On the Capacity of the Peak Power Constrained Vector Gaussian Channel: An Estimation Theoretic Perspective

Alex Dytso*, Mert Al†, H. Vincent Poor‡ and Shlomo Shamai (Shitz)§

*,†,‡Department of Electrical Engineering, Princeton University

§Department of Electrical Engineering, Technion – Israel Institute of Technology

Email: adytso@princeton.edu*, merta@princeton.edu†, poor@princeton.edu‡ and sshlomo@ee.technion.ac.il§

***Abstract*—This paper studies the capacity of an n-dimensional vector Gaussian noise channel subject to the constraint that an input must lie in the ball of radius R centered at the origin. It is known that in this setting the optimizing input distribution is supported on a finite number of concentric spheres. However, the number, the positions and the probabilities of the spheres are generally unknown. This paper characterizes necessary and sufficient conditions on the constraint R such that the input distribution supported on a single sphere is optimal. The maximum $\bar{R}_n$, such that using only a single sphere is optimal, is shown to be a solution of an integral equation. Moreover, it is shown that $\bar{R}_n$ scales as $\sqrt{n}$ and the exact limit of $\frac{\bar{R}_n}{\sqrt{n}}$ is found.**

I. Introduction

We consider an additive noise channel for which the input-output relationships are given by

$$Y = X + Z, \tag{1}$$

where $X \in \mathbb{R}^n$ is independent of $Z \in \mathbb{R}^n$ and where $Z \sim \mathcal{N}(0, I_n)$. We are interested in finding the capacity of the channel in (1) subject to the constraint that $X \in \mathcal{B}_0(R)$ where $\mathcal{B}_0(R)$ is an n-ball centered at 0 of radius R (amplitude or peak power constraint), that is

$$\max_{X \in \mathcal{B}_0(R)} I(X;Y). \tag{2}$$

In general the capacity in (2) is an open problem and only some special cases have been solved. In this work the capacity in (2) will be characterized for all R that are smaller than roughly $\sqrt{n}$.

The necessity of characterization of the capacity with a peak power constraint on the input is self-evident. Many practical systems inherently have a peak power constraint due to the limited range of operations of electronic equipment. Some channels (e.g., the direct detection photon channel [1]) have well defined ranges of operations where average power constraints are not relevant and peak power constraints must be used.

This work was supported in part by the U. S. National Science Foundation under Grant CNS-1702808, and in part by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement no. 694630.

A. Prior Work

For the case of $n = 1$ Smith in his seminal work [2], using convex optimization techniques, has shown that the maximizing distribution in (2) must be discrete with finitely many points. In [3], for the case of $n = 2$, the maximizing input distribution has been shown to be supported on finitely many concentric spheres. The generalization to an arbitrary n can be found in [4], [5] and [6].

This paper can be considered as an n-dimensional generalization of the work in [7] where, in the case of $n = 1$ and under the conjecture that the number of mass points, as we vary R, increases by at most one, a two point input distribution uniform on $\pm R$ has been shown to be optimal if and only if $R \le 1.665$, and a three point input distribution on $\{-R, 0, R\}$ has been shown to be optimal if and only if $1.665 \le R \le 2.786$. However, unlike the approach in [7], the proof strategy used in this work relies on very different methods (rooted in estimation theory) and, for every dimension n, recovers the exact condition for the optimality of an input supported on a single sphere. Moreover, our proof does not require the assumption of the conjecture that the number of points increases by at most one as we vary R.

The fact that a uniform distribution on a single sphere is optimal as $\frac{R}{\sqrt{n}} \to 0$ has been shown in [5]. Moreover, the authors of [5] have observed via numerical results the fact that a distribution with the support on a single sphere can be optimal for non-vanishing values of R. In addition, the authors of [5] have computed the maximum values of R, for which a single sphere is optimal, up to $n = 20$.

A number of works have also focused on deriving lower and upper bounds on (2). The authors in [8] derived an asymptotically tight upper bound on the capacity as $R \to \infty$ by using the dual representation of channel capacity. In [9] the authors derived an upper bound on the capacity, by using a maximum entropy principle under L_p moment constraints, that is tight for small values of R. See also [9] and [10] for asymptotically tight lower bounds on the capacity.

The interested reader is also referred to [11] where in addition to the amplitude input constraint the authors also considered an average power constraint on the input and characterized the amplitude-to-power ratio of good codes.

B. Paper Outline and Contributions

The paper outline and contributions are as follows:

1) Section II reviews some known facts about the optimal input distribution in (2) (e.g., the support is given by concentric spheres);
2) Section II-A gives the definition of the "small amplitude" regime as the regime in which a uniform probability distribution supported on a single sphere is optimal;
3) Section III, Theorem 2, presents our main result, which is an exact characterization of the size of the small amplitude regime. The proof of the main result is postponed to Section V-A;
4) Section IV, for an input distribution on X uniformly distributed on a sphere of radius R, computes the output distribution, the conditional expectation of the input X given the output Y, the mutual information between X and Y and the minimum mean square error (MMSE) of estimating X from Y;
5) Section V presents new conditions for the optimality of the distribution on a single sphere. The new conditions have an advantage of being easier to verify than the classical conditions presented in Section II. The key ingredients for the proof of the new conditions are the change of sign lemma due to Karlin [12], the I-MMSE relationship [13] and the point-wise I-MMSE relationship [14]. To the best of our knowledge this is the first application of the point-wise I-MMSE relationship to a capacity problem;
6) Section V-A presents the proof of the main result;
7) Section VI gives an alternative proof, using yet another information estimation identity, that $R \le \sqrt{n}$ is sufficient for the optimality of the distribution on a single sphere; and
8) Section VII concludes the paper by discussing connections between maximization of the mutual information and maximization of the MMSE (i.e., the theory of finding least favorable prior distributions). In particular, we discuss conditions under which least favorable distributions are also capacity achieving.

C. Definitions and Notation

The volume of the unit n-ball and the unit $(n-1)$-sphere are denoted and given by

$$V_n \coloneqq \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}, \tag{3}$$

$$S_{n-1} \coloneqq \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}. \tag{4}$$

We denote the $(n-1)$-sphere of radius r centered at the origin as follows:

$$\mathcal{C}(r) \coloneqq \{x : \|x\| = r\}. \tag{5}$$

$Q(\cdot)$ denotes the tail distribution function of the standard normal distribution. The modified Bessel function of the first kind of order v is denoted by $\mathsf{I}_v(x)$. We also use the following commonly encountered ratio of Bessel functions:

$$\mathsf{h}_v(x) \coloneqq \frac{\mathsf{I}_v(x)}{\mathsf{I}_{v-1}(x)}. \tag{6}$$

We denote the distribution of a random variable X by P_X. Moreover, we say that a point x is in the support of the distribution P_X if for every open set $\mathcal{O}$ such that $x \in \mathcal{O}$ we have that $P_X(\mathcal{O}) > 0$ and denote the collection of the support points of P_X as $\mathsf{supp}(P_X)$.

At times it will be convenient to use the following parametrization of the mutual information in terms of the input distribution P_X:

$$I(P_X) \coloneqq I(X;Y). \tag{7}$$

We also define the following quantity that is akin to the information density:

$$i(x, P_X) \coloneqq \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\|y-x\|^2}{2}} \log \frac{1}{f_Y(y)} \mathrm{d}y - h(Z) \tag{8}$$

$$= \mathbb{E}\left[\log\left(\frac{f_{Y|X}(Y|X)}{f_Y(Y)}\right) \mid X = x\right], \tag{9}$$

where $f_Y(y)$ is the output probability density function (pdf) of Y induced by $X \sim P_X$ and $h(Z)$ is the entropy of Gaussian noise. Moreover, note that

$$\mathbb{E}[i(X, P_X)] = I(P_X). \tag{10}$$

The MMSE of estimating the input X from the output Y will be denoted as follows:

$$\mathrm{mmse}(X \mid Y) \coloneqq \mathbb{E}\left[\|X - \mathbb{E}[X \mid Y]\|^2\right]. \tag{11}$$

II. Optimizing The Input Distribution

The optimal input distribution in (2) can be characterized by using the method presented in [2] and its extension to the complex channel (i.e., $n = 2$) given in [3]; see also [4], [5] and [6] for a detailed solution for any $n \ge 2$.

Theorem 1. (Characterization of the Optimal Input Distribution) *Suppose $P_X^\star$ is an optimizer in* (2)*. Then, $P_X^\star$ satisfies the following properties:*

- *$P_X^\star$ is unique;*
- *$P_X^\star$ is optimal if and only if the following two conditions are satisfied:*

$$i(x, P_X^\star) = I(P_X^\star),\ x \in \mathsf{supp}(P_X^\star), \tag{12a}$$

$$i(x, P_X^\star) \le I(P_X^\star),\ x \in \mathcal{B}_0(R);\ \textit{and} \tag{12b}$$

- *the support of the optimal input distribution is given by*

$$\mathsf{supp}(P_X^\star) = \bigcup_{i=1}^{N} \mathcal{C}(r_i), \tag{12c}$$

where $N < \infty$ (finite).

An example of the support of distributions in (12c) for $n = 2$ is shown in Fig. 1.

Note that for $n = 1$ the optimal inputs are discrete with finitely many points. For $n > 1$ the optimal input probability

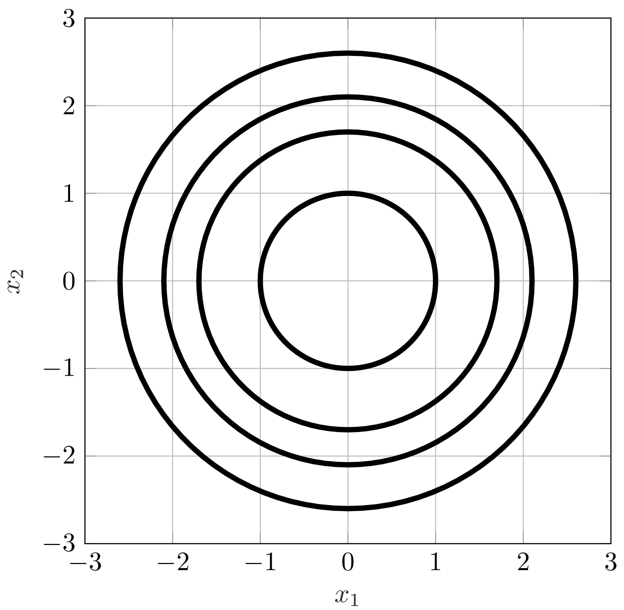


Fig. 1: An example of a support of an optimal input distribution for $n = 2$.

distributions are no-longer discrete but singular, however, the magnitude of the optimal input distribution $\|X\|$ is discrete with finitely many points.

A. Small Amplitude Regime

In this paper the small amplitude regime has the following definition.

Definition 1. Let $X_R \sim P_{X_R}$ be uniform on $\mathcal{C}(R)$. The capacity in (2) is said to be in *the small amplitude regime* if $R \leq \bar{R}_n$ where

$$\bar{R}_n := \max\{R : P_{X_R} = \arg\max \max_{X \in \mathcal{B}_0(R)} I(X;Y)\}. \quad (13)$$

In words, $\bar{R}_n$ is the largest radius R for which P_X uniformly distributed on $\mathcal{C}(R)$ is the capacity achieving distribution in (2).

In this work we are interested in exactly characterizing $\bar{R}_n$.

III. Main Result

The following theorem, which is the main result of this paper, gives a complete characterization of the small amplitude regime.

Theorem 2. (Characterization of the Small Amplitude Regime) *The input X_R is optimal in* (2) *(i.e., capacity achieving) if and only if $R \leq \bar{R}_n$ where $\bar{R}_n$ is given as the solution of the following equation:*

$$\int_0^1 \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma}R\|Z\|\right)\right] + \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma}R\|\sqrt{\gamma}x + Z\|\right)\right] \mathrm{d}\gamma = 1, \quad (14a)$$

for any x such that $\|x\| = R$. In addition, it is sufficient to take $R \leq \sqrt{n}$ (i.e., $\sqrt{n} \leq \bar{R}_n$), and

$$\lim_{n\to\infty} \frac{\bar{R}_n}{\sqrt{n}} = c \approx 1.860935682, \quad (14b)$$

where c is the solution of the following equation:

$$\int_0^1 \frac{\gamma c^2}{\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2}\right)^2} + \frac{\gamma c^2(1+\gamma c^2)}{\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2(1+\gamma c^2)}\right)^2} \mathrm{d}\gamma = 1. \quad (14c)$$

Proof: See Section V-A. ■

Note that R_n is given as the solution of an integral equation in (14a) and does not have an exact analytical form and must be found using numerical methods. Similarly, while the integral in (14c) does have a closed form expression given (62), the resulting equation must be solved numerically. The numerical evaluation of $\bar{R}_n$ up to $n = 35$ is shown on Fig. 2 and the values of $\bar{R}_n$ are provided in Table I.

It is important to note that numerical computation of $\mathsf{h}_{\frac{n}{2}}(x)$ via direct evaluations of the Bessel functions may be unstable for large x. The interested reader is referred to Appendix A for a discussion of these stability issues and details on how values of $\bar{R}_n$ can be computed by using a known continued fraction expansion of $\mathsf{h}_{\frac{n}{2}}(x)$.

Remark 1. Recall that $\|Z + x\|^2$ in (14a) is distributed according to the non-central chi-square distribution of degree n with non-centrality parameter $\|x\|^2$; this fact becomes useful when numerically computing $\bar{R}_n$.

We can also give the following alternative characterization of $\bar{R}_n$ that does not require integration over γ as in (14a).

Theorem 3. (Alternative Characterization of $\bar{R}_n$) *The input X_R is optimal in* (2) *if and only if $R \leq \bar{R}_n$ where $\bar{R}_n$ is given as a positive zero of the following equation:*

$$\mathbb{E}\left[\frac{W_1}{\|W\|}\mathsf{h}_{\frac{n}{2}}\left(R\|W\|\right)\right] = \frac{1}{2}, \quad (15)$$

where W is a random vector of independent components such that $W_1 \sim \frac{Q(w-R)-Q(w)}{R}$ and $W_i \sim \mathcal{N}(0,1)$ for $2 \leq i \leq n$.

Proof: See Section V-B. ■

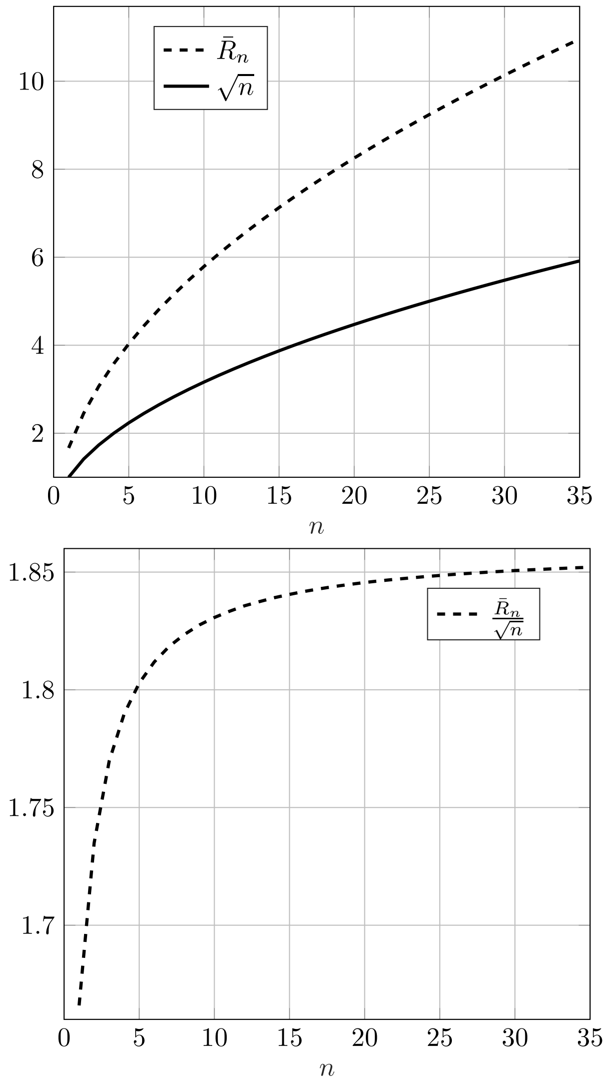
Remark 2. For the case of $n = 1$ using the fact that

$$R\frac{y}{|y|}\mathsf{h}_{\frac{1}{2}}\left(R|y|\right) = \mathbb{E}[X_R|Y = y] = R\tanh(Ry), \quad (16)$$

the expression in (15) simplifies to

$$\int_{\mathbb{R}} \left(Q(w - R) - Q(w)\right)\tanh(Rw)\mathrm{d}w = \frac{R}{2}. \quad (17)$$

The non-zero solution to (17) can be easily found numerically and is given by $\bar{R}_1 \approx 1.665925641$ as was already computed in [7]. However, interestingly, while the expression (17) is equivalent to the one presented in [7], it is not of the same



(b) Plot of $\bar{R}_n$ normalized by $\sqrt{n}$.

Fig. 2: Plots of $\bar{R}_n$ as defined in Theorem 2 vs. n.

form. In [7] $\bar{R}_1$ is instead given as a solution of the following equation:

$$\int_{\mathbb{R}} \left(\frac{\mathrm{e}^{-\frac{(y-R)^2}{2}} + \mathrm{e}^{-\frac{(y+R)^2}{2}}}{2} - \mathrm{e}^{-\frac{y^2}{2}} \right) \log\left(\mathrm{e}^{-Ry} + \mathrm{e}^{Ry}\right) \mathrm{d}y$$
$$= \frac{\sqrt{2\pi}R^2}{2}.$$

IV. Some Analytical Computations

In this section for the input X_R we compute the output pdf, the mutual information and the MMSE.

Proposition 1. (Output Distribution) *The pdf of the output distribution induced by the input X_R is given by*

$$f_Y(y) = \frac{\Gamma\left(\frac{n}{2}\right) \mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}}{2\pi^{\frac{n}{2}}} \frac{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}-1}}. \quad (18)$$

Proof: Let $\hat{X}_\epsilon$ have distribution with the pdf given on the annulus

$$f_{\hat{X}_\epsilon}(x) = \frac{1}{V_n\left(R^n - (R-\epsilon)^n\right)} 1_{\{R-\epsilon \le \|\hat{X}_\epsilon\| \le R\}}(x), \quad (19)$$

for some $\epsilon > 0$. Observe that $\hat{X}_\epsilon \to X_R$ in distribution as $\epsilon \to 0$ and, therefore, by the Dominated Convergence Theorem the output pdf can be written as follows:

$$f_Y(y) = \lim_{\epsilon \to 0} \mathbb{E}\left[\frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\|y-\hat{X}_\epsilon\|^2}{2}} \right]$$
$$= \frac{1}{V_n (2\pi)^{\frac{n}{2}}} \lim_{\epsilon \to 0} \frac{\int_{R-\epsilon \le \|x\| \le R} \mathrm{e}^{-\frac{\|y-x\|^2}{2}} \mathrm{d}x}{(R^n - (R-\epsilon)^n)}. \quad (20)$$

To compute the limit in (20) we will need the following integral [15]:

$$\int_{\|x\|=1} \mathrm{e}^{x^T y R} \mathrm{d}x = \left(\frac{\|y\|R}{2} \right)^{1-\frac{n}{2}} S_{n-1} \Gamma\left(\frac{n}{2}\right) \mathsf{I}_{\frac{n}{2}-1}(\|y\|R). \quad (21)$$

The derivation of the limit in (20) now proceeds as follows:

$$\lim_{\epsilon \to 0} \frac{\int_{R-\epsilon \le \|x\| \le R} \mathrm{e}^{-\frac{\|y-x\|^2}{2}} \mathrm{d}x}{(R^n - (R-\epsilon)^n)}$$
$$\overset{a)}{=} \lim_{\epsilon \to 0} \frac{\int_{R-\epsilon}^{R} \int_{S_{n-1}} \mathrm{e}^{-\frac{\|y-r\Theta\|^2}{2}} r^{n-1} \mathrm{d}\Theta \mathrm{d}r}{(R^n - (R-\epsilon)^n)}$$
$$\overset{b)}{=} \lim_{\epsilon \to 0} \frac{\int_{R-\epsilon}^{R} f(r,y) \mathrm{d}r}{(R^n - (R-\epsilon)^n)}$$
$$\overset{c)}{=} \frac{\frac{d}{d\epsilon} \int_{R-\epsilon}^{R} f(r,y) \mathrm{d}r |_{\epsilon=0}}{\frac{d}{d\epsilon} (R^n - (R-\epsilon)^n) |_{\epsilon=0}}$$
$$\overset{d)}{=} \frac{f(R,y)}{nR^{n-1}}$$
$$= \frac{\int_{S_{n-1}} \mathrm{e}^{-\frac{\|y-R\Theta\|^2}{2}} R^{n-1} \mathrm{d}\Theta}{nR^{n-1}}$$
$$= \frac{\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}} \int_{S_{n-1}} \mathrm{e}^{R\Theta^T y} \mathrm{d}\Theta}{n}$$
$$\overset{e)}{=} \frac{\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}} \left(\frac{\|y\|R}{2} \right)^{1-\frac{n}{2}} S_{n-1} \Gamma\left(\frac{n}{2}\right) \mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{n}, \quad (22)$$

where the labeled equalities follow from: a) changing to spherical coordinates; b) defining $f(r) := \int_{S_{n-1}} \mathrm{e}^{-\frac{\|y-r\Theta\|^2}{2}} r^{n-1} \mathrm{d}\Theta$; c) applying L'Hôpital's rule; d) applying the Fundamental Theorem of Calculus; and e) using the integral in (21).

Putting together (20) and (22) and using (21) we have that the output pdf is given by

$$f_Y(y)$$
$$= \frac{1}{V_n (2\pi)^{\frac{n}{2}}} \frac{\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}} \left(\frac{\|y\|R}{2} \right)^{1-\frac{n}{2}} S_{n-1} \Gamma\left(\frac{n}{2}\right) \mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{n}$$
$$= \frac{\Gamma\left(\frac{n}{2}\right) \mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}}{2\pi^{\frac{n}{2}}} \frac{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}-1}}.$$

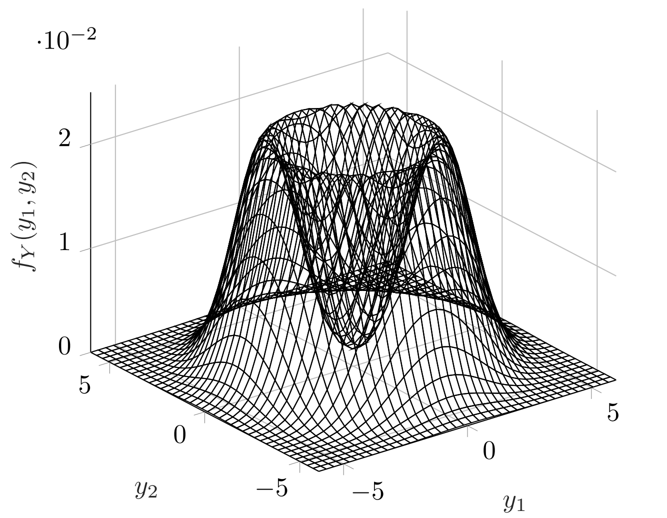


Fig. 3: The output pdf in (18) for $n = 2$ and $R = 3$.

This concludes the proof. ∎

For $n = 1$ using the identity $\mathsf{I}_{-\frac{1}{2}}(x) = \left(\frac{2}{\pi x}\right)^{\frac{1}{2}} \cosh(x)$ we have that

$$f_Y(y) = \frac{\mathrm{e}^{-\frac{R^2+|y|^2}{2}}}{\sqrt{2\pi}} \cosh(|y|R)$$
$$= \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(R+|y|)^2}{2}} + \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(R-|y|)^2}{2}}\right).$$

For $n = 2$ the output distribution is shown in Fig. 3 and is given by

$$f_Y(y) = \frac{\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}}{2\pi^2} \int_0^\pi \mathrm{e}^{\|y\|R\cos(\theta)}\mathrm{d}\theta;$$

for $n = 3$ using the identity $\mathsf{I}_{\frac{1}{2}}(x) = \left(\frac{2}{\pi x}\right)^{\frac{1}{2}} \sinh(x)$ we have that

$$f_Y(y) = \frac{\sqrt{2}}{8\pi^{\frac{3}{2}}} \frac{1}{\|y\|R}\left(\mathrm{e}^{-\frac{(R-\|y\|)^2}{2}} - \mathrm{e}^{-\frac{(R+\|y\|)^2}{2}}\right).$$

Using the expression for the pdf in (18) we can now also compute the conditional expectation $\mathbb{E}[X|Y]$.

Proposition 2. (Conditional Expectation) *For every* $R > 0$

$$\mathbb{E}[X_R \mid Y = y] = \frac{Ry}{\|y\|}\mathsf{h}_{\frac{n}{2}}\left(\|y\|R\right). \tag{23}$$

Proof: Using the identity between the conditional expectation and score function [16] we have that

$$\mathbb{E}[X_R \mid Y = y] = y + \frac{\nabla_y f_Y(y)}{f_Y(y)}, \tag{24}$$

and due to the symmetry of $f_Y(y)$ we have that

$$\nabla_y f_Y(y) = \frac{y}{\|y\|}\frac{d}{d\|y\|} f_Y(\|y\|), \tag{25}$$

where

$$\frac{d}{d\|y\|} f_Y(\|y\|)$$
$$= \frac{d}{d\|y\|}\frac{\Gamma\left(\frac{n}{2}\right)\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}}{2\pi^{\frac{n}{2}}}\frac{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}-1}}$$
$$= -\frac{\Gamma\left(\frac{n}{2}\right)\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}\|y\|}{2\pi^{\frac{n}{2}}}\frac{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}-1}}$$
$$+ \frac{\Gamma\left(\frac{n}{2}\right)\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}}{2\pi^{\frac{n}{2}}}\frac{d}{d\|y\|}\frac{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}-1}}$$
$$= -\|y\|f_Y(y) + \frac{\Gamma\left(\frac{n}{2}\right)\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}R}{2\pi^{\frac{n}{2}}}$$
$$\cdot\left(\frac{(\|y\|R)\mathsf{I}'_{\frac{n}{2}-1}(\|y\|R) - \left(\frac{n}{2}-1\right)\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}}}\right) \tag{26}$$
$$\overset{a)}{=} -\|y\|f_Y(y) + \frac{\Gamma\left(\frac{n}{2}\right)\mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}}{2\pi^{\frac{n}{2}}}\left(\frac{R^2\|y\|\mathsf{I}_{\frac{n}{2}}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}}}\right)$$
$$\overset{b)}{=} -\|y\|f_Y(y) + \frac{Rf_Y(y)\mathsf{I}_{\frac{n}{2}}(\|y\|R)}{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}$$
$$= -\|y\|f_Y(y) + Rf_Y(y)\mathsf{h}_{\frac{n}{2}}(\|y\|R), \tag{27}$$

where the labeled equalities follow from: a) using the well-known recurrence relation $x\mathsf{I}_{v+1}(x) = x\mathsf{I}'_v(x) - v\mathsf{I}_v(x)$ [17]; and b) using the expression for $f_Y(y)$ in (18).

The proof of (23) is completed by combining (24), (25) and (27). ∎

Remark 3. The proof of Proposition 2 relies on the following identity between the conditional expectation and the output pdf [16]:

$$\mathbb{E}[X \mid Y = y] = y + \frac{\nabla_y f_Y(y)}{f_Y(y)}, \tag{28}$$

in which the quantity $\frac{\nabla_y f_Y(y)}{f_Y(y)}$ is commonly known as the score function. The application of the identity in (28) considerably simplifies the computation of $\mathbb{E}[X \mid Y]$ as we do not need to derive the conditional distribution $P_{X|Y}$ and only use properties of the output pdf $f_Y(y)$.

Examples of shapes of the conditional expectation for $n = 1$ and $n = 2$ are shown on Fig. 4.

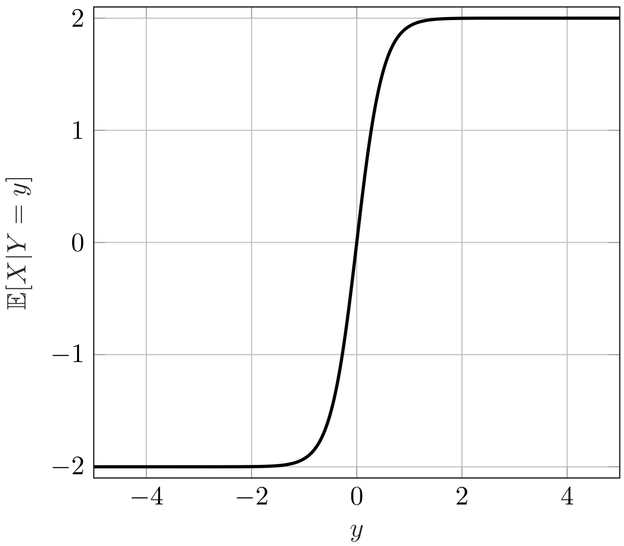
The mutual information and the MMSE of X_R are given next.

Proposition 3. (MMSE and Mutual Information) *For every* $R > 0$

$$I(X_R; Y) = R^2\log(\mathrm{e}) + \log\left(\frac{2^{1-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\right)$$
$$- \mathbb{E}\left[\log\left(\frac{\mathsf{I}_{\frac{n}{2}-1}(\|Z + x\|R)}{(\|Z + x\|R)^{\frac{n}{2}-1}}\right)\right], \tag{29}$$

and

$$\mathrm{mmse}(X_R \mid Y) = R^2 - R^2\mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(R\|x + Z\|\right)\right], \tag{30}$$



(a) Case of $n = 1$. The conditional expectation $\mathbb{E}[X_R|Y=y] = R\tanh(Ry)$ where $R = 2$.

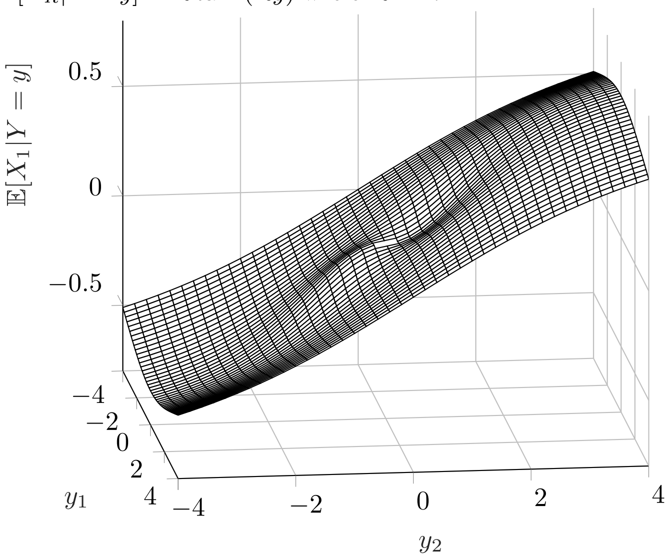


Fig. 4: Examples of the conditional expectation in (23) for $n = 1$ and $n = 2$.

for any $\|x\| = R$.

Proof: First observe that due to the symmetry of $i(x, P_{X_R})$ and X_R we have that

$$I(X_R; Y) = \mathbb{E}[i(X, P_{X_R})] = i(x, P_{X_R}),$$

where $\|x\| = R$. Therefore,

$$\begin{aligned} & I(X_R; Y) + h(Z) \\ &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\|y-x\|^2}{2}} \log \frac{1}{f_Y(y)} \mathrm{d}y \\ &= \log\left(\frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\right) + \frac{R^2 + \mathbb{E}[\|Z+x\|^2]}{2} \log(\mathrm{e}) \\ &+ \mathbb{E}\left[\log\left(\frac{(\|Z+x\|R)^{\frac{n}{2}-1}}{\mathsf{I}_{\frac{n}{2}-1}(\|Z+x\|R)}\right)\right] \end{aligned}$$

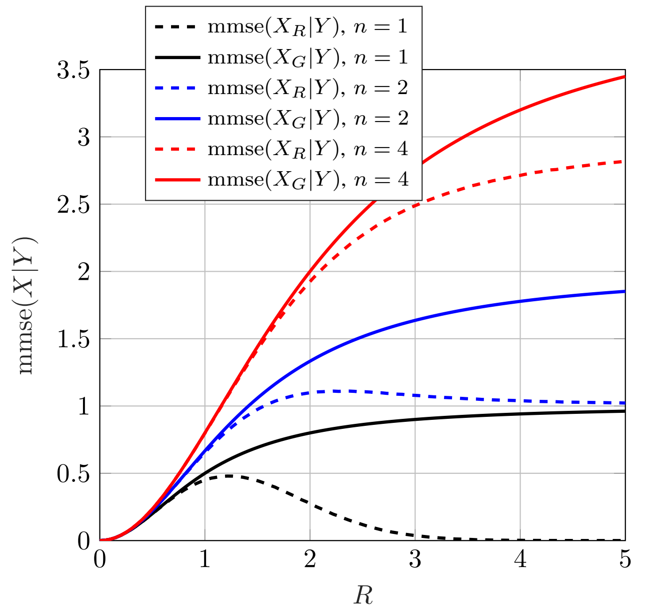


Fig. 5: Comparison of the MMSE of X_R (dashed line) and the MMSE of $X_G \sim \mathcal{N}(0, R^2 I_n)$ (solid line).

$$\begin{aligned} &= \log\left(\frac{2(\pi\mathrm{e})^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\right) + R^2 \log(\mathrm{e}) \\ &+ \mathbb{E}\left[\log\left(\frac{(\|Z+x\|R)^{\frac{n}{2}-1}}{\mathsf{I}_{\frac{n}{2}-1}(\|Z+x\|R)}\right)\right]. \end{aligned}$$

This concludes the proof of (29). To show (30) observe that the MMSE can be written as

$$\begin{aligned} \mathrm{mmse}(X_R|Y) &= \mathbb{E}[\|X_R\|^2] - \mathbb{E}\left[\|\mathbb{E}[X_R \mid Y]\|^2\right] \\ &= R^2 - \mathbb{E}\left[\|\mathbb{E}[X_R \mid Y]\|^2 \mid \|X_R\| = R\right] \\ &= R^2 - R^2 \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(R\|x+Z\|\right)\right], \end{aligned}$$

where $\|x\| = R$. This concludes the proof. ■

Fig. 5 shows the MMSE of X_R in (30) vs. the MMSE of $X_G \sim \mathcal{N}(0, R^2 I_n)$ which is given by

$$\mathrm{mmse}(X_G|Y) = n \frac{\frac{1}{n}R^2}{1 + \frac{1}{n}R^2}. \tag{31}$$

V. A New Condition for Optimality in the Small Amplitude Regime

In this section an equivalent optimality condition to that in Theorem 1 is derived. The new condition has the advantage of being easier to verify than the condition in Theorem 1.

The following two lemmas would be useful in our analysis.

Lemma 1. *The function* $x \mapsto i(x, P_{X_R})$ *is a function only of* $\|x\|$.

Proof: The proof follows from the symmetry of the Gaussian distribution and the symmetry of P_{X_R}. ■

The next lemma was shown in [12, Theorem 3].

Lemma 2. *Let the pdf $f(x,\omega)$ be a positive-definite kernel that can be differentiated n times with respect to x for all ω, and let $\eta(\omega)$ be a function that changes sign n times. If*

$$M(x) \coloneqq \int \eta(\omega) f(x,\omega) \mathrm{d}\omega, \tag{32}$$

can be differentiated n times, then $M(x)$ changes sign at most n times.

Theorem 4. (A New Optimality Condition) *P_{X_R} is optimal if and only if for all $\|x\| = R$*

$$i(0, P_{X_R}) \le i(x, P_{X_R}), . \tag{33}$$

Proof: Since by Lemma 1 $i(x, P_{X_R})$ is a function only of $\|x\|$ let

$$g(\|x\|) \coloneqq i(x, P_{X_R}). \tag{34}$$

The goal now is to show that the maximum of $g(\|x\|)$ for $x \in \mathcal{B}_0(R)$ occurs either at $\|x\| = 0$ or $\|x\| = R$. This would simplify the two conditions in (12a) and (12b) to only one condition

$$g(0) \le g(R). \tag{35}$$

In order to show this claim, we prove that the derivative of $g(\|x\|)$ makes only one sign change, and that sign change is from negative to positive. Hence, $g(\|x\|)$ has only one local minimum and must be maximized only at the boundaries $\|x\| = 0$ and $\|x\| = R$.

Because $g(\|x\|)$ depends on x only through $\|x\|$, there is no loss of generality in taking $x = [x_1, 0, ..., 0]$. Consider the derivative of $g(x_1)$ with respect to x_1

$$\begin{aligned} &g'(x_1) \\ &= \frac{d}{dx_1} \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\sum_{i=2}^n y_i^2 + (y_1 - x_1)^2}{2}} \log \frac{1}{f_Y(y)} \mathrm{d}y \\ &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\sum_{i=2}^n y_i^2 + (y_1 - x_1)^2}{2}} (y_1 - x_1) \log \frac{1}{f_Y(y)} \mathrm{d}y \\ &= -\int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\sum_{i=2}^n y_i^2 + (y_1 - x_1)^2}{2}} (y_1 - x_1) \log f_Y(y) \mathrm{d}y. \end{aligned}$$

Integrating by parts with respect to y_1 we have that

$$g'(x_1) = -\int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\sum_{i=2}^n y_i^n + (y_1 - x_1)^2}{2}} \rho(y) \mathrm{d}y,$$

where

$$\rho(y) \coloneqq \frac{d}{dy_1} \log f_Y(y) = \frac{\frac{d}{dy_1} f_Y(y)}{f_Y(y)}.$$

Next using the chain rule of differentiation we have that

$$\begin{aligned} \frac{d}{dy_1} f_Y(y) &= \frac{d}{d\|y\|} f_Y(y) \frac{y_1}{\|y\|} \\ &= \left(-\|y\| f_Y(y) + \frac{R f_Y(y) \mathsf{I}_{\frac{n}{2}}(\|y\|R)}{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)} \right) \frac{y_1}{\|y\|}, \end{aligned}$$

where in the last step we have used (27). Therefore,

$$\begin{aligned} \rho(y) &= \left(-\|y\| + R \mathsf{h}_{\frac{n}{2}}(\|y\|R)\right) \frac{y_1}{\|y\|} \\ &\coloneqq M(\|y\|) \frac{y_1}{\|y\|}. \end{aligned}$$

Next by transforming to spherical coordinates we have that

$$g'(x_1) = -2x_1 \int_0^\infty M(r) \mathrm{e}^{-\frac{r^2 + x_1^2}{2}} \frac{1}{2} \left(\frac{r}{x_1}\right)^{\frac{n}{2}} \mathsf{I}_{\frac{n}{2}}(x_1 r) \mathrm{d}r \tag{36}$$

$$= -2x_1 \mathbb{E}\left[M(\sqrt{V^2})\right], \tag{37}$$

where V^2 is the non-central chi-square distribution with $n+2$ degrees of freedom and non-centrality x_1^2; see Appendix B for the derivation (36) and (37).

Another fact which is not difficult to check is that for large enough x_1 the function $g'(x_1)$ is positive. This is shown next

$$\begin{aligned} -2x_1 \mathbb{E}\left[M\left(\sqrt{V^2}\right)\right] &= -2x_1 \mathbb{E}\left[-\sqrt{V^2} + R \mathsf{h}_{\frac{n}{2}}\left(\sqrt{V^2} R\right)\right] \\ &= 2x_1 \left(\mathbb{E}\left[\sqrt{V^2}\right] - \mathbb{E}\left[R \mathsf{h}_{\frac{n}{2}}(\sqrt{V^2} R)\right]\right) \\ &\overset{a)}{\ge} 2x_1 \left(\sqrt{\mathbb{E}[V^2]} - R\right) \\ &\overset{b)}{\ge} 2x_1 \left(\sqrt{n + 2 + x_1^2} - R\right), \end{aligned} \tag{38}$$

where the labeled (in)-equalities follow from: a) using $\mathsf{h}_{\frac{n}{2}}(\sqrt{V^2} R) \le 1$ and $\mathbb{E}\left[\sqrt{V^2}\right] \ge \sqrt{\mathbb{E}[V^2]}$; and b) using the expression of the mean of the non-central chi-square distribution with $n+2$ degrees of freedom and non-centrality x_1^2. Therefore, in view of the bound in (38), the expression in (37) is positive for x_1 large enough.

Next observe that in (36) the function

$$M(r) = -r + R \mathsf{h}_{\frac{n}{2}}(rR),$$

changes sign at most once for $r > 0$, which follows from the fact that $\mathsf{h}_{\frac{n}{2}}(x)$ is increasing and concave (see [15]) and $\mathsf{h}_{\frac{n}{2}}(0) = 0$. Hence, using Lemma 2 we have that for $x_1 > 0$ the function $g'(x_1)$ changes sign at most once, and since $g'(x_1) > 0$ for large enough x_1, we conclude that the sign change can only be from negative to positive. Therefore, for $x_1 > 0$ the function $g(x_1)$ has only one local minimum, no local maxima, and $g(\|x\|)$ is maximized only at the boundaries. This concludes the proof. ■

Remark 4. Condition (33) significantly simplifies the necessary and sufficient conditions for optimality in (12). For instance, we do not have to verify the conditions in (12b) for all $x \in \mathcal{B}_0(R)$ and instead need only to check the points satisfying $\|x\| = 0$ and $\|x\| = R$.

Moreover, the condition in (33) implies that as we increase R the new points of support cannot appear for $0 < \|x\| < R$ and shows that a new probability mass, as we transition beyond $\bar{R}_n$, can only appear at $\|x\| = 0$.

Fig. 6 shows $i(x, X_R)$ vs. x for $n = 1$. Note that, as expected, when $R = \bar{R}_1$ we have that $i(0, P_{X_R}) = i(R, P_{X_R})$.

Moreover, for $R > \bar{R}_1$ as X_R is no longer optimal, we have that $i(0, P_{X_R}) > i(R, P_{X_R})$

Next, we rewrite $i(0, P_{X_R})$ and $i(x, P_{X_R})$ in terms of estimation theoretic measures which facilitates the computation of $\bar{R}_n$.

Lemma 3. *For every $R > 0$ and $\|x\| = R$*

$$i(x, P_{X_R}) = \frac{1}{2} \int_0^1 \mathbb{E}\left[\|X_R - \mathbb{E}[X_R \mid Y_\gamma]\|^2 \mid \|X_R\| = R\right] \mathrm{d}\gamma, \tag{39}$$

$$i(0, P_{X_R}) = \frac{1}{2} \int_0^1 \mathbb{E}\left[\|X_R - \mathbb{E}[X_R \mid Y_\gamma]\|^2 \mid \|X_R\| = 0\right] \mathrm{d}\gamma, \tag{40}$$

where $Y_\gamma = \sqrt{\gamma} X_R + Z$.

Proof: The proof of (39) follows by a symmetry argument used in Proposition 3 and the I-MMSE relationship [13]

$$I(X; Y) = \frac{1}{2} \int_0^1 \mathbb{E}\left[\|X - \mathbb{E}[X \mid Y_\gamma]\|^2\right] \mathrm{d}\gamma. \tag{41}$$

To show (40) we use the point-wise I-MMSE formula [14]

$$\log \frac{f_{Y|X}(Y|X)}{f_Y(Y)} - \frac{1}{2} \int_0^1 \|X - \mathbb{E}[X \mid Y_\gamma]\|^2 \mathrm{d}\gamma = \int_0^1 (X - \mathbb{E}[X \mid Y_\gamma]) \cdot \mathrm{d}W_\gamma, \text{ a.s.} \tag{42}$$

where the integral on the right hand side of (42) is the Itô integral with respect to W_γ. The proof of the representation of $i(0, P_{X_R})$ now goes as follows:

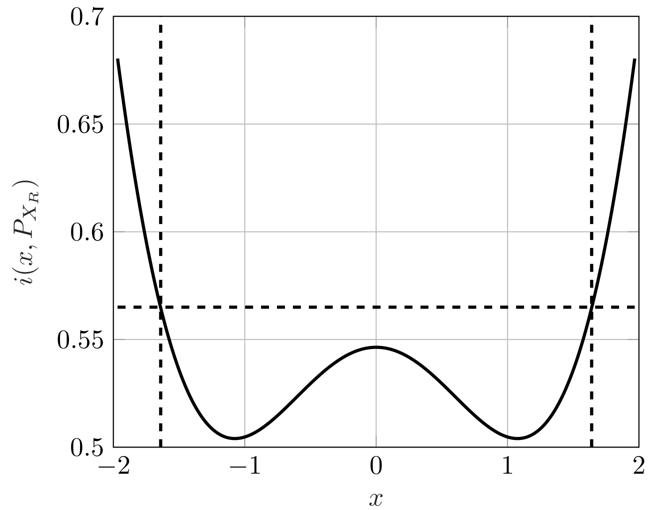
$$\begin{aligned}
2i(0, P_{X_R}) &= \mathbb{E}\left[\log \frac{f_{Y|X_R}(Y|X_R)}{f_Y(Y)} \mid \|X_R\| = 0\right] \\
&\overset{a)}{=} \mathbb{E}\left[\int_0^1 \|X_R - \mathbb{E}[X_R \mid Y_\gamma]\|^2 \mathrm{d}\gamma \mid \|X_R\| = 0\right] \\
&\quad - \mathbb{E}\left[\int_0^1 (X_R - \mathbb{E}[X_R \mid Y_\gamma]) \cdot \mathrm{d}W_\gamma \mid X_R = 0\right] \\
&\overset{b)}{=} \mathbb{E}\left[\int_0^1 \|X_R - \mathbb{E}[X_R \mid Y_\gamma]\|^2 \mathrm{d}\gamma \mid X_R = 0\right] \\
&= \int_0^1 \mathbb{E}\left[\|X_R - \mathbb{E}[X_R \mid Y_\gamma]\|^2 \mid X_R = 0\right] \mathrm{d}\gamma,
\end{aligned}$$

where the labeled equalities follow from: a) using the point-wise formula in (42); and b) using the symmetry of X_R to conclude that $\mathbb{E}[\mathbb{E}[X_R|Y_\gamma]|X_R = 0] = 0$. This concludes the proof. ■

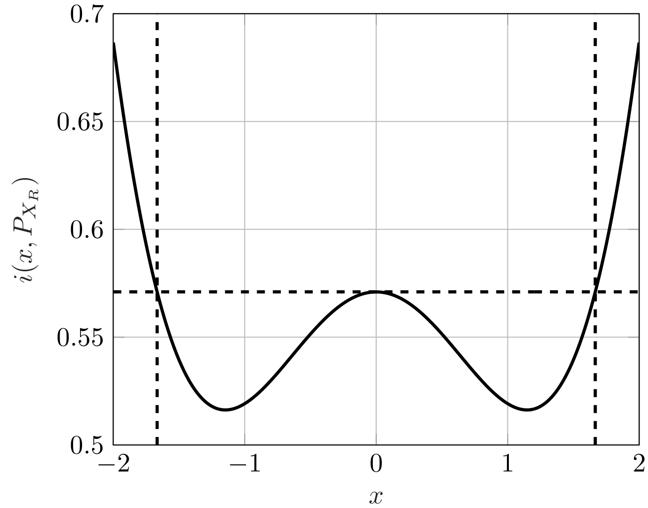
A. Proof of Theorem 2

Combining Lemma 3 and the optimality condition in Theorem 4 we arrive at

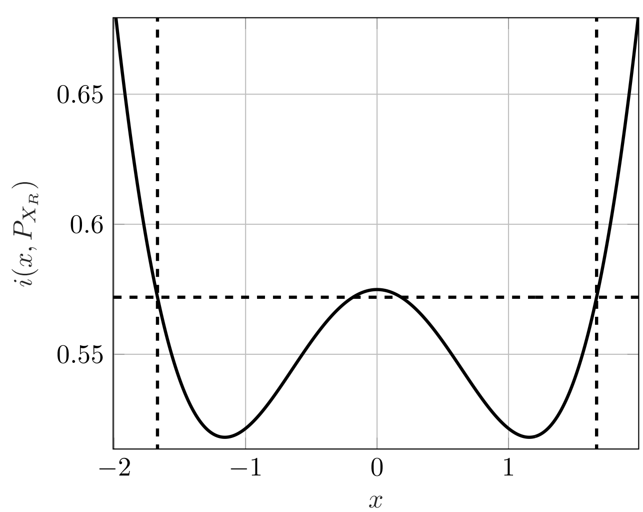
$$\begin{aligned}
0 &\geq i(0, P_{X_R}) - i(x, P_{X_R}) \\
&= \int_0^1 \mathbb{E}\left[\|X_R - \mathbb{E}[X_R \mid Y_\gamma]\|^2 \mid \|X_R\| = 0\right] \\
&\quad - \mathbb{E}\left[\|X_R - \mathbb{E}[X_R \mid Y_\gamma]\|^2 \mid \|X_R\| = R\right] \mathrm{d}\gamma
\end{aligned}$$



(a) Plot of $i(x, P_{X_R})$ vs. x for $R = 1.64$.



(b) Plot of $i(x, P_{X_R})$ vs. x for $R = \bar{R}_1 = 1.665925641$.



(c) Plot of $i(x, P_{X_R})$ vs. x for $R = 1.67$.

Fig. 6: Plot of $i(x, P_{X_R})$ vs. x for $n = 1$. Solid lines are $i(x, P_{X_R})$ and vertical dashed lines are $x = R$ and $x = -R$.

$$= \int_0^1 R^2 \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma}R\|Z\|\right)\right] - R^2 - R^2\mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma}R\|\sqrt{\gamma}x + Z\|\right)\right]\mathrm{d}\gamma, \quad (43)$$

where in the last step we have used the expression for the conditional expectation in (23) and the expression for the MMSE in (30). Now the condition in (43) is equivalent to

$$\int_0^1 \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma}R\|Z\|\right)\right] + \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma}R\|\sqrt{\gamma}x + Z\|\right)\right]\mathrm{d}\gamma \le 1, \quad (44)$$

where $\|x\| = R$. The value of $\bar{R}_n$ would now be a solution of (44) which concludes the proof of (14a).

To show the second part of Theorem 2 let $R = c\sqrt{n}$. We will also need the following bounds on $\mathsf{h}_v(x)$ [18], [19]:

$$\mathsf{h}_v(x) \ge \frac{x}{v + \sqrt{v^2 + x^2}}, \text{ for } v > 0, \quad (45)$$

$$\mathsf{h}_v(x) \le \frac{x}{\frac{2v-1}{2} + \sqrt{\frac{(2v-1)^2}{4} + x^2}}, \text{ for } v > \frac{1}{2}. \quad (46)$$

Moreover, if we let $\bar{x} = [c\sqrt{n}, 0, 0, ...]$ and define

$$V_n := \frac{1}{\sqrt{n}}\|cZ + c\sqrt{\gamma}\bar{x}\|, \quad (47)$$

$$W_n := \frac{1}{\sqrt{n}}\|cZ\|, \quad (48)$$

then the two terms on the left hand side of (44) can be lower and upper bounded as follows:

$$\mathbb{E}\left[\left(\frac{\sqrt{\gamma}W_n}{\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma W_n^2}}\right)^2\right] \le \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(c\sqrt{n}\sqrt{\gamma}\|Z\|\right)\right] \le \mathbb{E}\left[\left(\frac{\sqrt{\gamma}W_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma W_n^2}}\right)^2\right], \quad (49)$$

and

$$\mathbb{E}\left[\left(\frac{\sqrt{\gamma}V_n}{\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma V_n^2}}\right)^2\right] \le \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(c\sqrt{n}\sqrt{\gamma}\|\sqrt{\gamma}\bar{x} + Z\|\right)\right] \le \mathbb{E}\left[\left(\frac{\sqrt{\gamma}V_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma V_n^2}}\right)^2\right], \quad (50)$$

where the lower bounds hold for $n \ge 1$ and the upper bounds hold for $n > 1$.

In view of the fact that $u \mapsto \frac{u}{(a+\sqrt{a^2+u})^2}$ is a concave function for $a > 0$, using Jensen's inequality, we can further upper bound the expressions in (49) and (50) as follows:

$$\mathbb{E}\left[\left(\frac{\sqrt{\gamma}W_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma W_n^2}}\right)^2\right] \le \frac{\gamma\mathbb{E}[W_n^2]}{\left(\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma\mathbb{E}[W_n^2]}\right)^2} = \frac{\gamma c^2}{\left(\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma c^2}\right)^2}, \quad (51)$$

and

$$\mathbb{E}\left[\left(\frac{\sqrt{\gamma}V_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma^2 V_n^2}}\right)^2\right] \le \frac{\gamma\mathbb{E}[V_n^2]}{\left(\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma\mathbb{E}[V_n^2]}\right)^2} = \frac{\gamma c^2(1+\gamma c^2)}{\left(\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma c^2(1+\gamma c^2)}\right)^2}, \quad (52)$$

where we have also used that $\mathbb{E}[W_n^2] = c^2$ and $\mathbb{E}[V_n^2] = c^2(1+\gamma c^2)$. Applying the bounds in (51) and (52) to a necessary and sufficient condition in (44) we arrive at the following sufficient condition for optimality:

$$\int_0^1 \frac{\gamma c^2}{\left(\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma c^2}\right)^2} + \frac{\gamma c^2(1+\gamma c^2)}{\left(\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma c^2(1+\gamma c^2)}\right)^2}\mathrm{d}\gamma \le 1. \quad (53)$$

Next, we verify that it is sufficient to take $R \le \sqrt{n}$ which is equivalent to verifying that the inequality in (53) holds for $c = 1$. Choosing $c = 1$ in (53) and letting $a = \frac{n-1}{2n}$ we arrive at the following inequality:

$$2a\log(2a+1) - 2a\log\left(2\sqrt{a^2+2}+3\right) - 4a^2\tanh^{-1}\left(8a^2-1\right) + 4a^2\tanh^{-1}\left(\frac{4a\sqrt{a^2+2}-1}{3}\right) \le 1. \quad (54)$$

The solution to the above inequality can be found numerically and is given by $0.2358 \le a = \frac{n-1}{2n}$ or $n \ge 1.892$. Therefore, for $n \ge 2$ it is sufficient to take $c = 1$ or $R \le \sqrt{n}$.

Next, we find the exact limiting behavior of c. Observe that the lower and upper bounds in (49) and (50) are equal as $n \to \infty$ and, therefore, we focus only on the upper bounds in (49) and (50). By the strong law of large numbers almost surely we have the following limits:

$$\lim_{n\to\infty} V_n^2 = \lim_{n\to\infty} \frac{1}{n}\|cZ + c\sqrt{\gamma}\bar{x}\|^2$$
$$= \lim_{n\to\infty} \frac{1}{n}(cZ_1 + c^2\sqrt{\gamma}\sqrt{n})^2 + \lim_{n\to\infty} \frac{1}{n}\sum_{i=2}^{n}(cZ_i)^2$$
$$= c^2(1+\gamma c^2), \tag{55}$$

and similarly

$$\lim_{n\to\infty} W_n^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(cZ_i)^2 = c^2. \tag{56}$$

Now to show that the limit and the expectation can be interchanged observe that

$$\left| \frac{\sqrt{\gamma}V_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma V_n^2}} \right| \le 1, \tag{57}$$

$$\left| \frac{\sqrt{\gamma}W_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma W_n^2}} \right| \le 1, \tag{58}$$

and by the Dominated Convergence Theorem the limits are give by

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{c\sqrt{n}\sqrt{\gamma}\|Z+\sqrt{\gamma}\bar{x}\|}{\frac{n-1}{2} + \sqrt{\frac{(n-1)^2}{4} + c^2 n\gamma\|Z+\sqrt{\gamma}\bar{x}\|^2}}\right)^2\right]$$
$$= \lim_{n\to\infty} \mathbb{E}\left[\left(\frac{\sqrt{\gamma}V_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma V_n^2}}\right)^2\right]$$
$$= \left(\frac{\sqrt{\gamma}c\sqrt{1+\gamma c^2}}{\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2(1+\gamma c^2)}}\right)^2, \tag{59}$$

and

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{c\sqrt{n}\sqrt{\gamma}\|Z\|}{\frac{n-1}{2} + \sqrt{\frac{(n-1)^2}{4} + c^2 n\sqrt{\gamma}\|Z\|^2}}\right)^2\right]$$
$$= \lim_{n\to\infty} \mathbb{E}\left[\left(\frac{\sqrt{\gamma}W_n}{\frac{n-1}{2n} + \sqrt{\frac{(n-1)^2}{4n^2} + \gamma W_n^2}}\right)^2\right]$$
$$= \left(\frac{\sqrt{\gamma}c}{\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2}}\right)^2. \tag{60}$$

Therefore, the condition for optimality is given by

$$\int_0^1 \frac{\gamma c^2}{\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2}\right)^2} + \frac{\gamma c^2(1+\gamma c^2)}{\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2(1+\gamma c^2)}\right)^2} \mathrm{d}\gamma = 1. \tag{61}$$

The integral in (61) does have a closed form expression given in (62), however, the resulting equation must be solved numerically. Using numerical methods it is not difficult to verify that the solution to the equation in (61) is given by $c \approx 1.860935682$. This concludes the proof.

B. Proof of Theorem 3

First we compute the difference $i(x, P_X) - i(0, P_X)$ where we take $x = [x_1, 0, ..., 0]$

$$i(x, P_X) - i(0, P_X)$$
$$= \int_{\mathbb{R}^n} \left(\frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\|y-x\|^2}{2}} - \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\|y\|^2}{2}} \right) \log \frac{1}{f_Y(y)} \mathrm{d}y. \tag{63}$$

Next considering only the integral with respect to y_1, we have

$$-\int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(y_1-x_1)^2}{2}} - \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{y_1^2}{2}} \right) \log f_Y(y) \mathrm{d}y_1 \tag{64}$$
$$\overset{a)}{=} \int_{\mathbb{R}} \left(Q(y_1) - Q(y_1 - x_1)\right) \frac{\frac{d}{dy_1} f_Y(y)}{f_Y(y)} \mathrm{d}y_1$$
$$\overset{b)}{=} \int_{\mathbb{R}} \left(Q(y_1) - Q(y_1 - x_1)\right) \left(\mathbb{E}[X_1 \mid Y = y] - y_1\right) \mathrm{d}y_1$$
$$\overset{c)}{=} \int_{\mathbb{R}} \left(Q(y_1) - Q(y_1 - x_1)\right) \mathbb{E}[X_1 \mid Y = y] \mathrm{d}y_1 + \frac{x_1^2}{2}, \tag{65}$$

where the labeled equalities follow from: a) using integration by parts; b) using the identity in (28); and c) using the integral

$$\int_{\mathbb{R}} y\left(Q(y_1) - Q(y_1 - x_1)\right) \mathrm{d}y = -\frac{x_1^2}{2}. \tag{66}$$

Next, observing that $\frac{Q(y_1)-Q(y_1-x_1)}{-x_1}$ is a pdf since

$$\int_{\mathbb{R}} \frac{Q(y_1) - Q(y_1 - x_1)}{-x_1} dy_1 = 1, \tag{67}$$

and putting (63) and (65) together we have that

$$i(x, P_X) - i(0, P_X) = -x_1 \mathbb{E}[\mathbb{E}[X_1 \mid Y = W]] + \frac{x_1^2}{2}, \tag{68}$$

where W is a random vector such that $W_1 \sim \frac{Q(y_1)-Q(y_1-x_1)}{-x_1}$ and $W_i \sim \mathcal{N}(0,1)$ for $2 \le i \le n$.

Combining (65) with the conditional expectation of X_R in (23) and choosing $x_1 = R$ the difference in (63) is given by

$$i(R, P_{X_R}) - i(0, P_{X_R}) = -R^2 \mathbb{E}\left[\frac{W_1}{\|W\|} \mathsf{h}_{\frac{n}{2}}\left(R\|W\|\right)\right] + \frac{R^2}{2}. \tag{69}$$

The proof is concluded by applying (69) to the sufficient and necessary condition in (33).

VI. An Alternative Proof of the Lower Bound $\sqrt{n} \le \bar{R}_n$

In this section we give an alternative proof of the lower bound $\sqrt{n} \le \bar{R}_n$. The main idea is to show that $x \mapsto i(x, P_X)$ is a subharmonic function where the notion of subharmonic functions is defined next.

Definition 2. *(Subharmonic Function)* Suppose that the function f is twice continuously differentiable on an open set

$$\frac{\log\left(\sqrt{4c^2+1}+1\right)-\log\left(c^2+1\right)-\log\left(2\right)-\sqrt{4c^2+1}+2\,c^2+1}{c^2}=1 \tag{62}$$

$G \in \mathbb{R}^n$. Then f is *subharmonic* if $\nabla^2 f \geq 0$ on G where ∇^2 is the Laplacian[1].

We will use an important property that a subharmonic function always attains its maximum on the boundary of a set as shown in the following theorem.

Theorem 5. (Maximum Principle of Subharmonic Functions) *Suppose that G is a connected open set. If f is subharmonic and attains a global maximum value in the interior of G, then f is constant in G.*

To show the desired bound we will use Theorem 5 together with yet another identity that relates estimation and information measures [20, Property 3].

Lemma 4. *Denote the likelihood function by*

$$\ell(y) := \frac{f_Y(y)}{\frac{1}{(2\pi)^{\frac{n}{2}}}\mathrm{e}^{-\frac{\|y\|^2}{2}}}. \tag{70}$$

Then,

$$\begin{aligned}\nabla^2 \log \ell(y) &= \mathbb{E}[\|X-\mathbb{E}[X \mid Y]\|^2 \mid Y=y] \\ &:= \mathrm{Var}(X \mid Y=y). \end{aligned} \tag{71}$$

The next result shows that the function $x \mapsto i(x, P_X)$ is subharmonic if X is contained in a small enough neighborhood.

Theorem 6. *Suppose that $X \in \mathcal{B}_0(R)$. Then, for $R \leq \sqrt{n}$ and all $x \in \mathcal{B}_0(R)$ the function $x \mapsto i(x, P_X)$ is subharmonic.*

Proof: Observe that $i(x, P_X)$ can be written in terms of the log-likelihood function as follows:

$$\begin{aligned} & i(x,P_X) \\ &= -\mathbb{E}\left[\log f_Y(x+Z)\right] - h(Z) && (72) \\ &= -\mathbb{E}\left[\log \ell(x+Z)\right] - \mathbb{E}\left[\log \frac{1}{(2\pi)^{\frac{n}{2}}}\mathrm{e}^{-\frac{\|x+Z\|^2}{2}}\right] - h(Z) && (73) \\ &= -\mathbb{E}\left[\log \ell(x+Z)\right] + \frac{\|x\|^2}{2}. && (74) \end{aligned}$$

Therefore, using the fact that $X \in \mathcal{B}_0(R)$ we have that

$$\begin{aligned} \nabla^2 i(x,P_X) &= -\mathbb{E}\left[\mathrm{Var}(X \mid Y) \mid X=x\right] + n && (75) \\ &\geq -R^2 + n, && (76) \end{aligned}$$

where in the last step we have used the bound $\mathrm{Var}(X \mid Y) \leq R^2$. This concludes the proof. ■

As a consequence of Theorem 5 and Theorem 6 we have the following corollary.

[1] If f is twice differentiable its Laplacian is given by by $\nabla^2 f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$.

Corollary 1. *For $R \leq \sqrt{n}$*

$$\max_{X \in \mathcal{B}_0(R)} I(X; X+Z) = I(X_R; X_R+Z), \tag{77}$$

or, alternatively, $\sqrt{n} \leq \bar{R}_n$.

The proof of Theorem 6 gives yet another example of the utility of identities between estimation and information measures.

VII. Discussion

In this work we have characterized conditions under which an input distribution uniformly distributed over a single sphere achieves the capacity of a vector Gaussian noise channel with a constraint that the input must lie in the n-ball of radius R. We have also shown that the largest radius $\bar{R}_n$ for which it is still optimal to use a single sphere grows as $\sqrt{n}$. Moreover, the exact limit of $\frac{\bar{R}_n}{\sqrt{n}}$ as $n \to \infty$ is found to be ≈ 1.861.

A number of methods that we have used throughout the paper relied on using estimation theoretic representations of information measures such as the I-MMSE relationship. The path via estimation theoretic arguments allows us to contrast optimization of the mutual information with that of a similar problem of optimizing the MMSE, that is

$$\max_{X \in \mathcal{B}_0(R)} \mathrm{mmse}(X|Y). \tag{78}$$

Distributions that maximize (78) are referred to as *least favorable prior distributions* and have been shown to have a spherical structure similar to that of the distributions that maximize the mutual information; an interested reader is referred to [21] and [22] and references therein. Moreover, the conditions for the optimality of a single sphere distribution (i.e., the maximum radius $\bar{R}_n^{\mathrm{MMSE}}$) in (78) have been found in [21] and [23] and are given by a solution to the following equation:

$$\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(R\|Z\|\right)\right] + \mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(R\|x+Z\|\right)\right] = 1, \tag{79}$$

where $\|x\| = R$. It is pleasing to see the similarity between the optimality condition for the MMSE in (79) and the optimality condition for the mutual information in (14a). Note, however, that (14a) could not have been derived directly from (79) or vice versa. The values of $\bar{R}_n^{\mathrm{MMSE}}$ are shown in Table I. The code for the numerical computation of $\bar{R}_n^{\mathrm{MMSE}}$ and $\bar{R}_n$ can be found at [24].

It is also interesting to point out that that $\bar{R}_n^{\mathrm{MMSE}}$ is always lagging behind $\bar{R}_n$ as we increase n as shown in Fig. 7.

TABLE I: Values of $\bar{R}_n$ and $\bar{R}_n^{\text{MMSE}}$.

Dimension n	$\bar{R}_n$	$\bar{R}_n^{\text{MMSE}}$
1	1.666	1.057
2	2.454	1.535
3	3.065	1.908
4	3.580	2.223
5	4.031	2.501
6	4.438	2.751
7	4.811	2.981
8	5.158	3.195
9	5.483	3.395
10	5.789	3.585
11	6.080	3.765
12	6.359	3.936
13	6.625	4.101
14	6.881	4.259
15	7.128	4.412
16	7.367	4.560
17	7.598	4.702
18	7.823	4.841
19	8.041	4.976
20	8.254	5.107
21	8.461	5.235
22	8.663	5.360
23	8.860	5.483
24	9.054	5.602
25	9.243	5.719
26	9.428	5.834
27	9.610	5.946
28	9.789	6.056
29	9.964	6.165
30	10.136	6.271
31	10.306	6.376
32	10.472	6.479
33	10.636	6.580
34	10.798	6.680
35	10.957	6.779

Notably this behavior persists even as $n \to \infty$ since for the MMSE[2]

$$\lim_{n\to\infty} \frac{\bar{R}_n^{\text{MMSE}}}{\sqrt{n}} = c_{\text{MMSE}} \approx 1.151, \tag{80}$$

where c_{MMSE} is the solution of the following equation:

$$\frac{c^2}{\left(\frac{1}{2}+\sqrt{\frac{1}{4}+c^2}\right)^2} + \frac{c^2(1+c^2)}{\left(\frac{1}{2}+\sqrt{\frac{1}{4}+c^2(1+c^2)}\right)^2} = 1, \tag{81}$$

while for the mutual information according to Theorem 2

$$\lim_{n\to\infty} \frac{\bar{R}_n}{\sqrt{n}} \approx 1.861. \tag{82}$$

[2] The exact limit is given by $\lim_{n\to\infty} \frac{\bar{R}_n^{\text{MMSE}}}{\sqrt{n}} = c_{\text{MMSE}} = \frac{\sqrt{\sqrt[3]{9-\sqrt{69}}+\sqrt[3]{9+\sqrt{69}}}}{\sqrt[6]{2}\,\sqrt[3]{3}} \approx 1.15096$; see [21] and [25] for the details.

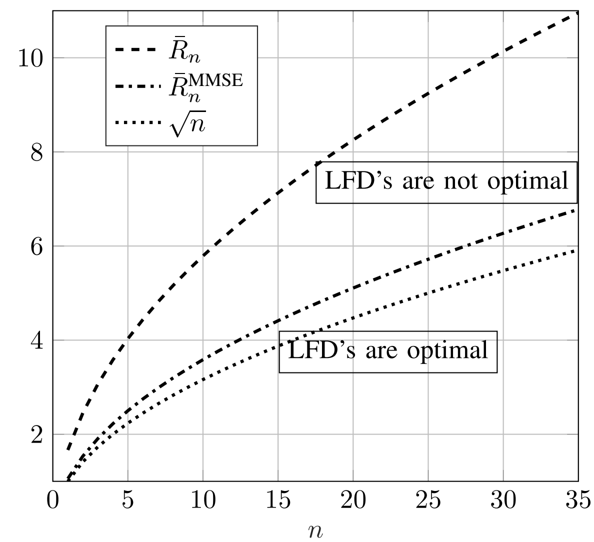


Fig. 7: Comparison of $\bar{R}_n$, $\bar{R}_n^{\text{MMSE}}$ and $\sqrt{n}$. For $R \le \bar{R}_n^{\text{MMSE}}$ the least favorable distributions (LPF's) are capacity achieving (optimal for short) and not capacity achieving for $R > \bar{R}_n^{\text{MMSE}}$.

Again, note the similarity between (81) and (14c).

The lagging of $\bar{R}_n^{\text{MMSE}}$ behind $\bar{R}_n$ also points out that the following bounding technique, which relies on the I-MMSE, results in a tight bound if $R \le \bar{R}_n^{\text{MMSE}}$ and is not tight if $\bar{R}_n^{\text{MMSE}} \le R \le \bar{R}_n$:

$$\max_{X\in\mathcal{B}_0(R)} I(X;Y) = \max_{X\in\mathcal{B}_0(R)} \frac{1}{2}\int_0^1 \text{mmse}(X|Y_\gamma)\text{d}\gamma \tag{83}$$

$$\le \frac{1}{2}\int_0^1 \max_{X\in\mathcal{B}_0(R)} \text{mmse}(X|Y_\gamma)\text{d}\gamma, \tag{84}$$

Such a condition for tightness of the bound via the I-MMSE relation was already pointed out in [26] for $n = 1$. Interestingly for several multiuser problems [27]–[29], with a second moment constraint on the input instead of an amplitude constraint, such lagging vanishes as $n \to \infty$ and bounds via the I-MMSE of the type in (84) (i.e., where the maximum and the integral are interchanged) are tight. The fundamental difference is that in the aforementioned works the Gaussian distribution is optimal in the limit of n, while in (2) and (78) Gaussian inputs are not optimal even as $n \to \infty$.

The optimality of an input distribution on a single sphere also has important practical implications as it suggests that phase modulation is optimal. Note that in practice, however, the continuous sphere would have to be discretized. The accuracy of such a discretization can potentially be evaluated by using the fact that the mutual information is continuous in the Wasserstein metric over a set of distributions with compact support [30].

An ambitious future direction is to consider an extension of the result in this paper to a general MIMO channel. For a recent survey on discrete inputs in MIMO systems the interested reader is referred to [31, Corollary 4].

Another interesting direction is to consider an input average power constraint (i.e., $\mathbb{E}[\|X\|^2] \le P$) together with the input amplitude constraint analyzed in this paper.

Appendix A
On the Numerical Evaluation of the Integrals in (15) and (79)

Evaluation of the expectations in (15) and (79) for any given R may be performed using Monte-Carlo methods. The ratios of Bessel functions in the expectations can be evaluated precisely thanks to the known continued fraction expansion [32]

$$\mathsf{h}_n(x) = \frac{\mathsf{I}_n(x)}{\mathsf{I}_{n-1}(x)} = \frac{1}{b_1 + \frac{1}{b_2 + \cdots}}, \tag{85}$$

where

$$b_k = \frac{2(n+k-1)}{x}. \tag{86}$$

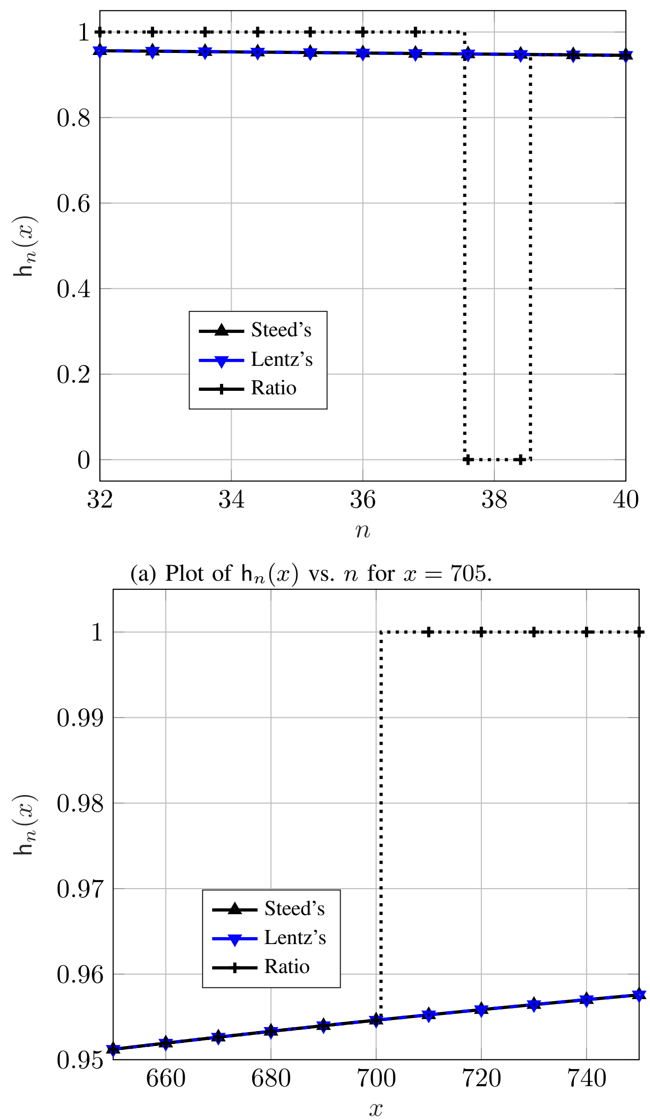
The continued fraction can be evaluated via Steed's or Lentz's algorithm [33], either of which gives stable and accurate results, whereas the direct evaluation of the ratio of Bessel functions may lead to floating-point overflows at high values of x. An example of the overflow for double precision values is shown on Fig. 8. Note that in Fig. 8a the zero values of the function $\mathsf{h}_n(x) = \frac{\mathsf{I}_n(x)}{\mathsf{I}_{n-1}(x)}$ around $n = 38$ correspond to denominator overflows while the one values for smaller values of n correspond to both numerator and denominator overflows. Moreover, observe that neither Steed's or Lentz's algorithm experiences this issue.

Since $h_n(x)$ is a monotonically increasing function of x, the integrands in (15) and (79) are monotonically increasing functions of R. Hence, given lower and upper bounds on R, the zeroes of (15) and (79) may be obtained via binary searches. In our simulations, we set the lower and upper bounds to $\sqrt{n}$ and $3\sqrt{n}$, respectively. Note that this interval includes both the set $[1.6659\sqrt{n}, 1.8609\sqrt{n}]$ for the maximum mutual information setting, and the set $[1.0567\sqrt{n}, 1.151\sqrt{n}]$ for the maximum MMSE setting.

We sampled 10^8 chi-square samples while evaluating the expectations for the binary searches. During the evaluation of (15), we integrated directly over the distribution of W_1, and we distributed the chi-square samples uniformly across the effective domain of W_1, which was set as $[-7, R+7]$ to capture all but a negligible amount of the probability mass. During the evaluation of (79), we sampled evenly from the central and non-central chi-square distributions[3].

The $\bar{R}_n$ and $\bar{R}_n^{\text{MMSE}}$ values reported in Table I have consistently led to residuals well below 10^{-4} during the Monte-Carlo evaluations of the integral equations. In our experience, multiple binary searches in this setting do not lead to changes in $\bar{R}_n$ and $\bar{R}_n^{\text{MMSE}}$ values beyond the fourth digit after the decimal point. Interested readers may refer to our implementation and simulation data found at [24].

[3]Alternatively, more samples can be taken from the non-central chi-square due to its higher variance.



(a) Plot of $\mathsf{h}_n(x)$ vs. n for $x = 705$.

(b) Plot of $\mathsf{h}_n(x)$ vs. x for $n = 33$.

Fig. 8: Comparison of values of $\mathsf{h}_n(x)$ obtained via Steed's algorithm, Lentz's algorithm and direct evaluation of the ratio of Bessel functions.

Appendix B
Converting to Spherical Coordinates in (36)

Using that

$$\rho(y) = \left(-\|y\| + \frac{R\mathsf{I}_{\frac{n}{2}}(\|y\|R)}{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}\right)\frac{y_1}{\|y\|} = M(\|y\|)\frac{y_1}{\|y\|},$$

we have

$$-g'(x_1) = \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{\sum_{i=2}^n y_i^2 + (y_1 - x_1)^2}{2}} M(\|y\|)\frac{y_1}{\|y\|}\mathrm{d}y.$$

Transforming $y_1, y_2, \ldots, y_n$ to the spherical coordinates $r, \phi_1, \ldots, \phi_{n-1}$ where $r \ge 0$, $0 < \phi_1 \le 2\pi$ and $0 < \phi_i \le \pi$

for $i=2,\ldots,n-1$ we have that

$$y_1 = r\cos(\phi_1), \tag{87a}$$

$$y_i = r\cos(\phi_i)\prod_{k=1}^{i-1}\sin\phi_k,\ i=2,\ldots,n-1, \tag{87b}$$

$$y_n = r\prod_{k=1}^{n-1}\sin\phi_k, \tag{87c}$$

and the Jacobian is given by

$$\mathrm{d}y = r^{n-1}\prod_{k=1}^{n-1}(\sin\phi_k)^{n-1-k}\,\mathrm{d}r\mathrm{d}\phi_1\ldots\mathrm{d}\phi_{n-1}. \tag{87d}$$

Therefore, the derivative can be written as follows:

$$\begin{aligned}
&-g'(x_1)\\
&\overset{a)}{=}\prod_{k=2}^{n-1}\int_0^{\pi}(\sin\phi_k)^{n-1-k}\,\mathrm{d}\phi_k\int_0^{2\pi}\int_0^{\infty}\frac{\mathrm{e}^{-\frac{r^2-2rx_1\cos(\phi_1)+x_1^2}{2}}}{(2\pi)^{\frac{n}{2}}}\\
&\cdot M(r)\cos(\phi_1)r^{n-1}(\sin\phi_1)^{n-1}\,\mathrm{d}r\mathrm{d}\phi_1\\
&\overset{b)}{=}S_{n-2}\int_0^{\infty}\frac{\mathrm{e}^{-\frac{r^2+x_1^2}{2}}}{(2\pi)^{\frac{n}{2}}}\frac{M(r)2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}{(x_1r)^{\frac{n-2}{2}}}\mathsf{I}_{\frac{n}{2}}(x_1r)r^{n-1}\mathrm{d}r\\
&\overset{c)}{=}x\int_0^{\infty}\mathrm{e}^{-\frac{r^2+x_1^2}{2}}M(r)\left(\frac{r}{x}\right)^{\frac{n}{2}}\mathsf{I}_{\frac{n}{2}}(x_1r)\mathrm{d}r\\
&=2x\int_0^{r}M(r)\mathrm{e}^{-\frac{r^2+x_1^2}{2}}\frac{1}{2}\left(\frac{r}{x}\right)^{\frac{n}{2}}\mathsf{I}_{\frac{n}{2}}(x_1r)\mathrm{d}r\\
&\overset{d)}{=}2x\mathbb{E}\left[M\left(\sqrt{V^2}\right)\right],
\end{aligned}$$

where the labeled equalities follow from: a) using spherical coordinates in (87); b) using that

$$\prod_{k=2}^{n-1}\int_0^{\pi}(\sin\phi_k)^{n-1-k}\,\mathrm{d}\phi_k = S_{n-1},$$

and that

$$\begin{aligned}
&\int_0^{2\pi}\mathrm{e}^{rx_1\cos(\phi_1)}\cos(\phi_1)(\sin\phi_1)^{n-1}\,\mathrm{d}\phi_1\\
&=\frac{2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}{(x_1r)^{\frac{n-2}{2}}}\mathsf{I}_{\frac{n}{2}}(x_1r)r^{n-1};
\end{aligned}$$

c) using that $S_{n-2}2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right) = (2\pi)^{\frac{n}{2}}$; and observing that $\frac{1}{2}\mathrm{e}^{-\frac{r^2+x_1^2}{2}}\left(\frac{r}{x}\right)^{\frac{n}{2}}\mathsf{I}_{\frac{n}{2}}(x_1r)$ is the pdf of a chi-square random variable of degree $n+2$ with non-centrality x_1^2.

This concludes the proof.

References

[1] S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proc. I (Communications, Speech and Vision)*, vol. 137, no. 6, pp. 424–430, 1990.

[2] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, no. 3, pp. 203–219, April 1971.

[3] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, July 1995.

[4] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

[5] B. Rassouli and B. Clerckx, "On the capacity of vector Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6884–6903, December 2016.

[6] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "When are discrete channel inputs optimal? - Optimization techniques and some new results," in *Proc. Conf. on Inf. Sci. and Sys.*, Princeton, NJ, USA, March 2018, pp. 1–6.

[7] N. Sharma and S. Shamai (Shitz), "Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels," *Probl. Inf. Transm.*, vol. 46, no. 4, pp. 283–299, 2010.

[8] A. Thangaraj, G. Kramer, and G. Böcherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4172–4182, July 2017.

[9] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor, "Upper and lower bounds on the capacity of amplitude-constrained MIMO channels," in *Proc. IEEE Global Commun. Conf.*, Singapore, December 2017, pp. 1–6.

[10] A. ElMoslimany and T. M. Duman, "On the capacity of multiple-antenna systems and parallel Gaussian channels with amplitude-limited inputs," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2888–2899, July 2016.

[11] Y. Polyanskiy and Y. Wu, "Peak-to-average power ratio of good codes for Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7655–7660, December 2014.

[12] S. Karlin, "Pólya type distributions, II," *Ann. of Math. Stat.*, vol. 28, no. 2, pp. 281–308, 1957.

[13] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[14] K. Venkat and T. Weissman, "Pointwise relations between information and estimation in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6264–6281, October 2012.

[15] G. S. Watson, *Statistics on Spheres*. Wiley-Interscience, 1983, vol. 6.

[16] R. Esposito, "On a relation between detection and estimation in decision theory," *Inf. Control*, vol. 12, no. 2, pp. 116–120, February 1968.

[17] G. N. Watson, *A Treatise on the Theory of Bessel Functions*. Cambridge University Press, 1995.

[18] J. Segura, "Bounds for ratios of modified Bessel functions and associated Turán-type inequalities," *J. Math. Anal. Appl.*, vol. 374, no. 2, pp. 516–528, 2011.

[19] Á. Baricz, "Bounds for Turánians of modified Bessel functions," *Expo. Math.*, vol. 33, no. 2, pp. 223–251, 2015.

[20] C. Hatsell and L. Nolte, "Some geometric properties of the likelihood ratio (corresp.)," *IEEE Trans. Inf. Theory*, vol. 17, no. 5, pp. 616–618, 1971.

[21] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai (Shitz), "On the structure of the least favorable prior distributions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, June 2018, to appear.

[22] E. Marchand and W. E. Strawderman, "Estimation in restricted parameter spaces: A review," in *A Festschrift for Herman Rubin*. Institute of Mathematical Statistics, 2004, pp. 21–44.

[23] J. C. Berry, "Minimax estimation of a bounded normal mean vector," *J. Multivariate Anal.*, vol. 35, no. 1, pp. 130–139, 1990.

[24] A. Dytso, M. Al, H. V. Poor, and S. Shamai (Shitz), "Release of peak-power capacity," August 2018. [Online]. Available: https://doi.org/10.5281/zenodo.1401556

[25] É. Marchand and F. Perron, "On the minimax estimator of a bounded normal mean," *Statist. Probab. Lett.*, vol. 58, no. 4, pp. 327–333, 2002.

[26] M. Raginsky, "On the information capacity of Gaussian channels under small peak power constraints," in *Proc. 46th Annual Allerton Conf. Commun., Control and Comp.*, Monticello, IL, USA, 2008, pp. 286–293.

[27] R. Bustin and S. Shamai (Shitz), "MMSE of 'bad' codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 733–743, February 2013.

[28] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, H. V. Poor, and S. Shamai (Shitz), "On communication through a Gaussian channel with an MMSE disturbance constraint," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 513–530, January 2018.

[29] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai (Shitz), "A view of information-estimation relations in Gaussian networks," *Entropy*, vol. 19, no. 8, p. 409, August 2017.

[30] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, March 2012.

[31] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, "A survey on MIMO transmission with finite input signals: Technical challenges, advances, and future trends," *Proc. IEEE*, no. 99, pp. 1–55, 2018.

[32] A. A. Cuyt, V. Petersen, B. Verdonk, H. Waadeland, and W. B. Jones, *Handbook of Continued Fractions for Special Functions*. Springer Science & Business Media, 2008.

[33] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge University Press, 2007.