# Blahut-Arimoto Algorithms for Computing Channel Capacity and Rate-Distortion With Side Information


Frédéric Dupuis
Dept. of Electrical and Computer Eng.
University of Toronto
Toronto, Ontario, Canada M5S 3G4
e-mail: dupontd@ecf.utoronto.ca

Wei Yu
Dept. of Electrical and Computer Eng.
University of Toronto
Toronto, Ontario, Canada M5S 3G4
e-mail: weiyu@comm.utoronto.ca

Frans M. J. Willems
Faculteit Electrotechniek
Eindhoven University of Technology
Eindhoven, The Netherlands
e-mail: F.M.J.Willems@tue.nl



*Abstract* — **This paper presents numerical algorithms for the computation of the capacity for channels with non-causal transmitter side information (the Gel'fand-Pinsker problem) and the rate-distortion function for source coding with decoder side information (the Wyner-Ziv problem). The algorithms are based on the reformulation of the mutual information expressions in terms of Shannon strategies.**


## I. Introduction

Consider the channel with non-causal side information described by $p(y|x,s)$ where $x \in \mathcal{X}$ is the input, $y \in \mathcal{Y}$ is the output and $s \in \mathcal{S}$ are the possible states of the channel. The sequence of states is i.i.d. according to $p(s)$ and the entire sequence is known at the transmitter but not at the receiver. Gel'fand and Pinsker [1] showed that the capacity of this channel is

$$C = \max_{p(u|s),p(x|u,s)} I(U;Y) - I(U;S) \tag{1}$$

where $U$ is an auxiliary random variable.

This paper gives a Blahut-Arimoto algorithm for the numerical computation of the capacity for the Gel'fand-Pinsker channel. Note that the mutual information expression above is concave in $p(u|s)$ and convex in $p(x|u,s)$. Thus, standard convex optimization techniques do not apply. However, because of the convexity of $I(U;Y) - I(U;S)$ in $p(x|u,s)$, the optimal $p(x|u,s)$ must take values of 0 or 1 only. Thus, the input $x$ must be a deterministic function $x = f(u,s)$. The main idea of this paper is that such a $p(x|u,s)$ can be eliminated from the problem altogether by expanding the input alphabet to be the set of all strategies $t : \mathcal{S} \to \mathcal{X}$. This is reminiscent of Shannon's strategy for channels with causal side information [2].

## II. Channel Capacity Computation

The convexity argument gives rise to the following alternative characterization of capacity for the Gel'fand-Pinsker channel:

**Theorem 1** *The capacity of a channel with non-causal side information at the transmitter is given by* $C = \max_{p(t|s)} I(T;Y) - I(T;S)$*, where* $p(t|s)$ *is a probability distribution over the set of all possible functions* $t : \mathcal{S} \to \mathcal{X}$ *and the input symbol* $X$ *is selected using* $x = t(s)$*.*

The characterization of channel capacity in terms of strategies $t : \mathcal{S} \to \mathcal{X}$ allows the computation of capacity via an Arimoto-Blahut algorithm. First, the mutual information is re-written as:

$$I(T;Y) - I(T;S) = \sum_{s,y,t} p(s)q(t|s)p(y|t,s) \log \frac{Q(t|y)}{q(t|s)} \tag{2}$$


[1] This work was supported by NSERC, Bell Canada and CITO.


where $p(y|t,s) = p(y|x = t(s), s)$ is fixed by the channel transition probability. The idea of the algorithm is to regard $q(t|s)$ and $Q(t|y)$ as independent variables and to optimize alternately between the two. It can readily be seen that since the objective function is concave in both $Q(t|y)$ and $q(t|s)$, the alternating optimization scheme must converge to the global maximum from any initial distribution. Following a standard derivation of Arimoto-Blahut algorithm, the optimal $q^*(t|s)$ and $Q^*(t|y)$ in each iteration can be obtained as follows:

$$q^*(t|s) = \frac{\prod_y Q(t|y)^{p(y|t,s)}}{\sum_{t'} \prod_y Q(t'|y)^{p(y|t',s)}} \tag{3}$$

$$Q^*(t|y) = \frac{\sum_s p(s)q(t|s)p(y|t,s)}{\sum_{s,t'} p(s)q(t'|s)p(y|t',s)} \tag{4}$$

Since the optimal $Q^*(t|y)$ is just the conditional marginal distribution over the joint distribution $p(s)q^*(t|s)p(y|t,s)$, the algorithm reaches the desired optimal solution after it converges. Note that the algorithm expands the input alphabet from $\mathcal{X}$ to a set of cardinality $|\mathcal{X}|^{|\mathcal{S}|}$. Thus, the memory requirement for the proposed algorithm is exponential in the number of channel states.

In terms of complexity, the proposed algorithm improves upon a previous algorithm proposed by El Gamal and Heegard [3] for the computation of capacity for the Gel'fand-Pinsker channel. Instead of relying on the set of all Shannon strategies, the algorithm in [3] included an additional step to optimize over $p(x|u,s)$ as well as methods to ensure convergence which, in this case, is harder to achieve. The result is an algorithm that is significantly slower but avoids the exponential use of memory. The slowdown is due to the fact that the algorithm in [3] has to potentially search over all possible functions $\mathcal{U} \times \mathcal{S} \to \mathcal{X}$, whereas the algorithm proposed in this paper only searches over functions $\mathcal{S} \to \mathcal{X}$.

Shannon strategy can also be applied to the computation of the Wyner-Ziv rate-distortion function for the source coding problem with decoder side information. This is described in [4].

## References


[1] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[2] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Dev.*, 2:289–293, 1958.

[3] A. El Gamal C. Heegard. On the capacity of computer memories with defects. *IEEE Trans. Inform. Theory*, IT-29:731–739, 1983.

[4] F. M. J. Willems. Computation of the Wyner-Ziv rate-distortion function Katholieke Universiteit Leuven, Departement Wiskunde, October 1982. Also Eindhoven University of Technology Research Reports, July 1983.