The Impact of Constellation Cardinality on Gaussian Channel Capacity

Yihong Wu
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: yihongwu@princeton.edu

Sergio Verdú
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

***Abstract*—Denote by $C_m(\mathsf{snr})$ the Gaussian channel capacity with signal-to-noise ratio snr and input cardinality m. We show that as m grows, $C_m(\mathsf{snr})$ approaches $C(\mathsf{snr}) = \frac{1}{2}\log(1+\mathsf{snr})$ exponentially fast. Lower and upper bounds on the exponent are given as functions of snr. We propose a family of input constellations based on the roots of the Hermite polynomials which achieves exponential convergence.**

I. Introduction

We define $C_m(\mathsf{snr})$ as the capacity of the additive-noise Gaussian channel channel with average power and input cardinality constraints:

$$C_m(\mathsf{snr}) = \sup_{\substack{\mathbb{E}[X^2]\leq 1 \\ |\mathrm{supp}(P_X)|\leq m}} I(X, \mathsf{snr}) \tag{1}$$

where

$$I(X, \mathsf{snr}) \triangleq I(X; \sqrt{\mathsf{snr}}X + N) \tag{2}$$

with $\mathsf{snr} > 0$ and N a standard normal random variable independent of X. The operational meaning of $C_m(\mathsf{snr})$ is the maximal reliable communication rate over the Gaussian channel using the best m-point constellation (modulation scheme). It is practically relevant to investigate $C_m(\mathsf{snr})$ since the constellation size is usually a proxy for its complexity.

It is easy to show that as the constellation cardinality m grows, the capacity tends the usual Gaussian channel capacity, that is,

$$C_m(\mathsf{snr}) \nearrow C(\mathsf{snr}) \triangleq \frac{1}{2}\log(1+\mathsf{snr}) \tag{3}$$

for any $\mathsf{snr} \geq 0$. The fundamental question of how fast we can approach the Gaussian channel capacity at a given SNR by increasing constellation size was raised in [1]. To address this question, we define the capacity gap as

$$D_m(\mathsf{snr}) = C(\mathsf{snr}) - C_m(\mathsf{snr}). \tag{4}$$

Note that the difference in mutual information achieved by a given input and its Gaussian counterpart can be expressed in terms of relative entropy between outputs:

$$D(\sqrt{\mathsf{snr}}X + N \,||\, \sqrt{\mathsf{snr}}\Phi_X + N) = I(\Phi_X, \mathsf{snr}) - I(X, \mathsf{snr}) \tag{5}$$

where Φ_X is a Gaussian random variable with the same mean and variance as X. Then

$$D_m(\mathsf{snr}) = \inf_{\substack{\mathbb{E}[X^2]=1 \\ |\mathrm{supp}(P_X)|\leq m}} D(\sqrt{\mathsf{snr}}X + N \,||\, \mathcal{N}(0, 1+\mathsf{snr})). \tag{6}$$

Observe that the density of $\sqrt{\mathsf{snr}}X + N$ is the mixture of m translated standard normal density. The reference measure in (6) is the Gaussian distribution with the same mean and variance as the first distribution. This is known also as the non-Gaussianness. Therefore, determining $D_m(\mathsf{snr})$ boils down to a *non-linear approximation* problem, whose goal is to approximate the optimal output distribution $\mathcal{N}(0, 1+\mathsf{snr})$, a flatter Gaussian distribution, using an m-term standard *Gaussian location mixture*, and the approximation error is gauged by the *relative entropy*.

Using a simple MMSE bound, it is shown in [1, Section VI] that $D_m(\mathsf{snr}) = O\left(\frac{\log m}{m}\right)$, achieved by uniformly quantizing a truncated Gaussian. The goal of this paper is to study the optimal construction and show that the optimal convergence rate to the Gaussian capacity is in fact exponential. Lower and upper bounds on the optimal exponent are given as follows:

Theorem 1.

$$2\log\left(1+\frac{1}{\mathsf{snr}}\right) \leq \liminf_{m\to\infty} \frac{1}{m}\log\frac{1}{D_m(\mathsf{snr})} \tag{7}$$

$$\leq \limsup_{m\to\infty} \frac{1}{m}\log\frac{1}{D_m(\mathsf{snr})} \tag{8}$$

$$\leq 2\log\left(1+\frac{2}{\mathsf{snr}}\right) \tag{9}$$

To appreciate the implication of Theorem 1, let us suppose the optimal exponent $E(\mathsf{snr}) \triangleq \lim_{m\to\infty}\frac{1}{m}\log\frac{1}{D_m(\mathsf{snr})}$ exists. As snr grows, more points are needed to maintain the same order of approximation. Therefore $E(\cdot)$ must be a strictly decreasing function, such that $E(0+) = \infty$ and $E(\infty) = 0$. Both the lower and upper bounds in Theorem 1 satisfy these intuitive requirements. Moreover, we have $E(\mathsf{snr}) = \Theta\left(\frac{1}{\mathsf{snr}}\right)$ in the high-SNR regime.

Approximation by location mixture dates back to Wiener's Tauberian theorem [2], which states that the linear subspace spanned by translates of a given function is dense in $L^2(\mathbb{R}^d)$ if and only if the zeros of its Fourier transform have zero Lebesgue measure. This result applies in particular to Gaussian mixtures. The order of approximation and constructive algorithms are studied in approximation theory, neural network and statistics community, for example, [3], [4], [5], [6], [7], etc. Barron [3] studied approximation by location and scale mixture of sigmoidal functions and showed that the worst case error of approximating a class of functions on $\mathbb{R}^d$ by m-term mixtures is $O\left(\frac{1}{\sqrt{m}}\right)$, independent of the dimension. Using moment matching, Ghosal and van der Vaart [6] showed that Gaussian mixtures can approximate the convolution of a Gaussian and a sub-Gaussian with exponentially small error term. However, no explicit construction is given nor is the achievable exponent analyzed. Moreover, it is unknown whether exponential convergence is optimal.

Finding a good input distribution amounts to approximating the standard Gaussian distribution by discrete distributions, while the quality of approximation is measured by the non-Gaussianness of the convolution of the discrete distribution with a standard Gaussian. Convolution blurs the difference between distributions. It is because of the smoothing effect of the *Gaussian convolution* that exponentially small approximation error can be achieved. In contrast, directly approximating a Gaussian distribution by discrete distributions is much harder, in the sense that the relative entropy or total variation distance is always at extreme (infinity or two respectively). In terms of other weaker distances (*e.g.*, Kolmogorov distance or Wasserstein distance), the approximation error also decays much more slowly according to $\Theta\left(\frac{1}{m}\right)$ [8].

II. Optimal Quadrature

In this section we give a brief introduction to optimal quadrature in a probabilistic setup. This construction plays a key role in finite-constellation problems. Let X_m be a simple random variable with m atoms, whose distribution is given by

$$P_{X_m} = \sum_{i=1}^{m} w_{im}\delta_{x_{im}}, \tag{10}$$

specified by the support $\boldsymbol{x}_m = (x_{1m}, \ldots, x_{mm}) \in \mathbb{R}^m$ and weights $\boldsymbol{w}_m = (w_{1m}, \ldots, w_{mm}) \in \mathbb{R}_+^m$ such that $\sum_{i=1}^m w_{im} = 1$.

Given $m \in \mathbb{N}$ and a real-valued random variable X with probability density w, we want to find an X_m of the form (10) to match as many moments of X as possible. Formally, let Π_N denote the collection of all polynomials of degree no more than N. Define $N^*(m)$ to be the maximal N such that there exists some X_m such that

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(X_m)\right] \tag{11}$$

holds for all $f \in \Pi_N$. In the language of numerical analysis, the distribution of X_m gives a *quadrature rule*, an approximate way to compute integration with respect to w

$$\int_{\mathbb{R}} f(x)w(x)\mathrm{d}x = \sum_{i=1}^{m} w_{im} f(x_{im}) \tag{12}$$

m	$H_m(x)$	$\boldsymbol{x}_m$	$\boldsymbol{w}_m$
1	x	1	1
2	x^2-1	$(-1,1)$	$(1/2,1/2)$
3	x^3-3x	$(-\sqrt{3},0,\sqrt{3})$	$(\frac{1}{6},\frac{2}{3},\frac{1}{6})$

TABLE I
The Hermite polynomials and Gauss quadrature.

that is *exact* for polynomials of degree no more than N.

In one-dimensional space, this problem was solved by Gauss, who showed that

$$N^*(m) = 2m - 1 \tag{13}$$

and the optimal quadrature is constructed by placing the atoms at the roots of the m^{th} orthogonal polynomial with respect to weight w, known as the *Gauss quadrature* [9, Section 3.6]. The optimality of (13) can be understood by matching the number of equations to the degrees of freedom in X_m.

Next we focus on the case where X is standard Gaussian with density $\varphi(x) \triangleq \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$. Let H_m denote the m^{th} Hermite polynomial[1] [10, Section 5.5]:

$$H_m(x) = \frac{(-1)^m}{\varphi(x)}\frac{\mathrm{d}^m\varphi(x)}{\mathrm{d}x^m} \tag{14}$$

$$= \sum_{k=0}^{\lfloor\frac{m}{2}\rfloor}(-1)^k(2k-1)!!\binom{m}{2k}x^{m-2k}. \tag{15}$$

The sequence $\{H_m\}$ forms an orthogonal basis for $L^2(\mathbb{R}, \varphi(x)\mathrm{d}x)$, with

$$\int_{\mathbb{R}} H_m(x)H_n(x)\varphi(x)\mathrm{d}x = m!\,\delta_{m,n}. \tag{16}$$

For each even (odd resp.) m, H_m is an even (odd resp.) function, with m real simple roots [9, Theorem 3.6.10].

The Gauss quadrature with respect to φ is uniquely given as follows [9, Theorem 3.6.12], [10, (15.3.6)]:

Theorem 2. *Let X_m^{Q} be distributed according to (10), where $x_{1m}, \ldots, x_{mm}$ denote the roots of H_m and*

$$w_{im} = \frac{(m-1)!}{mH_{m-1}^2(x_{im})}. \tag{17}$$

Then (11) holds for all $f \in \Pi_{2m-1}$.

Due to the symmetry of the Hermite polynomials, X_m^{Q} is also symmetric, with a bell-shaped distribution [11]. For each odd m, there is an atom at zero. The distribution of X_m^{Q} for $m \leq 3$ are given in Table I. Fig. 1 shows the seven-point quadrature. For higher-order quadrature formulae, see [12, Table 25.10].

Remark 1. Some asymptotic properties of the Gauss quadrature X_m^{Q} is summarized as follows:

[1]The sequence $\{H_m\}$ are called the probabilists' Hermite polynomials, to avoid confusion with the orthogonal polynomials weighted by e^{-x^2}, known as the physicists' Hermite polynomials [10].

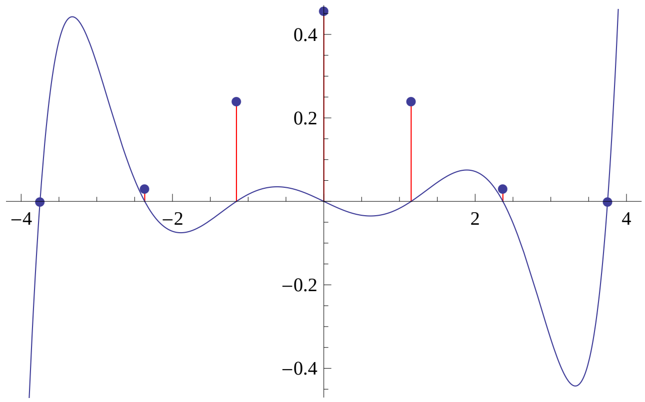


Fig. 1. The seven-point Gauss quadrature and scaled version of H_7.

- The peak amplitude $\left\|X_m^{\mathsf{Q}}\right\|_\infty$ is given by the largest root of H_m, which satisfies $2\sqrt{m} + \mathrm{o}\left(\frac{1}{\sqrt{m}}\right)$ [10, Theorem 6.32, p. 131].
- Because X_m^{Q} can be understood as an m-point fine quantization of a Gaussian random variable, its entropy grows according to $H(X_m^{\mathsf{Q}}) = \frac{1}{2}\log m(1 + \mathrm{o}(1))$ [13].

III. Properties of Finite-constellation Capacity

A. Existence of capacity-achieving input distribution

Denote by $\mathcal{M}$ be the collection of all probability measures on $(\mathbb{R}, \mathcal{B})$. Let

$$\mathcal{M}_m = \{P \in \mathcal{M} : \mathbb{E}[X] = 0, \mathbb{E}\left[X^2\right] \le 1, |\mathrm{supp}(P_X)| \le m\}. \tag{18}$$

It can be shown that $\mathcal{M}_m$ is weakly compact. Since

$$\sup_{P_X \in \mathcal{M}_m} \mathbb{E}\left[X^2\right] \le 1, \tag{19}$$

by [1, Theorem 9], $P_X \mapsto I(X, \mathsf{snr})$ is weakly continuous restricted on $\mathcal{M}_m$, hence achieving its maximum. Therefore, we have

Theorem 3. *For each m and each $\mathsf{snr} > 0$, $C_m(\mathsf{snr}) = \max_{P_X \in \mathcal{M}_m} I(X, \mathsf{snr})$.*

It should be noted that Theorem 3 does not state that the optimal input distribution is unique. Had uniqueness been established, it would follow that the optimal input distribution is symmetric.

B. Monotonicity and convergence

The following properties are straightforward from definitions and bounds on mutual information.

Theorem 4. *$(m, \mathsf{snr}) \mapsto C_m(\mathsf{snr})$ is increasing in each argument when the other argument is fixed, upper bounded by*

$$C_m(\mathsf{snr}) \le \min\left\{\log m, \frac{1}{2}\log(1 + \mathsf{snr})\right\}. \tag{20}$$

As a result of the I-MMSE relationship [14], the monotonicity of $\mathsf{snr} \mapsto C_m(\mathsf{snr})$ is strict, because for any non-deterministic X,

$$\frac{\mathrm{d}I(X, \mathsf{snr})}{\mathsf{dsnr}} = \frac{1}{2}\mathsf{mmse}(X|\sqrt{\mathsf{snr}}X + N) > 0. \tag{21}$$

We conjecture that $m \mapsto C_m(\mathsf{snr})$ is also strictly increasing.

The next result establishes the asymptotic normality of the capacity-achieving input distribution as the constellation size grows. The proof hinges on the weak lower semicontinuity of relative entropy [15].

Theorem 5. *For fixed $\mathsf{snr} > 0$,*

$$\lim_{m\to\infty} C_m(\mathsf{snr}) = \frac{1}{2}\log(1 + \mathsf{snr}). \tag{22}$$

Moreover, as $m \to \infty$, the optimal input distribution $P^*_{m,\mathsf{snr}} \to \mathcal{N}(0, 1)$ weakly.

IV. Low and High-SNR Asymptotics of Finite-constellation Capacity

In this section we consider the asymptotics of finite-constellation capacity and optimal input distribution in both low and high-SNR regimes when the constellation size is fixed. The following result shows that the optimal constellation in the high-SNR limit is the *equilattice*, proposed by Ungerboeck in [16] and analyzed subsequently by Ozarow and Wyner [17] in terms of mutual information. We establish its optimality in the sense that it achieves the high-SNR finite-constellation capacity.

Theorem 6. *For fixed $m \ge 2$, as $\mathsf{snr} \to 0$,*

$$\log m - C_m(\mathsf{snr}) = \mathrm{O}\left(\sqrt{\mathsf{snr}} \cdot \mathrm{e}^{-\frac{3\mathsf{snr}}{4(m^2-1)}}\right). \tag{23}$$

Moreover, as $\mathsf{snr} \to \infty$, $P^*_{m,\mathsf{snr}}$ converges weakly to the equiprobable distribution U_m on a uniformly spaced constellation $E_m \subset \mathbb{R}$, given by

$$E_m = \begin{cases} \left\{2i\Delta_m : i = \frac{1-m}{2}, \ldots, 0, \ldots, \frac{m-1}{2}\right\} & m \text{ odd} \\ \left\{(2i+1)\Delta_m : i = -\frac{m}{2}, \ldots, 0, \ldots, \frac{m-2}{2}\right\} & m \text{ even}, \end{cases} \tag{24}$$

where

$$\Delta_m = \sqrt{\frac{3}{m^2 - 1}}. \tag{25}$$

Proof sketch: Since $I(X, \mathsf{snr}) \to H(X)$ as $\mathsf{snr} \to \infty$, the optimal input must be equiprobable. Let $\hat{X}(Y)$ denote the optimal detector of X given Y. Then

$$\mathbb{P}\left\{X \neq \hat{X}\right\} = \mathrm{O}\left(Q(-d_{\min}\sqrt{\mathsf{snr}}/2)\right), \tag{26}$$

where $d_{\min}$ denotes the minimum pairwise distance in the constellation of X. Therefore X must be supported on the m-point configuration that maximizes the minimum distance subject to the average power constraint, which is the uniformly spaced constellation E_m. The rest of the proof follows from applying Fano's inequality. ■

Next we show that the Gaussian quadrature is optimal in the low-SNR regime and give an asymptotic expansion finite-constellation capacity.

Theorem 7. *For fixed $m \geq 2$, as $\mathsf{snr} \to \infty$,*

$$C(\mathsf{snr}) - C_m(\mathsf{snr}) = \Theta(\mathsf{snr}^{2m}). \tag{27}$$

Moreover, as $\mathsf{snr} \to \infty$, $P^*_{m,\mathsf{snr}}$ converges weakly to the m-point Gauss quadrature defined in Theorem 2.

Proof sketch:

Step 1. Using the I-MMSE relationship, it can be shown that $I(X, \cdot)$ defined in (2) is smooth on $\mathbb{R}_+$ if and only of X has all moments [18], which, in particular, holds for discrete random variable with finite support. This allows us to write $I(X, \mathsf{snr})$ as the Taylor expansion at $\mathsf{snr} = 0$ up to arbitrarily high order.

Step 2. Prove that the snr^k coefficient is a polynomial of the first k moments of X. This is the key argument of the proof.

Step 3. Since $X^* \sim \mathcal{N}(0, 1)$ is the natural maximizer of $I(X, \mathsf{snr})$ for any $\mathsf{snr} > 0$, it maximizes all coefficients simultaneously. To put it in other terms, the distribution that maximizes the m^{th} coefficient subject to the first $m - 1$ coefficients being optimal is the distribution that has the same m^{th} moment as the Gaussian.

Step 4. Now given m points, the optimal distribution maximizes the mutual information by matching as many moments of Gaussian as possible. By Theorem 2, the m-point Gauss quadrature X^{Q}_m match the first $2m-1$ moments. Therefore we have $C(\mathsf{snr}) - C_m(\mathsf{snr}) = O\left(\mathsf{snr}^{2m}\right)$. Since it can be computed that

$$\mathbb{E}\left[(X^*)^{2m}\right] - \mathbb{E}\left[(X^{\mathsf{Q}}_m)^{2m}\right] = m!, \tag{28}$$

we conclude that $C(\mathsf{snr}) - C_m(\mathsf{snr}) = \Omega(\mathsf{snr}^{2m})$.

∎

By integrating the Taylor expansion of MMSE in [18, (61)], we obtain the Taylor expansion of $I(X, \mathsf{snr})$ as follows: given $\mathbb{E}[X] = 0$ and $\mathbb{E}\left[X^2\right] = 1$,

$$\begin{aligned} I(X, \mathsf{snr}) = \frac{\log \mathrm{e}}{2}\Big[&\mathsf{snr} - \frac{1}{2}\mathsf{snr}^2 + (2 - (\mathbb{E}X^3)^2)\frac{\mathsf{snr}^3}{6} \\ &-\left(15 - 12(\mathbb{E}X^3)^2 - 6\,\mathbb{E}X^4 + (\mathbb{E}X^4)^2\right)\frac{\mathsf{snr}^4}{24}\Big] + O\left(\mathsf{snr}^5\right) \end{aligned} \tag{29}$$

For $m = 2$, it is easy to see that the optimal input is equiprobably distributed on $\{\pm 1\}$ for all snr, which is exactly the 2-point Gauss quadrature X^{Q}_2. In view of (28) and (29), we have

$$C(\mathsf{snr}) - C_2(\mathsf{snr}) = \frac{\log \mathrm{e}}{12}\mathsf{snr}^4 + O\left(\mathsf{snr}^5\right). \tag{30}$$

Note that the snr^4 coefficient is quadratic in $\mathbb{E}\left[X^4\right]$ and uniquely maximized by $\mathbb{E}\left[X^4\right] = 3$, the fourth moment of the standard Gaussian. For $m = 3$, the 3-point Gauss quadrature has the same first five moments and achieves a capacity gap of $\Theta(\mathsf{snr}^6)$.

Remark 2. As Shannon [19] observed, BPSK (antipodal signaling) achieves capacity in the low-SNR regime. Theorem 7 provides a finer quantitative justification of its optimality, in the sense that BPSK coincides with the 2-point Gauss quadrature and achieves the Gaussian channel capacity up to the third-order term.

Theorem 7 gives an optimality characterization of the Gauss-quadrature constellation in the low-SNR regime. Surprisingly, when the cardinality is large, this constellation achieves exponentially small gap to the capacity for all SNR (see Section V).

V. Lower bounds on finite-constellation Capacity

In order to bound the relative entropy, we define the following distances between probability measures [20]:

- The Hellinger distance between P and Q is

$$H(P, Q) = \sqrt{\int (\sqrt{\mathrm{d}P} - \sqrt{\mathrm{d}Q})^2}. \tag{31}$$

- The χ^2-distance between P and Q is

$$\chi^2(P, Q) = \int \left(\frac{\mathrm{d}P}{\mathrm{d}Q} - 1\right)^2 \mathrm{d}Q. \tag{32}$$

- The total variation distance between P and Q is

$$V(P, Q) = \int |\mathrm{d}P - \mathrm{d}Q|, \tag{33}$$

  which is equal to the L_1 distance between their respective densities with respect to a dominating measure.

Except for the χ^2-distance, the above definitions are metrics on the space of probability measures. Together with the relative entropy, they satisfy the following bounds:

Lemma 1.

- *[6, (2.4)]*

$$H^2(P, Q) \leq V(P, Q) \leq 2H(P, Q). \tag{34}$$

- *Csiszár-Kullback-Kemperman inequality:*

$$D(P||Q) \geq \frac{\log \mathrm{e}}{2} V^2(P, Q). \tag{35}$$

- *[20, p. 429]*

$$H^2(P, Q) \log \mathrm{e} \leq D(P||Q) \leq \chi^2(P, Q) \log \mathrm{e}. \tag{36}$$

Next we will work with the χ^2-distance which allows us to take a Hilbert-space approach and facilitates computation.

A. Upper bounds

For $P_{X_m} \in \mathcal{M}_m$, denote the density of the channel output $Y_m = \sqrt{\mathsf{snr}} X_m + N$ by

$$p_m(y) = \mathbb{E}\left[\varphi(y - \sqrt{\mathsf{snr}} X_m)\right]. \tag{37}$$

Then the ideal output density ($m = \infty$) is given by

$$p_\infty(y) = \mathbb{E}\left[\varphi(y - \sqrt{\mathsf{snr}} X_\infty)\right] = \varphi_{1+\mathsf{snr}}(y), \tag{38}$$

where $X_\infty \sim \mathcal{N}(0,1)$ and $\varphi_{\sigma^2}(y) \triangleq \frac{1}{\sigma}\varphi\left(\frac{y}{\sigma}\right)$ denotes the density of $\mathcal{N}(0,\sigma^2)$.

Let $\{\gamma_k\}_{k\geq 0}$ be an orthonormal basis of $L^2(\mathbb{R}, \mathrm{d}y)$. Then $\psi_k \triangleq \sqrt{p_\infty}\gamma_k$ forms an orthonormal basis for $L^2\left(\mathbb{R}, \frac{1}{p_\infty}\mathrm{d}y\right)$. Observe that $p_m \in L^2\left(\mathbb{R}, \frac{1}{p_\infty}\mathrm{d}y\right)$ for each m. This is obvious for $m = \infty$. For $m \in \mathbb{N}$, note that the likelihood ratio is upper bounded by

$$\frac{p_m}{p_\infty}(y)$$
$$= \mathbb{E}\left[\exp\left(-\frac{(y-X_m)^2}{2} + \frac{y^2}{2(1+\mathsf{snr})}\right)\right] \tag{39}$$
$$= \mathbb{E}\left[\exp\left(-\frac{\mathsf{snr}}{2(1+\mathsf{snr})}\left(y - \frac{1+\mathsf{snr}}{\sqrt{\mathsf{snr}}}X_m\right)^2 + \frac{X_m^2}{2}\right)\right] \tag{40}$$
$$\leq \mathbb{E}\left[\exp\left(\frac{X_m^2}{2}\right)\right], \tag{41}$$

hence

$$\int \frac{p_m^2}{p_\infty}\mathrm{d}y \leq \mathbb{E}\left[\exp\left(\frac{X_m^2}{2}\right)\right] < \infty. \tag{42}$$

Expanding p_m under the basis $\{\psi_k\}$, we have

$$\chi^2(p_m, p_\infty) = \|p_m - p_\infty\|^2_{L^2\left(\mathbb{R}, \frac{1}{p_\infty}\mathrm{d}y\right)} \tag{43}$$
$$= \sum_{k\geq 0} |\langle p_m, \psi_k\rangle - \langle p_\infty, \psi_k\rangle|^2. \tag{44}$$

Note that by Fubini's theorem and (37) and (38),

$$\langle p_m, \psi_k\rangle = \int p_m \sqrt{p_\infty}\gamma_k \frac{1}{p_\infty}\mathrm{d}y \tag{45}$$
$$= \int \mathbb{E}\left[\varphi(y - \sqrt{\mathsf{snr}}X_m)\right]\frac{\gamma_k}{\sqrt{p_\infty}}(y)\mathrm{d}y \tag{46}$$
$$= \mathbb{E}\left[\eta_k(\sqrt{\mathsf{snr}}X_m)\right], \tag{47}$$

where we define

$$\eta_k \triangleq \varphi * \frac{\gamma_k}{\sqrt{p_\infty}}. \tag{48}$$

Plugging (47) into (44), we have

$$\chi^2(p_m, p_\infty) = \sum_{k\geq 0}\left|\mathbb{E}\left[\eta_k(\sqrt{\mathsf{snr}}X_m)\right] - \mathbb{E}\left[\eta_k(\sqrt{\mathsf{snr}}X_\infty)\right]\right|^2, \tag{49}$$

which implies that in order for the output density p_m to approximate the Gaussian density p_∞ in the sense of χ^2-distance, the expectation of each η_k under X_m must approximate that under the standard Gaussian X_∞.

We choose the basis $\{\gamma_k\}$ carefully so that it is easy to evaluate the convolution in (48) as well as the expectation of η_k under the Gaussian measure. In view of (16), we choose the following orthonormal basis:

$$\gamma_k(y) = \sqrt{\frac{\varphi_{1+\mathsf{snr}}(y)}{k!}}H_k\left(\frac{y}{\sqrt{1+\mathsf{snr}}}\right). \tag{50}$$

Then by the convolution formula of Hermite polynomials [21, 7.374.8, p. 804], the convolution in (48) is given by

$$\eta_k(y) = \frac{1}{\sqrt{k!}}\left(\frac{\mathsf{snr}}{1+\mathsf{snr}}\right)^{\frac{k}{2}}H_k\left(\frac{y}{\sqrt{\mathsf{snr}}}\right). \tag{51}$$

By the orthogonality of the Hermite polynomials, for all $k \geq 1$, we have $\mathbb{E}\left[H_k(X_\infty)\right] = 0$, hence

$$\mathbb{E}\left[\eta_k(\sqrt{\mathsf{snr}}X_\infty)\right] = 0. \tag{52}$$

Also, $\eta_0 \equiv 1$. In view of (36) and (51), plugging (52) into (49) yields

$$D(p_m\|p_\infty) \leq \chi^2(p_m, p_\infty)\log \mathrm{e} \tag{53}$$
$$= \sum_{k\geq 1}\frac{\log \mathrm{e}}{k!}\left(\frac{\mathsf{snr}}{1+\mathsf{snr}}\right)^k\left|\mathbb{E}\left[H_k(X_m)\right]\right|^2. \tag{54}$$

B. Achievability by Gauss Quadrature

Since H_k is a polynomial of degree k, we immediately see from (54) that choosing the input to match the Gaussian moments will yield a small non-Gaussianness. To fulfill this requirement, a natural choice is the Gauss quadrature. Next we prove a non-asymptotic upper bound on the capacity gap based on this scheme.

Theorem 8. *For any $m \in \mathbb{N}$ and $\mathsf{snr} > 0$,*

$$D_m(\mathsf{snr}) \leq C(\mathsf{snr}) - I(X_m^{\mathsf{Q}}, \mathsf{snr}) \tag{55}$$
$$\leq 4(1+\mathsf{snr})\left(\frac{\mathsf{snr}}{1+\mathsf{snr}}\right)^{2m} \tag{56}$$

Proof: Note that $\mathbb{E}\left[H_k(X_m^{\mathsf{Q}})\right] = 0$ for all odd k by symmetry. By definition of Gauss quadrature, $\mathbb{E}\left[H_k(X_m^{\mathsf{Q}})\right] = 0$ for all $k \leq 2m-1$. Therefore by (54), we have

$$\chi^2(p_m, p_\infty) = \sum_{k\geq m}\frac{1}{(2k)!}\left(\frac{\mathsf{snr}}{1+\mathsf{snr}}\right)^{2k}\left|\mathbb{E}\left[H_{2k}(X_m^{\mathsf{Q}})\right]\right|^2. \tag{57}$$

Next we estimate the error term on higher-order moments of the Gaussian quadrature. By Cramér's inequality [21, p. 997],

$$|H_k(y)| \leq \kappa\sqrt{k!}\mathrm{e}^{\frac{y^2}{4}}, \tag{58}$$

where $\kappa \approx 1.086$ is an absolute constant. Therefore

$$\left|\mathbb{E}\left[H_{2k}(X_m^{\mathsf{Q}})\right]\right| \leq \kappa\sqrt{(2k)!}\,\mathbb{E}\left[\exp\left(\frac{(X_m^{\mathsf{Q}})^2}{4}\right)\right]. \tag{59}$$

Note that $(X_m^{\mathsf{Q}})^2 \to X_\infty^2$ in distribution and X_∞^2 is χ_1^2-distributed with density $p_{X_\infty^2}(z) = \frac{1}{\sqrt{2\pi z}}\mathrm{e}^{-\frac{z}{2}}$. Since X_∞^2 has an exponential tail with exponent $\frac{1}{2}$, by [22, Theorem

1], the moment generating function $\mathbb{E}\left[\exp\left(t(X_m^{\mathsf{Q}})^2\right)\right]$ converges to $\mathbb{E}\left[\exp\left(tX_\infty^2\right)\right] = \frac{1}{\sqrt{1-2t}}$ for all $t < \frac{1}{2}$. Therefore $\mathbb{E}\left[\exp\left(\frac{(X_m^{\mathsf{Q}})^2}{4}\right)\right] \to \sqrt{2}$. Moreover, it can be shown that the convergence is from below. Thus in view of (59) and (57), we have

$$\chi^2(p_m, p_\infty) \le 2\kappa^2 \sum_{k \ge m} \left(\frac{\mathsf{snr}}{1+\mathsf{snr}}\right)^{2k}, \tag{60}$$

where $2\kappa^2 \approx 2.36$. ∎

Fig. 2 shows the numerical value of the exponent achieved by the Gauss quadrature, that is,

$$\lim_{m\to\infty} \frac{1}{m} \log \frac{1}{C(\mathsf{snr}) - I(X_m^{\mathsf{Q}}, \mathsf{snr})}, \tag{61}$$

together with the lower and upper bounds in Theorem 1. We see that the performance of Gauss quadrature is near optimal. It is possible to improve the upper bound in Theorem 8, since using the Cramér inequality in (59) overestimates the error term of higher-order moments. More refined analysis entails exploiting the asymptotic expansion of Hermite polynomials and its oscillatory behavior.

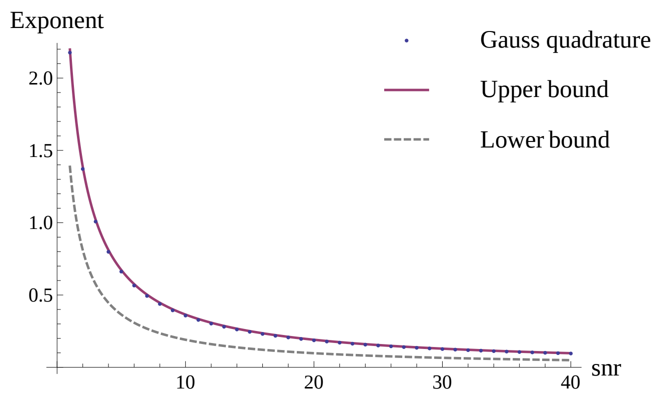


Fig. 2. Bounds on the optimal exponent $\lim_{m\to\infty} \frac{1}{m} \log \frac{1}{D_m(\mathsf{snr})}$.

VI. Converses

In this section we give several lower bounds on the capacity gap and sketch the proof of the converse part of Theorem 1.

A. Lower bound via L^2 distances

Note that (36) provides a lower bound on the output non-Gaussianness via the Hellinger distance, which, however, does not work well with the convolution structure of the output density in (37). Instead, we use the L^2 distance which allows us to take a similar orthogonal-expansion approach as in Section V-A.

Note that for bounded densities, L^2 distance is dominated by the Hellinger distance. From (37) and (38), we see that $\|p_m\|_\infty = \sup_y p_m(y) \le \frac{1}{\sqrt{2\pi}}$. Thus,

$$\|p_m - p_\infty\|_2^2 = \int (p_m - p_\infty)^2 \mathrm{d}y \tag{62}$$

$$\le \frac{4}{\sqrt{2\pi}} \int (\sqrt{p_m} - \sqrt{p_\infty})^2 \mathrm{d}y \tag{63}$$

$$= \frac{4}{\sqrt{2\pi}} H(p_m, p_\infty)^2, \tag{64}$$

which implies

$$D(p_m \| p_\infty) \ge \frac{\sqrt{2\pi}\log \mathrm{e}}{4} \|p_m - p_\infty\|_2^2. \tag{65}$$

Next we show that

$$\limsup_{m\to\infty} \frac{1}{m} \log \frac{1}{\|p_m - p_\infty\|_2^2} \le 2\log\left(1 + \frac{2}{\mathsf{snr}}\right), \tag{66}$$

which, together with (65), leads to the converse part of Theorem 1.

Proof sketch: Let

$$\alpha_k(y) = \sqrt{\frac{\varphi_{\frac{1}{2}}(y)}{k!}} H_k(\sqrt{2}y). \tag{67}$$

Then $\{\alpha_k\}_{k\ge 0}$ forms an orthonormal basis on $L^2(\mathbb{R}, \mathrm{d}y)$. Since p_m is square integrable, similar to the derivation in Section V-A, we have

$$\|p_m - p_\infty\|_2^2$$

$$= \sum_{k\ge 0} |\langle p_m, \alpha_k\rangle - \langle p_\infty, \alpha_k\rangle|^2 \tag{68}$$

$$= \sum_{k\ge 0} \left|\mathbb{E}\left[\varphi * \alpha_k(\sqrt{\mathsf{snr}}X_m)\right] - \mathbb{E}\left[\varphi * \alpha_k(\sqrt{\mathsf{snr}}X_\infty)\right]\right|^2 \tag{69}$$

$$= \sum_{k\ge 0} \frac{1}{\sqrt{\pi}2^{k+1}k!} \left|\mathbb{E}\left[\zeta_k(\sqrt{\mathsf{snr}}X_m)\right] - \mathbb{E}\left[\zeta_k(\sqrt{\mathsf{snr}}X_\infty)\right]\right|^2, \tag{70}$$

where we have used [21, 7.374.6, p. 803],

$$\varphi * \alpha_k(y) = \frac{1}{2^{\frac{k+1}{2}} \pi^{\frac{1}{4}} \sqrt{k!}} \zeta_k(y) \tag{71}$$

with $\zeta_k(y) \triangleq y^k \mathrm{e}^{-\frac{y^2}{4}}$. Note that ζ_k reaches its unique maximum

$$\zeta_k^* = \left(\frac{2k}{\mathrm{e}}\right)^{\frac{k}{2}} \tag{72}$$

at $\sqrt{2k}$. As k grows, ζ_k becomes increasingly concentrated at $\sqrt{2k}$. To see this, define

$$g(\alpha) = \mathrm{e}^{\frac{1-\alpha^2}{2}} \alpha. \tag{73}$$

Then $\zeta_k(\alpha\sqrt{2k}) = \zeta_k^* g(\alpha)^k$. Let $\mathbb{P}\{X_m = x_{im}\} = w_{im}$ for $i = 1, \ldots, m$. Then

$$\mathbb{E}\left[\zeta_k(\sqrt{\mathsf{snr}}X_\infty)\right] = \zeta_k^* \sum_{i=1}^m w_{im} \left[g\left(\frac{x_{im}}{\sqrt{2k}}\right)\right]^k. \tag{74}$$

On the other hand, for odd k, $\mathbb{E}\left[\zeta_k(\sqrt{\mathsf{snr}}X_\infty)\right] = 0$. For even k, direct calculation shows that

$$\mathbb{E}\left[\zeta_k(\sqrt{\mathsf{snr}}X_\infty)\right] = \sqrt{\frac{2}{\mathsf{snr}+2}}\left(\frac{\mathsf{snr}}{\mathsf{snr}+2}\right)^{\frac{k}{2}}\binom{k}{\frac{k}{2}}\left(\frac{k}{2}\right)! \tag{75}$$

$$\sim \left(\frac{\mathsf{snr}}{\mathsf{snr}+2}\right)^{\frac{k}{2}}\zeta_k^*. \tag{76}$$

Plugging (72), (74) and (76) into (70) and using Stirling's approximation, we have

$$\|p_m - p_\infty\|_2^2 \sim \sum_{k\geq 1}\frac{1}{\sqrt{k}}\left|\sum_{i=1}^m w_{im}\left[g\left(\frac{x_{im}}{\sqrt{2k}}\right)\right]^k - \left(\frac{\mathsf{snr}}{\mathsf{snr}+2}\right)^{\frac{k}{2}}\right|^2. \tag{77}$$

In order for the right-hand side (77) to be exponentially small, each sum $\sum_{i=1}^m w_{im}\left[g\left(\frac{x_{im}}{\sqrt{2k}}\right)\right]^k$ must has the same exponent $\sqrt{\frac{\mathsf{snr}}{\mathsf{snr}+2}}$. This imposes two restrictions on the distribution of X_m:

1) The weights need to behave like Gaussian asymptotically: $w_{im} \sim \exp\left(-\frac{x_{im}^2}{2}\right)$.
2) There is at least one atom in $[\sqrt{k}, \sqrt{k+2})]$ for $k = 0, 2, 4, \ldots$.

Since X_m only has m atoms, the $2m^{\text{th}}$ term in (77) cannot have the desired exponent, hence $\|p_m - p_\infty\|_2^2 \gtrsim \left(\frac{\mathsf{snr}}{\mathsf{snr}+2}\right)^{2m}$. ■

B. Lower bound via peak amplitude

The next result gives another lower bound on the capacity gap based on the peak amplitude:

Theorem 9. *Let $\|X_m\|_\infty \to \infty$. Then*

$$\limsup_{m\to\infty}\frac{1}{\|X_m\|_\infty^2}\log\frac{1}{D(p_m\|p_\infty)} \leq 2\limsup_{m\to\infty}\frac{1}{\|X_m\|_\infty^2}\log\frac{1}{V(p_m,p_\infty)} \tag{78}$$

$$\leq \frac{\mathsf{snr}}{(\sqrt{1+\mathsf{snr}}-1)^2}. \tag{79}$$

As a consequence of Theorem 9, in order to achieve exponential convergence to the capacity, it is necessary to have

$$\|X_m\|_\infty = \Omega(\sqrt{m}). \tag{80}$$

According to (79) and Remark 1, the exponent of the Gauss quadrature scheme is upper bounded by $\frac{4\mathsf{snr}}{(\sqrt{1+\mathsf{snr}}-1)^2}$, which is larger than the upper bound $2\log\left(1+\frac{2}{\mathsf{snr}}\right)$ in Theorem 1.

Proof: Inequality (78) follows from (35). To show (79), we prove the following lemma:

Lemma 2. *For any $0 < a < a'$,*

$$V(p_m,p_\infty) \geq 4Q\left(\frac{a'}{\sqrt{1+P}}\right) - 4Q(a'-a) - 2\mathbb{P}\left\{|X_m| > \frac{a}{\sqrt{\mathsf{snr}}}\right\}. \tag{81}$$

Choosing $a = \|X_m\|_\infty$ and $a' > \frac{\sqrt{1+P}}{\sqrt{1+P}-1}a$ arbitrarily close to equality yields the desired upper bound on the exponent in (79). To prove Lemma 2, note that

$$\mathbb{P}\{|Y_\infty| > a'\} = 2Q\left(\frac{a'}{\sqrt{1+\mathsf{snr}}}\right). \tag{82}$$

On the other hand, by union bound we have

$$\mathbb{P}\{|Y_m| > a'\} \leq \mathbb{P}\{|N| > a'-a\} + \mathbb{P}\left\{\sqrt{\mathsf{snr}}|X_m| \geq a\right\} \tag{83}$$

$$= 2Q(a'-a) + \mathbb{P}\left\{|X_m| \geq \frac{a}{\sqrt{\mathsf{snr}}}\right\}. \tag{84}$$

By definition of the total variation distance, we have

$$V(p_m,p_\infty) = 2\sup_A|\mathbb{P}\{Y\in A\} - \mathbb{P}\{Y_\infty\in A\}| \tag{85}$$

$$\geq 2\,\mathbb{P}\{|Y_\infty| > a'\} - 2\,\mathbb{P}\{|Y_m| > a'\}. \tag{86}$$

Substituting (82) and (84) into (86) yields (81). ■

VII. Comparison of Various Constellations

In this section we compare the performance of several achievable schemes, including

A. *Equilattice* constellation U_m defined in Theorem 6, which is capacity-achieving in the high-SNR regime.
B. *Gauss quadrature* X_m^{Q} defined in Theorem 2, which is capacity-achieving in the low-SNR regime.
C. *Quantized*: uniformly divide the Gaussian CDF into m segments and define an equiprobable input distribution with atoms given by [23]

$$x_{im} = \mathbb{E}\left[X_\infty|\alpha_{i,m} \leq X_\infty \leq \alpha_{i+1,m}\right], \quad i = 1,\ldots,m, \tag{87}$$

where $\alpha_{j,m} = \Phi^{-1}((j-1)/m)$, $j = 1,\ldots,m+1$. The constellation is then scaled to have unit variance.

D. *CLT*: let $\{Z_k\}$ be i.i.d. equiprobable on $\{\pm 1\}$. Define the normalized random walk:

$$\hat{X}_m = \frac{1}{\sqrt{m-1}}\sum_{k=1}^{m-1}Z_k, \tag{88}$$

$$\overset{\mathrm{D}}{=} \frac{2}{\sqrt{m-1}}\left(B_m - \frac{m-1}{2}\right), \tag{89}$$

where $B_m \sim \text{Binomial}(m-1, 1/2)$. It follows that $\hat{X}_m$ has m equally-spaced atoms. By the central limit theorem (CLT), $\hat{X}_m$ is asymptotically normal.

A. Performance analysis

As $m \to \infty$, U_m converges weakly to the uniform distribution U on $[-\sqrt{3}, \sqrt{3}]$. Therefore the equilattice constellation does not achieve the Gaussian capacity. In the high-SNR-large-constellation limit, the gap is given by

$$\lim_{\mathsf{snr}\to\infty} \lim_{m\to\infty} C(\mathsf{snr}) - I(U_m; \mathsf{snr})$$
$$= D(U \,||\, \mathcal{N}(0,1)) = \frac{1}{2} \log \frac{\pi \mathrm{e}}{6} \approx 0.25 \text{ bits}, \qquad (90)$$

Therefore at high SNR, the effective SNR loss due to using equilattice is $\frac{\pi e}{6} \approx 1.53$ dB (*e.g.*, [17], [24]).

Because of their asymptotic normality, for the constellation B, C and D, the capacity gap vanishes as $m \to \infty$. However, except for the Gauss quadrature, the other constellations achieve only polynomial convergence according to $\Theta\left(\frac{1}{m^2}\right)$. To see this, note that each moment is matched with $\frac{1}{m}$ precision, which, in view of (54), implies the convergence speed is $\mathrm{O}\left(\frac{1}{m^2}\right)$. On the other hand, the expectation of $\zeta_2(y) = y^2 \mathrm{e}^{-\frac{y^2}{4}}$ also has an error of $\frac{1}{m}$. Therefore the convergence rate is $\Omega\left(\frac{1}{m^2}\right)$, in view of (65) and (70).

B. Numerical Results

In Fig. 3 – 4 we plot the capacity gap $C_m(\mathsf{snr}) - I(X_m, \mathsf{snr})$ versus constellation size m for various input distributions; At $\mathsf{snr} = 0$ dB (Fig. 3), we see that the Gauss quadrature dominates all other schemes. At $\mathsf{snr} = 10$ dB (Fig. 4(a)), $\hat{X}_m$ seems to outperform the Gauss quadrature. This is because when snr is higher, more points are required for the exponential convergence of the Gauss quadrature. Indeed, as shown in the blowup in Fig. 4(b), Gauss quadrature achieves the smallest gap for $m \geq 25$ with exponential rate, while the convergence rate achieved by $\hat{X}_m$ is polynomially decaying. Nevertheless, empirical evidence suggests that $\hat{X}_m$ suffers from a relatively small capacity gap in the whole SNR range.

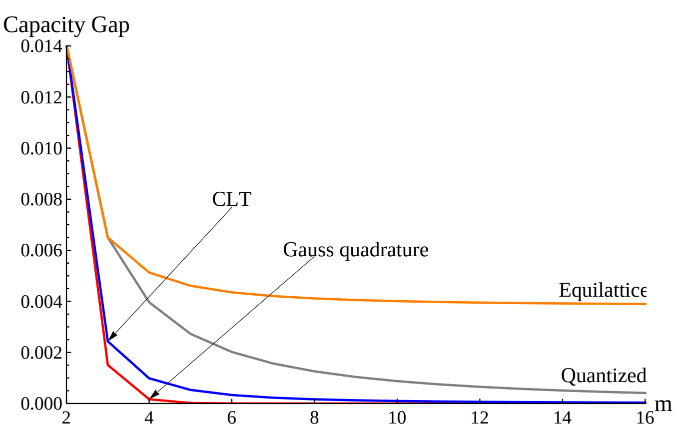
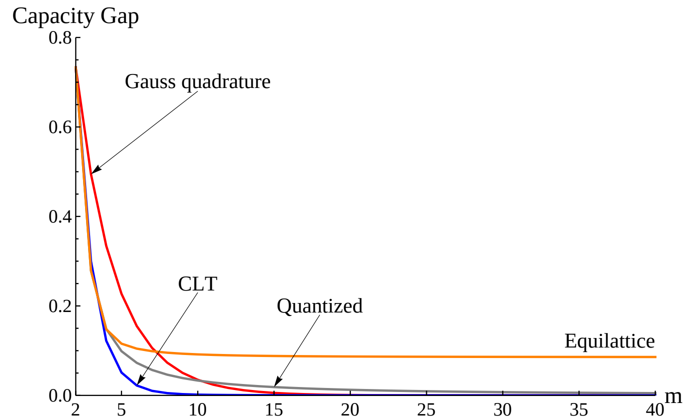
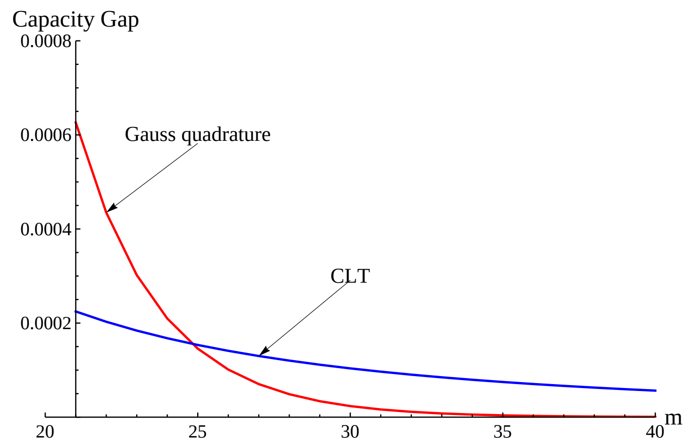


Fig. 3. Capacity gap $C_m(\mathsf{snr}) - I(X_m, \mathsf{snr})$ at $\mathsf{snr} = 0$ dB.



(a) $2 \leq m \leq 40$.



(b) $m \geq 20$.

Fig. 4. Capacity gap $C_m(\mathsf{snr}) - I(X_m, \mathsf{snr})$ at $\mathsf{snr} = 10$ dB.

Fig. 5 shows the capacity of the equilattice U_m and the Gauss quadrature X_m^{Q} as functions of SNR. We see that equilattice constellation suffers an SNR loss of 1.53 dB, while the Gauss quadrature has zero loss. On the other hand, according to Remark 1, $H(X_m^{\mathsf{Q}}) \sim \frac{1}{2}\log m = \frac{1}{2} H(U_m)$. Therefore for fixed m and large SNR, the capacity of the Gauss quadrature lies below that of equilattice, which is optimal in the high-SNR regime by Theorem 6.

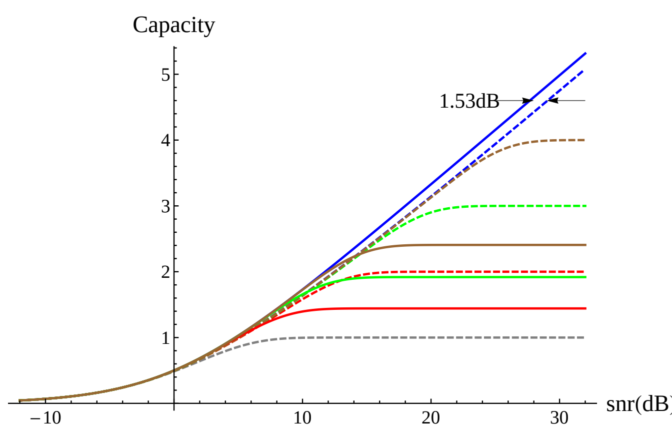


Fig. 5. Capacity of the Gaussian channel, the equilattice (dashed) and the Gauss quadrature (solid) constellations. Constellations of size 2, 4, 8, 16 and ∞ are denoted by gray, red, green, brown and blue respectively.

VIII. Concluding remarks

Apart from the point-to-point real-valued AWGN channel, our results can be generalized to complex field. However, it should be remarked that the optimal quadrature problem in higher-dimension is still open. For instance, optimal constructions for the two-dimensional Gaussian weight are unknown for more than 20 points. Even the asymptotics of $N^*(m)$ when m is large is unknown [25], [26]. Thus finding tight bounds on the optimal exponent appear to be challenging in the complex field.

Another interesting direction is the (coherent) fading channel. If the channel gain is known only at the receiver, the optimal input distribution is still standard Gaussian. If the encoder also has access to the channel gain, the optimal input is a Gaussian distribution with variance given by the water-filling solution. It is interesting to study whether exponential convergence still holds in these setups.

We conclude the paper by collecting a few open problems.

- Strict monotonicity of $m \to C_m(\mathsf{snr})$;
- Uniqueness and symmetry of the optimal input distribution.
- How does the peak power of the optimal input scales with constellation size? This question is closely related to Smith's classical result [27], which states that the optimal input distribution for AWGN channel with both amplitude and average power constraint is finitely supported. However, little is known about the cardinality of the support. Suppose $\sqrt{m}$, the peak amplitude of Gauss quadrature, is the optimal behavior. This would imply that the support increases quadratically as the peak constraint grows.
- Finding the optimal input support under finite-constellation constraint is a challenging problem because of its non-convexity. On a related note, Huang and Meyn [28] proposed an iterative algorithm to find the optimal input distribution for AWGN channel with both amplitude and average power constraint. It might be possible to apply their cutting-plane method to finite-constellation capacity problem. On the other hand, finding the optimal weights given the locations is a convex problem which can solved efficiently using the Blahut-Arimoto algorithm.
- In accordance to the practice of allocating larger constellation under higher SNR, it is interesting to consider $C_{m_{\mathsf{snr}}}(\mathsf{snr})$ with m_{snr} increasing with snr, *e.g.*, $m_{\mathsf{snr}} = \mathsf{snr}^{\alpha}$. In view of (20), the case of $\alpha = 1/2$ is particularly interesting.

References

[1] Y. Wu and S. Verdú, "Functional Properties of MMSE," in *Proceedings of IEEE International Symposium on Information Theory*, Austin, TX, June 2010.

[2] N. Wiener, "Tauberian theorems," *Annals of mathematics*, vol. 33, no. 1, pp. 1–100, 1932.

[3] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Transactions on Information theory*, vol. 39, no. 3, pp. 930–945, 1993.

[4] P. Ferreira, "Neural networks and approximation by superposition of Gaussians," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 1997, pp. 3197–3200.

[5] V. Maz'ya and G. Schmidt, "On approximate approximations using Gaussian kernels," *IMA Journal of Numerical Analysis*, vol. 16, no. 1, pp. 13 – 29, 1996.

[6] S. Ghosal and A. van der Vaart, "Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities," *The Annals of Statistics*, vol. 29, no. 5, pp. 1233–1263, 2001.

[7] C. R. Genovese and L. Wasserman, "Rates of convergence for the Gaussian mixture sieve," *Annals of Statistics*, vol. 28, no. 4, pp. 1105–1127, 2000.

[8] J. A. Cuesta-Albertos, "Shape of a distribution through the L_2-Wasserstein distance," in *Proceedings of Distributions with Given Marginals and Statistical Modelling*. Kluwer Academic Publishers, 2002, pp. 51 – 62.

[9] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 3rd ed. New York, NY: Springer-Verlag, 2002.

[10] G. Szegö, *Orthogonal polynomials*, 4th ed. Providence, RI: American Mathematical Society, 1975.

[11] G. Nikolov, "The Christoffel function for the Hermite weight is bell-shaped," *Journal of Approximation Theory*, vol. 125, no. 2, pp. 145–150, 2003.

[12] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York, NY: Wiley-Interscience, 1984.

[13] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, no. 1 – 2, March 1959.

[14] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261 – 1283, April 2005.

[15] P. Dupuis and R. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley-Interscience, 1997.

[16] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, 1982.

[17] L. H. Ozarow and A. D. Wyner, "On the capacity of the Gaussian channel with a finite number of input levels," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1426–1428, 1990.

[18] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," submitted to *IEEE Transactions on Information Theory*, 2010.

[19] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379 – 423, 623–656, 1948.

[20] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International Statistical Review*, vol. 70, no. 3, pp. 419–435, 2002.

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products*, 7th ed. New York, NY: Academic, 2007.

[22] W. Kozakiewicz, "On the convergence of sequences of moment generating functions," *The Annals of Mathematical Statistics*, pp. 61–69, 1947.

[23] F.-W. Sun and H. C. A. van Tilborg, "Approaching capacity by equiprobable signaling on the Gaussian channel," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1714 – 1716, 1993.

[24] G. D. Forney Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2384–2415, 1998.

[25] R. Cools, "Constructing cubature formulae: the science behind the art," *Acta Numerica*, vol. 6, pp. 1–54, 1997.

[26] ——, "Advances in multidimensional integration," *Journal of Computational and Applied Mathematics*, vol. 149, no. 1, pp. 1–12, 2002.

[27] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar gaussian channels," *Information and Control*, vol. 18, pp. 203 – 219, 1971.

[28] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2336 – 2351, 2005.