On 2×2 MIMO Gaussian Channels with a Small Discrete-Time Peak-Power Constraint

Alex Dytso*, Luca Barletta†, and Gerhard Kramer**
* Qualcomm Flarion Technologies, Bridgewater, NJ 08807, USA. Email: odytso2@gmail.com
† Politecnico di Milano, Milano, 20133, Italy. Email: luca.barletta@polimi.it
** Technical University of Munich, Munich, 80333, Germany. Email: gerhard.kramer@tum.de

***Abstract*—A multi-input multi-output (MIMO) Gaussian channel with two transmit antennas and two receive antennas is studied that is subject to an input peak-power constraint. The capacity and the capacity-achieving input distribution are unknown in general. The problem is shown to be equivalent to a channel with an identity matrix but where the input lies inside and on an ellipse with principal axis length r_p and minor axis length r_m. If $r_p \leq \sqrt{2}$, then the capacity-achieving input has support on the ellipse. A sufficient condition is derived under which a two-point distribution is optimal. Finally, if $r_m < r_p \leq \sqrt{2}$, then the capacity-achieving distribution is discrete.**

I. Introduction

The capacity of multi-input multi-output (MIMO) Gaussian noise channels with an average-power constraint is known. However, in practice one also has a peak-power constraint, and determining the corresponding capacity is challenging. The goal of this work is to study the capacity under a (total) peak-power constraint $\|\mathbf{X}\|^2 \leq 1$, namely

$$\max_{P_{\mathbf{X}}:\|\mathbf{X}\|^2 \leq 1} I(\mathbf{X};\mathbf{H}\mathbf{X}+\mathbf{Z}) \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{2\times 2}$ is a channel matrix known to both the receiver and the transmitter, $\mathbf{Z} \in \mathbb{R}^2$ is a standard normal vector independent of the input $\mathbf{X} \in \mathbb{R}^2$, and $\|\cdot\|$ is the Euclidean norm.

A. Problem Reformulation

We reformulate the problem as follows:

$$\mathsf{C}(r_p, r_m) = \max_{P_{\mathbf{X}}:\mathbf{X}\in\mathcal{E}(r_p,r_m)} I(\mathbf{X};\mathbf{X}+\mathbf{Z}),\ r_p, r_m \geq 0, \tag{2}$$

where $\mathcal{E}(r_p, r_m)$ is the set of points inside and on the two-dimensional ellipse:

$$\mathcal{E}(r_p, r_m) = \left\{ (x_1, x_2) \in \mathbb{R}^2 : \frac{x_1^2}{r_p^2} + \frac{x_2^2}{r_m^2} \leq 1 \right\}, \tag{3}$$

where we assume $r_p \geq r_m$ and refer to r_p as a primary radius and r_m as a minor radius. The equivalence of (1) and (2) is shown in [1, App. A].

To motivate our discussion, suppose $r_m \to 0$, i.e., the channel becomes one-dimensional with constraint $|X_1| \leq r_p$. The capacity-achieving input distribution is discrete with finite support [2]. Moreover, if $r_p \leq \bar{r}_p \approx 1.6$ then the optimal input distribution is (discrete) uniform on $\{-r_p, r_p\}$ [3]. Suppose next that $r_p = r_m$. The optimal input distribution is uniform on a circle of radius r_p for all $r_p \leq \tilde{r}_p \approx 2.4$ [4], [5]. One goal of this work is to explore what happens when $0 < r_m < r_p \leq \tilde{r}_p$.

B. Literature Review

The single-input single-output (SISO) version of the channel was considered in [2] and the capacity-achieving input distribution is discrete with finite support. The optimality of the two-point distribution was characterized in [3]. For further insights into the structure of the capacity-achieving distribution, the reader is referred to [6]. Upper and lower bounds on the capacity appear in [7]–[10].

The MIMO problem seems considerably more difficult than the SISO problem. For example, the support of the capacity-achieving distribution is, in general, not a union of isolated points but a union of lower-dimensional manifolds [11]–[13]. Little is known about the structure of these manifolds. The structure is known when the channel matrix is proportional to an identity matrix, in which case the support consists of concentric shells [4], [5], [14], [15]. Upper and lower bounds on the capacity of general MIMO channels for both total and per-antenna peak-power constraints were considered in [16]–[19].

C. Outline and Contributions

This paper is organized as follows. Sec. I-D describes the notation. Sec. II presents our main results. Specifically, Sec. II-B shows that for a sufficiently small ellipse the capacity-achieving distribution is supported on the boundary of the input space. Sec. II-C establishes symmetry properties and Sec. II-D proves a sufficient condition for optimality of a two-point distribution. Sec. II-E shows that, if the ellipse is small enough, then the capacity-achieving distribution is discrete. Sec. III provides proofs. Finally, Sec. IV concludes with a discussion and future directions.

Some of the proofs are omitted and can be found in [1].

D. Notation

The probability of a measurable subset $\mathcal{A} \subseteq \mathbb{R}^n$ is written as $P_{\mathbf{X}}(\mathcal{A}) = \mathbb{P}[\mathbf{X} \in \mathcal{A}]$. The support of $P_{\mathbf{X}}$ is

$$\mathsf{supp}(P_{\mathbf{X}}) = \{\mathbf{x} : P_{\mathbf{X}}(\mathcal{D}) > 0 \text{ for every open set } \mathcal{D} \ni \mathbf{x}\ \}.$$

If $\mathbf{X}$ is discrete, we write $P_{\mathbf{X}}(\mathbf{x})$ for its probability mass function (pmf). The probability density function (pdf) of a standard Gaussian is denoted by ϕ. The relative entropy of P and Q is $\mathsf{D}(P\,\|\,Q)$. For two random vectors $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{X} \overset{d}{=} \mathbf{Y}$ denotes equality in distribution.

II. Main Results

A. Karush-Kuhn-Tucker (KKT) Conditions

Lemma 1. *The optimal input distribution $P_{\mathbf{X}^\star}$ satisfies*

$$\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}^\star}\right) \le \mathsf{C}(r_p, r_m), \quad \mathbf{x} \in \mathcal{E}(r_p, r_m) \tag{4}$$

$$\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}^\star}\right) = \mathsf{C}(r_p, r_m), \quad \mathbf{x} \in \mathrm{supp}(P_{\mathbf{X}^\star}) \tag{5}$$

where $P_{\mathbf{Y}^\star}$ is the optimal output distribution.

Proof. This is an extension of the case $r_p = r_m$ treated in [4]. ■

Note that the KKT conditions imply that the set $\mathrm{supp}(P_{\mathbf{X}^\star})$ is the set of global maxima of the function $\mathbf{x} \mapsto \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}^\star}\right)$. This fact will be used several times.

B. When is the Support Concentrated on the Ellipse?

The theorem below identifies a sufficient condition such that the support of the optimal input distribution lies on the ellipse. To state the theorem, we need the following definition and lemma.

Definition 1. *Let f be a real-valued function that is twice continuously differentiable on an open set $\mathcal{G} \subset \mathbb{R}^n$. Then f is* subharmonic *if $\nabla^2 f \ge 0$ on $\mathcal{G}$, where ∇^2 is the Laplacian, i.e., the trace of the Hessian matrix.*

Lemma 2. *If $\|\mathbf{X}\| \le \sqrt{2}$, then the function $\mathbf{x} \mapsto \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}}\right)$ is subharmonic.*

Proof. The Hatsel-Nolte identity [20], [21] gives

$$\mathsf{H}_{\mathbf{y}} \log f_{\mathbf{Y}}(\mathbf{y}) = \mathsf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y}) - \mathbf{I}, \tag{6}$$

where $f_{\mathbf{Y}}$ is the output probability density function, $\mathsf{H}_{\mathbf{y}} f(\mathbf{y})$ is the Hessian matrix of function f with respect to $\mathbf{y}$, and $\mathbf{I}$ is an identity matrix. We further have

$$\begin{aligned}\mathsf{H}_{\mathbf{x}} \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}}\right) &= -\mathsf{H}_{\mathbf{x}} \mathbb{E}\left[\log f_{\mathbf{Y}}(\mathbf{x}+\mathbf{Z})\right] \\ &= \mathbf{I} - \mathbb{E}\left[\mathsf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{x}+\mathbf{Z})\right]. \end{aligned} \tag{7}$$

Taking the trace of (7), we arrive at

$$\begin{aligned}\nabla^2_{\mathbf{x}} \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}}\right) &= 2 - \mathbb{E}\left[\mathrm{Tr}(\mathsf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{x}+\mathbf{Z}))\right] \\ &\ge 2 - 2 = 0\end{aligned}$$

where we used $\|\mathbf{X}\| \le \sqrt{2}$. ■

Theorem 1. *For $r_m \le r_p \le \sqrt{2}$, we have*

$$\mathrm{supp}(P_{\mathbf{X}^\star}) \subseteq \partial\mathcal{E}(r_p, r_m), \tag{8}$$

where $\partial\mathcal{E}(r_p, r_m)$ is the boundary (an ellipse) of $\mathcal{E}(r_p, r_m)$.

Proof. Lemma 2 shows that $\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}}\right)$ is subharmonic if the support of $P_{\mathbf{X}}$ is in a sufficiently small ball. Next, note that non-constant subharmonic functions are maximized on the boundary of a closed non empty domain [22]. Since $\mathbf{x} \mapsto \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}}\right)$ is subharmonic for $r_p \le \sqrt{2}$, the maximum of $\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}}\right)$ over $\mathcal{E}(r_p, r_m)$ occurs on the boundary of $\mathcal{E}(r_p, r_m)$ for $r_m \le r_p \le \sqrt{2}$. ■

Remark 1. Lemma 2 generalizes to vector channels of any dimension n if $\|\mathbf{X}\| \le \sqrt{2}$ is replaced with $\|\mathbf{X}\| \le \sqrt{n}$.

We now focus on $r_m \le r_p \le \sqrt{2}$, which we call the *small peak-power regime*. Theorem 1 simplifies the KKT conditions.

Lemma 3. *$P_{\mathbf{X}^\star}$ is optimal if and only if*

$$\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}^\star}\right) \le \mathsf{C}(r_p, r_m), \quad \mathbf{x} \in \partial\mathcal{E}(r_p, r_m) \tag{9}$$

$$\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}^\star}\right) = \mathsf{C}(r_p, r_m), \quad \mathbf{x} \in \mathrm{supp}(P_{\mathbf{X}^\star}) \tag{10}$$

Note that in (9), we check only on the ellipse rather than on the region enclosed by the ellipse.

C. Symmetry

We prove the following symmetry result.

Proposition 1. *The marginals $P_{X_1^\star}$ and $P_{X_2^\star}$ are symmetric. Also, if $0 < r_m \le r_p \le \sqrt{2}$, then for $\mathbf{x} \in \mathcal{E}(r_p, r_m)$, we have*

$$\begin{aligned} &P_{X_2^\star|X_1^\star=x_1}(x_2) \\ &= \frac{1}{2}\delta\left(x_2 - r_m\sqrt{1-\frac{x_1^2}{r_p^2}}\right) + \frac{1}{2}\delta\left(x_2 + r_m\sqrt{1-\frac{x_1^2}{r_p^2}}\right)\end{aligned} \tag{11}$$

where $\delta(x)$ is the Dirac measure at x.

Proof: Observe that

$$\begin{aligned}\mathsf{C}(r_p, r_m) &= I\left(\mathbf{X}^\star; \mathbf{X}^\star + \mathbf{Z}\right) \\ &= I\left((-1)\mathbf{X}^\star; (-1)(\mathbf{X}^\star + \mathbf{Z})\right) \\ &= I\left(-\mathbf{X}^\star; -\mathbf{X}^\star + \mathbf{Z}\right)\end{aligned} \tag{12}$$

where (12) follows because $-\mathbf{Z} \overset{d}{=} \mathbf{Z}$. Note that $-\mathbf{X}^\star \in \mathcal{E}(r_p, r_m)$, and (12) implies the input $-\mathbf{X}^\star$ is also capacity-achieving. However, since the capacity-achieving distribution is unique [13, Thm. 1], we have $-\mathbf{X}^\star \overset{d}{=} \mathbf{X}^\star$. Next, marginalizing with respect to $X_2^\star$ gives $X_1^\star \overset{d}{=} -X_1^\star$, i.e., $X_1^\star$ is symmetric. Similarly, the marginal of $X_2^\star$ is symmetric.

To show (11), we have $\mathbf{X}^\star \in \partial\mathcal{E}(r_p, r_m)$ for $r_m \le r_p \le \sqrt{2}$. Given $X_1^\star = x_1$, this implies

$$\begin{aligned} &P_{X_2^\star|X_1^\star=x_1}(x_2) \\ &= p\,\delta\left(x_2 - r_m\sqrt{1-\frac{x_1^2}{r_p^2}}\right) + q\,\delta\left(x_2 + r_m\sqrt{1-\frac{x_1^2}{r_p^2}}\right)\end{aligned} \tag{13}$$

for some $0 \le p, q \le 1$ such that $p+q=1$. Finally, $p = q = \frac{1}{2}$ follows by the symmetry of $X_2^\star$. ■

D. When are Two Mass Points Optimal?

We characterize the regime where two points are optimal.

Theorem 2. *If $r_p \le \sqrt{2}$ then*

$$P_{\mathbf{X}^\star}(-r_p, 0) = P_{\mathbf{X}^\star}(r_p, 0) = \frac{1}{2} \tag{14}$$

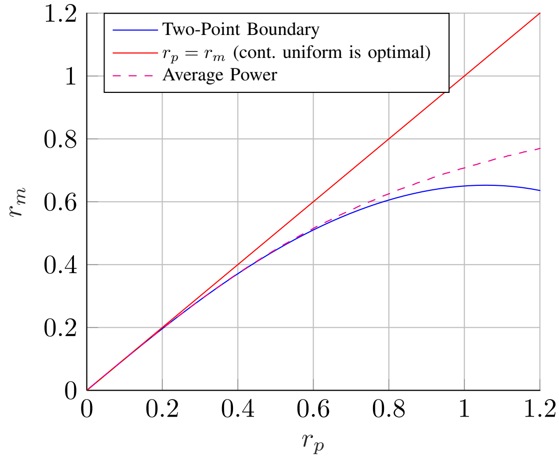

Fig. 1. Region of optimality of two points.

is optimal if and only if $\frac{r_m^2}{2}$ *is less than*

$$\int_{-\infty}^{\infty} (\phi(y - r_p) - \phi(y)) \log \frac{1}{\phi(y - r_p) + \phi(y + r_p)} \mathrm{d}y. \tag{15}$$

Proof. See Section III-A. ∎

Fig. 1 plots the boundary (in blue) of the region for which two mass points are optimal.

Remark 2. It is interesting to compare the result in Theorem 2 to having an average-power constraint. The capacity with an average-power constraint is

$$\max_{\mathbb{E}[\|\mathbf{X}\|^2] \leq 1} I(\mathbf{X}; \mathbf{D}\mathbf{X} + \mathbf{Z}), \tag{16}$$

where $\mathbf{D}$ may be chosen as diagonal because the singular value decomposition of $\mathbf{H}$ and the isotropic pdf of $\mathbf{Z}$ permit this restriction. The capacity-achieving distribution is Gaussian with a diagonal covariance matrix with $\mathbb{E}[(X_1^\star)^2] = P_1$ and $\mathbb{E}[(X_2^\star)^2] = P_2$ and where P_1 and P_2 are given by the waterfilling solution

$$P_i = \max\left(\nu - \frac{1}{D_{ii}^2}, 0\right), \quad i \in \{1, 2\}, \tag{17}$$

where ν is chosen so that $P_1 + P_2 = 1$. To compare, let $D_{11}^2 = r_p^2$ and $D_{22}^2 = r_m^2$; We are interested in the regime where only one antenna is used (i.e., $P_2 = 0$). The red dashed curve of Fig. 1 depicts the boundary of the region for which using only one antenna is optimal. Observe that the curves for the peak-power constraint and the average-power constraint at first follow each other but then they diverge.

Remark 3. Consider the r_p, r_m for which a four-point distribution is optimal. Unfortunately, finding a simple condition similar to (15) for four points does not seem possible. For example, such condition is not known even for $n = 1$. Also, by symmetry, the optimal input distribution is

$$P_{\mathbf{X}^\star}(\mathbf{x}) = \begin{cases} \frac{p}{2} & \mathbf{x} = (r_p, 0) \\ \frac{p}{2} & \mathbf{x} = (-r_p, 0) \\ \frac{1-p}{2} & \mathbf{x} = (0, r_m) \\ \frac{1-p}{2} & \mathbf{x} = (0, -r_m) \end{cases} \tag{18}$$

for some $p \in [0, 1]$. Because the distribution is parameterized by three parameters (p, r_p, r_m), it seems feasible to check the KKT conditions. However, because of the multivariate nature of the problem, numerical methods suffer from instabilities, and we were unable to reliably compute $P_{\mathbf{X}^\star}$.

E. On the Structure of $P_{\mathbf{X}^\star}$

We next resolve the question posed in Section I-A, i.e., we show that the capacity-achieving input in the small peak-power regime is discrete with finitely many points if $r_p \neq r_m$ and is uniform on the circle with radius r_p if $r_p = r_m$.

Theorem 3. *Suppose* $0 \leq r_m \leq r_p < \sqrt{2}$. *Then we have*

- $\mathsf{supp}(P_{\mathbf{X}^\star})$ *is a finite set for* $r_m \neq r_p$ *or* $r_m = r_p = 0$;
- $P_{\mathbf{X}^\star}$ *is uniformly distributed on* $\partial\mathcal{E}(r_p, r_p)$ *for* $0 < r_m = r_p$, *i.e.,* $\mathsf{supp}(P_{\mathbf{X}^\star})$ *has infinite cardinality.*

Proof. See Section III-B. ∎

The next theorem shows that the points $(\pm r_p, 0)$ are part of the optimal input distribution.

Theorem 4. *If* $0 \leq r_m \leq r_p$ *then we have*

$$\{(-r_p, 0), (r_p, 0)\} \subseteq \mathsf{supp}(P_{\mathbf{X}^\star}). \tag{19}$$

Proof. See Section III-C. ∎

III. Proofs

A. Proof of Theorem 2

To check the optimality of $P_{\mathbf{X}^\star}$, we check the KKT conditions in Lemma 1. We have

$$P_{\mathbf{Y}^\star}(y_1, y_2) = P_{Y_1^\star}(y_1) P_{Y_2|X_2}(y_2|0) \tag{20}$$

so the KL divergence simplifies to

$$\begin{aligned} &\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|\mathbf{x}) \,\middle\|\, P_{\mathbf{Y}^\star}\right) \\ &= \mathsf{D}\left(P_{Y_1|X_1}(\cdot|x_1) \,\middle\|\, P_{Y_1^\star}\right) + \mathsf{D}\left(P_{Y_2|X_2}(\cdot|x_2) \,\middle\|\, P_{Y_2|X_2}(\cdot|0)\right) && (21) \\ &= \mathsf{D}\left(P_{Y_1|X_1}(\cdot|x_1) \,\middle\|\, P_{Y_1^\star}\right) + \frac{x_2^2}{2} && (22) \\ &= \mathsf{D}\left(P_{Y_1|X_1}(\cdot|x_1) \,\middle\|\, P_{Y_1^\star}\right) + \frac{r_m^2\left(1 - \frac{x_1^2}{r_p^2}\right)}{2}, && (23) \end{aligned}$$

where the last step uses Theorem 1, i.e., $\frac{x_1^2}{r_p^2} + \frac{x_2^2}{r_m^2} = 1$. Now let

$$\begin{aligned} f(x_1) = \mathsf{D}\left(P_{Y_1|X_1}(\cdot|x_1) \,\middle\|\, P_{Y_1^\star}\right) + \frac{r_m^2\left(1 - \frac{x_1^2}{r_p^2}\right)}{2} \\ - \mathsf{D}\left(P_{Y_1|X_1}(\cdot|r_p) \,\middle\|\, P_{Y_1^\star}\right), \end{aligned} \tag{24}$$

where, due to symmetry, we have $\mathsf{D}\left(P_{Y_1|X_1}(\cdot|r_p) \,\middle\|\, P_{Y_1^\star}\right) = I(X_1^\star; Y_1^\star)$. Furthermore, the KKT conditions imply that if $P_{X_1^\star}$ is optimal, then

$$\mathrm{supp}(P_{X_1^\star}) \subseteq \{ \text{ maxima of } f(x_1) \text{ on } [-r_p, r_p]\}. \tag{25}$$

Therefore, we study the maxima of f by finding the derivative of f, which is

$$\begin{aligned}
f'(x_1) &= -\frac{\mathrm{d}}{\mathrm{d}x_1}\mathbb{E}[\log f_{Y_1^\star}(Z_1 + x_1)] - \frac{r_m^2 x_1}{r_p^2} && (26)\\
&= x_1 - \mathbb{E}[\mathbb{E}[X_1^\star | Y_1^\star = Z_1 + x_1]] - \frac{r_m^2 x_1}{r_p^2} && (27)\\
&= \left(1 - \frac{r_m^2}{r_p^2}\right) x_1 - r_p \mathbb{E}[\tanh(r_p(Z_1 + x_1))] && (28)\\
&= \mathbb{E}\left[\left(1 - \frac{r_m^2}{r_p^2}\right) Z_{x_1} - r_p \tanh(r_p Z_{x_1})\right] && (29)\\
&= \mathbb{E}\left[\eta(Z_{x_1})\right], && (30)
\end{aligned}$$

where in (27) we have used Tweedie's formula [23]

$$\mathbb{E}[X|Y = y] = y + \frac{\mathrm{d}}{\mathrm{d}y} \log f_Y(y). \tag{31}$$

In (28) we used $\mathbb{E}[X_1^\star | Y_1^\star = y] = r_p \tanh(r_p y)$; in (30) we defined $Z_{x_1} = Z_1 + x_1$ and

$$\eta(y) = \left(1 - \frac{r_m^2}{r_p^2}\right) y - r_p \tanh(r_p y). \tag{32}$$

Now observe the following about the function f':

1) at $x_1 = 0$, f' has a sign change. This follows from $f'(0) = 0$ and that f is even and, therefore, f' is odd and there must be a sign change at $x_1 = 0$.
2) the function f' can have at most three sign changes. This follows from because $\eta(y)$ can have at most three sign changes and using Karlin's oscillation theorem (see, e.g., [6]) which states that the number of sign changes of f' is upper-bounded by the number of sign changes of η.
3) if $r_m > 0$ and the function f' has exactly three sign changes, then sign changes must be in the order $-, +, -, +$. This follows because there is a sign change at $x_1 = 0$ and $\lim_{x_1 \to -\infty} f'(x_1) = -\infty$ and $\lim_{x_1 \to \infty} f'(x_1) = \infty$. Therefore, the sign changes at $x_1 \neq 0$ must be from negative to positive, and the sign change at $x_1 = 0$ must be from positive to negative.

In what follows, we focus on $x_1 \geq 0$. The above properties imply that only the following three cases are possible.

Case 1: f' is non-negative on $x_1 \geq 0$. This implies that, the function f is increasing and, therefore, is maximized at $x_1 = r_p$.

Case 2: f' is non-positive on $x_1 \geq 0$. In this case, the function f is decreasing and is maximized at $x_1 = 0$.

Case 3: On $x_1 > 0$, f' has a single sign change from negative to positive. Thus, f has only one local minimum and no local maxima, and f is maximized at $x_1 = 0$ or $x_1 = r_p$.

The above shows the function f on $x_1 \geq 0$ can have maxima only at either $x_1 = 0$ or $x_1 = r_p$. This implies $\mathrm{supp}(P_{X_1}^\star) = \{-r_p, r_p\}$ if and only if

$$\begin{aligned}
&f(0) < f(r_p)\\
&\Leftrightarrow \mathsf{D}\left(P_{Y_1|X_1}(\cdot|0) \,\middle\|\, P_{Y_1^\star}\right) + \frac{r_m^2}{2} < \mathsf{D}\left(P_{Y_1|X_1}(\cdot|r_p) \,\middle\|\, P_{Y_1^\star}\right) && (33)\\
&\Leftrightarrow \frac{r_m^2}{2} < \int_{-\infty}^{\infty} (\phi(y - r_p) - \phi(y)) \log \frac{1}{\phi(y - r_p) + \phi(y + r_p)} \mathrm{d}y && (34)
\end{aligned}$$

B. *Proof of Theorem 3*

Before proving the theorem, we present definitions and ancillary results.

Definition 2. *Let $\mathbf{V} \in \mathbb{R}^2$ have moment generating function $M_{\mathbf{V}}(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^2$. Then the cumulant generating function is*

$$K_{\mathbf{V}}(\mathbf{t}) = \log(M_{\mathbf{V}}(\mathbf{t})), \, \mathbf{t} \in \mathbb{R}^2. \tag{35}$$

The following lemma is useful.

Lemma 4. *For $x_1 \in [-r_p, r_p]$ let*

$$f(x_1; P_{\mathbf{X}^\star}) = \mathbb{E}\left[K_{\bar{\mathbf{X}}}\left(Z_1 + x_1, Z_2 + r_m \sqrt{1 - \frac{x_1^2}{r_p^2}}\right)\right], \tag{36}$$

where $\bar{\mathbf{X}}$ is distributed as $\mathrm{d}P_{\bar{\mathbf{X}}} = \frac{\mathrm{e}^{-\frac{\|\mathbf{x}\|^2}{2}}}{\mathbb{E}[\mathrm{e}^{-\frac{\|\mathbf{X}^\star\|^2}{2}}]} \mathrm{d}P_{\mathbf{X}^\star}$ (a tilted distribution). The KKT conditions (9)–(10) *are equivalent to*

$$0 \geq \frac{x_1^2}{2}\left(1 - \frac{r_m^2}{r_p^2}\right) - f(x_1; P_{\mathbf{X}^\star}) - a, \quad x_1 \in [-r_p, r_p], \tag{37}$$

$$0 = \frac{x_1^2}{2}\left(1 - \frac{r_m^2}{r_p^2}\right) - f(x_1; P_{\mathbf{X}^\star}) - a, \quad x_1 \in \mathsf{supp}(P_{X_1^\star}). \tag{38}$$

where $a = \mathsf{C}(r_p, r_m) - \log(2\pi) + \log\left(\mathbb{E}\left[\mathrm{e}^{-\frac{\|\mathbf{X}^\star\|^2}{2}}\right]\right) + h(\mathbf{Z}) - 1 - \frac{r_m^2}{2}$.

Proof. See [1, App. C]. ■

Definition 3. *For $z \in \mathbb{C}$ the principle square root function is*

$$z^{\frac{1}{2}} = |z|^{\frac{1}{2}} \mathrm{e}^{\frac{i \arg(z)}{2}}. \tag{39}$$

Remark 4. The principle square root function is holomorphic on $\mathbb{C} \setminus \{z : \mathrm{Re}(z) \leq 0, \mathrm{Im}(z) = 0\}$ [24].

Lemma 5. *The function $\mathbf{x} \mapsto \mathbb{E}\left[K_{\bar{\mathbf{X}}}(\mathbf{Z} + \mathbf{x})\right]$ has the complex extension*

$$\mathbb{E}_{\mathbf{Z}}\left[K_{\bar{\mathbf{X}}}(\mathbf{Z} + \mathbf{z})\right] = \mathbb{E}_{\mathbf{Z}}\left[\log \mathbb{E}_{\bar{\mathbf{X}}}\left[\mathrm{e}^{(\mathbf{Z}+\mathbf{z})^T \bar{\mathbf{X}}}\right]\right], \, \mathbf{z} \in \mathbb{C}^2. \tag{40}$$

Moreover, the function in (40) *is an entire function, i.e., analytic on $\mathbb{C}^2$. Consequently, the function $x_1 \mapsto f(x_1; P_{\mathbf{X}^\star})$ has a complex extension*

$$\tilde{f}(z; P_{\mathbf{X}^\star})$$

$$= \mathbb{E}_{\mathbf{Z}} \left[K_{\bar{\mathbf{X}}} \left(Z_1 + z, Z_2 + r_m \sqrt{\left| 1 - \frac{z^2}{r_p^2} \right| \mathrm{e}^{\frac{i \arg\left(1 - \frac{z^2}{r_p^2}\right)}{2}}} \right) \right] \tag{41}$$

which is analytic on $z \in \mathbb{C} \setminus \{z : |\mathrm{Re}(z)| \geq r_p, \mathrm{Im}(z) = 0\}$.

Proof. See [1, App. D]. ∎

The next lemma provides an upper bound on the peak power of $\left|\tilde{f}(z; P_{\mathbf{X}^\star})\right|$ along the imaginary axis.

Lemma 6. *For all* $u \in \mathbb{R}$

$$\left|\tilde{f}(iu; P_{\mathbf{X}^\star})\right| \leq \frac{r_p^2}{2} + \frac{r_m^2}{2} + r_m^2 \sqrt{1 + \frac{u^2}{r_p^2}} + 2\pi. \tag{42}$$

Proof. See [1, App. E]. ∎

Proof of Theorem 3: Suppose $\mathsf{supp}(P_{X_1^\star})$ has infinite cardinality. Starting with one of the KKT equations in (38), we have

$$0 = \frac{x_1^2}{2}\left(1 - \frac{r_m^2}{r_p^2}\right) - f(x_1; P_{\mathbf{X}^\star}) - a,\ x_1 \in \mathsf{supp}(P_{X_1^\star}) \tag{43}$$

$$\Rightarrow\ 0 = \frac{x_1^2}{2}\left(1 - \frac{r_m^2}{r_p^2}\right) - f(x_1; P_{\mathbf{X}^\star}) - a,\ x_1 \in (-r_p, r_p) \tag{44}$$

$$\Rightarrow\ 0 = \frac{z^2}{2}\left(1 - \frac{r_m^2}{r_p^2}\right) - \tilde{f}(z; P_{\mathbf{X}^\star}) - a, \quad z \in \mathbb{C} \setminus \{z : |\mathrm{Re}(z)| \geq r_p, \mathrm{Im}(z) = 0\} \tag{45}$$

$$\Rightarrow 0 = -\frac{u^2}{2}\left(1 - \frac{r_m^2}{r_p^2}\right) - \tilde{f}(iu; P_{\mathbf{X}^\star}) - a,\ u \in \mathbb{R}, \tag{46}$$

where (44) follows because (Bolzano-Weierstrass Theorem with the Identity Theorem) an analytic function that is equal to zero on a closed bounded set of infinite cardinality is equal to zero everywhere on its domain; and (45) follows because $\tilde{f}(z; P_{\mathbf{X}^\star})$ is the analytic extension of $f(x; P_{\mathbf{X}^\star})$. Now, Lemma 6 shows that $|\tilde{f}(iu; P_{\mathbf{X}^\star})| \in O(u)$, together with (46), leads to a contradiction.

C. Proof of Theorem 4

By symmetry of the channel (see Proposition 1), we may focus on the case $x_1 > 0$.

Lemma 7. *Suppose* $0 \leq r_m \leq r_p$ *and* $\Pr(X_1 > t) = 0$ *for some* $0 < t < r_p$. *Then for* $x_1 > t$, *we have*

$$\frac{\partial}{\partial x_1} \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|(x_1, x_2)) \,\middle\|\, P_{\mathbf{Y}}\right) > 0, \qquad \forall x_2 \in \mathbb{R}. \tag{47}$$

Proof. For all $x_2 \in \mathbb{R}$, we can write

$$\begin{aligned}
&\frac{\partial}{\partial x_1} \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|(x_1, x_2)) \,\middle\|\, P_{\mathbf{Y}}\right) \\
&= -\frac{\partial}{\partial x_1} \mathbb{E}\left[\log f_{\mathbf{Y}}(\mathbf{x} + \mathbf{Z})\right] && (48)\\
&= -\mathbb{E}\left[\nabla \log f_{\mathbf{Y}}(\mathbf{x} + \mathbf{Z})\right]^{\mathsf{T}} \cdot \frac{\partial}{\partial x_1}\mathbf{x} && (49)\\
&= -\mathbb{E}\left[\mathbb{E}\left[\mathbf{X}|\mathbf{Y} = \mathbf{x} + \mathbf{Z}\right] - \mathbf{x} + \mathbf{Z}\right]^{\mathsf{T}} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} && (50)\\
&= -\mathbb{E}\left[\mathbb{E}\left[X_1|\mathbf{Y} = \mathbf{x} + \mathbf{Z}\right]\right] + x_1 && (51)\\
&\geq -t + x_1, && (52)
\end{aligned}$$

where (50) follows from Tweedie's formula [23] and (52) due to $\Pr(X_1 > t) = 0$. Finally, combine (52) with $x_1 > t$. ∎

Arguing by contradiction, let $t < r_p$ be the largest value of the $\mathsf{supp}(P_{X_1^\star})$. Then, Lemma 7 implies

$$\mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|(x_1, x_2)) \,\middle\|\, P_{\mathbf{Y}^\star}\right) > \mathsf{D}\left(P_{\mathbf{Y}|\mathbf{X}}(\cdot|(t, x_2)) \,\middle\|\, P_{\mathbf{Y}^\star}\right) \tag{53}$$

for all $x_1 > t$ and $x_2 \in \mathbb{R}$. This would violate the KKT condition (4). As a consequence, we necessarily have $(r_p, 0) \in \mathsf{supp}(P_{\mathbf{X}^\star})$ and, by symmetry, $(-r_p, 0) \in \mathsf{supp}(P_{\mathbf{X}^\star})$.

IV. Discussion and Conclusion

We studied the capacity-achieving distribution of 2×2 MIMO channels subject to a peak-power constraint. We characterized conditions for the optimality of a two-point distribution and, by transforming the problem into a one-dimensional problem, we derived sufficient conditions for the discreteness of capacity-achieving distributions.

There are several interesting directions for future work. For example, an ambitious goal is extending the results to any number of antennas. It would also be intriguing to address questions in the 2×2 setting. For instance, let $P_{\mathbf{X}^\star_{r_p,r_m}}$ be the optimal input distribution for the specified values of r_p and r_m. Since capacity is a concave function of input constraint [25] (see also [26, Ch. 3.3]) and since concave functions are continuous [27], we have

$$\begin{aligned}
&\lim_{r_m \to r_p} \mathsf{C}(r_p, r_m) = \mathsf{C}(r_p, r_p) \\
&\iff \lim_{r_m \to r_p} I(\mathbf{X}^\star_{r_p,r_m}; \mathbf{X}^\star_{r_p,r_m} + \mathbf{Z}) = I(\mathbf{X}^\star_{r_p}; \mathbf{X}^\star_{r_p} + \mathbf{Z}),
\end{aligned} \tag{54}$$

where $\mathbf{X}^\star_{r_p} = \mathbf{X}^\star_{r_p,r_p}$. Note that $\mathbf{X}^\star_{r_p}$ is uniformly distributed on the circle of radius r_p. Furthermore, by the continuity of mutual information in the distribution [28], we have

$$\lim_{r_m \to r_p} \mathbf{X}^\star_{r_p,r_m} \stackrel{d}{=} \mathbf{X}^\star_{r_p}. \tag{55}$$

Note that $P_{\mathbf{X}^\star_{r_p}}$ has infinite support while $P_{\mathbf{X}^\star_{r_p,r_m}}$ has finite support for all $r_m < r_p \leq \sqrt{2}$ by Theorem 3. Thus, an interesting open problem is the rate at which the cardinality of the set $\mathsf{supp}(P_{\mathbf{X}^\star_{r_p,r_m}})$ approaches infinity as $r_m \to r_p$.

References

[1] A. Dytso, L. Barletta, and G. Kramer, "On 2 × 2 MIMO Gaussian channels with a small discrete-time peak-power constraint," *Extended version. Available on arXiv: https://arxiv.org/abs/2401.17084*, 2024.

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

[3] N. Sharma and S. Shamai, "Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels," *Prob. Inf. Transm.*, vol. 46, pp. 283–299, 2010.

[4] 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, 1995.

[5] A. Dytso, M. Al, H. V. Poor, and S. Shamai Shitz, “On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective,” *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3907–3921, 2019.

[6] A. Dytso, S. Yagli, H. V. Poor, and S. S. Shamai, “The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points,” *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2006–2022, 2019.

[7] 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, 2017.

[8] B. Rassouli and B. Clerckx, “An upper bound for the capacity of amplitude-constrained scalar AWGN channel,” *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 1924–1926, 2016.

[9] A. L. McKellips, “Simple tight bounds on capacity for the peak-limited discrete-time channel,” in *IEEE Int. Symp. Inf. Theory*, 2004, pp. 348–348.

[10] A. Dytso, M. Goldenbaum, H. V. Poor, and S. S. Shitz, “A generalized Ozarow-Wyner capacity bound with applications,” in *IEEE Int. Symp. Inf. Theory*. IEEE, 2017, pp. 1058–1062.

[11] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), “Amplitude constrained MIMO channels: Properties of optimal input distributions and bounds on the capacity,” *Entropy*, vol. 21, no. 2, 2019. [Online]. Available: https://www.mdpi.com/1099-4300/21/2/200

[12] J. Eisen, R. R. Mazumdar, and P. Mitran, “Capacity-achieving input distributions of additive vector Gaussian noise channels: Even-moment constraints and unbounded or compact support,” *Entropy*, vol. 25, no. 8, p. 1180, 2023.

[13] 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, 2005.

[14] 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, 2016.

[15] A. Favano, M. Ferrari, M. Magarini, and L. Barletta, “The capacity of the amplitude-constrained vector Gaussian channel,” in *IEEE Int. Symp. Inf. Theory*. IEEE, 2021, pp. 426–431.

[16] A. El Moslimany 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, 2016.

[17] A. Dytso, M. Goldenbaum, S. Shamai, and H. V. Poor, “Upper and lower bounds on the capacity of amplitude-constrained MIMO channels,” in *IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[18] A. Favano, M. Ferrari, M. Magarini, and L. Barletta, “A sphere packing bound for vector Gaussian fading channels under peak amplitude constraints,” *IEEE Trans. Inf. Theory*, vol. 69, no. 1, pp. 238–250, 2022.

[19] ——, “The capacity of fading vector Gaussian channels under amplitude constraints on antenna subsets,” in *2022 IEEE Information Theory Workshop (ITW)*, 2022, pp. 178–183.

[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, H. V. Poor, and S. S. Shamai, “Conditional mean estimation in Gaussian noise: A meta derivative identity with applications,” *IEEE Trans. Inf. Theory*, vol. 69, no. 3, pp. 1883–1898, 2022.

[22] E. M. Stein and R. Shakarchi, *Complex Analysis*. Princeton University Press, 2010, vol. 2.

[23] H. E. Robbins, “An empirical Bayes approach to statistics,” in *Breakthroughs in Statistics: Foundations and basic theory*. Springer, 1992, pp. 388–394.

[24] J. Bak, D. J. Newman, and D. J. Newman, *Complex Analysis*. Springer, 2010, vol. 8.

[25] C. E. Shannon *et al.*, “Coding theorems for a discrete source with a fidelity criterion,” *IRE Nat. Conv. Rec*, vol. 4, no. 142-163, p. 1, 1959.

[26] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[27] R. Webster, *Convexity*. Oxford university press, 1994.

[28] 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, 2011.