USING THE METHOD OF NEARBY MEASURES IN SUPERPOSITION CODING WITH A BERNOULLI DICTIONARY

Cynthia Rush and Andrew R. Barron,

Department of Statistics, Yale University, New Haven, CT 06511
email: {cynthia.rush, andrew.barron}@yale.edu

ABSTRACT

The work discussed in this presentation is motivated by the problem of decoding superposition codes when using $\{-1, +1\}$-valued dictionary elements. Under this encoding scheme, the output, Y, is a linear combination of independent Bernoulli random variables and noise. Statistical decoding then requires the study of the conditional distribution of the Bernoulli random variables given the output, Y. We find this distribution analogous to that of summands given the sum of independent random variables as studied by Cover and Campenhout and Csiszár. Motivated by this work we wish to bound the Rényi relative entropy distance between the true conditional distribution and the product of tilted distributions in order to show that events that are exponentially rare under the tilted product distribution remain rare under the true distribution.

1. INTRODUCTION

Distributional analysis of the summands given the sum of independent random variables arises in science and engineering applications. Superposition encoding over the Gaussian white noise channel, discussed here, is one such example. This conditional distribution has been studied extensively in both statistical mechanics and mathematical statistics, and we now know that given the sum, the summands are approximately independent with exponentially tilted distributions. We present the method of nearby measures as way to use this information. By bounding the Rényi relative entropy between the true distribution and the independent, exponentially tilted distribution it can be shown that events which are rare under the approximate distribution are also rare under the true distribution, allowing for calculations to be computed using the approximate, usually much simpler, distribution. The following section discusses background information on superposition coding, Section 3 presents the method of nearby measures and the Rényi relative entropy, Section 4 establishes upper-bounds on the Shannon relative entropy and the Rényi relative entropy of order α between the true distribution and the approximate distribution, and finally a short conclusion is supplied in Section 5. All Theorem proofs are collected in the appendices.

2. BACKGROUND

When using superposition encoding, codewords are sparse linear combinations of L out of $N = LM$ columns of a dictionary X of size $n \times N$ for which, in our case, each entry is an independent Bernoulli $\{-1, +1\}$-valued random variable,

$$\text{codeword} = \sum_{l=1}^{L} \sqrt{P_l} X_l \tag{1}$$

where P_l are weights such that $\sum_{l=1}^{L} P_l = P$ with power constraint P and X_l is the l^{th} selected column of X. By sparse, we mean that the subset of L columns used in the codeword is much smaller than the total number of columns N of the dictionary. The input specifies which L columns are used to form the codeword.

When sending information across an additive Gaussian white noise channel, the output Y is an $n \times 1$ column vector

$$Y = \sum_{l=1}^{L} \sqrt{P_l} X_l + \epsilon \tag{2}$$

for $\epsilon \sim N(0, \sigma^2 \mathbb{I}_{n\times n})$. Decoding using the adaptive successive decoder, developed by Joseph and Barron [1], requires the study of the conditional distribution of the columns $(X_1, ..., X_L)$ given the output Y. Notice that each entry in Y is independent of the others, so we focus on a single row of the output,

$$Y_i = \sum_{l=1}^{L} \sqrt{P_l} X_{i,l} + \epsilon_i \tag{3}$$

and the conditional distribution of $(X_{i,1}, ..., X_{i,L})$ given the output Y_i. In what follows we drop the subscript i when discussing the one-dimensional random variables $X_1, ..., X_L, Y$.

3. THE NEARBY MEASURE

The distribution of summands given the sum of independent random variables has been studied extensively in statistical mechanics motivated by the original work of Boltzmann (see, for example, Lanford [2]) and others as well as in statistics and information theory by Cover and van Campenhout [3], Csiszár [4], and others. This work states that conditionally given the sum, the summands are distributed approximately independently according to the the maximum entropy distribution subject to the mean constraint. This takes the form of exponentially tilted distributions. This situation is analogous to the statistical decoding problem involving the conditional distribution of $(X_1, ..., X_L)$ given the output Y. Motivated by this work, we hope to be able to approximate the true conditional distribution of $(X_1, ..., X_L)$ given Y by the product of independent, exponentially tilted Bernoulli ± 1 distributions, meaning that if an event is rare under the approximate distribution then it remains rare under the true distribution.

Consider the true distribution of $X_1, ..., X_L$ which we define by independent q such that

$$q(x_l) = \begin{cases} \frac{1}{2}, & \text{if } x_l = 1 \\ \frac{1}{2}, & \text{if } x_l = -1 \end{cases} \tag{4}$$

for each $l \in (1, 2, ..., L)$. Observation of Y gives rise to $q_{X|Y}$ which we approximate as the product of independent, exponentially tilted distributions. We let $q^a_{X|Y}$ be the tilted distributions given Y. For each $l \in (1, 2, ..., L)$,

$$q^a_{X_l|Y}(x_l) = \begin{cases} \frac{e^{aY\sqrt{P_l}}}{e^{aY\sqrt{P_l}}+e^{-aY\sqrt{P_l}}}, & \text{if } x_l = 1 \\ \frac{e^{-aY\sqrt{P_l}}}{e^{aY\sqrt{P_l}}+e^{-aY\sqrt{P_l}}}, & \text{if } x_l = -1 \end{cases} \tag{5}$$

where a is an appropriate constant. Let us define $\mathbb{Q}_L$ as the measure associated with the true joint distribution (joint probability mass function) of $(X_1, ..., X_L, Y)$. Similarly, let $\mathbb{Q}^a_L$, a for approximate, be the measure associated with the joint probability distribution of $(X_1, ..., X_L, Y)$ when the conditional distribution of $(X_1, ..., X_L)$ given Y is the product of exponentially tilted distributions. Finally, let q_L and q^a_L be the probability distributions associated with each measure. Specifically q^a_L is the product

$$p_Y(y) \prod_{l=1}^{L} q^a_{X_l|Y}(x_l)$$

where $p_Y(y)$ is the probability mass function of Y. Let us define the Rényi relative entropy of order $\alpha > 1$ between these measures, denoted $D_\alpha(\mathbb{Q}_L||\mathbb{Q}^a_L)$, as

$$\frac{1}{\alpha-1} \log \mathbb{E}_{\mathbb{Q}_L} \left(\frac{q_L(X_1, ..., X_L, Y)}{q^a_L(X_1, ..., X_L, Y)} \right)^{\alpha-1}. \tag{6}$$

If $D_\alpha(\mathbb{Q}_L||\mathbb{Q}^a_L)$ is finite for some $\alpha > 1$, then we can relate probabilities under the true measure $\mathbb{Q}_L$ to probabilities under the approximate measure $\mathbb{Q}^a_L$. This relationship is summarized in the following Lemma.

Lemma 3.1 *Consider an event A. If the Rényi relative entropy between the two measures is finite for some order $\alpha > 1$, meaning $D_\alpha(\mathbb{Q}_L||\mathbb{Q}^a_L) \le c_0$ for some constant c_0, then the probability of the event under the true measure is upper-bounded using the probability under the approximate distribution with the following inequality.*

$$\mathbb{Q}_L(A) \le \left(e^{c_0}\mathbb{Q}^a_L(A)\right)^{\frac{\alpha-1}{\alpha}}. \tag{7}$$

Proof

$$\begin{aligned} \mathbb{Q}_L(A) &= \int q_L(\underline{x}) \mathbb{1}_{\{\underline{x}\in A\}} d\underline{x}, && (8)\\ &= \int \frac{q_L(\underline{x})}{q^a_L(\underline{x})} q^a_L(\underline{x}) \mathbb{1}_{\{\underline{x}\in A\}} d\underline{x}, && (9)\\ &\le \left(\int \frac{(q_L(\underline{x}))^\alpha}{q^a_L(\underline{x})} d\underline{x} \right)^{\frac{1}{\alpha}} \left(\mathbb{Q}^a_L(A)\right)^{\frac{\alpha-1}{\alpha}}, && (10)\\ &= \left(e^{D_\alpha(\mathbb{Q}_L||\mathbb{Q}^a_L)} \mathbb{Q}^a_L(A) \right)^{\frac{\alpha-1}{\alpha}}. && (11) \end{aligned}$$

Step (10) follows from Holder's inequality. ∎

In the following section we demonstrate how to obtain bounds for both the Shannon relative entropy and the Rényi relative entropy between the two measures for all signal-to-noise ratios.

4. BOUNDING THE RÉNYI RELATIVE ENTROPY

The Shannon relative entropy between the true distribution $\mathbb{Q}_L$ and the approximate distribution $\mathbb{Q}^a_L$ is defined to be

$$D(\mathbb{Q}_L||\mathbb{Q}^a_L) = \mathbb{E}_{\mathbb{Q}_L} \log \frac{q_L(X_1, ..., X_L, Y)}{q^a_L(X_1, ..., X_L, Y)} \tag{12}$$

using the base e logarithm. This is also the Rényi relative entropy of order $\alpha = 1$. Because the Rényi relative entropy is continuous in α, the upper-bound for α just above 1 should be close to the Shannon entropy between the two measures. Before we demonstrate the bound acquired for the Rényi relative entropy, we show that the Shannon relative entropy upper-bound is finite for all signal-to-noise ratios where we define the signal-to-noise ratio to be

$$snr = \frac{P}{\sigma^2}.$$

4.1. Shannon entropy upper-bound

Consider the true joint distribution $q_L(x_1, ..., x_L, y)$ which equals

$$q(x_1)...q(x_L)\phi_\epsilon\left(y - \sum_{l=1}^{L}\sqrt{P_l}x_l\right), \tag{13}$$

where ϕ_ϵ is the probability mass function associated with $\epsilon \sim N(0, \sigma^2)$, and the approximate joint distribution

$$\begin{aligned} q^a_L(x_1, ..., x_L, y) &= q^a_{X_1|Y}(x_1)...q^a_{X_L|Y}(x_L)p_Y(y), \\ &= \prod_{l=1}^{L} \left(\frac{q(x_l)e^{aY\sqrt{P_l}x_l}}{\frac{1}{2}e^{aY\sqrt{P_l}} + \frac{1}{2}e^{-aY\sqrt{P_l}}} \right) p_Y(y). \end{aligned}$$

The following theorem provides an upper-bound for the Shannon relative entropy between these two distributions.

Theorem 4.1 *For any constant a,*

$$D(\mathbb{Q}_L||\mathbb{Q}_L^a) \leq \frac{1}{2}\log(1+snr) + \frac{1}{2}a^2P(\sigma^2+P) - aP, \quad (14)$$

which is minimized by choosing $a = \frac{1}{\sigma^2+P}$ making the upper-bound

$$D(\mathbb{Q}_L||\mathbb{Q}_L^a) \leq \frac{1}{2}\log(1+snr) - \frac{snr}{2(1+snr)}. \quad (15)$$

We prove Theorem 4.1 in Appendix 1. Notice that the upper-bound stated in Theorem 4.1 is positive for all values of snr, as we would expect of the Shannon relative entropy. This is can be seen by remembering that $\log(1+x) \leq x$ for all $x > -1$ and so

$$\log(1+snr) = -\log(1-\frac{snr}{1+snr}) \geq \frac{snr}{1+snr}. \quad (16)$$

We next demonstrate bounds for the Rényi relative entropy.

4.2. Rényi relative entropy upper-bound

We first choose work with the Rényi relative entropy of order $\alpha = 2$ for simplicity. The order $\alpha = 2$ relative entropy is defined to be

$$D_2(\mathbb{Q}_L||\mathbb{Q}_L^a) = \log \mathbb{E}_{\mathbb{Q}_L}\left(\frac{q_L(X_1,...,X_L,Y)}{q_L^a(X_1,...,X_L,Y)}\right). \quad (17)$$

The following Theorem upper-bounds this relative entropy. The proof can be found in Appendix 2.

Theorem 4.2 *For any $snr \leq .58$, there exists a range of γ values in the interval $0 < \gamma < 1 - \frac{snr}{(1+snr)^2}$ such that*

$$D_2(\mathbb{Q}_L||\mathbb{Q}_L^a) \leq \log\frac{20}{3} + \left(1+\frac{1}{\gamma}\right)2snr - \frac{1}{2}\log\left(1-\gamma-\frac{snr}{(1+snr)^2}\right). \quad (18)$$

While the Shannon entropy upper-bound held for all snr, the Rényi relative entropy upper-bound at order $\alpha = 2$ is limited to only small signal-to-noise ratios. In allowing α to approach 1, the Rényi relative entropy approaches the Shannon relative entropy, and bounds are obtained for all signal-to-noise ratios. The following Theorem bounds the Rényi relative entropy for all values of the signal-to-noise ratio by allowing α to be arbitrarily small.

Theorem 4.3 *For any snr and any γ in the range $0 < \gamma < \frac{1}{2}$, there exists a $\delta = \alpha - 1 > 0$ such that*

$$D_\alpha(\mathbb{Q}_L||\mathbb{Q}_L^a) \leq \log\frac{4(5)^{1/\delta}}{3} + \left(1+\frac{1}{\gamma}\right)2snr - \frac{1}{2\delta}\log(\delta(1-\gamma-a^2\sigma^2P)). \quad (19)$$

The proof of Theorem 4.3 can be found in Appendix 3. Using this bound and Lemma 3.1, we are able to demonstrate an upper-bound on the error accrued when approximating the true distribution with the tilted distribution.

5. CONCLUSION

Using knowledge of the distributional behavior of the summands given the sum of independent random variables, and the closeness of measures established by finite Rényi relative entropy, we are able to approximate a distribution which is statistically difficult to analyze with a much simpler distribution with a constant error rate, thus simplifying statistical decoding of superposition coding over the Gaussian white noise channel.

6. APPENDIX 1 PROOF OF THEOREM 4.1

We want to consider the expectation under the true joint distribution of the log ratio of the distributions,

$$\mathbb{E}\log\frac{q(X_1)...q(X_L)\phi_\epsilon(Y-\sum_{l=1}^L\sqrt{P_l}X_l)}{q^a_{X_1|Y}(X_1)...q^a_{X_L|Y}(X_L)p_Y(Y)}. \quad (20)$$

Letting the codeword, $\sum_{l=1}^L\sqrt{P_l}X_l$, be called W, expression (20) can be simplified to

$$\mathbb{E}\log\frac{\phi_\epsilon(Y-W)\prod_{l=1}^L\left(\frac{1}{2}e^{aY\sqrt{P_l}}+\frac{1}{2}e^{-aY\sqrt{P_l}}\right)}{e^{aYW}p_Y(Y)}. \quad (21)$$

We use the following Lemma.

Lemma 6.1 *For any value $x \in \mathbb{R}$,*

$$\frac{1}{2}e^x + \frac{1}{2}e^{-x} \leq e^{\frac{x^2}{2}}.$$

Proof Using the MacLaurin expansion $e^x = \sum_{k=0}^\infty \frac{1}{k!}x^k$,

$$\begin{aligned}\frac{1}{2}e^x + \frac{1}{2}e^{-x} &= \frac{1}{2}\sum_{k=0}^\infty\frac{1}{k!}(x^k+(-x)^k), &(22)\\ &= \sum_{k'=0}^\infty\frac{1}{(2k')!}x^{2k'}, &(23)\\ &\leq \sum_{k'=0}^\infty\frac{1}{k'!}\left(\frac{x^2}{2}\right)^{k'}, &(24)\\ &= e^{\frac{x^2}{2}}, &(25)\end{aligned}$$

with Step (24) following from the fact that $(2k)! \geq 2^k k!$.

Using Lemma 6.1, an upper-bound for (21) is given by

$$-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\mathbb{E}(Y-W)^2 + \frac{a^2P}{2}\mathbb{E}Y^2 - a\mathbb{E}YW - \mathbb{E}\log p_Y(Y). \quad (26)$$

Finally, we know that $-\mathbb{E}\log p_Y(Y)$ is the entropy of Y which is less than or equal to the entropy of a normal random variable with the same variance (see, for example Thomas and Cover [5]). This means that

$$-\mathbb{E}\log p_Y(Y) \leq \frac{1}{2}\log(2\pi(\sigma^2+P)) + \frac{1}{2}. \qquad (27)$$

Applying this to bound (26) and taking the expectation of the remaining terms gives the desired upper-bound

$$-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2} + \frac{a^2(\sigma^2+P)P}{2} - aP + \frac{1}{2}\log(2\pi(\sigma^2+P)) + \frac{1}{2}. \qquad (28)$$

Calculating the minimizing value of a is straightforward.

7. APPENDIX 2 PROOF OF THEOREM 4.2

We want to consider the log of the expectation under the true joint distribution of the ratio of the distributions,

$$\log \mathbb{E}\frac{q(X_1)...q(X_L)\phi_\epsilon(Y-\sum_{l=1}^L\sqrt{P_l}X_l)}{q^a_{X_1|Y}(X_1)...q^a_{X_L|Y}(X_L)p_Y(Y)}. \qquad (29)$$

Again we let the codeword, $\sum_{l=1}^L\sqrt{P_l}X_l$, be called W, and then expression (29) can be simplified to

$$\log \mathbb{E}\frac{\phi_\epsilon(Y-W)\prod_{l=1}^L\left(\frac{1}{2}e^{aY\sqrt{P_l}}+\frac{1}{2}e^{-aY\sqrt{P_l}}\right)}{e^{aYW}p_Y(Y)}. \qquad (30)$$

Applying Lemma 6.1, we find that (30) is upper-bounded by

$$\log \mathbb{E}\frac{\phi_\epsilon(Y-W)e^{\frac{a^2Y^2P}{2}}}{e^{aYW}p_Y(Y)}. \qquad (31)$$

We make use of the following Lemma, which is a generalization of a result given by Brown [6], to upper-bound the probability density function of Y.

Lemma 7.1 *Let Y be defined as $Y = W + \epsilon$ where W is the codeword $\sum_{l=1}^L\sqrt{P_l}X_l$ and $\epsilon \sim N(0,\sigma^2)$, then for any $\gamma > 0$*

$$P_Y(Y) \geq \frac{3}{4}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(1+\gamma)}{2\sigma^2}Y^2}e^{-(1+\frac{1}{\gamma})2snr}.$$

Proof We first supply a quick proof of the inequality $(A+B)^2 \leq (1+\gamma)A + (1+\frac{1}{\gamma})B$, for $\gamma > 0$. Notice that $(A+B)^2 = A^2+B^2+2AB$, so it suffices to show that $2AB \leq \gamma A + \frac{1}{\gamma}B$. We know that this is true since $0 \leq (\sqrt{\gamma}A - \frac{1}{\sqrt{\gamma}B})^2 = \gamma A + \frac{1}{\gamma}B - 2AB$.

Notice that by Chebyshev's Inequality,

$$\mathbb{P}(|W| \leq 2\sqrt{P}) \geq \frac{3}{4}.$$

We use both these facts below.

$$\begin{aligned} P_Y(Y) &= \int_{-\infty}^{\infty} p_W(s)\phi_\epsilon(Y-s)ds, & (32)\\ &\geq \int_{-2\sqrt{P}}^{2\sqrt{P}} p_W(s)\phi_\epsilon(Y-s)ds, & (33)\\ &\geq \frac{3}{4}\min_{s:|s|\leq 2\sqrt{P}}\phi_\epsilon(Y-s), & (34)\\ &\geq \frac{3}{4}\phi_\epsilon(|Y|+2\sqrt{P}), & (35)\\ &= \frac{3}{4}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-1}{2\sigma^2}(|Y|+2\sqrt{P})^2}, & (36)\\ &\geq \frac{3}{4}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-1}{2\sigma^2}((1+\gamma)Y^2+(1+\frac{1}{\gamma})4P)}. & (37)\end{aligned}$$

∎

Applying Lemma 7.1 gives the following upper-bound to (31).

$$\log\frac{4e^{(1+\frac{1}{\gamma})2snr}}{3}\mathbb{E}\frac{e^{\frac{-1}{2\sigma^2}(Y-W)^2}e^{\frac{a^2Y^2P}{2}}}{e^{aYW}e^{\frac{-(1+\gamma)}{2\sigma^2}Y^2}}. \qquad (38)$$

Using the fact that the expectation of the true distribution equals the expectation taken first over $Y|W$ and then over W, the expectation in expression (38) equals

$$\mathbb{E}_W e^{\frac{-1}{\sigma^2}W^2}\int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma^2}[y^2(1-\gamma-a^2\sigma^2P)-2yW(2-a\sigma^2)]}dy, \qquad (39)$$

and so we simplify expression (38) using the representation in (39) to give

$$\log\frac{4e^{(1+\frac{1}{\gamma})2snr}}{3\sqrt{1-\gamma-a^2\sigma^2P}}\mathbb{E}_W e^{W^2\left(\frac{1}{2\sigma^2}\left(\frac{(2-a\sigma^2)^2}{1-\gamma-a^2\sigma^2P}\right)-\frac{1}{\sigma^2}\right)}. \qquad (40)$$

We must now impose the restriction $0 < \gamma < 1 - a^2\sigma^2P$ to be sure that the integral in expression (39) is finite. Letting

$$c^* = \frac{1}{2\sigma^2}\left(\frac{(2-\frac{1}{1+snr})^2}{1-\gamma-\frac{snr}{(1+snr)^2}}\right) - \frac{1}{\sigma^2},$$

we wish to upper-bound $\mathbb{E}_W e^{c^*W^2}$. The following Lemma, a result from Pollard [7], is used to supply an upper-bound for this expectation.

Lemma 7.2 *For a random variable Z, if $\mathbb{E}e^{tZ} \leq e^{\frac{c^2t^2}{2}}$ for some constant c and for all real t, then for all $\tilde{c} \geq c$,*

$$\mathbb{E}e^{\frac{Z^2}{4\tilde{c}^2}} \leq 5. \qquad (41)$$

Proof

$$\mathbb{E}e^{\frac{Z^2}{4c^2}} - 1 = \mathbb{E}\int_0^\infty \mathbb{1}\left\{0 \le t \le \frac{Z^2}{4c^2}\right\} e^t dt, \quad (42)$$
$$\le \int_0^\infty \mathbb{E}e^{\frac{|Z|\sqrt{t}}{c} - t} dt, \quad (43)$$
$$\le \int_0^\infty \mathbb{E}\left(e^{\frac{Z\sqrt{t}}{c}} + e^{\frac{-Z\sqrt{t}}{c}}\right) e^{-t} dt, \quad (44)$$
$$\le 2\int_0^\infty e^{\frac{-t}{2}} dt, \quad (45)$$
$$= 4. \quad (46)$$

where step (43) follows by Markov's inequality and (45) follows from the fact that $\mathbb{E}e^{tZ} \le e^{\frac{c^2t^2}{2}}$ for all t. ▮

Using Lemma 7.2, we see that because $\mathbb{E}e^{tW} \le e^{\frac{t^2P}{2}}$ (by Lemma 6.1), then $\mathbb{E}e^{\frac{W^2}{4P}} \le 5$. Therefore, whenever $c^* \le \frac{1}{4P}$ the expectation in expression (40) is upper-bounded by 5. We will show that for any $snr < .58$, there exists a range of γ values in the interval $0 < \gamma < 1 - a^2\sigma^2 P$, which will make $c^* < \frac{1}{4P}$. To see this notice that $c^* < \frac{1}{4P}$ whenever

$$\frac{snr}{2}\left(\frac{(1+2snr)^2}{(1-\gamma)(1+2snr+snr^2) - snr}\right) - snr < \frac{1}{4}. \quad (47)$$

For $\gamma = 0$, the left-hand side of (47) equals $\frac{1}{4}$ when $snr \approx .58$. Therefore, for snr values strictly less than .58, there exists a range of γ values close to 0 making the inequality hold. Then upper-bound from (40) is then

$$\log\frac{20}{3\sqrt{1-\gamma-a^2\sigma^2P}} + \left(1 + \frac{1}{\gamma}\right)2snr. \quad (48)$$

8. APPENDIX 3 PROOF OF THEOREM 4.3

Proof Remember that the Rényi relative entropy of order α equals

$$\frac{1}{\alpha - 1}\log \mathbb{E}_{\mathbb{Q}_L}\left(\frac{q_L(X_1, ..., X_L, Y)}{q_L^a(X_1, ..., X_L, Y)}\right)^{\alpha-1}. \quad (49)$$

Letting $\delta = \alpha - 1$, we want to obtain an upper-bound for

$$\frac{1}{\delta}\log \mathbb{E}\left(\frac{q(X_1)...q(X_L)\phi_\epsilon(Y - \sum_{l=1}^L \sqrt{P_l}X_l)}{q^a_{X_1|Y}(X_1)...q^a_{X_L|Y}(X_L)p_Y(Y)}\right)^\delta. \quad (50)$$

Again let the codeword, $\sum_{l=1}^L \sqrt{P_l}X_l$, be called W, and then (50) can be simplified to

$$\frac{1}{\delta}\log \mathbb{E}\left(\frac{\phi_\epsilon(Y - W)\prod_{l=1}^L\left(\frac{1}{2}e^{aY\sqrt{P_l}} + \frac{1}{2}e^{-aY\sqrt{P_l}}\right)}{e^{aYW}p_Y(Y)}\right)^\delta. \quad (51)$$

Applying Lemma 6.1 and Lemma 7.1, expression (51) is bounded by

$$\frac{1}{\delta}\log\frac{4^\delta e^{\delta(1+\frac{1}{\gamma})2snr}}{3^\delta}\mathbb{E}\frac{e^{\frac{-\delta}{2\sigma^2}(Y-W)^2}e^{\frac{\delta a^2Y^2P}{2}}}{e^{\delta aYW}e^{\frac{-\delta(1+\gamma)}{2\sigma^2}Y^2}}. \quad (52)$$

Using the fact that the expectation of the true distribution equals the expectation taken first over $Y|W$ and then over W, the expectation in expression (52) equals

$$\mathbb{E}_W e^{\frac{-\delta}{\sigma^2}W^2}\int_{-\infty}^\infty e^{\frac{-\delta}{2\sigma^2}[y^2(1-\gamma-a^2\sigma^2P) - 2yW(2-a\sigma^2)]}dy. \quad (53)$$

Using the representation in (53), expression (52) can be simplified to

$$\frac{1}{\delta}\log\frac{4^\delta e^{\delta(1+\frac{1}{\gamma})2snr}}{3^\delta\sqrt{\delta(1-\gamma-a^2\sigma^2P)}}\mathbb{E}_W e^{c^*_\delta W^2}, \quad (54)$$

where

$$c^*_\delta = \frac{\delta}{2\sigma^2}\left(\frac{(2-a\sigma^2)^2}{1-\gamma-a^2\sigma^2P}\right) - \frac{\delta}{\sigma^2}.$$

We must again impose the restriction $0 < \gamma < 1 - a^2\sigma^2P$ to be sure that the integral in (53) is finite. Another appeal to Lemma 7.2 is made in order to obtain an upper-bound for $\mathbb{E}_W e^{c^*_\delta W^2}$. We see that because $\mathbb{E}e^{tW} \le e^{\frac{t^2P}{2}}$ (by Lemma 6.1), then $\mathbb{E}e^{\frac{W^2}{4P}} \le 5$. Therefore, whenever $c^*_\delta \le \frac{1}{4P}$ the expectation in (54) is bounded by 5. This occurs whenever

$$\delta\left(\frac{1}{2}snr\left(\frac{(1+2snr)^2}{(1-\gamma)(1+snr)^2 - snr}\right) - snr\right) \le \frac{1}{4}.$$

Since we can take δ arbitrarily close to 0, it is obvious that there is a small enough δ for this inequality to hold for any γ and snr pair.

Then upper-bound from (54) is

$$\frac{1}{\delta}\log\frac{4^\delta 5}{3^\delta} + \left(1 + \frac{1}{\gamma}\right)2snr - \frac{1}{2\delta}\log(\delta(1-\gamma-a^2\sigma^2P)). \quad (55)$$

9. REFERENCES

[1] A. Joseph and A. Barron, "Fast sparse superposition codes have exponentially small error probability for R < C," *arXiv preprint arXiv:1207.2406*, 2012.

[2] O. Lanford, "Entropy and equilibrium states in classical statistical mechanics," in *Statistical mechanics and mathematical problems*, pp. 1–113. Springer, 1973.

[3] J. Van Campenhout and T. Cover, "Maximum entropy and conditional probability," *Information Theory, IEEE Transactions on*, vol. 27, no. 4, pp. 483–489, 1981.

[4] I. Csiszár, "Sanov property, generalized I-projection and a conditional limit theorem," *The Annals of Probability*, pp. 768–793, 1984.

[5] T. Cover, J. Thomas, and J. Kieffer, "Elements of information theory," *SIAM Review*, vol. 36, no. 3, pp. 509–510, 1994.

[6] L. Brown, "A proof of the central limit theorem motivated by the Cramér-Rao inequality," *Statistics and probability: essays in honor of CR Rao*, pp. 141–148, 1982.

[7] D. Pollard, "Asymptopia, book in progress," 2000.