Note that this is close to the singleton bound for low rates (up to constant factors) for smaller alphabet sizes than Reed-Solomon codes.

2 Decoding distance-amplified Tanner codes

2.1 Algorithm

This algorithm decodes up to $\frac{1-\gamma}{2}$ fraction of errors.

1. For a codeword $y \in \Sigma^n$ on the right set of G, form $z \in \mathbb{F}_2^n$ by setting z_i to be majority of the votes for each bit from its r neighbors.
2. Decode the resulting codeword $z \in \mathbb{F}_2^n$ to some codeword c if there exists one within $\tau n \approx \frac{\delta_0 n}{4}$.

Note that this algorithm is linear time (in the RAM model) with some careful implementation.

2.2 Analysis

Let S be the set of errors on the left set and let T be the correct tuples on the right set. Note that a tuple is considered incorrect even if it contains a single flipped bit.
Since we are correcting up to $\frac{1-\gamma}{2}$ errors, $|T| \geq \left(\frac{1+\gamma}{2}\right) n$. Also notice that $S = \{u \mid z_u \neq c_u\}$

(for the unique closest codeword c output by the algorithm) and $u \in S \implies u$ has at most $r/2$ neighbors in T (else the majority would have returned the correct bit).
We now have (using ϵ-pseudorandomness)

$$\begin{aligned} \frac{r|S|}{2} &\geq |E(S,T)| \geq \frac{r|S||T|}{n} - \epsilon r\sqrt{|S||T|} \\ \frac{r|S|}{2} &\geq r|S|\frac{(1+\gamma)}{2} - \epsilon r\sqrt{|S||T|} \\ \implies |S| &\leq \frac{4\epsilon^2}{\gamma^2}n \leq \tau n \end{aligned}$$

when $\epsilon^2 \leq \frac{1}{4}\tau\gamma^2$.

Note that for obtaining the distance of the distance-amplified code, it sufficed to have a graph with $r = O(1/\gamma)$, however for decoding we need $r = O(1/\gamma^2)$.
This code is hence decodable up to $\frac{1-\gamma}{2}n$ with rate $\Omega(\gamma^2)$.

2.3 Improvements

MDS codes: Maximal Distance Separable codes are codes that achieve the singleton bound for a fixed alphabet size (ex: Reed-Solomon codes)

There exist explicit "near-MDS" codes of rate R, distance $\delta = 1 - R - \gamma$ on an alphabet of size $\exp(\text{poly}(1/\gamma))$ that are linear-time encodable and decodable up to a $(1-R-\gamma)/2$ fraction of errors constructed in [GI05]. We can also obtain linear-time decodable codes up to the Zyablov bound via concatenation codes.
Note that "linear" is technically $O_{1/\gamma}(n) = O(n \cdot \text{poly}(1/\gamma))$ since operations on the alphabet take time at least linear in the description of the alphabet.

Currently, linear-time decoding seem to be possible only using graph theoretic techniques like in Tanner codes [SS96, Zem01].
For encoding, using the generator matrix can take up to $O(n^2)$ time, while Reed-Solomon codes can be encoded in $O(n \log n)$ time via FFTs. There also exist explicit codes - Spielman codes [Spi96] that can be encoded in linear time.

3 Random errors - the Shannon approach

The field of information theory was started by Claude Shannon's seminal paper [Sha48] in 1948, and complemented by Hamming's paper [Ham50] which was from a combinatorial viewpoint (worst-case errors).

So far, we have considered the Hamming approach and constructed codes and decoders to handle worst-case error patterns. However, for $p < 1/4$ fraction of errors, we do not know the best rate codes in this setting, but only bounds (this can be fixed using list decoding, however).

3.1 Binary Symmetric Channel

The binary symmetric channel BSC_p flips each individual bit in a codeword with probability p independently; it is a memoryless channel. WLOG $p \in [0, 1/2]$.

The received erroneous codeword is a random variable $y = c + e$ where $e \sim Ber(p)^{\otimes n}$ is a vector where each coordinate $e_i \sim Ber(p)$.

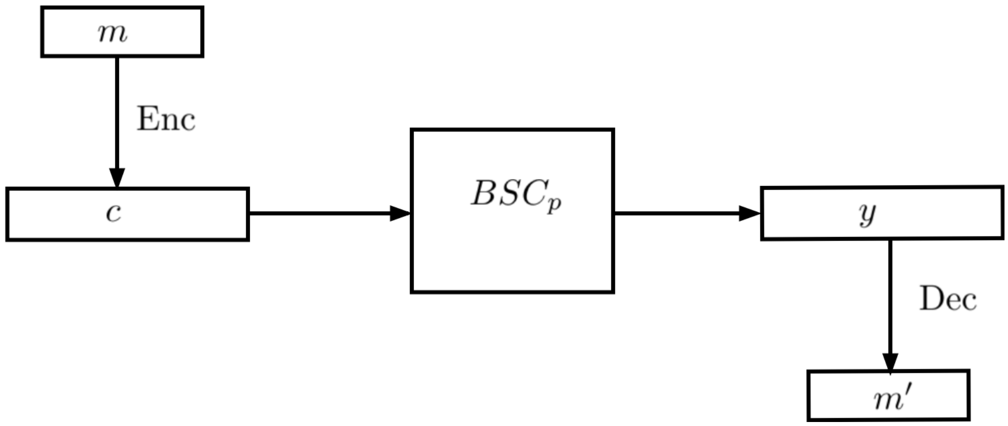

Figure 10.1: Channel coding

Note that $wt(e) \sim Bin(n, p)$. Hence, we can use a code which corrects p fraction of errors in the worst-case here, which would work most of the time.

3.2 Shannon's capacity theorem for BSC

We want to get the best possible rate (bounded away from 0) for an encoder Enc and decoder Dec such that for all messages m,

$$Pr_{e \in Ber(p)^{\otimes n}}[\mathsf{Dec}(\mathsf{Enc}(m) + e) \neq m] \to 0 \text{ as } n \to \infty$$

Since the rate is bounded away from zero and the decoder works better for larger n, this essentially gives a tradeoff between the reliability of decoding and the delay in receiving the entire message as $n \to \infty$.

Consider a ball of radius pn around a codeword $c = E(m)$ (which contains $\approx 2^{h(p)n}$ points). Notice that in this error model, most of the codewords are close to the boundary, and we want all of these to decode to m - which also means that a ball of this radius around any other codeword must be near-disjoint. This relaxation of disjoint-ness allows for better packing of codewords.

Hence, any code satisfying the decoding requirement above must have rate $R(C) \leq 1 - h(p) + \epsilon$ for all $\epsilon > 0$ (This is the *Converse theorem* to Shannon's capacity theorem).

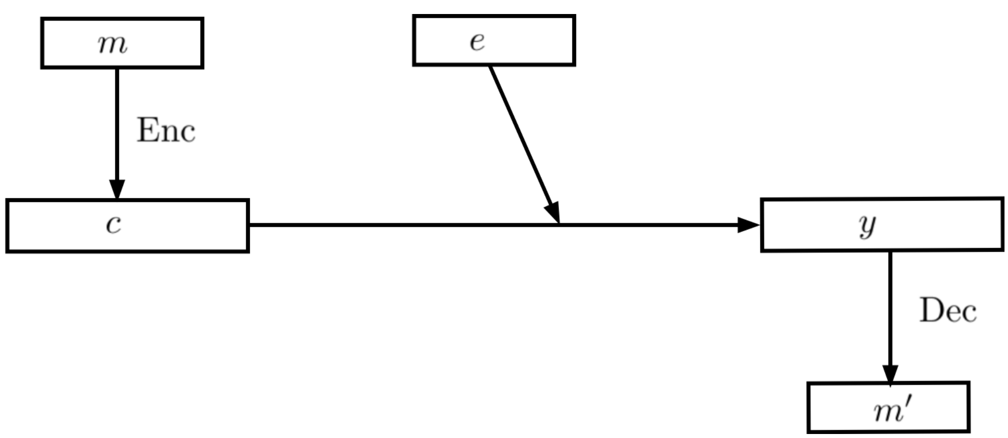

Figure 10.2: Converse theorem

Intuitively, this can be seen by checking the information transfer through the channel. Let e be chosen by the sender to contain $h(p)n$ bits of information, and m contains k bits of information. Then, if decoding is done correctly on the noisy codeword c (which is a lossy version of c), the receiver obtains m and can compute e as well. Hence, we must have that

$$\begin{aligned} k + h(p)n &\leq n + \epsilon n \\ \frac{k}{n} &\leq 1 - h(p) + \epsilon \end{aligned}$$

It turns out that we can actually obtain codes with this rate, unlike in the worst-case (Hamming) model.

Theorem 10.1 (Shannon's capacity theorem for BSC_p)**.** *$\forall p \in (0, 1/2)$, $\forall \gamma \in (0, 1/2 - p)$, there exists $\xi = \xi(\gamma, p)$ for all large enough $n \approx 1/\gamma^2$ and there exist $\mathsf{Enc} : \{0,1\}^k \to \{0,1\}^n$ with $k \geq (1 - h(p+\gamma))n$ and $\mathsf{Dec} : \{0,1\}^n \to \{0,1\}^k$ such that for all messages m,*

$$Pr_{e \in Ber(p)^{\otimes n}}[\mathsf{Dec}(\mathsf{Enc}(m) + e) \neq m] \leq 2^{-\xi n}$$

Proof. Fix $0 < p < 1/2$, $0 < \gamma < 1/2 - p$ and $k = (1 - h(p+\gamma))n$. Let $\mathsf{Enc}(m) = Gm$ for a random $n \times k$ matrix G. Assume that G has rank k (which is true with high probability).

Define the decoding algorithm Dec such that $\mathsf{Dec}(y)$ outputs m if m is the *unique* message that satisfies $\Delta(\mathsf{Enc}(m), y) \leq (p + \epsilon)n$ for some small $\epsilon = \epsilon(\gamma)$ and 0^k otherwise.

Notice that for any m, $Pr[\mathsf{Dec}$ fails for $m] = Pr[\mathsf{Dec}$ fails for $0^k]$. This is true due to linearity of the code - given an event of successful decoding of a vector in a ball around a codeword, we can translate the vector to 0^k and the associated noisy codeword will be decodable successfully as well (this is because the probability is over choices of the error pattern e, which is translation-invariant). Hence, it suffices to prove the theorem for $m = 0^k$.

The decoder fails in one of two ways:

1. $wt(e) > (p + \epsilon)n$
2. $\Delta(e, \mathsf{Enc}(m)) \leq (p + \epsilon)n$ (when there is another codeword closer to e than 0^k)

The first case happens with negligible probability due to the Chernoff bound - $Pr \leq 2^{-\Omega(\epsilon^2 n)}$.
For the second case, notice that for a random matrix G and fixed $m_0 \neq 0$, Gm_0 is distributed uniformly at random. Hence,

$$Pr[\Delta(e, \mathsf{Enc}(m_0)) \leq (p + e)n] \leq \frac{2^{h(p+\epsilon)n}}{2^n}$$

Taking the union bound over all 2^k messages,

$$Pr[\Delta(e, \mathsf{Enc}(m)) \leq (p + e)n] \leq 2^k \cdot \frac{2^{h(p+\epsilon)n}}{2^n} \leq 2^{-(h(p+\gamma) - h(p+\epsilon))n}$$

Combining both events, we can choose $\xi = \xi(\gamma, \epsilon)$ to get

$$Pr[\mathsf{Dec}(\mathsf{Enc}(m) + e) \neq m] \leq 2^{-\Omega(\epsilon^2 n)} + 2^{-(h(p+\gamma) - h(p+\epsilon))n} \leq 2^{-\xi n}$$

$\square$

References

[ABN$^+$92] N. Alon, J. Bruck, J. Naor, M. Naor, and R.M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38(2):509–516, 1992. 1

[GI05] V. Guruswami and P. Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Transactions on Information Theory*, 51(10):3393–3400, 2005. 2.3

[Ham50] R. W. Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160, 1950. 3

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. 3

[Spi96] D.A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 42(6):1723–1731, 1996. 2.3

[SS96] M. Sipser and D.A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6):1710–1722, 1996. 2.3

[Zem01] G. Zemor. On expander codes. *IEEE Transactions on Information Theory*, 47(2):835–837, 2001. 2.3