The Error Exponent of Random Gilbert-Varshamov Codes

Anelia Somekh-Baruch
Bar-Ilan University
somekha@biu.ac.il

Jonathan Scarlett
National University of Singapore
scarlett@comp.nus.edu.sg

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

We consider transmission over a discrete memoryless channel (DMC) $W(y|x)$ with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. It is assumed that an (n, M_n)-codebook $\mathcal{M}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M_n}\}$ with rate $R_n = \frac{1}{n}\log M_n$ is used for transmission. The type-dependent maximum-metric decoder estimates the transmitted message as

$$\hat{m} = \operatorname*{arg\,max}_{\boldsymbol{x}_i \in \mathcal{M}_n} q(\hat{P}_{\boldsymbol{x}_i,\boldsymbol{y}}), \tag{1}$$

where $\hat{P}_{\boldsymbol{x},\boldsymbol{y}}$ is the joint empirical distribution [1, Ch. 2] of the pair $(\boldsymbol{x},\boldsymbol{y})$ and the metric $q : \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \to \mathbb{R}$ is continuous. Maximum-likelihood (ML) decoding is a special case of (1), but the decoder may in general be *mismatched* [2], [3].

We construct the code $\mathcal{M}_n$ such that any two distinct codewords $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{M}_n$ satisfy $d(\boldsymbol{x},\boldsymbol{x}') > \Delta$ for a given distance function $d(\cdot,\cdot)$ and $\Delta \in \mathbb{R}$. This guarantees that the minimum distance of the codebook exceeds Δ. Similar constructions are used to prove the Gilbert-Varshamov bound in Hamming spaces [4], [5]. Our construction depends on an input distribution $P \in \mathcal{P}(\mathcal{X})$, and we let P_n denote an arbitrary type [1, Ch. 2] whose entries are $\frac{1}{n}$-close to P. The set of sequences with type P_n is denoted by $\mathcal{T}(P_n)$.

Fixing n, M_n, an input distribution $P \in \mathcal{P}(\mathcal{X})$, a distance function $d(\cdot,\cdot)$, and constants $\delta > 0, \Delta \in \mathbb{R}$, the construction is described by the following steps:

1) The first codeword, $\boldsymbol{x}_1$, is drawn uniformly over $\mathcal{T}_1(P_n)$, given by $\mathcal{T}_1(P_n) = \mathcal{T}(P_n)$;
2) The second codeword $\boldsymbol{x}_2$ is uniformly drawn from
$$\mathcal{T}_2(P_n,\boldsymbol{x}_1) = \{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n) : d(\bar{\boldsymbol{x}},\boldsymbol{x}_1) > \Delta\}, \tag{2}$$
the set of sequences of composition P_n whose distance to $\boldsymbol{x}_1$ exceeds Δ;
3) The i-th codeword $\boldsymbol{x}_i$ is drawn uniformly from
$$\begin{aligned}&\mathcal{T}_i(P_n,\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1})\\ &= \{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n) : d(\bar{\boldsymbol{x}},\boldsymbol{x}_j) > \Delta,\, j = 1\ldots,i-1\}.\end{aligned} \tag{3}$$

In order to ensure that the above procedure generates the desired number of codewords $M_n = \mathrm{e}^{nR_n}$ (i.e., the sets $\mathcal{T}_i$ are non-empty for $i = 1,\ldots,M_n$), set Δ and δ such that

$$\mathrm{e}^{n(R_n+\delta)}\mathrm{vol}_{\boldsymbol{x}}(\Delta) \le |\mathcal{T}(P_n)|, \tag{4}$$

where

$$\mathrm{vol}_{\boldsymbol{x}}(\Delta) = |\{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n) : d(\bar{\boldsymbol{x}},\boldsymbol{x}) \le \Delta\}| \tag{5}$$

is the volume of a ball of radius Δ according to distance $d(\cdot,\cdot)$ centered at $\boldsymbol{x} \in \mathcal{T}(P_n)$. If the distance d is type-dependent, $\mathrm{vol}_{\boldsymbol{x}}(\Delta)$ does not depend on $\boldsymbol{x} \in \mathcal{T}(P_n)$ since all $\bar{\boldsymbol{x}}$ in (5) belong to the same type class.

Our main result is as follows, namely, a single-letter lower bound for the error exponent of the RGV construction. Let

$$\begin{aligned}&E_{\mathrm{RGV}}(R,P,W,q,d,\Delta) =\\ &\quad\min_{V \in \mathcal{T}_{d,q,P}(\Delta)} D(V_{Y|X}\|W|P) + \big|I(\widetilde{X};Y,X) - R\big|_+,\end{aligned} \tag{6}$$

and

$$\begin{aligned}\mathcal{T}_{d,q,P}(\Delta) \triangleq \Big\{&V_{X\widetilde{X}Y} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}) : V_X = V_{\widetilde{X}} = P,\\ &q(V_{\widetilde{X}Y}) \ge q(V_{XY}),\, d(P_{X\widetilde{X}}) \ge \Delta\Big\}.\end{aligned} \tag{7}$$

Theorem 1. *For all $P \in \mathcal{P}(\mathcal{X})$, $\delta > 0$, $\Delta \in \mathbb{R}$, $d \in \Omega$, and $R > 0$ satisfying*

$$R \le \min_{P_{X\widetilde{X}} : d(P_{X\widetilde{X}}) \le \Delta,\, P_X = P_{\widetilde{X}} = P} I(X;\widetilde{X}) - 2\delta, \tag{8}$$

the ensemble average error probability $\bar{P}_{\mathrm{e}}^{(n)}$ of the RGV construction with parameters (n,R,P,d,Δ,δ) and the continuous type-dependent decoding metric $q(\cdot)$ over DMC W satisfies

$$\bar{P}_{\mathrm{e}}^{(n)} \;\dot{\le}\; \mathrm{e}^{-nE_{\mathrm{RGV}}(R,P,W,q,d,\Delta)}. \tag{9}$$

In addition, if q is an additive decoding metric, then

$$\bar{P}_{\mathrm{e}}^{(n)} \;\dot{\ge}\; \mathrm{e}^{-nE_{\mathrm{RGV}}(R,P,W,q,d,\Delta+\epsilon)} \tag{10}$$

for arbitrarily small $\epsilon > 0$.

While Theorem 1 states the error exponent, the central part of the analysis is in arriving at the following asymptotic expression for the ensemble average probability of error:

$$\begin{aligned}\bar{P}_{\mathrm{e}}^{(n)} \doteq &\sum_{\boldsymbol{x}\in\mathcal{T}(P_n),\boldsymbol{y}} \frac{1}{|\mathcal{T}(P_n)|} W^n(\boldsymbol{y}|\boldsymbol{x})\\ &\cdot \min\Bigg\{1, (M_n-1)\sum_{\substack{\boldsymbol{x}'\in\mathcal{T}(P_n) : q^n(\boldsymbol{x}',\boldsymbol{y}) \ge q^n(\boldsymbol{x},\boldsymbol{y})\\ d(\boldsymbol{x}',\boldsymbol{x}) \ge \Delta}} \frac{1}{|\mathcal{T}(P_n)|}\Bigg\}.\end{aligned} \tag{11}$$

This work was supported in part by the European Research Council under Grant 725411, and the Spanish Ministry of Economy and Competitiveness under Grant TEC2016-78434-C3-1-R.

This can be interpreted as a stronger (albeit asymptotic) analog of the *random coding union* bound [6] that achieves not only the random coding exponent, but also the low-rate improvements of the expurgated exponent.

The following corollary shows that when the distance function $d(\cdot,\cdot)$ is optimized, and Δ is chosen appropriately, the exponent in Theorem 1 recovers the exponent of [7], denoted by $E_q(R,P,W)$, known to be at least as large as the maximum of the random-coding and expurgated exponents.

Corollary 1. *Setting $d(P_{X\widetilde{X}}) = -I(X;\widetilde{X})$, $\Delta = -(R+2\delta)$ gives that for sufficiently small $\delta > 0$ and $\epsilon > 0$*

$$E_{\mathrm{RGV}}(R,P,W,q,d,\Delta) \geq E_q(R,P,W) - \epsilon. \tag{12}$$

Lastly, we show that the non-universal distance function $d(P_{X\tilde{X}}) = \beta_{R,W,q}(P_{X\widetilde{X}})$ also achieves the exponent of Csiszár and Körner, where

$$\beta_{R,W,q}(P_{X\widetilde{X}}) \triangleq \min_{V_{X\widetilde{X}Y} \in \mathcal{T}'(P_{X\widetilde{X}})} \Gamma(V_{X\tilde{X}Y}), \tag{13}$$

with

$$\Gamma(V_{X\tilde{X}Y}) \triangleq D(V_{Y|X} \| W | V_X) + \big| I(\widetilde{X};Y,X) - R \big|_+, \tag{14}$$

and

$$\mathcal{T}'(P_{X\widetilde{X}}) \triangleq \Big\{ V_{X\widetilde{X}Y} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}) : \\ V_{X\widetilde{X}} = P_{X\widetilde{X}}, q(V_{\widetilde{X}Y}) \geq q(V_{XY}) \Big\}. \tag{15}$$

We first provide a corollary characterizing the exponent of Theorem 1 with $d(\cdot) = \beta_{R,W,q}(\cdot)$, and then prove its equivalence to $E_q(R,P,W)$.

Corollary 2. *If the pair (R,Δ) satisfies (8) with $d(\cdot) = \beta_{R,W,q}(\cdot)$, then, the ensemble average error probability $\bar{P}_{\mathrm{e}}^{(n)}$ of the RGV construction with parameters $(n,R,P,\beta_{R,W,q},\Delta,\delta)$ using the continuous type-dependent decoding rule $q(\cdot)$ over the channel W satisfies $\bar{P}_{\mathrm{e}}^{(n)} \mathrel{\dot{\leq}} \mathrm{e}^{-n\Delta}$.*

Proposition 1. *For any $P \in \mathcal{P}(\mathcal{X})$, the achievable rate-exponent pairs (R,E) resulting from Theorem 1 (i.e., taking the union over all $\delta > 0$ and $\Delta > 0$) are identical for the choices $d(P_{X\widetilde{X}}) = -I(X;\widetilde{X})$ and $d(P_{X\widetilde{X}}) = \beta_{R,W,q}(P_{X\widetilde{X}})$.*

References

[1] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[2] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[3] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[4] E. N. Gilbert, "A comparison of signalling alphabets," *Bell Labs Tech. J.*, vol. 31, no. 3, pp. 504–522, 1952.

[5] R. R. Varshamov, "Estimate of the number of signals in error correcting codes," in *Dokl. Akad. Nauk SSSR*, vol. 117, no. 5, 1957, pp. 739–741.

[6] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[7] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, 1981.