Finding the Capacity of a Quantized Binary-Input DMC

Brian M. Kurkoski
School of Information Science
Japan Advanced Institute of Science and Technology
Nomi, Japan 923-1292
kurkoski@jaist.ac.jp

Hideki Yagi
Dept. of Information and Communication Engineering
University of Electro-Communications
Tokyo, Japan 182-8585
h.yagi@uec.ac.jp

***Abstract*—Consider a binary-input, M-output discrete memoryless channel (DMC) where the outputs are quantized to K levels, with $K < M$. The subject of this paper is the maximization of mutual information between the input and quantizer output, over both the input distribution and channel quantizer. This can be regarded as finding the capacity of a quantized DMC. An algorithm is given, which either finds the optimal input distribution and corresponding quantizer, or declares a failure.**

I. INTRODUCTION

Consider a binary-input, M-output discrete memoryless channel (DMC) where the outputs are quantized to K levels, with $K < M$. The subject of this paper is the maximization of mutual information between the input and quantizer output, over both the input distribution and channel quantizer. This can be regarded as finding the capacity of a quantized DMC. An algorithm is given, which either finds the optimal input distribution and corresponding quantizer, or declares a failure.

Concretely, let the DMC input be X, let the DMC output be Y, and let the quantized output be Z. The alphabet sizes of X, Y and Z are J, M and K, respectively. Here, $K < M$ is of interest, since $K \geq M$ implies no reduction in mutual information due to quantization.

Let C denote the quantizer function, that is:

$$C : \{1, 2, \ldots, M\} \rightarrow \{1, 2, \ldots, K\}. \quad (1)$$

Let $C^{-1}(k)$ denote the subset of $\{1, \ldots, M\}$ that maps to channel quantizer k. Denote the channel input distribution as:

$$p_j = \Pr(\mathsf{X} = j),$$

denote the channel probability transition probability matrix as P, with elements:

$$P_{m|j} = \Pr(\mathsf{Y} = m|\mathsf{X} = j),$$

and denote the input-to-quantizer output transition probability matrix as T, with elements:

$$T_{k|j} = \Pr(\mathsf{Z} = k|\mathsf{X} = j) = \sum_{m \in C^{-1}(k)} P_{m|j}.$$

B.K. was supported in part by the Ministry of Education, Science, Sports and Culture; Grant-in-Aid for Scientific Research (C) number 23560439. H.Y. was supported in part by the Ministry of Education, Science, Sports and Culture; Grant-in-Aid for Young Scientists (B) number 22760270.

This paper makes the restriction that $J = 2$. Since $J = 2$, $p_1 + p_2 = 1$, and in the sequel the input distribution is $p = p_1$. The mutual information between X and Y is:

$$I(\mathsf{X};\mathsf{Y}) = \sum_j \sum_k p_j P_{m|j} \log \frac{P_{m|j}}{\sum_{j'} p_{j'} P_{m|j'}}. \quad (2)$$

The sum $\sum_m$, etc. means the sum over the whole alphabet $\sum_{m=1}^{M}$, etc. Mutual information $I(\mathsf{X};\mathsf{Y})$ is convex (lower convex) in $P_{m|j}$, for fixed p_j. Similarly, it is concave (upper convex) in p_j for fixed $P_{m|j}$ [1, Theorem 2.7.4].

It is well-known that the channel capacity is:

$$\max_p I(\mathsf{X};\mathsf{Y}) \quad (3)$$

and clearly for any fixed quantizer C the capacity of the quantized channel is $\max_p I(\mathsf{X};\mathsf{Z})$. Furthermore, the celebrated Arimoto-Blahut algorithm [2] [3] finds the capacity-achieving input distribution:

$$p^* = \arg\max_p I(\mathsf{X};\mathsf{Z}). \quad (4)$$

Since P and C uniquely determine T, the capacity-achieving input distribution of the quantized channel will be denoted as a function CAPACITY: $p^* = \text{CAPACITY}(P, C)$.

On the other hand, for any fixed p, it is possible [4] [5] to find the quantizer C^* which maximizes the mutual information,

$$C^* = \arg\max_C I(\mathsf{X};\mathsf{Z}) \quad (5)$$

which will be denoted as a function QUANT: $C^* = \text{QUANT}(P, p)$.

The objective in this paper is to find the jointly optimal input distribution p^* and channel quantizer C^* which maximizes the mutual information:

$$\max_{p,C} I(\mathsf{X};\mathsf{Z}). \quad (6)$$

This expression is regarded as the capacity of a quantized DMC. The algorithm described in this paper either finds a jointly optimal p^* and C^*, or it declares a failure.

The rest of this paper is outlined as follows. Section II describes previous work on the capacity of quantized channels and summarizes the contribution of this paper. Section III

978-1-4673-2579-0/12/$31.00

describes two key concepts: partial mutual information, a partial sum of mutual information; and a certificate, the range of input distributions over which a particular quantizer is known to be optimal. Section IV describes the main algorithm which either gives the quantizer and channel input distribution which maximizes mutual information, or declares a failure. Section V gives some numerical results that illustrate the algorithm. Section VI is the conclusion.

II. Previous Work and Contribution

The importance of designing channel quantizers has long been recognized as a topic of interest in information theory with practical applications. In the 1960s, Wozencraft and Kennedy suggested using the cut-off rate as a criteria for quantizer optimization [6], and design algorithms for both binary-input channels and non-binary inputs channels were described around that time [7] [8]. But the first known reference to using mutual information to design channel quantizers came in 2002 [9]. An important application is the design of analog-to-digital converters for communication receivers, and codes for such systems.

Singh et al. considered the capacity of quantized channels, but of continuous output channels, partticularly the AWGN channel [10]. For a fixed quantizer, optimal input distributions can be found using a cutting-plane algorithm. Since certain two-bit quantizers can be characterized by one parameter (for symmetrical channels), joint optimization of the input distribution and quantizer can be performed in a brute-force manner. But for three-bit quantization, it was necessary to resort to an optimization approach that involves alternating between finding the capacity-achieving input and the optimal quantizer, but this was not proved globally optimal.

By considering a DMC rather than a continuous output channel, further progress can be made on this problem. For a fixed input distribution, there exists a polynomial-time algorithm which gives the quantizer which maximizes mutual information [4] [5], for a binary input channel. When the channel outputs satisfy:

$$\log \frac{P_{1|1}}{P_{1|2}} < \log \frac{P_{2|1}}{P_{2|2}} < \cdots < \log \frac{P_{M|1}}{P_{M|2}}, \tag{7}$$

then for the optimal quantizer, each quantizer output consists of a convex subset of channel outputs. This quantization problem is an example of impurity partitions from machine learning, where convex subsets are known to be optimal [11]. While restricted to binary-input DMCs, this approach finds the optimal quantizer for otherwise arbitrary channels.

It is also worth noting that the optimal quantizer is known to be deterministic. That is, for a continuous-output channel there is no advantage to using dithered quantization [12]. And for a DMC, probabilistic quantizers are suboptimal [4].

Thus, various optimization problems have been considered. The channel capacity is a straightforward convex minimization problem (Arimoto-Blahut). Finding the optimal channel quantizer is a concave minimization problem which is NP-hard in general, but has polynomial complexity when attention is restricted to binary-input channels. So joint optimization of the input distribution and the quantizer is a convex-concave optimization problem, and provably optimal methods remain elusive.

The contribution of this paper is an algorithm which finds the jointly optimal input distribution and channel quantizer, for a given binary-input DMC, or declares a failure. The basic approach is to augment the quantization algorithm for a fixed input distribution [4], by adding a "certificate" property. The certificate is a range of input distributions over which the channel quantizer (or a partial quantization of the channel) is known to be optimal. The algorithm can be seen fom a dynamic programming perspective. Dynamic programming decompositions are an effective way to show the optimality of algorithms. Distinct from previous work, this approach can find the capacity of arbitrary channels.

III. Partial Mutual Information and Its Certificate

This section considers a partially quantized channel, by developing the concepts of partial mutual information and a certificate for a partially quantized channel. After these preliminary concepts are established, the algorithm to compute the DMC capacity is given in the following section.

A. Partial Mutual Information

The objective function in (6) is the mutual information between X and Z:

$$I(\mathsf{X};\mathsf{Z}) = \sum_j \sum_k p_j T_{k|j} \log \frac{T_{k|j}}{\sum_{j'} p_{j'} T_{k|j'}}. \tag{8}$$

This can be written as:

$$\sum_j \sum_k p_j \sum_{m \in C^{-1}(k)} P_{m|j} \log \frac{\sum_{m \in C^{-1}(k)} P_{m|j}}{\sum_{j'} p_{j'} \sum_{m \in C^{-1}(k)} P_{m|j'}}.$$

A partial sum of mutual information is called *partial mutual information* in this paper. Partial mutual information is a function of the input distribution p. The partial mutual information ι_k for output k and quantizer C is:

$$\iota_k(p; C) = \sum_j p_j \sum_{m \in C^{-1}(k)} P_{m|j} \log \frac{\sum_{m \in C^{-1}(k)} P_{m|j}}{\sum_{j'} p_{j'} \sum_{m \in C^{-1}(k)} P_{m|j'}}.$$

For quantizer outputs 1 to k, quantized using C, the partial mutual information is:

$$\iota(p; C, k) = \sum_{k'=1}^{k} \iota_{k'}(p; C). \tag{9}$$

So the total mutual information for some input distribution p' is the sum of all the partial mutual information terms evaluated at p':

$$I(\mathsf{X};\mathsf{Z}) = \iota(p', C, K) = \sum_k \iota_k(p', C).$$

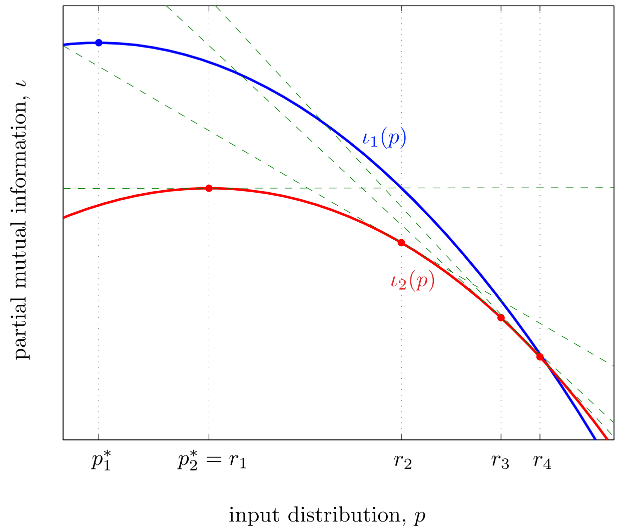

Fig. 1: Finding the certificate for $\iota_1(p)$, the region where $\iota_1(p) \geq \iota_2(p)$.

B. Certificate on Maximum Partial Mutual Information

Consider distinct quantizers C_1, C_2, ..., $C_i, \ldots$ and some fixed k. The partial mutual information is: $\iota(p, C_i, k)$. The input which maximizes mutual information can be found:

$$p_i^* \quad = \quad \arg\max_p \iota(p, C_i, k). \tag{10}$$

Without loss of generality, assume that the channels are ordered such that:

$$\iota(p_1^*, C_1, k) \geq \iota(p_2^*, C_2, k) \geq \cdots \tag{11}$$

Let $\mathcal{P}_i, i = 2, 3, \ldots$ be the domain of p for which C_1 has higher partial mutual information than C_i:

$$\mathcal{P}_i \quad = \quad \Big\{ p \;\Big|\; \iota(p, C_1, k) > \iota(p, C_i, k) \Big\}. \tag{12}$$

A *certificate* for C_1, denoted $\mathcal{L}$, is a domain of $\iota(p; C_1, k)$ for which C_1 achieves the maximum partial mutual information over all other quantizations:

$$\mathcal{L} \quad \subseteq \quad \mathcal{P}_2 \cap \mathcal{P}_3 \cap \cdots \tag{13}$$

Since $\mathcal{L}$ is a line segment, it is sufficient to represent $\mathcal{L}$ by the two values ℓ and r:

$$\mathcal{L} = [\ell, r] \quad = \quad \{p | \ell \leq p \leq r\}. \tag{14}$$

C. Finding the Certificate

The following subsection gives an explicit method for finding $\mathcal{P}_2$, but can of course be applied for any P_i. Then the certificate $\mathcal{L}$ is found as the intersection $\mathcal{P}_2 \cap \mathcal{P}_3 \cap \cdots$.

The algorithm input is two partially quantized channels with partial mutual information $\iota(p; C_1, k)$ and $\iota(p; C_2, k)$. Since C_1, C_2 and k are fixed, from here, write $\iota(p; C_i, k)$ as $\iota_i(p)$ since C and k are fixed. The corresponding derivatives are $\iota_1'(p)$ and $\iota_2'(p)$. The derivative of partial mutual information is:

$$\iota'(p; C, k) \quad = \quad \sum_{k'=1}^{k} \Big(-(T_{k'|1} - T_{k'|2})(\frac{1}{\ln 2} + \log_2 f(p)) + T_{k'|1} \log_2 T_{k'|1} - T_{k'|2} \log_2 T_{k'|2} \Big), \tag{15}$$

where $f(p) = pT_{k'|1} + (1-p)T_{k'|2}$.

1) Input: two partially-quantized channels with partial mutual information $\iota_1(p)$ and $\iota_2(p)$ with $\iota_1(p_1^*) > \iota_2(p_2^*)$
2) Initialize: $\ell_1 = p_2^*$ and $r_1 = p_2^*$.
3) For $i = 1, 2, \ldots$, find ℓ_{i+1}, the solution in p to:
$$\iota_2'(\ell_i)\big(p - \ell_i\big) + \iota_2(\ell_i) = \iota_1(p)$$
for which $\ell_{i+1} \leq \ell_i$. Repeat until a sufficiently accurate solution ℓ is obtained.
4) For $i = 1, 2, \ldots$, find r_{i+1}, the solution in p to
$$\iota_2'(r_i)\big(p - r_i\big) + \iota_2(r_i) = \iota_1(p)$$
for which $r_{i+1} \geq r_i$. Repeat until a sufficiently accurate solution r is obtained.
5) Output: $\mathcal{L} = [\ell, r]$.

The operation of the algorithm is illustrated in Fig. 1. The key point is that for any r, the line tangent to $\iota_2(p)$ at r is greater than or equal to $\iota_2(p)$:

$$\iota_2(p)(p - r) + \iota_2(r) \quad \geq \iota_2(p), \tag{16}$$

for $0 \leq p \leq 1$, and equality at $r = p$. In the region where this line is less than $\iota_1(p)$:

$$\{p | \iota_2(p)(p - r) + \iota_2(r) \leq \iota_1(p)\}, \tag{17}$$

the inequality $\iota_2(p) \leq \iota_1(p)$ holds.

D. Newton-Raphson Method

The Newton-Raphson method is an iterative technique for finding a root of $f(p)$ which has a derivative $f'(p)$. Beginning with an initial value p_1, compute:

$$p_{i+1} \quad = \quad p_i - \frac{f(p_i)}{f'(p_i)}, \tag{18}$$

iteratively until a sufficiently accurate value is obtained.

Here, $f(p)$ is the difference between the partial mutual information function $\iota(p)$ and a line. Since $\iota(p)$ is strictly convex, the equality $f(p) = 0$ has at most two solutions. However, which of the two solution found by the Newton-Raphson method depends upon the initial value. For Step 3, use an initial value of 0. For Step 4, use an initial value of 1.

To deal with this, the above iteration step is modified as:

$$q \quad = \quad p_i - \frac{f(p_i)}{f'(p_i)}, \tag{19}$$

and

$$p_{i+1} \quad = \quad \begin{cases} 0 & \text{if } q < 0 \\ q & \text{if } 0 \leq q \leq 1 \\ 1 & \text{if } q > 1 \end{cases} . \tag{20}$$

Because $f(x)$ is convex, this modification does not change the convergence.

IV. Algorithm to Compute the Capacity of a Quantized DMC

This section describes an algorithm which computes the capacity of the quantized DMC. A dynamic programming approach is used. In this way, it is possible to show the optimality of the algorithm. Note that the algorithm may fail, but if it does produce a solution, it is an optimal solution.

A. Consideration of Optimality

In dynamic programming, a problem exhibits optimal substructure if the optimal solution contains optimal solutions to subproblems [13]. The subproblem is as follows. For m channel outputs quantized to k quantizer outputs (with $m \leq M$ and $k \leq K$ and $k \leq m$), find $C_{m,k}$ with certificate $\mathcal{L}_{m,k}$ (that is, $C_{m,k}$ is known to be optimal over input distributions in the set $\mathcal{L}_{m,k}$). Assume that the optimal quantization:

$$C_{k-1,k-1}, C_{k,k-1}, \ldots, C_{n-1,k-1} \tag{21}$$

is known, and each has a corresponding certificate $\mathcal{L}_{k-1,k-1}$, $\mathcal{L}_{k,k-1}$, ..., $\mathcal{L}_{n-1,k-1}$.

The solution to the subproblem, forming the iterative step of the algorithm, is as follows. For some fixed m and k, and for some $n < m$, consider a candidate quantizer for channel outputs 1 to m, denoted $C^{(n)}_{m,k}$. This can be formed by combining the known-optimal quantizer $C_{n,k-1}$, with the quantization of channel outputs $n+1$ to m to the single output k. The candidate quantizer is given by:

$$C^{(n)}_{m,k}(m') = \begin{cases} C_{n,k-1}(m') & \text{if } 1 \leq m' \leq n \\ k & \text{if } n < m' \leq m \end{cases}. \tag{22}$$

For each n, compute the input distribution which achieves the maximum partial mutual information. Then select n^* for the quantizer $C^{(n*)}_{m,k}$ which has maximum mutual information (here n^* corresponds to 1 of the previous section) with certificate $\mathcal{K}$. Then, the optimal quantizer is $C^{(n*)}_{m,k}$ with certificate:

$$\mathcal{L}_{m,k} = \mathcal{L}_{n*,k-1} \cap \mathcal{K}. \tag{23}$$

Note that if $\mathcal{L}_{m,k} = \emptyset$, then a valid certificate cannot be found with this method, and the algorithm declares a failure.

B. Algorithm

The algorithm to compute the capacity of the quantized DMC is as follows.

1) Inputs:
   - Binary-input discrete memoryless channel $P_{m|j}$. If necessary, modify labels to satisfy (7).
   - The number of quantizer outputs K.
2) Initialize $C_{0,0} = \emptyset$ and $\mathcal{L}_{0,0} = [0, 1]$
3) For each $k \in \{1, \ldots, K\}$, and for each $m \in \{k, \ldots, k + M - K\}$
   a) for each $n \in \{k-1, \ldots, m-1\}$:
      - Find $C^{(n)}_{m,k}$ according to (22).
      - Compute $p^*_n = \text{Capacity}(P, C^{(n)}_{m,k})$.
   b) Select $n^* = \arg\max_n \iota(p^*_n; C^{(n)}_{m,k}, k)$
   c) Find $\mathcal{K}$, the certificate for n^*
   d) The locally optimal quantizer is:
      $$C_{m,k} = C^{(n*)}_{m,k}$$
      and the certificate is:
      $$\mathcal{L}_{m,k} = \mathcal{L}_{n*,k-1} \cap \mathcal{K}.$$
   e) If $\mathcal{L}_{m,k} = \emptyset$ then declare a failure. Stop.
4) Outputs. The globally optimal quantizer C^* is $C_{M,K}$. The capacity-achieving input distribution is:
   $$p^* = \text{Capacity}(P, C_{M,K}).$$

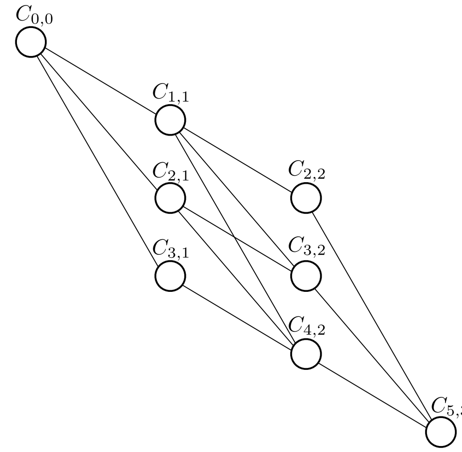

Fig. 2: The relationship between subproblems for case of $M = 5$ and $K = 3$.

The algorithm has polynomial complexity. In step 3, it can be seen there are three "for each" loops which which contribute M^3 operations. For each of these, it is necessary to find p^*, which is also polynomial complexity. Note that finding the certificate has complexity linear in M.

The relationship between the subproblems is illustrated in a trellis-type diagram in Fig. 2, for $M = 5$ and $K = 3$. For any node $C_{m,k}$, lines indicate those $C_{n,k-1}, n = k-1, \ldots, m-1$ which are used in step 3.

C. An Alternating Algorithm

An alternating algorithm is presented, which is based upon the principles similar to Singh et al [10]. By alternating between the DMC Quantization algorithm [4] (for a fixed p) and the Arimoto-Blahut algorithm (for a fixed C), this approach is straightforward:

1) Initialize with $i = 1$ and $p_1 = 0.5$.
2) $C_i = \text{Quant}(P, p_i)$
3) $p_{i+1} = \text{Capacity}(P, C_i)$
4) If $C_i = C_{i-1}$ and $i > 1$ then stop. Output quantizer C_i and distribution p_{i+1}.
5) $i \leftarrow i + 1$. Goto step 2.

This algorithm is considerably simpler, but it is not guaranteed to find optimum p^* and C^*. In particular, the capacity maximization may find an input distribution p which is locally optimal for all possible quantizers, but is distinct from the global optimal p^*.

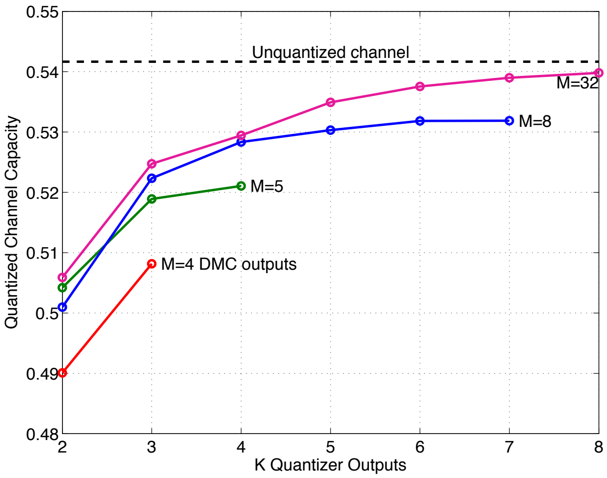

Fig. 3: For a test channel, the capacity of the quantized channel for various DMC outputs M and various quantizer outputs K.

V. Numerical Results

To illustrate, the following test channel is used. A BPSK channel with data-dependent noise is used, where Gaussian noise with variance 4 is added to -1 and Gaussian noise with variance 0.1 is added to $+1$. A DMC is formed by uniformly quantizing this between -1 and $+1$ to M levels.

A. Quantized Channel Capacity for the Test Channel

The quantized channel capacity of the test channel is shown in Fig. 3 for various values of M and K, with $K < M$. The unquantized channel capacity is also shown. Generally, a DMC with a larger number of outputs M has a greater channel capacity, for fixed K. Note an exception for $M = 5$ and $M = 8$, where the later has greater channel capacity. This may be attributed to relatively coarse channel quantization, where the boundaries for the $M = 5$ test channel are more suitable for quantization to $K = 2$.

For this particular test channel, the alternating algorithm of Subsec. IV-C produced the same quantizer and input distribution.

B. Algorithm Failure

The algorithm is able to certify the output. That is, if an output is produced it is known to be optimal. Otherwise, the algorithm declares a failure. Table I lists various combinations of M and K for the test channel. Cases where the algorithm failed are marked "F" (and success is marked "-"). The algorithm is more likely to fail when attempting to resolve small differences between competing quantizers. In these cases, the new certificate is relatively short, and has no overlap with the prior certificate, which may also be short. When there is no intersection, the capacity-achieving input distribution cannot be found.

VI. Conclusion

This paper has presented an algorithm which computes the capacity of quantized discrete memoryless channels. For a quantizer with K outputs, the algorithm may find the input distribution and quantizer which maximizes the mutual information. If it does not find these, then it declares a failure.

	K											
M	2	3	4	5	6	7	8	10	12	14	16	20
4	-	-										
5	-	-	-									
6	-	-	-	-								
7	-	-	-	-	-							
8	-	-	-	-	-	-						
10	-	-	-	-	-	-	-					
12	-	-	-	-	-	-	-	-				
16	-	-	-	-	-	-	-	-	-	-		
32	-	-	-	-	-	-	-	F	F	F	F	F
64	-	-	F	-	-	F	F	F	F	F	F	F

Table I: Test channels which failed are marked "F", and "-" indicates an optimal solution was provided.

Jointly maximizing mutual information in both the input distribution and the quantizer is concave-convex optimization problem; this class of problem is NP-hard in general. However, by exploiting the properties of efficient quantizers (described in II), polynomial complexity is possible. The optimality of the solution produced by the algorithm can be shown by dynamic programming techniques.

References

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

[2] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 14–20, January 1972.

[3] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, pp. 460–473, July 1972.

[4] B. M. Kurkoski and H. Yagi, "Quantization of binary-input discrete memoryless channels, with applications to LDPC decoding." Submitted to IEEE Transactions on Information Theory. Available http://arxiv.org/abs/1107.5637.

[5] B. Kurkoski and H. Yagi, "Concatenation of a discrete memoryless channel and a quantizer," in *Proceedings of the IEEE Information Theory Workshop*, (Cairo, Egypt), pp. 160–164, January 2010.

[6] J. M. Wozencraft and R. S. Kennedy, "Modulation and demodulation for probabilistic coding," *IEEE Transactions on Information Theory*, vol. 12, pp. 291–297, July 1966.

[7] J. L. Massey, "Coding and modulation in digital communications," in *Proceedings of International Zurich Seminar on Digital Communications*, (Zurich, Switzerland), pp. E2(1)–E2(4), 1974.

[8] L. Lee, "On optimal soft-decision demodulation," *IEEE Transactions on Information Theory*, vol. 22, pp. 437 – 444, July 1976.

[9] X. Ma, X. Zhang, H. Yu, and A. Kavcic, "Optimal quantization for soft-decision decoding revisited," in *International Symposium on Information Theory and its Applications*, (Xian, China), October 2002.

[10] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Transactions on Communications*, vol. 57, pp. 3629–3639, December 2009.

[11] D. Burshtein, V. D. Pietra, D. Kanevsky, and A. Nadas, "Minimum impurity partitions," *The Annals of Statistics*, vol. 20, no. 3, pp. 1637–1646, 1992.

[12] J. Singh, O. Dabeer, and U. Madhow, "Communication limits with low-precision analog-to-digital conversion at the receiver," in *International Conference on Communications, Circuits and Systems*, IEEE, 2007.

[13] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Second Edition*. The MIT Press, 2001.