Capacity of the Vector Gaussian Channel in the Small Amplitude Regime

Alex Dytso*, 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*, 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 where the input-output relationships are give by

$$Y = X + Z, \quad (1)$$

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 a ball centered at 0 of radius R, that is

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

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 $\sqrt{n}$.

a) Prior Work: For the case of $n = 1$ Smith in his seminal work [1], using convex optimization techniques, has shown that the maximizing distribution in (2) must be discrete with finitely many points. In [2], 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 [3] and [4].

This paper can be considered as an n-dimensional extension of the work in [5] where, in the case of $n = 1$, a two point input distribution uniform on $\pm R$ has been shown to be optimal if and only if $R \leq 1.665$, and a three point input distribution on $\{-R, 0, R\}$ has been shown to be optimal if and only if $1.665 \leq R \leq 2.786$. However, unlike the approach in [5], the proof strategy used in this work relies on very different methods (rooted in estimation theory) and, for every dimension n, recovers an exact condition for the optimality of an input supported on a single sphere.

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

A number of works have also focused on deriving lower and upper bounds on (2). The authors in [6] derived an asymptotically tight upper bound on the capacity as $R \to \infty$ by using the dual representation of channel capacity. In [7] 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 [7] and [8] for asymptotically tight lower bounds on the capacity.

An interested reader is also referred to [9] 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) and gives the definition of the "small amplitude" regime as the regime in which a uniform probability distribution supported on a single sphere is optimal;
2) Section III, Theorem 2, presents our main result, which is an exact characterization of the size of the small amplitude regime;
3) Section IV, for an input distribution 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) in estimating X from Y;
4) 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 from [10], the I-MMSE relationship [11] and the point-wise I-MMSE relationship [12]; and
5) Section VI 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.

Due to space limitations, some of the proofs are omitted and can be found in an extended version of this paper [13].

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

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

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:

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

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:

$$I(P_X) \coloneqq I(X;Y).$$

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

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

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

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

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

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

II. Optimizing The Input Distribution

The optimal input distribution in (2) can be characterized by using the method presented in [1] and its extension to the complex channel (i.e., $n = 2$) given in [2]; see also [3] and [4] for a detailed solution for any $n \geq 2$.

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:*

$$i(x, P_X^\star) = I(P_X^\star),\ x \in \mathsf{supp}(P_X^\star), \tag{3a}$$

$$i(x, P_X^\star) \leq I(P_X^\star),\ x \in \mathcal{B}_0(R);\ \text{and} \tag{3b}$$

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

$$\mathsf{supp}(P_X^\star) = \bigcup_{i=1}^{N} \mathcal{C}(r_i), \tag{3c}$$

where $N < \infty$ (finite).

Note that for $n = 1$ the optimal inputs are discrete with finitely many points. For $n > 1$ the optimal input probability distributions are no-longer discrete but singular, however, the magnitude of the optimal input distribution $\|X\|$ is discrete with finitely many points.

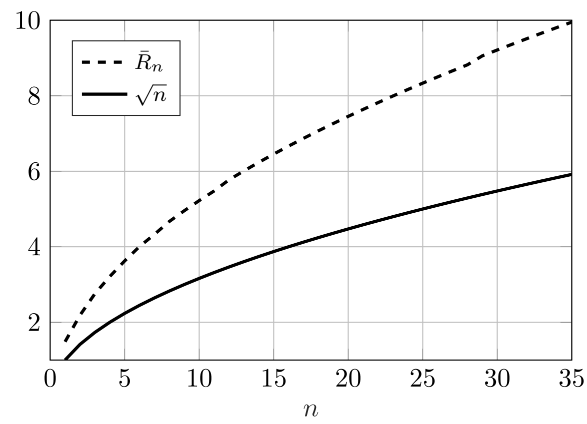

Fig. 1: Comparison of $\bar{R}_n$ and $\sqrt{n}$.

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

$$\bar{R}_n \coloneqq \max\{R : P_{X_R} = \arg\max \max_{X \in \mathcal{B}_0(R)} I(X;Y)\}. \tag{4}$$

In words, $\bar{R}_n$ is the largest 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 the maximum radius $\bar{R}_n$.

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:*

$$\int_0^1 \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma} R \|Z\|\right)\right] + \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(\sqrt{\gamma} R \|\sqrt{\gamma} x + Z\|\right)\right] \mathrm{d}\gamma = 1, \tag{5a}$$

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

$$\lim_{n \to \infty} \frac{\bar{R}_n}{\sqrt{n}} = c \approx 1.860935682, \tag{5b}$$

where c is the solution of the following equation:

$$\int_0^1 \frac{\gamma c^2}{\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2}\right)^2} + \frac{\gamma c^2(1 + \gamma c^2)}{\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \gamma c^2(1 + \gamma c^2)}\right)^2} \mathrm{d}\gamma = 1. \tag{5c}$$

Note that R_n is given as the solution of an integral equation in (5a) and (5c) and does not have an exact analytical form and must be found using numerical methods. The evaluation of $\bar{R}_n$ up to $n = 35$ is shown on Fig. 1.

Remark 1. Recall that $\|Z + x\|^2$ in (5a) 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 $\bar{R}_n$.

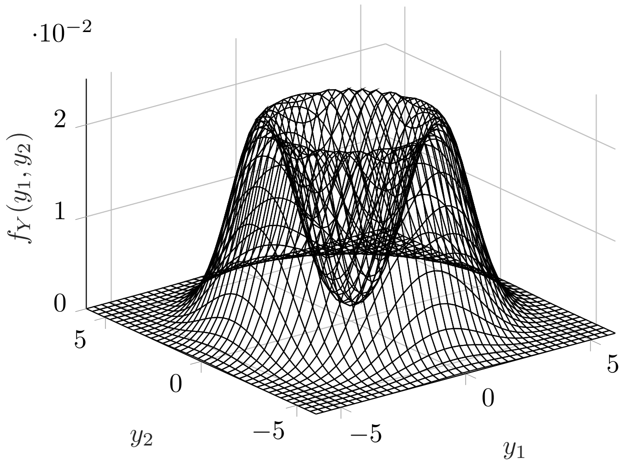

Fig. 2: The output pdf in (6) for $n = 2$ and $R = 3$.

IV. Some Analytic 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*

$$f_Y(y) = \frac{\Gamma\left(\frac{n}{2}\right) \mathrm{e}^{-\frac{R^2+\|y\|^2}{2}}}{2\pi^{\frac{n}{2}}} \frac{\mathsf{I}_{\frac{n}{2}-1}(\|y\|R)}{(\|y\|R)^{\frac{n}{2}-1}}. \quad (6)$$

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

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

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

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

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 (6) we can now also compute the conditional expectation $\mathbb{E}[X|Y]$.

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

$$\mathbb{E}[X_R|Y = y] = \frac{Ry}{\|y\|}\mathsf{h}_{\frac{n}{2}}(\|y\|R). \quad (7)$$

Proof: Using the identity between the conditional expectation and score function [14] we have that

$$\mathbb{E}[X_R|Y = y] = y + \frac{\nabla_y f_Y(y)}{f_Y(y)}, \quad (8)$$

and due to the symmetry of $f_Y(y)$ we have that $\nabla_y f_Y(y) = \frac{y}{\|y\|}\frac{d}{d\|y\|} f_Y(\|y\|)$ where

$$\frac{d}{d\|y\|} f_Y(\|y\|) = -\|y\| f_Y(y) + R f_Y(y) \mathsf{h}_{\frac{n}{2}}(\|y\|R). \quad (9)$$

The proof of (7) is completed by combining (8) and (9). ∎

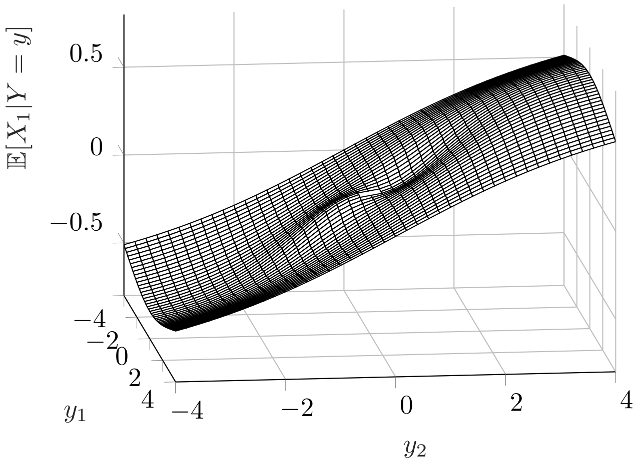

Fig. 3: The conditional expectation $\mathbb{E}\left[X_1|Y = y\right]$ where $X_R = [X_1, X_2]$ with $R = 2.5$ and where $y = [y_1, y_2]$.

Remark 2. The proof of Proposition 2 relies on the following identity between the conditional expectation and the output pdf [14]:

$$\mathbb{E}[X|Y = y] = y + \frac{\nabla_y f_Y(y)}{f_Y(y)}, \quad (10)$$

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 (10) considerably simplifies the computation of $\mathbb{E}[X|Y]$ as we do not need to derive the conditional distribution $P_{X|Y}$ and only use properties of the output pdf $f_Y(y)$.

An example of the shape of the conditional expectation for $n = 2$ is shown on Fig. 3.

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

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

$$I(X_R; Y) = R^2 \log(\mathrm{e}) + \log\left(\frac{2^{1-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\right) - \mathbb{E}\left[\log\left(\frac{\mathsf{I}_{\frac{n}{2}-1}(\|Z + x\|R)}{(\|Z + x\|R)^{\frac{n}{2}-1}}\right)\right], \quad (11)$$

and

$$\mathrm{mmse}(X_R|Y) = R^2 - R^2\mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}(R\|x + Z\|)\right], \quad (12)$$

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

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 lemmas would be useful in our analysis.

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

The next lemma was shown in [10, 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 $h(\omega)$ be a function that changes sign n times. If*

$$M(x) := \int h(\omega) f(x, \omega) \mathrm{d}\omega, \quad (13)$$

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

Theorem 3. (New Optimality Condition) *P_{X_R} is optimal if and only if for all $\|x\| = R$*

$$i(0, P_{X_R}) \le i(x, P_{X_R}). \tag{14}$$

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

$$g(\|x\|) \coloneqq i(x, P_{X_R}). \tag{15}$$

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 (3a) and (3b) to only one condition

$$g(0) \le g(R). \tag{16}$$

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 which after integrating by parts is given by

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

where

$$\begin{aligned} \rho(y) &\coloneqq \frac{d}{dy_1} \log f_Y(y) \\ &= \left(-\|y\| + R \mathsf{h}_{\frac{n}{2}}(\|y\|R)\right) \frac{y_1}{\|y\|} \coloneqq M(\|y\|) \frac{y_1}{\|y\|}. \end{aligned}$$

Next by transforming to spherical coordinates we have that

$$g'(x_1) = -2x_1 \int_0^\infty M(r) \mathrm{e}^{-\frac{r^2 + x_1^2}{2}} \frac{1}{2} \left(\frac{r}{x_1}\right)^{\frac{n}{2}} \mathsf{I}_{\frac{n}{2}}(x_1 r) \mathrm{d}r \tag{17}$$

$$= -2x_1 \mathbb{E}[M(V^2)], \tag{18}$$

where V^2 is the non-central chi-square distribution with $n+2$ degrees of freedom and non-centrality x_1^2.

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

Next observe that in (17) 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 3. Condition (14) significantly simplifies the necessary and sufficient conditions for optimality in (3). For instance, we do not have to verify the conditions in (3b) for all $x \in \mathcal{B}_0(R)$ and instead need only to check points satisfying $\|x\| = 0$ and $\|x\| = R$.

Moreover, the condition in (14) 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 $\|x\| = 0$.

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 $\|x\| = R$*

$$i(x, P_{X_R}) = \frac{1}{2} \int_0^1 \mathbb{E}\left[\|X_R - \mathbb{E}[X_R|Y_\gamma]\|^2 \mid \|X_R\| = R\right] \mathrm{d}\gamma, \tag{19}$$

$$i(0, P_{X_R}) = \frac{1}{2} \int_0^1 \mathbb{E}\left[\|X_R - \mathbb{E}[X_R|Y_\gamma]\|^2 \mid \|X_R\| = 0\right] \mathrm{d}\gamma, \tag{20}$$

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

Proof: The proof of (19) follows by a symmetry argument and the I-MMSE relationship [11]

$$I(X;Y) = \frac{1}{2} \int_0^1 \mathbb{E}[\|X - \mathbb{E}[X|Y_\gamma]\|^2] \mathrm{d}\gamma. \tag{21}$$

To show (20) we use the point-wise I-MMSE formula [12]

$$\begin{aligned} &\log \frac{f_{Y|X}(Y|X)}{f_Y(Y)} - \frac{1}{2} \int_0^1 \|X - \mathbb{E}[X|Y_\gamma]\|^2 \mathrm{d}\gamma \\ &= \int_0^1 (X - \mathbb{E}[X|Y_\gamma]) \cdot \mathrm{d}W_\gamma, \text{ a.s.} \end{aligned} \tag{22}$$

where the integral on the right hand side of (22) is the Itô integral with respect to W_γ. The proof of the representation of $i(0, P_{X_R})$ now goes as follows:

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

where the labeled equalities follow from: a) using the point-wise formula in (22); 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. ∎

VI. 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 ≈ 1.86.

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

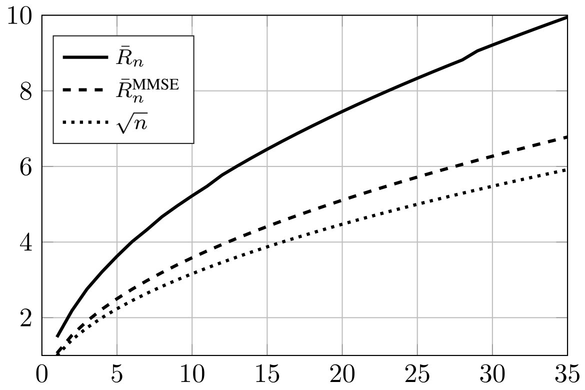

Fig. 4: Comparison of $\bar{R}_n$, $\bar{R}_n^{\text{MMSE}}$ and $\sqrt{n}$.

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

$$\max_{X \in \mathcal{B}_0(R)} \text{mmse}(X|Y). \tag{23}$$

Distributions that maximize (23) 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 [16] and [17] and references therein. Moreover, the condition for the optimality of a single sphere distribution (i.e., the maximum radius $\bar{R}_n^{\text{MMSE}}$) in (23) has been found in [16] and [18] and is given by the solution to the equation:

$$\mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(R\|Z\|\right)\right] + \mathbb{E}\left[\mathsf{h}^2_{\frac{n}{2}}\left(R\|x+Z\|\right)\right] = 1, \tag{24}$$

where $\|x\| = R$. It is pleasing to see the similarity between the optimality condition for the MMSE in (24) and the optimality condition for the mutual information in (5a). Note, however, that (5a) could not have been derived directly from (24) or vice versa.

It is also interesting to point out that that $\bar{R}_n^{\text{MMSE}}$ is always lagging behind $\bar{R}_n$ as we increase n as shown in Fig. 4. Notably this behavior persists even as $n \to \infty$ since for the MMSE (see [16] and [19])

$$\lim_{n\to\infty} \frac{\bar{R}_n^{\text{MMSE}}}{\sqrt{n}} \approx 1.1509, \tag{25}$$

while for the mutual information according to Theorem 2

$$\lim_{n\to\infty} \frac{\bar{R}_n}{\sqrt{n}} \approx 1.8609, \tag{26}$$

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 $\bar{R}_n^{\text{MMSE}} \le R \le \bar{R}_n$:

$$\max_{X \in \mathcal{B}_0(R)} I(X;Y) = \max_{X \in \mathcal{B}_0(R)} \frac{1}{2}\int_0^1 \text{mmse}(X|Y_\gamma)\text{d}\gamma \tag{27}$$

$$\le \frac{1}{2}\int_0^1 \max_{X \in \mathcal{B}_0(R)} \text{mmse}(X|Y_\gamma)\text{d}\gamma. \tag{28}$$

Such a condition for tightness of the bound via the I-MMSE relation was already pointed out in [20] for $n = 1$. Interestingly for several multiuser problems [21]–[23], 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 (28) (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 (23) Gaussian inputs are not optimal as $n \to \infty$.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] A. Dytso, H. V. Poor, and S. Shamai (Shitz), "On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective," 2018. [Online]. Available: https://arxiv.org/abs/1804.08524

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

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

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.