The Blahut-Arimoto Algorithm for the Calculation of the Capacity of a Discrete Memoryless Channel

Lawrence Ip
lip@eecs.berkeley.edu

December 10 1999

Abstract

The capacity of a discrete memoryless channel was shown by Shannon to be given by the value of a convex optimization problem. Blahut and Arimoto developed a simple and efficient algorithm to solve this convex optimization problem numerically by exploiting structure in the problem. We implement the Blahut-Arimoto algorithm using MATLAB and thus demonstrate how an algorithm exploiting the structure of a particular problem may be much simpler to implement than a general optimization algorithm like the ellipsoidal algorithm.

1 Introduction

One of the central concepts in information theory is that of channel capacity. Loosely speaking a channel takes an input and produces an output which has some probability distribution induced by the input. If A wants to communicate information to B, A transmits an input sequence. B observes the output sequence which has some distribution induced by A's input sequence. B then attempts to determine what input sequence A transmitted.

If A can transmit any possible input sequence it may happen that more than one input sequence corresponds to the observed output sequence. When that happens, it is impossible to recover the original sequence. So to ensure that B can recover A's message, A restricts his input sequences to a subset of the possible input sequences. Provided the input sequences are chosen "far enough apart", with very high probability, there will only be one highly probable input sequence corresponding to the output sequence observed by B.

By choosing a subset of the input sequences in this way, A reduces the rate at which he is transmitting. The supremum of the rates at which he can transmit with an arbitrarily low error probability (by taking the length of the input sequence long enough) is the capacity of the channel. This process of only transmitting a subset of the possible input sequences is known as *coding.*

We shall be considering discrete memoryless channels. Formally, a *discrete channel* is a system consisting of an input set $\{1, \ldots, n\}$, an output set $\{1, \ldots, m\}$ and a probability transition matrix $P = \{p(i|j)\}$ that gives the probability of observing output i given that the input j was transmitted. (Note that since P is a probability matrix, $\sum_{i=1}^{m} p(i|j) = 1$.) If the probability distribution of the output conditioned on the input at the current time is independent of previous channel inputs or outputs, the channel is said to be *memoryless.*

For discrete memoryless channels, Shannon showed that the capacity C of the channel could be expressed as the optimum value of a convex optimization problem with objective

$$C = \max_{p_1, \ldots, p_n} I(X; Y)$$

where X, Y are random variables representing the input and output respectively [5]. The optimization is taken over all input probability distributions $p = (p_1, \ldots, p_n)$. Thus we also have the constraints

$$\begin{aligned} p_j &\geq 0, \qquad j = 1, \ldots, n \\ \sum_{j=1}^{n} p_j &= 1. \end{aligned}$$

The mutual information $I(X; Y)$ is defined as

$$I(X; Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} p(i|j) p_j \log \frac{p(i|j)}{\sum_{k=1}^{n} p(i|k) p_k}.$$

It can be shown that $I(X; Y)$ is concave in p and thus the problem is a convex optimization problem.

The optimal p gives the distribution on the input symbols required to achieve capacity. Shannon's proof uses the probabilistic method to show the existence of a coding scheme with input distribution p. The proof is non-constructive but enables us to compute an upper bound on the performance of actual codes.

For simple channels it is possible to determine the capacity analytically but once the number of size of the input and output alphabets (n and m) becomes large it quickly becomes intractable. Thus it is important to be able to determine the capacity of the channel numerically. In 1972, Arimoto and Blahut independently discovered an iterative algorithm for computing the capacity of a discrete memoryless channel [1] [2].

Their algorithm exploits the special structure of the problem and converges monotonically towards the channel capacity. Arimoto showed that if p is non-unique the error can be bounded by $\log(n)/t$, where t is the number of iterations. Note that this bound depends only on the size of the problem and not on the characteristics of the actual channel. If p is unique he showed that the error decreases exponentially with t.

2 Description of the Algorithm

We reformulate the optimization problem by introducing additional variables $\phi(j|i)$ and changing our objective to

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

For fixed p,

$$\max_{\phi} J(p,P,\phi) = \sum_{i=1}^{m}\sum_{j=1}^{n} p(i|j)p_j \log\frac{p(i|j)}{\sum_{k=1}^{n} p(i|k)p_k},$$

attained at

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

Thus

$$C = \max_{p}\max_{\phi} J(p,P,\phi).$$

For fixed ϕ,

$$\max_p J(p, P, \phi)$$

is maximized when

$$p_j = \frac{\exp(\sum_{k=1}^m p(k|j) \log \phi(j|k))}{\sum_{l=1}^n \exp(\sum_{k=1}^m p(k|l) \log \phi(j|l))}.$$

The Blahut-Arimoto algorithm consists of alternately finding the optimal ϕ for a given p and then the optimal p for a given ϕ. The algorithm consists of the following steps

1. Choose an initial probability vector p^1. Then iterate the following steps with $t = 1, 2, \ldots$.

2. Maximize $J(p^t, P, \phi)$ with respect to ϕ. The maximizing ϕ is

$$\phi_j^t = \frac{p(i|j) p_j^t}{\sum_{k=1}^n p(i|k) p_k^t}, \qquad j = 1, \ldots, n.$$

3. Maximize $J(p, P, \phi^t)$ with respect p. The maximizing p is given by

$$p_j^{t+1} = \frac{r_j^t}{\sum_{k=1}^n r_k^t}, \qquad j = 1, \ldots, n,$$

where

$$r_j^t = \exp\left(\sum_{i=1}^m p(i|j) \log \phi^t(j|i)\right), \qquad j = 1, \ldots, n.$$

3 Implementation

```
%Note that we need to perturb the zero values in the
%probability transition matrix P by epsilon because
%MATLAB doesn't know that 0 log 0 should be treated as 0.

%Various example channels

%binary erasure channel p.187 Cover & Thomas
```

```
%optimal p=(0.5, 0.5) C=1-0.2=0.8
%P=[0.8-eps  0+eps
%   0.2      0.2
%   0+eps    0.8-eps];

%Z channel
%Ex 8.9 p.221 Cover & Thomas
%optimal p=(0.6, 0.4), C=0.322
%P=[1-eps 0.5
%   0+eps 0.5];

%Spring 99 EE229 Homework 9.3
%optimal p=(0.5, 0, 0.5), C=0.66667
%P=[2/3+eps 1/3 0+eps
%   1/3     1/3 1/3
%   0+eps   1/3 2/3-eps];

%Arimoto 1972 p.19 (68)
%optimal p=(0.501735, 0, 0.498265), C=0.161631
P=[0.6 0.7 0.5
   0.3 0.1 0.05
   0.1 0.2 0.45];


[m,n]=size(P);
p=ones(n,1)/n;
for t=1:200
   A=P*p*ones(1,n);
   B=ones(m,1)*p';
   phi=(P.*B)./A;
   r=exp(diag(log(phi')*P));
   s=sum(r);
   p=r/s;
end
p
C=sum(sum(P.*(ones(m,1)*p').*log2(P./(P*p*ones(1,n)))))
```

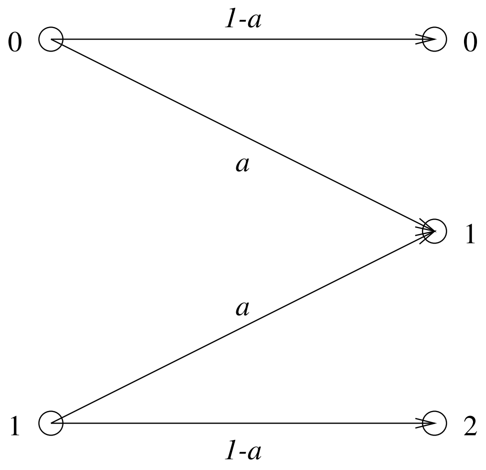

Figure 1: Binary erasure channel.

4 Various Examples

The implementation was tested out on the following channels.

In the binary erasure channel, a fraction a of the bits are erased [3]. See Figure 1. The probability transition matrix is

$$P = \begin{pmatrix} 1-a & 0 \\ a & a \\ 0 & 1-a \end{pmatrix}.$$

The receiver knows which bits have been erased. It can be shown that the capacity of this channel is $1-a$. This has an intuitive interpretation, since a proportion a of the bits have been lost, we can recover at most a proportion $1-a$ of the bits. If we allow the sender to obtain feedback from the receiver, then the sender could resend any lost bits and it is clear that a rate of $1-a$ is achievable. In fact, this rate is achievable *without* feedback. This result is completely general—the capacity of a discrete memoryless channel is not increased by feedback.

In the Z channel, a transmitted 0 always gets through, but a transmitted 1 has a 1/2 chance of being corrupted. See Figure 2. The probability transition matrix is

$$P = \begin{pmatrix} 1 & 0.5 \\ 0 & 0.5 \end{pmatrix}.$$

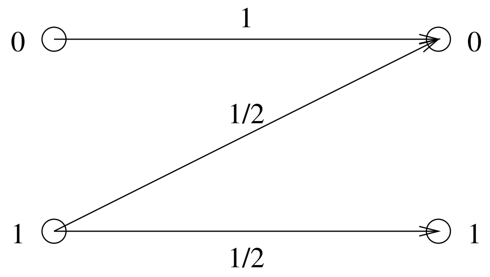

Figure 2: Z channel.

Thus you would expect to transmit more 0s than 1s. It can be shown that the optimal input distribution is $p = (0.6, 0.4)$ and the capacity is 0.322.

The other two channels we tested the algorithm on had probability transition matrices

$$P = \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

with optimal $p = (1/2, 0, 1/2)$, capacity 2/3, and

$$P = \begin{pmatrix} 0.6 & 0.7 & 0.5 \\ 0.3 & 0.1 & 0.05 \\ 0.1 & 0.2 & 0.45 \end{pmatrix}$$

with optimal $p = (0.501735, 0, 0.498265)$ and capacity 0.161631. The last channel was one given in Arimoto's paper.

For all four channels the algorithm converged to within 0.0001 of the actual capacity very quickly. Within 10-20 iterations for the first three channels and within 200 iterations for the last channel.

The algorithm is very cheap to implement, as it only involves a few matrix multiplications at each iteration. Thus the complexity is $O(mn^2t)$.

5 Further Work

Further work might involve considering channels where there are additional constraints. For example the input may be power limited.

Arimoto's convergence results suggest that often the convergence should be exponential. However, he was unable to give a priori bounds (other than the weak $\log(n)/\epsilon$ bound) on how many iterations would be required to obtain a given accuracy. Jürgensen gave a termination condition that may reduce the number of iterations required to guarantee that the value is within a given ϵ of the optimal value [4]. However he was unable to show that this termination condition was always better than the $\log(n)/\epsilon$ bound given by Arimoto. It would be interesting to investigate this further and perhaps develop termination conditions for the algorithm.

6 Conclusion

This project has given an example of a problem where by exploiting the structure in the problem, it was possible to develop an algorithm that is very straightforward to implement, only requiring a dozen lines of MATLAB code.

References

[1] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inform. Theory*, 18:14–20, January 1972.

[2] R. E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inform. Theory*, 18:460–473, July 1972.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[4] H. Jürgensen. A note on the arimoto-blahut algorithm for computing the capacity of discrete memoryless channels. *IEEE Trans. Inform. Theory*, 30:376–377, March 1984.

[5] C. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, 623–656, 1948.