ℓ_p-norms of codewords from capacity- and dispersion-achieveing Gaussian codes.

Yury Polyanskiy

***Abstract*—It is demonstrated that codewords of good codes for the additive white Gaussian noise (AWGN) channel become more and more isotropically distributed (in the sense of evaluating quadratic forms) and resemble white Gaussian noise (in the sense of ℓ_p norms) as the code approaches closer to the fundamental limits. In particular, it is shown that the optimal Gaussian code must necessarily have peak-to-average power ratio (PAPR) of order $\log n$.**

***Index Terms*—Shannon theory, empirical statistics, channel capacity, channel dispersion, concentration of measure, additive white Gaussian noise.**

I. Introduction

The problem of constructing good error-correcting codes has been one of the main focuses of information and coding theories. In this paper we investigate some of the properties that the optimal codes must necessarily posses. Such characterization facilitates the search for the good codes and may prove useful for establishing converse bounds in multiuser communication problems, as well as being of theoretical importance.

Specifically, in this paper we focus on the additive white Gaussian noise (AWGN) channel. After introducing the notation in Section I-A, we characterize the degree of isotropy of good constellations in Section II, and the possible ranges of ℓ_p norms of these constellations in Section III. Note that studying ℓ_p norms is a natural mathematical generalization of the concept of peak-to-average power ratio (PAPR), which corresponds to $p=\infty$. Thus one motivation of this work is to understand PAPR requirements of good channel codes.

A. Definitions

A random transformation $P_{Y|X}: \mathcal{X} \to \mathcal{Y}$ is a Markov kernel acting between a pair of measurable spaces. An $(M,\epsilon)_{avg}$ code for the random transformation $P_{Y|X}$ is a pair of random transformations $\mathsf{f}\colon \{1,\ldots,M\} \to \mathcal{X}$ and $\mathsf{g}\colon \mathcal{Y} \to \{1,\ldots,M\}$ such that

$$\mathbb{P}[\hat{W} \neq W] \le \epsilon\,, \tag{1}$$

where in the underlying probability space $X = \mathsf{f}(W)$ and $\hat{W} = \mathsf{g}(Y)$ with W equiprobable on $\{1,\ldots,M\}$, and $W, X, Y, \hat{W}$ forming a Markov chain:

$$W \xrightarrow{\mathsf{f}} X \xrightarrow{P_{Y|X}} Y \xrightarrow{\mathsf{g}} \hat{W}\,. \tag{2}$$

An $(M,\epsilon)_{max}$ code is defined similarly except that (1) is replaced with the more stringent maximal probability of error criterion:

$$\max_{1\le j\le M} \mathbb{P}[\hat{W} \neq W | W = j] \le \epsilon\,. \tag{3}$$

A code is called deterministic, denoted $(M,\epsilon)_{det}$, if the encoder f is a functional (non-random) mapping.

A channel is a sequence of random transformations, $\{P_{Y^n|X^n}, n = 1,\ldots\}$ indexed by the parameter n, referred to as the blocklength. An (M,ϵ) code for the n-th random transformation is called an (n,M,ϵ) code. The non-asymptotic fundamental limit of communication is defined as[1]

$$M^*(n,\epsilon) = \max\{M : \exists (n,M,\epsilon)\text{-code}\}\,. \tag{4}$$

In this paper we study the $AWGN(P)$ channel, that is defined as a sequence of random transformations $P_{Y^n|X^n} : \mathcal{X}_n \to \mathbb{R}^n$, where the n-th input space $\mathcal{X}^n$ is

$$\mathcal{X}_n = \{x^n \in \mathbb{R}^n : \sum x_i^2 \le nP\} \tag{5}$$

and

$$P_{Y^n|X^n=x} = \mathcal{N}(x, \mathbf{I}_n)\,. \tag{6}$$

By [1, Theorem 54], for this channel one has for any $0 < \epsilon < 1$:

$$\log M^*(n,\epsilon) = nC(P) - \sqrt{nV(P)}Q^{-1}(\epsilon) + O(\log n)\,, \tag{7}$$

where

$$C(P) = \frac{1}{2}\log(1+P), \quad V(P) = \frac{\log^2 e}{2}\frac{P(P+2)}{(P+1)^2} \tag{8}$$

are the channel capacity and dispersion, and $Q^{-1}(\cdot)$ is the functional inverse of the complementary Gaussian CDF: $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$.

In this paper we consider the following increasing degrees of optimality for sequences of (n, M_n, ϵ) codes:

1) A code sequence is called capacity-achieving, or $o(n)$-achieving, if
$$\frac{1}{n}\log M_n \to C\,. \tag{9}$$
2) A code sequence is called $O(\sqrt{n})$-achieving if
$$\log M_n = nC + O(\sqrt{n})\,. \tag{10}$$
3) A code sequence is called capacity-dispersion achieving, or $o(\sqrt{n})$-achieving, if
$$\log M_n = nC - \sqrt{nV}Q^{-1}(\epsilon) + o(\sqrt{n})\,. \tag{11}$$

Y. Polyanskiy is with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 02139 USA, e-mail: yp@mit.edu.

This work supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.

[1] Additionally, one should also specify which probability of error criterion, (1) or (3), is used.

4) A code sequence is called $O(\log n)$-achieving if

$$\log M_n = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n)\,. \qquad (12)$$

We quote several results from [2]:

Theorem 1 ([2]): Consider a random transformation $P_{Y|X}$, a distribution P_X induced by an $(M,\epsilon)_{max,det}$ code and an auxiliary output distribution Q_Y. Suppose

$$\sup_x \mathrm{Var}\left[\log \frac{dP_{Y|X=x}}{dQ_Y}(Y)\middle| X=x\right] \le S_m \qquad (13)$$

for some constant $S_m \ge 0$, then we have[2]

$$D(P_{Y|X}||Q_Y|P_X) \ge \log M - \sqrt{\frac{2S_m}{1-\epsilon}} + \log\frac{1-\epsilon}{2e}\,. \quad (14)$$

Theorem 2 ([2]): For any $0 < \epsilon < 1$ and $P > 0$ there exists $a = a(\epsilon, P) > 0$ such that the output distribution P_{Y^n} of any $(n, M_n, \epsilon)_{max,det}$ code for the $AWGN(P)$ channel satisfies

$$D(P_{Y^n}||P^*_{Y^n}) \le nC - \log M + a\sqrt{n}\,, \qquad (15)$$

where Y^{*n} and its distribution $P^*_{Y^n}$ are given by

$$Y^{*n} \sim P^*_{Y^n} \overset{\triangle}{=} \mathcal{N}(0, (1+P)\mathbf{I}_n). \qquad (16)$$

i.e. Y^{*n} is distributed according to the capacity achieving output distribution (caod) of the AWGN channel.

II. Quadratic forms

We denote the canonical inner product on $\mathbb{R}^n$ as

$$(\mathbf{a}, \mathbf{b}) = \sum_{j=1}^n a_j b_j\,, \qquad (17)$$

and write the quadratic form corresponding to matrix A as

$$(A\mathbf{x}, \mathbf{x}) = \sum_{j=1}^n \sum_{i=1}^n a_{i,j} x_i x_j\,, \qquad (18)$$

or (for a random $\mathbf{x}$) as (AX^n, X^n). Note that when $X^n \sim \mathcal{N}(0,P)^n$ we have trivially

$$\mathbb{E}\,[(AX^n, X^n)] = P \operatorname{tr} A\,, \qquad (19)$$

where tr is the trace operator. Therefore, the next result shows that good codes must be close to isotropic Gaussians, at least in the sense of evaluating quadratic forms:

Theorem 3: For any $P > 0$ and $0 < \epsilon < 1$ there exists a constant $b = b(P, \epsilon) > 0$ such that for all $(n, M, \epsilon)_{max,det}$ codes and all quadratic forms A such that

$$-I_n \le A \le I_n \qquad (20)$$

we have

$$|\mathbb{E}\,[(AX^n, X^n)] - P \operatorname{tr} A| \le b_1\sqrt{n}\sqrt{nC - \log M + b\sqrt{n}} \qquad (21)$$

for $b_1 = \frac{2(1+P)}{\sqrt{\log e}}$ and (a refinement for $A = I_n$)

$$|\sum_{j=1}^n \mathbb{E}\,[X_j^2] - nP| \le \frac{2(1+P)}{\log e}(nC - \log M + b\sqrt{n})\,. \quad (22)$$

Remark 1: It is possible to modify the proof slightly and demonstrate that (21) holds on a per-codeword basis for an overwhelming majority of codewords.

Proof: Denote

$$\begin{aligned} \mathbf{\Sigma} &= \mathbb{E}\,[\mathbf{x}\mathbf{x}^T]\,, & (23)\\ \mathbf{V} &= (\mathbf{I}_n + \mathbf{\Sigma})^{-1}\,, & (24)\\ Q_{Y^n} &= \mathcal{N}(0, \mathbf{I}_n + \mathbf{\Sigma})\,, & (25)\\ R(\mathbf{y}|\mathbf{x}) &= \log \frac{dP_{Y^n|X^n=\mathbf{x}}}{dQ_{Y^n}}(\mathbf{y})\,, & (26)\\ d(\mathbf{x}) &= \mathbb{E}\,[R(Y^n|\mathbf{x})|X^n = \mathbf{x}]\,, & (27)\\ v(\mathbf{x}) &= \mathrm{Var}[R(Y^n|\mathbf{x})|X^n = \mathbf{x}]\,. & (28) \end{aligned}$$

Denote also the spectrum of $\mathbf{\Sigma}$ by $\lambda_i, i = 1, \ldots, n$ and its eigenvectors by $\mathbf{v}_i, i = 1, \ldots, n$. We have then

$$\begin{aligned} |\mathbb{E}\,[(AX^n, X^n)] - P \operatorname{tr} A| &= \left|\sum_{i=1}^n (\lambda_i - P)(A\mathbf{v}_i, \mathbf{v}_i)\right| & (29)\\ &\le \sum_{i=1}^n |\lambda_i - P|\,, & (30) \end{aligned}$$

where (29) is by computing the trace in the eigenbasis of $\mathbf{\Sigma}$ and (30) is by (20).

For the log-Radon-Nikodym derivative we have:

$$R(\mathbf{y}|\mathbf{x}) = \frac{\log e}{2}\left(\ln\det(\mathbf{I}_n + \mathbf{\Sigma}) + (\mathbf{V}\mathbf{y}, \mathbf{y}) - ||\mathbf{y} - \mathbf{x}||^2\right) \qquad (31)$$

and thus under $P_{Y^n|X^n=\mathbf{x}}$ we have that $R(Y^n|\mathbf{x})$ is distributed as

$$\frac{\log e}{2}\left(\ln\det(\mathbf{I}_n + \mathbf{\Sigma}) - ||Z^n||^2 + (\mathbf{V}(\mathbf{x} + Z^n), \mathbf{x} + Z^n)\right) \qquad (32)$$

from where

$$d(\mathbf{x}) = \frac{\log e}{2}\left(\ln\det(\mathbf{I}_n + \mathbf{\Sigma}) + (\mathbf{V}\mathbf{x}, \mathbf{x}) + \operatorname{tr}(\mathbf{V} - \mathbf{I}_n)\right) \qquad (33)$$

and thus

$$D(P_{Y^n|X^n}||Q_{Y^n}|P_{X^n}) \;=\; \frac{1}{2}\sum_{j=1}^n \log(1 + \lambda_j)\,. \quad (34)$$

By using

$$\mathrm{Var}[A + B + C] \le 3(\mathrm{Var}[A] + \mathrm{Var}[B] + \mathrm{Var}[C]) \qquad (35)$$

we estimate

$$\begin{aligned} v(\mathbf{x}) \le 3\log^2 e\Bigg(&\frac{1}{4}\mathrm{Var}[||Z^n||_2^2] + \\ &\frac{1}{4}\mathrm{Var}[(VZ^n, Z^n)] + \mathrm{Var}[(V\mathbf{x}, Z^n)]\Bigg)\,. \qquad (36) \end{aligned}$$

Since $\mathbf{V} \le \mathbf{I}_n$ we have $\operatorname{tr}\mathbf{V}^2 \le n$ and

$$||\mathbf{V}\mathbf{x}||_2^2 \le ||\mathbf{x}||^2 \le nP\,. \qquad (37)$$

Plugging these estimates into (36) and computing expectations over $Z^n \sim \mathcal{N}(0, \mathbf{I}_n)$ we get

$$\sup_{\mathbf{x}} v(\mathbf{x}) \le n\left(\frac{9}{4} + 3P\right)\log^2 e \overset{\triangle}{=} nb_1^2\,. \qquad (38)$$

[2] The right-hand side of (14) may be improved by an additive constant $\log e$ if instead of the proof in [2] one invokes Augustin's strong converse [3, Satz 7.3 and 8.2], [4, Section 2].

Finally from Theorem 1 applied with $S_m = b_1^2 n$ and (34) we have

$$\begin{aligned} \frac{1}{2}\sum_{j=1}^n \log(1+\lambda_j) &\geq \log M - b_1\sqrt{n} - \log\frac{2}{1-\epsilon} \quad (39)\\ &\geq \log M - b\sqrt{n} \quad (40)\\ &= \frac{n}{2}(\log(1+P)-\delta)\,, \quad (41) \end{aligned}$$

where we defined

$$\begin{aligned} b &= \sqrt{\frac{2\left(\frac{9}{4}+3P\right)}{1-\epsilon}}\log e + \log\frac{2}{1-\epsilon} \quad (42)\\ \delta &= 2(nC + b\sqrt{n} - \log M)\,. \quad (43) \end{aligned}$$

To derive (22) consider the chain:

$$\begin{aligned} -\delta &\leq \frac{1}{n}\sum_{j=1}^n \log\frac{1+\lambda_i}{1+P} \quad (44)\\ &\leq \log\left(\frac{1}{n}\sum_{j=1}^n \frac{1+\lambda_i}{1+P}\right) \quad (45)\\ &\leq \frac{\log e}{n1+P}\sum_{j=1}^n(\lambda_i - P) \quad (46)\\ &= \frac{\log e}{n1+P}(\mathbb{E}\,[||X^n||_2^2] - nP)\,, \quad (47) \end{aligned}$$

where (44) is (41) divided by n, (45) is by Jensen's inequality, (46) is by $\log x \leq (x-1)\log e$ and (47) is trivial. Since $||\mathbf{x}||_2^2 - nP \leq 0$, we conclude that (22) holds. Finally, (21) follows from (30) and the next Lemma. ■

Lemma 4: Let $\lambda_1, \ldots, \lambda_n > -1$ be such that

$$\begin{aligned} \sum_{i=1}^n \lambda_i &\leq nP\,, \quad (48)\\ \sum_{i=1}^n \log(1+\lambda_i) &\geq n\log(1+P) - n\delta\,. \quad (49) \end{aligned}$$

Then

$$\sum_{i=1}^n |\lambda_i - P| \leq n(1+P)\sqrt{\frac{2\delta}{\log e}} \tag{50}$$

Proof: Define two probability distributions on $n+1$ integers $\{0, \ldots, n\}$ as follows:

$$\begin{aligned} p_i &= \frac{1}{n}\,, \qquad i = 1, \ldots, n \quad (51)\\ q_i &= \begin{cases} 1 - \frac{1}{n(1+P)}\sum_{j=1}^n (1+\lambda_j)\,, & i = 0\,,\\ \frac{1+\lambda_i}{1+P}\,, & i = 1, \ldots, n \end{cases} \quad (52) \end{aligned}$$

Then, by (48) we have $D(P||Q) \leq \delta$ and (50) follows from Pinsker-Csiszar inequality after noticing

$$\|P - Q\|_{TV} \geq \frac{1}{2n(1+P)}\sum_{j=1}^n |\lambda_i - P| \tag{53}$$

■

The previous proof relied on a direct application of Theorem 1 and is independent of the relative entropy estimates in Theorem 2. At the expense of a more technical proof we could derive a similar result using concentration properties of Lipschitz functions demonstrated in [2, Corollary 8]. Indeed, notice that because $\mathbb{E}\,[Z^n] = 0$ we have

$$\mathbb{E}\,[(AY^n, Y^n)] = \mathbb{E}\,[(AX^n, X^n)] + \operatorname{tr} A\,. \tag{54}$$

Thus, (21) follows if we can show

$$\begin{aligned} &|\mathbb{E}\,[(AY^n, Y^n)] - \mathbb{E}\,[(AY^{*n}, Y^{*n})]| \leq \\ &\qquad \mathrm{const}\sqrt{n}\sqrt{nC - \log M + b\sqrt{n}}\,, \end{aligned} \tag{55}$$

where Y^{*n} is defined in (16). This is precisely what [2, Corollary 8] would imply if the function $\mathbf{y} \mapsto (A\mathbf{y}, \mathbf{y})$ was Lipschitz with constant $O(\sqrt{n})$. Of course $(A\mathbf{y}, \mathbf{y})$ is generally not Lipschitz when considered on the entire of $\mathbb{R}^n$. However, it is clear that from the point of view of evaluation of both the $\mathbb{E}\,[(AY^n, Y^n)]$ and $\mathbb{E}\,[(AY^{*n}, Y^{*n}]$ only vectors of norm $O(\sqrt{n})$ are important, and when restricted to the ball $S = \{\mathbf{y} : \|\mathbf{y}\|_2 \leq b\sqrt{n}\}$ the function does have a required Lipschitz constant of $O(\sqrt{n})$. This approximation idea can be made precise using Kirzbraun's theorem (see [5] for a short proof) to extend $(A\mathbf{y}, \mathbf{y})$ beyond the ball S preserving the maximum absolute value and the Lipschitz constant $O(\sqrt{n})$. Another method of showing (55) is by using Bobkov-Götze extension of Gaussian concentration results to non-Lipschitz functions [6, Theorem 1.2] to estimate the moment generating function of (AY^{*n}, Y^{*n}). Both methods yield (55), and hence (21), but with less sharp constants than in Theorem 3.

III. Behavior of $||\mathbf{x}||_q$

The next natural question is to go to polynomials of higher degree. The simplest example of such polynomials are $F(\mathbf{x}) = \sum_{j=1}^n x_j^q$ for some power q, to analysis of which we proceed now. To formalize the problem, consider $1 \leq q \leq \infty$ and define the q-th norm of the input vector in the usual way

$$||\mathbf{x}||_q \triangleq \left(\sum_{i=1}^n |x_i|^q\right)^{\frac{1}{q}}\,. \tag{56}$$

The aim of this section is to investigate the values of $||\mathbf{x}||_q$ for the codewords of good codes for the AWGN channel. Notice that when coordinates of $\mathbf{x}$ are independent Gaussians we expect to have

$$\sum_{i=1}^n |x_i|^q \approx n\mathbb{E}\,[|Z|^q]\,, \tag{57}$$

where $Z \sim \mathcal{N}(0,1)$. In other words, there exists a sequence of capacity achieving codes and constants B_q, $1 \leq q \leq \infty$ such that every codeword $\mathbf{x}$ at every blocklength n satisfies[3]:

$$||\mathbf{x}||_q \leq B_q n^{\frac{1}{q}} = O(n^{\frac{1}{q}}) \qquad 1 \leq q < \infty\,, \tag{58}$$

and

$$||\mathbf{x}||_\infty \leq B_\infty\sqrt{\log n} = O(\sqrt{\log n})\,. \tag{59}$$

[3] This does not follow from a simple random coding argument since we want the property to hold for every codeword, which constitutes exponentially many constraints. However, the claim can indeed be shown by invoking the $\kappa\beta$-bound [1, Theorem 25] with a suitably chosen constraint set F.

Can we prove that (58)-(59) *hold (with possibly different constants) for* any *good code?*

It turns out that the answer depends on the range of q and on the degree of optimality of the code. Our findings are summarized in Table I. The precise meaning of each entry will be clear from Theorems 5, 6, 9 and their corollaries. The main observation is that the closer the code's cardinality comes to $M^*(n,\epsilon)$, the better ℓ_q-norms reflect those of random Gaussian codewords (58)-(59). Loosely speaking, very little can be said about ℓ_q-norms of capacity-achieving codes, while $O(\log n)$-achieving codes are almost indistinguishable from the random Gaussian ones. In particular, we see that, for example, for capacity-achieving codes it is not possible to approximate expectations of polynomials of degrees higher than 2 (or 4 for dispersion-achieving codes) by assuming Gaussian inputs, since even the asymptotic growth rate with n can be dramatically different. The question of whether we can approximate expectations of arbitrary polynomials for $O(\log n)$-achieving codes remains open.

We proceed to clarify the statements made in Table I. First, we show that all the entries except one are the best possible.

Theorem 5: Each estimate in Table I, except $n^{\frac{1}{q}} \log^{\frac{q-4}{2q}} n$, is tight in the following sense: if the entry is n^α, then there exists a constant B_q and a sequence of $O(\log n)$-, dispersion-, $O(\sqrt{n})$-, or capacity-achieving $(n, M_n, \epsilon)_{max,det}$ codes such that <u>each</u> codeword $\mathbf{x} \in \mathbb{R}^n$ satisfies for all $n \geq 1$

$$\|\mathbf{x}\|_q \geq B_q n^\alpha \,. \tag{60}$$

If the entry in the table states $o(n^\alpha)$ then for <u>any</u> sequence $\tau_n \to 0$ there exists a sequence of $O(\log n)$-, dispersion-, $O(\sqrt{n})$-, or capacity-achieving $(n, M_n, \epsilon)_{max,det}$ codes such that each codeword satisfies for all $n \geq 1$

$$\|\mathbf{x}\|_q \geq B_q \tau_n n^\alpha \,. \tag{61}$$

Proof: First, notice that a code from any row is an example of a code for the next row, so we only need to consider each entry which is worse than the one directly above it. Thus it suffices to show the tightness of $o(n^{\frac14})$, $n^{\frac14}$, $o(n^{\frac12})$ and $n^{\frac12}$.

To that end recall that by [1, Theorem 54] the maximum number of codewords $M^*(n,\epsilon)$ at a fixed probability of error ϵ for the AWGN channel satisfies

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n)\,, \tag{62}$$

where $V(P) = \frac{\log^2 e}{2}\frac{P(P+2)}{(P+1)^2}$ is the channel dispersion. Next, we fix a sequence $\delta_n \to 0$ and construct the following sequence of codes. The first coordinate $x_1 = \sqrt{n\delta_n P}$ for every codeword and the rest $(x_2, \ldots, x_n)$ are chosen as coordinates of an optimal AWGN code for the blocklength $n-1$ and power-constraint $P_n = (1-\delta_n)P$. Following the argument of [1, Theorem 67] the number of codewords M_n in such a code will be at least

$$\log M_n = (n-1)C(P_n) - \sqrt{(n-1)V(P_n)}Q^{-1}(\epsilon) + O(1) \tag{63}$$

$$= nC(P) - \sqrt{nV(P)}Q^{-1}(\epsilon) + O(n\delta_n) \tag{64}$$

assuming $\delta_n \gg \frac{1}{n}$. At the same time, because x_1 of each codeword $\mathbf{x}$ is abnormally high we have

$$\|\mathbf{x}\|_q \geq \sqrt{n\delta_n P}\,. \tag{65}$$

So all the examples are constructed by choosing a suitable δ_n as follows:

- Row 1: see (58)-(59).
- Row 2: nothing to prove.
- Row 3: for entries $o(n^{\frac14})$ taking $\delta_n = \frac{\tau_n^2}{\sqrt{n}}$ yields a dispersion-achieving code according to (64); the estimate (61) follows from (65).
- Row 4: for entries $n^{\frac14}$ taking $\delta_n = \frac{1}{\sqrt{n}}$ yields an $O(\sqrt{n})$-achieving code according to (64); the estimate (60) follows from (65).
- Row 5: for entries $o(n^{\frac12})$ taking $\delta_n = \tau_n^2$ yields a capacity-achieving code according to (64); the estimate (61) follows from (65).
- Row 6: for entries $n^{\frac12}$ we can take a codebook with one codeword $(\sqrt{nP}, 0, \ldots, 0)$.

■

Remark 2: The proof can be modified to show that in each case there are codes that simultaneously achieve all entries in the respective row of Table I (except $n^{\frac1q}\log^{\frac{q-4}{2q}} n$).

We proceed to proving upper bounds. Notice simple relations between the ℓ_q norms of vectors in $\mathbb{R}^n$. To estimate a lower-q norm in terms of a higher one, we invoke Holder's inequality:

$$\|\mathbf{x}\|_q \leq n^{\frac1q - \frac1p}\|\mathbf{x}\|_p\,, \qquad \forall 1 \leq q \leq p \leq \infty\,. \tag{66}$$

To provide estimates for $q > p$, notice that obviously

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_p\,. \tag{67}$$

Then, we can extend to $q < \infty$ via the following chain:

$$\|\mathbf{x}\|_q \leq \|\mathbf{x}\|_\infty^{1-\frac{p}{q}}\|\mathbf{x}\|_p^{\frac{p}{q}} \tag{68}$$

$$\leq \|\mathbf{x}\|_p\,, \qquad \forall q \geq p \tag{69}$$

Trivially, for $q = 2$ the answer is given by the power constraint

$$\|\mathbf{x}\|_2 \leq \sqrt{nP} \tag{70}$$

Thus by (66) and (69) we get: *Each codeword of code for the $AWGN(P)$ must satisfy*

$$\|\mathbf{x}\|_q \leq \sqrt{P}\cdot\begin{cases} n^{\frac1q}, & 1 \leq q \leq 2\,, \\ n^{\frac12}, & 2 < q \leq \infty\,. \end{cases} \tag{71}$$

This proves entries in the first column and the last row of Table I.

Before proceeding to upper bounds for $q > 2$ we point out an obvious problem with trying to estimate $\|\mathbf{x}\|_q$ for *each* codeword. Given any code whose codewords lie exactly on the power sphere, we can always apply an orthogonal transformation to it so that one of the codewords becomes $(\sqrt{nP}, 0, 0, \ldots 0)$. For such a codeword we have

$$\|\mathbf{x}\|_q = \sqrt{nP} \tag{72}$$

and the upper-bound (71) is tight. Therefore, to improve upon the (71) we must necessarily consider subsets of codewords

TABLE I: Behavior of ℓ_q norms $\|\mathbf{x}\|_q$ of codewords of good codes for the AWGN channel.

Code	$1 \le q \le 2$	$2 < q \le 4$	$4 < q < \infty$	$q = \infty$
random Gaussian	$n^{\frac{1}{q}}$	$n^{\frac{1}{q}}$	$n^{\frac{1}{q}}$	$\sqrt{\log n}$
any $O(\log n)$-achieving	$n^{\frac{1}{q}}$	$n^{\frac{1}{q}}$	$n^{\frac{1}{q}} \log^{\frac{q-4}{2q}} n$	$\sqrt{\log n}$
any dispersion-achieving	$n^{\frac{1}{q}}$	$n^{\frac{1}{q}}$	$o(n^{\frac{1}{4}})$	$o(n^{\frac{1}{4}})$
any $O(\sqrt{n})$-achieving	$n^{\frac{1}{q}}$	$n^{\frac{1}{q}}$	$n^{\frac{1}{4}}$	$n^{\frac{1}{4}}$
any capacity-achieving	$n^{\frac{1}{q}}$	$o(n^{\frac{1}{2}})$	$o(n^{\frac{1}{2}})$	$o(n^{\frac{1}{2}})$
any code	$n^{\frac{1}{q}}$	$n^{\frac{1}{2}}$	$n^{\frac{1}{2}}$	$n^{\frac{1}{2}}$

Note: All estimates, except $n^{\frac{1}{q}} \log^{\frac{q-4}{2q}} n$, are shown to be tight.

of a given code. For simplicity below we show estimates for the *half* of all codewords.

The following result, proven in the Appendix, takes care of the sup-norm:

Theorem 6 ($q = \infty$): For any $0 < \epsilon < 1$ and $P > 0$ there exists a constant $b = b(P, \epsilon)$ such that for any[4] $n \ge N(P, \epsilon)$ and any $(n, M, \epsilon)_{max,det}$-code for the $AWGN(P)$ channel at least half of the codewords satisfy

$$\|\mathbf{x}\|_\infty^2 \le \frac{4(1+P)}{\log e}\left(nC - \sqrt{nV}Q^{-1}(\epsilon) + \log\frac{2bn^2}{M}\right), \tag{73}$$

where C and V are the capacity-dispersion pair for the channel. In particular, the expression in brackets is non-negative for all codes and blocklengths.

Remark 3: What puts Theorem 6 aside from other results in this paper and [2]. is its sensitivity to whether the code achieves the dispersion term.

From Theorem 6 the explanation of the entries in the last column of Table I becomes obvious: the more terms the code achieves in the asymptotic expansion of $\log M^*(n,\epsilon)$ the closer its estimate of $\|\mathbf{x}\|_\infty$ becomes to the $O(\sqrt{\log n})$, which arises from a random Gaussian codeword (59). To be specific, we give exact statements:

Corollary 7 ($q = \infty$ for $O(\log n)$-codes): For any $0 < \epsilon < 1$ and $P > 0$ there exists a constant $b = b(P, \epsilon)$ such that for any $(n, M_n, \epsilon)_{max,det}$-code for the $AWGN(P)$ with

$$\log M_n \ge nC - \sqrt{nV}Q^{-1}(\epsilon) - K\log n \tag{74}$$

for some $K > 0$ we have that at least half of the codewords satisfy

$$\|\mathbf{x}\|_\infty \le \sqrt{(b+K)\log n} + b\,. \tag{75}$$

Corollary 8 ($q = \infty$ for capacity-achieving codes): For any capacity-achieving sequence of $(n, M_n, \epsilon)_{max,det}$-codes there exists a sequence $\tau_n \to 0$ such that for at least half of the codewords we have

$$\|\mathbf{x}\|_\infty \le \tau_n n^{\frac{1}{2}}\,. \tag{76}$$

Similarly, for any dispersion-achieving sequence of $(n, M_n, \epsilon)_{max,det}$-codes there exists a sequence $\tau_n \to 0$ such that for at least half of the codewords we have

$$\|\mathbf{x}\|_\infty \le \tau_n n^{\frac{1}{4}}\,. \tag{77}$$

Remark 4: By Theorem 5 sequences τ_n are necessarily code-dependent.

For the $q = 4$ we have the following estimate (see Appendix for the proof):

Theorem 9 ($q = 4$): For any $0 < \epsilon < \frac{1}{2}$ and $P > 0$ there exist constants $b_1 > 0$ and $b_2 > 0$, depending on P and ϵ, such that for any $(n, M, \epsilon)_{max,det}$-code for the $AWGN(P)$ channel at least half of the codewords satisfy

$$\|\mathbf{x}\|_4^2 \le \frac{2}{b_1}\left(nC + b_2\sqrt{n} - \log\frac{M}{2}\right), \tag{78}$$

where C is the capacity of the channel. In fact, we also have a lower bound

$$\mathbb{E}\left[\|\mathbf{x}\|_4^4\right] \ge 3nP^2 - (nC - \log M + b_3\sqrt{n})n^{\frac{1}{4}}\,, \tag{79}$$

for some $b_3 = b_3(P, \epsilon) > 0$.

Remark 5: Note that $\mathbb{E}\left[\|\mathbf{z}\|_4^4\right] = 3nP^2$ for $\mathbf{z} \sim \mathcal{N}(0, P)^n$.

We can now complete the proof of results in Table I:

1) Row 2: $q = 4$ is Theorem 9; $2 < q \le 4$ follows by (66) with $p = 4$; $q = \infty$ is Corollary 7; for $4 < q < \infty$ we apply interpolation via (68) with $p = 4$.
2) Row 3: $q \le 4$ is treated as in Row 2; $q = \infty$ is Corollary 8; for $4 < q < \infty$ apply interpolation (68) with $p = 4$.
3) Row 4: $q \le 4$ is treated as in Row 2; $q \ge 4$ follows from (69) with $p = 4$.
4) Row 5: $q = \infty$ is Theorem 8; for $2 < q < \infty$ we apply interpolation (68) with $p = 2$.

The upshot of this section is that we cannot approximate values of non-quadratic polynomials in $\mathbf{x}$ (or $\mathbf{y}$) by assuming iid Gaussian entries, unless the code is $O(\sqrt{n})$-achieving, in which case we can go up to degree 4 but still will have to be content with Gaussian lower bounds only such as (79).[5]

Before closing this discussion we demonstrate the sharpness of the arguments in this section by considering the following example. Suppose that a power of a codeword $\mathbf{x}$ from a capacity-dispersion optimal code is measured by an imperfect tool, such that its reading is described by

$$\mathcal{E} = \frac{1}{n}\sum_{i=1}^{n}(x_i)^2 H_i\,, \tag{80}$$

[4] $N(P,\epsilon) = 8(1 + 2P^{-1})(Q^{-1}(\epsilon))^2$ for $\epsilon < \frac{1}{2}$ and $N(P,\epsilon) = 1$ for $\epsilon \ge \frac{1}{2}$.

[5] Using quite similar methods, (79) can be extended to certain bi-quadratic forms, i.e. 4-th degree polynomials $\sum_{i,j} a_{i-j}x_i^2x_j^2$, where $A = (a_{i-j})$ is a Toeplitz positive semi-definite matrix.

where H_i's are i.i.d bounded random variables with expectation and variance both equal to 1. For large blocklengths n we expect $\mathcal{E}$ to be Gaussian with mean P and variance $\frac{1}{n}\|\mathbf{x}\|_4^4$. On the one hand, Theorem 9 shows that the variance will not explode; (79) shows that it will be at least as large as that of a Gaussian codebook. Finally, to establish the asymptotic normality rigorously, the usual approach based on checking Lyapunov condition will fail as shown by Theorem 5, but the Lindenberg condition does hold as a consequence of Theorem 8. If in addition, the code is $O(\log n)$-achieving then

$$\mathbb{P}[|\mathcal{E} - \mathbb{E}[\mathcal{E}]| > \delta] \le e^{-\frac{n\delta^2}{b_1 + b_2\delta\sqrt{\log n}}}\,.$$

Appendix

In this appendix we prove results from Section III.

To prove Theorem 6 our basic intuition is that any codeword which is abnormally peaky (i.e., has a high value of $\|\mathbf{x}\|_\infty$) is wasteful in terms of allocating its power budget. Thus a good capacity- or dispersion-achieving codebook cannot have too many of such wasteful codewords. Formalization of this intuitive argument is as follows:

Lemma 10: For any $\epsilon \le \frac{1}{2}$ and $P > 0$ there exists a constant $b = b(P, \epsilon)$ such that given any $(n, M, \epsilon)_{max,det}$ code for the $AWGN(P)$ channel, we have for any $0 \le \lambda \le P$:[6]

$$\mathbb{P}[\|\mathbf{x}\|_\infty \ge \sqrt{\lambda n}] \le \frac{b}{M}\exp\left\{nC(P') - \sqrt{nV(P')}Q^{-1}(\epsilon) + 2\log n\right\} \qquad (81)$$

where $P' = P - \lambda$ and $C(P)$ and $V(P)$ are defined in (8).

Proof: Our method is to apply the meta-converse in the form of [1, Theorem 30] to a subcode $\{\|\mathbf{x}\|_\infty \ge \sqrt{\lambda n}\}$. Application of a meta-converse requires selecting a suitable auxiliary channel $Q_{Y^n|X^n}$. We specify this channel now. For any $\mathbf{x} \in \mathbb{R}^n$ let j^* be the first index s.t. $|x_j| = ||\mathbf{x}||_\infty$, then we set

$$Q_{Y^n|X^n}(y^n|\mathbf{x}) = P_{Y|X}(y_{j^*}|x_{j^*}) \prod_{j \ne j^*(\mathbf{x})} P_Y^*(y_j) \qquad (82)$$

We will show below that for some $b_1 = b_1(P)$ any M-code over this Q-channel has average probability of error ϵ' satisfying:

$$1 - \epsilon' \le \frac{b_1 n^{\frac{3}{2}}}{M}\,. \qquad (83)$$

On the other hand, writing the expression for $\log\frac{dP_{Y^n|X^n=\mathbf{x}}}{dQ_{Y^n|X^n=\mathbf{x}}}(Y^n)$ we see that it coincides with the expression for $\log\frac{dP_{Y^n|X^n=\mathbf{x}}}{dP^*_{Y^n}}$ except that the term corresponding to $j^*(\mathbf{x})$ will be missing; compare with [7, (4.29) and before]. Thus, one can repeat step by step the analysis in the proof of [1, Theorem 65] with the only difference that nP should be replaced by $nP - \|\mathbf{x}\|_\infty^2$ reflecting the reduction in the energy due to skipping of j^*. Then, we obtain for some $b_2 = b_2(\alpha, P)$:

$$\log\beta_{1-\epsilon}(P_{Y^n|X^n=\mathbf{x}}, Q_{Y^n|X^n=\mathbf{x}}) \ge - nC\left(P'\right) + \sqrt{nV\left(P'\right)}Q^{-1}(\epsilon) - \frac{1}{2}\log n - b_2\,, \qquad (84)$$

where $P' = P - \frac{\|\mathbf{x}\|_\infty^2}{n}$ and which holds simultaneously for all $\mathbf{x}$ with $\|\mathbf{x}\| \le \sqrt{nP}$. Two remarks are in order: first, the analysis in [1, Theorem 64] must be done replacing n with $n-1$, but this difference is absorbed into b_2. Second, to see that b_2 can be chosen independent of $\mathbf{x}$ notice that $B(P)$ in [1, (620)] tends to 0 with $P \to 0$ and hence can be bounded uniformly for all $P \in [0, P_{max}]$.

Denote the cardinality of the subcode $\{\|\mathbf{x}\|_\infty \ge \sqrt{\lambda n}\}$ by

$$M_\lambda = M\mathbb{P}[\|\mathbf{x}\|_\infty \ge \sqrt{\lambda n}]\,. \qquad (85)$$

Then according to [1, Theorem 30], we get

$$\inf_{\mathbf{x}} \beta_{1-\epsilon}(P_{Y^n|X^n=\mathbf{x}}, Q_{Y^n|X^n=\mathbf{x}}) \le 1 - \epsilon'\,, \qquad (86)$$

where the infimum is over the codewords of M_λ-subcode. Applying both (83) and (84) we get

$$-nC\left(P'\right) + \sqrt{nV\left(P'\right)}Q^{-1}(\epsilon) - \frac{1}{2}\log n - b_2 \le - \log M_\lambda + \log b_1 + \frac{3}{2}\log n \qquad (87)$$

Thus, overall

$$\log M_\lambda \le nC(P-\lambda) - \sqrt{nV(P-\lambda)}Q^{-1}(\epsilon) + 2\log n + b\,, \qquad (88)$$

for $b = b_1\exp\{b_2\}$.

It remains to show (83). Consider an $(n, M, \epsilon')_{avg,det}$-code for the Q-channel and let $M_j, j = 1, \ldots, n$ denote the cardinality of the set of all codewords with $j^*(\mathbf{x}) = j$. Let ϵ'_j denote the minimum possible average probability of error of each such codebook achievable with the maximum likelihood (ML) decoder (informed of the value of j). Since

$$1 - \epsilon' \le \frac{1}{M}\sum_{j=1}^{n} M_j(1 - \epsilon'_j) \qquad (89)$$

it suffices to prove

$$1 - \epsilon'_j \le \frac{\sqrt{\frac{2nP}{\pi}} + 2}{M_j} \qquad (90)$$

for all j. Without loss of generality assume $j = 1$ in which case observations Y_2^n are useless for determining the value of the true codeword. Moreover, ML decoding regions $D_i, i = 1, \ldots, M_j$ for each codeword are disjoint intervals in $\mathbb{R}^1$ (so that decoder outputs message estimate i whenever $Y_1 \in D_i$). Note that for $M_j \le 2$ there is nothing to prove, so assume otherwise. Denote the M_j message points by $x_i, i = 1, \ldots, M_j$ and assume (without loss of generality) that $-\sqrt{nP} \le x_1 \le x_2 \le \cdots \le x_{M_j} \le \sqrt{nP}$ and that $D_2, \ldots D_{M_j - 1}$ are finite intervals. Then the following estimate may be established by elementary arguments

$$1 - \epsilon'_j \le \frac{2}{M_j} + \frac{M_j - 2}{M_j}\left(1 - 2Q\left(\frac{\sqrt{nP}}{M_j - 2}\right)\right) \qquad (91)$$

$$\le \frac{2}{M_j} + \frac{\sqrt{\frac{2nP}{\pi}}}{M_j}\,, \qquad (92)$$

Thus, (92) completes the proof of (90), (83) and the theorem. ■

[6] For $\epsilon > \frac{1}{2}$ one must replace $V(P-\lambda)$ with $V(P)$ in (81). This does not modify any of the arguments required to prove Theorem 6.

Proof of Theorem 6: Notice that for any $0 \leq \lambda \leq P$ we have

$$C(P-\lambda) \leq C(P) - \frac{\lambda \log e}{2(1+P)} \,. \qquad (93)$$

On the other hand, by concavity of $\sqrt{V(P)}$ and since $V(0) = 0$ we have for any $0 \leq \lambda \leq P$

$$\sqrt{V(P-\lambda)} \geq \sqrt{V(P)} - \frac{\sqrt{V(P)}}{P}\lambda \,. \qquad (94)$$

Thus, taking $s = \lambda n$ in Lemma 10 we get with the help of (93) and (94):

$$\mathbb{P}[\|\mathbf{x}\|_\infty^2 \geq s] \leq \exp\left\{\Delta_n - (b_1 - b_2 n^{-\frac{1}{2}})s\right\} \,, \qquad (95)$$

where we denoted for convenience

$$b_1 = \frac{\log e}{2(1+P)} \,, \qquad b_2 = \frac{\sqrt{V(P)}}{P} Q^{-1}(\epsilon) \,, \qquad (96)$$

$$\Delta_n = nC(P) - \sqrt{nV(P)}Q^{-1}(\epsilon) + \log \frac{bn^2}{M} \,. \qquad (97)$$

Note that Lemma 10 only shows validity of (95) for $0 \leq s \leq nP$, but since for $s > nP$ the left-hand side is zero, the statement actually holds for all $s \geq 0$. Then for $n \geq N(P, \epsilon)$ we have

$$(b_1 - b_2 n^{-\frac{1}{2}}) \geq \frac{b_1}{2} \qquad (98)$$

and thus further upper-bounding (95) we get

$$\mathbb{P}[\|\mathbf{x}\|_\infty^2 \geq s] \leq \exp\left\{\Delta_n - \frac{b_1 s}{2}\right\} \,. \qquad (99)$$

Finally, if we had that for some code $\Delta_n < 0$ then (99) would imply that $\mathbb{P}[\|\mathbf{x}\|_\infty^2 \geq s] < 1$ for all $s \geq 0$, which is clearly impossible. Thus we must have $\Delta_n \geq 0$ for any $(n, M, \epsilon)_{max,det}$ code. The proof concludes by taking $s = \frac{2(\log 2 + \Delta_n)}{b_1}$ in (99). ■

Before proving Theorem 9 we state two auxiliary results.

Lemma 11: Let $f : \mathcal{Y} \to \mathbb{R}$ be a (single-letter) function such that for some $\theta > 0$ we have

$$m_1 = \mathbb{E}\left[\exp\{\theta f(Y^*)\}\right] < \infty \,, \qquad (100)$$

(one-sided Cramer condition) and

$$m_2 = \mathbb{E}\left[|f(Y^*)|^2\right] < \infty \,. \qquad (101)$$

Then there exists $b = b(m_1, m_2, \theta) > 0$ such that for all $n \geq \frac{16}{\theta^4}$ we have

$$\frac{1}{n}\sum_{j=1}^n \mathbb{E}\left[f(Y_j)\right] \leq \mathbb{E}\left[f(Y^*)\right] + \frac{D(P_{Y^n} || P^*_{Y^n}) + b\sqrt{n}}{n^{\frac{3}{4}}} \,, \qquad (102)$$

provided that $P^*_{Y^n} = (P^*_Y)^n$.

Proof: Follows by a straightforward application of Donsker-Varadhan inequality [8, Lemma 2.1] and technical estimates of $\mathbb{E}\left[\exp\{tf(Y^*)\}\right]$. ■

Theorem 12: Consider an $(M, \epsilon)_{avg}$ code for an arbitrary random transformation $P_{Y|X}$. Then for any Q_Y we have

$$\beta_\alpha(P_Y, Q_Y) \geq M\beta_{\alpha-\epsilon}(P_{XY}, P_X Q_Y) \qquad \epsilon \leq \alpha \leq 1 \,. \qquad (103)$$

Proof: On the probability space corresponding to a given $(M, \epsilon)_{avg}$ code, define the following random variable

$$Z = 1\{\hat{W}(Y) = W, Y \in E\} \,, \qquad (104)$$

where E is an arbitrary subset satisfying

$$P_Y[E] \geq \alpha \,. \qquad (105)$$

Precisely as in the original meta-converse [1, Theorem 26] the main idea is to use Z as a suboptimal hypothesis test for discriminating P_{XY} against $P_X Q_Y$. Following the same reasoning as in [1, Theorem 27] one notices that

$$(P_X Q_Y)[Z = 1] \leq \frac{Q_Y[E]}{M} \qquad (106)$$

and

$$P_{XY}[Z = 1] \geq \alpha - \epsilon \,. \qquad (107)$$

Therefore, by definition of β_α we must have

$$\beta_{\alpha-\epsilon}(P_{XY}, P_X Q_Y) \leq \frac{Q_Y[E]}{M} \,. \qquad (108)$$

Taking the infimum in (108) over all E satisfying (105) completes the proof of (103). ■

Proof of Theorem 9: To prove (78) we will show the following statement: *There exist two constants b_0 and b_1 such that for any (n, M_1, ϵ) code for the $AWGN(P)$ channel with codewords $\mathbf{x}$ satisfying*

$$\|\mathbf{x}\|_4 \geq bn^{\frac{1}{4}} \qquad (109)$$

we have an upper bound on the cardinality:

$$M_1 \leq \frac{4}{1-\epsilon} \exp\left\{nC + 2\sqrt{\frac{nV}{1-\epsilon}} - b_1(b - b_0)^2\sqrt{n}\right\} \,, \qquad (110)$$

provided $b \geq b_0(P, \epsilon)$. From here (78) follows by first upper-bounding $(b - b_0)^2 \geq \frac{b^2}{2} - b_0^2$ and then verifying easily that the choice

$$b^2 = \frac{2}{b_1\sqrt{n}}(nC + b_2\sqrt{n} - \log\frac{M}{2}) \qquad (111)$$

with $b_2 = b_0^2 b_1 + 2\sqrt{\frac{V}{1-\epsilon}} + \log\frac{4}{1-\epsilon}$ takes the right-hand side of (110) below $\log\frac{M}{2}$.

To prove (110), fix b, denote

$$S = b - \left(\frac{6}{1+\epsilon}\right)^{\frac{1}{4}} \qquad (112)$$

and assume b is large enough so that

$$\delta \triangleq S - 6^{\frac{1}{4}}\sqrt{1+P} > 0 \,. \qquad (113)$$

Then, on one hand we have

$$P_{Y^n}[\|Y^n\|_4 \geq Sn^{\frac{1}{4}}] \geq \mathbb{P}[\|X^n\|_4 - \|Z^n\|_4 \geq Sn^{\frac{1}{4}}] \qquad (114)$$

$$\geq \mathbb{P}[\|Z^n\|_4 \leq n^{\frac{1}{4}}(S - b)] \qquad (115)$$

$$\geq \frac{1+\epsilon}{2} \,, \qquad (116)$$

where (114) is by triangle inequality for $\|\cdot\|_4$, (115) is by the constraint (109) and (116) is by Chebyshev inequality applied to $\|Z^n\|_4^4 = \sum_{j=1}^n Z_j^4$. On the other hand, we have

$$P^*_{Y^n}[\|Y^n\|_4 \le Sn^{\frac{1}{4}}] = \tag{117}$$

$$\ge P^*_{Y^n}[\{\|Y^n\|_4 \le 6^{\frac{1}{4}}\sqrt{1+P}n^{\frac{1}{4}}\} + \{\|Y^n\|_4 \le \delta n^{\frac{1}{4}}\}] \tag{118}$$

$$\ge P^*_{Y^n}[\{\|Y^n\|_4 \le 6^{\frac{1}{4}}\sqrt{1+P}n^{\frac{1}{4}}\} + \{\|Y^n\|_2 \le \delta n^{\frac{1}{4}}\}] \tag{119}$$

$$\ge 1 - \exp\{-b_1\delta^2\sqrt{n}\}\,, \tag{120}$$

where (118) is by the triangle inequality for $\|\cdot\|_4$ which implies the inclusion

$$\{\mathbf{y} : \|\mathbf{y}\|_4 \le a+b\} \supset \{\mathbf{y} : \|\mathbf{y}\|_4 \le a\} + \{\mathbf{y} : \|\mathbf{y}\|_4 \le b\} \tag{121}$$

with $+$ denoting the Minkowski sum of sets, (119) is by (69) with $p = 2$, $q = 4$; and (120) hold for some $b_1 = b_1(P) > 0$ by the Gaussian isoperimetric inequality [9] which is applicable since

$$P^*_{Y^n}[\|Y^n\|_4 \le 6^{\frac{1}{4}}\sqrt{1+P}n^{\frac{1}{4}}] \ge \frac{1}{2} \tag{122}$$

by Chebyshev inequality applied to $\sum_{j=1}^n Y_j^4$ (note: $Y^n \sim \mathcal{N}(0, 1+P)^n$ under $P^*_{Y^n}$). As a side remark, we add that the estimate of the large-deviations of the sum of 4-th powers of iid Gaussians as $\exp\{-O(\sqrt{n})\}$ is order-optimal.

Together (116) and (120) imply

$$\beta_{\frac{1+\epsilon}{2}}(P_{Y^n}, P^*_{Y^n}) \le \exp\{-b_1\delta^2\sqrt{n}\}\,. \tag{123}$$

On the other hand, by [1, Lemma 59] we have for any $\mathbf{x}$ with $\|\mathbf{x}\|_2 \le \sqrt{nP}$ and any $0 < \alpha < 1$:

$$\beta_\alpha(P_{Y^n|X^n=\mathbf{x}}, P^*_{Y^n}) \ge \frac{\alpha}{2}\exp\left\{-nC - \sqrt{\frac{2nV}{\alpha}}\right\}\,, \tag{124}$$

where C and V are the capacity and the dispersion of the $AWGN(P)$ channel. Then, by convexity in α of the right-hand side of (124) and [7, Lemma 32] we have for any input distribution P_{X^n}:

$$\beta_\alpha(P_{X^nY^n}, P_{X^n}P^*_{Y^n}) \ge \frac{\alpha}{2}\exp\left\{-nC - \sqrt{\frac{2nV}{\alpha}}\right\}\,. \tag{125}$$

We complete the proof of (110) by invoking Theorem 12 (see below) with $Q_Y = P^*_{Y^n}$ and $\alpha = \frac{1+\epsilon}{2}$:

$$\beta_{\frac{1+\epsilon}{2}}(P_{Y^n}, P^*_{Y^n}) \ge M_1\beta_{\frac{1-\epsilon}{2}}(P_{X^nY^n}, P_{X^n}P^*_{Y^n})\,. \tag{126}$$

Applying to (126) bounds (123) and (125) we conclude that (110) is shown with

$$b_0 = \left(\frac{6}{1+\epsilon}\right)^{\frac{1}{4}} + 6^{\frac{1}{4}}\sqrt{1+P}\,. \tag{127}$$

Next, we proceed to the proof of (79). On one hand, we have

$$\begin{aligned} \sum_{j=1}^n \mathbb{E}\left[Y_j^4\right] &= \sum_{j=1}^n \mathbb{E}\left[(X_j+Z_j)^4\right] && (128)\\ &= \sum_{j=1}^n \mathbb{E}\left[X_j^4 + 6X_j^2Z_j^2 + Z_j^4\right] && (129)\\ &\le \mathbb{E}\left[\|\mathbf{x}\|_4^4\right] + 6nP + 3n\,, && (130) \end{aligned}$$

where (128) is by the definition of the AWGN channel, (129) is because X^n and Z^n are independent and thus odd terms vanish, (130) is by the power-constraint $\sum X_j^2 \le nP$. On the other hand, applying Lemma 11 with $f(y) = -y^4$, $\theta = 2$ and using Theorem 2 we obtain[7]

$$\sum_{j=1}^n \mathbb{E}\left[Y_j^4\right] \ge 3n(1+P)^2 - (nC - \log M + b_3\sqrt{n})n^{\frac{1}{4}}\,, \tag{131}$$

for some $b_3 = b_3(P, \epsilon) > 0$. Comparing (131) and (130) statement (79) follows.

We remark that by a straightforward extension of Lemma 11 to expectations $\frac{1}{n-1}\sum_{j=1}^{n-1}\mathbb{E}\left[Y_j^2Y_{j+1}^2\right]$, cf. [10, Section IV.E], we could provide a lower bound similar to (79) for more general 4-th degree polynomials in $\mathbf{x}$. For example, it is possible to treat the case of $p(\mathbf{x}) = \sum_{i,j} a_{i-j}x_i^2x_j^2$, where $A = (a_{i-j})$ is a Toeplitz positive semi-definite matrix. We would proceed as in (130), computing $\mathbb{E}\left[p(Y^n)\right]$ in two ways, with the only difference that the peeled off quadratic polynomial would require application of Theorem 3 instead of the simple power constraint. Finally, we also mention that the method (130) does not work for estimating $\mathbb{E}\left[\|\mathbf{x}\|_6^6\right]$ because we would need an *upper* bound $\mathbb{E}\left[\|\mathbf{x}\|_4^4\right] \lesssim 3nP^2$, which is not possible to obtain in the context of $O(\sqrt{n})$-achieving codes as counter-examples in Theorem 5 show. ■

ACKNOWLEDGEMENT

Author would like to thank Prof. S. Verdú for useful discussions.

REFERENCES

[1] Y. Polyanskiy, H. 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.

[2] Y. Polyanskiy and S. Verdú, "Relative entropy at the channel output of a capacity-achieving code," in *Proc. 2011 49th Allerton Conference*, Allerton Retreat Center, Monticello, IL, USA, Oct. 2011.

[3] U. Augustin, "Gedächtnisfreie kanäle für diskrete zeit," *Z. Wahrscheinlichkeitstheorie und Verw. Geb.*, vol. 6, pp. 10–61, 1966.

[4] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," *J. Comb. Inform. Syst. Sci*, vol. 7, no. 3, pp. 216–230, 1982.

[5] I. J. Schoenberg, "On a theorem of Kirzbraun and Valentine," *Am. Math. Monthly*, vol. 60, no. 9, pp. 620–622, Nov. 1953.

[6] S. Bobkov and F. Götze, "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities," *J. Functional Analysis*, vol. 163, pp. 1–28, 1999.

[7] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton Univ., Princeton, NJ, USA, 2010, available: http://people.lids.mit.edu/yp/homepage/.

[8] M. Donsker and S. Varadhan, "Asymptotic evaluation of certain markov process expectations for large time. i. ii." *Comm. Pure Appl. Math.*, vol. 28, no. 1, pp. 1–47, 1975.

[9] V. Sudakov and B. Tsirelson, "Extremal properties of half-spaces for spherically invariant measures," *Zap. Nauch. Sem. LOMI*, vol. 41, pp. 14–24, 1974.

[10] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," preprint. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/optcodes_journal.pdf

[7] Of course, a similar Gaussian lower bound holds for any cumulative sum, in particular for any power $\sum \mathbb{E}\left[|Y_j|^q\right], q \ge 1$.