THE ARIMOTO-BLAHUT ALGORITHM
FOR COMPUTATION OF CHANNEL CAPACITY

William A. Pearlman

2002

References: S. Arimoto - IEEE Trans. Inform. Thy., Jan. 1972

R. Blahut - IEEE Trans. Inform. Thy., July 1972

Recall the definition of capacity for a discrete, memoryless channel (DMC). Given channel transition probabilities $P(j/i),\ i = 1, 2, ..., m,\ j = 1, 2, ..., n$ and let $Q = (Q(1), Q(2), ..., Q(m))$ be the input probability vector, the capacity is

$$C = \max_Q \sum_{i=1}^{m} \sum_{j=1}^{n} P(j/i)Q(i) \log \frac{P(i/j)}{Q(i)} \tag{1}$$

where $P(i/j) = \frac{P(j/i)Q(i)}{\sum_k P(j/k)Q(k)}$.
The quantity $\sum_i \sum_j P(j/i)Q(i) \log \frac{P(i/j)}{Q(i)}$ is the average mutual information $I(X;Y)$ between the input and output ensembles of the channel. For a fixed channel, i.e., fixed $P(j/i)$, it is a convex function of the input probabilities Q. To emphasize the functional relationship on Q, let us call it $J(Q)$:

$$J(Q) \equiv \sum_i \sum_j P(j/i)Q(i) \log \frac{P(i/j)}{Q(i)} = I(X;Y)$$

and

$$C = \max_Q J(Q) \tag{2}$$

Now let $\Phi(i/j)$ be any set of conditional probabilities of input given output ($i = 1, 2, ..., m$; $j = 1, 2, ..., n$). Define now a generalization of $J(Q)$ by

$$J(Q, \Phi) = \sum_i \sum_j P(j/i) Q(i) \log \frac{\Phi(i/j)}{Q(i)} \tag{3}$$

The function $J(Q, \Phi)$ is a function of the input probabilities Q and conditional probabilities Φ. Through this function and a series of lemmas, we shall formulate an algorithm for calculating the channel capacity C. For the sake of simplicity, assume the logarithmic base is always e.

The lemmas will first be stated without proof, so as not to clutter the development. They will be proved as a final order of business.

Lemma 1

For any fixed Q,

$$J(Q, \Phi) \leq J(Q)$$

with equality iff

$$\Phi(i/j) = \frac{P(j/i)Q(i)}{\sum_k P(j/k)Q(k)} = P(i/j)$$

By this lemma, $J(Q) = \max_\Phi J(Q, \Phi)$. Substitution into (2) gives

$$C = \max_Q \max_\Phi J(Q, \Phi) \tag{4}$$

Lemma 2

For fixed $\Phi(i/j)$, consider Q variable

$$J(Q,\Phi) \leq \log(\sum_i r(i))$$
$$r(i) = \exp[\sum_j P(j/i) \log \Phi(i/j)]$$

with equality iff $Q(i) = \frac{r(i)}{\sum_k r(k)}$.

The double maximum in (4) can be taken in any order. Therefore,

$$C = \max_{\Phi} \max_{Q} J(Q,\Phi) = \max_{\Phi} \log(\sum_i r(i))$$
$$= \max_{\Phi} \log(\sum_i \exp \sum_j P(j/i) \log \Phi(i/j)) \quad (5)$$

The point of introducing $J(Q,\Phi)$ is that C can be calculated through a double maximization procedure (40¡ each step of which can be solved in closed form, as given in Lemmas 1 and 2. The maximization of $J(Q)$ with respect to Q has no closed-form solution. We can doubly maximize $J(Q,\Phi)$ in any order. For example, let us fix Q to some initial value Q^o. By Lemma 1, $\Phi^o(i/j) = \frac{P(j/i)Q^o(i)}{\sum_k P(j/k)Q^o(k)}$ maximizes $J(Q^o,\Phi)$ when $\Phi = \Phi^o$. Now consider $J(Q,\Phi^o)$ for Φ^o fixed and vary Q to produce a maximum. In Lemma 2, that maximum is

$$J(Q,\Phi^o) = \log(\sum_i r(i)) = \log \sum_i \exp \sum_j P(j/i) \log \Phi(i/j)$$

for $Q(i) = Q^1(i) = \frac{r(i)}{\sum_k r(k)}$. Then for that Q^1 we can find a maximum of $J(Q^1,\Phi)$ for $\Phi = \Phi^1$ as given in Lemma 1 for Q^1 and $P(j/i)$. Fix Φ^1 and maximize $J(Q,\Phi^1)$ for $Q = Q^2$ as given in Lemma 2. Now maximize $J(Q^2,\Phi)$ and so on. At each step, $J(Q,\Phi)$ is increasing until it eventually reaches capacity C.

So, we can now formulate the Arimoto-Blahut algorithm to accomplish the double maximization in (4).

Algorithm

ℓ is the iteration index

1. Set $\ell = 0$ and choose initial set of input probabilities $Q^o(i) > 0$, all i.
2. Compute

$$\Phi^\ell(i/j) = \frac{Q^\ell(i)P(j/i)}{\sum_k Q^\ell(k)P(j/k)} \quad \text{all } i, j$$
$$r^\ell(i) = \exp \sum_j P(j/i) \ln \Phi^\ell(i/j)$$
$$J(Q^{\ell+1}, \Phi^\ell) = \ln(\sum_i r^\ell(i))$$
$$Q^{\ell+1}(i) = \frac{r^\ell(i)}{\sum_k r^\ell(i)}$$

3. Set $\ell = \ell + 1$ and go to 2.

This algorithm accomplishes what was already explained, but not how to stop the recursion. The question now is when do we know that we are close to capacity. We now have to introduce one more lemma.

For any ℓ, $J(Q^{\ell+1}, \Phi^\ell) = \ln(\sum_i r^\ell(i)) \leq C$. Let us define

$$c^\ell(i) \equiv \frac{r^\ell(i)}{Q^\ell(i)}$$

Then

$$J(Q^{\ell+1}, \Phi^\ell) = \ln(\sum_i Q^\ell(i)c^\ell(i))$$

So $J(Q^{\ell+1}, \Phi^\ell)$ is the logarithm of the average of the $c^\ell(i)$'s.

Lemma 3

$$C \leq \max_i \ln c^\ell(i)$$

Therefore, channel capacity C is bounded from below and above as follows:

$$\ln(\sum_i Q^\ell(i)c^\ell(i)) \le C \le \max_i \ln c^\ell(i)$$

In order to find C within accuracy ϵ, stop the iteration when

$$\max_i \ln c^\ell(i) - \ln(\sum_i Q^\ell(i)c^\ell(i)) < \epsilon$$

or

$$J(Q^{\ell+1}, \Phi^\ell) > \max_i \ln c^\ell(i) - \epsilon$$

Insert in algorithm:

2′. Calculate $-J(Q^{\ell+1}, \Phi^\ell) + \max_i \ln \frac{r^\ell(i)}{Q^\ell(i)} = T$.

If $T > \epsilon$, go to 3.

If $T < \epsilon$, go to 4.

4. $C = J(Q^{\ell+1}, \Phi^\ell)$

It is interesting to note that $\ln c^\ell(i) = I(i; Y/Q^\ell)$; the average information that the output ensemble Y gives about the input event i, when the input probabilities are Q^ℓ. Recall that if

$$I(i; Y/Q) = \gamma \text{ for all } i \text{ s.t. } Q(i) > 0$$

$$I(i; Y/Q) \le \gamma \text{ for all } i \text{ s.t. } Q(i) = 0 \qquad \text{(Kuhn-Tucker conditions).}$$

$J(Q) = \overline{I(i; Y/Q)} = C$ and $C = \gamma$. So, when we are nearing capacity C, the information given by the output ensemble Y about each input event of nonzero probability is nearing equality, since

$$\sum_i Q^\ell(i) \ln c^\ell(i) \le \ln \sum_i Q^\ell(i) c^\ell(i) \le C \le \max_i \ln c^\ell(i)$$

by the convex $\cap$ property of ln. So we have

$$\overline{I(i; Y/Q^\ell)} \le C \le \max_i I(i; Y/Q^\ell)$$

We now prove the lemmas.

Proof of Lemma 1:

$$\begin{aligned} J(Q,\Phi) - J(Q) &= \sum_i \sum_j P(j/i) Q(i) \log \frac{\Phi(i/j)}{P(i/j)} \\ &\le \sum_i \sum_j P(j/i) Q(i) \left[\frac{\Phi(i/j)}{P(i/j)} - 1 \right] \\ &= \sum_i \sum_j [P(j)\Phi(i/j) - P(j/i)Q(i)] = 1 - 1 = 0 \end{aligned}$$

with equality iff $\Phi(i/j) = P(i/j)$ for all i and j.

Proof of Lemma 2:

$$
\begin{aligned}
J(Q,\Phi) &= \sum_i \sum_j P(j/i)Q(i)\log\frac{\Phi(i/j)}{Q(i)} \\
&= \sum_i \sum_j P(j/i)Q(i)\log\frac{1}{Q(i)} + \sum_i \sum_j P(j/i)Q(i)\log\Phi(i/j) \\
&= \sum_i Q(i)\log\frac{1}{Q(i)} + \sum_i Q(i)\sum_j P(j/i)\log\Phi(i/j) \\
J(Q,\Phi) &= \sum_i Q(i)\log\left(\frac{\exp\sum_j P(j/i)\log\Phi(i/j)}{Q(i)}\right) \\
&= \sum_i Q(i)\log\left(\frac{r(i)}{Q(i)}\right), \qquad r(i) = \exp\sum_j P(j/i)\log\Phi(i/j) \\
&= \sum_i Q(i)\log(\sum_k r(k)) + \sum_i Q(i)\log\frac{r(i)/\sum_k r(k)}{Q(i)} \\
&\leq \log(\sum_k r(k))
\end{aligned}
$$

The last step results from $\ln x \leq x-1$ in the second sum with equality iff $Q(i) = \frac{r(i)}{\sum_k r(k)}$.

Proof of Lemma 3.

Suppose Q^* achieves capacity. Then

$$
\begin{aligned}
C &= \sum_i \sum_j P(j/i)Q^*(i)\log\frac{P(j/i)}{P^*(j)}, \;\; P^*(j) = \sum_k P(j/k)Q^*(k) \\
C &= \sum_i \sum_j P(j/i)Q^*(i)\log\frac{P(j/i)}{P^\ell(j)} \times \frac{P^\ell(j)}{P^*(j)}
\end{aligned}
$$

where

$$
P^\ell(j) = \sum_k P(j/k)Q^\ell(k) \;\; \text{are the output probabilities for } Q^\ell(k).
$$

Continuing,

$$C = \sum_i \sum_j P(j/i) Q^*(i) \log \frac{P^\ell(j)}{P^*(j)} + \sum_i \sum_j P(j/i) Q^*(i) \log \frac{P(j/i)}{P^\ell(j)}$$
$$= \sum_j P^*(j) \log \frac{P^\ell(j)}{P^*(j)} + \sum_i Q^*(i) \sum_j P(j/i) \log \frac{P(j/i)}{P^\ell(j)}$$

Using $\ln x \le x - 1$, the first term is ≤ 0. In the second term, we can overbound the $\sum_j P(j/i) \log \frac{P(j/i)}{P^\ell(j)}$ by its maximum over i. Therefore,

$$C \le \max_i \sum_j P(j/i) \log \frac{P(j/i)}{p^\ell(j)} = \max_i I(i; Y/Q^\ell)$$

Recall

$$r^\ell(i) = \exp \sum_j P(j/i) \ln \Phi^\ell(i/j) = \exp \sum_j P(j/i) \ln \frac{Q^\ell(i) P(j/i)}{P^\ell(j)}$$
$$\ln c^\ell(i) = \ln r^\ell(i) - \ln Q^\ell(i) = \sum_j P(j/i) \ln \frac{Q^\ell(i) P(j/i)}{P^\ell(j)} - \sum_j P(j/i) \ln Q^\ell(i)$$
$$= \sum_j P(j/i) \ln \frac{P(j/i)}{P^\ell(j)} = I(i; Y/Q^\ell).$$

Therefore, the conclusion of the lemma can be expressed as

$$C \le \max_i \ln c^\ell(i), \quad \text{as given.}$$

Note the condition for equality. We must have $P^\ell(j) = P^*(j)$ or $Q^\ell(i) = Q^*(i)$ for nonzero $Q^*(i)$ <u>and</u> $I(i; Y/Q^\ell)$ be the same maximum value for all i such that $Q^\ell(i) > 0$. These conditions are both necessary and sufficient for equality. So we have corroborated the Kuhn-Tucker conditions.