ECE253/CSE208 Introduction to Information Theory

Lecture 12: Channel Coding & Feedback Capacity

Dr. Yu Zhang

ECE Department
University of California, Santa Cruz

- Chap 7 of *Elements of Information Theory (2nd Edition) by Thomas Cover & Joy Thomas*.
- Vembu-Verdu-Steinberg, The Source-Channel Separation Theorem Revisited, *IEEE Trans. on Information Theory*, Jan. 1995.

Channel Coding

- The channel coding theorem promises the existence of good codes for transmitting information reliably at rates below capacity, if the block length is large enough.
- However, it does not provide a way to construct good codes. It took people a few decades to find capacity-achieving codes.
- In this class, we only study a simple channel coding scheme (invented by Richard Hamming in 1950), which illustrates basic ideas underlying most codes.

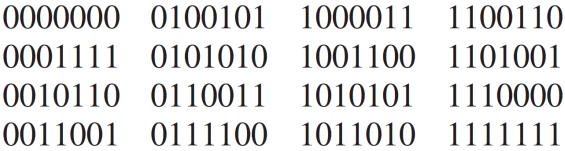
Repetition codes. Introduce redundancy naively: just repeat the codes a few times; e.g. $\{1 \rightarrow 11111;\ 0 \rightarrow 00000\}$. Do majority vote for decoding so that errors can be corrected as long as no more than 3 bits are corrupted. But this is not an efficient scheme since the rate is only $\frac{1}{5}$ bit per symbol.

Hamming Codes

$$\mathbf{H}_{3\times 7} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad \mathrm{rank}(\mathbf{H}) = 3, \quad \dim(\mathrm{null}(\mathbf{H})) = 4$$

Consider the matrix $\mathbf{H}$ (parity check matrix) whose columns collect all non-zero binary vectors of length 3. All 2^4 null space vectors of $\mathbf{H}$ form the set of codewords that are given as follows; i.e., (all operations are done modulo-2)

$$\left\{ \mathbf{c} \in \mathbb{B}^7 \mid \mathbf{Hc} = \mathbf{0} \right\}$$



0000000	0100101	1000011	1100110
0001111	0101010	1001100	1101001
0010110	0110011	1010101	1110000
0011001	0111100	1011010	1111111

Figure: (7,4,3) Hamming code.

Property of Hamming Codes

- Linear code: sum of any two codes is also a code.
- Minimum weight is 3: minimum number of 1's in any code is 3 (except for the all-zero code).
- Minimum distance is 3: minimum number of positions in which two codes differ.

Correct 1-digit error. If a code is corrupted only in one place, it can detected and corrected:

$$\mathbf{H}(\mathbf{c}+\mathbf{e}_i) = \mathbf{H}\mathbf{e}_i = \mathbf{H}(:, i) \ \leftarrow \ \text{the } i\text{-th column of } \mathbf{H},$$

where $\mathbf{e}_i$ is the i-th canonical vector.

Note that the first 4 bits of the 16 codes cycle through all 2^4 combinations of 4 bits, which can be used to represent 16 messages we want to send. The remaining 3 bits are parity check bits that are determined by the codes.

Generalization

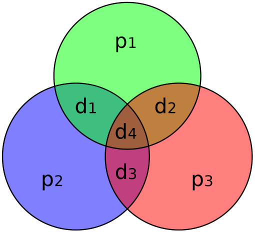


Figure: Hamming codes are a class of binary linear codes. The figure shows the Venn diagram of the (7,4,3) Hamming code: 4 information bits (d_1, d_2, d_3, d_4) are in the four intersection regions. The 3 parity bits (p_1, p_2, p_3) are added in the remaining regions so that the parity (number of 1's) of each circle is even. If one of the information bits is flipped, the parity of the two circles will change.

Generalization: For each integer $r \geq 2$ number of parity bits, we can construct a (n, k, d) Hamming code, where the block length $n = 2^r - 1$, message length $k = 2^r - r - 1$, and minimum distance $d = 3$. Hence, the rate of Hamming codes is $R = \frac{k}{n} = 1 - \frac{r}{2^r - 1}$ (the highest possible for codes with $d = 3$ and $n = 2^r - 1$). The parity check matrix $\mathbf{H}_{r \times n}$ is constructed by listing all non-zero vectors of length r.

Type of Channel Codes

- Block codes: map a block of information bits onto a codeword, no dependence on past info bits.

  A binary block code with minimum distance d can correct $\lfloor \frac{d-1}{2} \rfloor$ bit errors.

  Examples: `Hamming codes, Reed-Solomon (RS) codes, BCH codes, etc`.

- Convolutional codes: each output block depends also on some of the past inputs: Decoding can be efficiently implemented via the Viterbi algorithm.

  Examples: `Turbo codes, Polar codes, Low-density-parity-check (LDPC) codes, etc`.

Feedback Capacity

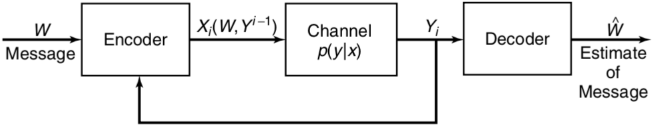


Figure: Discrete memoryless channel (DMC) with feedback.

For a DMC, feedback may help simplify encoding and decoding (e.g., for BEC), but will not increase the channel capacity.

Theorem (Feedback does not increase the channel capacity for a DMC)

For a DMC, $C_{FB} = C = \max_{p(x)} I(X;Y)$.

Proof: Clearly, $C_{FB} \geq C$. We need to show $C_{FB} \leq C$.

$$\begin{aligned} nR = H(W|\hat{W}) + I(W;\hat{W}) &\leq 1 + P_e^{(n)} nR + I(W;\hat{W}) \\ &\leq 1 + P_e^{(n)} nR + I(W;Y^n) \quad \text{[by DPI]} \end{aligned}$$

Proof Cont'd

$$I(W;Y^n) = H(Y^n) - H(Y^n|W) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},W)$$

$$= H(Y^n) - \sum_{i=1}^{n} H\left(Y_i|Y_1,\ldots,Y_{i-1},W,X_i(W,Y^{i-1})\right) \quad \text{[due to feedback]}$$

$$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) = \sum_{i=1}^{n} I(X_i;Y_i) \leq nC \implies$$

$$nR \leq 1 + P_e^{(n)} nR + I(W;Y^n) \leq 1 + P_e^{(n)} nR + nC \implies$$

$$\boxed{R \leq \frac{1}{n} + P_e^{(n)} R + C.}$$

- Taking $n \to \infty$ and $P_e^{(n)} \to 0$, we get $R \leq C$.
- $C_{\text{FB}} = C$: **A higher rate with feedback cannot be achieved**.

Joint Source-Channel Coding Theorem (Separation Theorem)

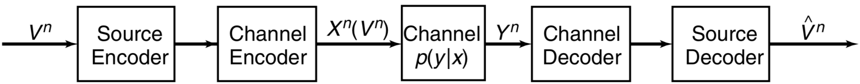


Figure: Separate source and channel coding.

Source coding aims to reduce redundancy as much as possible while channel coding (error-correction coding) is to combat channel noise by introducing structured redundancy.

- **Source coding theorem**: As long as the source symbols are compressed to $R_s > H$ information bits/source symbol, then lossless data compression is possible.
- **Channel coding theorem**: As long as $R_c < C$ information bits are transmitted per channel use, error-free transmission is possible.
- By using a two-step procedure (compression followed by encoding), we can send a source with entropy H reliably through a channel with capacity C provided $H < C$.

Joint Source-Channel Coding Theorem (Cont'd)

Q: Can we do better if we use a single-step joint source-channel coding procedure?

A: No, justified by the following joint source-channel coding theorem.

Theorem (A source with entropy rate H can be sent reliably over DMC iif $H < C$.)

If $V_1, V_2, \ldots V_n$ is a finite alphabet stochastic process that satisfies the AEP and $H(\mathcal{V}) < C$, there exists a source-channel code with probability of error $\Pr(\hat{V}^n \neq V_n) \to 0$. Conversely, for any stationary stochastic process, if $H(\mathcal{V}) > C$, the probability of error is bounded away from zero, and it is not possible to send the process over the channel with arbitrarily low probability of error.

Examples of finite alphabet stochastic processes satisfying the AEP:

1. A sequence of i.i.d. random variables
2. A stationary irreducible Markov chain
3. Any stationary ergodic process (time average = ensemble average): [*Shannon-McMillan-Breiman Theorem*].

Joint Source-Channel Coding Theorem (Cont'd)

Engineering implications: Two-stage is as good as one-stage process.

- Asymptotic optimality can be achieved by separating source and channel coding.
- Design source codes for the most efficient representation of the data.
- *Separately and independently* design channel codes appropriate for the channel.

Caveat: Two-stage is not always optimal.

- Sending English text over an erasure channel: corrupted bits are difficult to decode.
- Redundancy in the source is suited to the channel: speech for the human ear.
- Multiuser channels.

Vembu-Verdu-Steinberg studied the validity of the separation theorem in the context of very general sources and channels: A finer look at the statistical structure of the channel and source is necessary.

Thank You!

Email: <zhangy@ucsc.edu>

Homepage: https://people.ucsc.edu/~yzhan419/