An Upgrading Algorithm with Optimal Power Law

Or Ordentlich[1] Ido Tal[2]

[1]Hebrew University [2]Technion

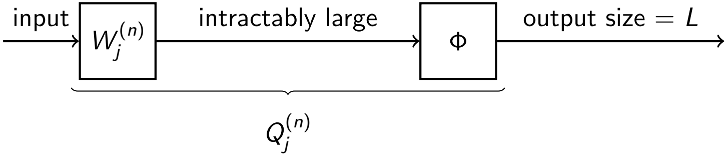
Big picture first

In this talk:

- An upgrading algorithm for channels with non-binary input
- Optimal power law
- Achieved by reduction to the binary-input case
- Important for constructing polar codes

Constructing vanilla polar codes

- Underlying channel: a binary-input symmetric and memoryless channel $W : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} = \{0, 1\}$
- Derive $N = 2^n$ synthetic channels $W_j^{(n)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{j-1}$, where $1 \leq j \leq N$.
- Constructing a vanilla polar code $\equiv$ finding which synthetic channels $W_j^{(n)}$ are 'almost noiseless'
- Problem: output alphabet $\mathcal{Y}^N \times \mathcal{X}^{j-1}$ is intractably large
- Solution:
  - Replace $W_j^{(n)}$ with $Q_j^{(n)}$ having output alphabet size L
  - Have $Q_j^{(n)}$ be (stochastically) degraded with respect to $W_j^{(n)}$



  - $Q_j^{(n)}$ almost noiseless $\Longrightarrow$ $W_j^{(n)}$ almost noiseless

Constructing vanilla polar codes

- We write $Q \leq W$ if Q is degraded with respect to W
- Alternatively, we write $W \geq Q$ and say that W is upgraded with respect to Q
- Previous slide:

$$Q_j^{(n)} \leq W_j^{(n)}$$

- We can also approximate $W_j^{(n)}$ "from above" by an upgraded channel $R_j^{(n)}$ having output alphabet size at most L.
- Sandwich property:

$$Q_j^{(n)} \leq W_j^{(n)} \leq R_j^{(n)}$$

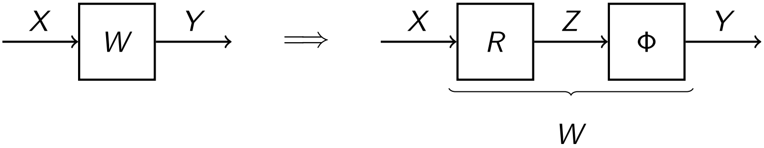
- In vanilla setting, $R_j^{(n)}$ has secondary importance...

Constructing generalized polar codes

- Polar codes have been generalized beyond vanilla setting
  - Asymmetric channels (with asymmetric input distribution)
  - Wiretap channels
  - Channels with memory (input distribution can have memory as well)
- In all these settings upgrading is as important as degrading for constructing the code
- For settings with memory, the "effective input alphabet" is non-binary

Problem statement

- Given: joint distribution of channel and input $P_{X,Y}(x,y)$
  - $x \in \mathcal{X}$, the input alphabet and $y \in \mathcal{Y}$, the output alphabet
  - $P_{X,Y}(x,y) = \underbrace{P_X(x)}_{\text{input distribution}} \cdot \underbrace{P_{Y|X}(y|x)}_{\text{channel}}$
- Find: $P^*_{X,Z,Y}(x,z,y)$ such that
  - Marginalization: $\sum_z P^*_{X,Z,Y}(x,z,y) = P_{X,Y}(x,y)$
  - Upgrading: $X - Z - Y$ is a Markov chain
  - Tractable output alphabet size: $z \in \mathcal{Z}$ and $|\mathcal{Z}| \leq L$



  - Figure of merit:

$$H(X|Y) - H(X|Z) = I(X;Z) - I(X;Y)$$

should be ‘small’

Power law

- Previous results:
  - Recall: output alphabet size of upgraded channel $|\mathcal{Z}| \leq L$
  - There exists a 'hard to upgrade' joint distribution $P(X, Y)$:

$$H(X|Y) - H(X|Z) = \Omega(L^{-2/(|\mathcal{X}|-1)})$$

  - For binary input, $|\mathcal{X}| = 2$, and any $P_{X,Y}$, there exists an upgrading algorithm such that

$$H(X|Y) - H(X|Z) = O(L^{-2}) = O(L^{-2/(|\mathcal{X}|-1)})$$

- New result:
  - Also for non-binary input, we can upgrade any $P_{X,Y}$ and achieve

$$H(X|Y) - H(X|Z) = O(L^{-2/(|\mathcal{X}|-1)})$$

- Main idea: use binary-input as a black-box (reduction)

One-hot representation

- Denote $q = |\mathcal{X}|$. Assume

$$\mathcal{X} = \{1, 2, \ldots, q\}$$

- For $x \in \mathcal{X}$, define

$$g(x) = (x_1, x_2, \ldots, x_{q-1}) \ ,$$

  the one-hot representation:

$$\begin{aligned} g(1) &= (1, 0, 0 \ldots 0, 0) \\ g(2) &= (0, 1, 0 \ldots 0, 0) \\ &\vdots \\ g(q-1) &= (0, 0, 0 \ldots 0, 1) \\ g(q) &= (0, 0, 0 \ldots 0, 0) \end{aligned}$$

- Abuse notation and write $x = g(x) = (x_1, x_2, \ldots, x_{q-1})$

$P_{X,Y} \Longrightarrow \alpha^{(i)} \Longrightarrow \beta^{(i)} \Longrightarrow \gamma^{(i)} \Longrightarrow P^*_{X,Z,Y}$

- We are given $P_{X,Y}$, where $|\mathcal{X}| = q$
- Need to produce $P^*_{X,Z,Y}$ by reducing to binary-input upgrading
- Denote $\mathcal{X}' = \{0, 1\}$
- Let $X = (X_1, X_2, \ldots, X_{q-1})$ and Y be distributed according to $P_{X,Y}$
- First step: define, for $1 \le i \le q-1$ the joint distribution

$$\alpha^{(i)}_{X_i,Y}(x', y) = P(X_i = x', Y = y | X_1^{i-1} = 0_1^{i-1})$$

- The joint distribution $\alpha^{(i)}_{X_i,Y}(x', y)$ has binary input, $x' \in \mathcal{X}'$
- We may apply our binary-input upgrading procedure

$P_{X,Y} \Longrightarrow \alpha^{(i)} \Longrightarrow \beta^{(i)} \Longrightarrow \gamma^{(i)} \Longrightarrow P^*_{X,Z,Y}$

- Recall our binary-input joint distribution: for $1 \le i \le q-1$,

$$\alpha^{(i)}_{X_i,Y}(x', y) = P(X_i = x', Y = y | X_1^{i-1} = 0_1^{i-1})$$

- Define

$$\Lambda = \left\lfloor L^{1/(q-1)} \right\rfloor .$$

- Second step:
  - Apply our binary-input upgrading procedure to $\alpha^{(i)}_{X_i,Y}(x', y)$, resulting in

$$\beta^{(i)}_{X_i,Z_i,Y}(x', z, y) ,$$

    where

$$|\mathcal{Z}_i| \le \Lambda$$

  - Difference in entropies is $O(\Lambda^{-2})$

$$P_{X,Y} \Longrightarrow \alpha^{(i)} \Longrightarrow \beta^{(i)} \Longrightarrow \gamma^{(i)} \Longrightarrow P^*_{X,Z,Y}$$

- Recall that we have produced $\beta^{(i)}_{X_i,Z_i,Y}(x', z, y)$, where $x' \in \mathcal{X}'$ is binary
- Third step: define the conditional distribution

$$\gamma^{(i)}_{X_i|Z_i,X_1^{i-1}}(x_i|z_i, x_1^{i-1}) = \begin{cases} \beta^{(i)}_{X_i|Z_i}(x_i|z_i) & \text{if } x_1^{i-1} = 0_1^{i-1} \,, \\ 1 & \text{if } x_1^{i-1} \neq 0_1^{i-1} \text{ and } x_i = 0 \,, \\ 0 & \text{otherwise} \,. \end{cases}$$

- That is, if x_1^{i-1} is non-zero, force x_i to zero, in accordance with the one-hot representation
- Otherwise, if x_1^{i-1} is zero, use $\beta^{(i)}_{X_i|Z_i}$

$$P_{X,Y} \Longrightarrow \alpha^{(i)} \Longrightarrow \beta^{(i)} \Longrightarrow \gamma^{(i)} \Longrightarrow P^*_{X,Z,Y}$$

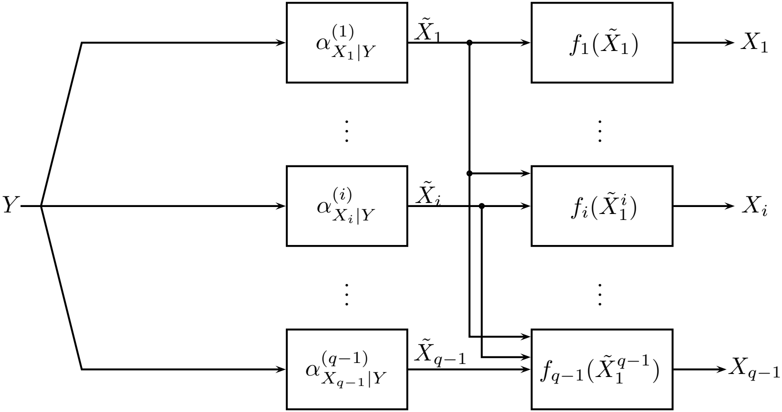
- Last step: define

$$P^*_{X,Z,Y}(x,z,y) = P_Y(y) \cdot \left(\prod_{i=1}^{q-1} \beta^{(i)}_{Z_i|Y}(z_i|y) \right) \cdot \left(\prod_{i=1}^{q-1} \gamma^{(i)}_{X_i|Z_i,X_1^{i-1}}(x_i|z_i,x_1^{i-1}) \right)$$

- A valid upgrade, with optimal power law:

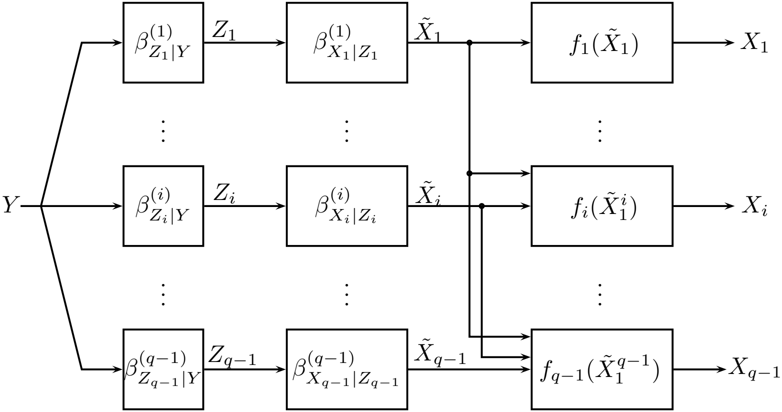
$$H(X|Y) - H(X|Z) = O(L^{-2/(|\mathcal{X}|-1)})$$

A graphical description of $P_{X,Y}$



where $f_i(\tilde{x}_1^i) \triangleq \tilde{x}_i \cdot \mathbb{1}_{\{\tilde{x}_1^{i-1}=0_1^{i-1}\}}$

A graphical description of $P_{X,Z,Y}$



where $f_i(\tilde{x}_1^i) \triangleq \tilde{x}_i \cdot \mathbb{1}_{\{\tilde{x}_1^{i-1}=0_1^{i-1}\}}$