A Largrange–Dual Lower Bound to the Error Exponent of the Typical Random Codes

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering

Technion – Israel Institute of Technology

Haifa 3200003, Israel

ITA 2020, San Diego, CA, U.S.A., February 2020

Typical Random Codes

Traditional random coding error exponents are defined as

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

We define typical–code error exponents 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, $E_{\mathsf{typ}}(R) \geq E_{\mathsf{r}}(R)$.
- $E_{\mathsf{r}}(R)$ – dominated by bad codes; $E_{\mathsf{typ}}(R)$ – by typical codes.

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

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

Otoh, $E_{\mathsf{typ}}(R) = \sum_E P(\mathcal{G}_E)\cdot E = E_0$, where $P[\mathcal{G}_{E_0}] \to 1$.

Motivation

- $E_{\mathsf{typ}}(R)$ is never worse than $E_{\mathsf{r}}(R)$.
- Code selected once and for all: no LLN to support $\mathbf{E}P_{\mathsf{e}}(\mathcal{C}_n)$.
- Once selected, w.h.p. $P_{\mathsf{e}}(\mathcal{C}_n) \sim e^{-nE_0}$, forever.
- Theoretical framework for random–like codes (Battail, 1995).
- Analogy: phys. of disordered sys. - quenched vs. annealed average.

Q: Why wasn't it explored long time before?

A: Not so easy to analyze (also in physics) ....

Related Work

- Barg & Forney (2002): i.i.d. random coding, BSC:

  At low rates: $E_{\mathsf{typ}}(R) = E_{\mathsf{ex}}(2R) + R$.

- Nazari (2011); Nazari, Anastasopoulos & Pradhan (2014):

  upper and lower bounds for the α–decoder.

- Stat. phys. literature: Kabashima (2008), Mora & Riviore (2006), ...:

  LDPC codes - replica analysis and cavity method.

- Battail (1995):

  random–like codes.

- Merhav (2018, 2019):

  Exact formula, the colored Gaussian channel; trellis codes.

- Averbuch *et al.* (2019):

  Large deviations of $\log P_{\mathsf{e}}$; Slepian–Wolf codes.

Main Contribution

A Lagrange–dual lower bound to the typical–code error exponent for a mismatched likelihood decoder

$$P(\hat{m} = m|\boldsymbol{y}) \propto \tilde{W}^{\beta}(\boldsymbol{y}|\boldsymbol{x}_m), \qquad \beta > 0$$

Advantages:

- ♡ Optimization over 5 vs. $|\mathcal{X}|^2 \cdot |\mathcal{Y}| + (|\mathcal{X}| - 1)| \cdot |\mathcal{Y}| - 1$ parameters.
- ♡ One vs. $|\mathcal{X}|^2 \cdot |\mathcal{Y}| - 1$ parameters for minimization.
- ♡ Several insights are gained from the resulting expression.

The Csiszár–Style Expression

Let

$$\alpha(R, Q_Y) = \sup[g(Q_{XY}) - I_Q(X;Y)] + R,$$

where supremum is over $\{Q_{X|Y} : \ I_Q(X;Y) \le R, \ Q_X = P_X\}$.

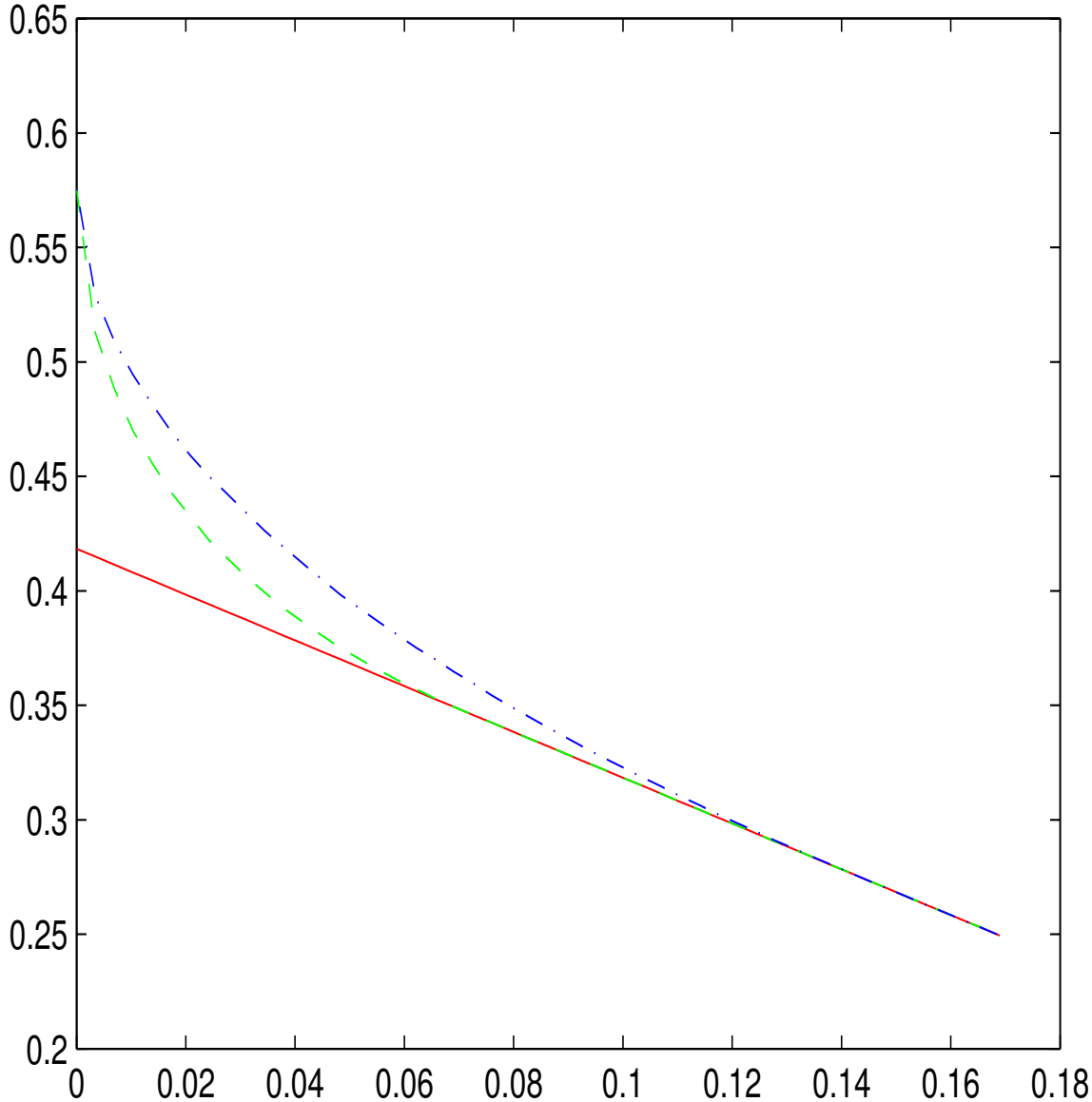
$$\begin{aligned}\Gamma(Q_{XX'}, R) \quad = \quad & \inf_{Q_{Y|XX'}} \{D(Q_{Y|X} \| W | P_X) + I_Q(X';Y|X) + \\ & [g(Q_{XY}) \wedge \alpha(R, Q_Y) - g(Q_{X'Y})]_+\},\end{aligned}$$

where $g(Q) = \mathbf{E}_Q \ln \tilde{W}(Y|X)$.

The typical error exponent is

$$E_{\text{typ}}(R) = \inf\{\Gamma(Q_{XX'}, R) + I_Q(X;X')\} - R,$$

where the infimum is over
$\{Q_{XX'} : \ I_Q(X;X') \le 2R, \ Q_X = Q_{X'} = P_X\}$.



rand. coding, expurg. and typical exponents for z–channel.

The Largrange–Dual (Gallager–style) Expression

Define

$$A(x,x',\sigma,\tau,\lambda)=\sum_y W(y|x)\cdot\frac{\tilde{W}^{\sigma+\tau}(y|x')}{\tilde{W}^{\sigma}(y|x)\left[\sum_{x''}P(x'')\tilde{W}^{1/\lambda}(y|x'')\right]^{\lambda\tau}},$$

$$B(x,\theta,\sigma,\tau,\lambda)=\sum_{x'}P(x')[A(x,x',\sigma,\tau,\lambda)]^{1/(1+\theta)},$$

and

$$C(\zeta,\theta,\sigma,\tau,\lambda)=\sum_x P(x)[B(x,\theta,\sigma,\tau,\lambda)]^{(1+\theta)/\zeta}.$$

Then,

$$E_{\mathsf{typ}}(R) \geq \sup_{0\leq\sigma\leq\beta}\sup_{0\leq\tau\leq\beta-\sigma}\inf_{\lambda\geq 0}\sup_{\theta\geq 0}\sup_{\zeta\geq 1+\theta}\left\{-\zeta\ln C(\zeta,\theta,\sigma,\tau,\lambda)-(\zeta+\theta-\lambda\tau)R\right\}.$$

Some Observations

In the inner–most expression,

$$A(x,x',\sigma,\tau,\lambda)=\sum_y W(y|x)\cdot\left[\frac{\tilde{W}(y|x')}{\tilde{W}(y|x)}\right]^{\sigma}\cdot\left[\frac{\tilde{W}(y|x')}{\left\{\sum_{x''}P(x'')\tilde{W}^{1/\lambda}(y|x'')\right\}^{\lambda}}\right]^{\tau},$$

red part $=$ Chernoff bound of pairwise error: x' beats x;

blue part – x' beats all other wrong codewords, $\{x''\}$.

- ♣ At low R, pairwise errors dominate $\Leftrightarrow\ \tau=0$.
- ♣ As R grows, more weight is assigned to the blue part.
- ♣ λ depends on R.

The Matched Case: $\tilde{W} = W$, $\beta \to \infty$

- The Lagrange–dual formula generalizes both $E_{\mathsf{sp}}(R)$ and $E_{\mathsf{ex}}(R)$.
- Low rates.

Let ϱ be the achiever of $\sup_{\rho \geq 1}[E_{\mathsf{x}}(\rho) - 2\rho R]$.

Let $\sigma = \frac{1}{2}$, $\tau = 0$, $\zeta = \varrho$, and $\theta = \varrho - 1$.

Then,

$$E_{\mathsf{trc}}(R, P) \geq E_{\mathsf{ex}}(2R, P) + R, \qquad R \leq \dot{E}_{\mathsf{x}}(1)/2.$$

The Matched Case: $\tilde{W} = W$, $\beta \to \infty$ (Cont'd)

♠ High rates.

Let ϱ be the achiever of $\sup_{\rho \geq 0}[E_0(\rho, P) - \rho R]$.

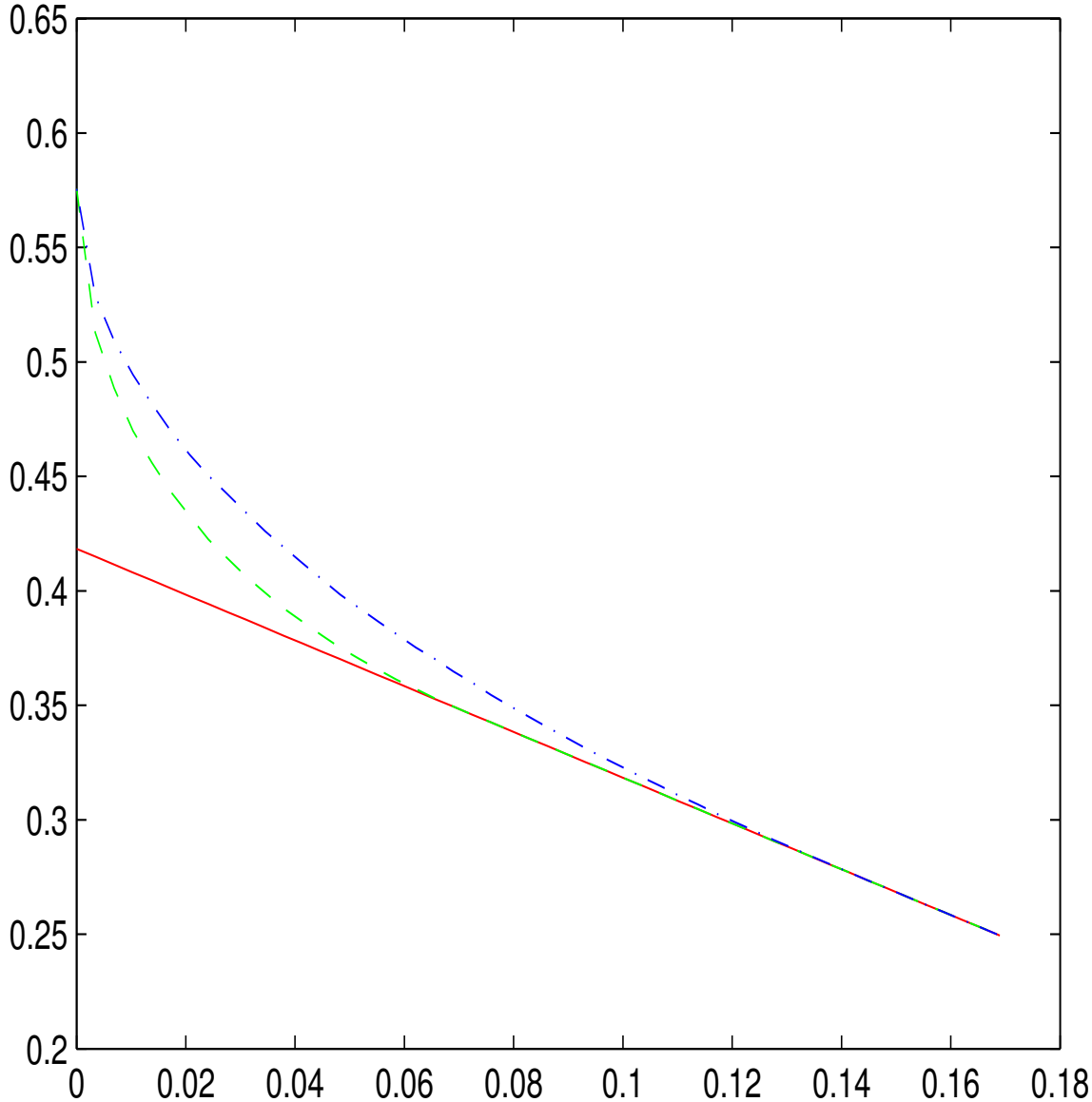
Let $\sigma = \frac{\varrho}{1+\varrho}$, $\tau = \frac{1-\varrho}{1+\varrho}$, $\zeta = 1$, and $\theta = 0$.

Then, $\lambda^* = 1 + \varrho$ and

$$E_{\mathsf{trc}}(R, P) \geq E_{\mathsf{sp}}(R, P).$$

♠ Moderate rates.

Same as high rates, but with $\varrho = 1$.



rand. coding, expurg. and typical exponents for z–channel.

The Role of β

Recall that we are considering the stochastic decoder

$$P(\hat{m} = m|\boldsymbol{y}) \propto \tilde{W}^{\beta}(\boldsymbol{y}|\boldsymbol{x}_m), \qquad \beta > 0.$$

The parameter β appears in red:

$$E_{\text{typ}}(R) \geq \sup_{0\leq\sigma\leq\beta} \sup_{0\leq\tau\leq\beta-\sigma} \inf_{\lambda\geq 0} \sup_{\theta\geq 0} \sup_{\zeta\geq 1+\theta} \{-\zeta \ln C(\zeta,\theta,\sigma,\tau,\lambda) - (\zeta+\theta-\lambda\tau)R.\}\,.$$

A few observations:

- $P_{\text{e}}\{\tilde{W} = W,\ \beta = 1\} \leq 2P_{\text{e}}^* \rightarrow\ \forall\beta \geq 1$ is optimal.
- $\beta \rightarrow \infty$, $\tilde{W} = W$, low R: $\sigma^* = \frac{1}{2}$, $\tau^* = 0 \rightarrow\ \forall\beta \geq \frac{1}{2}$ is optimal.
- $\beta \rightarrow \infty$, $\tilde{W} = W$, high R: $\sigma^* = \frac{\varrho}{1+\varrho}$, $\tau^* = \frac{1-\varrho}{1+\varrho} \rightarrow\ \forall\beta \geq \frac{1}{1+\varrho}$ is optimal.
- Even if $\tilde{W} \neq W$, the error exponent is non–decreasing in β.

Future Directions

- Analogues in source coding (e.g., Slepian–Wolf).
- Source–channel coding.
- Multi-user situations: MAC, BC, etc.
- Other (more structured) ensembles: allowing dependencies.