My Little Toolbox for Code Ensemble Performance Analysis

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering

Technion—Israel Institute of Technology

Haifa, Israel

IZS 2018, Zurich, Switzerland, February 22, 2018.

A Very Quick Historical Overview

- Shannon ('48): random coding as a simple tool for proving $\exists$ good codes.
- Elias ('55,'56); Fano ('61); Gallager ('65, '68): exponential error bounds.
- Shannon, Gallager, Berlekamp ('67): lower bounds: SP, SLB.
- Csiszár & Körner ('81): the method of types.
- Many: extensions, improvements; ensembles of structured codes.

Random coding – a paradigm on its own right.

Traditional Bounding Techniques

- $P_{\mathsf{e}}(\text{ML decoder}) \le P_{\mathsf{e}}(\text{another (easier) decoder})$.
- Jensen's inequality: $\mathbf{E}Z^{\rho} \le (\mathbf{E}Z)^{\rho}$, $0 \le \rho \le 1$ (Gallager–style bounds).
- $\mathcal{I}\{P(\boldsymbol{y}|\boldsymbol{x}_j) \ge P(\boldsymbol{y}|\boldsymbol{x}_i)\} \le [P(\boldsymbol{y}|\boldsymbol{x}_j)/P(\boldsymbol{y}|\boldsymbol{x}_i)]^{\lambda}$ (Chernoff bound).
- Simple union bound.
- Union bound with truncation: $P[\cup_j \mathcal{A}_j] \le \min\{1, \sum_j P[\mathcal{A}_j]\}$.
- Union bound with a power parameter: $P[\cup_j \mathcal{A}_j] \le (\sum_j P[\mathcal{A}_j])^{\rho}$, $0 \le \rho \le 1$.
- Union bound with intersection: $P[\cup_j \mathcal{A}_j] \le \sum_j P[\mathcal{A}_j \cap \mathcal{G}] + P[\mathcal{G}^{\mathsf{c}}]$.
- "Power distribution" inequality: $(\sum_i a_i)^s \le \sum_i a_i^s$, $0 \le s \le 1$ (Forney '68).

All these tools facilitate the analysis a great deal but at the risk of compromising exponential tightness.

Main message of this talk: It is often possible to preserve exponential tightness by bypassing some of the above inequalities.

My Little Toolbox

- *Type class enumeration (on top of the MoT).*
- *Analogue of the MoT for infinite alphabets.*

Difficulty: Summations of Exponentially Many Terms

Many derivations are associated with summations of exponentially many terms, e.g.,

$$\overline{P_{\mathsf{e}}} \le \sum_{\boldsymbol{y}} \mathbf{E}\left\{P(\boldsymbol{y}|\boldsymbol{X})^{1/(1+\rho)}\right\} \cdot \mathbf{E}\left[\sum_{m} P(\boldsymbol{y}|\boldsymbol{X}_m)^{1/(1+\rho)}\right]^{\rho},$$

$$\overline{P_{\mathsf{c}}} = \frac{1}{M}\mathbf{E}\left\{\sum_{\boldsymbol{y}} \max_{m} P(\boldsymbol{y}|\boldsymbol{X}_m)\right\} = \frac{1}{M} \lim_{\beta\to\infty} \sum_{\boldsymbol{y}} \mathbf{E}\left\{\left[\sum_{m} P(\boldsymbol{y}|\boldsymbol{X}_m)^{\beta}\right]^{1/\beta}\right\}.$$

In some situations (e.g., the BC, the IFC, the GPC, the wiretap channel, erasure/list decoding), the optimal likelihood function $=$ sum of exponentially many terms,

Broadcast channel: $\quad \text{score}_i = \sum_{m} P(\boldsymbol{y}|\boldsymbol{x}_{m,i})$

Interference channel: $\quad \text{score}_i = \sum_{m} P(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{x}_m)$

A Natural Remedy: Type Class Enumerators

The idea:

$$\sum_{m} P(\boldsymbol{y}|\boldsymbol{X}_m)^\beta = \sum_{Q} N_{\boldsymbol{y}}(Q) \cdot P(\boldsymbol{y}|\boldsymbol{x}_Q)^\beta = \sum_{Q} N_{\boldsymbol{y}}(Q) \cdot e^{n\beta f(Q)},$$

where

$N_{\boldsymbol{y}}(Q) =$ number of $\boldsymbol{X}_m$ in a given type Q of $\boldsymbol{x}$ given $\boldsymbol{y}$.

What have we gained?

- $\sum$ of exponentially many terms $\rightarrow$ $\sum$ of polynomially few terms.
- $N_{\boldsymbol{y}}(Q) \sim$ Binomial$(e^{nR}, e^{-nI(Q)})$ – easy to handle.
- Marginals of $\{N_{\boldsymbol{y}}(Q)\}$ almost always suffice; Pairs are $\sim$ independent.

Consequence: Avoiding the Use of Jensen's Inequality

$$
\begin{aligned}
\mathbf{E}\left\{\left[\sum_m P(\boldsymbol{y}|\boldsymbol{X}_m)^\beta\right]^{1/\beta}\right\} &= \mathbf{E}\left[\sum_Q N_{\boldsymbol{y}}(Q)\cdot e^{n\beta f(Q)}\right]^{1/\beta} \\
&\doteq \mathbf{E}\left[\max_Q N_{\boldsymbol{y}}(Q)\cdot e^{n\beta f(Q)}\right]^{1/\beta} \\
&= \mathbf{E}\left\{\max_Q [N_{\boldsymbol{y}}(Q)]^{1/\beta}\cdot e^{nf(Q)}\right\} \\
&\doteq \mathbf{E}\left\{\sum_Q [N_{\boldsymbol{y}}(Q)]^{1/\beta}\cdot e^{nf(Q)}\right\} \\
&= \sum_Q \mathbf{E}\{[N_{\boldsymbol{y}}(Q)]^{1/\beta}\}\cdot e^{nf(Q)}.
\end{aligned}
$$

- We just have to know how to assess moments of $N_{\boldsymbol{y}}(Q)$.
- Equivalently, deal with the large deviations behavior.

Properties of $N \sim$ Binomial(e^{nA}, e^{-nB})

Drastic difference between $A > B$ and $A < B$: phase transition at $A = B$.

Moments:

$$E\{N^s\} \doteq \begin{cases} \exp\{ns(A-B)\} & A > B \\ \exp\{n(A-B)\} & A < B \end{cases}$$

Intuition:

- $A > B$: double–exponential concentration of N around its mean $e^{n(A-B)}$.
- $A < B$: $E\{N^s\} = \sum_{n \geq 1} n^s P[N = n] \doteq 1^s P[N = 1] \doteq e^{n(A-B)}$.

Properties of $N \sim$ Binomial(e^{nA}, e^{-nB}) (Cont'd)

Large deviations behavior:

$$\text{Pr}\{N \geq e^{\lambda n}\} \doteq e^{-nE},$$

with

$$E = \begin{cases} [B-A]_+ & [A-B]_+ \geq \lambda \\ \infty & \text{elsewhere} \end{cases}$$

Intuition – “interesting” for $A < B$ and $\lambda \leq 0$: $P[N \geq 1] \doteq e^{-n(B-A)}$.

$$\text{Pr}\{N \leq e^{\lambda n}\} \doteq \begin{cases} 1 & A \leq B + [\lambda]_+ \\ 0 & \text{elsewhere} \end{cases}$$

Example – Exponentially Tight Evaluation of $\overline{P_c}$

Consider the BSC with crossover probability p. Using the relation

$$\mathbf{E}\{N_{\boldsymbol{y}}(Q)^{1/\beta}\} \doteq \begin{cases} \exp\{n[R - I_Q(X;Y)]/\beta\} & R > I_Q(X;Y) \\ \exp\{n[R - I_Q(X;Y)]\} & R < I_Q(X;Y) \end{cases}$$

plugging it to the expression of $\overline{P_{\mathsf{c}}}$, and using the MoT, we get

$$\begin{aligned} \overline{P_{\mathsf{c}}} &\doteq \exp\{-nD(\delta_{\mathsf{GV}}(R)\|p)\} \\ &= \exp\left\{-n\left[\delta_{\mathsf{GV}}(R)\ln\frac{1}{p} + (1-\delta_{\mathsf{GV}}(R))\ln\frac{1}{1-p} - h_2(\delta_{\mathsf{GV}}(R)\right]\right\} \end{aligned}$$

where $\delta_{\mathsf{GV}}(R)$ is the (smaller) solution to the equation

$$\ln 2 - h_2(\delta) = R.$$

Example (Cont'd)

It is interesting to compare it to the result of using Jensen's inequality:

$$\begin{aligned}\overline{P_{\mathsf{c}}} &= \frac{1}{M}\lim_{\beta\to\infty}\sum_{\boldsymbol{y}}\mathbf{E}\left\{\left[\sum_m P(\boldsymbol{y}|\boldsymbol{X}_m)^\beta\right]^{1/\beta}\right\} \\ &\le \frac{1}{M}\lim_{\beta\to\infty}\sum_{\boldsymbol{y}}\left[\mathbf{E}\sum_m P(\boldsymbol{y}|\boldsymbol{X}_m)^\beta\right]^{1/\beta} \\ &\doteq \exp\left(-n\left[\min\left\{\ln\frac{1}{p},\ln\frac{1}{1-p}\right\} - h_2(\delta_{\mathsf{GV}}(R)\right]\right)\end{aligned}$$

Reminder: the red expression should be compared to

$$\delta_{\mathsf{GV}}(R)\ln\frac{1}{p} + (1-\delta_{\mathsf{GV}}(R))\ln\frac{1}{1-p}$$

of the exponentially tight evaluation of the previous slide.

Application to Random Binning

Consider the process of random binning:

Each $\boldsymbol{x} \in \mathcal{X}^n$ is randomly assigned to a bin $z = f(\boldsymbol{x}) \sim$ Unif$\{1, \ldots, e^{nR}\}$.

At the decoder

$$\hat{\boldsymbol{x}}(\boldsymbol{y}, z) = \arg\max_{\boldsymbol{x} \in f^{-1}(z)} P(\boldsymbol{x}|\boldsymbol{y})$$

Then,

$$\begin{aligned}\overline{P_{\mathrm{e}}} &= \mathrm{Pr} \bigcup_{\boldsymbol{x}' \neq \boldsymbol{x}} \{f(\boldsymbol{x}') = f(\boldsymbol{x}),\ P(\boldsymbol{x}'|\boldsymbol{y}) \geq P(\boldsymbol{x}|\boldsymbol{y})\} \\ &\doteq \sum_{\boldsymbol{x}\boldsymbol{y}} P(\boldsymbol{x}, \boldsymbol{y}) \sum_{Q_{X'Y} \in \mathcal{E}} \mathrm{Pr}\,\{N(Q_{X'Y}, f(\boldsymbol{x})) \geq 1\}\end{aligned}$$

where $\mathcal{E}$ is the class of all $\{Q_{X'Y}\}$ with $\mathbf{E}_{Q'} \ln P(X'|Y) \geq \mathbf{E}_Q \ln P(X|Y)$ and where the type class enumerator

$$N(Q_{X'Y}, z) = |\mathcal{T}(Q_{X'Y}|\boldsymbol{y}) \cap f^{-1}(z)| \sim \text{Binomial}(|\mathcal{T}(Q_{X'Y}|\boldsymbol{y})|, e^{-nR}).$$

Avoid Bounding Indicator Functions by Chernoff Bounds

Consider the error+erasure event a la Forney ('68): Instead of

$$\Pr\{\mathcal{E}_1\} = \Pr\left\{\frac{P(\boldsymbol{y}|\boldsymbol{x}_m)}{\sum_{m'\neq m} P(\boldsymbol{y}|\boldsymbol{X}_{m'})} < e^{nT}\right\} \leq e^{nsT}\mathbf{E}\left\{\left(\sum_{m'\neq m}\frac{P(\boldsymbol{y}|\boldsymbol{X}_{m'})}{P(\boldsymbol{y}|\boldsymbol{x}_m)}\right)^s\right\},$$

$$\begin{aligned}
\text{use}: \ \Pr\{\mathcal{E}_1\} &= \Pr\left\{\sum_{m'\neq m} P(\boldsymbol{y}|\boldsymbol{X}_{m'}) > e^{-nT}P(\boldsymbol{y}|\boldsymbol{x}_m)\right\} \\
&= \Pr\left\{\sum_{Q} N_{\boldsymbol{y}}(Q)e^{nf(Q)} > e^{-nT}e^{nf(Q_m)}\right\} \\
&\doteq \Pr\left\{\max_{Q} N_{\boldsymbol{y}}(Q)e^{nf(Q)} > e^{-nT}e^{nf(Q_m)}\right\} \\
&\doteq \Pr\bigcup_{Q}\left\{N_{\boldsymbol{y}}(Q)e^{nf(Q)} > e^{n[f(Q_m)-T]}\right\} \\
&\doteq \max_{Q}\Pr\left\{N_{\boldsymbol{y}}(Q) > e^{n[f(Q_m)-f(Q)-T]}\right\}
\end{aligned}$$

and now the large deviaions properties of a single $N_{\boldsymbol{y}}(Q)$ are invoked..

What if Those Sums Appear Also in the Denominator?

Consider the likelihood decoder that randomly selects $\hat{m}$ under the posterior:

$$\overline{P_{\mathsf{e}|m=0}} = \mathbf{E}\left\{\frac{\sum_{m=1}^{M-1} P(\boldsymbol{Y}|\boldsymbol{X}_m)}{\sum_{m=0}^{M-1} P(\boldsymbol{Y}|\boldsymbol{X}_m)}\right\}.$$

$$\begin{aligned}
& \mathbf{E}\left\{\frac{\sum_{m=1}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m)}{P(\boldsymbol{y}|\boldsymbol{x}_0) + \sum_{m=1}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m)}\right\} \\
= \quad & \int_0^1 \mathrm{d}s \cdot \mathrm{Pr}\left\{\frac{\sum_{m=1}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m)}{P(\boldsymbol{y}|\boldsymbol{x}_0) + \sum_{m=1}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m)} \geq s\right\} \\
= \quad & n \cdot \int_0^\infty \mathrm{d}\theta e^{-n\theta} \mathrm{Pr}\left\{\frac{\sum_{m=1}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m)}{P(\boldsymbol{y}|\boldsymbol{x}_0) + \sum_{m=1}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m)} \geq e^{-n\theta}\right\} \\
\doteq \quad & \int_0^\infty \mathrm{d}\theta e^{-n\theta} \mathrm{Pr}\left\{\sum_{m=1}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m) \geq e^{-n\theta} P(\boldsymbol{y}|\boldsymbol{x}_0)\right\}
\end{aligned}$$

and the rest is as before.

What if ... in the Denominator? (Cont'd)

Sometimes random denominators can be handled using transform methods.
For example, let $X_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, n$, be independent. Then,

$$\mathbf{E}\left\{\frac{1}{\sum_{i=1}^{n} X_i^2}\right\} \quad = \quad ???$$

What if ... in the Denominator? (Cont'd)

Sometimes random denominators can be handled using transform methods. For example, let $X_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, n$, be independent. Then,

$$
\begin{aligned}
\mathbf{E}\left\{\frac{1}{\sum_{i=1}^{n} X_i^2}\right\} &= \mathbf{E}\left\{\int_0^\infty \mathrm{d}t \cdot \exp\left[-t\sum_{i=1}^{n} X_i^2\right]\right\} \\
&= \int_0^\infty \mathrm{d}t \cdot \mathbf{E}\left\{\exp\left[-t\sum_{i=1}^{n} X_i^2\right]\right\} \\
&= \int_0^\infty \frac{\mathrm{d}t}{(1+2\sigma^2 t)^{n/2}} \\
&= \begin{cases} \infty & n \le 2 \\ \frac{1}{(n-2)\sigma^2} & n > 2 \end{cases}
\end{aligned}
$$

Analogue of the MoT for Infinite Alphabets

In the memoryless finite–alphabet (FA) case, we usually think of the type class of a given $\boldsymbol{x}$ as the set of all $\boldsymbol{x}'$

- with the same empirical distribution as $\boldsymbol{x}$,
- that are permutations of $\boldsymbol{x}$.

These definitions are specific to the FA memoryless case.

An alternative definition that lends itself to extensions:

$$\mathcal{T}(\boldsymbol{x}) = \left\{\boldsymbol{x}' : \ P(\boldsymbol{x}') = P(\boldsymbol{x}) \text{ for every memoryless source } \ P\right\}.$$

For a general parametric family of sources $\{P_\theta, \ \theta \in \Theta\}$:

$$\mathcal{T}(\boldsymbol{x}) = \left\{\boldsymbol{x}' : \ P_\theta(\boldsymbol{x}') = P_\theta(\boldsymbol{x}) \text{ for every } \theta \in \Theta\right\}.$$

Analogue of the MoT for Infinite Alphabets (Cont'd)

If $\{P_\theta,\ \theta \in \Theta\}$ is an exponential family:

$$P_\theta(\boldsymbol{x}) = \frac{\exp\left\{-\sum_{i=1}^{k} \theta_i \phi_i(\boldsymbol{x})\right\}}{Z(\theta)},$$

then

$$\mathcal{T}(\boldsymbol{x}) = \left\{\boldsymbol{x}' :\ \phi_i(\boldsymbol{x}') = \phi_i(\boldsymbol{x}),\ \ i = 1, 2, \ldots, k\right\}.$$

FA memoryless: $\phi_i(\boldsymbol{x}) = \sum_{t=1}^{n} \mathcal{I}\{x_t = i\}$

FA Markov: $\phi_{ij}(\boldsymbol{x}) = \sum_{t=1}^{n} \mathcal{I}\{x_t = i, x_{t+1} = j\}$

Gaussian memoryless: $\phi_1(\boldsymbol{x}) = \sum_{t=1}^{n} x_t;\ \ \phi_2(\boldsymbol{x}) = \sum_{t=1}^{n} x_t^2.$

Zero–mean, Gaussian AR(p): $\phi_i(\boldsymbol{x}) = \sum_{t=1}^{n} x_t x_{t+i},\ \ i = 0, 1, \ldots, k$

Analogue of the MoT for Infinite Alphabets (Cont'd)

The main building blocks (just like in the ordinary MoT):

- A computable expression for $|\mathcal{T}(\boldsymbol{x})|$, or Vol$\{\mathcal{T}(\boldsymbol{x})\}$.
- Make sure that number of different types is not too large.

If $\mathcal{X} = \mathbb{R}$ (say, the Gaussian case), we have two problems:

- Vol$\{\mathcal{T}(\boldsymbol{x})\} = 0$.
- The space is unbounded $\rightarrow$ infinitely many types.

First problem – allow some tolerance ϵ:

$$\mathcal{T}_\epsilon(\boldsymbol{x}) = \left\{\boldsymbol{x}' : \; |\phi_i(\boldsymbol{x}') - \phi_i(\boldsymbol{x})| \le \epsilon, \;\; i = 1, 2, \ldots, k\right\}.$$

But this still does not resolve the second problem.

Second problem – confine attention to a bounded region in $\mathbb{R}^n$ (say, a sphere), outside of which the probability decays with a large enough exponent.

Analogue of the MoT for Infinite Alphabets (Cont'd)

To assess the exponent of Vol$\{\mathcal{T}(\boldsymbol{x})\}$:

$$1 \geq \int_{\mathcal{T}_\epsilon(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}' \cdot P_\theta(\boldsymbol{x}') \doteq \mathrm{Vol}\{\mathcal{T}_\epsilon(\boldsymbol{x})\} \cdot P_\theta(\boldsymbol{x}),$$

$$\text{leading to } \mathrm{Vol}\{\mathcal{T}_\epsilon(\boldsymbol{x})\} \stackrel{\cdot}{\leq} \frac{1}{P_\theta(\boldsymbol{x})} = \exp\left\{\ln Z(\theta) + \sum_{i=1}^{k} \theta_i \phi_i(\boldsymbol{x})\right\}$$

$$\text{and since this is } \forall\theta: \quad \mathrm{Vol}\{\mathcal{T}_\epsilon(\boldsymbol{x})\} \leq \min_\theta \exp\left\{\ln Z(\theta) + \sum_{i=1}^{k} \theta_i \phi_i(\boldsymbol{x})\right\}.$$

Exponentially tight as the minimizer θ^* assigns $P_{\theta^*}\{\mathcal{T}_\epsilon(\boldsymbol{x})\} \approx 1$ (WLLN).
The same idea applies to assess volumes to conditional types:

$$\mathcal{T}_\epsilon(\boldsymbol{x}|\boldsymbol{y}) = \left\{\boldsymbol{x}' : \ |\phi_i(\boldsymbol{x}', \boldsymbol{y}) - \phi_i(\boldsymbol{x}, \boldsymbol{y})| \leq \epsilon, \ \ i = 1, 2, \ldots, k\right\}.$$

Here one defines an exponential family of channels.

Analogue of the MoT for Infinite Alphabets (Cont'd)

A challenge (relevant to ISI channels) is to assess the volume of a conditional type defined by both $\frac{1}{n}\sum_t g(x_t, y_t)$ and $\sum_{t=1}^n h(x_t, x_{t+j})$, $j = 1, \ldots, k$, for some g and h. For example, the volume of

$$\mathcal{T}(\phi, \psi, \mu|\boldsymbol{y}) = \left\{\boldsymbol{x}: \ \sum_{t=1}^n f(x_t) = n\phi, \ \sum_{t=1}^n, h(x_t, x_{t-1}) = n\psi, \ \sum_{t=1}^n g(x_t, y_t) = n\mu\right\}$$

is $\displaystyle\int_{\mathbb{R}^n} \mathrm{d}\boldsymbol{x}\delta\left(\sum_{t=1}^n f(x_t) - n\phi\right)\delta\left(\sum_{t=1}^n h(x_t, x_{t+1}) - n\psi\right)\delta\left(\sum_{t=1}^n g(x_t, y_t) - n\mu\right)$.

Next, represent $\displaystyle\delta(A) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\{i\omega A\}\mathrm{d}\omega$

then interchange the integrations, and finally, use the saddle–point method. Such a derivation is doable at least when f, g and h are all quadratic (the Gaussian case), as this is a Gaussian integral (Huleihel, Salamatian, Merhav & Médard, 2017).

Some Results ...

Example 1: List Decoding (IT, Nov. 2014)

- A code $\mathcal{C} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$, $M = e^{nR}$, is selected at random.
- The marginal of each codeword $\boldsymbol{x}_i \in \mathcal{X}^n$ is Unif$\{\mathcal{T}(Q)\}$.
- The channel $P(\boldsymbol{y}|\boldsymbol{x})$ is a DMC.
- The index I of the transmitted message $\boldsymbol{x}_I$ is Unif$\{0, 1, \ldots, M-1\}$.
- The decoder outputs the indices of the L most likely messages.
- Error event: I is not on the list.
- Regimes: fixed list size (FLS) and exponential list size (ELS).

Example 1: List Decoding (Cont'd)

A general, non–asymptotic bound:

Theorem: The average probability of list error, $\overline{P_e}$, associated with the optimal list decoder, is upper bounded by

$$\overline{P_e} \leq \sum_{\boldsymbol{x}.\boldsymbol{y}} P(\boldsymbol{x})P(\boldsymbol{y}|\boldsymbol{x}) \exp\left\{-nL\left[\hat{I}_{\boldsymbol{xy}}(X;Y) + \frac{\ln L}{n} - R - O\left(\frac{\log n}{n}\right)\right]_+\right\},$$

where $P(\boldsymbol{x})$ is the uniform distribution over $\mathcal{T}(Q)$ and $\hat{I}_{\boldsymbol{xy}}(X;Y)$ is the empirical mutual information induced by $(\boldsymbol{x}, \boldsymbol{y})$.

The proof is by a large deviations analysis of the binomial RV

$$N(\boldsymbol{x}, \boldsymbol{y}) = \sum_{m=1}^{M-1} \mathcal{I}\{P(\boldsymbol{y}|\boldsymbol{X}_m) \geq P(\boldsymbol{y}|\boldsymbol{x})\}.$$

Example 1: List Decoding (Cont'd)

The dependence on L appears twice:

$$\overline{P\mathrm{e}} \leq \sum_{\boldsymbol{x},\boldsymbol{y}} P(\boldsymbol{x})P(\boldsymbol{y}|\boldsymbol{x}) \exp\left\{-\underbrace{nL}_{\text{FLS}}\left[\hat{I}_{\boldsymbol{xy}}(X;Y) + \frac{\overbrace{\ln L}^{\text{ELS}}}{n} - R - O\left(\frac{\log n}{n}\right)\right]_+\right\},$$

In the FLS regime, $\frac{\ln L}{n} \to 0$, and averaging $\exp\{-nL[\hat{I}_{\boldsymbol{xy}}(X;Y) - R]_+\}$ yields

$$\overline{P_e} \stackrel{\cdot}{\leq} e^{-nE(R,L,Q)}, \quad \text{where}$$

$$E(R,L,Q) \stackrel{\triangle}{=} \min_{\tilde{P}_{Y|X}} \{D(\tilde{P}_{Y|X} \| P_{Y|X}|Q) + L \cdot [\tilde{I}(X;Y) - R]_+\},$$

The best exponent is obtained by maximizing over Q to yield

$$E(R,Q) = \max_Q E(R,L,Q).$$

Example 1: List Decoding (Cont'd)

$$\overline{P_{\mathsf{e}}} \le \sum_{\boldsymbol{x},\boldsymbol{y}} P(\boldsymbol{x})P(\boldsymbol{y}|\boldsymbol{x}) \exp\left\{-nL\left[\hat{I}_{\boldsymbol{x}\boldsymbol{y}}(X;Y) + \frac{\ln L}{n} - R - O\left(\frac{\log n}{n}\right)\right]_+\right\},$$

In the ELS regime, $\frac{\ln L}{n} = \lambda$. By defining

$$\mathcal{E} = \left\{(\boldsymbol{x},\boldsymbol{y}):\ \hat{I}_{\boldsymbol{x}\boldsymbol{y}}(X;Y) + \lambda - R \ge \epsilon\right\}.$$

we see that the conribution of $\mathcal{E}$ is $\le \exp(-n\epsilon e^{\lambda n}) \doteq e^{-n\infty}$, and so,

$$\begin{aligned}\overline{P_e} \stackrel{\cdot}{\le} \mathrm{Pr}\{\mathcal{E}^c\} &\doteq \exp\left\{-n \min_{\{\tilde{P}_{Y|X}:\ \tilde{I}(X;Y)\le R-\lambda\}} D(\tilde{P}_{Y|X}\|P_{Y|X}|Q)\right\} \\ &\stackrel{\triangle}{=} \exp\{-nE_{\mathsf{sp}}(R-\lambda,Q)\}\end{aligned}$$

which, for the optimum Q, becomes $\exp\{-nE_{\mathsf{sp}}(R-\lambda)\}$ — meeting the converse bound of Shannon–Gallager–Berlekamp ('67).

Example 2: Erasure/List S–W Decoding (2014)

Let $(\boldsymbol{X}, \boldsymbol{Y}) \sim \prod_{i=1}^{n} P(x_i, y_i)$.

$\boldsymbol{x}$ – source to be encoded.

$\boldsymbol{y}$ – side info @ decoder.

Encoder: $f : \mathcal{X}^n \to \{0, 1, \ldots, M-1\}$, $M = e^{nR}$.

$$z = f(\boldsymbol{x}).$$

Random binning: For every $\boldsymbol{x} \in \mathcal{X}^n$, z is selected independently at random from $\{0, 1, \ldots, M-1\}$.

Example 2: Erasure/List S–W Decoding (Cont'd)

Erasure/list decoder: Given $\boldsymbol{y} \in \mathcal{Y}^n$ and z, calculate for all $\hat{\boldsymbol{x}} \in f^{-1}(z)$:

$$\frac{P(\hat{\boldsymbol{x}}, \boldsymbol{y})}{\sum_{\boldsymbol{x}' \in f^{-1}(z) \setminus \{\hat{\boldsymbol{x}}\}} P(\boldsymbol{x}', \boldsymbol{y})}.$$

If $\geq e^{nT}$, $\hat{\boldsymbol{x}}$ is a candidate.

- If there are no candidates – an erasure is declared.
- If there is exactly one candidate – ordinary decoding: $\hat{\boldsymbol{x}} =$candidate.
- If there is more than one candidate – a list is of all candidates is created.

Define $\mathcal{E}_1$ as the event where the real $\boldsymbol{x}$ is not a candidate.
Let $E_1(R,T) =$ exponent of Pr$\{\mathcal{E}_1\}$. The other exponent

$$E_2(R,T) = \begin{cases} \text{decoding error exp} & \text{erasure mode} \\ \text{expected list size exp} & \text{list mode} \end{cases} = E_1(R,T) + T.$$

Example 2: Erasure/List S–W Decoding (Cont'd)

Model: A double–BSS with a BSC(p) in between.

$E_1^{\text{tce}}(R,T) \geq E_1^{\text{Forney}}(R,T)$ always.

For some regions in the plane R—T, $E_1^{\text{tce}}(R,T)$ may be larger than $E_1^{\text{Forney}}(R,T)$ by an arbitrarily large factor!

1. For $R > h(p)$ and $T < \ln \frac{p}{1-p}$:

$$E_1^{\text{Forney}}(R,T) \leq R + |T| < \infty; \qquad E_1^{\text{tce}}(R,T) = \infty.$$

2. Consider the case of very weakly correlated sources, i.e., $p = \frac{1}{2} - \epsilon$, $|\epsilon| \ll 1$.

For $R \in [h(p), \ln 2]$ and $T = -\tau\epsilon^2$ with $\tau > 4$:

$$E_1^{\text{Forney}}(R,T) \leq (\tau + 2)\epsilon^2, \quad E_1^{\text{tce}}(R,T) \geq \left[\frac{\tau(\tau+8)}{16} - 1\right]\epsilon^2.$$

Example 3: Typical Random Codes (2017)

While traditional random coding error exponents are defined as

$$E_{\mathsf{r}}(R) = \lim_{n\to\infty}\left[-\frac{\ln \mathbf{E} P_{\mathsf{e}}(\mathcal{C}_n)}{n}\right],$$

typical–code error exponents are defined as

$$E_{\mathsf{typ}}(R) = \lim_{n\to\infty}\left[-\frac{\mathbf{E} \ln P_{\mathsf{e}}(\mathcal{C}_n)}{n}\right].$$

- By Jensen's inequality, $E_{\mathsf{typ}}(R) \ge E_{\mathsf{r}}(R)$.
- $E_{\mathsf{r}}(R)$ – dominated by bad codes; $E_{\mathsf{typ}}(R)$ dominated by typical codes.

Let $\mathcal{G}_E = \{\mathcal{C} : \;\; P_{\mathsf{e}}(\mathcal{C}) \doteq e^{-nE}\}$.

$$\overline{P_{\mathsf{e}}(\mathcal{C})} \doteq \sum_E P(\mathcal{G}_E)\cdot e^{-nE} \doteq P(\mathcal{G}_E^*)\cdot e^{-nE^*},$$

whereas $E_{\mathsf{typ}}(R) = E_0$, where $P[\mathcal{G}_{E_0}] \to 1$.

Example 3: Typical Random Codes (Cont'd)

We derive the exact typical–code error exponent for a class of stochastic decoders,

$$P(\hat{m} = m|\boldsymbol{y}) \propto \exp\{ng(\hat{P}_{\boldsymbol{x}_m \boldsymbol{y}})\}.$$

and show that

$$E_{\mathsf{typ}}(R) = E_{\mathsf{ex}}(2R) + R,$$

Extending Barg & Forney (2002) in several directions:

- General DMC is considered, not merely the BSC.
- Covering a wider family of decoders.
- Ensemble of constant composition codes – optimal PI distribution.
- Relation to expurgated exponent – for all R and a general decoder.
- The analysis technique is applicable also to more general scenarios.

Example 3: Typical Random Codes (Cont'd)

Particularizing to ML decoding, the error exponent formula includes minimization subject to the constraint,

$$\mathbf{E}_Q \ln W(Y|X') \geq \max\{\mathbf{E}_Q \ln W(Y|X), D(R, Q_Y)\},$$

$$D(R, Q_Y) = \sup\{\mathbf{E}_Q \ln W(Y|X'') : \ I_Q(X''; Y) \leq R, \ (Q_Y \times Q_{X''|Y})_X = Q_X\},$$

being the typical highest score of an incorrect message.

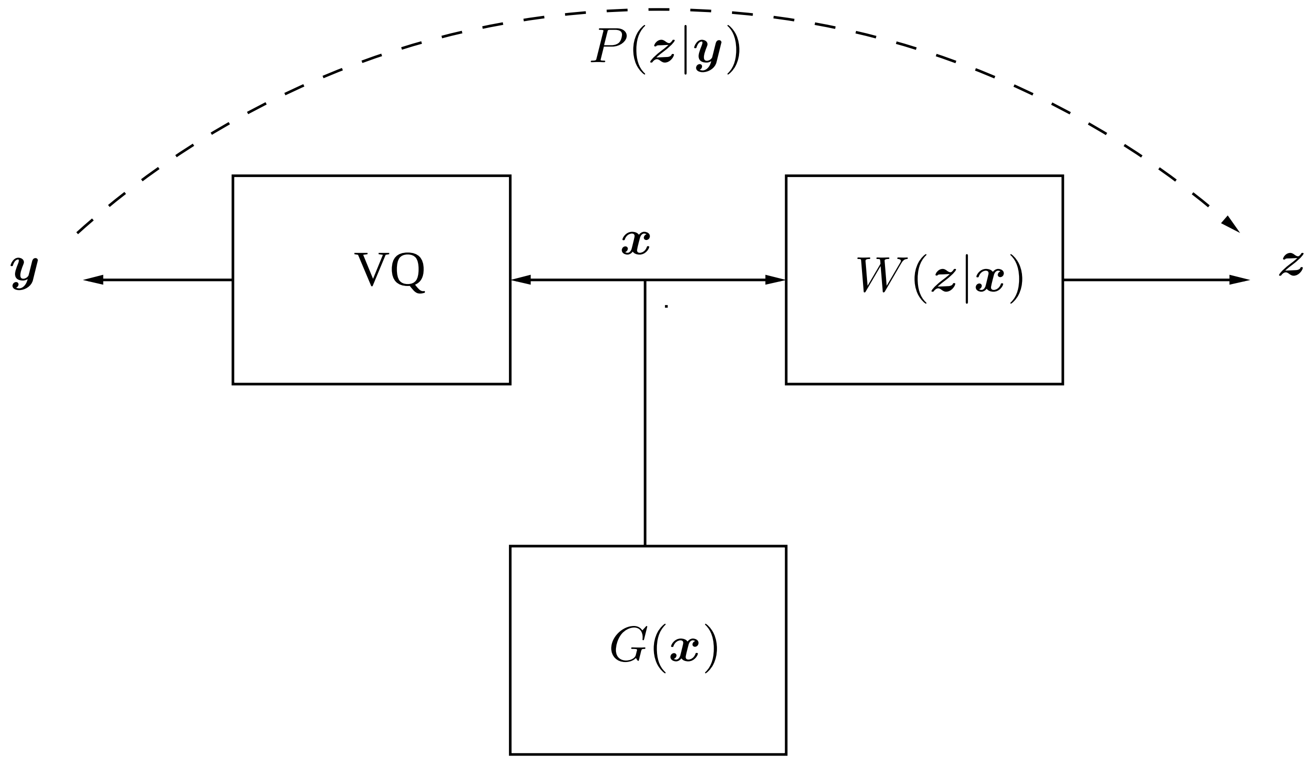
A technical issue: handling summations of exponentially many fractions with random denominators – exploit concentration properties.

$$\mathbf{E}\left[\frac{1}{M}\sum_{m}\sum_{m'\neq m}\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{X}_m) \cdot \frac{P(\boldsymbol{y}|\boldsymbol{X}_{m'})}{P(\boldsymbol{y}|\boldsymbol{X}_m) + \sum_{\tilde{m}\neq m} P(\boldsymbol{y}|\boldsymbol{X}_{\tilde{m}})}\right]^{\rho}.$$

Example 4: Broadcast Channels (R. Averbuch)

- Exact exponents for the weak and strong user with optimal decoders.
- Universal decoders for both users, achieving the same error exponents.
- Significant improvement and simplification of earlier results.
- Gallager–style lower bounds for both users.
- Expurgated exponents (joint work also with N. Weinberger).

Example 5: Channel Decoding with VQ'ed 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:

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

Example 5: Decoding with VQ'ed Codewords (Cont'd)

Main contributions:

- Exponentially tight bound on the ensemble performance.
- Improvement relative to 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 $\{P(\boldsymbol{z}|\boldsymbol{y})\}$.

Example 5: Decoding with VQ'ed Codewords (Cont'd)

Ensemble of VQ's:

- $\forall$ 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 $M(\boldsymbol{x}, \boldsymbol{y})$.

Example 5: Decoding with VQ'ed Codewords (Cont'd)

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.

Some Other Works

- Improved bounds for erasure/list decoding (2008).
- The interference channel (w. Etkin & Ordentlich, 2010).
- The broadcast channel (w. Kaspi, 2011).
- Exact bounds for erasure/list decoding (w. Somekh–Baruch, 2011).
- Expurgation (w. Scarlett, Peng, Guillén i. Fabregas, Martinéz, 2014).
- Erasure/list for S–W decoding (2014).
- Codeword or noise? (w. Weinberger, 2014).
- Optimal bin index decoding (2014).
- Correct wiretapper decoding (2014).
- Statistical physics of random binning (2015).
- Universal source/channel with SI (2016).
- Simplified erasure/list decoding (w. Weinberger, 2017).
- Improved exponents for the IFC (w. Huleihel, 2017).

Some Other Works (Cont'd)

- Joint channel detection & coding (w. Weinberger, 2017).
- Generalized likelihood decoder (2017).
- Exact secrecy exponents (w. Bastani-Parizi & Telatar, 2017).
- Universal decoding for VQ'ed codewords (2017).
- Exact exponents & universal decoding for the ABC (w. Averbuch, 2017).
- Ensemble performance of biometric ident. systems (2017).
- Mismatched ISI channels (w. Huleihel, Salamatian & Médard, 2017).
- V–L codes with single–bit feedback (w. Ginzach & Sason, 2017).
- Typical–code random coding exponents (2017).
- Expurgated bounds for the ABC (w. Averbuch & Weinberger, 2017).
- 2nd order & moderate deviations in error+erasure (Hayashi & Tan, 2015).
- Residual uncertainties under Rényi entropies (Hayashi & Tan, 2016).
- Mismatched decoding (Scarlett, Ph.D. thesis, 2014).

Future Challenges and Open Problems

- Handling ensembles of linear/lattice/convolutional/LDPC codes, etc.
- Further results on typical random codes (multi-user configurations).
- Simplify optimization problems (e.g., Gallager–style bounds).
- A more solid theory for the extended MoT (for exponential families).

Thank U 4 Coming & Listening!