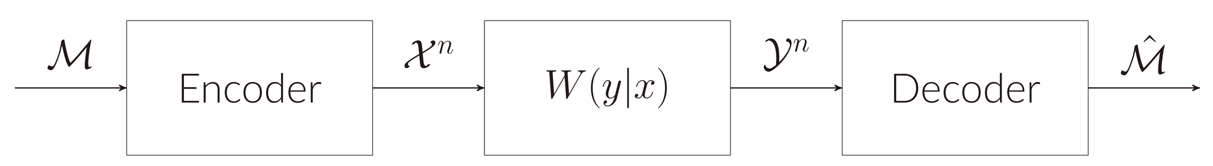
On the Zero-error Shannon Capacity of Graphs

Researcher: XXXXXXXXXX†, ♯ Supervisor: Amin Gohari †

The Chinese University of Hong Kong

Introduction

A point-to-point noisy channel models communication from a sender X to a receiver Y, shown below.

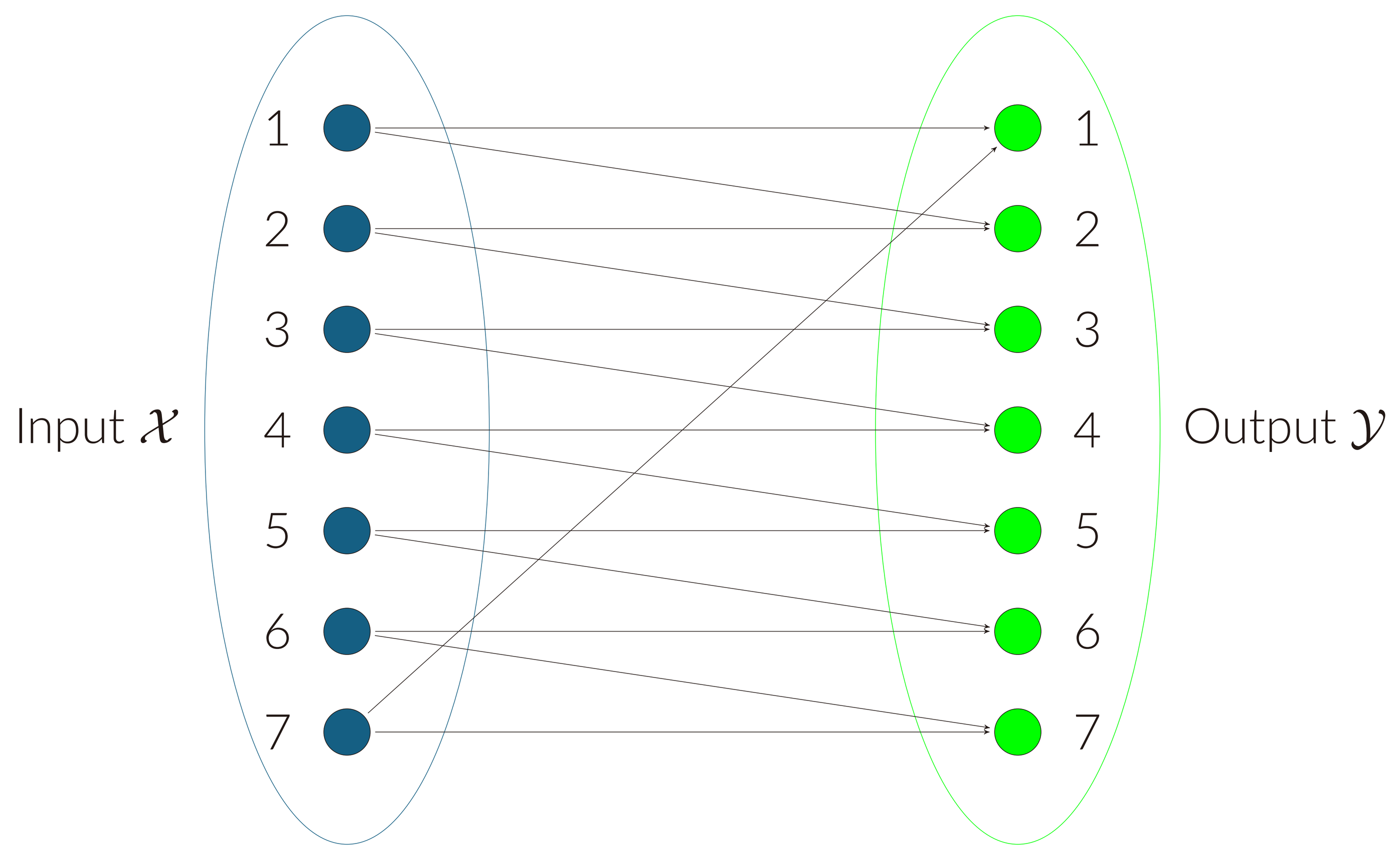


Vanishing-error Channel Capacity. An (n, R)-code is achievable if the probability of error $P(M \neq \hat{M}) \to 0$ when $n \to +\infty$. $C = \sup\{\text{achievable } R\}$ is called the vanishing-error channel capacity.

Zero-error Channel Capacity. An (n, R_0)-code is achievable if the probability of error $P(M \neq \hat{M}) = 0$ when $n \to +\infty$. $C_0 = \sup\{\text{achievable } R_0\}$ is called the zero-error channel capacity.

Research Objectives and Existing Results

In this project, we mainly consider the 7-letter noisy typewriter, shown below.



Main Problem. Find the most effective **zero-error** code for the above channel.

Equivalent Problem. Find the **largest** subset $\mathcal{X}_0 \subseteq \mathcal{X}$ such that the neighbourhood of vertices in $\mathcal{X}_0$ are **mutually disjoint**.

Existing Lower Bound The best known lower bound is $367^{\frac{1}{5}} \approx \mathbf{3.258}$ (result in 2019). The previous best known lower bound is $108^{\frac{1}{4}} \approx 3.224$

Existing Upper Bound The best known upper bound is $\mathbf{3.318}$, given by the Lóvasz theta function $\vartheta(G)$.

Proposed Methodologies and Difficulties

Method 1: Entropic Characterization. This method makes use of two information functionals in information theory.

- **Entropy** $H(X) = -\sum_x p(x) \log p(x)$. When the distribution p_X is uniform on a subset of X with cardinality M, then $H(X) = \log M$.
- **Conditional Entropy** $H(X|Y) = -\sum p(x, y) \log p(x|y)$. It has the property that $H(X|Y) = 0$ if and only if X is a **function** of Y.

Upper Bound Formulation. Using the properties above, we formulate the following upper bound for the general code-length n,

$$C_0 \leq \min_{\lambda \geq 0} \max_{p(x^n)} \min_{W_{Y^n|X^n}} \frac{1}{n}(H(X^n) - \lambda H(X^n|Y^n))$$

Difficulties. 1) Optimization over the probability simplex (i.e., all $p(x^n)$) is an **NP-Hard** problem. 2) Dimension grows **exponentially** with respect to n.

Method 2: Single-letterization. This method re-investigates the upper bound formulated above. The goal is to explore the connection between the n-letter expression and the single-letter expression.

Difficulties. 1) The n-letter expression has **numerous** way to perform expansion or groupings. It is hard to find an **optimal** way to express the single-letterized expression. 2) Some information functionals may **not** be **single-letterizable**.

Result from Method 1

Initially, we know that when $\lambda = 1$, the optimizer is located at the uniform distribution. Therefore, we would like to find the largest λ^* such that the optimizer is still located at the uniform distribution, i.e., largest λ^* such that $\nabla^2\Big(\frac{1}{n}\big(H(X^n) - \lambda^* H(X^n|Y^n)\big)\Big) \preceq 0.$

The **second order condition** of $\frac{1}{n}(H(X^n) - \lambda^* H(X^n|Y^n))$ being concave is equivalent to

$$(\lambda - 1)\mathbb{E}[L^2] - \lambda \mathbb{E}\big[\mathbb{E}[L|Y]^2\big] \leq 0 \Leftrightarrow 1 - \frac{1}{\lambda} = \min_{L:\mathbb{E}[L]=0, \mathbb{E}[L^2]\neq 0} \frac{\mathbb{E}\big[\mathbb{E}[L|Y]^2\big]}{\mathbb{E}[L^2]}$$

The value of the minimization problem on R.H.S. can be characterized by the **smallest singular value** of the matrix of size $|\mathcal{X}| \times |\mathcal{X}|$,

$$Q_{x,x'} = \sum_y p(y) \frac{p(x|y)p(x'|y)}{\sqrt{p(x)p(x')}}$$

The best upper bound we can get from this method is $\mathbf{3.376}$.

Result from Method 2

Denote $(X_1, \ldots, X_n)$ by X^n. We identified the generalized **Markov structure** under our background settings. Moreover, we single-letterized the two-letter expression with some suitable auxiliary terms and relaxation,

$$\begin{aligned} H(X^2) - \lambda H(X^2|Y^2) \leq& \sum_{i=1}^{2}(1-\lambda)H(X_i) + \lambda H(Y_i) - \lambda H(X_i|Y_i) \\ &+ \sup_{U \to X_1 \to Y_1} \mu I(X_1; U) - \lambda I(Y_1; U) + \sup_{V_1 \to X_2 \to Y_2} (\lambda - 1) I(V_1; X_2) - \mu(V_1; Y_2) \end{aligned}$$

We attempted to single-letterized the three-letter expression,

$$\begin{aligned} H(X^3) - \lambda H(X^3|Y^3) \leq& \sum_{i=1}^{3}(1-\lambda)H(X_i) + \lambda H(Y_i) - \lambda H(X_i|Y_i) \\ &+ \sup_{U_1 \to X_1 \to Y_1} \mu_1 I(X_1; U_1) - \lambda I(Y_1; U_1) + \sup_{U_3 \to X_1 \to Y_1} \mu_3 I(X_1; U_3) - \lambda I(Y_1; U_3) \\ &+ \sup_{U_2 \to X_2 \to Y_2} \mu_2 I(X_2; U_2) - \lambda I(Y_2; U_2) + \sup_{V_2 \to X_2 \to Y_2} (\lambda - 1) I(V_2; X_2) - \mu_2(V_2; Y_2) + \\ &\sup_{V_3 \to X_3 \to Y_3} (\lambda - 1) I(V_3; X_2) - \mu_3 I(V_3; Y_2) + \sup_{V_2 \to X_3 \to Y_3} (\lambda - 1) I(V_2; X_2) - \mu_2 I(V_2; Y_2) \\ &+ (1 - \lambda) I(X_1; X_2; X_3) + \lambda I(Y_1; Y_2; Y_3) \end{aligned}$$

The last two terms $I(X_1; X_2; X_3)$ and $I(Y_1; Y_2; Y_3)$ are not Shannon information measures. In general, we don't know how to single-letterize $I(X_1; \ldots; X_n)$.

Extension from Method 1

In Method 1, the largest λ^* is usually very close to 1, which indicates the function **lost concavity** for some small λ^*. Therefore, we suggested replacing the entropy functions with some **generalized** entropy functions and observing if there is any improvement.

We attempted the above simulation with two classes of entropy functions:
1) Rényi entropy $H_\alpha(X)$; 2) f-entropy $H_f(X)$.
Several simulations are performed, and no significant improvement giving a tighter upper bound.

Conclusion and Future Direction

In this project, we investigated a well-known open combinatorial problem - the value of C_7 and transformed it into a probabilistic optimization problem. We recovered some upper bound which is closed to the best known bound, but not yet improved beyond it.

In the future, we should investigate and try to derive more hidden properties of the objective function $H(X^n) - \lambda H(X^n|Y^n)$, which could probably simplify the optimization problem and lead to a better bound.

References

[1] L. Lovász, 'On the Shannon capacity of a graph,' *IEEE Transactions on Information theory*, vol. 25, no. 1, pp. 1-7, 1979.

[2] S. C. Polak and A. Schrijver, 'New lower bound on the Shannon capacity of C7 from circular graphs,' *Information Processing Letters*, vol. 143, pp. 37-40, 2019.

[3] S. Masiha, A. Gohari, S. Kamath and C Nair, 'On Maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover,' *arXiv preprint arXiv:1304.6133*, 2013.

†: Department of Information Engineering

♯: Department of Mathematics