New bounds on the mutual information for discrete constellations and application to wireless channel estimation

A. Taufiq Asyhari

School of Computing and Digital Technology, Birmingham City University, Millennium Point, Birmingham, B4 7XG, UK

ARTICLE INFO

Keywords:
Channel estimation
Coded modulation
Discrete constellations
Mutual information
Wireless communications

ABSTRACT

The lack of closed-form expressions of the mutual information for discrete constellations has limited its uses for analyzing reliable communication over wireless fading channels. In order to address this issue, this paper proposes analytically-tractable lower bounds on the mutual information based on Arithmetic-Mean-Geometric-Mean (AM-GM) inequality. The new bounds can apply to a wide range of discrete constellations and reveal some insights into the rate behavior at moderate to high Signal-to-Noise Ratio (SNR) values. The usability of the bounds is further demonstrated to approximate the optimum pilot overhead in stationary fading channels.

1. Introduction

Mutual information has been widely used to compute the ultimate data rate for reliable communication over wireless channels, which proves to be instrumental in various aspects of system design [1–3]. Consider a simple discrete-time model of wireless channels in which the channel output is given by

\[ Y = H\sqrt{\rho}X + Z \]

(1)

where \( Z \) is a Gaussian noise with the distribution \( f_c(0, 1) \), \( H \) is the fading coefficient and \( \sqrt{\rho}X \) is the input symbol scaled by the square root of signal power \( \rho \), which directly represents the Signal-to-Noise Ratio (SNR). Given model (1), the most common formula of the mutual information in the literature (including [1–3]) is given by a simple logarithmic function

\[ I(\sqrt{\rho}X; Y | H) = \log_2(1 + |H|^2 \rho) \]

(2)

which is derived based on employing capacity-achieving Gaussian inputs to communicate the message. Gaussian constellations, however, are rarely used in realistic scenarios due to their unbounded support and infinite alphabet size characteristics.

In practice, an input symbol \( X \) is commonly drawn from a discrete constellation \( \mathcal{X} \) that is made up of fixed and finite alphabets for digital processing reasons. In such a setting, assuming \( |\mathcal{X}| = 2^M \) equiprobable symbols, the mutual information for channel (1) is given by

\[ I\left(\sqrt{\rho}X; Y | H\right) = M - \frac{1}{2M} \sum_{x \in \mathcal{X}} E_Z \left[ \log_2 \left( 1 + \sum_{x' \neq x} e^{-|\sqrt{\rho}\sqrt{H - \sigma^2} - x'|^2 / \sigma^2} \right) \right] \]

(3)

where notation \( E_A[.|] \) denotes the expectation over random variable \( A \).

Formula (3) can be accurately used to compute the data rates for various modulation techniques, such as Quadrature Amplitude Modulation (QAM), Phase/Frequency Shift Keying (P/FSK), Pulse Amplitude Modulation (PAM), etc.

To the best of our knowledge, the lack of a closed-form expression of the expectation over \( Z \) is a major drawback of applying formula (3). Computationally-demanding Monte Carlo simulation is often required to evaluate the expectation. Only for small-sized constellations, this simulation can be simplified using Gauss-Hermite quadratures [4].

In order to simplify the rate computation in (3), this paper proposes tractable lower bounds that can reveal some insights on the behavior of the discrete-input mutual information from moderate to high SNR values. The bounds are derived based on the Arithmetic-Mean-Geometric-Mean (AM-GM) inequality [5] and are valid for a wide range of discrete constellations. We show how the bounds can be effectively used to approximate the optimal pilot overhead for wireless channel estimation.

As a point of departure, Section 2 reviews some existing bounds in the literature. In Section 3, we discuss our proposed bounds and their derivations. We then apply the bounds to approximate the optimum pilot overhead for fading channel estimation and conclude the findings in Section 5.

E-mail address: taufiq-a@ieee.org.

https://doi.org/10.1016/j.dcan.2020.04.001

Received 20 January 2019; Received in revised form 2 February 2020; Accepted 2 April 2020

Available online 11 April 2020

2352-8648/© 2020 Chongqing University of Posts and Telecommunications. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
2. Review of existing bounds

Based on the basic properties of the mutual information, we can directly obtain the upper and lower bounds to $I(\sqrt{\rho}X; Y | H)$ as

$$0 \leq I(\sqrt{\rho}X; Y | H) \leq H(\sqrt{\rho}X) \leq M$$

(4)

where $H(\cdot)$ denotes the entropy function [6]. The lower bound, i.e., $I(\sqrt{\rho}X; Y | H) \geq 0$, comes from the non-negativity property of the mutual information. The upper bound can be obtained using

$$I(\sqrt{\rho}X; Y | H) = H(\sqrt{\rho}X | H) - H(\sqrt{\rho}X, Y | H)$$

(5)

$$\leq H(\sqrt{\rho}X | H) \leq H(\sqrt{\rho}X) \leq M$$

(6)

where the order of inequalities in (6) follows from the non-negativity of the discrete entropy $H(\sqrt{\rho}X | Y, H)$, the use of conditioning that reduces the entropy, i.e., $H(\sqrt{\rho}X) \leq H(\sqrt{\rho}X)$, and equiprobable symbols in $\mathcal{X}$ maximize the entropy.

As analyzed in Ref. [7], refining the upper bound in (4) is possible by considering that for channel (1) with a fixed fading coefficient and SNR $\rho$, Gaussian inputs with the same second order moment as the discrete inputs maximize the mutual information over all input distributions. A tighter upper bound can thus be obtained by combining (2) with (6) to yield

$$I(\sqrt{\rho}X; Y | H) \leq \min \left\{ H(\sqrt{\rho}X), \log_2(1 + |H|^2 \rho) \right\}$$

(7)

$$\leq \min \{ M, \log_2(1 + |H|^2 \rho) \}$$

(8)

The Right-Hand Side (RHS) of (8) has been used in Ref. [7] to provide a tight approximation of the outage probability at a reduced computational complexity. The limitation of bound (8) is that, when $|H|^2 \rho \geq 2^M - 1$, the expression becomes the basic upper bound in (4). Therefore, bound (8) may not fully reveal the high-SNR behavior of the mutual information, specifically the gain when the SNR is amplified.

In some analyses, it is sensible to approximate the behavior of $I(\sqrt{\rho}X; Y | H)$ when $|H|^2 \rho$ goes to the asymptotic limit. For $|H|^2 \rho \ll 1$, typical approximations apply Taylor series expansion [8], which demonstrates the tightness of the refined upper-bound (8). For $|H|^2 \rho \gg 1$, Refs. [9,10] applied a mutual information-Minimum Mean-Squared Error (MMSE) relationship to characterize the rate gain when increasing the SNR. While those approximations can reasonably be tight to the actual $I(\sqrt{\rho}X; Y | H)$, they are only relevant and accurate in the asymptotic regimes of interest.

3. The new bounds

In this section, we introduce a lemma from the AM-GM inequality [5] that assists in obtaining a non-trivial upper bound to the expectation in (3).

**Lemma 1.** For any random variable $W$, the expectation of its logarithm can be bounded as

$$E[\log W] \leq \theta \log E[W]$$

(9)

**Proof.** Define an independent and identically distributed (i.i.d.) sequence $\{\sqrt{W_n}\}$, where $W_n$ has the same distribution as $W$ and $\theta \in \mathbb{Z}$. By applying AM-GM inequality [5] to this sequence and expressing the geometric-mean in terms of an exponential function, we have

$$\sqrt[n]{n} = 1 \sqrt{W_n} = \exp \left( \frac{1}{n \theta} \sum_{n=1}^{n} \log_W \right) \leq \frac{1}{n} \sum_{n=1}^{n} \log W_n$$

(10)

As $n \to \infty$, applying the strong law of large numbers to inequality (10) yields

$$\exp \left( \frac{1}{\theta} \log E[W] \right) \leq E[\sqrt[\theta]{W}]$$

(11)

Taking logarithm on both sides of (11) yields (9).

The new bound can then be stated in the following proposition.

**Proposition 1.** Consider a discrete constellation $\mathcal{X}$ that consists of equiprobable symbols with cardinality of $|\mathcal{X}| = 2^M < \infty$ for transmission over channel (1). For a given fading $H$, the lower bound of the mutual information can be given by:

$$I(\sqrt{\rho}X; Y | H) \geq \left[ M - \frac{1}{\theta n} \sum_{\mathcal{X} \neq x} \log_2 \left( 1 + \sum_{x' \neq x, x \in \mathcal{X}} e^{-|H|^2 |x-x'|^2} \right) \right]$$

(12)

$$\triangleq I'(\rho, H)$$

where $[a]_+ \triangleq \max\{0,a\}$.

**Proof.** Consider a fixed $x \in \mathcal{X}$ and the expectation over $Z$ on the RHS of (3). By specifying $\theta \in [1, \infty)$ and assigning

$$W = 1 + \sum_{x' \neq x} e^{-|H(\sqrt{\rho}(x-x')| + |x'|^2}$$

(13)

in Lemma 1, we apply inequality (9) to obtain an upper bound to the expectation over $Z$ in (3), i.e.,

$$E_X \left[ \log_2 \left( 1 + \sum_{x' \neq x} e^{-|H(\sqrt{\rho}(x-x')| + |x'|^2} \right) \right] \leq \theta \log E_Z \left[ \left( 1 + \sum_{x' \neq x} e^{-|H(\sqrt{\rho}(x-x')| + |x'|^2} \right)^\theta \right]$$

(14)

The expectation on the RHS of (14) is given by the integral

$$\int_{x \in \mathcal{X}} \left( 1 + \sum_{x' \neq x} e^{-|H(\sqrt{\rho}(x-x')| + |x'|^2} \right)^\theta \frac{e^{-|x|^2}}{\pi} dx$$

(15)

Let $v_x = e^{-|H(\sqrt{\rho}(x-x')| + |x'|^2}$ be the exponential term in the sum. Applying the triangle inequality to the integrand in (15), i.e.,

$$\left( 1 + \sum_{x' \neq x} v_x \right)^\theta \leq 1 + \sum_{x' \neq x} v_x^\theta$$

(16)

for $v_x \geq 0$

and evaluating the corresponding integral lead to

$$\int_{x \in \mathcal{X}} \left( 1 + \sum_{x' \neq x} e^{-|H(\sqrt{\rho}(x-x')| + |x'|^2} \right)^\theta \frac{e^{-|x|^2}}{\pi} dx \leq 1 + \sum_{x' \neq x} \int_{x \in \mathcal{X}} e^{-|y|^2} \frac{e^{-|x'|^2}}{\pi} dx$$

(17)
Combining (18) and (14)–(15) with (3) yields

\[
E_p \left[ \log_2 \left( 1 + \sum_{x_i} e^{-|H(x_i-x')|2} \right) \right] \\
\leq \theta \log_2 \left( 1 + \sum_{x_i} e^{-|H(x_i-x')|2} \right)
\]

(19)

Bound (19) can be optimized by minimizing over \( \theta \geq 1 \), which is analytically difficult due to the involvement of \( \log_2(\cdot) \) and \( \exp(\cdot) \) functions in solving the zeros of the derivative. In order to keep a tractable bound, a suboptimal \( \theta = 2 \) is obtained from minimizing the term inside the \( \log_2(\cdot) \) function only, which is equivalent to minimizing \( \frac{\|x\|_{\ell_2}^2}{\|x\|_{\ell_2}^2} \).

By setting \( \theta = \tilde{\theta} = 2 \) in (19) and noting the non-negativity property of the mutual information, we obtain the lower bound (12).

Remark 1. Applying Lemma 1 with \( \theta = 1 \) in the Proof of Proposition 1 coincides with Jensen’s inequality [6] and only leads to a trivial non-negative lower bound on \( I(\sqrt{\rho}X; Y|H) \) as captured by (4).

Proposition 2. For any binary constellations, i.e., \( |x|^2 = 2 \), a tractable and tighter lower bound on the mutual information is given by \( I^{iv}(\rho,H) \) such that

\[
I(\sqrt{\rho}X; Y|H) \geq I^{iv}(\rho,H) \geq I(\rho,H),
\]

where \( Q(\cdot) \) is the standard Gaussian Q function [4].

Proof. An improved bound \( I^{iv}(\rho,H) \) can be obtained by replacing the RHS of triangle inequality (16) with appropriate piecewise functions over different intervals of \( z \) in (15). For binary constellations, the sum over \( x' \neq x \) in (16) consists of \( v_x \) and we can then replace (16) with

\[
(1 + v_x)^2 \leq \begin{cases} 
1 + \frac{v_x}{d}, & \text{for } \frac{\|H(x-x')\|_2}{\|H\|_{\ell_2}} \geq \frac{d}{\sqrt{2}} \\
2^\theta - 1 + v_x^\theta, & \text{for } \frac{\|H(x-x')\|_2}{\|H\|_{\ell_2}} < \frac{d}{\sqrt{2}}
\end{cases}
\]

(21)

where \( \mathbb{R}\{\cdot\} \) denotes the real part of the argument, \( \{\cdot\}^\dagger \) denotes the conjugate and \( d \) is the Euclidean distance of the two binary points. Then, following the same steps as the Proof of Proposition 1, the improved bound can be obtained in (20). In Fig. 1, the lower bounds \( I(\rho,H) \) and \( I^{iv}(\rho,H) \) for Binary Phase-Shift Keying (BPSK) and On-Off Keying (OOK) constellations are plotted along with the exact \( I(\sqrt{\rho}X; Y|H) \) (from Monte Carlo simulation) and the Gaussian-input mutual information against SNR \( \rho \). In terms of accuracy, the bounds can provide some insights on the behavior of achievable rates, especially at moderate to high SNR values (i.e., \( \rho \geq 8 \) dB in this plot). At a low SNR, however, the bounds tend to be less accurate. Indeed, in the case of \( I(\rho,H) \) below certain SNR thresholds, the bound is simply reduced to \( I(\sqrt{\rho}X; Y|H) \geq I(\rho,H) = 0 \), which can not capture the dependency of the rate on \( \rho \). The improved bound \( I^{iv}(\rho,H) \) is observed from the square marked line and dashed line in Fig. 1, which are tighter to \( I(\sqrt{\rho}X; Y|H) \) than \( I(\rho,H) \).

4. Application in the pilot overhead optimization

We next illustrate the usability of the new bound (12) to optimize the pilot overhead for channel estimation where there exist trade-offs between the quality of estimates and the amount of transmitted information per unit time [11–13]. We focus on bound (12) due to its simplicity and generality compared with (20).

Consider channel (1) where the magnitude of the fading follows Rayleigh distribution. The time dynamics [11–13] is governed by an ideal low-pass spectral density [13] with a normalized Doppler bandwidth \( \lambda_d \equiv f_d T_d \), where \( f_d \) is the maximum Doppler shift and \( T_d \) is the symbol time. In order to accurately estimate the fading \( H_k \), \( n_p \) consecutive pilot symbols are transmitted in every period of \( \frac{1}{\lambda_d} \) time instants, where \( \lfloor \cdot \rfloor \) denotes the floor operation. Data symbols are then inserted in the remaining \( \frac{1}{\lambda_d} - n_p \) time instants per period. The pilot overhead can thus be defined as

\[
\beta \equiv \frac{n_p}{\frac{1}{\lambda_d}}
\]

(22)

where \( n_p \) can take any integer value from 1, \ldots, \( \lfloor \frac{1}{\lambda_d} \rfloor \). Denote \( \mathcal{P} \) as a set of time indices for pilot transmission. The time-\( k \) channel estimate \( \hat{H}_k \) can be obtained from a linear estimator that considers observations at times \( k \in \mathcal{P} \), i.e.,

\[
\hat{H}_k = \sum_{k \in \mathcal{P}} \omega_k Y_k = \sum_{k \in \mathcal{P}} \omega_k \left( \sqrt{\rho} H_k + Z_k \right)
\]

(23)

where the pilot symbols are assumed to be in unity. The accuracy of this estimator can be measured from its MMSE [12], i.e.,

\[
\rho^2 = E \left[ \left| H_k - \hat{H}_k \right|^2 \right] = \left( 1 + \frac{\beta}{2 \lambda_d} \rho \right)^{-1}
\]

(24)

For a reasonably wide class of inputs (e.g., Gaussian and equi-energy discrete constellations), the maximum achievable rate \( R \) with the nearest neighbor decoding is given in Refs. [11,13] as

\[
R = (1 - \beta) C(\rho_d)
\]

(25)

where \( C(\rho_d) \) is the average mutual information \( I(\sqrt{\rho_d}X; Y|H) \), averaged over \( H \), with the effective SNR.

![Fig. 1. Mutual Information (MI) and bounds versus the SNR.](image-url)
and applying Jensen
Monte Carlo simulation is thus required to obtain the value of
around
open problem, but a very accurate approximation using Taylor expansion
respectively given by
rivatives of
C
closed-form expression of the mutual information
the other hand, when
A.T. Asyhari Digital Communications and Networks 6 (2020) 542
¼
C
≜
¼
d
* /C18
(30)
A closed-form solution of the optimal β that maximizes R in (25) is an
open problem, but a very accurate approximation using Taylor expansion
around λd < 0 is derived in Ref. [13] as
β' = max \{2λd, \sqrt{(1 + ρ) \frac{C(ρ)}{C(β)}} \lambda_d - \lambda_d \left(1 + \frac{C(ρ)}{C(β)} \right) 2 λ_d - \epsilon \left( \frac{2}{d} \right) \} 
(27)
where \( C(\cdot) \) and \( Ĉ(\cdot) \) are respectively the first and second order deriva-
tives of \( C(\cdot) \) with respect to β.
When the constellation is Gaussian, the mutual information \( C_G(ρ_{th}) \)
and overhead \( \overline{ρ} \) can be analytically computed as shown in Ref. [13]. On
the other hand, when \( \mathcal{X} \) is an equi-energy discrete constellation, a
closed-form expression of the mutual information \( C_x(ρ_{th}) = E_{H'}[I(ρ_{th} X; Y|H)] \)
is not available to the best of our knowledge. A resource-intensive
Monte Carlo simulation is thus required to obtain the value of \( \overline{ρ}_x \).
To simplify the computation, we can approximate \( \overline{ρ}_x \) by replacing \( C_x(ρ_{th}) \)
with a further relaxation of \( E_{H'}[I'(ρ_{th} X; Y)] \), where \( I'(\cdot, \cdot) \) is given by the
proposed bound (12), i.e.,
\[
C_x(ρ_{th}) ≥ E_{H'}[I'(ρ_{th}, H)] 
\geq M - \frac{2}{M} \sum_{x \in \mathcal{X}} \log_2 \left( 1 + \sum_{x \in \mathcal{X}} \rho_{th} |x|^2 \right) 
\geq M - \frac{2}{M} \sum_{x \in \mathcal{X}} \log_2 \left( 1 + \sum_{x \in \mathcal{X}} 4 ρ_{th} |x|^2 + 4 \right) \tag{28}
\]
Inequality (28) follows from exchanging the order of \( E_{H'}[\cdot] \) and \( [\cdot]_+ \),
and applying Jensen’s inequality to \( \log_2(\cdot) \) function.
By considering only ρ when \( C_x'(\cdot) \) is strictly positive, it can be shown that
the first and second order derivatives of \( C_x'(ρ) \) with respect to ρ are
respectively given by
\[
C_x'(ρ) = \sum_{x \in \mathcal{X}} Q(x) \left[ 1 + G(x) \right] 
\]
closely matches with the actual optimal rate obtained from the numerical evaluation for all constellations.

5. Conclusion

We have developed new bounds on the mutual information for discrete constellations and demonstrated the usability to approximate the optimal pilot overhead in the channel estimation. The strength of the main proposed bound is given by its validity across various constellations, and it is thus promising for assisting performance evaluation across different applications.

Acknowledgements

The help from Colin Stephen (Coventry University, UK) in proof-reading the paper is gratefully acknowledged. Part of this work was conducted while the author was with Coventry University, UK, and the University of Bradford, UK.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dcan.2020.04.001.

References