Universal Decoding Using a Noisy Codebook

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering

Technion—Israel Institute of Technology

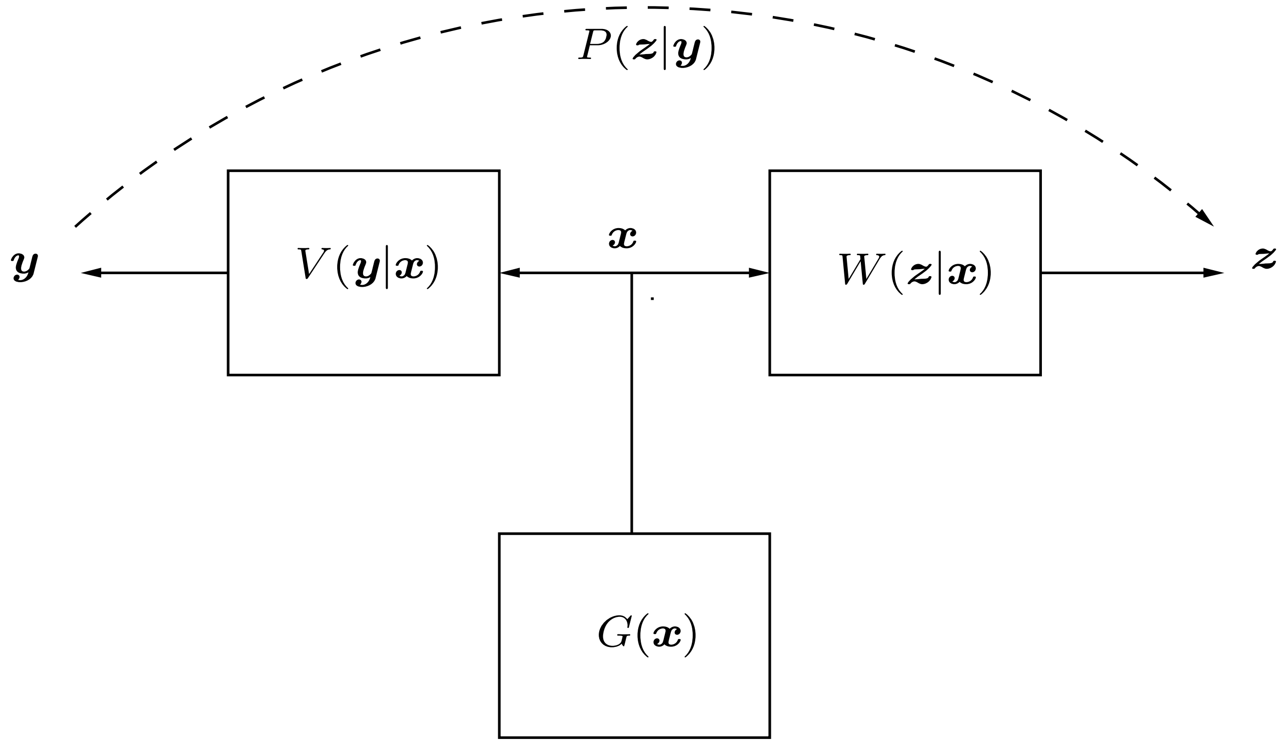
Haifa 3200004, Israel

ISIT 2017, Aachen, Germany, June 2017.

Motivation and Background

- Ziv (1985): universal decoding for unifilar FS channels.
- Lapidoth & Ziv (1998): extension to “hidden Markov” channels.
- In both: universal decoder achieves optimal random coding exponents.
- Random coding distribution = uniform.
- This work: further extension to noisy codewords.

Communication System with Noisy Codewords

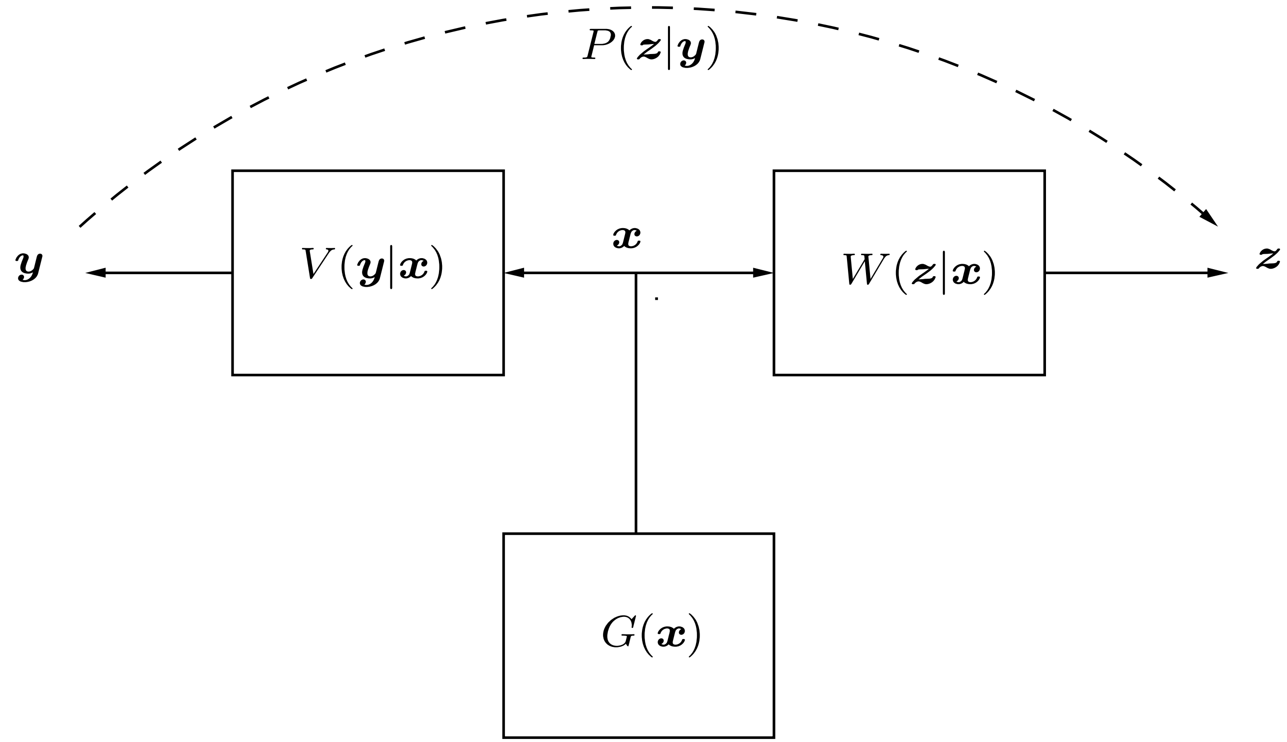


$$G(\boldsymbol{x}) = \sum_{\boldsymbol{\omega}} \prod_{i=1}^{n} G(x_i, \omega_i | \omega_{i-1})$$

$$W(\boldsymbol{z}|\boldsymbol{x}) = \sum_{\boldsymbol{\sigma}} \prod_{i=1}^{n} W(z_i, \sigma_i | x_i, \sigma_{i-1})$$

$$V(\boldsymbol{y}|\boldsymbol{x}) = \sum_{\boldsymbol{\theta}} \prod_{i=1}^{n} V(y_i, \theta_i | x_i, \theta_{i-1})$$

Communication System with Noisy Codewords



- Non-uniform effective random coding dist. –$P(\boldsymbol{y}) = \sum_{\boldsymbol{x}} G(\boldsymbol{x})V(\boldsymbol{y}|\boldsymbol{x})$.
- Effective channel – $P(\boldsymbol{z}|\boldsymbol{y}) = \sum_{\boldsymbol{x}} G(\boldsymbol{x})V(\boldsymbol{y}|\boldsymbol{x})W(\boldsymbol{z}|\boldsymbol{x})$ – not FS.
- A rate-R “codebook” of $\boldsymbol{y}$–vectors generated by corresponding $\boldsymbol{x}$’s.
- Motivation: biometric identification (enrollment vs. authentication).
- Code randomness – part of the model.
- Assumption: $\{P(\boldsymbol{y})\}$ is known (estimable from the noisy codebook).

Universal Decoding Metric – Conditional LZ Parsing

Given $(\boldsymbol{y}, \boldsymbol{z}) = [(y_1, z_1), \ldots, (y_n, z_n)]$, apply LZ parsing to this sequence pair.

- $c(\boldsymbol{y}, \boldsymbol{z}) =$ number of phrases.
- $c(\boldsymbol{z}) =$ number of distinct phrases of $\boldsymbol{u}$.
- $\boldsymbol{z}(l) =$ the lth distinct $\boldsymbol{u}$–phrase, $l = 1, 2, ..., c(\boldsymbol{z})$.
- $c_l(\boldsymbol{y}|\boldsymbol{z}) =$ number of $\boldsymbol{z}(l)$ in parsing of $\boldsymbol{z}$.

$$\text{conditional “code length”} = \sum_{l=1}^{c(\boldsymbol{z})} c_l(\boldsymbol{y}|\boldsymbol{z}) \log c_l(\boldsymbol{y}|\boldsymbol{z}).$$

For example, $n = 6$ and

$$\begin{aligned} \boldsymbol{y} &= 0 \mid 1 \mid 0\,0 \mid 0\,1| \\ \boldsymbol{z} &= 0 \mid 1 \mid 0\,1 \mid 0\,1| \end{aligned}$$

then

$$c(\boldsymbol{y}, \boldsymbol{z}) = 4, \;\; c(\boldsymbol{z}) = 3, \;\; \boldsymbol{z}(1) = 0, \; \boldsymbol{z}(2) = 1, \; \boldsymbol{z}(3) = 01,$$

$$c_1(\boldsymbol{y}|\boldsymbol{z}) = c_2(\boldsymbol{y}|\boldsymbol{z}) = 1, \;\; c_3(\boldsymbol{y}|\boldsymbol{z}) = 2.$$

Universal Decoding Metric

Consider the decoding metric

$$u(\boldsymbol{y}, \boldsymbol{z}) = \log P(\boldsymbol{y}) + \underbrace{\sum_{l=1}^{c(\boldsymbol{z})} c_l(\boldsymbol{y}|\boldsymbol{z}) \log c_l(\boldsymbol{y}|\boldsymbol{z})}_{\text{Ziv's metric}}$$

The proposed decoder is

$$\hat{m} = \text{argmin}_m u(\boldsymbol{y}_m, \boldsymbol{z}).$$

Discussion: A kind of a generalized MMI decoder:

$$\hat{m} = \text{argmax}_m \left\{ \underbrace{\frac{1}{n} \log \left[\frac{1}{P(\boldsymbol{y}_m)} \right]}_{\hat{H}(\boldsymbol{y}_m)} - \underbrace{\frac{1}{n} \sum_{l=1}^{c(\boldsymbol{z})} c_l(\boldsymbol{y}|\boldsymbol{z}) \log c_l(\boldsymbol{y}|\boldsymbol{z})}_{\hat{H}(\boldsymbol{y}_m|\boldsymbol{z})} \right\}$$

Main Result

Theorem. Under certain assumptions, the universal decoder $\hat{m}$ satisfies

$$\overline{\mathsf{P}}_{\mathsf{e,u}} \leq e^{n\epsilon(n)} \overline{\mathsf{P}}_{\mathsf{e,o}},$$

where

$$\overline{\mathsf{P}}_{\mathsf{e,u}} = \text{expected error probability of } \hat{m},$$

$$\overline{\mathsf{P}}_{\mathsf{e,o}} = \text{expected error probability of the ML decoder},$$

and

$$\epsilon(n) = O\left(\frac{\log \log n}{\log n}\right),$$

with a leading term that is linear in $\log |\mathcal{Y} \times \mathcal{Z}|$.

Technical assuption: The hidden Markov source $P(\boldsymbol{y})$ has strictly positive transition probabilities, $\{p(y, \omega, \theta|\omega', \theta')\}$.

Comments About the Proof

The proof follows essentially the same lines of thought as in [Ziv85] and [Lapidoth-Ziv98], but there are two main extra difficulties:

- The effective channel, $\{P(\boldsymbol{z}|\boldsymbol{y})\}$, is not finite–state, in general:

$$
\begin{aligned}
P(\boldsymbol{z}|\boldsymbol{y}) &= \frac{P(\boldsymbol{y},\boldsymbol{z})}{P(\boldsymbol{y})} \\
&= \frac{\sum_{\boldsymbol{x}} G(\boldsymbol{x})V(\boldsymbol{y}|\boldsymbol{x})W(\boldsymbol{z}|\boldsymbol{x})}{\sum_{\boldsymbol{x}} G(\boldsymbol{x})V(\boldsymbol{y}|\boldsymbol{x})}
\end{aligned}
$$

- The random coding distribution, $\{P(\boldsymbol{y})\}$ is not uniform, in general (HMM).

Not a FS Channel

- While $P(\boldsymbol{z}|\boldsymbol{y})$ is not a FS channel, as said,

$$P(\boldsymbol{z}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}, \boldsymbol{z})}{P(\boldsymbol{y})}$$

  and $P(\boldsymbol{y}, \boldsymbol{z})$ and $P(\boldsymbol{y})$ are both finite–state (hidden Markov).

- Approximating both $P(\boldsymbol{y}, \boldsymbol{z})$ and $P(\boldsymbol{y})$ as product measures using their (possibly different) dominant state sequences.
- Applying combinatorial arguments (like the method of types) of permuting phrases w.r.t. both state sequences.

Not a Uniform Random Coding Distribution

- In [Lapidoth–Ziv98], there is a technical lemma, where one of the key steps is the calculation of

$$\sum_{\boldsymbol{y}} \frac{1}{M_{\mathsf{o}}(\boldsymbol{y}, \boldsymbol{z})},$$

where $M_{\mathsf{o}}(\boldsymbol{y}, \boldsymbol{z})$ is the rank of $\boldsymbol{y}$ according to $\{P(\boldsymbol{z}|\boldsymbol{y})\}$.

- Since $\{M_{\mathsf{o}}(\boldsymbol{y}, \boldsymbol{z})\}$ take on the values $1, 2, \ldots$ in some order, then

$$\sum_{\boldsymbol{y}} \frac{1}{M_{\mathsf{o}}(\boldsymbol{y}, \boldsymbol{z})} = \sum_{i=1}^{N} \frac{1}{i} \leq 1 + \ln N.$$

Not a Uniform Distribution (Cont'd)

- Here, with a non–uniform input, the above quanity is generalized to

$$\sum_{\boldsymbol{y}} \frac{P(\boldsymbol{y})}{P[\mathcal{E}_{\mathsf{o}}(\boldsymbol{y}, \boldsymbol{z})]},$$

  where

$$\mathcal{E}_0(\boldsymbol{y}, \boldsymbol{z}) = \{\boldsymbol{y}' : \ M_{\mathsf{o}}(\boldsymbol{y}', \boldsymbol{z}) \le M_{\mathsf{o}}(\boldsymbol{y}, \boldsymbol{z})\}.$$

- Fortunately enough, there is a relatively simple trick to bound this quantity by a sub–exponential quantity in a rather general manner.

Since There is Still a Minute, Here is the Trick:

Let $a_i \geq 0$ and $A_i = \sum_{j=1}^{i} a_j$. Then,

$$
\begin{aligned}
\sum_{i=1}^{N} \frac{a_i}{A_i} &= 1 + \sum_{i=2}^{N} \frac{a_i}{A_{i-1} + a_i} \\
&= 1 + \sum_{i=2}^{N} \frac{a_i/A_{i-1}}{1 + a_i/A_{i-1}} \\
&\leq 1 + \sum_{i=2}^{N} \ln\left(1 + \frac{a_i}{A_{i-1}}\right) \\
&= 1 + \sum_{i=2}^{N} \ln\left(\frac{A_{i-1} + a_i}{A_{i-1}}\right) \\
&= 1 + \sum_{i=2}^{N} \ln\left(\frac{A_i}{A_{i-1}}\right) \\
&= 1 + \ln\left(\frac{A_N}{A_1}\right)
\end{aligned}
$$