Correction to "The Generalized Stochastic Likelihood Decoder: Random Coding and Expurgated Bounds"

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

This purpose of this letter is to handle a gap that was found in the proof of Theorem 2 in the paper "The generalized stochastic likelihood decoder: random coding and expurgated bounds."

1 Introduction

In a recent article [1], random coding error exponents and expurgated exponents were analyzed for the generalized likelihood decoder (GLD), where the decoded message is randomly selected under a probability distribution that is proportional to a general exponential function of the empirical joint distribution of the codeword and the channel output vectors. In Section V of [1], Theorem 2 provides an expurgated exponent which is applicable to this decoder (and hence also to the optimal maximum likelihood decoder). The proof of that theorem is based on two steps of a certain expurgation procedure. Nir Weinberger has brought to my attention that there is a certain gap in that proof, as the second expurgation step might interfere with the first step (more details will follow in Section 2 of this letter). The purpose of this letter is to provide an alternative proof to the above mentioned theorem.

2 Setup and Background

Consider a discrete memoryless channel (DMC), designated by a matrix of single–letter input–output transition probabilities $\{W(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$. Here the channel input symbol x takes

on values in a finite input alphabet $\mathcal{X}$, and the channel output symbol y takes on values in a finite output alphabet $\mathcal{Y}$. When the channel is fed by a vector $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, it outputs a vector $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ according to

$$W(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^{n} W(y_t|x_t). \tag{1}$$

A code $\mathcal{C}_n \subseteq \mathcal{X}^n$ is a collection of $M = e^{nR}$ channel input vectors, $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$, R being the coding rate in nats per channel use. It is assumed that all messages, $m = 0, 1, \ldots M-1$, are equally likely.

As is very common in the information theory literature, we consider the random coding regime. The random coding ensemble considered is the ensemble of constant composition codes, where each codeword is drawn independently under the uniform distribution within a given type class $\mathcal{T}(Q_X)$, i.e., the set of all vectors in $\mathcal{X}^n$ whose empirical distribution is given by Q_X. Once the code has been randomly selected, it is revealed to both the encoder and the decoder.

When the transmitter wishes to convey a message m, it transmits the corresponding code-vector $\boldsymbol{x}_m$ via the channel, which in turn, stochastically maps it into an n–vector $\boldsymbol{y}$ according to (1). Upon receiving $\boldsymbol{y}$, the stochastic generalized likelihood decoder randomly selects the estimated message $\hat{m}$ according to a generalized version of the induced posterior distribution of the transmitted message, i.e.,

$$\Pr\{\hat{m} = m_0|\boldsymbol{y}\} = \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m_0}\boldsymbol{y}})\}}{\sum_{m=0}^{M-1}\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}}, \tag{2}$$

where $\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}$ is the empirical joint distribution induced by $(\boldsymbol{x}_m, \boldsymbol{y})$ and $g(\cdot)$ is an arbitrary continuous function. For example,

$$g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}) = \sum_{x,y}\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}(x,y)\ln W(y|x) \tag{3}$$

corresponds to the ordinary likelihood decoder, where (2) is the correct underlying posterior probability of message m_0. This framework also allows additional important stochastic decoders, where g corresponds to a mismatched metric $\tilde{W}$ or to the empirical mutual information, as discussed in [1].

As mentioned above, in Section V of [1], an expurgated error exponent is derived. Specifically, letting Q_{XY} denote a generic joint distribution over $\mathcal{X}\times\mathcal{Y}$, and letting $I_Q(X;Y)$ denote the mutual information induced by Q_{XY}, we define the following. Let

$$\alpha(R, Q_Y) = \sup_{\{Q_{X|Y}:\ I_Q(X;Y)\le R\}}[g(Q_{XY}) - I_Q(X;Y)] + R, \tag{4}$$

and

$$\Gamma(Q_{XX'}, R) \quad = \quad \inf_{Q_{Y|XX'}}\Big\{D(Q_{Y|X}\|W|Q_X) + I_Q(X';Y|X) +$$

$$[\max\{g(Q_{XY}), \alpha(R, Q_Y)\} - g(Q_{X'Y})]_+\} \tag{5}$$

$$\equiv \inf_{Q_{Y|XX'}} \{\boldsymbol{E}_Q \log[1/W(Y|X)] - H(Y|X, X') + [\max\{g(Q_{XY}), \alpha(R, Q_Y)\} - g(Q_{X'Y})]_+\}, \tag{6}$$

where $D(Q_{Y|X}\|W|Q_X)$ is defined in the usual manner (see also [1]). The main result in [1, Section V] is the following:

Theorem 1 *There exists a sequence of constant composition codes, $\{\mathcal{C}_n,\ n = 1, 2, \ldots\}$, with composition Q_X, such that*

$$\liminf_{n\to\infty} \left[-\frac{\log P_{e|m}(\mathcal{C}_n)}{n}\right] \geq E_{ex}^{gld}(R, Q_X), \tag{7}$$

where

$$E_{ex}^{gld}(R, Q_X) = \inf[\Gamma(Q_{XX'}, R) + I_Q(X; X')] - R, \tag{8}$$

where the infimum is over all joint distributions $\{Q_{XX'}\}$ such that $I_Q(X; X') \leq R$ and $Q_{X'} = Q_X$.

The proof in [1] contains two main steps of expurgation. In the first step, we confine attention to the subset of constant composition codes $\{\mathcal{C}_n\}$ with the property

$$\sum_{m'\neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\} \geq \exp\{n\alpha(R - \epsilon, \hat{P}_{\boldsymbol{y}})\} \quad \forall\ m, \boldsymbol{y} \tag{9}$$

where $\epsilon > 0$ is arbitrarily small. It is proved in [1, Appendix B] that the vast majority of constant composition codes satisfy (9) for large n. In the second expurgation step (see [1, Appendix C]), at most $M \cdot (n+1)^{|\mathcal{X}|^2} e^{-n\epsilon/2}$ "bad" codewords are eliminated from the codebook in order to guarantee the desired maximum error probability performance for the remaining part of the code.

The gap in the proof of [1, Theorem 2] is in the following point: after the second expurgation step, it is no longer guaranteed that eq. (9) still holds for every m and $\boldsymbol{y}$, since the summation on the left–hand side of (9) is now taken over a smaller number of codewords.

Fortunately enough, Theorem 2 of [1] is still correct (as will be proved in the next section in a completely different manner) at least when $g(Q_{XY})$ is an affine functional of Q_{XY}, which is the case of the ordinary matched/mismatched stochastic likelihood decoder (3) with or without a "temperature" parameter (see the discussion around eqs. (5)–(7) of [1]). This affinity assumption is used only at the last step of our derivation below. Thus, when $g(Q_{XY})$ is not affine, one merely backs off from the last step of the derivation, and considers the second to the last expression as the formula of the expurgated exponent.

3 Corrected Proof of [1, Theorem 2]

Assuming that message m was transmitted, the probability of error of the GLD, for a given code $\mathcal{C}_n$, is given by

$$P_{\mathrm{e}|m}(\mathcal{C}_n) = \sum_{m' \neq m} \sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}} \tag{10}$$

and so, for $\rho \geq 1$,

$$[P_{\mathrm{e}|m}(\mathcal{C}_n)]^{1/\rho} \leq \sum_{m' \neq m} \left[\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}} \right]^{1/\rho}, \tag{11}$$

where we have used the inequality $(\sum_i a_i)^s \leq \sum_i a_i^s$, which holds whenever $s \leq 1$ and $a_i \geq 0$ for all i [2, Exercise 4.15(f)]. Let $\mathcal{G}_\epsilon = \mathcal{B}_\epsilon^c$ be defined as in [1], that is, the set of codes for which (9) holds, and consider the fact (proved in Appendix B therein), that $\Pr\{\mathcal{B}_\epsilon\} \leq \exp(-e^{n\epsilon} + n\epsilon + 1)$. We now take the expectation over the randomness of the (incorrect part of the) codebook, $\mathcal{C}_n^m = \mathcal{C}_n \setminus \{\boldsymbol{x}_m\}$ (where all wrong codewords are drawn from a given type Q_X), except $\boldsymbol{x}_m$, which is kept fixed for now. When dealing with the pairwise error probability from m to m', we do this in two steps: we first average over all codewords except $\boldsymbol{x}_m$ and $\boldsymbol{x}_{m'}$, and then average over the randomness of $\boldsymbol{x}_{m'}$.

$$\begin{aligned}
& \boldsymbol{E}\left\{ [P_{\mathrm{e}|m}(\mathcal{C}_n)]^{1/\rho} \middle| \boldsymbol{x}_m \right\} \\
\leq\ & \sum_{m' \neq m} \sum_{\mathcal{C}_n^m} P(\mathcal{C}_n^m) \left[\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}} \right]^{1/\rho} \\
=\ & \sum_{m' \neq m} \sum_{\mathcal{C}_n^m \in \mathcal{G}_\epsilon} P(\mathcal{C}_n^m) \left[\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}} \right]^{1/\rho} + \\
& \sum_{m' \neq m} \sum_{\mathcal{C}_n^m \in \mathcal{B}_\epsilon} P(\mathcal{C}_n^m) \left[\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \sum_{m' \neq m} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}} \right]^{1/\rho} \\
\leq\ & \sum_{m' \neq m} \sum_{\mathcal{C}_n^m \in \mathcal{G}_\epsilon} P(\mathcal{C}_n^m) \left[\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \min\left\{ 1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}} \right\} \right]^{1/\rho} + \\
& \sum_{m' \neq m} \sum_{\mathcal{C}_n^m \in \mathcal{B}_\epsilon} P(\mathcal{C}_n^m) \cdot 1^{1/\rho} \\
\leq\ & \sum_{m' \neq m} \sum_{\mathcal{C}_n^m} P(\mathcal{C}_n^m) \left[\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \min\left\{ 1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}} \right\} \right]^{1/\rho} + \\
& e^{nR} \cdot \exp(-e^{n\epsilon} + n\epsilon + 1) \\
\dot{\leq}\ & \sum_{m' \neq m} \boldsymbol{E}\left(\left[\sum_{\boldsymbol{y}} W(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \min\left\{ 1, \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}}\boldsymbol{y})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m}\boldsymbol{y})\} + \exp\{n\alpha(R-\epsilon, \hat{P}_{\boldsymbol{y}})\}} \right\} \right]^{1/\rho} \middle| \boldsymbol{x}_m \right)
\end{aligned}$$

$$
\begin{aligned}
&\doteq \sum_{m' \neq m} \boldsymbol{E}\left\{\exp[-n\Gamma(\hat{P}_{\boldsymbol{x}_m \boldsymbol{x}'_m})/\rho] \middle| \boldsymbol{x}_m\right\} \\
&\doteq \sum_{Q_{X'|X}} \boldsymbol{E}\{N_m(Q_{X'|X})|\boldsymbol{x}_m\} \exp\{-n\Gamma(Q_{XX'})/\rho\} \\
&\doteq \max_{Q_{X'|X}} \exp\{n[R - I_Q(X;X')]\} \cdot \exp\{-n\Gamma(Q_{XX'})/\rho\} \\
&= \exp\left\{-n \min_{Q_{X'|X}} [\Gamma(Q_{XX'})/\rho + I_Q(X;X') - R]\right\}, \qquad (12)
\end{aligned}
$$

where $I_Q(X;X')$ is the mutual information induced by $Q_{XX'}$ and $N_m(Q_{X'|X}) = |\mathcal{T}(Q_{X'|X}|\boldsymbol{x}_m) \cap \mathcal{C}_m|$, $\mathcal{T}(Q_{X'|X}|\boldsymbol{x}_m)$ being the conditional type class pertaining to $Q_{X'|X}$ given $\boldsymbol{x}_m$. Since this bound is independent of $\boldsymbol{x}_m$, it also holds for the unconditional expectation, $\boldsymbol{E}[P_{e|m}(\mathcal{C}_n)]^{1/\rho}$. Now, for a given code $\mathcal{C}_n$, index the message $\{m\}$ according to decreasing order of $\{P_{e|m}(\mathcal{C}_n)\}$. Then,

$$
\frac{1}{M}\sum_{m=1}^{M}[P_{e|m}(\mathcal{C}_n)]^{1/\rho} \geq \frac{1}{M}\sum_{m=1}^{M/2}[P_{e|m}(\mathcal{C}_n)]^{1/\rho} \geq \frac{1}{M}\cdot\frac{M}{2}[P_{e|M/2}(\mathcal{C}_n)]^{1/\rho} = \frac{1}{2}\cdot[\max_m P_{e|m}(\mathcal{C}'_n)]^{1/\rho}, \quad (13)
$$

where $\mathcal{C}'_n$ is the good half of $\mathcal{C}_n$. Thus,

$$
\begin{aligned}
\boldsymbol{E}\left\{[\max_m P_{e|m}(\mathcal{C}'_n)]^{1/\rho}\right\} &\leq 2\boldsymbol{E}\left\{\frac{1}{M}\sum_{m=1}^{M} P_{e|m}(\mathcal{C}_n)^{1/\rho}\right\} \\
&\doteq \exp\left\{-n \min_{Q_{X'|X}} [\Gamma(Q_{XX'})/\rho + I_Q(X;X') - R]\right\} \qquad (14)
\end{aligned}
$$

which means that there exists a code of size $M/2$ with

$$
[\max_m P_{e|m}(\mathcal{C}'_n)]^{1/\rho} \leq \exp\left\{-n \min_{Q_{X'|X}} [\Gamma(Q_{XX'})/\rho + I_Q(X;X') - R]\right\}, \qquad (15)
$$

or equivalently,

$$
\max_m P_{e|m}(\mathcal{C}'_n) \leq \exp\left(-n \min_{Q_{X'|X}} \{\Gamma(Q_{XX'}) + \rho[I_Q(X;X') - R]\}\right), \qquad (16)
$$

and since this holds for every $\rho \geq 1$, we have

$$
\max_m P_{e|m}(\mathcal{C}'_n) \leq \exp\left(-n \sup_{\rho \geq 1} \min_{Q_{X'|X}} \{\Gamma(Q_{XX'}) + \rho[I_Q(X;X') - R]\}\right). \qquad (17)
$$

Now, consider the exponent,

$$
\begin{aligned}
E_{\mathrm{ex}}(R, Q_X) &\stackrel{\Delta}{=} \sup_{\rho \geq 1} \min_{Q_{X'|X}} \{\Gamma(Q_{XX'}) + \rho[I_Q(X;X') - R]\} \qquad (18) \\
&= \sup_{\rho \geq 0} \min_{Q_{XX'}} \{\Gamma(Q_{XX'}) + I_Q(X;X') - R + \rho[I_Q(X;X') - R]\}, \qquad (19)
\end{aligned}
$$

where the marginals of $Q_{XX'}$ are constrained to the given fixed composition, Q_X. Using the definitions in [1],

$$\begin{aligned}\Gamma(Q_{XX'})+I_Q(X;X') &= \inf_{Q_{Y|XX'}}\{-\boldsymbol{E}_Q\ln W(Y|X)-H(Y|X,X')+\\ &\quad I_Q(X;X')+[\max\{g(Q_{XY}),\alpha(R,Q_Y)\}-g(Q_{X'Y})]_+\}\\ &= \inf_{Q_{Y|XX'}}\{-\boldsymbol{E}_Q\ln[W(Y|X)Q(X)Q(X')]-H_Q(X,X',Y)+\\ &\quad +[\max\{g(Q_{XY}),\alpha(R,Q_Y)\}-g(Q_{X'Y})]_+\}, \end{aligned}\tag{20}$$

thus,

$$\begin{aligned}&\min_{Q_{XX'}}\{\Gamma(Q_{XX'})+I_Q(X;X')-R+\rho[I_Q(X;X')-R]\}\\ =\ &\min_{Q_{XX'Y}}\{-\boldsymbol{E}_Q\ln[W(Y|X)Q(X)Q(X')]-H_Q(X,X',Y)+\\ &\rho[I_Q(X;X')-R]+[\max\{g(Q_{XY}),\alpha(R,Q_Y)\}-g(Q_{X'Y})]_+\}. \end{aligned}\tag{21}$$

Now, the first term on the right–most side is linear (and hence convex) in $Q_{XX'Y}$ since Q_X is fixed, the second term is convex, and the third term is convex for a given Q_X. As for the fourth term, it is convex at least in the case where g is affine in Q (e.g., matched/mismatched likelihood metric with/without a temperature parameter) because the function $f(x)=[x]_+$ is monotonic and convex and we argue that $\alpha(R,Q_Y)$ is also convex since it is given by the supremum over a family of convex functions of Q_Y (as g is linear and $-I_Q(X;X')$ is convex in Q_Y for a given $Q_{X|Y}$). The maximum between two convex functions is convex. Since the objective is affine (and hence concave) in ρ, we can interchange the minimization and the maximization to obtain,

$$\begin{aligned}E_{\rm ex}(R,Q_X) &= \inf_{Q_{XX'}}\left\{\Gamma(Q_{XX'})+I_Q(X;X')-R+\sup_{\rho\geq 0}\rho[I_Q(X;X')-R]\right\}\\ &= \inf_{\{Q_{XX'}:\ I_Q(X;X')\leq R\}}[\Gamma(Q_{XX'})+I_Q(X;X')-R]\\ &= E_{\rm ex}^{\rm gld}(R,Q_X). \end{aligned}\tag{22}$$

If the supremum and the minimum cannot be interchanged, then, of course, the formula of the expurgated exponent remains as in (18).

Acknowledgment

I would like to thank Dr. Nir Weinberger for drawing my attention to the gap in the proof of Theorem 2 in [1].

References

[1] N. Merhav, "The generalized stochastic likelihood decoder: random coding and expurgated bounds," *IEEE Trans. Inform. Theory*, vol. 63, no. 8, pp. 5039–5051, August 2017.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.