Finding the Capacity of a Quantized Binary-Input DMC

Brian M. Kurkoski
Japan Advanced Institute of Science and Technology

and

Hideki Yagi
University of Electro-Communications
Tokyo, Japan

July 2, 2012
International Symposium on Information Theory
MIT, Cambridge, Massachusetts, USA
Information Theoretic Limits on Channel Quantization

- Detection and decoding algorithms: algorithms with real-valued numbers
- VLSI implementation: numbers are converted to bits — quantization
- Power consumption, cost, etc. of receivers increases with the number of bits
- **Broad goal**: Reduce the number of bits without sacrificing performance.
  - What are the information theoretic limits?

\[
p \sim X \xrightarrow{\text{channel}} Y \xrightarrow{\text{quantizer } Q} Z
\]

Two simpler questions:

1. Find channel quantizer to maximize the information rate

\[
Q^* = \arg \max_Q I(X; Z)
\]

2. Find the quantized channel capacity

\[
p^*, Q^* = \arg \max_{p,Q} I(X; Z)
\]
Quantization of Continuous-Output Channels

- 1960s and 70s: The cut-off rate as a criterion for channel quantizer design
- Mutual information is a better criterion (capacity-achieving LDPC codes)

1. Channel Quantization
   - BI-AWGN quantized to 3 levels [Ma et. al, 2002]
   - “Locally optimal” quantization algorithm [Liveris and Georghiades 2003]
   - Quantization of flash memory channel models [Wang et al 2011]

2. Quantized Channel Capacity
   - Singh et al 2007, 2009: Continuous channels quantized to $K$ levels
     - input distribution: $K$ or $K+1$ discrete levels sufficient
     - locally optimal/brute force channel quantization
   - Quantization of channels with memory [Zeitler, Singer and Kramer 2010, 2011]

These are hard problems! Few proofs of optimality
(Highly Simplified)

Channel Quantization is Concave Optimization

Mutual information $I(X; Z)$: function of

- input distribution $p$
- and quantizer $Q$

<table>
<thead>
<tr>
<th>Optimization conditions</th>
<th>Known as</th>
<th>Class of problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximize over $p$, fixed channel</td>
<td>Classical channel capacity</td>
<td>convex optimization</td>
</tr>
<tr>
<td>Fixed $p$, minimize over $Q$</td>
<td>Classical rate-distortion</td>
<td>Arimoto-Blahut algorithm [1972]</td>
</tr>
<tr>
<td>Fixed $p$, maximize over $Q$</td>
<td>1. Channel quantization</td>
<td>concave optimization, NP hard</td>
</tr>
<tr>
<td>Maximize over $p$ and over $Q$</td>
<td>2. Quantized channel capacity</td>
<td>convex-concave optimization</td>
</tr>
</tbody>
</table>
Contributions of This Work

Consider discrete memoryless channels, rather than continuous channels

\[ p \sim X \xrightarrow{\text{DMC}} Y \xrightarrow{Q} Z \]

1. Channel Quantization
   - Maximize mutual information \( I(X;Z) \), provably optimal
   - Polynomial-complexity algorithm, dynamic programming approach
2. Quantized Channel Capacity
   - Find the jointly optimal input distribution/quantizer or declare failure
   - also polynomial complexity

Applies to arbitrary DMCs with **binary inputs**
**Optimal quantizer is convex**

Consider a DMC with $P_{i|j}$. Goal: $\max I(X; Z)$

The channel outputs are points in a 1-D space

$$\frac{P_{i|1}}{P_{i|2}},$$

for $i = 1, 2, \ldots, M$

Lemma: If the outputs are labeled according to:

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

then the mutual-information maximizing quantizer is convex on $\{1,2,\ldots, M\}$. 
**Optimal quantizer is convex**

Consider a DMC with $P_{i|j}$. Goal: $\max I(X; Z)$

The channel outputs are points in a 1-D space

$$\frac{P_{i|1}}{P_{i|2}},$$

for $i = 1, 2, \ldots, M$

Lemma: If the outputs are labeled according to:

$$\frac{P_{1|1}}{P_{1|2}} < \frac{P_{2|1}}{P_{2|2}} < \ldots < \frac{P_{M|1}}{P_{M|2}},$$

then the mutual-information maximizing quantizer is convex on $\{1, 2, \ldots, M\}$. 

$M$ channel outputs

$K$ quantizer outputs
Statistical Learning Theory

Optimal Quantizer is Convex: Proof Sketch

Proof sketch: \[ I(X; Z) = H(X) - H(X|Z) \]
\[ H(X|Z) = -E_Z \left[ E_{X|Z} \left[ \log p(X|Z) \right] \right] \]

Statistical/Machine Learning Theory

- Classification: from observation, make a classification (e.g. optical character recognition)
- Minimize some loss function (supervised learning)
- Impurity (or risk) is the expectation of an expectation of a loss function:

\[
R(q) = \sum_{t\in\mathcal{T}} P(t) \cdot E[\ell(Y, \hat{y}(t)) | t], \quad (2)
\]

[Burshtein et al., 1992]

Broad class of loss functions, preimage of optimum classification mapping forms a convex set:

2. Results.

**Theorem 1.** For any \( C: \mathcal{X} \rightarrow \mathcal{C} \) there exists a \( \hat{C}: \mathcal{Y} \rightarrow \mathcal{C} \) such that \( \Psi(\hat{C}(Y)) \leq \Psi(C) \) and such that \( \hat{C}^{-1}(c) \) is convex for all \( c \in \mathcal{C} \).

[Burshtein et al., 1992]

Identify \( H(X|Z) \) as an impurity. Quantizer (mapping) image forms a convex set.
1. Channel Quantization
   Possible algorithmic approaches

Algorithms ($M$ channel outputs, $K$ quantizer outputs):

1. Brute force (ignore convexity): complexity $K^M$

2. Search all convex sets. Complexity:
   \[
   \binom{M-1}{K-1}
   \]
   - Worst-case complexity: exponential in $K$

3. Proposed Dynamic programming approach:
   - Avoid recomputing part of the total sum
   - Parts are called “partial mutual information”
   - Complexity $M^3$

\[
I(X;Z) = \sum_{k=1}^{K} \left( \sum_{j=1}^{2} p_{k,j} \log \frac{T_{k,j}}{\sum_{j=1}^{2} p_{k,j} T_{k,j}} \right)
\]
\[
I(X;Z) = \sum_{k=1}^{K} (\mu_k)
\]

Example: $M = 5$ channel outputs, $K = 3$ quantizer outputs

\[
\begin{align*}
\binom{4}{2} &= 6 \\
1 &\quad 2 &\quad 3 &\quad 4 &\quad 5 \\
1 &\quad 2 &\quad 3 &\quad 4 &\quad 5 \\
1 &\quad 2 &\quad 3 &\quad 4 &\quad 5 \\
1 &\quad 2 &\quad 3 &\quad 4 &\quad 5 \\
1 &\quad 2 &\quad 3 &\quad 4 &\quad 5 \\
k = 1 &\quad k = 2 &\quad k = 3
\end{align*}
\]
1. Channel Quantization
Algorithm for Optimal quantization

Dynamic programming: Optimal solution contains optimal solutions to subproblems

Subproblem find optimal quantization of ch. outputs 1...m to quantizer outputs 1...k

Key: use the optimal quantization 1...k-1 to find optimal quantization of 1...k

Example: how to quantize outputs 1-4 to two values

\[ i(1 \rightarrow 1) + i(2 \rightarrow 4) \]
\[ i(1 \rightarrow 2) + i(3 \rightarrow 4) \]
\[ i(1 \rightarrow 3) + i(4 \rightarrow 4) \]

partial sum of mutual information

\[ \sum_{i=a'+1}^{a} \sum_{j} p_{j}^{i} \log \frac{\sum_{i'=a'+1}^{a} P_{i'|j}}{\sum_{j'} P_{j'} \sum_{i'=a'+1}^{a} P_{i'|j'}}. \]
1. Channel Quantization
Algorithm for Optimal quantization

Dynamic programming: Optimal solution contains optimal solutions to subproblems

Subproblem find optimal quantization of ch. outputs 1...m to quantizer outputs 1...k

Key: use the optimal quantization 1...k-1 to find optimal quantization of 1...k

Example: how to quantize outputs 1-5 to three values
2. Quantized Channel Capacity
Jointly Optimal Quantizer and Input Distribution

Similar procedure to previous. For $m$ outputs quantized to $k$ levels:
- for each quantizer: maximum partial mutual information
- select the quantizer with the greatest partial mutual information
- for this one, find a range $L$ of locally optimal input distributions

This range is important
1. if final $p^*$ is range $L$, then it is known optimal.
2. otherwise, another quantizer may be optimal
   -> Declare a failure

Continue recursively:
- locally optimal quantizer and
- the range of input distributions
2. Quantized Channel Capacity
Jointly Optimal Quantizer and Input Distribution

- Similar procedure to previous. For $m$ outputs quantized to $k$ levels:
  - for each quantizer: maximum partial mutual information
  - select the quantizer with the greatest partial mutual information
  - for this one, find a range $L$ of locally optimal input distributions

- This range is important
  1. if final $p^*$ is range $L$, then it is known optimal.
  2. otherwise, another quantizer may be optimal
     -> Declare a failure

- Continue recursively:
  - locally optimal quantizer and
  - the range of input distributions
2. Quantized Channel Capacity

Example: Optimal Quantizers

Create a DMC:

- BPSK +1/−1
- data-dependent AWGN noise
- quantize uniformly $M$ levels

Example:

- $M = 16$ output channel
- quantize to $K = 4$ levels

Observations:

- data-independent noise: decisions cluster at cross-over
- data-dependent noise: decisions move towards reliable data
2. Quantized Channel Capacity

Example: Capacity

Asymmetric channel (var 0.1 & 4)
- Capacity increases in $K$, $M$
- Asymmetric optimal quantizer

 Failure to converge
- Large $M$: Many good candidate quantizers, so optimal range is narrow
Conclusion

Channel quantization is important for reducing complexity of receivers

- Maximization of mutual information is a highly suitable metric
- These concave optimization problems are NP-Hard

Easier to work with discrete problems than continuous problems

For arbitrary binary-input DMCs:

- 1. Channel quantization (fixed input distribution):
  - Maximize mutual information, provably optimal
  - Polynomial (cubic) complexity

- 2. Quantized channel capacity
  - Find the jointly optimal input distribution/quantizer or declare failure
  - Also polynomial-complexity