*MAJOR TECHNICAL PROJECT ON*

# TOTAL CAPACITY COMPUTATION FOR DISCRETE MEMORYLESS POINT TO POINT CHANNELS

*INTERIM PROGRESS REPORT*

*to be submitted by*

**NAME**
**Roll. No.**

*for the award of the degree*
*of*

**BACHELOR OF TECHNOLOGY IN**
**COMPUTER SCIENCE ENGINEERING**

**SCHOOL OF COMPUTING AND ELECTRICAL ENGINEERING**
**INDIAN INSTITUTE OF TECHNOLOGY MANDI, MANDI**

**September 2018**

# 1 Introduction

Shannon [1] considered the point-to-point communication system where a sender wishes to reliably communicate a message $M$ at a rate $R$ bits per transmission to a receiver over a noisy communication channel. Information theory answers a fundamental question in communication theory: what is the ultimate transmission rate $R$ of the message and the answer is the channel capacity. Channel capacity is defined as the supremum of all achievable transmission rates. If we transmit the message below channel capacity an error free and reliable communication is possible otherwise reliable communication is not possible.

A simple discrete memoryless channel (DMC) model $(\mathcal{X},\ p(y|x),\ \mathcal{Y})$ [2] that consists of a finite input set $\mathcal{X}$, a finite output set $\mathcal{Y}$, and a collection of conditional pmfs $p(y|x)$ on $\mathcal{Y}$ for every $x \in \mathcal{X}$. Thus, when an input symbol $x \in \mathcal{X}$ is transmitted, the probability of receiving an output symbol $y \in \mathcal{Y}$ is $p(y|x)$. Here the term discrete memoryless (DM) refers to "finite-alphabet and stationary memoryless".

For a discrete memoryless channel $p(y|x)$, capacity is defined as $C = \max_{p(x)} I(X;Y)$. Here, we have to notice that mutual information $I(X;Y)$ is a concave function of input distribution $p(x)$ for a fixed discrete memoryless channel $p(y|x)$. A symmetric channels is defined as, where transition probabilities of channel matrices $p(y|x)$ using inputs as rows and outputs as columns, has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.

To calculate the capacity of elementary channels i.e. binary symmetric channel (BSC), binary erasure channel (BEC) or any symmetric channels, we use the symmetry to guess that capacity is achieved with $p(0) = p(1) = 1/2$ i.e. with uniform input distribution. Even in case of semi-symmetric channels (where either each row is a permutation of each other row or each column is a permutation of each other column of the transition probabilities of channel matrices) one can guess that capacity is achieved with $p(0) = p(1) = 1/2$ i.e. with uniform input distribution. Now, one might sometimes be interested in finding the capacity of a channel that is not symmetric and for which the optimum $p(x)$ cannot be guessed. Then the procedure is to use a computational power to find the maximum. Since the mutual information function is a concave function of input distribution, we need an algorithm involves varying input distribution $p(x)$ to increase $I(X;Y)$ until a local (and, therefore, global) maximum is reached [3]. Capacity calculation of a discrete memoryless channels is a non-linear optimization problem with both equality and inequality constraints. The Blahut Arimoto algorithm [4]is an iterative algorithm devised for this purpose. The generalized Blahut Arimoto algorithm is used for calculating capacity for discrete memoryless multiple terminal channels.

## 2 Background

Now, we have to show that the mutual information $I(X;Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$. we expand the mutual information

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X = x) \tag{1}$$

If $p(y|x)$ is fixed, then $p(y)$ is a linear function of $p(x)$. Hence $H(Y)$, which is a concave function of $p(y)$, is a concave function of $p(x)$. The second term is a linear function of $p(x)$. Hence, the difference is a concave function of $p(x)$.

Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ be a $\boldsymbol{K}$ dimensional vector with real-valued components defined in a region $\boldsymbol{R}$ of vector space. We define a region $\boldsymbol{R}$ to be convex if for each vector $\boldsymbol{\alpha}$ in $\boldsymbol{R}$ and each vector $\boldsymbol{\beta}$ in $\boldsymbol{R}$, the vector $\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}$ is in $\boldsymbol{R}$ for $0 \leq \theta \leq 1$. Geometrically, as $\theta$ goes from 0 to 1, $\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}$ traces out a straight line from $\boldsymbol{\beta}$ to $\boldsymbol{\alpha}$. Thus a region is convex if, for each pair of points in the region, the straight line between those points stays in the region.

An example of a convex region for our purposes is the region of probability vectors. A vector is defined to be a probability vector if its components are all nonnegative and sum to 1. To show that this region is convex, let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be probability vectors and let $\boldsymbol{\gamma} = \theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}$ for $0 \leq \theta \leq 1$. Then

$$\gamma_k = \theta\alpha_k + (1-\theta)\beta_k \tag{2}$$

Thus $\gamma_k \geq 0$, and also we have

$$\sum_{k=1}^{K} \gamma_k = \theta \sum_{k=1}^{K} \alpha_k + (1-\theta) \sum_{k=1}^{K} \beta_k = 1 \tag{3}$$

Thus $\boldsymbol{\gamma}$ is a probability vector and hence it is prooved that the region of probability vectors is convex in nature.

Now, we can say that capacity calculation of a discrete memoryless channels is a problem of concave optimization of mutual information over the convex set of input distribution for a fixed channel with both equality and inequality constraints.

A real-valued function $f$ of a vector is defined to be concave over a convex region $\boldsymbol{R}$ of vector space if, for all $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ in $\boldsymbol{R}$, and $\theta$, $0 \leq \theta \leq 1$, the function satisfies

$$\theta f(\boldsymbol{\alpha}) + (1-\theta) f(\boldsymbol{\beta}) \leq f(\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}) \tag{4}$$

If the inequality can be replaced with strict inequality, $f(\boldsymbol{\alpha})$ is strictly concave.
Figure (1) sketches the two sides of (4) as a function of $\theta$. It can be seen that the geometric interpretation of (4) is that every chord connecting two points on the surface representing the function must lie below the surface. The reason for restricting the definition to a convex region is to guarantee that the vector on the right-hand side of (4) is in $\boldsymbol{R}$.

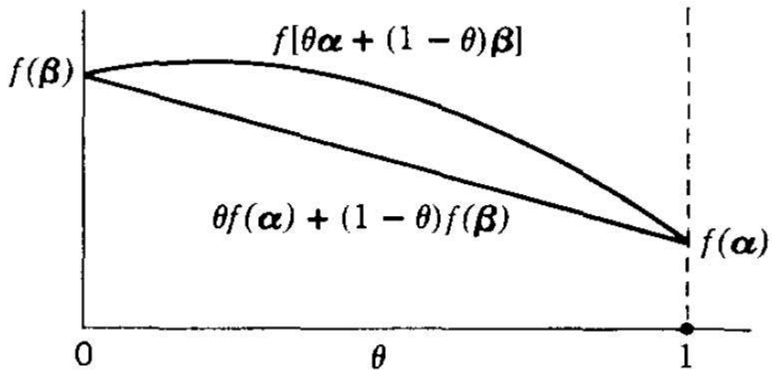


Figure 1: A concave function.

Now suppose that $f(\boldsymbol{\alpha})$ is a concave function in the region $\boldsymbol{R}$ where $\alpha_k \geq 0, 1 \leq k \leq K$. We might maximize $f(\boldsymbol{\alpha})$ over $\boldsymbol{R}$ by finding a stationary point of $f(\boldsymbol{\alpha})$, that is, an $\boldsymbol{\alpha}$ for which

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} = 0; \qquad 1 \leq k \leq K \tag{5}$$

The set of equations (5) might have no solution, and if there are solutions, they might not satisfy the constraints $\alpha_k \geq 0$. We now show that if there is an $f(\boldsymbol{\alpha})$ satisfying both (5) and the constraints, then that $f(\boldsymbol{\alpha})$ maximizes the function. To see this, we shall assume that the result is false: that there is some vector $f(\boldsymbol{\beta})$ in the region for which $f(\boldsymbol{\beta}) > f(\boldsymbol{\alpha})$. The chord joining $f(\boldsymbol{\beta})$ and $f(\boldsymbol{\alpha})$ is then increasing from $f(\boldsymbol{\alpha})$ to $f(\boldsymbol{\beta})$. As seen from Figure (1), the rate of increase of the function at $\boldsymbol{\alpha}$ in the direction of $\boldsymbol{\beta}$ is at least as large as that of the chord. Thus $\boldsymbol{\alpha}$ cannot be a stationary point of $f$ and we have arrived at a contradiction.
On the other hand, if $f$ is differentiable, still the maximum to be at a stationary point with respect to variation of the nonzero components of $\boldsymbol{\alpha}$. This suggests replacing (5) with the equations

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} = 0; \qquad \text{all } k \text{ such that } \alpha_k > 0 \tag{6}$$

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} \leq 0; \qquad \text{all } k \text{ such that } \alpha_k = 0 \tag{7}$$

if an $\boldsymbol{\alpha}$ in the region satisfies (6) and (7), it maximizes $f$, and, conversely, if $f$ is differentiable and has a maximum in the region at $\boldsymbol{\alpha}$, then (6) and (7) are satisfied.
Suppose that we want to maximize a convex function, $f(\boldsymbol{\alpha})$, over the region where $\boldsymbol{\alpha}$ is a probability vector, that is, where the components of $\boldsymbol{\alpha}$ are nonnegative and sum to 1. The constraint $\sum \alpha_k = 1$ suggests using a Lagrange multiplier, which implies maximizing $f(\boldsymbol{\alpha}) - \lambda \sum \alpha_k$ over the region where the components are nonnegative, and then choosing $\lambda$ so that the

maximum occurs at $\sum \alpha_k = 1$. Applying (6) and (7) to the function $f(\boldsymbol{\alpha}) - \lambda \sum \alpha_k$, we obtain

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} = \lambda; \qquad \text{all } k \text{ such that } \alpha_k > 0 \tag{8}$$

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} \leq \lambda; \qquad \text{all } k \text{ such that } \alpha_k = 0 \tag{9}$$

We shall see that (8) and (9) are indeed necessary and sufficient conditions on the probability vector $\boldsymbol{\alpha}$ that maximizes a concave differentiable function $f$. In other words, if a probability vector $\boldsymbol{\alpha}$ satisfies (8) and (9) for some value of $\lambda$, that $\boldsymbol{\alpha}$ maximizes $f$ over the region, and conversely if $\boldsymbol{\alpha}$ maximizes $f$ over the region, (8) and (9) are satisfied for some $\lambda$. Now, we move towards a proof of this result.

**Theorem 1.** Let $f(\boldsymbol{\alpha})$ be a concave function of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ over the region $\boldsymbol{R}$ when $\boldsymbol{\alpha}$ is a probability vector. Assume that the partial derivatives, $\partial f(\boldsymbol{\alpha})/\partial \alpha_k$ are defined and continuous over the region $\boldsymbol{R}$ with the possible exception that $\lim_{\alpha_k \to 0} \partial f(\boldsymbol{\alpha})/\partial \alpha_k$ may be $+\infty$. Then (8) and (9) are necessary and sufficient conditions on a probability vector $\boldsymbol{\alpha}$ to maximize $f$ over the region $\boldsymbol{R}$.

**Proof (Sufficiency)**. Assume that (8) and (9) are satisfied for some $\lambda$ and some probability vector $\boldsymbol{\alpha}$. We shall show that, for any other probability vector $\boldsymbol{\beta}$, $f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \leq 0$, thus establishing that $\boldsymbol{\alpha}$ maximizes $f$.
From the definition of concavity,

$$\theta f(\boldsymbol{\beta}) + (1-\theta) f(\boldsymbol{\alpha}) \leq f(\theta \boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}); \qquad 0 < \theta < 1 \tag{10}$$

Rearranging terms (10) becomes

$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \leq \frac{f[\theta \boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}] - f(\boldsymbol{\alpha})}{\theta} \tag{11}$$

Since (11) is valid for all $0 < \theta < 1$, but right hand term of the (11) is in indeterminate form for $\theta = 0$. Using L'Hospital's rule, we can pass to the limit as $\theta \to 0$, we have

$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \leq \frac{d(f[\theta \boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}] - f(\boldsymbol{\alpha}))}{d\theta} \bigg/ \frac{d\theta}{d\theta}\bigg|_{\theta=0} \tag{12}$$

$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \leq \frac{df[\theta \boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}]}{d\theta}\bigg|_{\theta=0} \tag{13}$$

Let, $\boldsymbol{x} = \theta \boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}$, then right hand side of (13) can be written as

$$\frac{df(\boldsymbol{x})}{d\theta} = \frac{df(\boldsymbol{x})}{d\boldsymbol{x}} . \frac{d\boldsymbol{x}}{d\theta}\bigg|_{\theta=0} \tag{14}$$

By differentiation, we obtain

$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \le \sum_k \frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k}(\beta_k - \alpha_k) \tag{15}$$

The existence of the derivative in (12) and the equivalence of (12) and (15) are guaranteed by the continuity of the partial derivatives. This continuity is given by hypothesis since (8) and (9) rule out the exceptional case where $\partial f(\boldsymbol{\alpha})/\partial \alpha_k = +\infty$. Observe now that

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k}(\beta_k - \alpha_k) \le \lambda(\beta_k - \alpha_k) \tag{16}$$

This follows from (8) if $\alpha_k > 0$. If $\alpha_k = 0$, we have, $\beta_k - \alpha_k \ge 0$, and (10) follows from (9). Substituting (16) into (15), we have

$$f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \le \lambda\left[\sum_k \beta_k - \sum_k \alpha_k\right] \tag{17}$$

Since $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are probability vectors, we have $f(\boldsymbol{\beta}) - f(\boldsymbol{\alpha}) \le 0$ for each $\boldsymbol{\beta}$ in the region.

**Necessity.** Let $\boldsymbol{\alpha}$ maximize $f$ over the region and assume, for the moment that the partial derivatives are continuous at $\boldsymbol{\alpha}$. Since $\boldsymbol{\alpha}$ maximizes $f$, we have

$$f(\theta\boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}) - f(\boldsymbol{\alpha}) \le 0 \tag{18}$$

for any probability vector $\boldsymbol{\beta}$ and any $\theta, 0 < \theta < 1$. Dividing by $\theta$, but left hand term of the (18) is in indeterminate form for $\theta = 0$. Using L'Hospital's rule, we can pass to the limit as $\theta \to 0$, we have

$$\left.\frac{df[\theta\boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}]}{d\theta} \bigg/ \frac{d\theta}{d\theta}\right|_{\theta=0} \le 0 \tag{19}$$

$$\left.\frac{df[\theta\boldsymbol{\beta} + (1-\theta)\boldsymbol{\alpha}]}{d\theta}\right|_{\theta=0} \le 0 \tag{20}$$

$$\sum_k \frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k}(\beta_k - \alpha_k) \le 0 \tag{21}$$

At least one component of $\boldsymbol{\alpha}$ is strictly positive, and we assume, for simplicity of notation, that $\alpha_1 > 0$. Let $i_k$ be a unit vector with a one in the $k$th position and zeros elsewhere and choose $\boldsymbol{\beta}$ as $\boldsymbol{\alpha} + \epsilon i_k - \epsilon i_1$. Since $\alpha_1 > 0$, $\boldsymbol{\beta}$ is a probability vector for $0 \le \epsilon \le \alpha_1$. Substituting this $\boldsymbol{\beta}$ in (21), we have

$$\epsilon\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} - \epsilon\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_1} \le 0 \tag{22}$$

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} \leq \frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_1} \tag{23}$$

If $\alpha_k > 0, \epsilon$ can also be chosen negative, in which case the inequality in (23) is reversed, yielding

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} = \frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_1}; \qquad \alpha_k > 0 \tag{24}$$

Finally, choosing $\lambda$ as $\partial f(\boldsymbol{\alpha})/\partial \alpha_1$, (23) and (24) become equivalent to (8) and (9). To complete the proof, we shall consider an $\boldsymbol{\alpha}$ for which $\partial f(\boldsymbol{\alpha})/\partial \alpha_k = +\infty$ for some $k$ and show that such an $\boldsymbol{\alpha}$ cannot maximize $f$. Assume, for simplicity of notation, that $\alpha_1 > 0$.

$$\begin{aligned}\frac{f(\boldsymbol{\alpha} + \epsilon i_k - \epsilon i_1) - f(\boldsymbol{\alpha})}{\epsilon} = & \frac{f(\boldsymbol{\alpha} + \epsilon i_k - \epsilon i_1) - f(\boldsymbol{\alpha} + \epsilon i_k)}{\epsilon} \\ & + \frac{f(\boldsymbol{\alpha} + \epsilon i_k) - f(\boldsymbol{\alpha})}{\epsilon}\end{aligned} \tag{25}$$

In the limit as $\epsilon \to 0$, the first term on the right-hand side of (25) remains bounded because of the continuity of $\partial f(\boldsymbol{\alpha})/\partial \alpha_1$. The second term blows up, so that the left side of (25) is positive for sufficiently small $\epsilon$. This shows that $\boldsymbol{\alpha}$ does not maximize $f$, completing the proof.

**Finding Channel Capacity for a Discrete Memoryless Channel.** A discrete memoryless channel is described by a probability transition matrix $Q = [Q_{k|j}]$ where $[Q_{k|j}]$ is the probability of receiving the $k$th output letter given that the $j$th input letter was transmitted. In general, $Q$ is not square. The capacity of the channel is defined as

$$C = \max_{p \in \boldsymbol{P}^n} I(p; Q) = \max_{p \in \boldsymbol{P}^n} \sum_j \sum_k p_j Q_{k|j} \log \frac{Q_{k|j}}{\sum_j p_j Q_{k|j}} \tag{26}$$

where,

$\boldsymbol{P}^n = \{p \in \boldsymbol{R}^n : p_j \geq 0 \; \forall j; \sum_j p_j = 1\}$

is the set of all probability distributions on the channel input, and $I(p, Q)$ is known as the mutual information between the channel input and channel output. It is usually convenient in applications to take base 2 so that $C$ is expressed in terms of bits-per-channel use; for theoretical work, natural logs are more convenient.

**Theorem 2.** Suppose the channel transition matrix $Q$ is $n \times m$. For any $m \times n$ transition matrix $P$, let

$$J(p, Q, P) = \sum_j \sum_k p_j Q_{k|j} \log \frac{P_{j|k}}{p_j} \tag{27}$$

Then the following is true.

(a) $C = \max_p \max_P J(p, Q, P)$

(b) For fixed $p$, $J(p,Q,P)$ is maximized by

$P_{j|k} = \frac{p_j Q_{k|j}}{\sum\limits_j p_j Q_{k|j}}$

(c) For fixed $p$, $J(p,Q,P)$ is maximized by

$P_j = \frac{\exp(\sum_k Q_{k|j} \log P_{j|k})}{\sum_j \exp(\sum_k Q_{k|j} \log P_{j|k})}$

**Theorem 3.** For any $p \in \boldsymbol{P}^n$, let

$$c_j(p) = \exp \sum_k Q_{k|j} \log \frac{Q_{k|j}}{\sum\limits_j p_j Q_{k|j}} \tag{28}$$

Then, if $p^0$ is any element of $\boldsymbol{P}^n$ with all components strictly positive, the sequence of probability vectors defined by

$$p_j^{r+1} = p_j^r \frac{c_j^r}{\sum\limits_j p_j^r c_j^r} \tag{29}$$

is such that $I(p^r, Q) \to C$ as $r \to \infty$.

# 3 Problem Statement

1. To understand and implement the iterative algorithms: (1) Blahut Arimoto algorithm [4] for the capacity calculation of point to point discrete memoryless channel and (2) generalized Blahut Arimoto algorithm for calculating capacity for discrete memoryless multiple access channel [5].

2. To understand and implement algorithms to compute inner and outer bounds for the MAC capacity region [6].

3. To implement new algorithms to compute inner and outer bounds for the MAC capacity region.

4. To numerically compare the inner and outer bounds.

# 4 Tentative Plan

Capacity calculation of discrete memoryless channels is a non-linear optimization problem with equality and inequality constraints. The Blahut Arimoto algorithm is an iterative algorithm devised for this purpose. The computation of channel capacity of a single user discrete memoryless channels can be done by various numerical optimization techniques but for multiple access channels optimization is non-convex. It is found in recent studies that KKT condition is sufficient to find the optimality or channel can be further decomposed into sub channels for

which this condition is sufficient for optimality in which at least one channel achieves the capacity of the original channel. The generalized BA algorithm can be used to find the total capacity of the channel. The generalized BA algorithm will be implemented in this project for the calculation of the total capacity of multiple terminal channels.
**Run the spell check before submitting the report**- Remove this comment in final stage.

# References


[1] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[2] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[4] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Information Theory*, vol. 18, pp. 460–473, 1972.

[5] M. Rezaeian and A. Grant, "A generalization of arimoto-blahut algorithm," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, p. 181, IEEE, 2004.

[6] E. Calvo, D. P. Palomar, J. R. Fonollosa, and J. Vidal, "On the computation of the capacity region of the discrete mac," *IEEE Transactions on Communications*, vol. 58, no. 12, pp. 3512–3525, 2010.