Computation of the Capacity for Discrete Memoryless Channels and Sum Capacity of Multiple Access Channels with Causal Side Information at Receiver

Mohammad Osmani Bojd
Department of Electrical Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
mohammad.osmani@gmail.com

Assadallah Sahebalam
Department of Electrical Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
a_sahebalam@yahoo.com

Ghosheh Abed Hodtani
Department of Electrical Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
ghodtani@gmail.com

Abstract—**Capacity analysis and computation for communication channels with side information at the receiver is an important area of research. Cover-Chiang have proved the increment of capacity with non-causal information at the receiver. We investigate this capacity increment for the channel with causal side information and proposed an efficient algorithm for computing the capacity of discrete memoryless channels and the sum capacity of multiple access channels with causal side information at the receiver.**

Keywords- Capacity computation, Causal side information, Discrete multiple access channel, Sum capacity

I. Introduction

A numerical algorithm is known as the Arimoto-Blahut algorithm, has been introduced in [1], [2], and [3] for computing the capacity of discrete memoryless channels. This algorithm has been successfully extended to the calculation of the sum capacity of discrete multiple access channels [4], [5], [6], [7]. In [8] an algorithm for computing channel capacity and rate-distortion with non causal side information is introduced and the capacity of channel with causal side information at the transmitter is computed in [9].

Fading coefficients, channel interference levels, states of a markov channel and channel gains are some examples of channel states and adaptive rate/power control over Rayleigh fading channels, MIMO beam-forming, precoding and multi-tone water filling are some scenarios and applications for improving the channel performance with side information.

In this paper, first, we propose an efficient algorithm for computing the channel capacity with causal side information at the receiver and obtain a closed form. Then we present another algorithm according to the previous algorithm for numerical computation of sum capacity of discrete multiple access channels with causal side information at the receiver.

II. Notations and Preliminaries

We use $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{s}$ to denote the input vector $(x_1 \dots x_n)$, output vector $(y_1 \dots y_n)$ and state vector $(s_1 \dots s_n)$, respectively, and allow $\mathbf{x}^i$, $\mathbf{y}^i$ and $\mathbf{s}^i$ to denote $(x_1 \dots x_i)$, $(y_1 \dots y_i)$ and $(s_1 \dots s_i)$, respectively. $P(s)$, $P(x)$, $P(x|y)$ represent probability distribution functions. Also, $\mathcal{X}, \mathcal{Y}$ and $\mathcal{S}$ represent the input alphabet, output alphabet and state alphabet, respectively. And finally, C^* is used for capacity of channels with side information at the receiver.

Causal Side Information at the receiver

We begin with a brief review of discrete memoryless channel with causal side information. Consider the channel depicted in Fig. 1. The channel is discrete with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and state alphabet $\mathcal{S}$, all of which are finite sets. The channel states are i.i.d with distribution $P(\mathbf{s})$ and independent of the input sequence. Furthermore, given the states, the channel is memoryless with transition distribution $P(\mathbf{y}|\mathbf{x},\mathbf{s})$. Hence, the conditional distribution of $\mathbf{y}$ and $\mathbf{s}$ given $\mathbf{x}$ can be written as:

$$P(\mathbf{y},\mathbf{s}|\mathbf{x}) = P(\mathbf{s})P(\mathbf{y}|\mathbf{x},\mathbf{s}) = \prod_{i=1}^{n} P(s_i)P(y_i|x_i,s_i) \quad (1)$$

The state sequence in this model plays the role of 'side information'. The receiver decodes the message ω from received vector $\mathbf{y}$ and $\mathbf{s}$ as,

$$\hat{\omega} = g(\mathbf{y},\mathbf{s}) \quad (2)$$

The capacity with causal side information at the receiver is given by [10],

$$C^* = \max_{P(x)} I(x;y|s) \quad (3),$$

$$P(s,x,y) = P(s,x)P(y|x,s) = P(x)P(s)P(y|x,s) \quad (4)$$

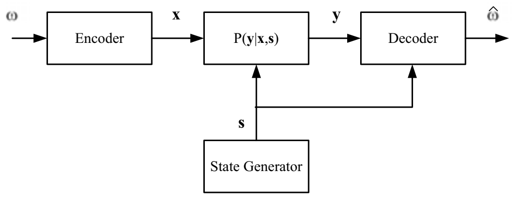

Figure 1. Channel with side information at the receiver

Annual International Conference on Information Theory and Applications (ITA 2011)
Edited by Prof. Gautam Sanyal and Dr. Kalpana Yadav

ISBN: 978-981-08-8113-9
doi:10.5176/978-981-08-8113-9_ITA07

III. COMPUTING THE CAPACITY FOR CHANNEL WITH CAUSAL SIDE INFORMATION AT THE RECEIVER

It can be clearly seen that the capacity of channel with side information at the receiver can be expressed as the following form

$$C^* = \max_{P(x)} I(x; y, s) \tag{5}$$

The above equality follow from the fact that the state s is independent from input of the channel i.e. x ; consequently, $I(x; y|s) = I(x; y, s)$.

We can obtain a lower band for the capacity of channel with causal side information at the receiver as following

$$I(x; y|s) = H(x|s) - H(x|y, s) \tag{6}$$

$$\geq H(x) - H(x|y) \tag{7}$$

$$= I(x; y)$$

where (6) and (7) follow from the fact that x and s are independent and the removing conditioning.

Hence,

$$C^* \geq \max_{P(x)} I(x; y) \tag{8}$$

where the above inequality shows that the capacity of channel with side information at the receiver is greater than the ordinary discrete memoryless channel without side information.

Theorem 1. Let $P(y|x, s)$ be a given conditional distribution for a channel with side information at the receiver and state distribution P(s). Then the distribution P(x) that maximizes the capacity of channel with causal side information at the receiver is,

$$P(x) = \frac{\prod_{y,s} P(x|y,s)^{P(s)P(y|x,s)}}{\sum_x \prod_{y,s} P(x|y,s)^{P(s)P(y|x,s)}} \tag{9}$$

And the capacity of this channel can be written as the following expression,

$$C^* = \sum_{x,y,s} P(x)P(s)P(y|x, s)\log\frac{P(x|y,s)}{P^*(x)} \tag{10}$$

Proof:

The best input distribution $P^*(x)$ is found by solving the mazimization problem using a Lagrange multiplier to constraint $\sum_X P^*(x) = 1$.

$$I(x; y, s) = \sum_{x,y,s} P(y|x, s)P(s)P(x)\log\frac{P(x|y,s)}{P(x)} \tag{11}$$

$$J = \sum_{x,y,s} P(y|x, s)P(s)P(x)\log\frac{P(x|y,s)}{P(x)} - \lambda(\sum_x P(x) - 1) \tag{12}$$

$$\frac{\partial J}{\partial P(x)} = \sum_{y,s} P(y|x, s)P(s)\log P(x|y, s) - \log P(x) - 1 - \lambda = 0 \tag{13}$$

From above equation, we obtain,

$$P(x) = e^{\sum_{y,s} P(y|x,s)P(s)\log P(x|y,s) - (\lambda+1)} \tag{14}$$

With substituting (14) into the constraint, we can obtain,

$$P(x) = \frac{e^{\sum_{y,s} P(y|x,s)P(s)\log P(x|y,s)}}{\sum_x e^{\sum_{y,s} P(y|x,s)P(s)\log P(x|y,s)}} \tag{15}$$

And finally,

$$P(x) = \frac{\prod_{y,s} P(x|y, s)^{P(s)P(y|x,s)}}{\sum_x \prod_{y,s} P(x|y, s)^{P(s)P(y|x,s)}}$$

where,

$$P(x|y, s) = \frac{P(y|x,s)P(x)P(s)}{\sum_x P(y|x,s)P(x)P(s)} \tag{16}$$

Since $P(x|y, s) = \frac{P(y|x,s)P(s)P(x)}{P(y,s)}$, we can write as following,

$$\log P(x|y, s) = \log\frac{P(s)P(y|x,s)}{P(y,s)} + \log P(x) \tag{17}$$

Therefore, according to (15) and (17), the following closed form expression for P(x) is obtained,

$$P^{t+1}(x) = P^t(x)\frac{e^{\sum_{y,s} P(y|x,s)P(s)\log\frac{P(y|x,s)P(s)}{P(y,s)}}}{\sum_x P^t(x)e^{\sum_{y,s} P(y|x,s)P(s)\log\frac{P(y|x,s)P(s)}{P(y,s)}}} \tag{18}$$

where $P^t(x)$ is the input distribution after 't' iteration.

The concept of the algorithm is to regard P(x|y,s) and P(x) as independent variables and optimize alternately between these two. This algorithm goes through these steps,

Step1: Start the algorithm with a guess of the maximizing distribution P(x).

Step 2: Calculate P(x) by using (18).

Step 3: Checking whether convergence criterion is met, if yes then go to next step; otherwise go to step 2.

Step 4: Calculate the capacity by using (10).

The above algorithm converges monotonically to the capacity of the channel as the same convergence of Arimoto-Blahut algorithm [1], [2].

For accelerating the convergence, we can use a coefficient, $\eta \geq 1$, in (18) as following,

$$P^{t+1}(x) = P^t(x)\frac{e^{\eta\sum_{y,s} P(y|x,s)P(s)\log\frac{P(y|x,s)P(s)}{P(y,s)}}}{\sum_x P^t(x)e^{\eta\sum_{y,s} P(y|x,s)P(s)\log\frac{P(y|x,s)P(s)}{P(y,s)}}} \tag{19}$$

Example 1. Consider the channel with following transition probability distribution; this channel has two states, **s** = (0, 1) with distribution P(s)=[0.25 0.75], the inputs of this channel are 0 or 1.

$$P(y|x,s) = \begin{bmatrix} 0.55 & 0.45 \\ 0.65 & 0.35 \\ 0.45 & 0.55 \\ 0.90 & 0.10 \end{bmatrix}$$

where the columns represent the different elements of $\mathcal{Y} = \{0,1\}$ and the rows correspond to the natural ordering of the x, s. The initial distribution for input is P(x)=[0.35 0.65] and our convergence criterion is $\|P^{t+1}(x) - P^{t}(x)\| < 0.0001$.

Fig. 2 shows convergence of the above algorithm with different values of η to C=0.0528 bits per channel use and we have only 4 iterations for $\eta = 10$. The best input distribution that maximizes the capacity is P(x)=[0.4645 0.5355]. It can be seen from Fig. 2 that the algorithm with $\eta \geq 20$ diverges and can not converge to the true capacity.

If this channel is used without side information, the capacity of the channel will be about 0.0232 bits per channel use. In [9] the capacity for this channel with side information at transmitter is about 0.04 bits per channel use.

Example 2. Consider the channel with following transition probability distribution, in this example $|\mathcal{X}| = |\mathcal{Y}| = 3$. The initial distribution for input is P(x)=[0.3 0.6 0.1]. The other specifications are the same as example 1.

$$P(y|x,s) = \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.25 & 0.30 & 0.45 \\ 0.15 & 0.50 & 0.35 \\ 0.40 & 0.20 & 0.40 \\ 0.35 & 0.40 & 0.25 \\ 0.30 & 0.40 & 0.30 \end{bmatrix}$$

Fig. 3 shows convergence of the above algorithm with three different of η to C = 0.0370 bits per channel. The optimal distribution that maximizes the capacity is P(x)=[0.1438 0.4547 0.4015]. The algorithm with $\eta \geq 33$ diverges.

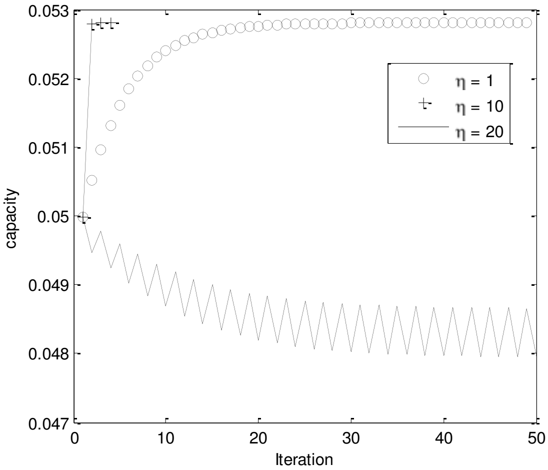

Figure 2. Optimized result for example 1

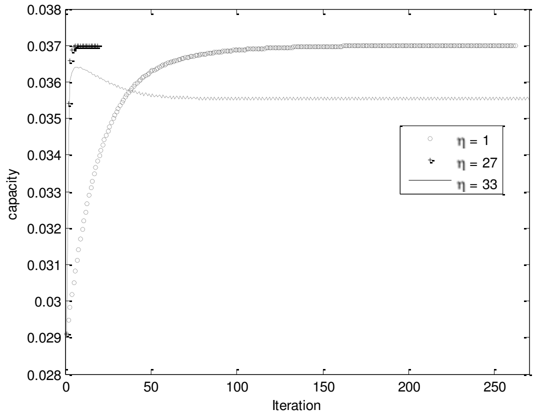

Figure 3. Optimized result for example 2

IV. COMPUTING THE SUM CAPACITY OF DISCRETE MULTIPLE ACCESS CHANNEL WITH CAUSAL SIDE INFORMATION AT THE RECEIVER

The capacity of discrete multiple access channel with causal side information at the receiver shown in Fig. 5, can be expressed as the following form [13],

$$C^* = \max_{P(x_1)p(x_2)\ldots P(x_m)} I(x_1, \ldots, x_m; y, s) \tag{20}$$

As the same channel with causal side information at the receiver we can obtain a lower band for the capacity of discrete multiple access channel with causal side information at the receiver as following,

$$I(y; x_1, \ldots, x_m, s) = I(y; x_1, \ldots, x_m) + I(y; s|x_1, \ldots, x_m) \tag{21}$$

$$= I(y; s) + I(y; x_1, \ldots, x_m|s) \tag{22}$$

$$I(x_1, \ldots, x_m; y|s) = I(y; x_1, \ldots, x_m) - I(y; s) + I(y; s|x_1, \ldots, x_m) \tag{23}$$

$$= I(y; x_1, \ldots, x_m) - I(y; s) + H(s|x_1, \ldots, x_m) - H(s|x_1, \ldots, x_m, y) \tag{24}$$

$$\geq I(y; x_1, \ldots, x_m) - I(y; s) + H(s) - H(s|y) \tag{25}$$

$$= I(y; x_1, \ldots, x_m) \tag{26}$$

Where (24) and (25) follow from the fact that $x_1, \ldots, x_m$ and s are independent and the removing conditioning.

Hence,

$$C^* \geq \max_{P(x_1)p(x_2)\ldots P(x_m)} I(x_1, \ldots, x_m; y) \tag{27}$$

where the above inequality shows that the capacity of discrete multiple access channel with side information at the receiver is greater than the ordinary discrete multiple access memoryless channel without side information.

Theorem 2. Let $P(y|x_1,\dots,x_m,s)$ be a given conditional distribution for a multiple access channel with side information at the receiver and the state distribution P(s). Then the distribution $P(x_1,\dots,x_m)$ that maximizes the capacity of this channel with side information at the receiver is,

$$P(x_1,\dots,x_m)=\frac{\prod_{y,s}P(x_1,\dots,x_m|y,s)^{P(s)P(y|x_1,\dots,x_m,s)}}{\sum_{x_1,\dots,x_m}\prod_{y,s}P(x_1,\dots,x_m|y,s)^{P(s)P(y|x,x_1,\dots,x_m)}} \quad (28)$$

where,

$$P(x_1,\dots,x_m|y,s)=\frac{P(y|x_1,\dots,x_m,s)P(x_1,\dots,x_m)P(s)}{\sum_{x_1,\dots,x_m}P(y|x_1,\dots,x_m,s)P(x_1,\dots,x_m)P(s)} \quad (29)$$

And the capacity of this channel can be written as the following expression,

$$C^*=\sum_{x_1,\dots,x_m,y,s}P(x_1,\dots,x_m)P(s)P(y|x_1,\dots,x_m,s)\log\frac{P(x_1,\dots,x_m|y,s)}{P(x_1,\dots,x_m)} \quad (30)$$

The closed form expression for P(x) is,

$$P^{t+1}(x_1,\dots,x_m)=$$

$$\frac{P^t(x_1,\dots,x_m)e^{\sum_{y,s}P(y|x_1,\dots,x_m,s)P(s)\log\frac{P(y|x_1,\dots,x_m,s)P(s)}{P(y,s)}}}{\sum_{x_1,\dots,x_m}P^t(x_1,\dots,x_m)e^{\sum_{y,s}P(y|x_1,\dots,x_m,s)P(s)\log\frac{P(y|x_1,\dots,x_m,s)P(s)}{P(y,s)}}} \quad (31)$$

The proof of this theorem is similar to theorem 1.

The algorithm for computing the total capacity of discrete multiple access memoryless channels with side information at receiver is similar to previous algorithm with some changes. This algorithm is as following,

Step1: Make a vector the elements of which are the initial input joint distribution, $P(x_1,\dots,x_m)$, as follows,

$$P(x_1,\dots,x_m)\Leftrightarrow\begin{bmatrix}P(x_1=i_{1,1},\dots,x_m=i_{m,1})\\ \cdot\\ \cdot\\ \cdot\\ P(x_1=i_{1,|\mathcal{X}_1|},\dots,x_m=i_{m,|\mathcal{X}_m|})\end{bmatrix} \quad (32)$$

Step 2: Calculate $P(x_1,\dots,x_m)$ by using (31).

Step 3: Checking whether convergence criterion is met, if yes then go to next step; otherwise go to step 2.

Step 4: After the convergence of the algorithm, we can change the optimal $P(x_1,\dots,x_m)$ to the joint distribution which is named $P^*(x_1,\dots,x_m)$. According to the rank of it, we have two situations:

- $Rank(P^*(x_1,\dots,x_m))=1$. At this case we can compute the sum capacity by following expression:

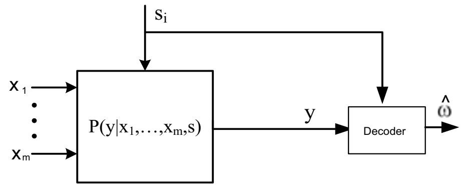

Figure 4. Discrete multiple access channel with causal side information at the receiver

$$C^*_{total}=\sum_{x_1,\dots,x_m,y,s}P^*(x_1,\dots,x_m)P(s)P(y|x_1,\dots,x_m,s)\log\frac{P(x_1,\dots,x_m|y,s)}{P^*(x_1,\dots,x_m)} \quad (33)$$

- $Rank(P^*(x_1,\dots,x_m))\neq 1$. This will be the case in general. We need to project $P^*(x_1,\dots,x_m)$ as a product distribution $(P_{product}(x_1,\dots,x_m))$ [14],[15]. Then we can compute the sum capacity with following expression;

$$C^*_{total}=\sum_{x_1,\dots,x_m,y,s}P_{product}(x_1,\dots,x_m)P(s)P(y|x_1,\dots,x_m,s)\log\frac{P(x_1,\dots,x_m|y,s)}{P_{product}(x_1,\dots,x_m)} \quad (34)$$

Example 3. Consider two-user binary discrete multiple access channel with two states, **s** = (0,1), with distribution P(s)=[0.25 0.75], the transition probability distribution is:

$$P(y|x_1,x_2,s)=\begin{pmatrix}0.25 & 0.75\\ 0.20 & 0.80\\ 0.30 & 0.70\\ 0.10 & 0.90\\ 0.40 & 0.60\\ 0.50 & 0.50\\ 0.65 & 0.35\\ 0.85 & 0.15\end{pmatrix}$$

where the columns represent the different elements of $\mathcal{Y}=\{0,1\}$ and the rows correspond to the natural ordering of the x_1, x_2 and s. This algorithm converges to $C^*_{total}=0.3668$ bits per channel use. If this channel is used without side information, the capacity of the channel will be about 0.3322 bits per channel use. In [9] the capacity for this channel with side information at transmitter is about 0.3324 bits per channel use.

V. Conclusion

We have given two algorithms for discrete memoryless channel and discrete multiple access channel with causal side information at the receiver. The computation results show that causal side information increases the capacity of discrete memoryless channels. The capacity of channels with causal side information at the receiver is greater than the capacity of ordinary discrete memoryless channel and channel with causal side information at the transmitter. These results hold for sum capacity of the discrete multiple access channels with side information at the receiver.

References

[1] S. Arimoto, "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channel," IEEE Trans. Inform. Theory, vol. IT-18,no. 1, pp. 14-20, Jan. 1972.

[2] R. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions," IEEE Trans. Inform. Theory, vol. IT-18, no. 4, pp. 460-472, Jul. 1972.

[3] I. Csiszar and G. Tusnady, “Information Geometry and Alternating Minimization Procedures,” Statistics and Decisions, Supplement Issue, pp. 205-237, 1984.

[4] M. Rezaeian and A. Grant, “A Generalization of the Arimoto-Blahut Algorithm,” ISIT 2004.

[5] Y. Watanabe, “The Total Capacity of Two-User Multiple-Access Channel with Binary Output”, IEEE Trans. Inform. Theory, vol. 42, pp. 1453-1465, Sept. 1996.

[6] Y. Watanabe and K. Kamoi, “The Total Capacity of Multiple-Access Channel,” ISIT 2002.

[7] M. Rezaeian and A. Grant, “Computation of Total Capacity for Discrete Memoryless Multiple-Access Channels”, IEEE Trans. Inform. Theory, vol.50, no. 11, pp. 2779-2784, Nov. 2004.

[8] F. Dupuis, W. Yu, and F. M. J. Willems, “Blahut-Arimoto Algorithms for Computing Channel Capacity and Rate-Distortion with Side Information,” ISIT 2004.

[9] A. Sahebalam, M. Osmani Bojd and G.A. Hodtani, “Numerical Computation of Sum Capacity for Discrete Multiple Access Channel with Side Information at the Transmitter,” 48th Annual Allerton Conference on communication, control and computing, Sept. 29 – Oct. 1, 2010.

[10] M. Salehi, “Capacity and coding for memories with real-time noisy defect information at encoder and decoder," IEE Proceedings -1, vol. 139, No. 2. pp.113-117. Apr. 1992.

[11] G. Keshet, Y. Steinberg, and N. Merhav, Channel Coding in the Presence of Side Information: Subject Review, Now Publisher, pp. 9-15.

[12] U. Erez and R. Zamir, “Capacity and Coding for Symmetric Channels with Side Information at the Transmitter,” Tech. Report EE-S-97-10, pp. 1-5, Oct. 1997.

[13] S. Ali Jafar, “Channel Capacity with Side Information -A Unified View,” 43th Annual Allerton Conference on Communication, Control, and Computing, Sept. 28-30, 2005.

[14] C. K. Chow and C. N. Liu, “Approximating discrete probability distributions with dependence trees,” IEEE Trans. Inform. Theory, vol. IT-14, no. 3, pp. 462-467, May 1968.

[15] H. Ku and S. Kullback, “Approximating discrete probability distributions,” IEEE Trans. Inform. Theory, vol. IT-15, no.4, pp. 444-447, July 1969.