ISIT 2003, Yokohama, Japan, June 29 – July 4, 2003

# A Generalized Blahut-Arimoto Algorithm

Pascal O. Vontobel[1]
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
e-mail: `vontobel@ifp.uiuc.edu`

*Abstract* — **Kavčić proposed in [1] an algorithm that apparently finds the mutual-information-rate-maximizing parameters of a Markov source at the input to an indecomposable finite-state channel. In this paper we prove that the stationary points of this algorithm indeed correspond one-to-one to the critical points of the information-rate curve. Kavčić's algorithm can be considered as a generalized Blahut-Arimoto algorithm, as it includes as special cases the classical Blahut-Arimoto algorithm for discrete memoryless channels (DMCs) and the solution to finding the capacity-achieving input distribution for finite-state channels with no noise [2].**

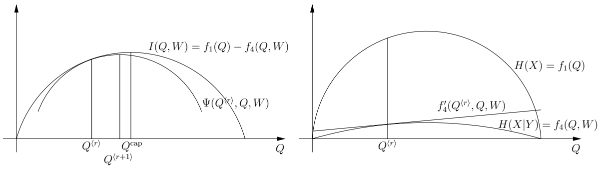
Recently, Arnold/Loeliger, Sharma/Singh, and Pfister et al. independently proposed a method for computing information rates of indecomposable finite-state channels [3] whose input is a Markov source. Subsequently, Kavčić [1] proposed an algorithm that apparently optimizes the parameters of such an input Markov source in order to maximize the information rate.

In the following, we first discuss the classical Blahut-Arimoto algorithm (BAA); it will be cast in a form in which also the generalized BAA can be stated. The main facts of the generalized BAA will then be given.

The capacity of a DMC with channel law $W(y|x)$ is defined to be $C(W) \triangleq \max_Q I(Q,W)$ where $I(Q,W) = I(X;Y)$ and where a maximizing channel input pmf $Q$ is called a capacity-achieving input pmf. The classical BAA [4, 5] iteratively solves the problem of finding numerically and in an efficent way a capacity-achieving input distribution of a DMC and therefore also solves the problem of finding numerically the capacity of a DMC. Blahut and Arimoto independently showed that the above maximization problem can be stated as a double-maximization problem; the BAA essentially consists of alternating between these two maximizations and it can be shown that this approach indeed leeds to the desired solution.

Alternatively, the BAA can be described as follows. Assume that at iteration $r$ we found some input pmf $Q^{\langle r \rangle}$ with corresponding information rate $I(Q^{\langle r \rangle}, W)$. At iteration $r+1$ we would like to find an input pmf $Q^{\langle r \rangle}$ such that $I(Q^{\langle r+1 \rangle}, W) \geq I(Q^{\langle r \rangle}, W)$ (see lhs of figure). To this end we introduce an auxiliary function $\Psi(Q^{\langle r \rangle}, Q, W)$ which locally (i.e. around $Q = Q^{\langle r \rangle}$) approximates $I(Q,W)$ (see lhs of figure). We require (1) that the auxiliary function assumes the same value at $Q = Q^{\langle r \rangle}$ as $I(Q,W)$ does, i.e. $\Psi(Q^{\langle r \rangle}, Q^{\langle r \rangle}, W) = I(Q^{\langle r \rangle}, W)$, and (2) that $\Psi(Q^{\langle r \rangle}, Q, W)$ is never above $I(Q,W)$. If we find such an auxiliary function that can easily be maximized (let us call the pmf where the maximum is achieved $Q^{\langle r+1 \rangle}$), we get easily a new input pmf $Q^{\langle r+1 \rangle}$ with $I(Q^{\langle r+1 \rangle}, W) \geq I(Q^{\langle r \rangle}, W)$ (see lhs of figure). Let $I(Q,W) = H(X) - H(X|Y) \triangleq f_1(Q) - f_4(Q,W)$. Then it can be shown that the auxiliary function $\Psi(Q^{\langle r \rangle}, Q, W) \triangleq f_1(Q) - f_4'(Q^{\langle r \rangle}, Q, W)$, where $f_4'(Q^{\langle r \rangle}, Q, W)$ is the linear approximation of $f_4(Q,W)$ at $Q = Q^{\langle r \rangle}$, fulfills the above requirements and leads to the classical BAA (see rhs of figure).

We now turn to the generalized BAA. With an appropriate parametrization $Q$ of a Markov source [6], the above algorithm can be generalized to find a Markov source of a given memory size that maximizes the information rate of an indecomposable finite-state channel. Again, the function $f_4'(Q^{\langle r \rangle}, Q, W)$ is a linear approximation of $f_4(Q,W) = H(X|Y)$ at $Q = Q^{\langle r \rangle}$; most importantly, the new $Q^{\langle r+1 \rangle}$ can again be found in a computationally efficient manner. From the properties of $f_4'(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ it follows that the stationary points of the generalized BAA correspond one-to-one to critical points of the information-rate curve [6]; but note that only maxima of the information-rate curve are stable stationary points of the generalized BAA.




It remains an open problem if $f_4(Q,W)$ is a concave function in $Q$ (which is the case for the DMC) from which it would follow that at each step $r$ the difference $I(Q^{\langle r+1 \rangle}, W) - I(Q^{\langle r \rangle}, W)$ is never negative. Showing additionally the concavity of $I(Q,W)$ (which holds in the case of the DMC) would lead to a statement about global convergence of the algorithm.

All the details of this paper can be found in the report [6].

## References

[1] A. Kavčić, "On the capacity of Markov sources over noisy channels," in *Proc. IEEE GLOBECOM*, (San Antonio, TX, USA), pp. 2997–3001, Nov. 2001.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July/Oct. 1948.

[3] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[4] R. E. Blahut, "Computation of channel capacity and rate distortion functions," *IEEE Trans. on Inform. Theory*, vol. IT–18, no. 4, pp. 460–473, 1972.

[5] S. Arimoto, "An algorithm for computing the capacity of arbitrary memoryless channels," *IEEE Trans. on Inform. Theory*, vol. IT–18, no. 1, pp. 14–20, 1972.

[6] P. O. Vontobel, *A Generalized Blahut-Arimoto Algorithm.* Laboratory for Signal and Information Processing, Internal Report INT200203, 2002. Available online under `http://www.isi.ee.ethz.ch/publications`.

[1] This work was partly supported by NSF Grant CCR 99-84515. A large part of this work was done while being at ETH Zurich, Switzerland.