ARBITRARY CHANNEL CAPACITY

Shashank V Maiya

University of Notre Dame

I. Introduction

Channel capacity is used to specify the asymptotic limit on the maximum rate at which information can be conveyed reliably over a channel. Any coding scheme that superficially appears to operate at a rate higher than C will cause enough data to be lost because of uncorrectable channel errors so that the actual information is not to be greater than C [5]. On the other hand, Rate Distortion Function $R(D)$ is a measure of the amount of information per unit of time that is required to specify the output of an ergodic message source within an average distortion D. This means that the channel capacity must be greater than or equal to $R(D)$ nats per unit of time if the message stream emitted by the source is to be transmitted to the destination with an average distortion D.
Evaluation of Channel Capacity, C, or the Rate distortion function $R(D)$ involves the solution of a convex programming problem. In most cases, analytic solutions cannot be found. An iterative algorithm is discussed in order to calculate the value of C [1] as well as $R(D)$ [2]. Few simulated results for the Capacity of the channel have been provided. Finally a few inequalities giving the upper and lower bounds on the capacity [3] as well as Rate Distortion Function [2] are presented.

II. Capacity of Unconstrained Discrete Channels

Denote a discrete memoryless channel with n inputs and m outputs symbols by a stochastic $m \times n$ matrix P.

$$P = \{p(i|j)\}, i = 1, \dots, m; j = 1, \dots, n \tag{1}$$

such that $p(i|j) \geq 0$ and $\sum_{i=1}^{m} p(i|j) = 1$; where $p(i|j)$ is the probability of receiving the i^{th} output symbol given that the j^{th} symbol was transmitted.
The Capacity of the channel is defined as

$$C = \max_{p \in \bar{D}^n} I(P;p) = max \sum_j \sum_i p_j p(i|j) \log \frac{p(i|j)}{\sum_j p_j p(i|j)} \tag{2}$$

where

$$\bar{D}^n = \{p = (p_1, p_2, \dots, p_n) | p_i \geq 0 \forall i; \textstyle\sum_{i=1}^{n} p_i = 1\}$$

Consider a stochastic $n \times m$ matrix ϕ such that

$$\phi = \{\phi(j|i)\}, j = 1, \dots, n; i = 1, \dots, m \tag{3}$$

such that $\phi(j|i) \geq 0$ and $\sum_{j=1}^{n} \phi(j|i) = 1$.
Let Φ be the set of all stochastic matrices satisfying these conditions. Then if

$$J(P;p,\phi) = -\textstyle\sum_{i=1}^{m} \sum_{j=1}^{n} p(i|j) p_j \log \phi(j|i)$$

$$C(P) = \max_{p \in \bar{D}^n} \max_{\phi \in \Phi} \left[H(p) - J(P;p,\phi)\right] \tag{4}$$

A. Iterative Procedure of Calculating Capacity

The iterative procedure of Calculating Capacity [1] comprises of the following steps :

1) Initially choose an arbitrary probability vector p^1 where p^1 = probability vector $(p_1^o, \ldots, p_n^o) \in \bar{D}_n$ (say uniform distribution $p_j^1 = \frac{1}{n}\ \forall j$. Then follow the two steps iteratively.
2) Maximize $H(p^t) - J(P; p^t, \phi)$ with respect to $\phi \in \Phi$ while fixing p^t. This results in

$$\phi^t(j|i) = \frac{p(i|j)p_j^t}{\sum_{k=1}^{n} p(i|k)p_k^t} \tag{5}$$

$$C(t,t) = H(p^t) - J(P; p^t, \phi^t) \tag{6}$$

3) Maximize $H(p) - J(P; p, \phi^t)$ with respect to $p \in \bar{D}^n$ while fixing ϕ^t. This results in

$$p_j^{t+1} = \frac{r_j^t}{\sum_{k=1}^{n} r_k^t} \tag{7}$$

$$r_j^t = exp\left[\sum_{i=1}^{m} p(i|j) \log \phi^t(j|i)\right] \tag{8}$$

$$C(t+1,t) = H(p^{t+1}) - J(P; p^{t+1}, \phi^t) \tag{9}$$

Proof:

For a fixed Input Probability Distribution p^t, Let

$$\phi^*(j|i) = \frac{p_j p(i|j)}{\sum_{k=1}^{n} p_k p(i|k)} \tag{10}$$

$$q_i = \sum_{k=1}^{n} p_k p(i|k) \tag{11}$$

Now consider

$$\begin{aligned}\sum_j \sum_i p_j p(i|j) \log \phi^*(j|i) - \sum_j \sum_i p_j p(i|j) \log \phi(j|i) &= \sum_j \sum_i q_i \phi^*(j|i) \log \frac{\phi^*(j|i)}{\phi(j|i)} \\ &= \sum_j \sum_i q_i \phi^*(j|i)\big(1 - \frac{\phi^*(j|i)}{\phi(j|i)}) \\ &= 0\end{aligned}$$

Therefore, $\phi^*(i|j) = \frac{p_j p(i|j)}{\sum_{k=1}^{n} p_k p(i|k)}$ maximizes $H(p^t) - J(P; p^t, \phi)$ with respect to ϕ for a constant p^t.

Maximizing with respect to ϕ^t; consider a Lagrange multiplier

$$f(p) = H(p) - J(P; p, \phi) + \lambda(1 - \sum_{j=1}^{n} p_j) \tag{12}$$

Maximizing by taking the derivative $\frac{\delta f(p)}{\delta p_j}|_{p=p_j^\star} = 0$ yields

$$p_j^\star = exp\big[-1 - \lambda + \sum_{i=1}^{m} p(i|j) \log \phi(j|i)\big], p_j^\star \geq 0 \tag{13}$$

Now taking into account $\sum p_j^\star = 1$ and finding λ from this condition, we get the required result.

B. *Convergence*

$$C(t+1,t) = \max\left[H(p^{t+1}) - J(P;p^{t+1},\phi^t)\right]$$

Substituting $p_j^{t+1} = \frac{r_j^t}{\sum_{k=1}^n r_k^t}$ in the above equation and simplifying [1], we get

$$C(t+1,t) = \log(\sum_{k=1}^{n} r_k^t) \tag{14}$$

Since we are maximizing capacity at each step, we have,

$$C(1,1) \leq C(2,1) \leq C(2,2) \cdots \leq C(t,t) \leq C(t+1,t) \tag{15}$$

Let p^o be one of the input probability vectors achieving capacity. Then

$$I(P;p^o) = C(P)$$
$$q_i^o = \sum_{j=1}^{n} p(i|j)p_j^o$$
$$q_i^t = \sum_{j=1}^{n} p(i|j)p_j^k$$

Now

$$\begin{aligned}\sum_{k=1}^{n} p_k^o \log(\frac{p_{t+1}^k}{p_t^k}) &= \sum_{k=1}^{n} p_k^o \log(\frac{r_t^k}{p_t^k \sum_j r_j^t}) \\ &= -C(t+1,t) + \sum_{k=1}^{n} p_k^o \big(\sum_{i=1}^{m} p(i|k) \log \frac{p(i|k)}{q_t^i}\big) \\ &= C(P) - C(t+1,t) + \sum_{i=1}^{m} q_i^o \log \frac{q_i^o}{q_t^i}\end{aligned}$$

Since

$$\begin{aligned}\sum_{i=1}^{m} q_i^o \log \frac{q_i^o}{q_t^i} &\geq \sum_{i=1}^{m} q_i^o (1 - \frac{q_i^t}{q_o^i}) \\ &= 0\end{aligned}$$

$$C(P) - C(t+1,t) \leq \sum_{k=1}^{n} p_k^o \log(\frac{p_{t+1}^k}{p_t^k}) \tag{16}$$

Summing up for $t = 1$ to N we get,

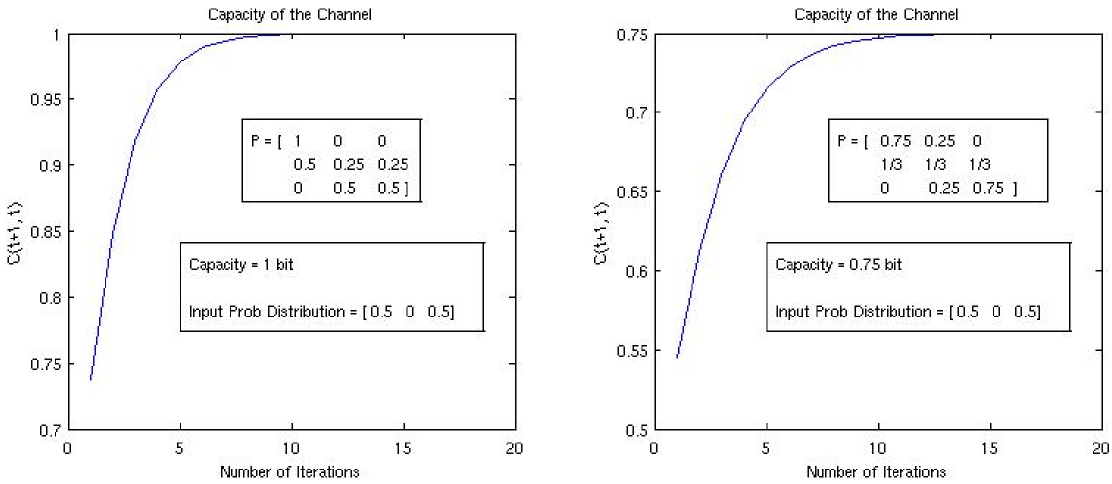
$$\sum_{t=1}^{N} (C(P) - C(t+1,t)) \leq \sum_{k=1}^{n} p_k^o \log(\frac{p_k^{N+1}}{p_k^1}) \tag{17}$$

$$\leq \sum_{k=1}^{n} p_k^o \log(\frac{p_k^o}{p_k^1}) \tag{18}$$

The right hand side of the inequality is a finite quantity. Since the sum of a non negative, non increasing series converges to a finite value,

$\lim_{t\to\infty} \big[C(p) - C(t+1,t)\big] = 0$ or in other words $C(t+1,t)$ converges to the Capacity.

C. Simulation Results



Plot of "Capacity" Vs "Number of Iterations" for two different Channels

D. Corollaries

1) The approximation error $e(t) = C(P) - C(t+1,t)$ is inversely proportional to the number of iterations. In particular if p^1 is chosen to be uniform, then

$$C(P) - C(t+1,t) \leq \frac{\log n - H(p^0)}{t} \tag{19}$$

2) $\lim_{n\to\infty} q^N = q^0$ That is the output distribution for any input capacity achieving distribution is the same. To prove this, consider a series $\omega(N) = \sum_{i=1}^{m} q_i^0 \log(\frac{q_i^0}{q_i^N})$.
Now $\sum_{t=1}^{N} \omega(t)$ converges to a finite value [1]. Hence $\omega(N) \to 0$ as $N \to \infty$

E. Bounds on Capacity

Here we make use of two results

1) The capacity of a cascade channels is never greater than that of the member with the smallest capacity. That is if $A = BC$, then

$$Cap(A) \leq Cap(B)$$
$$Cap(A) \leq Cap(C).$$

2) If A and B are n by m channels, then for any real number $0 \leq \alpha \leq 1$

$$Cap\big[\alpha A + (1-\alpha)B\big] \leq \alpha Cap(A) + (1-\alpha)Cap(B) \tag{20}$$

1) Lower Bound: Let $P = p_{ij}$ be an arbitrary $n \times m$ stochastic matrix. Define

$$\Delta_M = \text{Sum of the column maxima} = \sum_j \max_i p_{ij}$$
$$\Delta_m = \text{Sum of the column minima} = \sum_j \min_i p_{ij}$$

Now we choose a pair of stochastic matrices A and B, all of whose elements are either 1 or 0, in such a way that the sum of the elements on the main diagonal of the $n \times n$ stochastic matrix $W = APB$ is equal to Δ_M. Let $i_1, i_2 \ldots i_m$ be the indices of those rows of P containing column maxima. Let J_k be the set of column indices of

P which contains column maxima of i_k^{th} row.
Now by defining $A = (a_{ij})$ by

$$a_{ji_j} = 1 \qquad \text{for } j = 1, 2 \ldots m$$
$$\text{for } j > m, \text{ the row unities of A may be placed anywhere.}$$

and $B = (b_{ij})$ where

$$b_{ij} = 1 \qquad i \varepsilon J_j \; ; \; j = 1, 2, \ldots m,$$

we achieve the required results [3]. That is the resulting matrix W will have Δ_M as its trace.

Next consider the set $\{R_i\}$ of all $n!$ permutations matrices of order n, and the $n \times n$ stochastic matrix $Q = \sum_{i=1}^{n} (1/n!) R_i W R_i$. Then the matrix Q [3] will have

$$q_{ii} = \Delta_M / n \qquad \text{as the main diagonal elements}$$
$$q_{ij} = \frac{1 - \Delta_M/n}{n-1} \;, \qquad i \neq j \qquad \text{as the off diagonal elements.}$$

We could repeat the entire argument dealing with Δ_m instead of Δ_M
in which case $q_{ii} = \Delta_m / n$ and $q_{ij} = \frac{1 - \Delta_m/n}{n-1}$. Now,

$$\begin{aligned} Cap(Q) &= Cap(\sum_{i=1}^{n} (1/n!) R_i W R_i) \\ &\leq \sum_{i=1}^{n} (1/n!) Cap(W) \\ &= \sum_{i=1}^{n} (1/n!) Cap(APB) \\ Cap(Q) &\leq Cap(P) \end{aligned}$$

Therefore For any $n \times m$ channel P

$$Cap(P) \geq \max \begin{cases} \log(n) - (1 - \frac{\Delta_M}{n}) log(n-1) - H(\frac{\Delta_M}{n}) \\ \log(n) - (1 - \frac{\Delta_m}{n}) log(n-1) - H(\frac{\Delta_m}{n}) \end{cases}$$

2) Upper Bound: For a $n \times m$ channel P

$$Cap(P) \leq \log m - \delta(m-1) \log(m-1) - H(\delta(m-1)) \qquad (21)$$

where $\delta = \min_{i,j} p_{ij}$. Consider $P = RQ$ where

$$\begin{aligned} Q : q_{ij} &= \delta, & i \neq j \\ &= 1 - \delta(m-1), & i = j \end{aligned}$$

It can be shown that [3] all the rows of R sum up to unity and all the elements are non negative. Hence the matrix R is stochastic. Thus

$$Cap(P) \leq Cap(Q) = \log m - \delta(m-1) \log(m-1) - H(\delta(m-1))$$

Following similar procedure [3], another upper bound could be obtained for a $n \times m$ channel P

$$Cap(P) \leq \log n - \sigma(n-1) \log(n-1) - H(\sigma(n-1)) \qquad (22)$$

where $\sigma = \min_j \{ \frac{min(q_{ij})}{\sum_i q_{ij}} \}$

III. Rate Distortion Functions for Discrete Sources

A distortion matrix with elements ρ_{ji} specifies the distortion associated with reproducing the j^{th} source letter by the i^{th} reproducing letter. $0 \leq j \leq n-1$, $0 \leq i \leq m-1$ The rate distortion function is defined as

$$R(D) = \min_{P \in P_D} \sum_j \sum{}_i p_j p(i|j) \log \frac{p(i|j)}{\sum_i p_j p(i|j)} \tag{23}$$

$$= \min_{P \in P_D} I(p; P) \tag{24}$$

where $P_D = \{p \in R^n \times R^m : \sum_{i=1}^m p(i|j) = 1$ and $p(i|j) \geq 0, d(P) \leq D\}$
and $d(P) = \sum_j \sum_i p_j p(i|j)\rho_{ji}$

The investigation of Rage Distortion Functions is usually carried out parametrically in terms of a parameter s, which is introduced as a Lagrange multiplier. This parameter turns out to be equal to the slope of the Rate Distortion Curve at the point it parameterizes [4].

$$R(D) = \min_P \left[\sum_j \sum_i p_j p(i|j) \log \frac{p(i|j)}{\sum_j p_j p(i|j)} - s(\sum_j \sum_i p_j p(i|j)\rho_{ji} - D) \right] \tag{25}$$

$$D = \sum_j \sum_i p_j p^*(i|j)\rho_{ji} \tag{26}$$

A. *Iterative Algorithm for calculating Rate Distortion Function*

Let

$$F(p, P, q) = \sum_j \sum_i p_j p(i|j) \log \frac{p(i|j)}{q_i} - \sum_j p_j p(i|j) - s \sum_j \sum_i p_j p(i|j)\rho_{ji} \tag{27}$$

Then

$$R(D) = sD + \min_q \min_P F(p, P, q) \tag{28}$$

The iterative algorithm [2] involves the following steps

1) Initially choose an arbitrary output probability vector q^o (say uniform distribution). Then follow the 2 steps iteratively.
2) For a fixed P, $F(p, P, q)$ is minimized by

$$q_i = \sum_j p_j p(i|j) \tag{29}$$

3) For a fixed q, $F(p, P, q)$ is minimized by

$$p(i|j) = \frac{q_i exp(s\rho_{ji})}{\sum_i q_i exp(s\rho_{ji})} \tag{30}$$

For a fixed P, consider

$$\begin{aligned} \sum_j \sum_i p_j p(i|j) \log \frac{p(i|j)}{q_i} - I(p; P) &= \sum_j \sum_i p_j p(i|j) \log \frac{\sum_j p_j p(i|j)}{q_i} \\ &\geq \sum_j \sum_i p_j p(i|j) - \sum_j q_i \\ &= 0 \end{aligned}$$

with equality if $q_i = \sum_j p_j p(i|j)$
For a fixed q introducing a Lagrange multiplier to constrain $\sum_i p(i|j) = 1$ and minimizing with respect to $p(i|j)$ gives the reqired result [2].

B. Bounds on Rate Distortion

1) Upper Bound:

$$R(D) \leq sD - \sum_j \rho_j \log \sum_i A_{ji} q_i - \sum_i q_i c_i \log c_i \tag{31}$$

where

$$A_{ji} = exp(s\rho_{ji})$$
$$c_i = \sum_j \frac{p_j A_{ji}}{\sum_i A_{ji} q_i^r}$$

The proof [2] is a direct consequence of the result $R(D) \leq I(p; P)$

2) Lower Bound:

$$R(D) \geq sD - \sum_j \rho_j \log \sum_i A_{ji} q_i - \max_i \log c_i \tag{32}$$

A lower bound theorem for Rate Distortion states that

$$R(D) \geq sD + \sum_j p_j \log \lambda_j \tag{33}$$

where λ_j is such that $\sum_j p_j \lambda_j A_{ji} \leq 1$.
By letting $\lambda_j = (c_{max} \sum_k A_{ji} q_i)^{-1}$, we get the desired lower bound [2].

IV. References and Bibiliography

1) S. Arimoto, "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels". IEEE Trans. Inform. Theory. Jan 1972.
2) R. Blahut, "Computation of Channel Capacity and Rate Distortion Functions". IEEE Trans. Inform. Theory. July 1972.
3) H. J. Helgert, " A comparison of Arbitrary and Symettric Channels on the basis of Capacity". IEEE. Trans. Inform. Theory. vol IT-13, July 1967, pp 467-470
4) B. G. Haskell, " The Computation of Bounding of Rate Distortion Functions". IEEE Trans. Inform Theory. vol IT-15 pp 525-531, Sept 1969.
5) Thomas. M. Cover and Jay A Thomas, "Elements of Information Theory". Wiley Series in Telecommunications. pp 338-342 and pp 194-203