On the Construction of Polar Codes for Channels with Moderate Input Alphabet Sizes

Ido Tal

Problem: Construction of polar (LDPC) codes, for a channel with moderate input alphabet size q. Say, $q \geq 16$.

Punchline: Provably hard*†‡§.

*For a specific channel
†under a certain construction model
‡deterministically
§some more assumptions

Given:

- Underlying channel $\mathcal{W} : \mathcal{X} \to \mathcal{Y}_{\mathrm{und}}$
  - $|\mathcal{X}| = q$
  - Uniform input distribution is capacity achieving
- Codeword length $n = 2^m$

Goal:

- Assuming uniform input, calculate misdecoding probability of synthesized channels

$$\mathcal{W}_i^{(m)} : \mathcal{X} \to \mathcal{Y}_i \ , \quad 0 \le i < n$$

- Unfreeze channels with very low probability of misdecoding

$P_{\mathrm{U}(\mathcal{X})} \triangleq$ uniform distribution on input alphabet $\mathcal{X}$

Algorithm: Naive solution

---

input : Underlying channel $\mathcal{W}$, index $i = \langle b_1, b_2, \ldots, b_m \rangle_2$
output: $P_e(\mathcal{W}_i^{(m)}, P_{\mathrm{U}(\mathcal{X})})$

```
W ← 𝒲
for j = 1, 2, ..., m do
    if b_j = 0 then
        W ← W⁻
    else
        W ← W⁺
return P_e(W, P_U(𝒳))
```

Problem: $\mathcal{Y}_i$ grows exponentially with n.

$P_{\mathrm{U}(\mathcal{X})} \triangleq$ uniform distribution on input alphabet $\mathcal{X}$

Algorithm: Degrading solution

input : Underlying channel $\mathcal{W}$, index $i = \langle b_1, b_2, \ldots, b_m \rangle_2$
, bound on output alphabet size L

output: Upper bound on $P_e(\mathcal{W}_i^{(m)}, P_{\mathrm{U}(\mathcal{X})})$

```
Q ← degrading_merge(𝒲, L, P_U(𝒳))
for j = 1, 2, ..., m do
    if b_j = 0 then
        W ← Q⁻
    else
        W ← Q⁺
    Q ← degrading_merge(W, L, P_U(𝒳))
return P_e(Q, P_U(𝒳))
```

Question: How good of an approximation to W is degrading_merge(W, L, $P_{\mathrm{U}(\mathcal{X})}$)?

Notation:

- $W : \mathcal{X} \to \mathcal{Y}$ — generic memoryless channel
- $q = |\mathcal{X}|$ — input alphabet size
- P_X — input distribution
- $Q : \mathcal{X} \to \mathcal{Y}'$ — degraded version of W
- L — bound on new output alphabet size, $|\mathcal{Y}'| \leq L$
- X — input to W or Q
- Y — output of W
- Y' — output of Q

Goal: $\texttt{degrading\_merge}(W, L, P_X)$ must find $Q : \mathcal{X} \to \mathcal{Y}'$ such that

- Q degraded with respect to W
- $|\mathcal{Y}'| \leq L$
- $\Delta = I(X;Y) - I(X;Y')$ is "small"

An implementation of `degrading_merge`(W, L, P_X) exists [TalSharovVardy] for which

$$\Delta = I(X;Y) - I(X;Y') \leq O\left(\left(\frac{1}{L}\right)^{1/q}\right)$$

Apropos: similar behaviour in upgraded case [PeregTal]

Totally useless (at least in theory), for moderate q:

$$q = 16\ , \quad \Delta \leq 0.01 \quad \Longrightarrow \quad L \approx 10^{32}$$

Good luck...

An inherent difficulty?

What can be said about

$$\mathrm{DC}(q, L) \triangleq \sup_{W, P_X} \min_{\substack{Q \,:\, Q \prec W, \\ |\mathrm{out}(Q)| \leq L}} \left(I(W) - I(Q) \right) .$$

We already know that

$$\mathrm{DC}(q, L) \leq O\left(\left(\frac{1}{L} \right)^{1/q} \right)$$

Need: a lower bound on $\mathrm{DC}(q, L)$

Cut to the end

$$\mathrm{DC}(q,L) \triangleq \sup_{W,P_X} \min_{\substack{Q\,:\,Q \prec W,\\ |\mathrm{out}(Q)| \le L}} \left(I(W) - I(Q)\right)$$

We will shortly prove that

$$\mathrm{DC} \ge O\left(\left(\frac{1}{L}\right)^{\frac{2}{q-1}}\right)$$

Above attained for

- Uniform input distribution $P_X = P_{\mathrm{U}(\mathcal{X})}$
- Sequence $\mathcal{W}_1, \mathcal{W}_2, \ldots$ of "progressively hard channels"
- The capacity achieving input distribution of each $\mathcal{W}_M$ is the uniform distribution $P_{\mathrm{U}(\mathcal{X})}$

Consequences: Try and build a polar code for $\mathcal{W}_M \ldots$

Algorithm: Degrading solution

input : Underlying channel $\mathcal{W}$, index $i = \langle b_1, b_2, \ldots, b_m \rangle_2$
, bound on output alphabet size L

output: Upper bound on $P_e(\mathcal{W}_i^{(m)}, P_{\mathrm{U}(\mathcal{X})})$

$\mathsf{Q} \leftarrow \texttt{degrading\_merge}(\mathcal{W}, L, P_{\mathrm{U}(\mathcal{X})})$
for $j = 1, 2, \ldots, m$ **do**
  if $b_j = 0$ **then**
    $\mathsf{W} \leftarrow \mathsf{Q}^-$
  else
    $\mathsf{W} \leftarrow \mathsf{Q}^+$
  $\mathsf{Q} \leftarrow \texttt{degrading\_merge}(\mathsf{W}, L, P_{\mathrm{U}(\mathcal{X})})$
return $P_e(\mathsf{Q}, P_{\mathrm{U}(\mathcal{X})})$

Consequences: Try and build a polar code for $\mathcal{W}_M\ldots$

- Would like number of good channels to be

$$\approx n \cdot I(\mathcal{W}_M)$$

- However, number of good channels is upper bounded by

$$n \cdot I\left(\texttt{degrading\_merge}(\mathcal{W}_M, L, P_{\mathrm{U}(\mathcal{X})})\right) \geq n \cdot \left(I(\mathcal{W}_M) - O\left(\left(\frac{1}{L}\right)^{\frac{2}{q-1}}\right)\right)$$

For $q = 16$, in order to lose at most 0.01, need $L \approx 10^{15}$

LDPC:

Same problem when trying to design an LDPC code for $\mathcal{W}_M$

- Pick a code ensamble with rate close to $I(\mathcal{W}_M)$
- Use density evolution to asses code:
  1. Initialize
     - Assume all-zero codeword
     - Quantize output letters: letters with close posteriors are grouped together
  2. Main loop
     - Already hopeless at this point: main loop is with respect to quantized channel, which has mutual information below design rate

The channel $\mathcal{W}_M$:

For an integer $M \geq 1$, define $\mathcal{W}_M : \mathcal{X} \to \mathcal{Y}_M$ as follows:

- Input alphabet is $\mathcal{X} = \{1, 2, \ldots, q\}$
- Output alphabet is

$$\mathcal{Y}_M = \left\{ \langle j_1, j_2, \ldots, j_q \rangle : j_1, j_2, \ldots, j_q \geq 0 \; , \quad \sum_{x=1}^{q} j_x = M \right\} ,$$

  where j_x are non-negative integers summing to M
- Channel transition probabilities:

$$\mathcal{W}(\langle j_1, j_2, \ldots, j_q \rangle | x) = \frac{q \cdot j_x}{M \binom{M+q-1}{q-1}}$$

- Input distribution unifrom $\Longrightarrow$ all output letters equally likely

The channel $\mathcal{W}_M$:

- Posterior probabilities

$$P(X = x | Y = \langle j_1, j_2, \dots, j_q \rangle) = \frac{j_x}{M}$$

- Shorthand: output letter is labelled by posterior probabilities vector

$$\langle j_1, j_2, \dots, j_q \rangle \triangleq (j_1/M, j_2/M, \dots, j_q/M)$$

Optimal degrading:

Claim [KurkoskiYagi]:

- Let $W : \mathcal{X} \to \mathcal{Y}$, P_X, and L be given.
- Let $Q : \mathcal{X} \to \mathcal{Z}$ be an optimal degrading of W to a channel Q with $|\mathcal{Z}| \leq L$.
- That is, $I(X, Y) - I(X, Y')$ is minimized.
- Then, Q is gotten from W by defining a partition $(A_i)_{i=1}^{L}$ of $\mathcal{Y}$ and mapping with probability 1 all symbols in A_i to a single symbol $z_i \in \mathcal{Z}$

Let $(A_i)_{i=1}^{L}$ be such a partition with respect to $\mathcal{W}_M$

L_2 squared bound:

Lemma: For $A = A_i$ as above, let $\Delta(A)$ be the drop in mutual information incurred by merging all the letters in A_i into a single letter. Then,

$$\Delta(A) \geq \tilde{\Delta}(A) \ ,$$

where

$$\tilde{\Delta}(A) = \frac{1}{2\binom{M+q-1}{q-1}} \sum_{\mathbf{p} \in A} \|\mathbf{p} - \bar{\mathbf{p}}\|_2^2 \ , \quad \bar{\mathbf{p}} = \sum_{\mathbf{p} \in A} \frac{1}{|A|} \mathbf{p} \ .$$

Bounding in terms of $|A|$:

Lemma:

$$\sum_{i=1}^{L} \Delta(A_i) \geq \sum_{i=1}^{L} \tilde{\Delta}(A_i) \geq \mathrm{const}(q) \cdot \sum_{i=1}^{L} |A_i|^{\frac{q+1}{q-1}} + o(1) \,,$$

where the $o(1)$ is a function of M alone and goes to 0 as $M \to \infty$

Observation: Up to the $o(1)$, expression is convex in $|A_i|$. Thus, sum is lower bounded by setting $|A_i| = |\mathcal{Y}_M|/L$.

Theorem:

$$\mathrm{DC}(q, L) \geq \frac{q-1}{2(q+1)} \cdot \left(\frac{1}{\sigma_{q-1} \cdot (q-1)!}\right)^{\frac{2}{q-1}} \cdot \left(\frac{1}{L}\right)^{\frac{2}{q-1}} ,$$

where σ_{q-1} is the constant for which the volume of a sphere in $\mathbb{R}^{q-1}$ of radius r is $\sigma_{q-1} r^{q-1}$

Backup

- Just how representative is $\mathcal{W}_M$?
- What can be done?
- Channels $\mathcal{W}_M$ "converges" to
  - $\mathcal{W}_\infty : \mathcal{X} \to \mathcal{X} \times [0,1]^q$
  - Given an input x, the channel picks $\varphi_1, \varphi_2, \ldots, \varphi_q$, non-negative reals summing to 1. All possible choices are equally likely, Dirichlet(1,1,...,1)
  - Then, the input x is transformed into $x + i$ (with a modulo operation where appropriate) with probability φ_i
  - The transformed symbol along with the vector $(\varphi_1, \varphi_2, \ldots, \varphi_q)$ are the output of the channel