

Article

Amplitude Constrained MIMO Channels: Properties of Optimal Input Distributions and Bounds on the Capacity [†]

Alex Dytso ^{1,*}, Mario Goldenbaum ^{1,*}, H. Vincent Poor ^{1,*} and Shlomo Shamai (Shitz) ^{2,*}¹ Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA² Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel

* Correspondence: adytso@princeton.edu (A.D.); goldenbaum@princeton.edu (M.G.); poor@princeton.edu (H.V.P.); sshlomo@ee.technion.ac.il (S.S.)

[†] Parts of the material in this paper were presented at the 2017 IEEE Global Communications Conference (Singapore, 4–8 December 2017).

Received: 21 January 2019; Accepted: 13 February 2019; Published: 19 February 2019



Abstract: In this work, the capacity of multiple-input multiple-output channels that are subject to constraints on the support of the input is studied. The paper consists of two parts. The first part focuses on the general structure of capacity-achieving input distributions. Known results are surveyed and several new results are provided. With regard to the latter, it is shown that the support of a capacity-achieving input distribution is a small set in both a topological and a measure theoretical sense. Moreover, explicit conditions on the channel input space and the channel matrix are found such that the support of a capacity-achieving input distribution is concentrated on the boundary of the input space only. The second part of this paper surveys known bounds on the capacity and provides several novel upper and lower bounds for channels with arbitrary constraints on the support of the channel input symbols. As an immediate practical application, the special case of multiple-input multiple-output channels with *amplitude constraints* is considered. The bounds are shown to be within a constant gap to the capacity if the channel matrix is invertible and are tight in the high amplitude regime for arbitrary channel matrices. Moreover, in the regime of high amplitudes, it is shown that the capacity scales linearly with the minimum between the number of transmit and receive antennas, similar to the case of average power-constrained inputs.

Keywords: MIMO; channel capacity; amplitude constraint; input distribution; capacity bounds

1. Introduction

While the capacity of a multiple-input multiple-output (MIMO) channel with an average power constraint is well understood [1], there is surprisingly little known about the capacity of the more practically relevant case in which the channel inputs are subject to *amplitude constraints*. Shannon was the first who considered a channel that is constrained in its amplitude [2]. In that paper, he derived corresponding upper and lower bounds and showed that in the low-amplitude regime, the capacity behaves as that of channel with an average power constraint. The next major contribution to this problem was a seminal paper of Smith [3] published in 1971. Smith showed that, for the single-input single-output (SISO) Gaussian noise channel with an amplitude-constrained input, the capacity-achieving inputs are discrete with finite support. In [4], this result is extended to peak-power-constrained quadrature Gaussian channels. Using the approach of Shamai [4], it is shown in [5] that the input distribution that achieves the capacity of a MIMO channel with an identity channel matrix and a Euclidian norm constraint on the input vector is discrete. Even though the optimal input

distribution is known to be discrete, very little is known about the number or the optimal positions of the corresponding constellation points. To the best of our knowledge, the only case for which the input distribution is precisely known is considered in [6], where it is shown for the Gaussian SISO channel with an amplitude constraint that two point masses are optimal if amplitude values are smaller than 1.665 and three for amplitude values of up to 2.786. Finally, it has been shown very recently that the number of mass points in the support of the capacity-achieving input distribution of a SISO channel is of the order $O(A^2)$ with A the amplitude constraint.

Based on a dual capacity expression, in [7], McKellips derived an upper bound on the capacity of a SISO channel that is subject to an amplitude constraint. The bound is asymptotically tight; that is, for amplitude values that tend to infinity. By cleverly choosing an auxiliary channel output distribution in the dual capacity expression, the authors of [8] sharpened McKellips' upper bound and extended it to parallel MIMO channels with a Euclidian norm constraint on the input. The SISO version of the upper bound in [8] has been further sharpened in [9] by yet another choice of auxiliary output distribution. In [10], asymptotic lower and upper bounds for a 2×2 MIMO channel are presented and the gap between the bounds is specified.

In this work, we make progress on this open problem by deriving several new upper and lower bounds that hold for channels with *arbitrary constraints* on the support of the channel input distribution and then apply them to the practically relevant special case of MIMO channels that are subject to amplitude-constraints.

1.1. Contributions and Paper Organization

The remainder of the paper is organized as follows. The problem is stated in Section 2. In Section 3, we study properties of input distributions that achieve the capacity of input-constrained MIMO channels. The section reviews known results on the structure of optimal input distributions and presents several new results. In particular, Theorem 3 shows that the support of a capacity-achieving input distribution must necessarily be a small set both topologically and measure theoretically. Moreover, Theorem 8 characterizes conditions on the channel input space as well as on the channel matrix such that the support of the optimal input distribution is concentrated on the boundary of the channel input space.

In Section 4, we derive novel upper and lower bounds on the capacity of a MIMO channel that is subject to an arbitrary constraint on the support of the input. In particular, three families of upper bounds are proposed, which are based on: (i) the maximum entropy principle (see the bound in Theorem 9); (ii) the dual capacity characterization (see the bound in Theorem 10); and (iii) a relationship between mutual information and the minimum mean square error that is known as the I-MMSE relationship (see the bound in Theorem 11). On the other hand, Section 4 provides three different lower bounds. The first one is given in Theorem 12 and is based on the entropy power inequality. The second one (see Theorem 13) is based on a generalization of the celebrated Ozarow–Wyner bound [11] to the MIMO case. The third upper bound (see Theorem 14) is based on Jensen's inequality and depends on the characteristic function of the channel input distribution.

In Section 5, we evaluate the performance of our bounds by studying MIMO channels with invertible channel matrices. In particular, Theorem 17 states that our upper and lower bounds are within $n \log_2(\rho)$ bits, where ρ is the packing efficiency and n the number of transmit and receive antennas. For diagonal channel matrices, it is then shown (see Theorem 18) that the Cartesian product of simple pulse-amplitude modulation (PAM) constellations achieves the capacity to within $1.64n$ bits.

Section 6 is devoted to MIMO channels with arbitrary channel matrices. It is shown that, in the regime of high amplitudes, similar to the case of average power-constrained channel inputs, the capacity scales linearly with the minimum of the number of transmit and receive antennas.

In Section 7, our upper and lower bounds are applied to the SISO case, which are then compared with bounds known from the literature. Finally, Section 8 concludes the paper. Note that parts of the results in this paper were also published in [12].

1.2. Notation

Vectors are denoted as bold lowercase letters, random vectors as bold uppercase letters, and matrices as bold uppercase sans serif letters (e.g., \mathbf{x} , \mathbf{X} , \mathbf{X}). For any deterministic vector $\mathbf{x} \in \mathbb{R}^n$, $n \in \mathbb{N}$, we denote the Euclidian norm of \mathbf{x} by $\|\mathbf{x}\|$. For some random $\mathbf{X} \in \text{supp}(\mathbf{X}) \subseteq \mathbb{R}^n$ and any $p > 0$, we define

$$\|\mathbf{X}\|_p := \left(\frac{1}{n} \mathbb{E}[\|\mathbf{X}\|^p] \right)^{\frac{1}{p}}, \quad (1)$$

where $\text{supp}(\mathbf{X})$ denotes the support of \mathbf{X} . Note that for $p \geq 1$, the quantity in Equation (1) defines a norm and for $n = 1$ we simply have $\|X\|_p^p = \mathbb{E}[|X|^p]$.

The norm of a matrix $\mathbf{H} \in \mathbb{R}^{n \times n}$ is defined as

$$\|\mathbf{H}\| := \sup_{\mathbf{x}: \mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{H}\mathbf{x}\|}{\|\mathbf{x}\|},$$

whereas $\text{Tr}(\mathbf{H})$ is denoting its trace. The $n \times n$ identity matrix is represented as \mathbf{I}_n .

Let \mathcal{S} be a subset of \mathbb{R}^n . Then,

$$\text{Vol}(\mathcal{S}) := \int_{\mathcal{S}} d\mathbf{x}$$

denotes its volume. Moreover, the boundary of \mathcal{S} is denoted as $\partial\mathcal{S}$.

Let $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$. We define an n -dimensional ball or radius $r \in \mathbb{R}_+$ centered at $\mathbf{x} \in \mathbb{R}^n$ as the set

$$\mathcal{B}_{\mathbf{x}}(r) := \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{y}\| \leq r\}.$$

Recall that, for any $\mathbf{x} \in \mathbb{R}^n$ and $r \in \mathbb{R}_+$,

$$\text{Vol}(\mathcal{B}_{\mathbf{x}}(r)) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} r^n,$$

where $\Gamma(z)$ denotes the gamma function.

For any matrix $\mathbf{H} \in \mathbb{R}^{k \times n}$ and some $\mathcal{S} \subset \mathbb{R}^n$, we define

$$\mathbf{H}\mathcal{S} := \{\mathbf{y} \in \mathbb{R}^k : \mathbf{y} = \mathbf{H}\mathbf{x}, \mathbf{x} \in \mathcal{S}\}.$$

Note that for an invertible $\mathbf{H} \in \mathbb{R}^{n \times n}$, we have

$$\text{Vol}(\mathbf{H}\mathcal{S}) = |\det(\mathbf{H})| \text{Vol}(\mathcal{S})$$

with $\det(\mathbf{H})$ the determinant of \mathbf{H} . We define the maximum and minimum radius of a set $\mathcal{S} \subset \mathbb{R}^n$ that contains the origin as

$$\begin{aligned} r_{\max}(\mathcal{S}) &:= \min\{r \in \mathbb{R}_+ : \mathcal{S} \subset \mathcal{B}_0(r)\}, \\ r_{\min}(\mathcal{S}) &:= \max\{r \in \mathbb{R}_+ : \mathcal{B}_0(r) \subseteq \mathcal{S}\}. \end{aligned}$$

For a given vector $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}_+^n$, we define

$$\text{Box}(\mathbf{a}) := \{\mathbf{x} \in \mathbb{R}^n : |x_i| \leq a_i, i = 1, \dots, n\}$$

and the smallest box containing a given set $\mathcal{S} \subset \mathbb{R}^n$ as

$$\text{Box}(\mathcal{S}) := \inf\{\text{Box}(\mathbf{a}) : \mathcal{S} \subseteq \text{Box}(\mathbf{a})\},$$

respectively.

The entropy of any discrete random object \mathbf{X} is denoted as $H(\mathbf{X})$, whereas $h(\mathbf{X})$ (i.e., the differential entropy) is used whenever \mathbf{X} is continuous. The mutual information between two random objects \mathbf{X} and \mathbf{Y} is denoted as $I(\mathbf{X}; \mathbf{Y})$ and $\mathcal{N}(\mathbf{m}, \mathbf{C})$ denotes the multivariate normal distribution with mean vector \mathbf{m} and covariance matrix \mathbf{C} . Finally, $\log_a^+(x) := \max\{\log_a(x), 0\}$, for any base $a > 0$, $Q(x)$, $x \in \mathbb{R}$, denotes the Q-function, and $\delta_{\mathbf{x}}(\mathbf{y})$ the Kronecker delta, which is one for $\mathbf{x} = \mathbf{y}$ and zero otherwise.

2. Problem Statement

Consider a MIMO system with $n_t \in \mathbb{N}$ transmit and $n_r \in \mathbb{N}$ receive antennas. The corresponding n_r -dimensional channel output for a single channel use is of the form

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z},$$

for some fixed channel matrix $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ (considering a real-valued channel model is without loss of generality). Here and hereafter, we assume $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_r})$ is independent of the channel input $\mathbf{X} \in \mathbb{R}^{n_t}$ and \mathbf{H} is known to both the transmitter and the receiver.

Now, let $\mathcal{X} \subset \mathbb{R}^{n_t}$ be a *convex and compact* channel input space that contains the origin (i.e., the length- n_t zero vector) and let $F_{\mathbf{X}}$ denote the cumulative distribution function of \mathbf{X} . As of the writing of this paper, the capacity

$$C(\mathcal{X}, \mathbf{H}) := \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} I(\mathbf{X}; \mathbf{Y}) = \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}), \quad (2)$$

of this channel is unknown and we are interested in finding novel lower and upper bounds. Even though most of the results in this paper hold for arbitrary convex and compact \mathcal{X} , we are mainly interested in the two important special cases:

- (i) per-antenna amplitude constraints, i.e., $\mathcal{X} = \text{Box}(\mathbf{a})$ for some given $\mathbf{a} = (A_1, \dots, A_{n_t}) \in \mathbb{R}_+^{n_t}$; and
- (ii) n_t -dimensional amplitude constraint, i.e., $\mathcal{X} = \mathcal{B}_0(A)$ for some given $A \in \mathbb{R}_+$.

Remark 1. Note that determining the capacity of a MIMO channel with average per-antenna power constraints is also still an open problem and has been solved for some special cases only [13–17].

3. Properties of an Optimal Input Distribution

Unlike the special cases of real and complex-valued SISO channels (i.e., $n_t = n_r = 1$), the structure of the capacity-achieving input distribution, denoted as $F_{\mathbf{X}}^*$, is in general not known. To motivate why in this paper we are seeking for novel upper and lower bounds on the capacity (Equation (2)) instead of trying to solve the optimization problem directly, in this section we first summarize properties optimal input distributions must possess, which demonstrate how complicated the optimization problem actually is. Note that, whereas an optimal input distribution always exists, it does not necessarily need to be unique.

3.1. Necessary and Sufficient Conditions for Optimality

To study properties of an optimal input distribution, we need the notion of a point of increase of probability distribution.

Definition 1. (Points of Increase of a Distribution) A point $\mathbf{x} \in \mathbb{R}^n$, $n \in \mathbb{N}$, is said to be a point of increase of a given probability distribution $F_{\mathbf{X}}$ if for any open set $\mathcal{A} \subset \mathbb{R}^n$ containing \mathbf{x} , $F_{\mathbf{X}}(\mathcal{A}) > 0$.

The following result provides necessary and sufficient conditions for the optimality of a channel input distribution.

Theorem 1. Let F_X be some given channel input distribution and let $\mathcal{E}(F_X) \subset \mathcal{X}$ denote the set of points of increase of F_X . Then, the following holds:

- F_X is capacity-achieving if and only if the Karush–Kuhn–Tucker (KKT) conditions

$$h(\mathbf{x}; F_X) \leq h(\mathbf{H}\mathbf{x} + \mathbf{Z}), \mathbf{x} \in \mathcal{X}, \quad (3a)$$

$$h(\mathbf{x}; F_X) = h(\mathbf{H}\mathbf{x} + \mathbf{Z}), \mathbf{x} \in \mathcal{E}(F_X) \subset \mathcal{X}, \quad (3b)$$

are satisfied [3,18], where

$$h(\mathbf{x}; F_X) := - \int_{\mathbb{R}^{n_r}} \frac{1}{(2\pi)^{\frac{n_r}{2}}} e^{-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{2}} \log_2(f_Y(\mathbf{y})) d\mathbf{y}$$

with f_Y being the probability density of the channel output induced by the channel input $\mathbf{X} \sim F_X$.

- F_X is unique and symmetric if \mathbf{H} is left invertible [18].
- F_Y (i.e., the channel output distribution) is unique [18,19].

3.2. General Structure of Capacity-Achieving Input Distributions

Theorem 1 can be used to find general properties of the support of a capacity-achieving input distribution, which we will do in this subsection.

Remark 2. Fully characterizing an input distribution that achieves the capacity of a general MIMO channel with per-antenna or an n_t -dimensional amplitude constraint is still an open problem. To the best of our knowledge, the only general available for showing that discrete channel inputs are optimal was developed by Smith in [3] for the amplitude and variance-constrained Gaussian SISO channel. Since then, it has been useful to also characterize the optimal input distribution of several other SISO channels (see, for instance, [4,20–24]). The method relies on the following series of steps:

1. Towards a contradiction, it is assumed that the set of points of increase $\mathcal{E}(F_X)$ is infinite.
2. The assumption in Step 1 is then used to establish a certain property of the function $h(\mathbf{x}; F_X)$ on the input space \mathcal{X} . For example, by showing that $h(\mathbf{x}; F_X)$ has an analytic continuation to \mathbb{C} . Then, by means of the Identity Theorem of complex analysis and the Bolzano–Weierstrass Theorem [25], Smith was able to show that $h(\mathbf{x}; F_X)$ must be constant.
3. By using either the Fourier or Laplace transform of $h(\mathbf{x}; F_X)$ together with the property of $h(\mathbf{x}; F_X)$ established in Step 2, a new property of the channel output distribution F_Y is established. For example, Smith was able to show that F_Y must be constant.
4. A conclusion out of Step 3 is used to reach a contradiction. The contradiction implies that $\mathcal{E}(F_X)$ must be finite. For example, to reach a contradiction, Smith was using the fact that the channel output distribution F_Y results from a convolution with a Gaussian probability density, which cannot be constant.

Remark 3. Under the restriction that the output space, \mathcal{Y} , of a Gaussian SISO channel is finite and the channel input space, \mathcal{X} , is subject to an amplitude constraint, Witsenhausen has shown in [26] that the capacity-achieving input distribution is discrete with the number of mass points bounded as $|\mathcal{X}| \leq |\mathcal{Y}|$. The approach of Witsenhausen, however, does not use the variational technique of Smith and relies on arguments from convex analysis instead.

According to Remark 2, assuming in the MIMO case that $\mathcal{E}(F_X)$ is of infinite cardinality does not help (or at least it is not clear how this assumption should be used) in showing that the capacity-achieving input distribution is discrete and finite. However, by using the weaker assumption that $\mathcal{E}(F_X)$ contains a non-empty open subset in conjunction with the following version of the Identity Theorem, we can show that the support of the optimal input distribution is a small set in a certain topological sense.

Theorem 2. (Identity Theorem for Real-Analytic Functions [27]) For some $n \in \mathbb{N}$ let \mathcal{U} be a subset of \mathbb{R}^n and $f, g : \mathcal{U} \rightarrow \mathbb{R}$ be two real-analytic functions that agree on a set $\mathcal{A} \subseteq \mathcal{U}$. Then, f and g agree on \mathbb{R}^n if one of the following two conditions is satisfied:

- (i) \mathcal{A} is an open set.
- (ii) \mathcal{A} is a set of positive Lebesgue measure.

Furthermore, for $n = 1$, it suffices for \mathcal{A} to be an arbitrary set with an accumulation point.

We also need the definitions of a dense and a nowhere dense set.

Definition 2. (Dense and Nowhere Dense Sets) A subset $\mathcal{A} \subset \mathcal{X}$ is said to be dense in the set \mathcal{X} if every element $\mathbf{x} \in \mathcal{X}$ either belongs to \mathcal{A} or is an accumulation point of \mathcal{A} . A subset $\mathcal{A} \subset \mathcal{X}$ is said to be nowhere dense if for every nonempty open subset $\mathcal{U} \subset \mathcal{X}$, the intersection $\mathcal{A} \cap \mathcal{U}$ is not dense in \mathcal{U} .

With Theorem 2 at our disposal, we are now able to prove the following result on the structure of the support of the optimal input distribution.

Theorem 3. The set of points of increase $\mathcal{E}(F_{\mathbf{X}}^*)$ of an optimal input distribution $F_{\mathbf{X}}^*$ is a nowhere dense subset of \mathcal{X} that is of Lebesgue measure zero.

Proof. It is not difficult to show that $h(\mathbf{x}; F_{\mathbf{X}}^*)$ is a real-analytic function on \mathbb{R}^{n_t} ([18] Proposition 5). Now, in order to prove the result, we follow a series of steps similar to those outlined in Remark 2. Towards a contradiction, assume that the set of points of increase $\mathcal{E}(F_{\mathbf{X}}^*)$ of $F_{\mathbf{X}}^*$ is not a nowhere dense subset of \mathcal{X} . Then, according to Definition 2, there exists an open set $\mathcal{U} \subset \mathcal{X}$ such that $\mathcal{E}(F_{\mathbf{X}}^*) \cap \mathcal{U}$ is dense in \mathcal{U} .

By using the KKT condition in Equation (3b), we have that $h(\mathbf{x}; F_{\mathbf{X}}^*)$ is constant on the intersection $\mathcal{E}(F_{\mathbf{X}}^*) \cap \mathcal{U}$. Thus, as $\mathcal{E}(F_{\mathbf{X}}^*) \cap \mathcal{U}$ is dense in \mathcal{U} , it follows by the properties of continuous functions (real-analytic functions are continuous) that $h(\mathbf{x}; F_{\mathbf{X}}^*)$ is also constant on \mathcal{U} . Moreover, as \mathcal{U} is an open set, Theorem 2 implies that $h(\mathbf{x}; F_{\mathbf{X}}^*)$ must also be constant on \mathbb{R}^{n_t} . This, however, leads to a contradiction as $h(\mathbf{x}; F_{\mathbf{X}}^*)$ cannot be constant on all of \mathbb{R}^{n_t} , which can be shown by taking the Fourier transform of $h(\mathbf{x}; F_{\mathbf{X}}^*)$ and solving for the probability density $f_{\mathbf{Y}}(\mathbf{y})$ of the channel output (the reader is referred to [3] for details). Therefore, we conclude that $\mathcal{E}(F_{\mathbf{X}}^*)$ is a nowhere dense subset of \mathcal{X} .

Showing that $\mathcal{E}(F_{\mathbf{X}}^*)$ has Lebesgue measure zero follows along similar lines by assuming that $\mathcal{E}(F_{\mathbf{X}}^*)$ is a set of positive measure. Then, Property (ii) of Theorem 2 can be used to conclude that $h(\mathbf{x}; F_{\mathbf{X}}^*)$ must be zero on all of \mathcal{U} . This again leads to a contradiction, which implies that $\mathcal{E}(F_{\mathbf{X}}^*)$ must be of Lebesgue measure zero. \square

Remark 4. Note that if $\mathcal{X} = \mathcal{B}_0(A)$ for some $A \in \mathbb{R}_+$ and $h(\mathbf{x}; F_{\mathbf{X}}^*)$ is orthogonally equivariant (i.e., it only depends on $\|\mathbf{x}\|$), then $\mathcal{E}(F_{\mathbf{X}}^*)$ can be written as a union of concentric spheres. That is,

$$\mathcal{E}(F_{\mathbf{X}}^*) = \bigcup_j \mathcal{C}(A_j) \quad (4)$$

with $\mathcal{C}(A_j) := \{\mathbf{x} \in \mathbb{R}^{n_t} : \|\mathbf{x}\| = A_j\}$ for some $A_j \in \mathbb{R}_+$. To see this, let

$$g(\mathbf{x}) := h(\mathbf{x}; F_{\mathbf{X}}^*) - h(\mathbf{H}\mathbf{x} + \mathbf{Z})$$

and observe that if $\mathbf{x} \in \mathcal{E}(F_{\mathbf{X}}^*)$, then $g(\mathbf{x}) = 0$. Combining this with the symmetry of the function $\|\mathbf{x}\| \mapsto g(\|\mathbf{x}\|)$, we have that (We know that it is abuse of notation to use the same letter for the functions $\mathbf{x} \mapsto g(\mathbf{x})$ and $\|\mathbf{x}\| \mapsto g(\|\mathbf{x}\|)$ even if they are different. It is an attempt to say in a compact way that g is orthogonally equivariant.)

$$\forall \|\mathbf{x}\| = \|\mathbf{y}\| : \mathbf{x} \in \mathcal{E}(F_{\mathbf{X}}^*) \Rightarrow \mathbf{y} \in \mathcal{E}(F_{\mathbf{X}}^*).$$

Moreover, this implies that

$$\mathcal{E}(F_{\mathbf{X}}^*) = \bigcup_{j \in \mathcal{I}} \mathcal{C}(A_j),$$

where \mathcal{I} is possibly of infinite cardinality. In fact, \mathcal{I} has finite cardinality. To see this, note that, if $g(\mathbf{x})$ is real-analytic, then so is $g(\|\mathbf{x}\|)$. However, as $\|\mathbf{x}\| \mapsto g(\|\mathbf{x}\|)$ is a non-zero real-analytic function on \mathbb{R} , it can have at most finitely many zeros on an interval.

As an example consider the special case $n_r = n_t = n$ with $\mathbf{H} = \mathbf{I}_n$. Then, the union in Equation (4) implies that the cardinality of $\mathcal{E}(F_{\mathbf{X}}^*)$ is uncountable and that discrete inputs are in general not optimal. Therefore, Theorem 3 can generally not be improved in the sense that for $n > 1$, statements about the cardinality of $\mathcal{E}(F_{\mathbf{X}}^*)$ cannot be made. Note, however, that the magnitude of \mathbf{X} is discrete. An example of the corresponding optimal input distribution for the case of $n = 2$ is given in Figure 1.

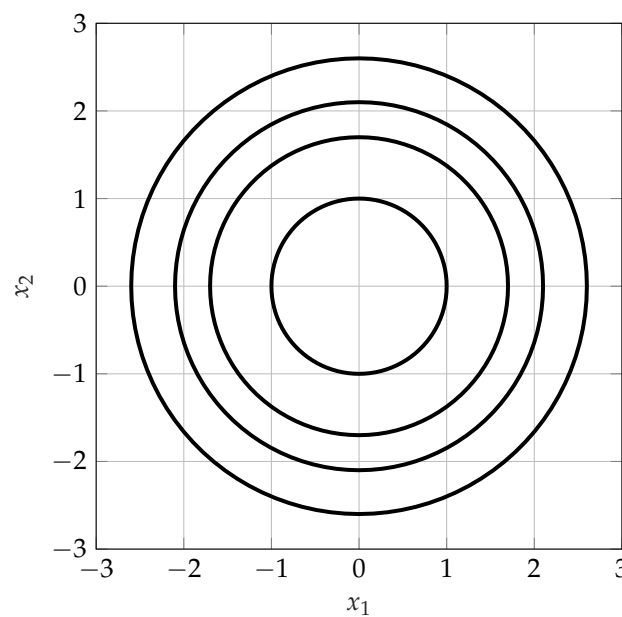


Figure 1. An example of a support of an optimal input distribution for the special case $n_t = n_r = n = 2$.

Even though Theorem 3 does not allow us to conclude that the optimal input distribution of an arbitrary MIMO channel is discrete and finite, for the special case of a SISO channel we have the following partial result.

Theorem 4. (Optimal Input Distribution of a SISO Channel [3,6,28]) For some fixed $h \in \mathbb{R}$ and $A \in \mathbb{R}_+$, consider the SISO channel $Y = hX + Z$ with input space $\mathcal{X} = [-A, A]$. Let $F_{\mathbf{X}}^*$ be an input distribution that achieves the capacity, $C(\mathcal{X}, h)$, of that channel. Then, $F_{\mathbf{X}}^*$ satisfies the following properties:

- $F_{\mathbf{X}}^*$ is unique.
- $F_{\mathbf{X}}^*$ is symmetric.
- $F_{\mathbf{X}}^*$ is discrete with the number of mass points being of the order $O(A^2)$.
- $F_{\mathbf{X}}^*$ contains probability mass points at $\{-A, A\}$.

Moreover, binary communication with mass points at $\{-A, A\}$ is optimal if and only if $A \leq \bar{A}$, where $\bar{A} \approx 1.665$.

Theorem 4 can now be used to also address the special cases of multiple-input single output (MISO) and single-input multiple output (SIMO) channels.

Theorem 5. Let $Y = \mathbf{h}^T \mathbf{X}^* + Z$ be a MISO channel with channel matrix $\mathbf{h}^T \in \mathbb{R}^{n_t}$ and some optimal input $\mathbf{X}^* \in \mathcal{X} \subset \mathbb{R}^{n_t}$. Then, the distribution of $\mathbf{h}^T \mathbf{X}^*$ is discrete with finitely many mass points. On the other hand, let $\mathbf{Y} = \mathbf{h} \mathbf{X}^* + \mathbf{Z}$ be a SIMO channel with channel matrix $\mathbf{h} \in \mathbb{R}^{n_r}$. Then, the optimal input $\mathbf{X}^* \in \mathcal{X} \subset \mathbb{R}$ has a discrete distribution with finitely many mass points.

Proof. For the MISO case, the capacity can be expressed as

$$\begin{aligned} C(\mathcal{X}, \mathbf{h}^T) &= \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} I(\mathbf{X}; \mathbf{h}^T \mathbf{X} + Z) \\ &= \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} I(\mathbf{h}^T \mathbf{X}; \mathbf{h}^T \mathbf{X} + Z) \\ &= \max_{F_S: S \in \mathbf{h}^T \mathcal{X}} I(S; S + Z) \\ &= \max_{F_S: |S| \leq r_{\max}(\mathbf{h}^T \mathcal{X})} I(S; S + Z). \end{aligned} \quad (5)$$

Using Theorem 5 we have that the maximizing distribution in Equation (5) F_S^* , where $S := \mathbf{h}^T \mathbf{X}$, is discrete with finitely many mass points.

For the SIMO case, the channel input distribution is discrete as a SIMO channel can be transformed into a SISO channel. Thus, let $A \in \mathbb{R}_+$ be finite. Then, the capacity of the SIMO channel can be expressed as

$$\begin{aligned} C(\mathcal{X}, \mathbf{h}) &= \max_{F_{\mathbf{X}}: |\mathbf{X}| \leq A} I(\mathbf{X}; \mathbf{h} \mathbf{X} + \mathbf{Z}) \\ &= \max_{F_{\mathbf{X}}: |\mathbf{X}| \leq A} I(\mathbf{X}; \|\mathbf{h}\| \mathbf{X} + Z). \end{aligned} \quad (6)$$

Again, it follows from Theorem 5 that the mutual information in Equation (6) is maximized by a channel input distribution, $F_{\mathbf{X}}^*$, that is discrete with finitely many mass points. This concludes the proof. \square

Remark 5. Note that in the MISO case, we do not claim $F_{\mathbf{X}}^*$ to be discrete with finitely many points. To illustrate the difficulty, let $\mathbf{h}^T = [1, -1]$ so that

$$\mathbf{h}^T \mathbf{X}^* = \mathbf{h}^T \begin{pmatrix} X_1^* \\ X_2^* \end{pmatrix} = X_1^* - X_2^*.$$

As X_1^* and X_2^* can be arbitrarily correlated, we cannot rule out cases in which $X_1^* = X - D$ and $X_2^* = X - 2D$, with D a discrete random variable and X of arbitrary distribution. Clearly the distribution of \mathbf{X}^* is not discrete.

Note that in general it can be shown that the capacity-achieving input distribution is discrete if the optimization problem in Equation (2) can be reformulated as an optimization over one dimensional distributions. This, for example, has been done in [5] for parallel channels with a total amplitude constraint.

3.3. Properties of Capacity-Achieving Input Distributions in the Small (But Not Vanishing) Amplitude Regime

In this subsection, we study properties of capacity-achieving input distribution in the regime of small amplitudes. To that end, we will need the notion of a subharmonic function.

Definition 3. (Subharmonic Function) 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 $\nabla^2 f \geq 0$ on \mathcal{G} , where ∇^2 denotes the Laplacian (if f is twice differentiable, its Laplacian is given by $\nabla^2 f(x_1, \dots, x_n) = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$).

We use the following Theorem, which states that a subharmonic function always attains its maximum on the boundary of its domain.

Theorem 6. (Maximum Principle of Subharmonic Functions [29]) *Let $\mathcal{G} \subset \mathbb{R}^n$ be a connected open set. If $f : \mathcal{G} \rightarrow \mathbb{R}$ is subharmonic and attains a global maximum in the interior of \mathcal{G} , then f is constant on \mathcal{G} .*

In addition to Theorem 6, we need the following result that has been proven in [30].

Lemma 1. *Let the likelihood function of the output of a MIMO channel be defined as*

$$\ell : \mathbb{R}^{n_r} \rightarrow \mathbb{R}, \ell(\mathbf{y}) = \frac{f_{\mathbf{Y}}(\mathbf{y})}{\frac{1}{(2\pi)^{\frac{n_r}{2}}} e^{-\frac{\|\mathbf{y}\|^2}{2}}}$$

and let $\mathbf{A}_{\mathbf{y}}$ denote the Hessian matrix of $\log_e(\ell(\mathbf{y}))$. Then, the Laplacian (or the trace of $\mathbf{A}_{\mathbf{y}}$) is given by

$$\nabla^2 \log_e(\ell(\mathbf{y})) = \text{Tr}(\mathbf{A}_{\mathbf{y}}) = \mathbb{E}[\|\mathbf{H}\mathbf{X}\|^2 | \mathbf{Y} = \mathbf{y}] - \|\mathbb{E}[\mathbf{H}\mathbf{X} | \mathbf{Y} = \mathbf{y}]\|^2. \quad (7)$$

Theorem 7. *Suppose that $r_{\max}^2(\mathbf{H}\mathcal{X}) \leq \log_2(e)$. Then, $\mathbf{x} \mapsto h(\mathbf{x}; F_{\mathbf{X}})$ is a subharmonic function for every $F_{\mathbf{X}}$.*

Proof. Let $F_{\mathbf{X}}$ be arbitrary and observe that

$$\begin{aligned} h(\mathbf{x}; F_{\mathbf{X}}) &= -\mathbb{E}[\log_2(f_{\mathbf{Y}}(\mathbf{H}\mathbf{x} + \mathbf{Z}))] \\ &= -\mathbb{E}[\ell(\mathbf{H}\mathbf{x} + \mathbf{Z})] - \mathbb{E}\left[\log_2\left(\frac{1}{(2\pi)^{\frac{n_r}{2}}} e^{-\frac{\|\mathbf{H}\mathbf{x} + \mathbf{Z}\|^2}{2}}\right)\right] \\ &= -\mathbb{E}[\ell(\mathbf{H}\mathbf{x} + \mathbf{Z})] + \mathbb{E}\left[\frac{\|\mathbf{H}\mathbf{x} + \mathbf{Z}\|^2}{2}\right] \log_2(e) + \log_2\left((2\pi)^{\frac{n_r}{2}}\right) \\ &= -\mathbb{E}[\ell(\mathbf{H}\mathbf{x} + \mathbf{Z})] + \frac{\|\mathbf{H}\mathbf{x}\|^2 + n_r}{2} \log_2(e) + \log_2\left((2\pi)^{\frac{n_r}{2}}\right). \end{aligned}$$

With this expression in hand, the Laplacian of $h(\mathbf{x}; F_{\mathbf{X}})$ with respect to \mathbf{x} can be bounded from below as follows:

$$\begin{aligned} \nabla^2 h(\mathbf{x}; F_{\mathbf{X}}) &= \nabla^2 \left(-\mathbb{E}[\ell(\mathbf{H}\mathbf{x} + \mathbf{Z})] + \frac{\|\mathbf{H}\mathbf{x}\|^2 + n_r}{2} \log_2(e) + \log_2\left((2\pi)^{\frac{n_r}{2}}\right) \right) \\ &= -\mathbb{E}[\nabla^2 \ell(\mathbf{H}\mathbf{x} + \mathbf{Z})] + \nabla^2 \frac{\|\mathbf{H}\mathbf{x}\|^2}{2} \log_2(e) \\ &\stackrel{(a)}{=} -\mathbb{E}[\text{Tr}(\mathbf{H}\mathbf{H}^T \mathbf{A}_{\mathbf{H}\mathbf{x} + \mathbf{Z}})] + \nabla^2 \frac{\|\mathbf{H}\mathbf{x}\|^2}{2} \log_2(e) \\ &= -\mathbb{E}[\text{Tr}(\mathbf{H}\mathbf{H}^T \mathbf{A}_{\mathbf{H}\mathbf{x} + \mathbf{Z}})] + \text{Tr}(\mathbf{H}\mathbf{H}^T) \log_2(e) \\ &\stackrel{(b)}{\geq} -\mathbb{E}[\text{Tr}(\mathbf{H}\mathbf{H}^T) \text{Tr}(\mathbf{A}_{\mathbf{H}\mathbf{x} + \mathbf{Z}})] + \text{Tr}(\mathbf{H}\mathbf{H}^T) \log_2(e) \\ &= \text{Tr}(\mathbf{H}\mathbf{H}^T) (-\mathbb{E}[\text{Tr}(\mathbf{A}_{\mathbf{H}\mathbf{x} + \mathbf{Z}})] + \log_2(e)) \\ &\stackrel{(c)}{\geq} \text{Tr}(\mathbf{H}\mathbf{H}^T) \left(-r_{\max}^2(\mathbf{H}\mathcal{X}) + \log_2(e) \right), \end{aligned} \quad (8)$$

where (a) follows from Equation (7) and the chain rule for the Hessian; (b) from using the well-known inequality

$$\text{Tr}(\mathbf{C}\mathbf{D})^2 \leq \text{Tr}(\mathbf{C})^2 \text{Tr}(\mathbf{D})^2$$

that holds for \mathbf{C} and \mathbf{D} positive semi-definite; and (c) from using the inequality

$$\begin{aligned}\text{Tr}(\mathbf{A}_y) &= \mathbb{E}[\|\mathbf{H}\mathbf{X}\|^2 | \mathbf{Y} = y] - \|\mathbb{E}[\mathbf{H}\mathbf{X} | \mathbf{Y} = y]\|^2 \\ &\leq \mathbb{E}[\|\mathbf{H}\mathbf{X}\|^2 | \mathbf{Y} = y] \\ &\leq r_{\max}^2(\mathbf{H}\mathcal{X}).\end{aligned}$$

Thus, according to the assumption that $r_{\max}^2(\mathbf{H}\mathcal{X}) \leq \log_2(e)$, the right-hand side of Equation (8) is nonnegative, which proves the result. \square

Now, knowing that $h(\mathbf{x}; F_{\mathbf{X}})$ is a subharmonic function allows us to characterize the support of an optimal input distribution of a MIMO channel provided that the radius of the channel input space, \mathcal{X} , is sufficiently small.

Theorem 8. *Let $F_{\mathbf{X}}^*$ be a capacity-achieving input distribution and $r_{\max}^2(\mathbf{H}\mathcal{X}) \leq \log_2(e)$. Then, $\mathcal{E}(F_{\mathbf{X}}^*) \subseteq \partial\mathcal{X}$.*

Proof. From the KKT conditions in Equation (3), we know that, if $\mathbf{x} \in \mathcal{E}(F_{\mathbf{X}}^*)$, then \mathbf{x} is a maximizer of $h(\mathbf{x}; F_{\mathbf{X}}^*)$. According to Theorem 7, we also know that $h(\mathbf{x}; F_{\mathbf{X}}^*)$ is subharmonic. Hence, from the Maximum Principle of Subharmonic Functions (i.e., Theorem 6), it follows $\mathcal{E}(F_{\mathbf{X}}^*) \subseteq \partial\mathcal{X}$. \square

Combining Theorem 8 with the observations made in Remark 4 leads to the following corollary.

Corollary 1. *Let $\mathcal{X} = \mathcal{B}_0(A)$ and $A \leq \frac{1}{\|\mathbf{H}\|} \log_2(e)$. Then,*

$$\mathcal{E}(F_{\mathbf{X}}^*) \subseteq \mathcal{C}(A),$$

where $\mathcal{C}(A) := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = A\}$ denotes a sphere of radius A .

We conclude this section by noting that for the special case $n_t = n_r = n$ with $\mathbf{H} = \mathbf{I}_n$, the exact value of A such that $\mathcal{E}(F_{\mathbf{X}}^*) = \mathcal{C}(A)$ has been characterized in terms of an integral equation in [31], which is approximately equal to $1.5\sqrt{n}$.

4. Upper and Lower Bounds on the Capacity

The considerations in the previous section have shown that characterizing the structure of an optimal channel-input distribution is a challenging question in itself that we could only partially answer. A full characterization, however, is a necessary prerequisite to narrow down the search space in Equation (2) to one that is tractable. Except for some special cases (i.e., special choices of \mathcal{X}), optimizing over the most general space of input distributions that consists of *all* continuous n_t -dimensional probability distributions $F_{\mathbf{X}}$ with $\mathbf{X} \in \mathcal{X}$, is prohibitive (Note that Dytso et al. [32] summarized methods of how to optimize functionals over the space of probability distributions that are constrained in their support). Thus, up to the writing of this paper, there is little hope in being able to solve the problem in Equation (2) in full generality so that in this section we are proposing novel lower and upper bounds on the capacity $C(\mathcal{X}, \mathbf{H})$. Nevertheless, these bounds will allow us to better understand how the capacity of such MIMO channels behaves.

Towards this end, in Section 4.1, we provide four upper bounds. The first is based on an upper bound on the differential entropy of a random vector that is constraint in its p th moment, the second and third bounds are based on duality arguments, and the fourth on the relationship between mutual information and the minimum mean square error (MMSE), I-MMSE relationship for short, known from [33]. The three lower bounds proposed in Section 4.2, on the other hand, are based on the celebrated entropy power inequality, a generalization of the Ozarow–Wyner capacity bound taken from [11], and on Jensen’s inequality.

4.1. Upper Bounds

To establish our first upper bound on Equation (2), we need the following result ([11] Th. 1).

Lemma 2. (Maximum Entropy Under p th Moment Constraint) *Let $n \in \mathbb{N}$ and $p \in (0, \infty)$ be arbitrary. Then, for any $\mathbf{U} \in \mathbb{R}^n$ such that $h(\mathbf{U}) < \infty$ and $\|\mathbf{U}\|_p < \infty$, we have*

$$h(\mathbf{U}) \leq n \log_2 \left(k_{n,p} n^{\frac{1}{p}} \|\mathbf{U}\|_p \right),$$

where

$$k_{n,p} := \frac{\sqrt{\pi} e^{\frac{1}{p}} \left(\frac{p}{n}\right)^{\frac{1}{p}} \Gamma\left(\frac{n}{p} + 1\right)^{\frac{1}{n}}}{\Gamma\left(\frac{n}{2} + 1\right)^{\frac{1}{n}}}.$$

Theorem 9. (Moment Upper Bound) *For any channel input space \mathcal{X} and any fixed channel matrix \mathbf{H} , we have*

$$C(\mathcal{X}, \mathbf{H}) \leq \bar{C}_M(\mathcal{X}, \mathbf{H}) := \inf_{p>0} n_r \log_2 \left(\frac{k_{n_r,p}}{(2\pi e)^{\frac{1}{2}}} n_r^{\frac{1}{p}} \|\tilde{\mathbf{x}} + \mathbf{Z}\|_p \right),$$

where $\tilde{\mathbf{x}} \in \mathbf{H}\mathcal{X}$ is chosen such that $\|\tilde{\mathbf{x}}\| = r_{\max}(\mathbf{H}\mathcal{X})$.

Proof. Expressing Equation (2) in terms of differential entropies results in

$$\begin{aligned} C(\mathcal{X}, \mathbf{H}) &= \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} h(\mathbf{H}\mathbf{X} + \mathbf{Z}) - h(\mathbf{Z}) \\ &\stackrel{(a)}{\leq} \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} n_r \log_2 \left(\frac{k_{n_r,p}}{(2\pi e)^{\frac{1}{2}}} n_r^{\frac{1}{p}} \|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p \right) \\ &\stackrel{(b)}{=} n_r \log_2 \left(\frac{k_{n_r,p}}{(2\pi e)^{\frac{1}{2}}} n_r^{\frac{1}{p}} \max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} \|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p \right), \end{aligned} \quad (9)$$

where (a) follows from Lemma 2 with the fact that $h(\mathbf{Z}) = \frac{n_r}{2} \log_2(2\pi e)$; and (b) from the monotonicity of the logarithm.

Now, notice that $\|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p$ is linear in $F_{\mathbf{X}}$ and therefore it attains its maximum at an extreme point of the set $\mathcal{F}_{\mathbf{X}} := \{F_{\mathbf{X}} : \mathbf{X} \in \mathcal{X}\}$ (i.e., the set of all cumulative distribution functions of \mathbf{X}). As a matter of fact [26], the extreme points of $\mathcal{F}_{\mathbf{X}}$ are given by the set of degenerate distributions on \mathcal{X} ; that is, $\{F_{\mathbf{X}}(\mathbf{y}) = \delta_{\mathbf{x}}(\mathbf{y}), \mathbf{y} \in \mathcal{X}\}_{\mathbf{x} \in \mathcal{X}}$. This allows us to conclude

$$\max_{F_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} \|\mathbf{H}\mathbf{X} + \mathbf{Z}\|_p = \max_{\mathbf{x} \in \mathcal{X}} \|\mathbf{H}\mathbf{x} + \mathbf{Z}\|_p.$$

Observe that the Euclidian norm is a convex function, which is therefore maximized at the boundary of the set $\mathbf{H}\mathcal{X}$. Combining this with Equation (9) and taking the infimum over $p > 0$ completes the proof. \square

The following Theorem provides two alternative upper bounds that are based on duality arguments.

Theorem 10. (Duality Upper Bounds) *For any channel input space \mathcal{X} and any fixed channel matrix \mathbf{H}*

$$C(\mathcal{X}, \mathbf{H}) \leq \bar{C}_{\text{Dual},1}(\mathcal{X}, \mathbf{H}) := \log_2 \left(c_{n_r}(d) + \frac{\text{Vol}(\mathcal{B}_0(d))}{(2\pi e)^{\frac{n_r}{2}}} \right), \quad (10)$$

where

$$d := r_{\max}(\mathbf{H}\mathcal{X}), c_{n_r}(d) := \sum_{i=1}^{n_r-1} \binom{n_r-1}{i} \frac{\Gamma\left(\frac{n_r-1}{2}\right)}{2^{\frac{n_r}{2}} \Gamma\left(\frac{n_r}{2}\right)} d^i,$$

and

$$C(\mathcal{X}, \mathbf{H}) \leq \bar{C}_{\text{Dual},2}(\mathcal{X}, \mathbf{H}) := \sum_{i=1}^{n_r} \log_2 \left(1 + \frac{2A_i}{\sqrt{2\pi e}} \right), \quad (11)$$

where $\mathbf{a} = (A_1, \dots, A_{n_r})$ such that $\text{Box}(\mathbf{a}) = \text{Box}(\mathbf{H}\mathcal{X})$.

Proof. Using duality bounds, it has been shown in [8] that for any centered n -dimensional ball of radius $r \in \mathbb{R}_+$

$$\max_{\mathbf{X}: \mathbf{X} \in \mathcal{B}_0(r)} I(\mathbf{X}; \mathbf{X} + \mathbf{Z}) \leq \log_2 \left(c_n(r) + \frac{\text{Vol}(\mathcal{B}_0(r))}{(2\pi e)^{\frac{n}{2}}} \right), \quad (12)$$

where $c_n(r) := \sum_{i=1}^{n-1} \binom{n-1}{i} \frac{\Gamma\left(\frac{n-1}{2}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} r^i$.

Now, observe that

$$\begin{aligned} C(\mathcal{X}, \mathbf{H}) &= \max_{\mathbf{F}_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} h(\mathbf{H}\mathbf{X} + \mathbf{Z}) - h(\mathbf{H}\mathbf{X} + \mathbf{Z} | \mathbf{H}\mathbf{X}) \\ &= \max_{\mathbf{F}_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} I(\mathbf{H}\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) \\ &= \max_{\mathbf{F}_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathbf{H}\mathcal{X}} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \\ &\stackrel{(a)}{\leq} \max_{\mathbf{F}_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathcal{B}_0(d), d := r_{\max}(\mathbf{H}\mathcal{X})} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \\ &\stackrel{(b)}{\leq} \log_2 \left(c_{n_r}(d) + \frac{\text{Vol}(\mathcal{B}_0(d))}{(2\pi e)^{\frac{n_r}{2}}} \right). \end{aligned} \quad (13)$$

where (a) follows from enlarging the optimization domain; and (b) from using the upper bound in Equation (12). This proves Equation (10).

To show the upper bound in Equation (11), we proceed with an alternative upper bound to Equation (13):

$$\begin{aligned} C(\mathcal{X}, \mathbf{H}) &= \max_{\mathbf{F}_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \mathbf{H}\mathcal{X}} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \\ &\stackrel{(a)}{\leq} \max_{\mathbf{F}_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \text{Box}(\mathbf{H}\mathcal{X})} I(\tilde{\mathbf{X}}; \tilde{\mathbf{X}} + \mathbf{Z}) \\ &\stackrel{(b)}{\leq} \max_{\mathbf{F}_{\tilde{\mathbf{X}}}: \tilde{\mathbf{X}} \in \text{Box}(\mathbf{H}\mathcal{X})} \sum_{i=1}^{n_r} I(\tilde{X}_i; \tilde{X}_i + Z_i) \\ &\stackrel{(c)}{=} \sum_{i=1}^{n_r} \max_{\mathbf{F}_{\tilde{X}_i}: |\tilde{X}_i| \leq A_i} I(\tilde{X}_i; \tilde{X}_i + Z_i) \\ &\stackrel{(d)}{\leq} \sum_{i=1}^{n_r} \log_2 \left(1 + \frac{2A_i}{\sqrt{2\pi e}} \right), \end{aligned}$$

where the (in)equalities follow from: (a) enlarging the optimization domain; (b) single-letterizing the mutual information; (c) choosing individual amplitude constraints $(A_1, \dots, A_{n_r}) =: \mathbf{a} \in \mathbb{R}_+^{n_r}$ such that $\text{Box}(\mathbf{a}) = \text{Box}(\mathbf{H}\mathcal{X})$; and (d) using the upper bound in Equation (12) for $n = 1$. This concludes the proof. \square

As mentioned at the beginning of the section, another simple technique for deriving upper bounds on the capacity is to use the I-MMSE relationship [33]

$$I(\mathbf{X}; \mathbf{X} + \mathbf{Z}) = \frac{\log_2(e)}{2} \int_0^1 \mathbb{E} \left[\|\mathbf{X} - \mathbb{E}[\mathbf{X} | \sqrt{\gamma}\mathbf{X} + \mathbf{Z}]\|^2 \right] d\gamma. \quad (14)$$

For any $\gamma \geq 0$, the quantity $\mathbb{E} [\|\mathbf{X} - \mathbb{E}[\mathbf{X} | \sqrt{\gamma}\mathbf{X} + \mathbf{Z}]\|^2]$ is known as the MMSE of estimating \mathbf{X} from the noisy observation $\sqrt{\gamma}\mathbf{X} + \mathbf{Z}$. An important fact that will be useful is that the conditional expected value $\mathbb{E}[\mathbf{X} | \sqrt{\gamma}\mathbf{X} + \mathbf{Z}]$ is the best estimator in the sense that it minimizes the mean square error over all measurable functions $f : \mathbb{R}^{n_r} \rightarrow \mathbb{R}^{n_t}$; that is, for any $\mathbf{Y} \in \mathbb{R}^{n_r}$ and $\mathbf{X} \in \mathbb{R}^{n_t}$

$$\mathbb{E} [\|\mathbf{X} - \mathbb{E}[\mathbf{X} | \mathbf{Y}]\|^2] = \inf_{f \text{ is measurable}} \mathbb{E} [\|\mathbf{X} - f(\mathbf{Y})\|^2]. \quad (15)$$

Theorem 11. (I-MMSE Upper Bound) For any channel input space \mathcal{X} and any fixed channel matrix \mathbf{H}

$$C(\mathcal{X}, \mathbf{H}) \leq \bar{C}_{\text{I-MMSE}}(\mathcal{X}, \mathbf{H}) = \log_2(e) \begin{cases} \frac{n_r}{2} + \frac{n_r}{2} \log_2 \left(\frac{r_{\max}^2(\mathbf{H}\mathcal{X})}{n_r} \right), & r_{\max}^2(\mathbf{H}\mathcal{X}) \geq n_r \\ \frac{r_{\max}^2(\mathbf{H}\mathcal{X})}{2}, & r_{\max}^2(\mathbf{H}\mathcal{X}) \leq n_r \end{cases}.$$

Proof. Fix some $\epsilon \in [0, 1]$ and observe that

$$\begin{aligned} \frac{2}{\log_2(e)} I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) &\stackrel{(a)}{=} \frac{2}{\log_2(e)} I(\mathbf{H}\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) \\ &\stackrel{(b)}{=} \int_0^1 \mathbb{E} [\|\mathbf{H}\mathbf{X} - \mathbb{E}[\mathbf{H}\mathbf{X} | \sqrt{\gamma}\mathbf{H}\mathbf{X} + \mathbf{Z}]\|^2] d\gamma \\ &= \int_0^\epsilon \mathbb{E} [\|\mathbf{H}\mathbf{X} - \mathbb{E}[\mathbf{H}\mathbf{X} | \sqrt{\gamma}\mathbf{H}\mathbf{X} + \mathbf{Z}]\|^2] d\gamma \\ &\quad + \int_\epsilon^1 \mathbb{E} [\|\mathbf{H}\mathbf{X} - \mathbb{E}[\mathbf{H}\mathbf{X} | \sqrt{\gamma}\mathbf{H}\mathbf{X} + \mathbf{Z}]\|^2] d\gamma \\ &\stackrel{(c)}{\leq} \int_0^\epsilon \mathbb{E} [\|\mathbf{H}\mathbf{X} - \mathbf{0}\|^2] d\gamma \\ &\quad + \int_\epsilon^1 \mathbb{E} \left[\left\| \mathbf{H}\mathbf{X} - \frac{1}{\sqrt{\gamma}} (\sqrt{\gamma}\mathbf{H}\mathbf{X} + \mathbf{Z}) \right\|^2 \right] d\gamma \\ &= \epsilon \mathbb{E} [\|\mathbf{H}\mathbf{X}\|^2] + \int_\epsilon^1 \frac{1}{\gamma} \mathbb{E} [\|\mathbf{Z}\|^2] d\gamma \\ &= \epsilon \mathbb{E} [\|\mathbf{H}\mathbf{X}\|^2] + \mathbb{E} [\|\mathbf{Z}\|^2] \log_e \left(\frac{1}{\epsilon} \right) \\ &= \epsilon \mathbb{E} [\|\mathbf{H}\mathbf{X}\|^2] + n_r \log_e \left(\frac{1}{\epsilon} \right), \end{aligned}$$

where the (in)equalities follow from: (a) using that $I(\mathbf{H}\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) = I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z})$ for any fixed \mathbf{H} ; (b) using the I-MMSE relationship in Equation (14); and (c) using the property that conditional expectation minimizes mean square error (i.e., (15)).

Now, notice that

$$\begin{aligned} \max_{\mathbf{F}\mathbf{X}; \mathbf{X} \in \mathcal{X}} I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) &\leq \max_{\mathbf{F}\mathbf{X}; \mathbf{X} \in \mathcal{X}} \frac{\log_2(e)}{2} \left(\epsilon \mathbb{E} [\|\mathbf{H}\mathbf{X}\|^2] + n_r \log_e \left(\frac{1}{\epsilon} \right) \right) \\ &\stackrel{(a)}{=} \frac{1}{2} \left(\epsilon \max_{\mathbf{x} \in \mathcal{X}} \|\mathbf{H}\mathbf{x}\|^2 + n_r \log_e \left(\frac{1}{\epsilon} \right) \right) \\ &\stackrel{(b)}{=} \frac{1}{2} \left(\epsilon r_{\max}^2(\mathbf{H}\mathcal{X}) + n_r \log_e \left(\frac{1}{\epsilon} \right) \right), \end{aligned} \quad (16)$$

where (a) follows from $\max_{\mathbf{X} \in \mathcal{X}} \mathbb{E} [\|\mathbf{H}\mathbf{X}\|^2] = \max_{\tilde{\mathbf{x}} \in \mathcal{X}} \|\mathbf{H}\tilde{\mathbf{x}}\|^2$ (the same argument was used in the proof of Theorem 9); and (b) from the definition of $r_{\max}^2(\mathbf{H}\mathcal{X})$.

Since ϵ is arbitrary, we can choose it to minimize the upper bound in Equation (16). Towards this end, we need the following optimization result

$$\min_{\epsilon \in [0,1]} \left(\epsilon a + b \log \left(\frac{1}{\epsilon} \right) \right) = \begin{cases} b + b \log \left(\frac{a}{b} \right), & a \geq b \\ a, & a \leq b \end{cases} \quad (17)$$

which is easy to show. Combining Equation (16) with Equation (17), we obtain the following upper bound on the capacity

$$\max_{\mathbf{X} \in \mathcal{X}} I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) \leq \begin{cases} \frac{n_r}{2} + \frac{n_r}{2} \log \left(\frac{r_{\max}^2(\mathbf{H}\mathcal{X})}{n_r} \right), & r_{\max}^2(\mathbf{H}\mathcal{X}) \geq n_r \\ r_{\max}^2(\mathbf{H}\mathcal{X}), & r_{\max}^2(\mathbf{H}\mathcal{X}) \leq n_r \end{cases}.$$

This concludes the proof. \square

Corollary 2. For any channel input space \mathcal{X} and any fixed channel matrix \mathbf{H}

$$\bar{C}_{\text{I-MMSE}}(\mathcal{X}, \mathbf{H}) \leq \log_2(e) \begin{cases} \frac{n_r}{2} + \frac{n_r}{2} \log_e \left(\frac{\|\mathbf{H}\|^2 r_{\max}^2(\mathcal{X})}{n_r} \right), & \|\mathbf{H}\|^2 r_{\max}^2(\mathcal{X}) \geq n_r \\ \frac{\|\mathbf{H}\|^2 r_{\max}^2(\mathcal{X})}{2}, & \|\mathbf{H}\|^2 r_{\max}^2(\mathcal{X}) \leq n_r \end{cases}.$$

Proof. The corollary follows by upper bounding Equation (16) using the fact that $r_{\max}^2(\mathbf{H}\mathcal{X}) \leq \|\mathbf{H}\| r_{\max}^2(\mathcal{X})$. \square

Remark 6. In the proof of Theorem 11, instead of using sub-optimal estimators $f(\mathbf{Y}) = 0$ and $f(\mathbf{Y}) = \frac{1}{\gamma} \mathbf{Y}$, we could have used an optimal linear estimator of the form $f(\mathbf{Y}) = \mathbf{K}_{\mathbf{X}\mathbf{Y}} \mathbf{K}_{\mathbf{Y}}^{-1} \mathbf{Y}$, where $\mathbf{K}_{\mathbf{X}\mathbf{Y}}$ denotes the cross-covariance matrix between \mathbf{X} and \mathbf{Y} and $\mathbf{K}_{\mathbf{Y}}$ the covariance matrix of \mathbf{Y} . This choice would result in the capacity upper bound

$$C(\mathcal{X}, \mathbf{H}) \leq \max_{\mathbf{K}_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} \frac{1}{2} \log_2 \left(\det(\mathbf{I}_{n_r} + \mathbf{H} \mathbf{K}_{\mathbf{X}} \mathbf{H}^T) \right) \quad (18)$$

with $\mathbf{K}_{\mathbf{X}}$ the covariance matrix of the channel input. While Equation (18) is a valid upper bound, as of the writing of this paper, it is not clear how to perform an optimization over covariance matrices of random variables with bounded support. One possibility to avoid this is to use the inequality between arithmetic and geometric mean and bound the determinant by the trace:

$$\det(\mathbf{I}_{n_r} + \mathbf{H} \mathbf{K}_{\mathbf{X}} \mathbf{H}^T) \leq \left(\frac{\text{Tr}(\mathbf{I}_{n_r} + \mathbf{H} \mathbf{K}_{\mathbf{X}} \mathbf{H}^T)}{n_r} \right)^{n_r} = \left(\frac{\|\mathbf{Z} + \mathbf{H}\mathbf{X}\|_2^2}{n_r} \right)^{n_r}. \quad (19)$$

However, combining Equation (19) with Equation (18) is merely a special case of the moment upper bound of Theorem 9 for $p = 2$. Therefore, the estimators in Theorem 11 are chosen to obtain a non-trivial upper bound avoiding the optimization over covariance matrices.

In Section 5, we present a comparison of the upper bounds of Theorems 9–11 by means of a simple example.

4.2. Lower Bounds

A classical approach to bound a mutual information from below is to use the entropy power inequality (EPI).

Theorem 12. (EPI Lower Bounds) For any fixed channel matrix \mathbf{H} and any channel input space \mathcal{X} with \mathbf{X} absolutely continuous, we have

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\text{EPI}}(\mathcal{X}, \mathbf{H}) := \max_{\mathbf{X}: \mathbf{X} \in \mathcal{X}} \frac{n_r}{2} \log_2 \left(1 + \frac{2^{\frac{2}{n_r} h(\mathbf{H}\mathbf{X})}}{2\pi e} \right). \quad (20)$$

Moreover, if $n_t = n_r = n$, $\mathbf{H} \in \mathbb{R}^{n \times n}$ invertible, and \mathbf{X} uniformly distributed over \mathcal{X} , then

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\text{EPI}}(\mathcal{X}, \mathbf{H}) := \frac{n}{2} \log_2 \left(1 + \frac{|\det(\mathbf{H})|^{\frac{2}{n}} \text{Vol}(\mathcal{X})^{\frac{2}{n}}}{2\pi e} \right). \quad (21)$$

Proof. By means of the EPI

$$2^{\frac{2}{n_r} h(\mathbf{H}\mathbf{X} + \mathbf{Z})} \geq 2^{\frac{2}{n_r} h(\mathbf{H}\mathbf{X})} + 2^{\frac{2}{n_r} h(\mathbf{Z})},$$

we conclude

$$2^{\frac{2}{n_r} C(\mathcal{X}, \mathbf{H})} \geq 1 + (2\pi e)^{-1} 2^{\frac{2}{n_r} \max_{\mathbf{X}: \mathbf{X} \in \mathcal{X}} h(\mathbf{H}\mathbf{X})},$$

which finalizes the proof of the lower bound in Equation (20).

To show the lower bound in Equation (21), all we need is to recall that

$$h(\mathbf{H}\mathbf{X}) = h(\mathbf{X}) + \log_2 |\det(\mathbf{H})|,$$

which is maximized for \mathbf{X} uniformly distributed over \mathcal{X} . However, if \mathbf{X} is uniformly drawn from \mathcal{X} , we have

$$2^{\frac{2}{n} h(\mathbf{H}\mathbf{X})} = \text{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}} = |\det(\mathbf{H})|^{\frac{2}{n}} \text{Vol}(\mathcal{X})^{\frac{2}{n}},$$

which completes the proof. \square

The considerations in Section 3 suggest that a channel input distribution that maximizes Equation (2) might be discrete. Therefore, there is a need for lower bounds that unlike the bounds in Theorem 12 rely on discrete inputs.

Theorem 13. (Ozarow–Wyner Type Lower Bound) Let $\mathbf{X}_D \in \text{supp}(\mathbf{X}_D) \subset \mathbb{R}^{n_t}$ be a discrete random vector of finite entropy, $g: \mathbb{R}^{n_r} \rightarrow \mathbb{R}^{n_t}$ a measurable function, and $p > 0$. Furthermore, let \mathcal{K}_p be a set of continuous random vectors, independent of \mathbf{X}_D , such that for every $\mathbf{U} \in \mathcal{K}_p$ we have $h(\mathbf{U}) < \infty$, $\|\mathbf{U}\|_p < \infty$, and

$$\text{supp}(\mathbf{U} + \mathbf{x}_i) \cap \text{supp}(\mathbf{U} + \mathbf{x}_j) = \emptyset \quad (22)$$

for all $\mathbf{x}_i, \mathbf{x}_j \in \text{supp}(\mathbf{X}_D)$, $i \neq j$. Then,

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\text{OW}}(\mathcal{X}, \mathbf{H}) := [H(\mathbf{X}_D) - \text{gap}]^+,$$

where

$$\text{gap} := \inf_{\substack{\mathbf{U} \in \mathcal{K}_p \\ g \text{ measurable} \\ p > 0}} (G_{1,p}(\mathbf{U}, \mathbf{X}_D, g) + G_{2,p}(\mathbf{U}))$$

with

$$G_{1,p}(\mathbf{U}, \mathbf{X}_D, g) := n_t \log_2 \left(\frac{\|\mathbf{U} + \mathbf{X}_D - g(\mathbf{Y})\|_p}{\|\mathbf{U}\|_p} \right), \quad (23)$$

$$G_{2,p}(\mathbf{U}) := n_t \log_2 \left(\frac{k_{n_t,p} n_t^{\frac{1}{p}} \|\mathbf{U}\|_p}{2^{\frac{1}{n_t} h(\mathbf{U})}} \right), \quad (24)$$

and $k_{n_t,p}$ as defined in Lemma 2, respectively.

Proof. The proof is identical to ([11] Theorem 2). To make the manuscript more self-contained, we repeat it here.

Let \mathbf{U} and \mathbf{X}_D be statistically independent. Then, the mutual information $I(\mathbf{X}_D; \mathbf{Y})$ can be lower bounded as

$$\begin{aligned} I(\mathbf{X}_D; \mathbf{Y}) &\stackrel{(a)}{\geq} I(\mathbf{X}_D + \mathbf{U}; \mathbf{Y}) \\ &= h(\mathbf{X}_D + \mathbf{U}) - h(\mathbf{X}_D + \mathbf{U} | \mathbf{Y}) \\ &\stackrel{(b)}{=} H(\mathbf{X}_D) + h(\mathbf{U}) - h(\mathbf{X}_D + \mathbf{U} | \mathbf{Y}). \end{aligned} \quad (25)$$

Here, (a) follows from the data processing inequality as $\mathbf{X}_D + \mathbf{U} \rightarrow \mathbf{X}_D \rightarrow \mathbf{Y}$ forms a Markov chain in that order; and (b) from the assumption in Equation (22). By using Lemma 2, we have that the last term in Equation (25) can be bounded from above as

$$h(\mathbf{X}_D + \mathbf{U} | \mathbf{Y}) \leq n_t \log_2 \left(k_{n_t,p} n_t^{\frac{1}{p}} \|\mathbf{X}_D + \mathbf{U} - g(\mathbf{Y})\|_p \right).$$

Combining this expression with Equation (25) results in

$$I(\mathbf{X}_D; \mathbf{Y}) \geq H(\mathbf{X}_D) - (G_{1,p}(\mathbf{U}, \mathbf{X}_D, g) + G_{2,p}(\mathbf{U})),$$

with $G_{1,p}$ and $G_{2,p}$ as defined in Equations (23) and (24), respectively. Maximizing the right-hand side over all $\mathbf{U} \in \mathcal{K}_p$, measurable functions $g : \mathbb{R}^{n_r} \rightarrow \mathbb{R}^{n_t}$, and $p > 0$ provides the bound. \square

Interestingly, the bound of Theorem 13 holds for arbitrary channels and is therefore not restricted to MIMO channels. The interested reader is referred to [11] for details.

We conclude the section by providing a lower bound that is based on Jensen's inequality and holds for arbitrary inputs.

Theorem 14. (Jensen's Inequality Lower Bound) *For any channel input space \mathcal{X} and any fixed channel matrix \mathbf{H} , we have*

$$C(\mathcal{X}, \mathbf{H}) \geq \underline{C}_{\text{Jensen}}(\mathcal{X}, \mathbf{H}) := \max_{\mathbf{F}_{\mathbf{X}}: \mathbf{X} \in \mathcal{X}} \log_2^+ \left(\left(\frac{2}{e} \right)^{\frac{n_r}{2}} \psi^{-1}(\mathbf{X}, \mathbf{H}) \right) \quad (26)$$

with

$$\psi(\mathbf{X}, \mathbf{H}) := \mathbb{E} \left[\exp \left(- \frac{\|\mathbf{H}(\mathbf{X} - \mathbf{X}')\|^2}{4} \right) \right] = \mathbb{E} \left[\left| \phi_{\mathbf{X}} \left(\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \right|^2 \right],$$

where \mathbf{X}' is an independent copy of \mathbf{X} and $\phi_{\mathbf{X}}$ denotes the characteristic function of \mathbf{X} .

Proof. To show the lower bound, we follow an approach of Dytso et al. [34]. Note that by Jensen's inequality

$$h(\mathbf{Y}) = -\mathbb{E}[\log_2 f_{\mathbf{Y}}(\mathbf{Y})] \geq -\log_2 \mathbb{E}[f_{\mathbf{Y}}(\mathbf{Y})] = -\log_2 \int_{\mathbb{R}^{n_r}} f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}. \quad (27)$$

Now, evaluating the integral in Equation (27) results in

$$\int_{\mathbb{R}^{n_r}} f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = \frac{1}{(2\pi)^{n_r}} \int_{\mathbb{R}^{n_r}} \mathbb{E} \left[e^{-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{X}\|^2}{2}} \right] \mathbb{E} \left[e^{-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{X}'\|^2}{2}} \right] d\mathbf{y}$$

$$\begin{aligned}
&\stackrel{(a)}{=} \frac{1}{(2\pi)^{n_r}} \mathbb{E} \left[\int_{\mathbb{R}^{n_r}} e^{-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{X}\|^2 + \|\mathbf{y}-\mathbf{H}\mathbf{X}'\|^2}{2}} d\mathbf{y} \right] \\
&\stackrel{(b)}{=} \frac{1}{(2\pi)^{n_r}} \mathbb{E} \left[e^{-\frac{\|\mathbf{H}\mathbf{X}-\mathbf{H}\mathbf{X}'\|^2}{4}} \int_{\mathbb{R}^{n_r}} e^{-\|\mathbf{y}-\frac{\mathbf{H}(\mathbf{X}+\mathbf{X}')}{2}\|^2} d\mathbf{y} \right] \\
&\stackrel{(c)}{=} \frac{1}{2^{n_r} \pi^{\frac{n_r}{2}}} \mathbb{E} \left[e^{-\frac{\|\mathbf{H}(\mathbf{X}-\mathbf{X}')\|^2}{4}} \right], \tag{28}
\end{aligned}$$

where (a) follows from the independence of \mathbf{X} and \mathbf{X}' and Tonelli's Theorem ([35] Chapter 5.9); (b) from completing a square; and (c) from the fact that $\int_{\mathbb{R}^{n_r}} e^{-\|\mathbf{y}-\frac{\mathbf{H}(\mathbf{X}-\mathbf{X}')}{2}\|^2} d\mathbf{y} = \int_{\mathbb{R}^{n_r}} e^{-\|\mathbf{y}\|^2} d\mathbf{y} = \pi^{\frac{n_r}{2}}$. Combining Equation (27) with Equation (28) and subtracting $h(\mathbf{Z}) = \frac{n_r}{2} \log_2(2\pi e)$ completes the proof of the first version of the bound.

To show the second version, observe that

$$\begin{aligned}
\mathbb{E} \left[e^{-\frac{\|\mathbf{H}(\mathbf{X}-\mathbf{X}')\|^2}{4}} \right] &\stackrel{(d)}{=} \mathbb{E} \left[\phi_{\frac{\mathbf{H}(\mathbf{X}-\mathbf{X}')}{\sqrt{2}}}(\mathbf{Z}) \right] \\
&\stackrel{(e)}{=} \mathbb{E} \left[\phi_{\frac{\mathbf{H}\mathbf{X}}{\sqrt{2}}}(\mathbf{Z}) \phi_{-\frac{\mathbf{H}\mathbf{X}'}{\sqrt{2}}}(\mathbf{Z}) \right] \\
&\stackrel{(f)}{=} \mathbb{E} \left[\phi_{\mathbf{X}} \left(\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \phi_{\mathbf{X}'} \left(-\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \right] \\
&\stackrel{(g)}{=} \mathbb{E} \left[\phi_{\mathbf{X}} \left(\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \phi_{\mathbf{X}} \left(-\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \right] \\
&\stackrel{(h)}{=} \mathbb{E} \left[\phi_{\mathbf{X}} \left(\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \phi_{\mathbf{X}}^* \left(\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \right] \\
&= \mathbb{E} \left[\left| \phi_{\mathbf{X}} \left(\frac{\mathbf{H}^T \mathbf{Z}}{\sqrt{2}} \right) \right|^2 \right],
\end{aligned}$$

where (d) follows from Parseval's identity ([35] Chapter 9.5) by noting that $\exp(-\|\cdot\|^2/2)$ is a characteristic function of \mathbf{Z} and $\phi_{\frac{\mathbf{H}(\mathbf{X}-\mathbf{X}')}{\sqrt{2}}}(\cdot)$ is a characteristic function of $\frac{\mathbf{H}(\mathbf{X}-\mathbf{X}')}{\sqrt{2}}$; (e) from using the property that the characteristic function of a sum of random vectors is equal to the product of its characteristic functions; (f) from using the fact that a characteristic function is a linear transformation; (g) from using that \mathbf{X} and \mathbf{X}' have the same characteristic function; and (h) from the fact that the characteristic function is Hermitian. This completes the proof. \square

Remark 7. As is evident from our examples in the following sections, in many cases, the Jensen's inequality lower bound of Theorem 14 performs remarkably well. The bound, however, is also useful for MIMO channels that are subject to an average power constraint. For example, evaluating Equation (26) with $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\mathbf{X}})$ results in

$$I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) \geq \frac{1}{2} \log_2^+ \left(\left(\frac{2}{e} \right)^{\min(n_r, n_t)} \det(\mathbf{I}_{n_r} + \mathbf{H}\mathbf{K}_{\mathbf{X}}\mathbf{H}^T) \right).$$

Note that this bound is within $\frac{\min(n_r, n_t)}{2} \log_2(\frac{2}{e})$ bits of the capacity of the power-constrained channel.

In Section 3, we discuss that the distributions that maximize mutual information in n_t -dimensions are typically singular, which means that they are concentrated on a set of Lebesgue measure zero. Singular distributions generally do not have a probability density, whereas the characteristic function always exists. This is why the version of Jensen's inequality lower bound in Theorem 14 that is based on the characteristic function of the channel input is especially useful for amplitude-constrained MIMO channels.

5. Invertible Channel Matrices

Consider the symmetric case of $n_t = n_r = n$ antennas with $\mathbf{H} \in \mathbb{R}^{n \times n}$ being invertible. In this section, we evaluate some of the lower and upper bounds proposed in the previous section for the special case of \mathbf{H} being also diagonal and then characterize the gap to the capacity for arbitrary invertible channel matrices.

5.1. Diagonal Channel Matrices

Suppose the channel inputs are subject to per-antenna or an n -dimensional amplitude constraint. Then, the duality upper bound $\bar{C}_{\text{Dual},2}(\mathcal{X}, \mathbf{H})$ of Theorem 10 takes on the following form.

Theorem 15. (Upper Bounds) Let $\mathbf{H} = \text{diag}(h_{11}, \dots, h_{nn}) \in \mathbb{R}^{n \times n}$ be fixed. If $\mathcal{X} = \text{Box}(\mathbf{a})$ for some $\mathbf{a} = (A_1, \dots, A_n) \in \mathbb{R}_+^n$, then

$$\bar{C}_{\text{Dual},2}(\text{Box}(\mathbf{a}), \mathbf{H}) = \sum_{i=1}^n \log_2 \left(1 + \frac{2|h_{ii}|A_i}{\sqrt{2\pi e}} \right). \quad (29)$$

Moreover, if $\mathcal{X} = \mathcal{B}_0(A)$ for some $A \in \mathbb{R}_+$, then

$$\bar{C}_{\text{Dual},2}(\mathcal{B}_0(A), \mathbf{H}) = \sum_{i=1}^n \log_2 \left(1 + \frac{2|h_{ii}|A}{\sqrt{n}\sqrt{2\pi e}} \right). \quad (30)$$

Proof. The bound in Equation (29) immediately follows from Theorem 10 by observing that $\text{Box}(\mathbf{H}\text{Box}(\mathbf{a})) = \text{Box}(\mathbf{H}\mathbf{a})$. The bound in Equation (30) follows from Theorem 10 by the fact that

$$\text{Box}(\mathbf{H}\mathcal{B}_0(A)) \subset \text{Box}(\mathbf{H}\text{Box}(\mathcal{B}_0(A))) = \text{Box}(\mathbf{h}),$$

where $\mathbf{h} := \frac{A}{\sqrt{n}}(|h_{11}|, \dots, |h_{nn}|)$. This concludes the proof. \square

For an arbitrary channel input space \mathcal{X} , the EPI lower bound of Theorem 12 and Jensen's inequality lower bound of Theorem 14 take on the following form.

Theorem 16. (Lower Bounds) Let $\mathbf{H} = \text{diag}(h_{11}, \dots, h_{nn}) \in \mathbb{R}^{n \times n}$ be fixed and \mathcal{X} arbitrary. Then,

$$\underline{C}_{\text{Jensen}}(\mathcal{X}, \mathbf{H}) = \log_2^+ \left(\left(\frac{2}{e} \right)^{\frac{n}{2}} \frac{1}{\psi(\mathbf{H}, \mathbf{b}^*)} \right) \quad (31)$$

with

$$\psi(\mathbf{H}, \mathbf{b}^*) := \min_{\mathbf{b} \in \mathcal{X}} \prod_{i=1}^n \varphi(|h_{ii}|B_i),$$

where $\mathbf{b} := (B_1, \dots, B_n)$ and $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$,

$$\varphi(x) := \frac{1}{x^2} \left(e^{-x^2} - 1 + \sqrt{\pi}x(1 - 2Q(\sqrt{2}x)) \right). \quad (32)$$

Moreover,

$$\underline{C}_{\text{EPI}}(\mathcal{X}, \mathbf{H}) = \frac{n}{2} \log_2 \left(1 + \text{Vol}(\mathcal{X})^{\frac{2}{n}} \frac{|\prod_{i=1}^n h_{ii}|^{\frac{2}{n}}}{2\pi e} \right). \quad (33)$$

Proof. For some given values $B_i \in \mathbb{R}_+$, $i = 1, \dots, n$, let the i th component of $\mathbf{X} = (X_1, \dots, X_n)$ be independent and uniformly distributed over the interval $[-B_i, B_i]$. Thus, the expected value appearing in the bound of Theorem 14 can be written as

$$\mathbb{E} \left[e^{-\frac{\|\mathbf{H}(\mathbf{X}-\mathbf{X}')\|^2}{4}} \right] = \mathbb{E} \left[e^{-\frac{\sum_{i=1}^n h_{ii}^2 (X_i - X'_i)^2}{4}} \right] = \mathbb{E} \left[\prod_{i=1}^n e^{-\frac{h_{ii}^2 (X_i - X'_i)^2}{4}} \right] = \prod_{i=1}^n \mathbb{E} \left[e^{-\frac{h_{ii}^2 (X_i - X'_i)^2}{4}} \right]. \quad (34)$$

Now, if \mathbf{X}' is an independent copy of \mathbf{X} , it can be shown that the expected value at the right-hand side of Equation (34) is of the explicit form

$$\mathbb{E} \left[e^{-\frac{h_{ii}^2 (x_i - x'_i)^2}{4}} \right] = \varphi(|h_{ii}|B_i)$$

with φ as defined in Equation (32). Finally, optimizing over all $\mathbf{b} = (B_1, \dots, B_n) \in \mathcal{X}$ results in the bound (31). The bound in Equation (33) follows by inserting $|\det(\mathbf{H})| = |\prod_{i=1}^n h_{ii}|$ into Equation (21), which concludes the proof. \square

In Figure 2, the upper bounds of Theorems 9 and 15 and the lower bounds of Theorem 16 are depicted for a diagonal 2×2 MIMO channel with per-antenna amplitude constraints. It turns out that the moment upper bound and the EPI lower bound perform well in the small amplitude regime while the duality upper bound and Jensen's inequality lower bound perform well in the high amplitude regime. Interestingly, for this specific example, the duality upper bound and Jensen's lower bound are asymptotically tight.

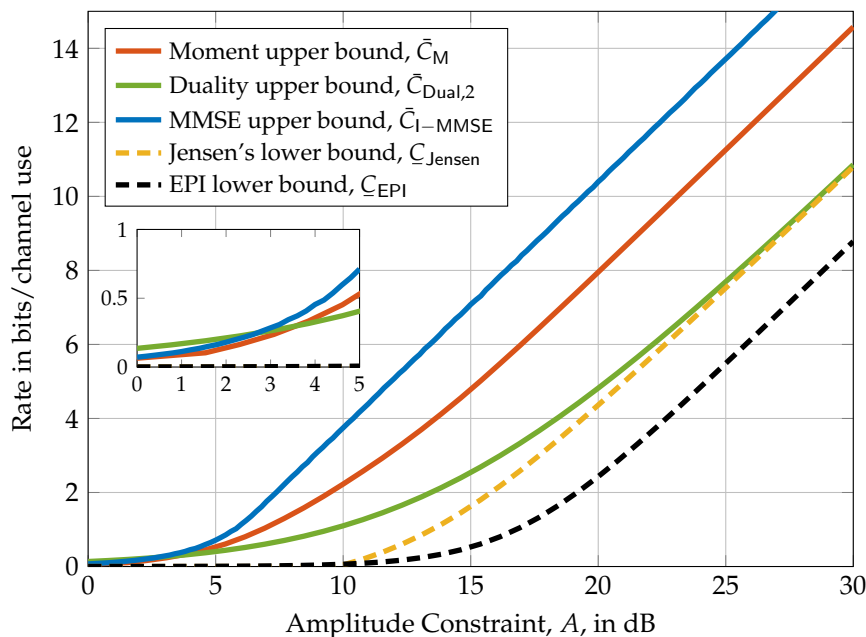


Figure 2. Comparison of the upper and lower bounds of Theorems 9, 11, 15, and 16 evaluated for a 2×2 MIMO system with per-antenna amplitude constraints $A_1 = A_2 = A$ (i.e., $\mathbf{a} = (A, A)$) and channel matrix $\mathbf{H} = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.1 \end{pmatrix}$. The nested figure represents a zoom into the region $0 \text{ dB} \leq A \leq 5 \text{ dB}$ to visualize the differences between the bounds at small amplitude constraints.

5.2. Gap to the Capacity

Our first result provides an upper bound to the gap between the capacity in Equation (2) and the lower bound in Equation (21).

Theorem 17. Let $\mathbf{H} \in \mathbb{R}^{n \times n}$ be of full rank and

$$\rho(\mathcal{X}, \mathbf{H}) := \frac{\text{Vol}(\mathcal{B}_0(r_{\max}(\mathbf{H}\mathcal{X})))}{\text{Vol}(\mathbf{H}\mathcal{X})}.$$

Then,

$$C(\mathcal{X}, \mathbf{H}) - \underline{C}_{\text{EPI}}(\mathcal{X}, \mathbf{H}) \leq \frac{n}{2} \log_2 \left((\pi n)^{\frac{1}{n}} \rho(\mathcal{X}, \mathbf{H})^{\frac{2}{n}} \right).$$

Proof. For notational convenience, let the volume of an n -dimensional ball of radius $r > 0$ be denoted as

$$V_n(r) := \text{Vol}(\mathcal{B}_0(r)) = V_n(1)r^n = \frac{\pi^{\frac{n}{2}} r^n}{\Gamma(\frac{n}{2} + 1)}.$$

Now, observe that, by choosing $p = 2$, the upper bound of Theorem 9 can further be upper bounded as

$$\begin{aligned} \bar{C}_M(\mathcal{X}, \mathbf{H}) &\leq n \log_2 \left(\frac{k_{n,2}}{(2\pi e)^{\frac{1}{2}}} n^{\frac{1}{2}} \|\tilde{\mathbf{x}} + \mathbf{Z}\|_2 \right) \\ &\stackrel{(a)}{=} \frac{n}{2} \log_2 \left(\frac{1}{n} \mathbb{E}[\|\tilde{\mathbf{x}} + \mathbf{Z}\|^2] \right) \\ &\stackrel{(b)}{=} \frac{n}{2} \log_2 \left(1 + \frac{1}{n} \|\tilde{\mathbf{x}}\|^2 \right), \end{aligned}$$

where (a) follows since $k_{n,2} = \sqrt{\frac{2\pi e}{n}}$; and (b) since $\mathbb{E}[\|\mathbf{Z}\|^2] = n$. Therefore, the gap between Equation (21) and the moment upper bound of Theorem 9 can be upper bounded as follows:

$$\begin{aligned} \bar{C}_M(\mathcal{X}, \mathbf{H}) - \underline{C}_{\text{EPI}}(\mathcal{X}, \mathbf{H}) &= \frac{n}{2} \log_2 \left(\frac{1 + \frac{1}{n} \|\tilde{\mathbf{x}}\|^2}{1 + \frac{\text{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}}}{2\pi e}} \right) \\ &\stackrel{(a)}{=} \frac{n}{2} \log_2 \left(\frac{1 + \frac{1}{n} \left(\frac{V_n(\|\tilde{\mathbf{x}}\|)}{V_n(1)} \right)^{\frac{2}{n}}}{1 + \frac{\text{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}}}{2\pi e}} \right) \\ &= \frac{n}{2} \log_2 \left(\frac{1 + \frac{1}{n} \left(\frac{\rho(\mathcal{X}, \mathbf{H}) \text{Vol}(\mathbf{H}\mathcal{X})}{V_n(1)} \right)^{\frac{2}{n}}}{1 + \frac{\text{Vol}(\mathbf{H}\mathcal{X})^{\frac{2}{n}}}{2\pi e}} \right) \\ &\stackrel{(b)}{\leq} \frac{n}{2} \log_2 \left(\frac{1}{n} 2\pi e \left(\frac{\rho(\mathcal{X}, \mathbf{H})}{V_n(1)} \right)^{\frac{2}{n}} \right) \\ &\stackrel{(c)}{\leq} \frac{n}{2} \log_2 \left((\pi n)^{\frac{1}{n}} \rho(\mathcal{X}, \mathbf{H})^{\frac{2}{n}} \right). \end{aligned}$$

where (a) is due to the fact that $\|\tilde{\mathbf{x}}\|$ is the radius of an n -dimensional ball; (b) follows from the inequality $\frac{1+c\chi}{1+x} \leq c$ for $c \geq 1$ and $x \in \mathbb{R}_+$; and (c) follows from using Stirling's approximation to obtain $\left(\frac{1}{V_n(1)} \right)^{\frac{2}{n}} \leq \frac{1}{2e\pi^{1-\frac{1}{n}}} n^{1+\frac{1}{n}}$. \square

The term $\rho(\mathcal{X}, \mathbf{H})$ is referred to as the *packing efficiency* of the set $\mathbf{H}\mathcal{X}$. In the following proposition, we present the packing efficiencies for important special cases.

Proposition 1. (Packing Efficiencies) Let $\mathbf{H} \in \mathbb{R}^{n \times n}$ be of full rank, $A \in \mathbb{R}_+$, and $\mathbf{a} := (A_1, \dots, A_n) \in \mathbb{R}_+^n$. Then,

$$\rho(\mathcal{B}_0(A), \mathbf{I}_n) = 1, \quad (35)$$

$$\rho(\mathcal{B}_0(A), \mathbf{H}) = \frac{\|\mathbf{H}\|^n}{|\det(\mathbf{H})|}, \quad (36)$$

$$\rho(\text{Box}(\mathbf{a}), \mathbf{I}_n) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} \frac{\|\mathbf{a}\|^n}{\prod_{i=1}^n A_i}, \quad (37)$$

$$\rho(\text{Box}(\mathbf{a}), \mathbf{H}) \leq \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} \frac{\|\mathbf{H}\|^n \|\mathbf{a}\|^n}{|\det(\mathbf{H})| \prod_{i=1}^n A_i}. \quad (38)$$

Proof. The packing efficiency in Equation (35) follows immediately. Note that

$$r_{\max}(\mathbf{H}\mathcal{B}_0(A)) = \max_{\mathbf{x} \in \mathcal{B}_0(A)} \|\mathbf{H}\mathbf{x}\| = \|\mathbf{H}\|A.$$

Thus, as \mathbf{H} is assumed to be invertible, we have $\text{Vol}(\mathbf{H}\mathcal{B}_0(A)) = |\det(\mathbf{H})|\text{Vol}(\mathcal{B}_0(A))$, which results in Equation (36). To show Equation (37), observe that

$$\text{Vol}(\mathcal{B}_0(r_{\max}(\mathbf{I}_n \text{Box}(\mathbf{a})))) = \text{Vol}(\mathcal{B}_0(\|\mathbf{a}\|)) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} \|\mathbf{a}\|^n.$$

The proof of Equation (37) is concluded by observing that $\text{Vol}(\mathbf{I}_n \text{Box}(\mathbf{a})) = \prod_{i=1}^n A_i$. Finally, observe that $\text{Box}(\mathbf{a}) \subset \mathcal{B}_0(\|\mathbf{a}\|)$ implies $r_{\max}(\mathbf{H}\text{Box}(\mathbf{a})) \leq r_{\max}(\mathbf{H}\mathcal{B}_0(\|\mathbf{a}\|))$ so that

$$\rho(\mathbf{H}, \text{Box}(\mathbf{a})) \leq \frac{\text{Vol}(\mathcal{B}_0(\|\mathbf{H}\|\|\mathbf{a}\|))}{\text{Vol}(\mathbf{H}\text{Box}(\mathbf{a}))} = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} \frac{\|\mathbf{H}\|^n \|\mathbf{a}\|^n}{|\det(\mathbf{H})| \prod_{i=1}^n A_i},$$

which is the bound in Equation (38). \square

We conclude this section by characterizing the gap to the capacity when \mathbf{H} is diagonal and the channel input space is the Cartesian product of n PAM constellations. In this context, $\text{PAM}(N, A)$ refers to the set of $N \in \mathbb{N}$ equidistant PAM-constellation points with amplitude constraint $A \in \mathbb{R}_+$ (see Figure 3 for an illustration), whereas $X \sim \text{PAM}(N, A)$ means that X is uniformly distributed over $\text{PAM}(N, A)$ [11].

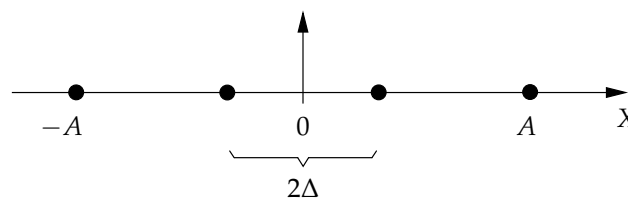


Figure 3. Example of a pulse-amplitude modulation constellation with $N = 4$ points and amplitude constraint A (i.e., $\text{PAM}(4, A)$), where $\Delta := A/(N - 1)$ denotes half the Euclidean distance between two adjacent constellation points. In the case N is odd, 0 is a constellation point.

Theorem 18. Let $\mathbf{H} = \text{diag}(h_{11}, \dots, h_{nn}) \in \mathbb{R}^{n \times n}$ be fixed and $\mathbf{X} = (X_1, \dots, X_n)$. Then, if $X_i \sim \text{PAM}(N_i, A_i)$, $i = 1, \dots, n$, for some given $\mathbf{a} = (A_1, \dots, A_n) \in \mathbb{R}_+^n$, it holds that

$$\bar{\mathbb{C}}_{\text{Dual},2}(\text{Box}(\mathbf{a}), \mathbf{H}) - \mathbb{C}_{\text{OW}}(\text{Box}(\mathbf{a}), \mathbf{H}) \leq c \cdot n \text{ bits}, \quad (39)$$

where $N_i := \left\lfloor 1 + \frac{2A_i|h_{ii}|}{\sqrt{2\pi e}} \right\rfloor$ and

$$c := 1 + \frac{1}{2} \log_2 \left(\frac{\pi e}{6} \right) + \frac{1}{2} \log_2 \left(1 + \frac{6}{\pi e} \right) \approx 1.64.$$

Moreover, if $X_i \sim \text{PAM}(N_i, A)$, $i = 1, \dots, n$, for some given $A \in \mathbb{R}_+$, it holds that

$$\bar{C}_{\text{Dual},2}(\mathcal{B}_0(A), \mathbf{H}) - \underline{C}_{\text{OW}}(\mathcal{B}_0(A), \mathbf{H}) \leq c \cdot n \text{ bits}, \quad (40)$$

where $N_i := \left\lfloor 1 + \frac{2A|h_{ii}|}{\sqrt{n}\sqrt{2\pi e}} \right\rfloor$.

Proof. Since the channel matrix is diagonal, letting the channel input \mathbf{X} be such that its elements X_i , $i = 1, \dots, n$, are independent, we have that

$$I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z}) = \sum_{i=1}^n I(X_i; h_{ii}X_i + Z_i).$$

Let $X_i \sim \text{PAM}(N_i, A_i)$ with $N_i := \left\lfloor 1 + \frac{2A_i|h_{ii}|}{\sqrt{2\pi e}} \right\rfloor$ and observe that half the Euclidean distance between any pair of adjacent points in $\text{PAM}(N_i, A_i)$ is equal to $\Delta_i := A_i / (N_i - 1)$ (see Figure 3), $i = 1, \dots, n$. To lower bound the mutual information $I(X_i; h_{ii}X_i + Z_i)$, we use the bound of Theorem 13 for $p = 2$ and $n_t = 1$. Thus, for some continuous random variable U that is uniformly distributed over the interval $[-\Delta_i, \Delta_i)$ and independent of X_i , we have that

$$I(X_i; h_{ii}X_i + Z_i) \geq H(X_i) - \frac{1}{2} \log_2 \left(\frac{\pi e}{6} \right) - \frac{1}{2} \log_2 \left(\frac{\mathbb{E}[(U + X_i - g(Y_i))^2]}{\mathbb{E}[U^2]} \right). \quad (41)$$

Now, note that the entropy term in Equation (41) can be lower bounded as

$$H(X_i) = \log_2 \left(\left\lfloor 1 + \frac{2A_i|h_{ii}|}{\sqrt{2\pi e}} \right\rfloor \right) \geq \log_2 \left(1 + \frac{2A_i|h_{ii}|}{\sqrt{2\pi e}} \right) + \log_2(2), \quad (42)$$

where we have used that $\lfloor x \rfloor \geq \frac{x}{2}$ for every $x \geq 1$. On the other hand, the last term in Equation (41) can be upper bounded by upper bounding its argument as follows:

$$\begin{aligned} \frac{\mathbb{E}[(U + X_i - g(Y_i))^2]}{\mathbb{E}[U^2]} &\stackrel{a)}{=} 1 + \frac{3\mathbb{E}[(X_i - g(Y_i))^2]}{\Delta_i^2} \\ &\stackrel{b)}{\leq} 1 + \frac{3\mathbb{E}[Z_i^2](N_i - 1)^2}{A_i^2|h_{ii}|^2} \\ &= 1 + \frac{3(N_i - 1)^2}{A_i^2|h_{ii}|^2} \\ &\stackrel{c)}{\leq} 1 + \frac{3 \left(\frac{2A_i|h_{ii}|}{\sqrt{2\pi e}} \right)^2}{A_i^2|h_{ii}|^2} \\ &= 1 + \frac{6}{\pi e}. \end{aligned} \quad (43)$$

where (a) follows from using that X_i and U are independent and $\mathbb{E}[U^2] = \frac{\Delta_i^2}{3}$; (b) from using the estimator $g(Y_i) = \frac{1}{h_{ii}} Y_i$; and (c) from $N_i = \left\lfloor 1 + \frac{2A_i|h_{ii}|}{\sqrt{2\pi e}} \right\rfloor \leq 1 + \frac{2A_i|h_{ii}|}{\sqrt{2\pi e}}$. Combining Equations (41), (42), and (43) results in the gap in (39).

The proof of the capacity gap in Equation (40) follows along similar lines, which concludes the proof. \square

We are also able to determine the gap to the capacity for a general invertible channel matrix.

Theorem 19. For any \mathcal{X} and any invertible \mathbf{H}

$$C(\mathcal{X}, \mathbf{H}) - \underline{C}_{\text{OW}}(\mathcal{X}, \mathbf{H}) \leq \log_2(\pi n) + \frac{n}{2} \log_2 \left(1 + 4n(4 + 4n) \left(\frac{\|\mathbf{H}^{-1}\|^2 r_{\max}^2(\mathbf{H}\mathcal{X})}{r_{\min}^2(\mathcal{X})} + \frac{n\|\mathbf{H}^{-1}\|^2}{r_{\min}^2(\mathcal{X})} \right) \right).$$

Proof. Let \mathbf{X} be uniformly distributed over a set constructed from an n -dimensional cubic lattice with the number of points equal to $N = \lfloor \|\tilde{\mathbf{x}} + \mathbf{Z}\|_2^n \rfloor$, where $\tilde{\mathbf{x}} \in \mathbf{H}\mathcal{X}$ is chosen such that $\|\tilde{\mathbf{x}}\| = r_{\max}(\mathbf{H}\mathcal{X})$, and scaled such that it is contained in the input space \mathcal{X} . Note that the minimum distance between point in \mathbf{X} are given by

$$d_{\min}(\text{supp}(\mathbf{X})) := \frac{r_{\min}(\mathcal{X})}{N^{\frac{1}{n}}}.$$

Now, we compute the difference between the moment upper bound of Theorem 9 and the Ozarow–Wyner lower bound of Theorem 13:

$$\begin{aligned} \bar{C}_{\text{M}}(\mathcal{X}, \mathbf{H}) - \underline{C}_{\text{OW}}(\mathcal{X}, \mathbf{H}) &\stackrel{(a)}{\leq} \log_2(\|\tilde{\mathbf{x}} + \mathbf{Z}\|_2) - H(\mathbf{X}) + \text{gap} \\ &= n \log_2(\|\tilde{\mathbf{x}} + \mathbf{Z}\|_2) - \log_2(\lfloor \|\tilde{\mathbf{x}} + \mathbf{Z}\|_2^n \rfloor) + \text{gap} \\ &\stackrel{(b)}{\leq} \log_2(2) + \text{gap}, \end{aligned} \quad (44)$$

where (a) follows from Theorem 9 by choosing $p = 2$; and (b) by using the bound $\lfloor x \rfloor \geq \frac{x}{2}$ for $x > 1$. The next step in the proof consists in bounding the gap term, which requires to upper bound the terms in Equations (23) and (24) individually. Towards this end, choose $p = 2$ and let \mathbf{U} be a random vector that is uniformly distributed over a ball of radius $d_{\min}(\mathbf{X})$. Thus, for (23) it follows

$$\begin{aligned} G_{1,2}(\mathbf{U}, \mathbf{X}, g) &= n \log_2 \left(\frac{\|\mathbf{U} + \mathbf{X} - g(\mathbf{Y})\|_2}{\|\mathbf{U}\|_2} \right) \\ &\stackrel{(a)}{=} n \log_2 \left(\frac{\|\mathbf{U} - \mathbf{H}^{-1}\mathbf{Z}\|_2}{\|\mathbf{U}\|_2} \right) \\ &= \frac{n}{2} \log_2 \left(1 + \frac{\|\mathbf{H}^{-1}\mathbf{Z}\|_2^2}{\|\mathbf{U}\|_2^2} \right) \\ &\stackrel{(b)}{=} \frac{n}{2} \log_2 \left(1 + \frac{4(4 + 4n)\|\mathbf{H}^{-1}\mathbf{Z}\|_2^2}{d_{\min}^2(\text{supp}(\mathbf{X}))} \right) \\ &\stackrel{(c)}{\leq} \frac{n}{2} \log_2 \left(1 + \frac{(4 + 4n)\|\mathbf{H}^{-1}\mathbf{Z}\|_2^2 \|\tilde{\mathbf{x}} + \mathbf{Z}\|_2^2}{r_{\min}^2(\mathcal{X})} \right) \\ &\stackrel{(d)}{=} \frac{n}{2} \log_2 \left(1 + \frac{(4 + 4n)\|\mathbf{H}^{-1}\mathbf{Z}\|_2^2 (r_{\max}^2(\mathbf{H}\mathcal{X}) + n)}{r_{\min}^2(\mathcal{X})} \right) \\ &\stackrel{(e)}{\leq} \frac{n}{2} \log_2 \left(1 + \frac{4(4 + 4n)\|\mathbf{H}^{-1}\|^2 \|\mathbf{Z}\|_2^2 (r_{\max}^2(\mathbf{H}\mathcal{X}) + n)}{r_{\min}^2(\mathcal{X})} \right) \\ &= \frac{n}{2} \log_2 \left(1 + \frac{4(4 + 4n)\|\mathbf{H}^{-1}\|^2 n (r_{\max}^2(\mathbf{H}\mathcal{X}) + n)}{r_{\min}^2(\mathcal{X})} \right). \end{aligned}$$

where (a) follows by choosing $g(\mathbf{Y}) = \mathbf{H}^{-1}\mathbf{Y}$; (b) by using $\|\mathbf{U}\|_2^2 = \frac{r^2}{4+2n}$, where $r = \frac{d_{\min}(\text{supp}(\mathbf{X}))}{2}$ is the radius of an n -dimensional ball; (c) from dropping the floor function in the expression for the minimum distance, i.e.,

$$d_{\min}^{-1}(\text{supp}(\mathbf{X})) = \frac{N^{\frac{1}{n}}}{r_{\min}(\mathcal{X})} = \frac{\lfloor \|\tilde{\mathbf{x}} + \mathbf{Z}\|_2^n \rfloor^{\frac{1}{n}}}{r_{\min}(\mathcal{X})} \leq \frac{\|\tilde{\mathbf{x}} + \mathbf{Z}\|_2}{r_{\min}(\mathcal{X})};$$

(d) follows by expanding $\|\tilde{\mathbf{x}} + \mathbf{Z}\|_2^2$ using that $\|\tilde{\mathbf{x}}\| = r_{\max}(\mathbf{H}\mathcal{X})$; and (e) from using the bound $\|\mathbf{H}^{-1}\mathbf{Z}\|_2 \leq \|\mathbf{H}^{-1}\| \|\mathbf{Z}\|_2$.

On the other hand, the term $G_{2,p}(\mathbf{U})$ can be bounded from above as follows ([36] Appendix L):

$$G_{2,p}(\mathbf{U}) = n \log_2 \left(\frac{k_{n,p} n^{\frac{1}{p}} \|\mathbf{U}\|_p}{2^{\frac{1}{n}h(\mathbf{U})}} \right) \leq n \log_2 \left((\pi n)^{\frac{1}{n}} \right).$$

Combining these two bounds with the one in (44) provides the result. \square

6. Arbitrary Channel Matrices

For an arbitrary MIMO channel with an average power constraint, it is well known that the capacity is achieved by a singular value decomposition (SVD) of the channel matrix (i.e., $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$) along with considering the equivalent channel model

$$\tilde{\mathbf{Y}} = \mathbf{\Lambda}\tilde{\mathbf{X}} + \tilde{\mathbf{Z}},$$

where $\tilde{\mathbf{Y}} := \mathbf{U}^T\mathbf{Y}$, $\tilde{\mathbf{X}} := \mathbf{V}^T\mathbf{X}$, and $\tilde{\mathbf{Z}} := \mathbf{U}^T\mathbf{Z}$, respectively.

To provide lower bounds for channels with amplitude constraints and SVD precoding, we need the following lemma.

Lemma 3. For any given orthogonal matrix $\mathbf{V} \in \mathbb{R}^{n_t \times n_t}$ and constraint vector $\mathbf{a} = (A_1, \dots, A_{n_t}) \in \mathbb{R}_+^{n_t}$, there exists a distribution $F_{\tilde{\mathbf{X}}}$ of $\tilde{\mathbf{X}}$ such that $\tilde{\mathbf{X}} = \mathbf{V}^T\mathbf{X}$ is uniformly distributed over $\text{Box}(\mathbf{a})$. Moreover, the components $\tilde{X}_1, \dots, \tilde{X}_{n_t}$ of $\tilde{\mathbf{X}}$ are mutually independent with \tilde{X}_i uniformly distributed over $[-A_i, A_i]$, $i = 1, \dots, n_t$.

Proof. Suppose that $\tilde{\mathbf{X}}$ is uniformly distributed over $\text{Box}(\mathbf{a})$; that is, the density of $\tilde{\mathbf{X}}$ is of the form

$$f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = \frac{1}{\text{Vol}(\text{Box}(\mathbf{a}))}, \quad \tilde{\mathbf{x}} \in \text{Box}(\mathbf{a}).$$

Since \mathbf{V} is orthogonal, we have $\mathbf{V}\tilde{\mathbf{X}} = \mathbf{X}$ and by the change of variable Theorem for $\mathbf{x} \in \mathbf{V}\text{Box}(\mathbf{a})$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|\det(\mathbf{V})|} f_{\tilde{\mathbf{X}}}(\mathbf{V}^T\mathbf{x}) = \frac{1}{|\det(\mathbf{V})| \text{Vol}(\text{Box}(\mathbf{a}))} = \frac{1}{\text{Vol}(\text{Box}(\mathbf{a}))}.$$

Therefore, such a distribution of \mathbf{X} exists. \square

Theorem 20. (Lower Bounds with SVD Precoding) Let $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ be fixed, $n_{\min} := \min(n_r, n_t)$, and $\mathcal{X} = \text{Box}(\mathbf{a})$ for some $\mathbf{a} = (A_1, \dots, A_{n_t}) \in \mathbb{R}_+^{n_t}$. Furthermore, let σ_i , $i = 1, \dots, n_{\min}$, be the i th singular value of \mathbf{H} . Then,

$$\underline{C}_{\text{Jensen}}(\text{Box}(\mathbf{a}), \mathbf{H}) = \log_2^+ \left(\left(\frac{2}{e} \right)^{\frac{n_{\min}}{2}} \frac{1}{\psi(\mathbf{H}, \mathbf{b}^*)} \right) \quad (45)$$

and

$$\underline{C}_{\text{EPI}}(\text{Box}(\mathbf{a}), \mathbf{H}) = \frac{n_{\min}}{2} \log_2 \left(1 + \frac{|\prod_{i=1}^{n_{\min}} A_i \sigma_i|^{\frac{2}{n_{\min}}}}{2\pi e} \right), \quad (46)$$

where

$$\psi(\mathbf{H}, \mathbf{b}^*) := \min_{\mathbf{b} \in \text{Box}(\mathbf{a})} \prod_{i=1}^{n_{\min}} \varphi(\sigma_i B_i)$$

with $\mathbf{b} := (B_1, \dots, B_{n_t})$ and φ as defined in Equation (32).

Proof. Performing the SVD, the expected value in Theorem 14 can be written as

$$\mathbb{E} \left[e^{-\frac{\|\mathbf{H}(\mathbf{x}-\mathbf{x}')\|^2}{4}} \right] = \mathbb{E} \left[e^{-\frac{\|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T(\mathbf{x}-\mathbf{x}')\|^2}{4}} \right] = \mathbb{E} \left[e^{-\frac{\|\mathbf{\Lambda}\mathbf{V}^T(\mathbf{x}-\mathbf{x}')\|^2}{4}} \right] = \mathbb{E} \left[e^{-\frac{\|\mathbf{\Lambda}(\tilde{\mathbf{x}}-\tilde{\mathbf{x}}')\|^2}{4}} \right].$$

By Lemma 3, there exists a distribution $F_{\tilde{\mathbf{X}}}$ such that the components of $\tilde{\mathbf{X}}$ are independent and uniformly distributed. Since $\mathbf{\Lambda}$ is a diagonal matrix, we can use Theorem 16 to arrive at Equation (45).

Note that by Lemma 3 there exists a distribution on \mathbf{X} such that $\tilde{\mathbf{X}}$ is uniform over $\text{Box}(\mathbf{a}) \subset \mathbb{R}^{n_t}$ and $\mathbf{\Lambda}\tilde{\mathbf{X}}$ is uniform over $\mathbf{\Lambda}\text{Box}(\mathbf{a}) \subset \mathbb{R}^{n_{\min}}$, respectively. Therefore, by the EPI lower bound given in Equation (20), we obtain

$$\begin{aligned} \underline{C}_{\text{EPI}}(\text{Box}(\mathbf{a}), \mathbf{H}) &= \frac{n_{\min}}{2} \log_2 \left(1 + \frac{2^{\frac{2}{n_{\min}} h(\mathbf{\Lambda}\tilde{\mathbf{X}})}}{2\pi e} \right) \\ &= \frac{n_{\min}}{2} \log_2 \left(1 + \frac{\text{Vol}(\mathbf{\Lambda}\text{Box}(\mathbf{a}))^{\frac{2}{n_{\min}}}}{2\pi e} \right) \\ &= \frac{n_{\min}}{2} \log_2 \left(1 + \frac{(\prod_{i=1}^{n_{\min}} A_i)^{\frac{2}{n_{\min}}} |\prod_{i=1}^{n_{\min}} \sigma_i|^{\frac{2}{n_{\min}}}}{2\pi e} \right), \end{aligned}$$

which is exactly the expression in Equation (46). This concludes the proof. \square

Remark 8. Notice that choosing the optimal \mathbf{b} for the lower bound in Equation (45) is an amplitude allocation problem, which is reminiscent of waterfilling in the average power constraint case. It would be interesting to study whether the bound in Equation (45) is connected to what is called mercury waterfilling in [37,38].

In Figure 4, the lower bounds of Theorem 20 are compared to the moment upper bound of Theorem 2 for the special case of a 3×1 MIMO channel. Similar to the example presented in Figure 2, the EPI lower bound performs well in the low amplitude regime, while Jensen's inequality lower bound performs well in the high amplitude regime.

We conclude this section by showing that for an arbitrary channel input space \mathcal{X} , in the large amplitude regime the capacity pre-log is given by $\min(n_r, n_t)$.

Theorem 21. Let \mathcal{X} be arbitrary and $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ fixed. Then,

$$\lim_{r_{\min}(\mathcal{X}) \rightarrow \infty} \frac{C(\mathcal{X}, \mathbf{H})}{\log_2 \left(1 + \frac{2r_{\min}(\mathcal{X})}{\sqrt{2\pi e}} \right)} = \min(n_r, n_t).$$

Proof. Notice that there always exists $\mathbf{a} \in \mathbb{R}_+^{n_t}$ and $c \in \mathbb{R}_+$ such that $\text{Box}(\mathbf{a}) \subseteq \mathcal{X} \subseteq c\text{Box}(\mathbf{a})$. Thus, without loss of generality, we can consider $\mathcal{X} = \text{Box}(\mathbf{a})$, $\mathbf{a} = (A, \dots, A)$, for sufficiently large $A \in \mathbb{R}_+$. To prove the result, we therefore start with enlarging the constraint set of the bound in Equation (11):

$$\begin{aligned} \text{Box}(\mathbf{H}\text{Box}(\mathbf{a})) &\subseteq \mathcal{B}_0(r_{\max}(\mathbf{H}\text{Box}(\mathbf{a}))) \\ &\subseteq \mathcal{B}_0(r_{\max}(\mathbf{H}\mathcal{B}_0(\sqrt{n_t}A))) \\ &= \mathcal{B}_0(r_{\max}(\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T\mathcal{B}_0(\sqrt{n_t}A))) \end{aligned}$$

$$\begin{aligned}
&= \mathcal{B}_0(r_{\max}(\mathbf{U}\mathbf{\Lambda}\mathcal{B}_0(\sqrt{n_t}A))) \\
&= \mathcal{B}_0(r_{\max}(\mathbf{\Lambda}\mathcal{B}_0(\sqrt{n_t}A))) \\
&\subseteq \mathcal{B}_0(r) \\
&\subseteq \text{Box}(\mathbf{a}'),
\end{aligned}$$

where $r := \sqrt{n_t}A\sqrt{\sum_{i=1}^{n_{\min}}\sigma_i^2}$ and $\mathbf{a}' := (\frac{r}{\sqrt{n_{\min}}}, \dots, \frac{r}{\sqrt{n_{\min}}}) \in \mathbb{R}_+^{n_{\min}}$. Therefore, by using the upper bound in Equation (11), it follows that

$$C(\text{Box}(\mathbf{a}), \mathbf{H}) \leq \sum_{i=1}^{n_r} \log_2 \left(1 + \frac{2A_i}{\sqrt{2\pi e}} \right) \leq n_{\min} \log_2 \left(1 + \frac{2}{\sqrt{2\pi e}} \frac{\sqrt{n_t}A\sqrt{\sum_{i=1}^{n_{\min}}\sigma_i^2}}{\sqrt{n_{\min}}} \right).$$

Moreover,

$$\lim_{A \rightarrow \infty} \frac{C(\text{Box}(\mathbf{a}), \mathbf{H})}{\log_2 \left(1 + \frac{2A}{\sqrt{2\pi e}} \right)} \leq n_{\min} \lim_{A \rightarrow \infty} \frac{\log_2 \left(1 + \frac{2}{\sqrt{2\pi e}} \frac{\sqrt{n_t}A\sqrt{\sum_{i=1}^{n_{\min}}\sigma_i^2}}{\sqrt{n_{\min}}} \right)}{\log_2 \left(1 + \frac{2A}{\sqrt{2\pi e}} \right)} = n_{\min}.$$

Next, using the EPI lower bound in Equation (46), we have that

$$\lim_{A \rightarrow \infty} \frac{\underline{C}_{\text{EPI}}(\text{Box}(\mathbf{a}), \mathbf{\Lambda})}{\log_2 \left(1 + \frac{2A}{\sqrt{2\pi e}} \right)} = n_{\min} \lim_{A \rightarrow \infty} \frac{\frac{1}{2} \log_2 \left(1 + \frac{A|\prod_{i=1}^{n_{\min}}\sigma_i|^{\frac{2}{n_{\min}}}}{2\pi e} \right)}{\log_2 \left(1 + \frac{2A}{\sqrt{2\pi e}} \right)} = n_{\min}.$$

This concludes the proof. \square

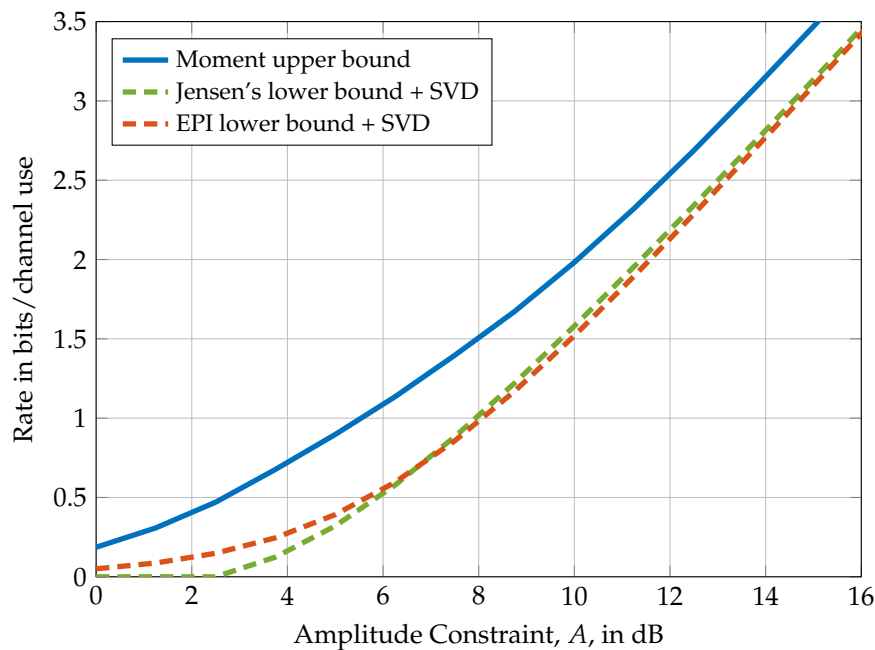


Figure 4. Comparison of the upper bound in Theorem 2 with the lower bounds of Theorem 20 for a 3×1 MIMO system with amplitude constraints $A_1 = A_2 = A_3 = A$ (i.e., $\mathbf{a} = (A, A, A)$) and channel matrix $\mathbf{h} = (0.6557, 0.0357, 0.8491)$.

7. The SISO Case

In this section, we apply the upper and lower bounds presented in the previous sections to the special case of a SISO channel that is subject to an amplitude constraint (i.e., $\mathcal{X} = [-A, A]$ for some $A \in \mathbb{R}_+$) and compare them with the state-of-the-art. More precisely, we are interested in upper and lower bounds to the capacity

$$C([-A, A], h) := \max_{F_X: X \in [-A, A]} I(X; hX + Y). \quad (47)$$

Without loss of generality, we assume $h = 1$ in all that follows.

7.1. Upper and Lower Bounds

As a starting point for our comparisons, the following Theorem summarizes bounds on the capacity (47) that are known from the literature. The bounds are all based on the duality approach that we generalize in Section 4 to the MIMO case.

Theorem 22. (Known Duality Upper Bounds) *Let $A > 0$ be arbitrary. Then, the following are valid upper bounds to the capacity of the amplitude-constrained SISO channel defined in Equation (47).*

- McKellips upper bound [7]:

$$C([-A, A], 1) \leq \bar{C}_{\text{McK}}([-A, A], 1) := \log_2 \left(1 + \frac{2A}{\sqrt{2\pi e}} \right). \quad (48)$$

- Thangaraj–Kramer–Böcherer upper bound ([8] Theorem 1):

$$C([-A, A], 1) \leq \bar{C}_{\text{TKB}}([-A, A], 1) := \begin{cases} \beta(A) \log_e \left(\sqrt{\frac{2}{\pi e}} A \right) + H_b(\beta(A)), & A^2 \leq 6.304 \text{ dB} \\ \bar{C}_{\text{McK}}([-A, A], 1), & \text{else,} \end{cases} \quad (49)$$

where $\beta(A) := \frac{1}{2} - Q(2A)$ and H_b denotes the binary entropy function.

- Rassouli–Clerckx upper bound [9]:

$$C([-A, A], 1) \leq \bar{C}_{\text{RC}}([-A, A], 1) := \bar{C}_{\text{TKB}}([-A, A], 1) + W(A), \quad (50)$$

where

$$W(A) := \frac{1}{2} \left(\log_e(\sigma^2(A)) + \frac{1}{\sigma^2(A)} - 1 \right) \left(\frac{1}{2} + Q(2A) \right) + \frac{g(2A)}{2\sigma^2(A)},$$

$$\sigma^2(A) := 1 + \frac{2g(2A)}{1 + 2Q(2A)},$$

and

$$g(x) := x^2 Q(x) - \frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Now, we apply the moment upper bound of Theorem 9 to the SISO case.

Theorem 23. (Moment Upper Bound) *Let $A > 0$ be arbitrary. Then,*

$$C([-A, A], 1) \leq \bar{C}_{\text{M}}([-A, A], 1) = \inf_{p>0} \log_2 \left(\frac{k_{1,p}}{(2\pi e)^{\frac{1}{2}}} \mathbb{E}[|A + Z|^p]^{\frac{1}{p}} \right), \quad (51)$$

where the expected value is of the explicit form

$$\mathbb{E}[|A + Z|^p] = \frac{2^{\frac{p}{2}} \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}} {}_1F_1\left(-\frac{p}{2}; \frac{1}{2}; -\frac{A^2}{2}\right)$$

with ${}_1F_1(a; b; z)$ being the confluent hypergeometric function of the first kind ([39] Chapter 13).

Proof. First, note that $r_{\max}([-A, A]) = A$. Then, by using the expression for the raw absolute moment of a Gaussian distribution given in [40], we have that

$$\max_{a \in [0, A]} \mathbb{E}[|a + Z|^p] = \max_{a \in [0, A]} \frac{2^{\frac{p}{2}} \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}} {}_1F_1\left(-\frac{p}{2}; \frac{1}{2}; -\frac{a^2}{2}\right).$$

The proof is concluded by observing that $f(a) := {}_1F_1\left(-\frac{p}{2}; \frac{1}{2}; -\frac{a^2}{2}\right)$ is an increasing function in a . \square

The following theorem establishes the EPI and the Jensen lower bound of Section 4.2 assuming the channel input symbols are uniformly distributed.

Theorem 24. (Lower Bounds with Uniform Inputs) Let $A > 0$ be arbitrary and the channel input X be uniformly distributed over $[-A, A]$. Then,

$$C([-A, A], 1) \geq \mathbb{C}_{\text{EPI}}([-A, A], 1) = \frac{1}{2} \log_2 \left(1 + \frac{2A^2}{\pi e} \right) \quad (52)$$

and

$$C([-A, A], 1) \geq \mathbb{C}_{\text{Jensen}}([-A, A], 1) = \log_2 \left(\frac{\sqrt{2}A^2}{e^{\frac{1}{2}} \left(e^{-A^2} - 1 + \sqrt{\pi}A \left(1 - 2Q(\sqrt{2}A) \right) \right)} \right). \quad (53)$$

Proof. The lower bound in Equation (52) follows from Theorem 12 by observing that $\text{Vol}(\mathcal{X}) = 2A$. To show the lower bound in Equation (53), consider Theorem 14 and let X and X' be independent and uniformly distributed over $[-A, A]$. Then, we have

$$\begin{aligned} \mathbb{E} \left[e^{-\frac{|X-X'|^2}{4}} \right] &= \frac{1}{4A^2} \int_{-A}^A \int_{-A}^A e^{-\frac{(x-x')^2}{4}} dx dx' \\ &= \frac{1}{A^2} \left(e^{-A^2} - 1 + \sqrt{\pi}A \left(1 - 2Q(\sqrt{2}A) \right) \right), \end{aligned}$$

which concludes the proof. \square

Restricting the channel inputs to be discrete allows for another set of lower bounds on Equation (47).

Theorem 25. (Lower Bounds with Discrete Inputs) Let $A > 1$ be arbitrary, $X_B \in \{-A, A\}$ equally likely, and $X_D \sim \text{PAM}(N)$ with $N = \left\lceil 1 + \frac{A}{\sqrt{2\pi e}} \right\rceil$. Then,

$$C([-A, A], 1) \geq \mathbb{C}_{\text{Binary}}([-A, A], 1) := I(X_B; X_B + Z) \quad (54)$$

$$= \frac{1}{\log_e(2)} \left(A^2 - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \log_e \left(\cosh(A^2 - Ay) \right) dy \right), \quad (55)$$

$$C([-A, A], 1) \geq \underline{C}_{\text{Jensen}}([-A, A], 1) = -\log_2 \left(\sqrt{\frac{e}{2}} \frac{1}{N^2} \sum_{(x_{D_i}, x_{D_j}) \in \text{PAM}(N)^2} e^{-\frac{(x_{D_i} - x_{D_j})^2}{4}} \right), \quad (56)$$

and

$$C([-A, A], 1) \geq \underline{C}_{\text{OW}}([-A, A], 1) = \bar{C}_{\text{McK}}([-A, A], 1) - 2. \quad (57)$$

Proof. The expression of the mutual information in Equation (54) for a uniform binary input $X_B \in \{-A, A\}$ is found in [41] by using the I-MMSE relationship. The bound in Equation (56) follows from using Theorem 14 and the bound in Equation (57) from Theorem 13, respectively. This concludes the proof. \square

Figure 5 compares the upper and lower bounds presented in this section in dependency of the amplitude constraint A . Observe that for values of A smaller than ≈ 1.665 (i.e., to the left of the gray vertical line), the lower bound (55) is in fact equal to the capacity. Up to constraints of $A \approx 1$, the moment upper bound in Equation (51) is the best after which the bound in Equation (50) becomes the tightest. The best lower bound for constraint values smaller than $A \approx 10$ is the bound in Equation (56) after which the lower bound in Equation (53) becomes the tightest. Note that all lower and upper bounds are asymptotically tight (i.e., for $A \rightarrow \infty$).

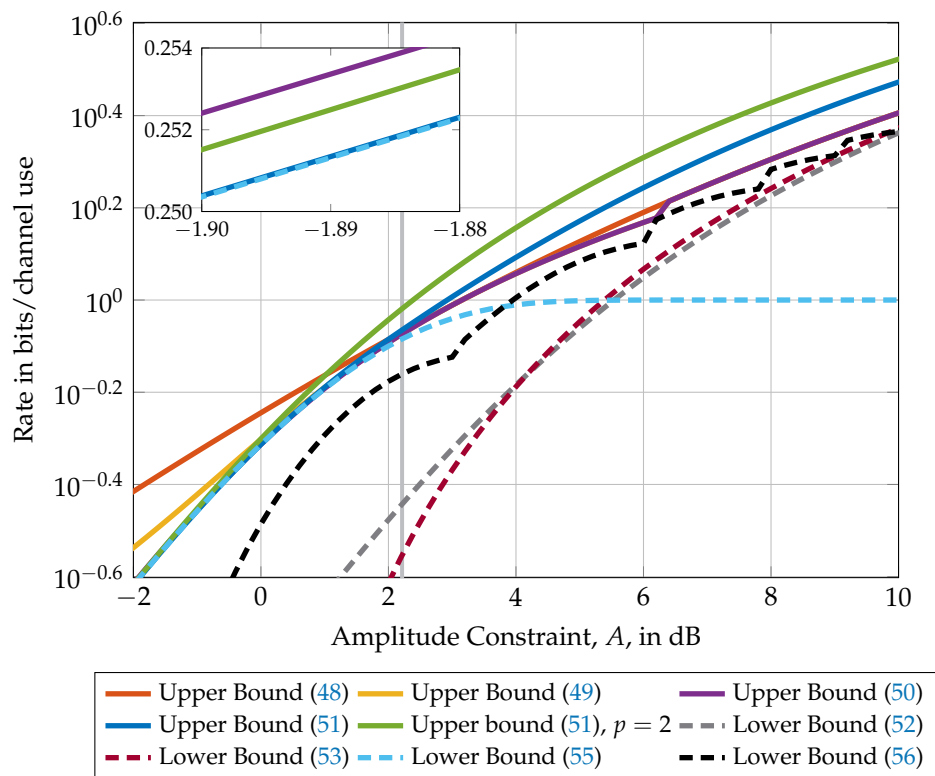


Figure 5. Comparison of upper and lower bounds on the capacity of a SISO channel with amplitude constraint A . The capacity of this channel is known for amplitudes smaller than $A \approx 10 \log_{10}(1.665) = 2.214$ dB only (i.e., to the left of the gray vertical line) and unknown elsewhere. The nested figure represents a zoom into the region $-1.9 \text{ dB} \leq A \leq -1.88 \text{ dB}$ to highlight the differences between the Moment upper bound (51), the Rassouli–Clerckx upper bound in Equation (50), and the lower bound with binary inputs in Equation (56).

7.2. High and Low Amplitude Asymptotics

In this subsection, we study how the capacity in Equation (47) behaves in the high and low amplitude regimes. To this end, we need the following expression

$$\bar{C}_{\text{AWGN}}([-A, A], 1) := \frac{1}{2} \log_2(1 + A^2),$$

which is either the capacity of a SISO channel with an average power constraint A^2 or the moment bound in Equation (51) evaluated for $p = 2$.

Theorem 26. (SISO High and Low Amplitude Asymptotics) *It holds*

$$\lim_{A \rightarrow 0} \frac{C([-A, A], 1)}{\bar{C}_{\text{AWGN}}([-A, A], 1)} = 1, \quad (58)$$

$$\lim_{A \rightarrow \infty} \frac{C([-A, A], 1)}{\bar{C}_{\text{McK}}([-A, A], 1)} = 1, \quad (59)$$

and

$$\lim_{A \rightarrow \infty} C([-A, A], 1) - \bar{C}_{\text{AWGN}}([-A, A], 1) = \frac{1}{2} \log_2\left(\frac{\pi e}{2}\right) \approx 1.044. \quad (60)$$

Proof. The capacity of an amplitude-constrained SISO channel in the regime of low amplitudes (i.e., for amplitudes smaller than $A \approx 1.655$) was given by Guo et al. [41]

$$C([-A, A], 1) = \frac{1}{\log_e(2)} \left(A^2 - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \log_e(\cosh(A^2 - Ay)) dy \right).$$

Now, observe that

$$\begin{aligned} & \lim_{A \rightarrow 0} \frac{1}{\frac{1}{2} \log_2(1 + A^2)} \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \log_e(\cosh(A^2 - Ay)) dy \\ &= \frac{2}{\log_2(e)} \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \frac{\log_e(\cosh(A^2 - Ay))}{\log_e(1 + A^2)} dy \\ &= \frac{2}{\log_2(e)} \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \lim_{A \rightarrow 0} \frac{\log_e(\cosh(A^2 - Ay))}{\log_e(1 + A^2)} dy \\ &= \frac{2}{\log_2(e)} \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \frac{y^2}{2} dy \\ &= \frac{1}{\log_2(e)}. \end{aligned}$$

Therefore, the limit in Equation (58) is given by

$$\begin{aligned} \lim_{A \rightarrow 0} \frac{C([-A, A], 1)}{\bar{C}_{\text{AWGN}}([-A, A], 1)} &= \lim_{A \rightarrow 0} \frac{A^2}{\log_e(2) \frac{1}{2} \log_2(1 + A^2)} - \frac{1}{\log_e(2) \log_2(e)} \\ &= \frac{2 \log_e(2)}{\log_e(2)} - \frac{1}{\log_e(2) \log_2(e)} \\ &= 2 - 1 = 1. \end{aligned}$$

The limit in Equation (59) follows from comparing the EPI lower bound $\underline{C}_{\text{EPI}}([-A, A], 1) = \frac{1}{2} \log_2(1 + \frac{2A^2}{\pi e})$ in (52) with the McKellips upper bound $\bar{C}_{\text{McK}}([-A, A], 1) = \log_2(1 + \frac{2A}{\sqrt{2\pi e}})$ given in Equation (48).

Finally, to show Equation (60), observe that

$$\begin{aligned} \lim_{A \rightarrow \infty} C([-A, A], 1) - \bar{C}_{\text{AWGN}}([-A, A], 1) &= \lim_{A \rightarrow \infty} \bar{C}_{\text{McK}}([-A, A], 1) - \bar{C}_{\text{AWGN}}([-A, A], 1) \\ &= \lim_{A \rightarrow \infty} \log_2\left(1 + \frac{2A}{\sqrt{2\pi e}}\right) - \frac{1}{2} \log_2(1 + A^2) \\ &= \frac{1}{2} \log_2\left(\frac{\pi e}{2}\right). \end{aligned}$$

This concludes the proof. \square

8. Conclusions

In this work, we studied the capacity of MIMO channels with bounded input spaces. Several new properties of input distributions that achieve the capacity of such channels have been provided. In particular, it is shown that the support of a capacity-achieving channel input distribution is a set that is small in a topological and measure theoretical sense. In addition to that, it is shown that, if the radius of the underlying channel input space, \mathcal{X} , is small enough, then the support of a corresponding capacity-achieving input distribution must necessarily be a subset of the boundary of \mathcal{X} . As the considerations on the input distribution have demonstrated that determining the capacity is a very challenging problem, we proposed several new upper and lower bounds that are shown to be tight in the high amplitude regime. An interesting future direction would be to study generalizations of our techniques to wireless optical MIMO channels [42] and other channels such as the wiretap channel [43].

Author Contributions: All authors contributed equally to this work.

Funding: This work was supported in part by the U. S. National Science Foundation under Grants CCF-093970 and CCF-1513915, by the German Research Foundation (DFG) under Grant GO 2669/1-1, and by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement No. 694630.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Telatar, I.E. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun.* **1999**, *10*, 585–595.
2. Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*.
3. Smith, J.G. The information capacity of amplitude-and variance-constrained scalar Gaussian channels. *Inf. Control* **1971**, *18*, 203–219.
4. Shamai (Shitz), S.; Bar-David, I. The capacity of average and peak-power-limited quadrature Gaussian channels. *IEEE Trans. Inf. Theory* **1995**, *41*, 1060–1071.
5. Rassouli, B.; Clerckx, B. On the capacity of vector Gaussian channels with bounded inputs. *IEEE Trans. Inf. Theory* **2016**, *62*, 6884–6903.
6. Sharma, N.; Shamai (Shitz), S. Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels. *Probl. Inf. Transm.* **2010**, *46*, 283–299.
7. McKellips, A.L. Simple tight bounds on capacity for the peak-limited discrete-time channel. In Proceedings of the 2004 IEEE International Symposium on Information Theory (ISIT), Chicago, IL, USA, 27 June–2 July 2004; pp. 348–348.
8. Thangaraj, A.; Kramer, G.; Böcherer, G. Capacity Bounds for Discrete-Time, Amplitude-Constrained, Additive White Gaussian Noise Channels. Available Online: <https://arxiv.org/abs/1511.08742> (accessed on 15 February 2019).
9. Rassouli, B.; Clerckx, B. An Upper Bound for the Capacity of Amplitude-Constrained Scalar AWGN Channel. *IEEE Commun. Lett.* **2016**, *20*, 1924–1926.

10. ElMoslimany, A.; Duman, T.M. On the Capacity of Multiple-Antenna Systems and Parallel Gaussian Channels With Amplitude-Limited Inputs. *IEEE Trans. Commun.* **2016**, *64*, 2888–2899.
11. Dytso, A.; Goldenbaum, M.; Poor, H.V.; Shamai (Shitz), S. A Generalized Ozarow-Wyner Capacity Bound with Applications. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 1058–1062.
12. Dytso, A.; Goldenbaum, M.; Shamai (Shitz), S.; Poor, H.V. Upper and Lower Bounds on the Capacity of Amplitude-Constrained MIMO Channels. In Proceedings of the 2017 IEEE Global Communications Conference (GLOBECOM), Singapore, 4–8 December 2017; pp. 1–6.
13. Cao, P.L.; Oechtering, T.J. Optimal transmit strategy for MIMO channels with joint sum and per-antenna power constraints. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017.
14. Loyka, S. The Capacity of Gaussian MIMO Channels Under Total and Per-Antenna Power Constraints. *IEEE Trans. Commun.* **2017**, *65*, 1035–1043.
15. Loyka, S. On the Capacity of Gaussian MIMO Channels under the Joint Power Constraints. Available online: <https://arxiv.org/abs/1809.00056> (accessed on 15 February 2019).
16. Tuninetti, D. On the capacity of the AWGN MIMO channel under per-antenna power constraints. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, NSW, Australia, 10–14 June 2014; pp. 2153–2157.
17. Vu, M. MISO capacity with per-antenna power constraint. *IEEE Trans. Commun.* **2011**, *59*, 1268–1274.
18. Chan, T.H.; Hranilovic, S.; Kschischang, F.R. Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs. *IEEE Trans. Inf. Theory* **2005**, *51*, 2073–2088.
19. Csiszar, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
20. Abou-Faycal, I.C.; Trott, M.D.; Shamai, S. The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. Inf. Theory* **2001**, *47*, 1290–1301.
21. Fahs, J.; Abou-Faycal, I. On properties of the support of capacity-achieving distributions for additive noise channel models with input cost constraints. *IEEE Trans. Inf. Theory* **2018**, *64*, 1178–1198.
22. Gursoy, M.C.; Poor, H.V.; Verdú, S. The noncoherent Rician fading channel-part I: Structure of the capacity-achieving input. *IEEE Trans. Wireless Commun.* **2015**, *4*, 2193–2206.
23. Katz, M.; Shamai, S. On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels. *IEEE Trans. Inf. Theory* **2004**, *50*, 2257–2270.
24. Ozel, O.; Ekrem, E.; Ulukus, S. Gaussian wiretap channel with amplitude and variance constraints. *IEEE Trans. Inf. Theory* **2015**, *61*, 5553–5563.
25. Bak, J.; Newman, D.J. *Complex Analysis*; Springer Science & Business Media: Berlin, Germany, 2010.
26. Witsenhausen, H.S. Some Aspects of Convexity Useful in Information Theory. *IEEE Trans. Inf. Theory* **1980**, *26*, 265–271.
27. Krantz, S.G.; Parks, H.R. *A Primer of Real Analytic Functions*; Springer Science & Business Media: New York, NY, USA, 2002.
28. Dytso, A.; Yagli, S.; Poor, H.V.; Shamai (Shitz), S. Capacity Achieving Distribution for the Amplitude Constrained Additive Gaussian Channel: An Upper Bound on the Number of Mass Points. Available online: <https://arxiv.org/abs/1901.03264> (accessed on 15 February 2019).
29. Ransford, T. *Potential Theory in the Complex Plane*; London Mathematical Society Student Texts; Cambridge University Press: Cambridge, UK, 1995; Volume 28.
30. Hatsell, C.; Nolte, L. Some geometric properties of the likelihood ratio (Corresp.). *IEEE Trans. Inf. Theory* **1971**, *17*, 616–618.
31. Dytso, A.; Poor, H.V.; Shamai (Shitz), S. On the Capacity of the Peak Power Constrained Vector Gaussian Channel: An Estimation Theoretic Perspective. Available online: <https://arxiv.org/abs/1804.08524> (accessed on 15 February 2019).
32. Dytso, A.; Goldenbaum, M.; Poor, H.V.; Shamai (Shitz), S. When are Discrete Channel Inputs Optimal?—Optimization Techniques and Some New Results. In Proceedings of the 2018 Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 21–23 March 2018; pp. 1–6.
33. Guo, D.; Shamai, S.; Verdú, S. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inf. Theory* **2005**, *51*, 1261–1282.

34. Dytso, A.; Tuninetti, D.; Devroye, N. Interference as Noise: Friend or Foe? *IEEE Trans. Inf. Theory* **2016**, *62*, 3561–3596.
35. Resnick, S.I. *A Probability Path*; Springer Science & Business Media: New York, NY, USA, 2013.
36. Dytso, A.; Bustin, R.; Tuninetti, D.; Devroye, N.; Shamai (Shitz), S.; Poor, H.V. On the Minimum Mean p -th Error in Gaussian Noise Channels and its Applications. *IEEE Trans. Inf. Theory* **2018**, *64*, 2012–2037.
37. Lozano, A.; Tulino, A.M.; Verdú, S. Optimum power allocation for parallel Gaussian channels with arbitrary input distributions. *IEEE Trans. Inf. Theory* **2006**, *52*, 3033–3051.
38. Pérez-Cruz, F.; Rodrigues, M.R.; Verdú, S. MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation. *IEEE Trans. Inf. Theory* **2010**, *56*, 1070–1084.
39. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th dover printing ed.; Dover Publications: New York, NY, USA, 1964.
40. Winkelbauer, A. Moments and Absolute Moments of the Normal Distribution. Available online: <https://arxiv.org/abs/1209.4340> (accessed on 15 February 2019).
41. Guo, D.; Wu, Y.; Shamai, S.; Verdú, S. Estimation in Gaussian Noise: Properties of the Minimum Mean-Square Error. *IEEE Trans. Inf. Theory* **2011**, *57*, 2371–2385.
42. Moser, S.M.; Mylonakis, M.; Wang, L.; Wigger, M. Asymptotic Capacity Results for MIMO Wireless Optical Communication. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 536–540.
43. Dytso, A.; Egan, M.; Perlaza, S.; Poor, H.V.; Shamai (Shitz), S. Optimal Inputs for Some Classes of Degraded Wiretap Channels. In Proceedings of the 2018 IEEE Information Theory Workshop (ITW), Guangzhou, China, 25–29 November 2018; pp. 1–5.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).