On Universal Decoding with the Krichevsky–Trofimov Estimator

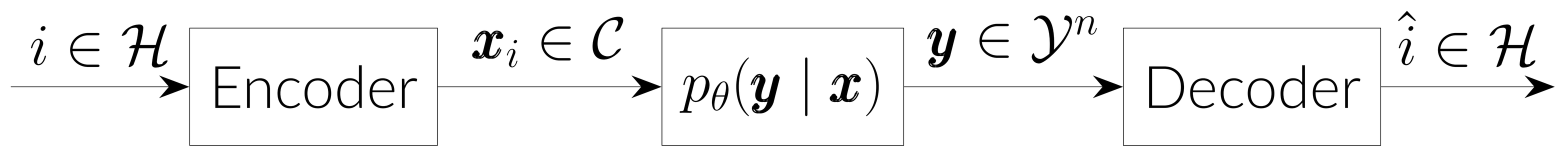
Henrique Miyamoto*, Sheng Yang

Laboratoire des Signaux et Systèmes (L2S) — Université Paris-Saclay, CNRS, CentraleSupélec

henrique.miyamoto@centralesupelec.fr

Context

Setup: a code $\mathcal{C} \coloneqq \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\} \subseteq \mathcal{X}^n$ is used for communication through a channel in the family $\mathcal{F} \coloneqq \{p_\theta : \theta \in \Theta\}$ of DMCs.



Maximum a posteriori (MAP) is the optimal decoding rule:

$$\phi_{\text{MAP}}(\boldsymbol{y}) = \arg\max_{i \in \mathcal{H}} p_\theta(\boldsymbol{x}_i \mid \boldsymbol{y}),$$

and it simplifies to the ML rule for equiprobable messages.

Random coding: codewords are selected independently and uniformly in a set $\mathcal{B}_n \subseteq \mathcal{X}^n$.

Universal decoding

***Definition* [1]:** Let $P_{\theta,\phi}(\text{error})$ denote the average probability of error (over messages and random codes) when decoder ϕ is used in channel θ. A sequence of decoders $(\phi_n)_{n\in\mathbb{N}}$ is said to be random-coding ***(weakly)* universal** for the family $\mathcal{F}$ and $(\mathcal{B}_n)_{n\in\mathbb{N}}$, if

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\frac{P_{\theta,\phi_n}(\text{error})}{P_{\theta,\text{ML}}(\text{error})} \right) = 0, \quad \forall \theta \in \Theta,$$

and random-coding ***strongly* universal**, if

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \frac{1}{n} \log \left(\frac{P_{\theta,\phi_n}(\text{error})}{P_{\theta,\text{ML}}(\text{error})} \right) = 0.$$

Memoryless channels

In a DMC, the posterior is memoryless w.r.t. $\boldsymbol{x}$:

$$p_\theta(\boldsymbol{x} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{x})}{p_\theta(\boldsymbol{y})} p_\theta(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{|\mathcal{B}_n|^{-1}}{p_\theta(\boldsymbol{y})} \prod_{x\in\mathcal{X}} \prod_{y\in\mathcal{Y}} p_\theta(y \mid x)^{a_{\boldsymbol{x}\boldsymbol{y}}(x,y)}$$

$$= \left(\frac{\prod_{y\in\mathcal{Y}} C_\theta(y)^{a_{\boldsymbol{y}}(y)}}{|\mathcal{B}_n| p_\theta(\boldsymbol{y})} \right) \prod_{x\in\mathcal{X}} \prod_{y\in\mathcal{Y}} q_\xi(x \mid y)^{a_{\boldsymbol{x}\boldsymbol{y}}(x,y)},$$

with $q_{\xi(\theta)}(x \mid y) \coloneqq \frac{p_\theta(y|x)}{C_\theta(y)}$ and $C_\theta(y) \coloneqq \sum_{x'\in\mathcal{X}} p_\theta(y \mid x')$.

KT-based decoder

The conditional *Krichevsky–Trofimov* (KT) estimator [2] for sequences $\boldsymbol{x} \coloneqq x_1 \cdots x_n \in \mathcal{X}^n$ and $\boldsymbol{y} \coloneqq y_1 \cdots y_n \in \mathcal{Y}^n$ is

$$p_{\text{KT}}(\boldsymbol{x} \mid \boldsymbol{y}) = \frac{\Gamma\left(\frac{|\mathcal{X}|}{2}\right)^{|\mathcal{Y}|}}{\Gamma\left(\frac{1}{2}\right)^{|\mathcal{X}||\mathcal{Y}|}} \frac{\prod_{x,y} \Gamma\left(a_{\boldsymbol{x}\boldsymbol{y}}(x,y) + \frac{1}{2}\right)}{\prod_{y\in\mathcal{Y}} \Gamma\left(a_{\boldsymbol{y}}(y) + \frac{|\mathcal{X}|}{2}\right)}. \quad (1)$$

- It only depends on the type (conditional distribution) of the sequences:

$$a_{\boldsymbol{x}\boldsymbol{y}}(x,y) = n\pi_{\boldsymbol{x}\boldsymbol{y}}(x,y) = \sum_{i=1}^{n} \mathbb{1}\{(x_i, y_i) = (x,y)\}.$$

- It corresponds to a mixture of i.i.d. distributions $q_\xi(x \mid y)$ with Dirichlet priors $\pi(\xi) \sim \text{Dir}\left(\frac{1}{2}, \cdots, \frac{1}{2}\right)$ (Jeffreys' prior).
- It can be sequentially computed:

$$p_{\text{KT}}(x_1^{n+1} \mid y_1^{n+1}) = \left(\frac{a_{x_1^n, y_1^n}(x_{n+1}, y_{n+1}) + \frac{1}{2}}{a_{y_1^n}(y) + \frac{|\mathcal{X}|}{2}} \right) p_{\text{KT}}(x_1^n \mid y_1^n).$$

- We have:

$$\log \left(\frac{\pi_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x} \mid \boldsymbol{y})}{p_{\text{KT}}(\boldsymbol{x} \mid \boldsymbol{y})} \right) \le \frac{(|\mathcal{X}|-1)|\mathcal{Y}|}{2} \log n + 2|\mathcal{Y}|. \quad (2)$$

***Theorem*:** Let $\mathcal{B}_n = \mathcal{X}^n$ or $\mathcal{B}_n = \mathcal{T}(\pi)$ the type class of some type π on $\mathcal{X}^n$. The decoder

$$\phi_{\text{KT}}(\boldsymbol{y}) = \arg\max_{i\in\mathcal{H}} p_{\text{KT}}(\boldsymbol{x}_i \mid \boldsymbol{y}) \quad (3)$$

is random-coding strongly universal for the family of DMCs.

Proof: Denote $\mathcal{E}_\phi(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \{\boldsymbol{x}' \in \mathcal{B}_n : p_\phi(\boldsymbol{x}' \mid \boldsymbol{y}) \ge p_\phi(\boldsymbol{x} \mid \boldsymbol{y})\}$.

$$\frac{P_{\theta,\text{KT}}(\text{error})}{P_{\theta,\text{ML}}(\text{error})} \overset{\text{(a)}}{\le} \max_{\boldsymbol{x},\boldsymbol{y}} \frac{|\mathcal{E}_{\text{KT}}(\boldsymbol{x},\boldsymbol{y})|}{|\mathcal{E}_{\text{ML}}(\boldsymbol{x},\boldsymbol{y})|} \overset{\text{(b)}}{\le} 2^{n\delta(n)} \left(\max_{\boldsymbol{x},\boldsymbol{y}} \frac{\pi_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x} \mid \boldsymbol{y})}{p_{\text{KT}}(\boldsymbol{x} \mid \boldsymbol{y})} \right) \overset{\text{(c)}}{\le} 2^{n\delta(n)} 2^{n\epsilon(n)},$$

where (a) follows from [1, Eq. (25)], (b) follows from [3, Lem. 11] and Markov inequality, and (c) follows from (2), with $\delta(n) \to 0$, $\epsilon(n) \to 0$. ■

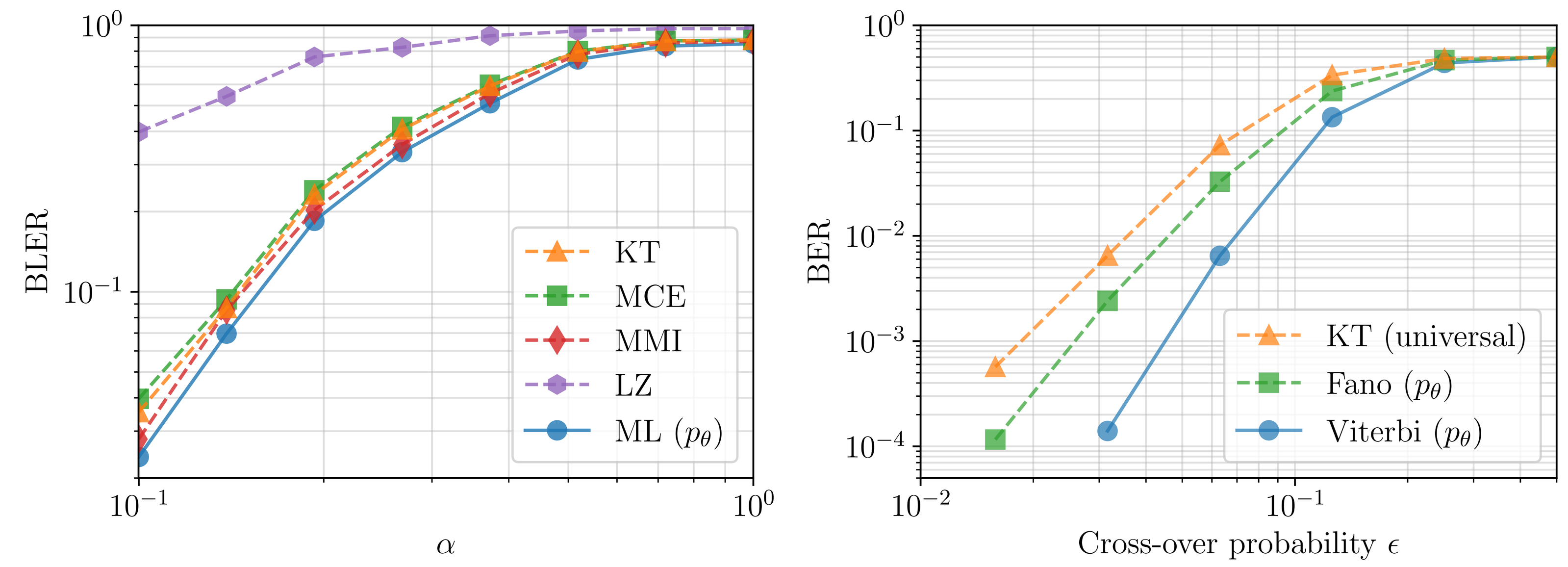
Decoding schemes with practical codes

1. (k,n)**-linear block codes** have to be adapted:

- The all-zero codeword is included, maximising (1) regardless of $\boldsymbol{y}$ → all codewords are shifted by a known amount.
- Antipodal codewords are assigned the same metric → they are identified, reducing the effective rate from $\frac{k}{n}$ to $\frac{k-1}{n}$.

2. For **convolutional codes**, use with a modified *stack decoder* [4]: keep a stack of searched paths in the trellis; at each step, extend the one with highest metric, until the end is reached. The Fano metric of a path from node i to j is replaced by

$$M_{\text{KT}}(\boldsymbol{x}_i^j, \boldsymbol{y}_i^j) \coloneqq \log p_{\text{KT}}(\boldsymbol{x}_i^j \mid \boldsymbol{y}_i^j) + l(j-i+1)\left[\log|\mathcal{X}| - (R+\Delta)\right].$$



(a) Golay code ($n = 24$) in binary channels. Cross-over probabilities are drawn uniformly in $[0, \alpha]$.

(b) Convolutional code of rate 1/2 and constraint length 7 in BSC with crossover probability $\epsilon \in [0, 1]$.

References

[1] M. Feder and A. Lapidoth. Universal decoding for channels with memory. *IEEE Trans. Inf. Theory*, vol. 44, pp. 1726–1745, 1998.

[2] R. Krichevsky and V. Trofimov. The performance of universal encoding. *IEEE Trans. Inf. Theory*, vol. 27, pp. 199–207, 1981.

[3] J. Ziv. Universal decoding for finite-state channels. *IEEE Trans. Inf. Theory*, vol. 31, pp. 453–460, 1985.

[4] A. Lapidoth and J. Ziv. On the decoding of convolutional codes on an unknown channel. *IEEE Trans. Inf. Theory*, vol. 45, pp. 2321–2332, 1999.

L2S Laboratoire Signaux & Systèmes — CentraleSupélec — université PARIS-SACLAY