Universal Decoding for Arbitrary Channels Relative to a Given Class of Decoding Metrics

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 32000, Israel

IZS 2014, Zurich, Switzerland, February 26–28, 2014.

Some Background on Universal Decoding

- Memoryless channels:
  - Goppa (1975) – MMI decoder achieves capacity.
  - Csiszár & Körner (1981) – MMI achieves random coding exponent.
  - Csiszár (1982) – minimum entropy decoder for linear codes.
  - Merhav (1993) – similar results for memoryless Gaussian channels.
- Channels with memory:
  - Ziv (1985) – LZ-based metric for unifilar finite–state channels.
  - Lapidoth & Ziv (1998) – extension to HMM channels.
  - Feder & Lapidoth (1998) – merged decoder.
  - Feder & Merhav (2002) – competitive minimax approach.

Some Background on Universal Decoding (Cont'd)

- Deterministic arbitrary channels ("individual" channels):
  - Lomnitz & Feder (2012) – empirical rate functions.
  - Misra & Weissman (2012) – porosity of additive noise channels.
  - Shayevtiz & Feder (2012) – binary additive channels with feedback.
  - Elkayam & Feder (2014) – following and very related to this work.

System Model and Problem Definition

- A rate–R code $\mathcal{C}$ selected at random.
- The marginal of each codeword $\boldsymbol{x}_i \in \mathcal{X}^n$ is Q.
- The channel $P(\boldsymbol{y}|\boldsymbol{x})$ is arbitrary and unknown (may be deterministic).
- We are given a class of decoding metrics $\mathcal{M} = \{m_\theta(\boldsymbol{x}, \boldsymbol{y}),\ \theta \in \Theta\}$.
- Decoder $\mathcal{D}_\theta$ picks the message i with highest $m_\theta(\boldsymbol{x}_i, \boldsymbol{y})$.
- $\overline{P_{e,\theta}(R,n)} \stackrel{\triangle}{=}$ average error probability of $\mathcal{D}_\theta$.
- We seek a universal decoding metric $u(\boldsymbol{x}, \boldsymbol{y})$ with

$$\overline{P_{e,u}(R,n)} \stackrel{\cdot}{\leq} \min_{\theta \in \Theta} \overline{P_{e,\theta}(R,n)}$$

for every channel $P(\boldsymbol{y}|\boldsymbol{x})$.

The Proposed Univeral Decoding Metric

For the given class $\mathcal{M}$ of decoding metrics, define

$$\mathcal{T}(\boldsymbol{x}|\boldsymbol{y}) \stackrel{\triangle}{=} \left\{\boldsymbol{x}' : \ \forall\theta \in \Theta \ \ m_\theta(\boldsymbol{x}', \boldsymbol{y}) = m_\theta(\boldsymbol{x}, \boldsymbol{y})\right\}.$$

Our universal decoding metric is defined as

$$u(\boldsymbol{x}, \boldsymbol{y}) \stackrel{\triangle}{=} -\frac{1}{n} \log Q[\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})].$$

For a given $\boldsymbol{y}$, $\{\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})\}$ are equivalence classes that partition $\mathcal{X}^n$. Define

$$K_n(\boldsymbol{y}) \stackrel{\triangle}{=} \text{number of distinct } \{\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})\} \text{ for a given } \boldsymbol{y}$$

$$\Delta_n \stackrel{\triangle}{=} \frac{\max_{\boldsymbol{y}} \log K_n(\boldsymbol{y})}{n}.$$

Δ_n is a measure for the richness of the class of metrics $\mathcal{M}$.

Basic Result and Discussion

Theorem: Let the randomly selected codewords in $\mathcal{C}$ be conditionally pairwise independent. Then,

$$\overline{P_{e,u}(R,n)} \le 2 \cdot e^{n\Delta_n} \cdot \min_{\theta\in\Theta} \overline{P_{e,\theta}(R,n)}$$

and

$$\overline{P_{e,u}(R,n)} \le 2 \cdot \min_{\theta\in\Theta} \overline{P_{e,\theta}(R+\Delta_n,n)}.$$

Discussion:

- $u(\boldsymbol{x},\boldsymbol{y})$ has a competitive error exponent if $\Delta_n \to 0$.
- The class $\mathcal{M}$ should not be too rich.
- In general, $\Delta = \lim_{n\to\infty} \Delta_n$ is the rate loss and the loss in error exponent.

Example

- $Q =$ uniform distribution within a single type class T_Q.
- $\mathcal{M}$ is the class of metrics of the form

$$m_\theta(\boldsymbol{x}, \boldsymbol{y}) = \sum_{t=1}^{n} \theta(x_t, y_t).$$

In this case, $\mathcal{T}(\boldsymbol{x}|\boldsymbol{y}) = T_{\boldsymbol{x}|\boldsymbol{y}}$, the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$. Thus,

$$\begin{aligned} u(\boldsymbol{x}, \boldsymbol{y}) &= -\frac{1}{n} \log Q[T_{\boldsymbol{x}|\boldsymbol{y}}] = -\frac{1}{n} \log[Q(\boldsymbol{x}) \cdot |T_{\boldsymbol{x}|\boldsymbol{y}}|] \\ &= \hat{H}_{\boldsymbol{x}}(X) - \hat{H}_{\boldsymbol{x}\boldsymbol{y}}(X|Y) + o(n) \approx \hat{I}_{\boldsymbol{x}\boldsymbol{y}}(X;Y), \end{aligned}$$

which is the MMI decoder. Here, $\Delta_n = O(\log n/n)$.

If Q is i.i.d.

$$u(\boldsymbol{x}, \boldsymbol{y}) = \hat{I}_{\boldsymbol{x}\boldsymbol{y}}(X;Y) + D(\hat{P}_{\boldsymbol{x}} \| Q).$$

Comparison with Elkayam & Feder (2014)

Elkayam & Feder propose a different universal metric:

$$\tilde{u}(\boldsymbol{x},\boldsymbol{y}) = -\frac{1}{n}\log\min_{\theta\in\Theta} Q\{m_\theta(\boldsymbol{X},\boldsymbol{y}) \geq m_\theta(\boldsymbol{x},\boldsymbol{y})\},$$

which satisfies the same theorem, provided that

$$\limsup_{n\to\infty}\frac{1}{n}\log\left(\max_{\boldsymbol{y}}\mathbf{E}_Q\{\exp[n\tilde{u}(\boldsymbol{X},\boldsymbol{y})]\}\right) = 0.$$

Plus: This condition of univerality is weaker than ours.

Minuses:

- The error exponent of $\tilde{u}$ cannot be better than that of u.
- Difficult to implement (even for the above example, which is elementary).
- The above condition is difficult to verify.

Useful Approximations of $u(\boldsymbol{x}, \boldsymbol{y})$

- For an explicit expression of $u(\boldsymbol{x}, \boldsymbol{y})$ – need to assess $Q[\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})]$.
- If Q is invariant within $\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})$, then $Q[\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})] = Q(\boldsymbol{x}) \cdot |\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})|$.
- In many cases, we can assess $|\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})|$ (method of types, stat. mech.,..).
- In other cases, it is difficult, but some approximatons might help.

Theorem: Suppose that $Q[\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})] = e^{-nu(\boldsymbol{x},\boldsymbol{y})}$ can be lower bounded by $e^{-nu'(\boldsymbol{x},\boldsymbol{y})}$ such that

$$\max_{\boldsymbol{y}} \mathbf{E}_Q\{\exp_2[nu'(\boldsymbol{X}, \boldsymbol{y})]\} \doteq 1.$$

Then, our earlier theorem applies to u' as well.

Example – Finite–State Decoding Metrics

Given $\boldsymbol{x}$ and $\boldsymbol{y}$, and $g : \mathcal{X} \times \mathcal{Y} \times \mathcal{S} \to \mathcal{S}$, consider the evolution of a finite–state machine $s_{t+1} = g(x_t, y_t, s_t)$, $t = 1, 2, \ldots, n-1$, and let $\mathcal{M}$ be the class of metrics

$$m_\theta(\boldsymbol{x}, \boldsymbol{y}) = \sum_{t=1}^{n} \theta(x_t, y_t, s_t).$$

Here, there is no simple expression for $|\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})|$ (even if g is known), but [Ziv 1985]:

$$|\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})| \geq e^{LZ(\boldsymbol{x}|\boldsymbol{y}) - o(n)}$$

and so, for $Q(\boldsymbol{x}) = |\mathcal{X}|^{-n}$, one can take

$$u'(\boldsymbol{x}, \boldsymbol{y}) = \log |\mathcal{X}| - \frac{LZ(\boldsymbol{x}|\boldsymbol{y})}{n},$$

which satisfies the condition since the LZ code satisfies Kraft's inequality.

Extension 1 – Feedback

In the presence of feedback, $Q(\boldsymbol{x})$ can be replaced by

$$Q(\boldsymbol{x}|\boldsymbol{y}) = \prod_{t=1}^{n} Q(x_t|x^{t-1}, y^{t-1})$$

and the results extend straightforwardly with $u(\boldsymbol{x}, \boldsymbol{y})$ being redefined as

$$u(\boldsymbol{x}, \boldsymbol{y}) = -\frac{1}{n} \log Q[\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})|\boldsymbol{y}].$$

For example, the LZ decoding metric would generalize (under certain conditions) to

$$u'(\boldsymbol{x}, \boldsymbol{y}) = -\frac{1}{n}[\log Q(\boldsymbol{x}|\boldsymbol{y}) + LZ(\boldsymbol{x}|\boldsymbol{y})].$$

Extension 2 – MAC

For a certain class of MAC's (e.g., $P(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) = W(\boldsymbol{y}|\boldsymbol{x}_1 \oplus \boldsymbol{x}_2)$), the following extension applies: Let $\mathcal{M} = \{m_\theta(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}),\ \theta \in \Theta\}$ be given and define

$$\begin{aligned}
\mathcal{T}(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{y}) &= \left\{(\boldsymbol{x}_1', \boldsymbol{x}_2') :\ \forall \theta \in \Theta\ m_\theta(\boldsymbol{x}_1', \boldsymbol{x}_2', \boldsymbol{y}) = m_\theta(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})\right\} \\
\mathcal{T}(\boldsymbol{x}_1|\boldsymbol{x}_2, \boldsymbol{y}) &= \left\{\boldsymbol{x}_1' :\ \forall \theta \in \Theta\ m_\theta(\boldsymbol{x}_1', \boldsymbol{x}_2, \boldsymbol{y}) = m_\theta(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})\right\}
\end{aligned}$$

and similarly $\mathcal{T}(\boldsymbol{x}_2|\boldsymbol{x}_1, \boldsymbol{y})$. Now, let

$$\begin{aligned}
u_0(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) &= -\frac{1}{n} \log\{(Q_1 \times Q_2)[\mathcal{T}(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{y})]\} \\
u_1(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) &= -\frac{1}{n} \log Q_1[\mathcal{T}(\boldsymbol{x}_1|\boldsymbol{x}_2, \boldsymbol{y})]
\end{aligned}$$

and similarly $u_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$. Finally, our decoding metric is:

$$u(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) = \min\{u_0 - R_1 - R_2,\ u_1 - R_1, u_2 - R_2\}.$$

Extension 3 – Continuous Alphabet Case

Our results can in principle be modified to the case $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ (with some caution): For example, let Q be zero–mean, Gaussian i.i.d. with variance σ^2, and let

$$m_\theta(\boldsymbol{x}, \boldsymbol{y}) = \theta_1 \sum_{t=1}^{n} x_t^2 + \theta_2 \sum_{t=1}^{n} x_t y_t.$$

Here we need to assess the volume of $\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})$, the set of all $\boldsymbol{x}'$ with (approximately) the same empirical variance and empirical correlation with $\boldsymbol{y}$ as that of $\boldsymbol{x}$. The resulting metric is:

$$u(\boldsymbol{x}, \boldsymbol{y}) = \frac{S(\boldsymbol{x})}{2\sigma^2} - \frac{1}{2} \ln[S(\boldsymbol{x})(1 - \rho^2_{\boldsymbol{x}\boldsymbol{y}})],$$

where

$$S(\boldsymbol{x}) = \frac{1}{n} \sum_{t=1}^{n} x_t^2, \quad \rho_{\boldsymbol{x}\boldsymbol{y}} = \frac{\frac{1}{n} \sum_{t=1}^{n} x_t y_t}{\sqrt{S(\boldsymbol{x}) S(\boldsymbol{y})}}.$$

Summary and Conclusion

- We have defined a general framework. Earlier results are special cases.
- If $\mathcal{M}$ is a singleton, this is mismatched decoding.
- Deterministic channels (“individual” channels) are included.
- The proof technique is simple and easy to extend.
- Implenentability relies on an expression of $|\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})|$ or a lower bound.