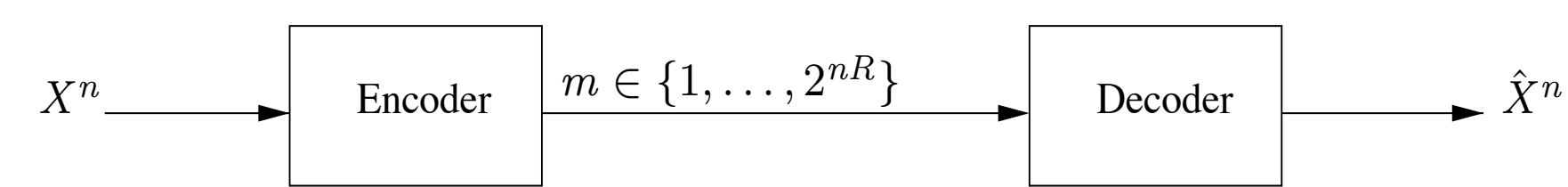


# An Implementable Scheme for Universal Lossy Compression

2nd Annual North American School of Information Theory

Shirin Jalali (Advisor : Tsachy Weissman)

Rate-Distortion Coding	Universal lossy compression	Empirical distribution	Conditional empirical entropy
------------------------	-----------------------------	------------------------	-------------------------------



- $\{X_i\}_{i \in \mathbb{Z}}$ : discrete-valued stochastic source
- To each coding scheme a distortion  $D$  is assigned as

$$D = \mathbb{E} d_n(X^n, \hat{X}^n) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{E} d(X_i, \hat{X}_i)$$

- The goal is to find a coding scheme that for each given  $\alpha > 0$  minimizes the Lagrangian cost  $R + \alpha D$

- The performance of each coding scheme depends on the source distribution.
- Does there exist a **universal** coding scheme that works well for any stationary ergodic source?
  - **Yes.**
  - Are there practical universal lossy compression algorithms?
  - For a long time it was believed that **no** such algorithm exists.

- Consider binary alphabet  $\mathcal{X} = \mathcal{X} = \{0, 1\}$
- Let  $\mathbf{m}(y^n)$  denote a  $2 \times 2^k$  matrix defining the  $(k+1)^{\text{th}}$  order empirical distribution of  $y^n$
- Example:  $k = 2$ ,

$$\mathbf{m}(y^n) = \begin{bmatrix} m_{0,00} & m_{0,01} & m_{0,10} & m_{0,11} \\ m_{1,00} & m_{1,01} & m_{1,10} & m_{1,11} \end{bmatrix}$$

For  $\beta \in \{0, 1\}$

$$m_{\beta,00}(y^n) = \frac{1}{n} |\{i : y_i = \beta, y_{i-2}^{i-1} = [0, 0]\}|$$

- **Conditional empirical entropy** of  $y^n$  is defined as  $H_k(y^n) = H(Y_{k+1} | Y^k)$  when  $Y^{k+1}$  is distributed according to  $\mathbf{m}$
- In the previous example:

$$H_2(y^n) = \sum_{b_1, b_2} (m_{0,[b_1, b_2]} + m_{1,[b_1, b_2]}) \mathcal{H}(m_{0,[b_1, b_2]}, m_{1,[b_1, b_2]})$$

- Let  $\alpha_1 = m_{0,[b_1, b_2]}$  and  $\alpha_2 = m_{1,[b_1, b_2]}$ . Then

$$\mathcal{H}(\alpha_1, \alpha_2) = H(X),$$

where  $X \sim \text{Bern}(\frac{\alpha_1}{\alpha_1 + \alpha_2})$

Universal lossy compression algorithm	Compression via Simulated Annealing	Another approach	Viterbi algorithm
---------------------------------------	-------------------------------------	------------------	-------------------

## An Exhaustive Search Algorithm:

For coding sequence  $x^n$ , find sequence  $\hat{x}^n$  such that

$$\hat{x}^n = \arg \min_{y^n} [H_k(y^n) + \alpha d_n(x^n, y^n)],$$

and describe it to the decoder using Lempel-Ziv algorithm.

### Theorem:

For any stationary ergodic source  $\mathbf{X}$  and  $k = o(\log n)$ ,

$$\mathbb{E} \left[ \ell_{\text{LZ}}(\hat{X}^n) + \alpha d(X^n, \hat{X}^n) \right] \rightarrow \min_D [R(\mathbf{X}, D) + \alpha D]$$

- **Drawback:** exponential complexity

- To each sequence  $y^n$  assign energy  $\mathcal{E}(y^n) = H_k(y^n) + \alpha d_n(x^n, y^n)$
- For  $T > 0$ , define a distribution on the space of all possible reconstructions as  $p_T(y^n) \propto e^{-\mathcal{E}(y^n)/T}$
- **Simulated annealing:** gradually decreasing temperature while sampling from the distribution; finds  $\hat{x}^n$  such that

$$\mathcal{E}(\hat{x}^n) \approx \min_{y^n} \mathcal{E}(y^n)$$

### Theorem: [JW08]

Let  $\mathbf{X}$  be a stationary and ergodic source. Then

$$\lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \mathbb{E} \left[ \frac{1}{n} \ell_{\text{LZ}}(\hat{X}_{\alpha, r}^n) + \alpha d_n(X^n, \hat{X}^n) \right] = \min_{D \geq 0} [R(\mathbf{X}, D) + \alpha D].$$

- **Drawback:** no bound on the required number of iterations

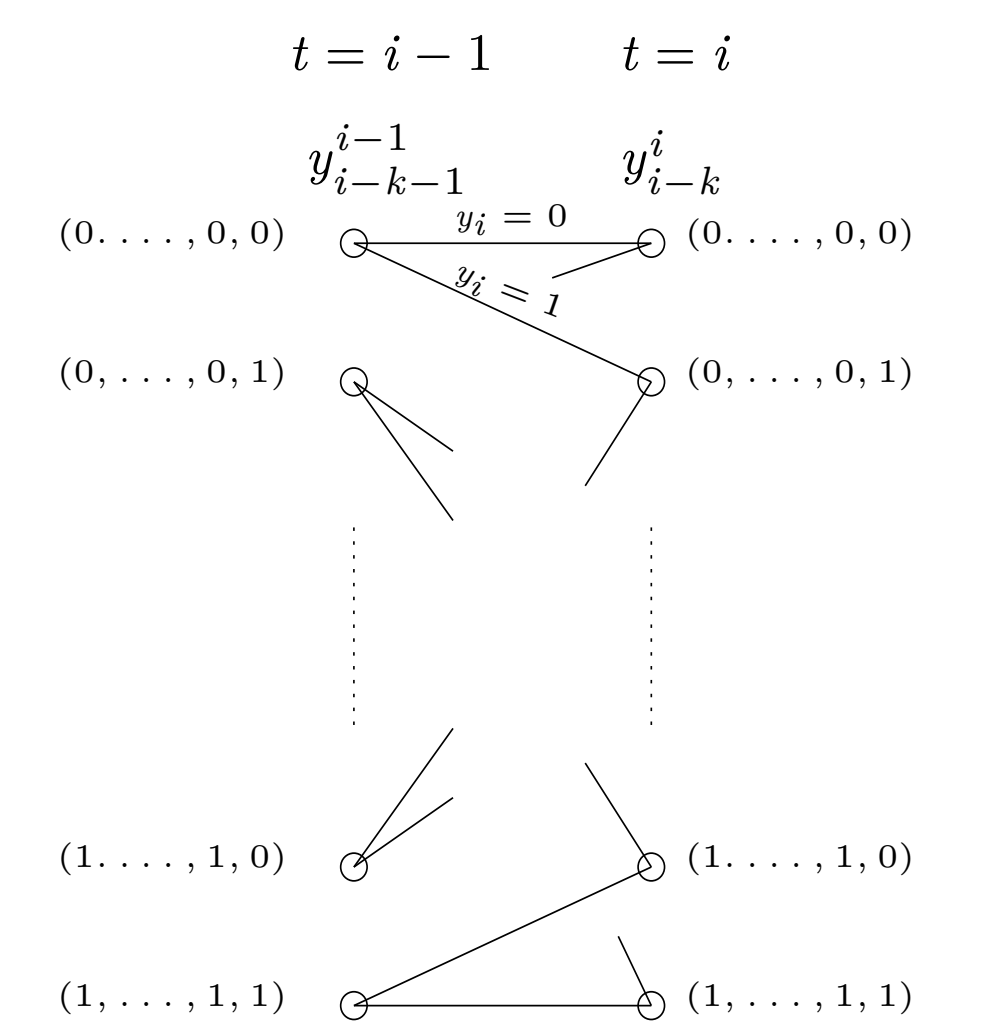
- Let
  - (P1) :  $\min_{y^n} [H_k(y^n) + \alpha d_n(x^n, y^n)]$
  - (P2) :  $\min_{y^n} \left[ \sum_{b_{k+1}, \mathbf{b}} \lambda_{b_{k+1}, \mathbf{b}} m_{b_{k+1}, \mathbf{b}}(y^n) + \alpha d_n(x^n, y^n) \right]$
- Assume that we could find a set of coefficients  $\{\lambda_{b_{k+1}, \mathbf{b}}\}_{b_{k+1}, \mathbf{b}}$  such that (P1) and (P2) have the same minimizers
- Why are we interested in this new representation?
- Note:

$$\begin{aligned} & \sum_{b_{k+1}, \mathbf{b}} [\lambda_{b_{k+1}, \mathbf{b}} m_{b_{k+1}, \mathbf{b}}(y^n) + \alpha d_n(x^n, y^n)] \\ &= \frac{1}{n} \sum_{b_{k+1}, \mathbf{b}} \left[ \lambda_{b_{k+1}, \mathbf{b}} \sum_{i=1}^n \mathbf{1}_{y_{i-k}^{i-1} = [b, b_{k+1}]} + \alpha d(x_i, y_i) \right] \\ &= \frac{1}{n} \sum_{i=1}^n [\lambda_{y_i, y_{i-k}^{i-1}} + \alpha d(x_i, y_i)]. \end{aligned}$$

- (P2) can be solved using **Viterbi algorithm**

- Let the state at time  $i$  be  $\mathbf{s}_i = (y_{i-k}^{i-1}, y_i)$
- One-to-one correspondence between the paths inside the trellis, sequence of states  $\{\mathbf{s}_i\}_{i=1}^n$ , and also sequences  $y^n$
- We want to find the path in the trellis minimizing  $\sum_{i=1}^n w_i(y_{i-k}^{i-1}, y_i)$ , where

$$w_i(\mathbf{s}_i) = \lambda_{y_i, y_{i-k}^{i-1}} + \alpha d(x_i, y_i)$$



Do such coefficients exist?	Practical implementation of the new algorithm	Simulation results	Simulation results
-----------------------------	---	--------------------	--------------------

- **Yes**

### Lemma

(P1) and (P2) have the same minimum values, if the coefficients are chosen according to

$$\lambda_{b_{k+1}, \mathbf{b}} = \frac{\partial}{\partial m_{b_{k+1}, \mathbf{b}}} H(\mathbf{m}) \Big|_{\mathbf{m}_n^*}$$

where  $\mathbf{m}_n^* = \mathbf{m}_n^*(\hat{x}^n)$ , where  $\hat{x}^n$  is a minimizer of (P1)

- We need to approximate the coefficients
- Solving the non-convex optimization problem is hard
- Try simpler approximations for  $\mathbf{m}^*$ , like  $\mathbf{m}(x^n)$ 
  - works well for small distortions
- For larger distortions, we can use iterative approach

- Do we have access to  $\hat{x}^n$ ? Of course not!
- But, we can find an approximate version of  $\mathbf{m}_n^*$  by solving a non-convex optimization problem.

