Reliability of Universal Decoding Based on Vector–Quantized Codewords

Neri Merhav

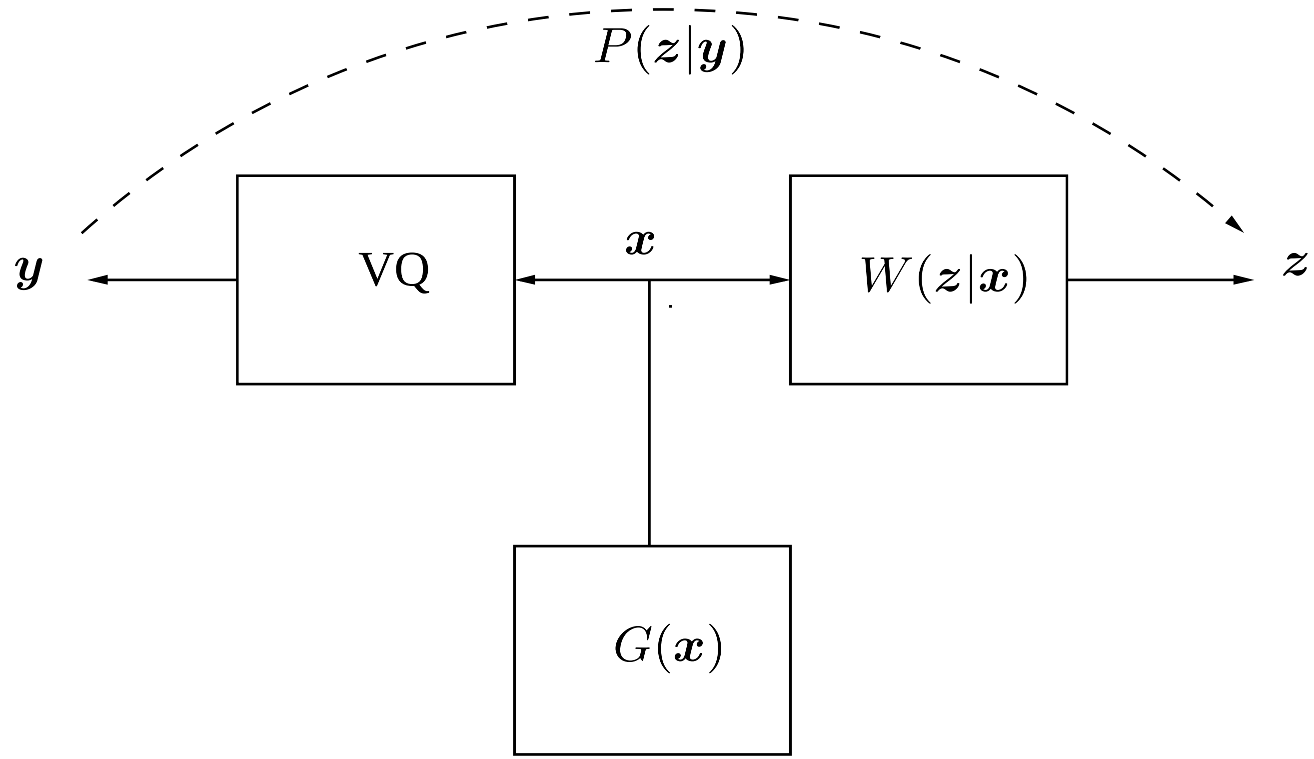
The Andrew & Erna Viterbi Faculty of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 3200004, Israel

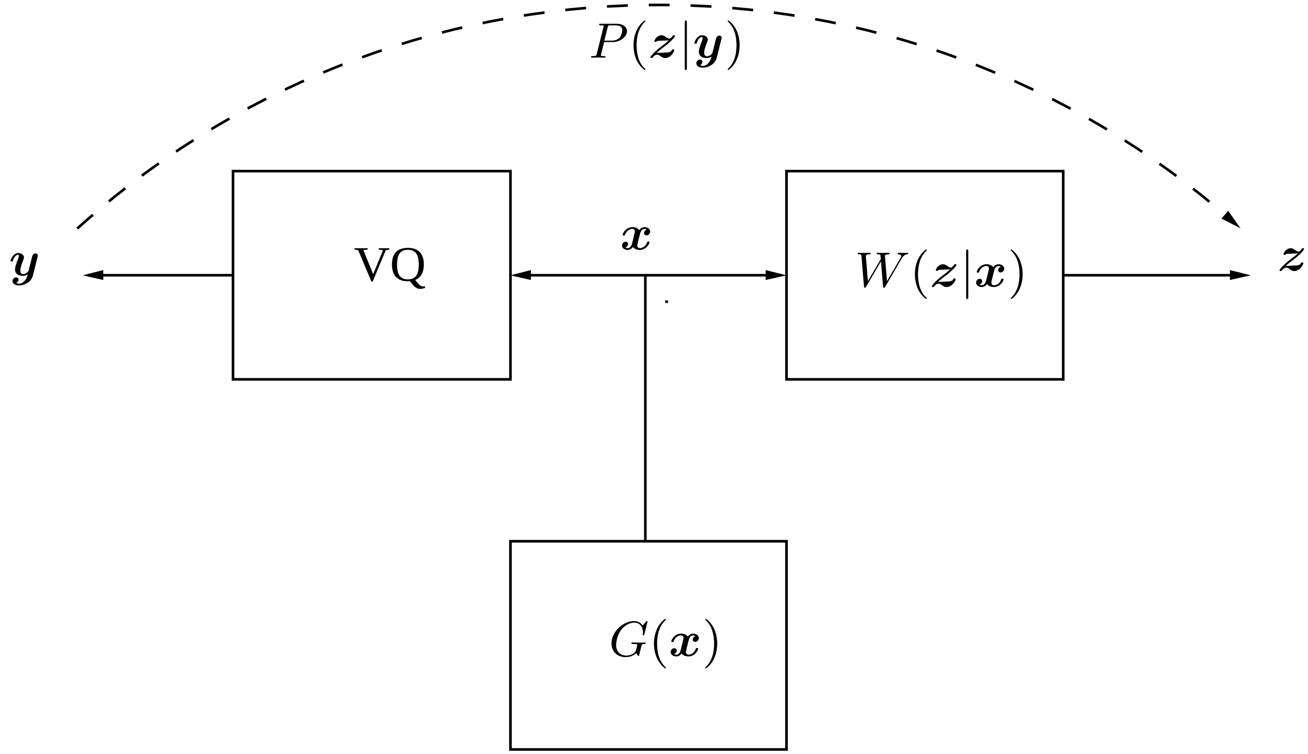
ISIT 2017, Aachen, Germany, June 2017.

Communication System with Quantized Codewords



$$
\begin{aligned}
G(\boldsymbol{x}) &= \prod_{i=1}^{n} G(x_i) \\
W(\boldsymbol{z}|\boldsymbol{x}) &= \prod_{i=1}^{n} W(z_i|x_i) \\
\boldsymbol{y} &= f(\boldsymbol{x})
\end{aligned}
$$

Communication System with Quantized Codewords



- Rate-R_c "codebook" of $\boldsymbol{y}$'s, quantized versions of corresponding $\boldsymbol{x}$'s.
- Motivation: biometric identification (enrollment vs. authentication).
- Objectives: ensemble performance; universal decoding.
- Dasarthy & Draper (2011): MMI decoder. Can we improve? Yes!
- Difficulty: the effective channel, $\{P(\boldsymbol{z}|\boldsymbol{y})\}$, is complicated:

$$P(\boldsymbol{z}|\boldsymbol{y}_m) = \frac{P(\boldsymbol{y}_m, \boldsymbol{z})}{P(\boldsymbol{y}_m)} = \frac{\sum_{\boldsymbol{x}} G(\boldsymbol{x}) W(\boldsymbol{z}|\boldsymbol{x}) \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{y}_m\}}{\sum_{\boldsymbol{x}} G(\boldsymbol{x}) \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{y}_m\}}$$

Main Contributions of This Work

- Exponentially tight bound on the ensemble performance.
- Comparison with Dasarathy & Draper (2011).
- Universal decoder a.g.a. ML decoder ($\forall\ x, z:\ W(z|x) > 0$).
- Also a.g.a. any decoder that depends on joint empirical statistics ($\forall\ W$).
- A good approximation to the channel $\{P(\boldsymbol{z}|\boldsymbol{y})\}$.

Ensemble of Vector–Quanitzers

- $\forall$ input type, Q_X, choose $Q_{Y|X}$ (s.t. compression constraints).
- Randomly draw e^{nR_Q} vectors from $\mathcal{T}(Q_Y)$, with $R_Q = I_Q(X;Y) + \Delta$.
- Randomly rank all members of every $\mathcal{T}(Q_{Y|X}|\boldsymbol{x})$.
- Let $M(\boldsymbol{x},\boldsymbol{y}) =$ rank of $\boldsymbol{y} \in \mathcal{T}(Q_{Y|X}|\boldsymbol{x})$.
- Code ensemble: random codebook + random rank function.
- Quantize $\boldsymbol{x}$ to $\boldsymbol{y} \in \mathcal{T}(Q_{Y|X}|\boldsymbol{x}) \cap$ code with the smallest $M(\boldsymbol{x},\boldsymbol{y})$.

Examples of Compression Constraints

- Expected length: $\mathbf{E}\{L(\boldsymbol{Y})\} \leq nR_{\mathsf{C}}$.
- Excess–length probability: $\Pr\{L(\boldsymbol{Y}) \geq nR_{\mathsf{C}}\} \leq e^{-nE_{\mathsf{C}}}$ for a given $E_{\mathsf{C}} > 0$.
- Exponential moment: $\mathbf{E}\{\exp[sL(\boldsymbol{Y})]\} \leq e^{n\Lambda}$ for given $s > 0$ and $\Lambda > 0$.

Why Not Ordinary MMI Decoding?

- Even without VQ, MMI is best only for random fixed composition codes.
- When $\boldsymbol{x}_m \sim G$ (i.i.d.), better use MMI metric + $D(\hat{P}_{\boldsymbol{x}_m} \| G)$ (prior info).
- Without VQ, the term $\hat{H}_{\boldsymbol{y}_m \boldsymbol{z}}(Y|Z)$ of MMI comes from $|\mathcal{T}(Q_{Y|Z}|\boldsymbol{z})|$.
- But with VQ, not all members of $\mathcal{T}(Q_{Y|Z}|\boldsymbol{z})$ are in the VQ codebook!

A Modified MMI Decoder

For most codes in the ensemble, we can approximate

$$P(\boldsymbol{y}_m) = \sum_{\boldsymbol{x}} G(\boldsymbol{x}) \cdot \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{y}_m\} \doteq \exp\{-n\alpha(\hat{P}_{\boldsymbol{y}_m})\},$$

where $\alpha(\cdot)$ has a certain single–letter formula.

The proposed modified MMI decoder is of the form

$$\hat{m} = \text{argmin}_m \left\{ \log N(\boldsymbol{y}_m|\boldsymbol{z}) - n\alpha(\hat{P}_{\boldsymbol{y}_m}) \right\},$$

where

$$N(\boldsymbol{y}_m|\boldsymbol{z}) = \left| \mathcal{T}(\boldsymbol{y}_m|\boldsymbol{z}) \cap \mathcal{C} \right|,$$

$\mathcal{C}$ being the VQ code.

Main Theorem

For a given choice of $Q_{Y|X}$ as a functional of Q_X:

(a) the random coding error exponent is given by

$$E(R_I) = \min_{Q_X} \min_{Q_{Z|Y}} \left\{ D(Q_X \| G) + \min_{\tilde{Q}_{X|YZ} \in \mathcal{U}(Q_{X|Y})} D(\tilde{Q}_{XZ|Y} \| Q_{X|Y} \times W | Q_Y) + \right.$$
$$\left. + \max\{[I_Q(Y;Z) - I_Q(X;Y)]_+, [I_Q(Y;Z) + D(Q_X \| G) - R_I]_+\} \right\},$$

where R_I is the identification rate, for a given Q_{YZ}, the set $\mathcal{U}(Q_{X|Y})$ is defined to consist of all $\{\tilde{Q}_{X|YZ}\}$ s.t. $\sum_{z \in \mathcal{Z}} \tilde{Q}_{X|YZ}(x|y,z) Q_{Z|Y}(z|y) = Q_{X|Y}(x|y)$ for every $(x,y) \in \mathcal{X} \times \mathcal{Y}$.

(b) Assuming that all $W(z|x) > 0$, the error exponent of the ML decoder, $\hat{m} =$ argmax$_m P(\boldsymbol{z}|\boldsymbol{y}_m)$, is the same.

The blue terms are the extra terms relative to Dasarathy and Draper (2011).

Discussion

- $\exists$ examples where the new decoder strictly improves upon MMI.
- New decoder better than $\forall$ decoder whose metric depends on $\hat{P}_{\boldsymbol{y}_m \boldsymbol{z}}$.
- For most codes,

$$P(\boldsymbol{z}|\boldsymbol{y}_m) \doteq \exp\{-n\gamma(\hat{P}_{\boldsymbol{y}_m \boldsymbol{z}})\}$$

  where $\gamma(\cdot)$ has a single–letter formula.
- Best to keep $Q_X \to Q_Y$ one–to–one (otherwise, perturb a little).

Analysis Tools

- Method of types.
- Focus on pairwise error probability analysis + truncated union bound.
- Properties of Binomial(e^{nA}, e^{-nB}) – “type class enumeration”.
- Lemmas from [Lapidoth-Ziv98], extended to general input assignments.