The Mutual Information In The Vicinity of Capacity-Achieving Input Distributions

Hao-Chung Cheng
Department of Electrical Engineering and
Graduate Institute of Communication Engineering,
Department of Mathematics, National Taiwan University
Taipei 10617, Taiwan (R.O.C.)
Hon Hai (Foxconn) Quantum Computing Centre
0000-0003-4499-4679

Barış Nakiboğlu
Department of Electrical and Electronics Engineering
Middle East Technical University
06800 Ankara, Turkey
0000-0001-7737-5423

***Abstract*—The mutual information is analyzed as a function of the input distribution using an identity due to Topsøe for channels with (possibly multiple) linear constraints and finite input and output sets. The mutual information is bounded above by a function decreasing quadratically with the distance to the set of all capacity-achieving input distributions for the case when the distance is less than a certain threshold. Explicit expressions for the threshold and the coefficient of the quadratic decrease are derived. A counter-example is provided demonstrating the non-existence of such a quadratic bound in the case of infinitely many linear cost constraints. Implications of these observations for the channel coding problem and applications of the proof technique to related problems are discussed.**

I. Introduction

In his seminal paper [1], Strassen proved for channels with finite input and output sets that there exist positive constants γ and δ for which the mutual information satisfies

$$I(p; W) \leq C - \gamma \|p - p_*\|^2 \qquad \text{if } \|p - p_*\| \leq \delta \qquad (1)$$

where p_* is the projection of p to the set of all capacity-achieving input distributions Π in the underlying Euclidean space, and hence $\|p - p_*\|$ is the distance of p to Π. Strassen's brief and elegant argument relies implicitly on the fact that for any $p \notin \Pi$, the direction $p - p_*$ cannot be simultaneously orthogonal to the gradient of mutual information at p_*, i.e., orthogonal to $D(W\|q_W)$, and in the kernel of the linear transformation relating the input distributions to the output distributions, i.e., in $\mathcal{K}_W$. Strassen's proof has a gap that can be fixed by using polyhedral convexity, see [2, Appendix A].

Strassen's bound (1), plays an important role in establishing sharp impossibility results for the channel coding theorem, see [1], [3], [4]. Determining an explicit expression for (δ, γ) pair for which (1) holds is also worthwhile because of this role.

One of the claims of Polyanskiy, Poor, and Verdú in [3] is to establish (1) with an explicit coefficient γ. They apply an orthogonal decomposition to write $p - p_* = v_0 + v^{\perp}$, where v_0 is the projection of $p - p_*$ to $\mathcal{K}_W$. Then they assert $\langle v_0, D(W\|q_W)\rangle \leq -\Gamma\|v_0\|$ for some $\Gamma > 0$, see [3, (500)]. This claim, however, is wrong for some p's on certain channels; see, for example, the channel described in §II-C.

In our judgment, the issue overlooked in [3] is the following. The projection of $p - p_*$ to the subspace of $\mathcal{K}_W$ that is also orthogonal to $D(W\|q_W)$ needs not be zero; the principle used by Strassen in [1] asserts merely that this projection cannot be the $p - p_*$ vector itself. This principle can be strengthened using convex analysis to assert that the angle between the $p - p_*$ vector and its projection cannot be less than a positive constant, determined by the channel. In §III, we establish this fact for the case with multiple linear constraints. In §IV, we use this observation to prove (1) with explicit expressions for γ and δ for channels with finitely many linear constraints using orthogonal decompositions, similar to [3].

Recently in [5], Cao and Tomamichel presented a proof of (1), in the spirit of [1]. First the cone generated by the vectors $p - p_*$ for $p \notin \Pi$ is proved to be closed, and then a second-order Taylor series expansion for the parametric family of functions $\{I(p_* + \tau(p - p_*); W)\}_{p \notin \Pi}$ at $\tau = 0$ with a uniform approximation error term for all $p \notin \Pi$ is obtained. Then (1) is established using the extreme value theorem, the fact that $p - p_*$ cannot be an element of $\mathcal{K}_W$ that is orthogonal to $D(W\|q_W)$, and the Taylor series expansion. Cao and Tomamichel, later generalized their analysis to the case with finitely many linear constraints, in [6].

In §II, we introduce our notation, and bound the Kullback–Leibler divergence from below and the mutual information from above using a Taylor series expansion. In §III, we review essential concepts and results from convex analysis and prove the positivity of the aforementioned minimum angle. In §IV, we prove (1) for any channel with possibly multiple linear constraints and finite input and output sets. We also present a channel with infinitely many linear cost constraints for which (1) does not hold. In §V, we discuss the implications of the analysis presented and possible applications of the proof techniques to some related problems.

II. Information Theoretic Preliminaries

For any finite set $\mathcal{X}$, we denote the set of all probability mass functions on $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$. A $p \in \mathcal{P}(\mathcal{X})$ is said to be absolutely continuous in a $q \in \mathcal{P}(\mathcal{X})$, i.e., $p \prec q$, if $p(x) = 0$ for all x satisfying $q(x) = 0$. The Kullback–Leibler divergence $D(p\|q)$ is defined for any $p, q \in \mathcal{P}(\mathcal{X})$ as

$$D(p\|q) := \begin{cases} \sum_x p(x) \ln \frac{p(x)}{q(x)} & p \prec q \\ \infty & p \not\prec q \end{cases}.$$

The Kullback–Leibler divergence is a non-negative function and $D(p\|q) = 0$ iff $p = q$.

We interpret real valued functions on a finite set $\mathcal{X}$ as the elements of a Euclidean vector space $\mathbb{R}^{\mathcal{X}}$. For any $|\mathcal{X}|$-by-$|\mathcal{X}|$ positive semidefinite matrix Λ, we define the inner product $\langle\cdot,\cdot\rangle_{\Lambda}:\mathbb{R}^{\mathcal{X}}\times\mathbb{R}^{\mathcal{X}}\to\mathbb{R}$ and the norm $\|\cdot\|_{\Lambda}:\mathbb{R}^{\mathcal{X}}\to\mathbb{R}_{\geq 0}$ as

$$\langle f,g\rangle_{\Lambda}:=f^{T}\Lambda g \qquad \forall f,g\in\mathbb{R}^{\mathcal{X}},$$
$$\|f\|_{\Lambda}:=\sqrt{\langle f,f\rangle_{\Lambda}} \qquad \forall f\in\mathbb{R}^{\mathcal{X}}.$$

When Λ is the identity matrix, we denote the inner product and the norm by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively.

We represent a channel with a finite input set $\mathcal{X}$ and a finite output set $\mathcal{Y}$ by an $|\mathcal{X}|$-by-$|\mathcal{Y}|$ right stochastic matrix W, i.e. the element in the row x and the column y is the probability of observing the output letter y when the input letter is x, which is commonly denoted by $W(y|x)$. Without loss of generality, we assume that any output letter y satisfies $W(y|x)>0$ for some $x\in\mathcal{X}$. The kernel of the channel W is denoted by $\mathcal{K}_{W}$:

$$\mathcal{K}_{W}:=\left\{v\in\mathbb{R}^{\mathcal{X}}:\left\|W^{T}v\right\|=0\right\}. \tag{2}$$

For $|\mathcal{X}|$-by-$|\mathcal{Y}|$ right stochastic matrix W and $q\in\mathcal{P}(\mathcal{Y})$, $D(W\|q)$ is a column vector whose rows are $D(W(\cdot|x)\|q)$'s. For any $p\in\mathcal{P}(\mathcal{X})$ we define the conditional Kullback–Leibler divergence $D(W\|q|p)$ as

$$D(W\|q|p):=\langle p,D(W\|q)\rangle.$$

A. The Mutual Information

For any $p\in\mathcal{P}(\mathcal{X})$ and $|\mathcal{X}|$-by-$|\mathcal{Y}|$ right stochastic matrix W, the mutual information $I(p;W)$ is defined as

$$I(p;W):=D(W\|q_p|p), \tag{3}$$

where $q_p\in\mathcal{P}(\mathcal{Y})$ is the output distribution of p, i.e.,

$$q_p:=W^{T}p. \tag{4}$$

The following identity, due to Topsøe [7], can be confirmed by substitution

$$D(W\|q|p)=I(p;W)+D(q_p\|q) \tag{5}$$

for all $p\in\mathcal{P}(\mathcal{X})$ and $q\in\mathcal{P}(\mathcal{Y})$. Let $p_*\in\mathcal{P}(\mathcal{X})$ be such that $q_p\prec q_{p_*}$, then as a result of (5) we have

$$\begin{aligned}I(p;W)&=D(W\|q_{p_*}|p)-D(q_p\|q_{p_*})\\&=I(p_*;W)+\langle p-p_*,D(W\|q_{p_*})\rangle-D(q_p\|q_{p_*}).\end{aligned} \tag{6}$$

To characterize the behavior of the identity in (6) in the vicinity of p_* we will bound $D(q_p\|q_{p_*})$ from below using the Taylor series expansion of $z\ln z$ around $z=1$.

$$z\ln z\geq z-1+\tfrac{1}{2}(z-1)^2-\tfrac{1}{6}(z-1)^3 \quad \forall z\in(0,\infty),$$

and the inequity is strict unless $z=1$. Thus

$$D(q_p\|q_{p_*})\geq\sum\nolimits_{y}\tfrac{1}{2}\tfrac{(q_p(y)-q_{p_*}(y))^2}{q_{p_*}(y)}-\tfrac{1}{6}\tfrac{(q_p(y)-q_{p_*}(y))^3}{(q_{p_*}(y))^2}, \tag{7}$$

for all $p,p_*\in\mathcal{P}(\mathcal{X})$ satisfying $q_p\prec q_{p_*}$ and the inequality is strict unless $q_p=q_{p_*}$. Furthermore,

$$\begin{aligned}\sum\nolimits_{y}\tfrac{(q_p(y)-q_{p_*}(y))^2}{q_{p_*}(y)}&=(q_p-q_{p_*})^{T}\mathrm{diag}\Big(\mathbb{1}_{\{q_{p_*}>0\}}\tfrac{1}{q_{p_*}}\Big)(q_p-q_{p_*})\\&=\|p-p_*\|^2_{\Lambda_{p_*}},\end{aligned}$$

where Λ_{p_*} is the $|\mathcal{X}|$-by-$|\mathcal{X}|$ matrix defined as follows

$$\Lambda_{p_*}:=W\mathrm{diag}\Big(\mathbb{1}_{\{q_{p_*}>0\}}\tfrac{1}{q_{p_*}}\Big)W^{T}. \tag{8}$$

On the other hand using Cauchy–Schwarz inequality we get

$$\begin{aligned}q_p(y)-q_{p_*}(y)&=\sum\nolimits_{x}W(y|x)(p(x)-p_*(x))\\&\leq\|W(y|\cdot)\|\cdot\|p-p_*\|,\end{aligned}$$

where $\|W(y|\cdot)\|$ is

$$\|W(y|\cdot)\|=\sqrt{\sum\nolimits_{x}(W(y|x))^2}. \tag{9}$$

Thus we can bound the second term in (7) to get

$$D(q_p\|q_{p_*})\geq\tfrac{1}{2}\|p-p_*\|^2_{\Lambda_{p_*}}-\tfrac{\kappa^3_{p_*}}{6}\|p-p_*\|^3, \tag{10}$$

where κ_{p_*} is defined as

$$\kappa_{p_*}:=\sqrt[3]{\sum\nolimits_{y}\tfrac{\|W(y|\cdot)\|^3}{(q_{p_*}(y))^2}}. \tag{11}$$

Then using (6), we can bound $I(p;W)$ from above in terms of $I(p_*;W)$ provided that $q_p\prec q_{p_*}$ holds, i.e.,

$$\begin{aligned}I(p;W)\leq I(p_*;W)&+\langle p-p_*,D(W\|q_{p_*})\rangle\\&-\tfrac{1}{2}\|p-p_*\|^2_{\Lambda_{p_*}}+\tfrac{\kappa^3_{p_*}}{6}\|p-p_*\|^3.\end{aligned} \tag{12}$$

B. Shannon Capacity and Center

For any closed and convex constraint set $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$, the Shannon capacity $C_{\mathcal{A}}$ is defined as

$$C_{\mathcal{A}}:=\sup\nolimits_{p\in\mathcal{A}}I(p;W).$$

As a result of the extreme value theorem we know that the supremum is achieved because $I(p;W)$ is continuous in p. More interestingly, [8], [9], there exists a unique Shannon center $q_{\mathcal{A}}\in\mathcal{P}(\mathcal{Y})$ satisfying,

$$D(W\|q_{\mathcal{A}}|p)\leq C_{\mathcal{A}} \qquad \forall p\in\mathcal{A}. \tag{13}$$

Furthermore, $D(q_{p_*}\|q_{\mathcal{A}})=0$ and thus $q_{p_*}=q_{\mathcal{A}}$ for any $p_*\in\mathcal{A}$ satisfying $I(p_*;W)=C_{\mathcal{A}}$ by (5).

We know that $q_p=q_{\mathcal{A}}$ and $D(W\|q_{\mathcal{A}}|p)=C_{\mathcal{A}}$ imply $I(p;W)=C_{\mathcal{A}}$ by (3). The existence of a unique Shannon center implies that the converse statement is true as well, i.e., $I(p;W)=C_{\mathcal{A}}$ implies $q_p=q_{\mathcal{A}}$ and $D(W\|q_{\mathcal{A}}|p)=C_{\mathcal{A}}$. Thus the capacity-achieving input distributions can be characterized as the elements of $\mathcal{A}$ satisfying certain linear constraints, i.e.,

$$\Pi_{\mathcal{A}}=\mathcal{A}\cap\mathcal{S}_{\mathcal{A}}, \tag{14}$$

where $\Pi_{\mathcal{A}}$ is the set of all capacity-achieving input distributions in $\mathcal{A}$ and $\mathcal{S}_{\mathcal{A}}$ is an affine subset of $\mathbb{R}^{\mathcal{X}}$ defined below

$$\Pi_{\mathcal{A}}:=\{p\in\mathcal{A}:I(p;W)=C_{\mathcal{A}}\}, \tag{15}$$
$$\mathcal{S}_{\mathcal{A}}:=\left\{v\in\mathbb{R}^{\mathcal{X}}:\langle v,D(W\|q_{\mathcal{A}})\rangle=C_{\mathcal{A}}\ \&\ W^{T}v=q_{\mathcal{A}}\right\}. \tag{16}$$

Both Λ_{p_*} and κ_{p_*}, defined in (8) and (11), respectively, are same for all $p_*\in\Pi_{\mathcal{A}}$ because $q_{p_*}=q_{\mathcal{A}}$ for all $p_*\in\Pi_{\mathcal{A}}$; hence we denote them by $\Lambda_{\mathcal{A}}$ and $\kappa_{\mathcal{A}}$. Furthermore, the non-negativity of the mutual information, (5), and (13), imply

$D(q_p \| q_{\mathcal{A}}) \leq C_{\mathcal{A}}$ and hence $q_p \prec q_{\mathcal{A}}$. Thus for any $p_* \in \Pi_{\mathcal{A}}$ the bound in (12) is

$$I(p; W) \leq C_{\mathcal{A}} + \langle p - p_*, D(W \| q_{\mathcal{A}}) \rangle - \tfrac{1}{2} \|p - p_*\|^2_{\Lambda_{\mathcal{A}}} + \tfrac{\kappa^3_{\mathcal{A}}}{6} \|p - p_*\|^3 \quad \forall p \in \mathcal{A}. \quad (17)$$

We denote the kernel of the second term in (17) by $\mathcal{K}^d_{\mathcal{A}}$:

$$\mathcal{K}^d_{\mathcal{A}} := \{ v \in \mathbb{R}^{\mathcal{X}} : \langle v, D(W \| q_{\mathcal{A}}) \rangle = 0 \}. \quad (18)$$

Note that the kernel of the last two terms in (17) are both equal to $\mathcal{K}_W$ because

$$\|v\|_{\Lambda_{\mathcal{A}}} \geq \| W^T v \| \geq \|v\|_{\Lambda_{\mathcal{A}}} \sqrt{\min_y q_{\mathcal{A}}(y)} \quad \forall v \in \mathbb{R}^{|\mathcal{X}|}. \quad (19)$$

Furthermore, $\|v\|_{\Lambda_{\mathcal{A}}}$ can be bounded from above in terms of $\|v\|$ for an arbitrary $v \in \mathbb{R}^{\mathcal{X}}$ using first the Cauchy–Schwarz inequality and then the concavity of the function $z^{2/3}$ in z together with the Jensen's inequality, as follows:

$$\begin{aligned} \|v\|^2_{\Lambda_{\mathcal{A}}} &= \sum_y \tfrac{(\sum_x W(y|x) v(x))^2}{q_{\mathcal{A}}(y)} \\ &\leq \sum_y \tfrac{\|W(y|\cdot)\|^2 \cdot \|v\|^2}{q_{\mathcal{A}}(y)} \\ &= \|v\|^2 \sum_y q_{\mathcal{A}}(y) \left(\tfrac{\|W(y|\cdot)\|^3}{(q_{\mathcal{A}}(y))^3} \right)^{\frac{2}{3}} \\ &\leq \|v\|^2 \cdot \kappa^2_{\mathcal{A}} \qquad \forall v \in \mathbb{R}^{\mathcal{X}}. \quad (20) \end{aligned}$$

Recall that $\kappa_{\mathcal{A}}$ is κ_{p_*} defined in (11) for any $p_* \in \Pi_{\mathcal{A}}$.

C. A Counter-Example for [3, (500)]

Example 1. Let W be a channel with 9 input letters and 8 output letters given in the following

$$W = \begin{bmatrix} \beta/3 \mathbb{1}_{5\times1} & \beta/3 \mathbb{1}_{5\times1} & \beta/3 \mathbb{1}_{5\times1} & (1-\beta)\mathbf{I}_5 \\ 1/2 & 1/3 & 1/6 & 0_{1\times5} \\ 1/6 & 1/2 & 1/3 & 0_{1\times5} \\ 1/3 & 1/6 & 1/2 & 0_{1\times5} \\ 1/3 & 1/2 & 1/6 & 0_{1\times5} \end{bmatrix},$$

where $\mathbb{1}_{5\times1}$ is a column vector of ones, $\mathbf{I}_5$ is 5-by-5 identity matrix, $0_{1\times5}$ is a row vector of zeros, and β is the unique solution of the equation $\sqrt{3}\sqrt[3]{0.002} = \beta 5^{-\beta}$ on $\beta \in (0, 1/2)$.

With a slight abuse of notation when $\mathcal{A} = \mathcal{P}(\mathcal{X})$, we denote the Shannon capacity by C_W and the Shannon center by q_W. Let us assume $\mathcal{A} = \mathcal{P}(\mathcal{X})$. Then the capacity-achieving input distribution is unique and it is the uniform distribution on the first 5 input letters. Furthermore,

$$C_W = (1-\beta)\ln 5 \quad \text{and} \quad q_W = \begin{bmatrix} \tfrac{\beta}{3} & \tfrac{\beta}{3} & \tfrac{\beta}{3} & \tfrac{1-\beta}{5}\mathbb{1}_{1\times5} \end{bmatrix}^T.$$

Note that $D(W(x) \| q_W) = C_W$ for all input letters x.

On the other hand $\mathcal{K}_W = \{\tau s : \tau \in \mathbb{R}\}$ where the vector s is given by

$$s = \begin{bmatrix} 0_{1\times5} & 2 & 2 & -1 & -3 \end{bmatrix}^T.$$

Note that $\langle s, D(W \| q_W) \rangle = 0$. Thus $\langle v_0, D(W \| q_W) \rangle = 0$ for any p, where v_0 is the projection of p–p_* onto $\mathcal{K}_W$ considered in [3]. On the other hand if p puts non-zero probability only on one of the last four input letters then $\|v_0\| \neq 0$. Consequently, $\langle v_0, D(W \| q_W) \rangle \leq -\Gamma \|v_0\|$, i.e., [3, (500)], cannot be true for any positive Γ.

III. Preliminaries on Convex Analysis

Let $\mathcal{A}$ be a closed convex subset of the Euclidean space $\mathbb{R}^n$. Then by [10, Proposition A.5.2.1], the *tangent cone* of $\mathcal{A}$ at $p_* \in \mathcal{A}$ is the closure of the cone generated by $\{p - p_* : p \in \mathcal{A}\}$:

$$\mathcal{T}_{\mathcal{A}}(p_*) = \mathtt{cl}(\mathtt{cone}(\mathcal{A} - p_*)).$$

The *normal cone* of $\mathcal{A}$ at a point $p_* \in \mathcal{A}$ is

$$\mathcal{N}_{\mathcal{A}}(p_*) := \{ s \in \mathbb{R}^n : s^T(p - p_*) \leq 0, \ \forall p \in \mathcal{A} \}.$$

Then by [10, p. 66]

$$\mathcal{T}_{\mathcal{A}}(p_*) \cap \mathcal{N}_{\mathcal{A}}(p_*) = \{0\} \qquad \forall p_* \in \mathcal{A}. \quad (21)$$

The *projection* of p onto $\mathcal{A}$ is the unique point satisfying

$$\mathtt{P}_{\mathcal{A}}(p) = \arg\min_{p_* \in \mathcal{A}} \|p - p_*\| \qquad \forall p \in \mathbb{R}^n,$$

see [10, p. 46]. Then by [10, Theorem A.3.1.1]

$$p_* = \mathtt{P}_{\mathcal{A}}(p) \iff \langle p - p_*, s - p_* \rangle \leq 0 \qquad \forall s \in \mathcal{A}. \quad (22)$$

A closed convex set $\mathcal{A} \subset \mathbb{R}^n$ is *polyhedral* iff there exists a finite index set $\mathcal{I}_{\mathcal{A}}$, vectors $\{f_\imath \in \mathbb{R}^n\}_{\imath \in \mathcal{I}_{\mathcal{A}}}$, and constants $\{b_\imath \in \mathbb{R}\}_{\imath \in \mathcal{I}_{\mathcal{A}}}$ such that

$$\mathcal{A} = \{ p \in \mathbb{R}^n : \langle f_\imath, p \rangle \leq b_\imath \quad \forall \imath \in \mathcal{I}_{\mathcal{A}} \}. \quad (23)$$

We denote the set of active constraints at p_* by $\mathcal{J}_{\mathcal{A}}(p_*)$, i.e.,

$$\mathcal{J}_{\mathcal{A}}(p_*) := \{ \imath \in \mathcal{I}_{\mathcal{A}} : \langle f_\imath, p_* \rangle = b_\imath \} \qquad \forall p_* \in \mathcal{A}. \quad (24)$$

Then the tangent cone and the normal cone at any $p_* \in \mathcal{A}$ can be characterized via $\mathcal{J}_{\mathcal{A}}(p_*)$ as follows,see [10, p. 67],

$$\mathcal{T}_{\mathcal{A}}(p_*) = \{ p \in \mathbb{R}^n : \langle f_\imath, p \rangle \leq 0 \quad \forall \imath \in \mathcal{J}_{\mathcal{A}}(p_*) \}, \quad (25)$$

$$\mathcal{N}_{\mathcal{A}}(p_*) = \mathtt{cone}(\{ f_\imath : \imath \in \mathcal{J}_{\mathcal{A}}(p_*) \}). \quad (26)$$

Thus both $\mathcal{T}_{\mathcal{A}}(p_*)$ and $\mathcal{N}_{\mathcal{A}}(p_*)$ are closed convex polyhedral sets, as well.

$\mathcal{S}$ is an affine subspace iff there exists a finite index set $\mathcal{I}_{\mathcal{S}}$, vectors $\{f_\imath\}_{\imath \in \mathcal{I}_{\mathcal{S}}}$, and constants $\{b_\imath\}_{\imath \in \mathcal{I}_{\mathcal{S}}}$ such that

$$\mathcal{S} = \{ p \in \mathbb{R}^n : \langle f_\imath, p \rangle = b_\imath \quad \forall \imath \in \mathcal{I}_{\mathcal{S}} \}. \quad (27)$$

Thus an affine subspace $\mathcal{S}$ can be interpreted as a closed convex polyhedral set for which all constraints are active at all points $p_* \in \mathcal{S}$. Hence, the tangent cone and the normal cone will not change from one point of $\mathcal{S}$ to the next and they can be denoted by $\mathcal{T}_{\mathcal{S}}$ and $\mathcal{N}_{\mathcal{S}}$ instead of $\mathcal{T}_{\mathcal{S}}(p_*)$ and $\mathcal{N}_{\mathcal{S}}(p_*)$. If $\mathcal{S}$ is non-empty then $\mathcal{T}_{\mathcal{S}}$ and $\mathcal{N}_{\mathcal{S}}$ are

$$\mathcal{T}_{\mathcal{S}} = \{ p \in \mathbb{R}^n : \langle f_\imath, p \rangle = 0 \quad \forall \imath \in \mathcal{I}_{\mathcal{S}} \}, \quad (28)$$

$$\mathcal{N}_{\mathcal{S}} = \mathtt{span}(\{ f_\imath : \imath \in \mathcal{I}_{\mathcal{S}} \}), \quad (29)$$

where $\mathtt{span}(\{ f_\imath : \imath \in \mathcal{I}_{\mathcal{S}} \})$ is the subspace spanned by $f_\imath$ vectors for $\imath \in \mathcal{I}_{\mathcal{S}}$.

Lemma 1. *Let $\mathcal{A}$ be a closed convex polyhedral subset of $\mathbb{R}^n$, $\mathcal{S}$ be an affine subspace, and Π be their intersection, i.e., $\Pi := \mathcal{A} \cap \mathcal{S}$. Then*

$$\mathcal{T}_{\Pi}(p_*) = \mathcal{T}_{\mathcal{A}}(p_*) \cap \mathcal{T}_{\mathcal{S}} \qquad \forall p_* \in \Pi, \quad (30)$$

$$\mathcal{N}_{\Pi}(p_*) = \mathcal{N}_{\mathcal{A}}(p_*) + \mathcal{N}_{\mathcal{S}} \qquad \forall p_* \in \Pi, \quad (31)$$

$$(\mathcal{T}_{\mathcal{A}}(p_*) \cap \mathcal{N}_{\Pi}(p_*)) \cap \mathcal{T}_{\mathcal{S}} = \{0\} \qquad \forall p_* \in \Pi. \quad (32)$$

The angle θ_{p_*}, defined in (33)*, is uniquely determined by the active constraints at p_* for $\mathcal{A}$ and $\mathcal{S}$, i.e. by $\{f_\imath\}_{\imath\in\mathcal{J}_{\mathcal{A}}(p_*)}$ and $\{f_\imath\}_{\imath\in\mathcal{I}_{\mathcal{S}}}$. Furthermore, θ_{p_*} is positive for all $p_*\in\Pi$ and the angle θ_Π, defined in* (34)*, is positive.*

$$\theta_{p_*}:=\inf_{v\in\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)}\arccos\left(\tfrac{\|\mathrm{P}_{\mathcal{T}_{\mathcal{S}}}(v)\|}{\|v\|}\right),\tag{33}$$

$$\theta_\Pi:=\inf\nolimits_{p_*\in\Pi}\theta_{p_*},\tag{34}$$

where we let $\theta_{p_*}:=\frac{\pi}{2}$ if $\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)=\{0\}$.

Note that θ_{p_*} is the minimum angle between $\mathcal{T}_{\mathcal{S}}$ and $\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)$. Similarly θ_Π is the minimum angle between $\mathcal{T}_{\mathcal{S}}$ and $\bigcup_{p_*\in\Pi}\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)$.

Proof of Lemma 1. Π is a closed convex polyhedral set because any affine subspace is a closed convex polyhedral set and intersection of two closed convex polyhedral sets is again a closed convex polyhedral set. Furthermore,

$$\mathcal{J}_\Pi(p_*)=\mathcal{J}_{\mathcal{A}}(p_*)\cup\mathcal{J}_{\mathcal{S}}(p_*)\qquad\forall p_*\in\Pi.\tag{35}$$

Note that (30) follows from (25), (28), and (35). The identity in (31) follows from (26), (29), and (35). Furthermore, (32) follows from (30) because $\mathcal{T}_\Pi(p_*)\cap\mathcal{N}_\Pi(p_*)=\{0\}$ by (21).

If $\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)=\{0\}$, then $\theta_{p_*}=\frac{\pi}{2}$ by definition, else $v=\mathrm{P}_{\mathcal{T}_{\mathcal{S}}}(v)+\mathrm{P}_{\mathcal{N}_{\mathcal{S}}}(v)$ and $\|\mathrm{P}_{\mathcal{N}_{\mathcal{S}}}(v)\|\neq 0$ for any v in $\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)$ because $v\notin\mathcal{T}_{\mathcal{S}}$ by (32). Thus $\frac{\|\mathrm{P}_{\mathcal{T}_{\mathcal{S}}}(v)\|}{\|v\|}<1$ whenever $v\in\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)$.

$$\begin{aligned}\sup_{v\in\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*)}\frac{\|\mathrm{P}_{\mathcal{T}_{\mathcal{S}}}(v)\|}{\|v\|}&=\sup_{v\in\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*):\|v\|=1}\|\mathrm{P}_{\mathcal{T}_{\mathcal{S}}}(v)\|\\&=\max_{v\in\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_\Pi(p_*):\|v\|=1}\|\mathrm{P}_{\mathcal{T}_{\mathcal{S}}}(v)\|.\end{aligned}$$

Note that we can replace the supremum with maximum because norm and projection are continuous and supremum is over a compact set.

There are only finitely many distinct possible $\mathcal{T}_{\mathcal{A}}(p_*)$ sets for $p_*\in\mathcal{A}$ and finitely many distinct possible $\mathcal{N}_\Pi(p_*)$ sets for $p_*\in\Pi$. Thus there are only finitely many distinct possible θ_{p_*} values for $p_*\in\Pi$. Consequently the infimum θ_Π is positive, as well. ☐

We apply Lemma 1 to $\Pi_{\mathcal{A}}$ defined in (15) and $\mathcal{S}_{\mathcal{A}}$ defined in (16). For any closed convex constraint set $\mathcal{A}$, (14) holds and $\mathcal{S}_{\mathcal{A}}$ is an affine subspace. Thus the hypotheses of Lemma 1 holds whenever $\mathcal{A}$ is determined by finite number of cost constraints, i.e., whenever $\mathcal{A}$ is polyhedral. Thus for any $\mathcal{A}$ determined by finite number of linear constraints the minimum angle $\theta_{\Pi_{\mathcal{A}}}$ between $\mathcal{T}_{\mathcal{S}_{\mathcal{A}}}$ and $\mathcal{T}_{\mathcal{A},\Pi_{\mathcal{A}}}$ is positive, i.e.,

$$\theta_{\Pi_{\mathcal{A}}}>0\tag{36}$$

where $\theta_{\Pi_{\mathcal{A}}}$ and $\mathcal{T}_{\mathcal{A},\Pi_{\mathcal{A}}}$ are

$$\mathcal{T}_{\mathcal{A},\Pi_{\mathcal{A}}}:=\bigcup\nolimits_{p_*\in\Pi_{\mathcal{A}}}\mathcal{T}_{\mathcal{A}}(p_*)\cap\mathcal{N}_{\Pi_{\mathcal{A}}}(p_*),\tag{37}$$

$$\theta_{\Pi_{\mathcal{A}}}:=\begin{cases}\frac{\pi}{2} & \text{if } \mathcal{T}_{\mathcal{A},\Pi_{\mathcal{A}}}=\emptyset\\ \inf\limits_{v\in\mathcal{T}_{\mathcal{A},\Pi_{\mathcal{A}}}}\arccos\left(\frac{\|\mathrm{P}_{\mathcal{T}_{\mathcal{S}}}(v)\|}{\|v\|}\right) & \text{if } \mathcal{T}_{\mathcal{A},\Pi_{\mathcal{A}}}\neq\emptyset\end{cases}.\tag{38}$$

Furthermore, as a result of (28) the tangent $\mathcal{T}_{\mathcal{S}_{\mathcal{A}}}$ of $\mathcal{S}_{\mathcal{A}}$ is

$$\mathcal{T}_{\mathcal{S}_{\mathcal{A}}}=\mathcal{K}^d_{\mathcal{A}}\cap\mathcal{K}_W,\tag{39}$$

where $\mathcal{K}_W$ and $\mathcal{K}^d_{\mathcal{A}}$ are defined in (2) and (18), respectively.

IV. Main Result

Theorem 1. *Let W be a $|\mathcal{X}|$-by-$|\mathcal{Y}|$ right stochastic matrix, i.e. a channel, and $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$ be a closed convex polyhedral constraint set, i.e. a constraint set that can be characterized by a finite number of linear constraints,*

$$I(p;W)\leq C_{\mathcal{A}}-\gamma_{\mathcal{A}}\|p-p_*\|^2+\tfrac{\kappa_{\mathcal{A}}^3}{6}\|p-p_*\|^3\quad\forall p\in\Pi^\delta_{\mathcal{A}},\tag{40}$$

where $p_*:=\mathrm{P}_{\Pi_{\mathcal{A}}}(p)$, the positive constant $\kappa_{\mathcal{A}}$ is defined in[1] *(11), and positive constants $\gamma_{\mathcal{A}}$ and δ, the set $\Pi^\delta_{\mathcal{A}}$ are*

$$\gamma_{\mathcal{A}}:=\tfrac{\sin^2\theta_{\Pi_{\mathcal{A}}}}{2}\inf_{v\in\mathcal{K}^d_{\mathcal{A}}\cap\mathcal{N}_{\mathcal{S}_{\mathcal{A}}}:\|v\|=1}\|v\|^2_{\Lambda_{\mathcal{A}}},\tag{41a}$$

$$\delta:=\left(\kappa^2_{\mathcal{A}}+\tfrac{\gamma_{\mathcal{A}}}{\sin^2\theta_{\Pi_{\mathcal{A}}}}\right)^{-1}\|D(W\|q_{\mathcal{A}})\|,\tag{41b}$$

$$\Pi^\delta_{\mathcal{A}}:=\left\{p\in\mathcal{A}:\|p-\mathrm{P}_{\Pi_{\mathcal{A}}}(p)\|\leq\delta\right\}.\tag{41c}$$

Proof of Theorem 1. Let v be $p-p_*$, and v_1, v_2, v_3 be v's projections to the orthogonal subspace $\mathcal{T}_{\mathcal{S}_{\mathcal{A}}}$, $\mathcal{K}^d_{\mathcal{A}}\cap\mathcal{N}_{\mathcal{S}_{\mathcal{A}}}$, and $\{\beta D(W\|q_{\mathcal{A}}):\beta\in\mathbb{R}\}$:

$$v:=p-p_*,\tag{42a}$$

$$v_1:=\mathrm{P}_{\mathcal{T}_{\mathcal{S}_{\mathcal{A}}}}(v),\tag{42b}$$

$$v_2:=\mathrm{P}_{\mathcal{K}^d_{\mathcal{A}}\cap\mathcal{N}_{\mathcal{S}_{\mathcal{A}}}}(v)\tag{42c}$$

$$v_3:=\tfrac{\langle v,D(W\|q_{\mathcal{A}})\rangle}{\|D(W\|q_{\mathcal{A}})\|^2}D(W\|q_{\mathcal{A}}).\tag{42d}$$

Note that $\mathtt{span}\big(\mathcal{T}_{\mathcal{S}_{\mathcal{A}}},\mathcal{K}^d_{\mathcal{A}}\cap\mathcal{N}_{\mathcal{S}_{\mathcal{A}}},D(W\|q_{\mathcal{A}})\big)=\mathbb{R}^{\mathcal{X}}$. Thus

$$v=v_1+v_2+v_3.\tag{43}$$

On the other hand using (17), we can bound $I(p;W)$ from above for any $p\in\mathcal{A}$ as follows

$$I(p;W)\leq C_{\mathcal{A}}+\langle v,D(W\|q_{\mathcal{A}})\rangle-\tfrac{1}{2}\|v\|^2_{\Lambda_{\mathcal{A}}}+\tfrac{\kappa^3_{\mathcal{A}}}{6}\|v\|^3.\tag{44}$$

Let us proceed with bounding the terms in (44). Note that the sign of the inner product $\langle v,D(W\|q_{\mathcal{A}})\rangle$ cannot be positive because otherwise (13) would be violated. Thus

$$\begin{aligned}\langle v,D(W\|q_{\mathcal{A}})\rangle&=\langle v_3,D(W\|q_{\mathcal{A}})\rangle\\&=-\|v_3\|\cdot\|D(W\|q_{\mathcal{A}})\|.\end{aligned}\tag{45}$$

[1]Recall that (11) defines κ_{p_*} but $q_{p_*}=q_{\mathcal{A}}$ for all $p_*\in\Pi_{\mathcal{A}}$ and thus κ_{p_*} has the same value for all $p_*\in\Pi_{\mathcal{A}}$, which we denote by $\kappa_{\mathcal{A}}$.

On the other hand,

$$
\begin{aligned}
\|v\|^2_{\Lambda_{\mathcal{A}}} &= \|v_2 + v_3\|^2_{\Lambda_{\mathcal{A}}} \\
&\overset{(a)}{\geq} \left(\|v_2\|_{\Lambda_{\mathcal{A}}} - \|v_3\|_{\Lambda_{\mathcal{A}}}\right)^2 \\
&\geq \|v_2\|^2_{\Lambda_{\mathcal{A}}} - 2 \cdot \|v_2\|_{\Lambda_{\mathcal{A}}} \cdot \|v_3\|_{\Lambda_{\mathcal{A}}} \\
&\overset{(b)}{\geq} \|v_2\|^2_{\Lambda_{\mathcal{A}}} - 2\kappa^2_{\mathcal{A}} \cdot \|v_2\| \cdot \|v_3\| \\
&\overset{(c)}{\geq} \tfrac{2\gamma_{\mathcal{A}}}{\sin^2 \theta_{\Pi_{\mathcal{A}}}} \|v_2\|^2 - 2\kappa^2_{\mathcal{A}} \cdot \|v_2\| \cdot \|v_3\| \\
&= \tfrac{2\gamma_{\mathcal{A}}}{\sin^2 \theta_{\Pi_{\mathcal{A}}}} \|v_2 + v_3\|^2 - 2\left(\tfrac{\gamma_{\mathcal{A}} \cdot \|v_3\|}{\sin^2 \theta_{\Pi_{\mathcal{A}}}} + \kappa^2_{\mathcal{A}} \cdot \|v_2\|\right) \cdot \|v_3\| \\
&\overset{(d)}{\geq} \tfrac{2\gamma_{\mathcal{A}}}{\sin^2 \theta_{\Pi_{\mathcal{A}}}} \|v_2 + v_3\|^2 - 2\|v\| \tfrac{\|D(W\|q_{\mathcal{A}})\|}{\delta} \|v_3\|, \\
&\overset{(e)}{\geq} 2\gamma_{\mathcal{A}} \cdot \|v\|^2 - 2\|v\| \tfrac{\|D(W\|q_{\mathcal{A}})\|}{\delta} \|v_3\|, \qquad (46)
\end{aligned}
$$

where (a) follows from the triangle inequality, (b) follows from (20), (c) follows from the definition of $\gamma_{\mathcal{A}}$ given in (41a), (d) follows from (41b) and $\|v_2\| \vee \|v_3\| \leq \|v\|$, and (e) follows from (38), which implies $\|v_1\| \leq \|v\| \cos \theta_{\Pi_{\mathcal{A}}}$ and thus $\|v_2 + v_3\| \geq \|v\| \sin \theta_{\Pi_{\mathcal{A}}}$.

(40) holds for all $p \in \Pi^{\delta}_{\mathcal{A}}$ as a result of (44), (45), and (46).

We are left with establishing the positivity of $\gamma_{\mathcal{A}}$. Note that $\{v \in \mathcal{K}^d_{\mathcal{A}} \cap \mathcal{N}_{\mathcal{S}_{\mathcal{A}}} : \|v\| = 1\}$ is a closed and bounded set, i.e., a compact set, thus the infimum in the definition of $\gamma_{\mathcal{A}}$ given in (41a) is a minimum, i.e., it is achieved by some v_*. If the minimum value in (41a) is zero then $v_* \in \mathcal{K}_W$ by (19); on the other hand $v_* \in \mathcal{K}^d_{\mathcal{A}}$ by hypothesis. Thus $v_* \in \mathcal{T}_{\mathcal{S}_{\mathcal{A}}}$ by (39). This, however, is a contradiction because $v_* \in \mathcal{N}_{\mathcal{S}_{\mathcal{A}}}$ by hypothesis. Hence, $\gamma_{\mathcal{A}}$ is positive. ☐

Theorem 1 assumes $\mathcal{A}$ to be polyhedral. The following example demonstrates that this assumption is not superficial.

Example 2. Let $s \in \mathcal{P}(\mathcal{X})$, $\mathcal{A}$, W, and $p_\beta \in \mathcal{P}(\mathcal{X})$ be

$$
s = \begin{bmatrix} \tfrac{2}{3} & \tfrac{1}{6} & \tfrac{1}{6} \end{bmatrix}^T \qquad \mathcal{A} = \left\{ p \in \mathcal{P}(\mathcal{X}) : \|p - s\| \leq \tfrac{1}{2\sqrt{6}} \right\}
$$

$$
W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad p_\beta = s + \tfrac{1}{12} \begin{bmatrix} -2\cos\beta \\ \cos\beta + \sqrt{3}\sin\beta \\ \cos\beta - \sqrt{3}\sin\beta \end{bmatrix}.
$$

Then $C_{\mathcal{A}} = \ln 2$, $\Pi_{\mathcal{A}} = \{p_0\}$, and $q_{\mathcal{A}} = \begin{bmatrix} \tfrac{1}{2} & \tfrac{1}{2} \end{bmatrix}^T$. Furthermore, the boundary $\mathcal{A}$ can be described parametrically as follows $\partial\mathcal{A} = \{p_\beta : \beta \in (-\pi, \pi]\}$. One can confirm by substitution the following closed-form expressions for $\|p_\beta - p_0\|$ and $I(p_\beta; W)$

$$
\|p_\beta - p_0\| = \tfrac{1}{\sqrt{6}} \left| \sin \tfrac{\beta}{2} \right|,
$$

$$
I(p_\beta; W) = \ln 2 - D(q_{p_\beta} \| q_{p_0}), \quad q_{p_\beta} = \begin{bmatrix} \tfrac{1}{2} + \tfrac{1}{3}\sin^2 \tfrac{\beta}{2} \\ \tfrac{1}{2} - \tfrac{1}{3}\sin^2 \tfrac{\beta}{2} \end{bmatrix}.
$$

Using $D(q_{p_\beta} \| q_{p_0}) \leq \|p_\beta - p_0\|_{\Lambda_{p_0}}$ together with (12) we get

$$
\ln 2 - \tfrac{1}{18}\sin^4 \tfrac{\beta}{2} + \tfrac{\sqrt{2}}{162}\sin^6 \tfrac{\beta}{2} \geq I(p_\beta; W) \geq \ln 2 - \tfrac{1}{9}\sin^4 \tfrac{\beta}{2}.
$$

Thus the decrease of mutual information with the distance from $\Pi_{\mathcal{A}}$ can not be claimed to be at least quadratic as in (1), because for the points on $\partial\mathcal{A}$, i.e., on the boundary of $\mathcal{A}$, decay of the mutual information is much slower, it is proportional with forth power of the distance, rather than the second power.

V. Discussion

We have bounded the mutual information from above by a function that is decreasing quadratically with the distance to the set of all capacity-achieving input distributions $\Pi_{\mathcal{A}}$, for channels with finite input and output sets and with a finite number of linear constraints, i.e., with a polyhedral constraint set $\mathcal{A}$, in Theorem 1.

We assumed the output set of the channel is finite, however, the same analysis applies for the case of countably infinite output sets provided that $\kappa_{\mathcal{A}}$ is finite. There are, however, channels with finite input sets and countably infinite output sets for which not only $\kappa_{\mathcal{A}}$, but also one or more of the entries of $\Lambda_{\mathcal{A}}$ are infinite, see [2, Example 3]. Unfortunately, for such channels we do not have the Hilbert space structure we had before. Nevertheless, one can use Lemma 1 together with Pinsker's inequality —in place of (7)— to establish (1) for all channels with finite input sets and countable output sets, see [2, Appendix B]. Evidently, this analysis extends to any channel with measurable output space and finite input set. This analysis recovers (1) for classical–quantum channels whose density operators are on a separable (rather than finite dimensional) Hilbert space, see [2, Appendix C]. Under appropriate technical assumptions, one can obtain (1) for Augustin information [11]–[14] using the same framework too. However, in each of these extensions one relies on tools that are specific to the problem at hand, as expected.

Example 2 demonstrates that if the convex constraint set $\mathcal{A}$ is not polyhedral, then the decrease might be slower than quadratic with the distance to $\Pi_{\mathcal{A}}$. One way to address this issue might be calculating the distance not to $\Pi_{\mathcal{A}}$ but to the hyperplane $\Pi_{\mathcal{A}} + \mathcal{T}_{\mathcal{S}_{\mathcal{A}}}$ or its intersection with the probability simplex. Such a modification recovers the quadratic decrease with the distance at least for Example 2. Another remedy to the issue raised by Example 2 is working with a distance related to the vector $D(W\|q_{\mathcal{A}})$ and the norm $\|\cdot\|_{\Lambda_{\mathcal{A}}}$ instead of the usual Euclidean distance. Alternatively, one can work with a distance related to the vector $D(W\|q_{\mathcal{A}})$ and the total variation norm of the corresponding output distributions. These approaches are inspired by characterization given in (6).

Another interesting question for this line of work is the determination of the best coefficient for the quadratic decrease for a given channel and polyhedral constraint.

Acknowledgment

Authors would like to thank Michael X. Cao and Marco Tomamichel for bringing to their attention the gap in Strassen's proof [1] and for the subsequent discussion on the topic. HC is supported under Grant NSTC 111-636-E-002-026, NTU-111V0904, and NTU-CC-111L894605.

References

[1] V. Strassen, “Asymptotische abschätzungen in Shannons Informationstheorie,” in *Trans. Third Prague Conf. Inf. Theory*, 1962, pp. 689–723, (https://pi.math.cornell.edu/~pmlut/strassen.pdf).

[2] H.-C. Cheng and B. Nakiboğlu, “The mutual information in the vicinity of capacity–achieving input distributions,” *arXiv:2304.14219v2 [cs.IT]*, 2023. [Online]. Available: https://arxiv.org/abs/2304.14219v2

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[4] M. Tomamichel and V. Y. F. Tan, “A tight upper bound for the third-order asymptotics for most discrete memoryless channels,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7041–7051, Nov 2013.

[5] M. X. Cao and M. Tomamichel, “On the quadratic decaying property of the information rate function,” *arXiv:2208.12945v1 [cs.IT]*, 2022. [Online]. Available: https://arxiv.org/abs/2208.12945v1

[6] ——, “Comments on “channel coding rate in the finite blocklength regime”: On the quadratic decaying property of the information rate function,” *arXiv:2208.12945v2 [cs.IT]*, 2023. [Online]. Available: https://arxiv.org/abs/2208.12945v2

[7] F. Topsøe, “An information theoretical identity and a problem involving capacity,” *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 291–292, 1967.

[8] J. H. B. Kemperman, “On the Shannon capacity of an arbitrary channel,” *Indagationes Mathematicae (Proceedings)*, vol. 77, no. 2, pp. 101–115, 1974.

[9] B. Nakiboğlu, “The Rényi Capacity and Center,” *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 841–860, Feb 2019, (arXiv:1608.02424 [cs.IT]).

[10] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, 1st ed., ser. Grundlehren Text Editions. Heidelberg: Springer-Verlag Berlin, 2001.

[11] U. Augustin, “Noisy channels,” Habilitation Thesis, Universität Erlangen-Nürnberg, 1978, (http://bit.ly/3bsWDgG).

[12] I. Csiszár, “Generalized cutoff rates and Rényi’s information measures,” *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 26–34, Jan 1995.

[13] B. Nakiboğlu, “The Augustin Capacity and Center,” *Problems of Information Transmission*, vol. 55, no. 4, pp. 299–342, October 2019, (arXiv:1803.07937 [cs.IT]).

[14] H.-C. Cheng, M. H. Hsieh, and M. Tomamichel, “Quantum sphere-packing bounds with polynomial prefactors,” *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 2872–2898, May 2019, (arXiv:1704.05703 [quant-ph]).