


Article

Belavkin–Staszewski Relative Entropy, Conditional Entropy, and Mutual Information

Yuan Zhai , Bo Yang and Zhengjun Xi *

The College of Computer Science, Shaanxi Normal University, Xi'an 710062, China; zhaiyuan@snnu.edu.cn (Y.Z.); byang@snnu.edu.cn (B.Y.)

* Correspondence: xizhengjun@snnu.edu.cn

Abstract: Belavkin–Staszewski relative entropy can naturally characterize the effects of the possible noncommutativity of quantum states. In this paper, two new conditional entropy terms and four new mutual information terms are first defined by replacing quantum relative entropy with Belavkin–Staszewski relative entropy. Next, their basic properties are investigated, especially in classical–quantum settings. In particular, we show the weak concavity of the Belavkin–Staszewski conditional entropy and obtain the chain rule for the Belavkin–Staszewski mutual information. Finally, the subadditivity of the Belavkin–Staszewski relative entropy is established, i.e., the Belavkin–Staszewski relative entropy of a joint system is less than the sum of that of its corresponding subsystems with the help of some multiplicative and additive factors. Meanwhile, we also provide a certain subadditivity of the geometric Rényi relative entropy.

Keywords: Belavkin–Staszewski relative entropy; geometric Rényi relative entropy; conditional entropy; mutual information; classical–quantum setting



Citation: Zhai, Y.; Yang, B.; Xi, Z. Belavkin–Staszewski Relative Entropy, Conditional Entropy, and Mutual Information. *Entropy* **2022**, *24*, 837. <https://doi.org/10.3390/e24060837>

Academic Editor: Gregg Jaeger

Received: 23 May 2022

Accepted: 16 June 2022

Published: 17 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rényi proposed an axiomatic approach to derive the Shannon entropy, and he found a family of entropies with parameter α ($\alpha \in [0, 1) \cup (1, \infty)$), called Rényi entropy. Meanwhile, the same axiomatic approach was extended to relative entropy and obtained Rényi relative entropy [1]. Relative entropy (or Kullback–Leibler divergence [2]) is a special case of Rényi relative entropy, which is an important ingredient for a mathematical framework of information theory. It has operational meaning in information theoretical tasks and can be used to describe the level of closeness between two random variables [3,4]. The axiomatic approach introduced by Rényi can be readily generalized to quantum settings [5,6]. Because of the non-commutativity of the quantum states, there are at least three different and special ways to generalize the classical Rényi relative entropy [6–10], such as Petz–Rényi relative entropy [11,12], sandwiched Rényi relative entropy [6,13] and geometric Rényi relative entropy [14,15]. These quantities are very meaningful in different information-theoretic tasks, including source coding, hypothesis testing, state merging, and channel coding.

The fact is that quantum relative entropy, by taking the limit as $\alpha \rightarrow 1$, is a special case of the Petz–Rényi and sandwiched Rényi relative entropies. However, the geometric Rényi relative entropy converges to the Belavkin–Staszewski (BS) relative entropy by taking the same limit. It is noteworthy that both the quantum and BS relative entropies are important variants of the classical relative entropy extension to quantum settings [16–18]. Quantum relative entropy, a direct generalization of the classical relative entropy, has been studied extensively in recent decades. BS relative entropy is also an enticing and crucial entropy used to process quantum information tasks, which can be used to describe the effects of possible noncommutativity of the quantum states (the quantum relative entropy can not work well for this). Additionally, BS relative entropy has recently attracted the attention of researchers. More precisely, Katariya and Wilde employed BS relative entropy to discuss quantum channel estimation and discrimination [19], Bluhm and Capel contributed a

strengthened data processing inequality for BS relative entropy [20]. This property was first established by Hiai and Petz [21]. Bluhm et al. produced some weak quasi-factorization results for BS relative entropy [22]. Fang and Fawzi studied quantum channel capacities with respect to geometric Rényi relative entropy [23].

It is commonly known that von Neumann entropy, quantum conditional entropy, and quantum mutual information play vital roles in quantum information theory. Apart from the above entropic measures derived from the quantum relative entropy, however, other useful entropy-like quantities have also been well studied recently, such as max-information [24], collision entropy [25], and min- and max-entropies [26–28]. All of these information measures were generated from quantum Rényi relative entropies by taking different limits.

BS relative entropy can be seen as a fresh and forceful tool to resolve some specific challenges of quantum information-processing tasks. Concurrently, the main use of the geometric Rényi and BS relative entropies is to establish upper bounds on the rates of feedback-assisted quantum communication protocols [29]. To our present knowledge, there is no systematic analysis and research for conditional entropy and mutual information defined from BS relative entropy. Therefore, this paper explores some basic but necessary results for BS relative entropy. More precisely, we first provide a class of new definitions of conditional entropy (called BS conditional entropy, see Definition 2) and new mutual information (called BS mutual information, see Definition 3) via BS relative entropy. Additionally, we showed that von Neumann entropy can be defined by BS relative entropy. Second, we built an order relation between the BS conditional entropy of the bipartite and tripartite quantum systems. Subsequently, since classical-quantum states play an essential role in quantum channel coding and classical data compression with quantum side information, we discussed some valuable properties of BS conditional entropy and BS mutual information in classical-quantum settings. We established the weak concavity of BS conditional entropy and obtained chain rules for BS mutual information. Last but not least, the subadditivity of the geometric Rényi and BS relative entropies is established with the help of some multiplicative and additive factors (the factors are different linear combinations of quantum max-relative entropy [30]), i.e., the geometric Rényi/BS relative entropy of a joint system is less than the sum of that of its corresponding subsystems.

This paper is organized as follows. In Section 2, we present the mathematical terminology and formal definitions necessary for the formulation of our results. Our results are shown in Section 3. The paper ends with a conclusion.

2. Basic Notations and Definitions

We denote a finite-dimensional Hilbert space by \mathcal{H} . Normalized quantum states are in the set $\mathcal{S}_=(\mathcal{H}) := \{\rho \in \mathcal{P}(\mathcal{H}) : \text{Tr}\rho = 1\}$, and subnormalized states are in the set $\mathcal{S}_\leq(\mathcal{H}) := \{\rho \in \mathcal{P}(\mathcal{H}) : 0 < \text{Tr}\rho \leq 1\}$. We use $\mathcal{P}_+(\mathcal{H})$ and $\mathcal{P}(\mathcal{H})$ to denote the set of positive definite operators and the set of positive semi-definite operators on \mathcal{H} , respectively. An identity operator is denoted by I . The Hilbert spaces corresponding to different physical systems are distinguished by different capital Latin letters as a subscript. A compound system is modeled using the Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. For a bipartite classical-quantum system \mathcal{H}_{XB} , the corresponding state ρ_{XB} is formalized as

$$\rho_{XB} = \sum_x p(x) |x\rangle\langle x|_X \otimes \rho_B^x, \tag{1}$$

where $\{|x\rangle\}$ corresponds to an orthonormal basis on the classical system \mathcal{H}_X , ρ_B^x is any quantum state on the quantum system \mathcal{H}_B , $p(x)$ is the probability distribution, and $\sum_x p(x) = 1$ [17,29]. We also refer to a tripartite classical-quantum state,

$$\rho_{XAB} = \sum_x p(x) |x\rangle\langle x|_X \otimes \rho_{AB}^x, \tag{2}$$

where ρ_{AB}^x is any quantum state on the quantum system \mathcal{H}_{AB} .

In quantum information theory, one can generalize Rényi relative entropy to the quantum case, these quantities depend on a parameter $\alpha \in (0, 1) \cup (1, \infty)$, and one can evaluate their values at $\alpha \in \{0, 1, \infty\}$ by taking different limits. For Petz-Rényi relative entropy [11,12] and sandwiched Rényi relative entropy [6,13], if one takes the limit as $\alpha \rightarrow 1$, we can obtain the well-known quantum relative entropy. For $\rho \in \mathcal{S}_=(\mathcal{H})$ and $\sigma \in \mathcal{S}_\leq(\mathcal{H})$, if $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, the quantum relative entropy of ρ and σ is defined as

$$D(\rho\|\sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)]; \tag{3}$$

otherwise, it is defined as $+\infty$. Throughout this paper, we take the logarithmic function to base 2. The quantum relative entropy is nonnegative and satisfies the data processing inequality, which has good applications in quantum hypothesis testing and quantum resource theory [5,31,32]. We now define the geometric Rényi relative entropy [5,14,15,29].

Definition 1. For all $\alpha \in (0, 1) \cup (1, \infty)$, $\rho \in \mathcal{S}_=(\mathcal{H})$ and $\sigma \in \mathcal{S}_\leq(\mathcal{H})$, the geometric Rényi relative entropy is defined as

$$\hat{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \left[\text{Tr} \left[\sigma (\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}})^\alpha \right] \right]. \tag{4}$$

The term of $\sigma^{\frac{1}{2}} (\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}})^\alpha \sigma^{\frac{1}{2}}$ is called the weighted matrix geometric mean of two positive definite operators ρ and σ , where α is the weight parameter [5,23,33]. The geometric Rényi relative entropy can be shown to be the maximal relative entropy among all quantum Rényi relative entropies satisfying the data processing inequality [5,15], so it is also called maximal quantum Rényi relative entropy [5]. The geometric Rényi relative entropy increases monotonically with respect to the parameter α . Specially, for the limit as $\alpha \rightarrow 1$, the geometric Rényi relative entropy converges to the BS relative entropy [15,19,29], i.e.,

$$\hat{D}(\rho\|\sigma) = \lim_{\alpha \rightarrow 1} \hat{D}_\alpha(\rho\|\sigma) = \text{Tr} \left[\rho \log (\rho^{\frac{1}{2}} \sigma^{-1} \rho^{\frac{1}{2}}) \right], \tag{5}$$

where $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$; otherwise, $\hat{D}(\rho\|\sigma) = +\infty$. Here, the inverse σ^{-1} is taken on the support of σ .

Similar to the quantum relative entropy [17], the BS relative entropy is non-negative and satisfies the data processing inequality [15,19,29]. For every quantum channel \mathcal{E} , we have

$$\hat{D}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq \hat{D}(\rho\|\sigma). \tag{6}$$

For more properties of the BS and geometric Rényi relative entropies, one can refer to [19,29]. In particular, the quantum relative entropy is never larger than the BS relative entropy [19,21], i.e.,

$$D(\rho\|\sigma) \leq \hat{D}(\rho\|\sigma). \tag{7}$$

Obviously, if ρ and σ can be commuted, the BS relative entropy will reduce to the quantum relative entropy. In this paper, we also need to employ the quantum max-relative entropy [5,6,30,34], which comes from the sandwiched Rényi relative entropy by taking the limit as $\alpha \rightarrow \infty$, and

$$D_{\max}(\rho\|\sigma) = \log \inf \{ \lambda : \rho \leq \lambda \sigma \}. \tag{8}$$

3. Main Results

Once again, the Petz-/sandwiched and geometric Rényi relative entropies are inconsistent when taking the limit as $\alpha \rightarrow 1$, which leads to many differences between the BS relative entropy $\hat{D}(\rho\|\sigma)$ (generated from the geometric Rényi relative entropy) and the quantum relative entropy $D(\rho\|\sigma)$ (generated from the Petz-/sandwiched Rényi relative entropy). However, we find that both the quantum relative entropy and the BS relative

entropy reduce to the von Neumann entropy for any quantum state ρ when one takes $\sigma = I$, i.e.,

$$\begin{aligned} \hat{D}(\rho\|I) &= \text{Tr}\left[\rho \log(\rho^{\frac{1}{2}}I^{-1}\rho^{\frac{1}{2}})\right] \\ &= \text{Tr}[\rho \log \rho] \\ &= -S(\rho). \end{aligned}$$

Thus, we have

$$\hat{D}(\rho\|I) = D(\rho\|I) = -S(\rho). \tag{9}$$

3.1. Belavkin–Staszewski Conditional Entropy

One of the significant properties of the relative entropy is that it can derive the conditional entropy and mutual information in information theory. The quantum relative entropy is the quantum analogue of Kullback–Leibler divergence. We know that there is no similar concept for the joint probability distribution of two variables with different time in quantum mechanics; in other words, there is no real conditional quantum state to process quantum information tasks. Thus, we can consider a formal definition of the quantum conditional entropy [17,31], i.e.,

$$S(A|B)_\rho = S(\rho_{AB}) - S(\rho_B), \tag{10}$$

where $\rho_B = \text{Tr}_A(\rho_{AB})$ is the reduced state for the bipartite quantum state ρ_{AB} . The quantum conditional entropy $S(A|B)_\rho$ can be denoted as the quantum relative entropy [5,29], i.e.,

$$S(A|B)_\rho = -D(\rho_{AB}\|I_A \otimes \rho_B). \tag{11}$$

In fact, from the basic properties of the quantum relative entropy, the above equation has another equivalent expression [5,29],

$$S(A|B)_\rho = \max_{\sigma_B} -D(\rho_{AB}\|I_A \otimes \sigma_B), \tag{12}$$

where the maximum is taken over all sub-normalized states on \mathcal{H}_B .

Combining Equation (9) with Equation (10), we have

$$S(A|B)_\rho = -\hat{D}(\rho_{AB}\|I_{AB}) + \hat{D}(\rho_B\|I_B). \tag{13}$$

However, from the property of Equation (7) and definition of Equation (11), intuitively, we find that conditional entropy defined by the BS relative entropy is different from the quantum conditional entropy of Equation (11) generally. Therefore, we define a new conditional entropy based on the BS relative entropy in the following: the so-called BS conditional entropy.

Definition 2. For any quantum state $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$, the BS conditional entropy is defined as

$$\hat{S}(A|B)_\rho = -\hat{D}(\rho_{AB}\|I_A \otimes \rho_B). \tag{14}$$

Similar to Equation (12), we can also define the alternative BS conditional entropy, i.e.,

$$\hat{S}_m(A|B)_\rho = \max_{\sigma_B} -\hat{D}(\rho_{AB}\|I_A \otimes \sigma_B), \tag{15}$$

where $\sigma_B \in \mathcal{S}_\leq(\mathcal{H}_B)$. In general, the optimal state is not necessarily the state ρ_B . We further have

$$\hat{S}_m(A|B)_\rho \leq \hat{S}(A|B)_\rho \leq S(A|B)_\rho, \tag{16}$$

from the relation of Equation (7). Additionally, if one considers the above relations in the bipartite classical-quantum systems, they remain equal, as follows.

Lemma 1. For any bipartite classical-quantum states ρ_{XB} , we have

$$\hat{S}_m(B|X)_\rho = \hat{S}(B|X)_\rho = S(B|X)_\rho. \tag{17}$$

Proof. Without a loss of generality, letting $\sigma_X = \sum_x q(x)|x\rangle\langle x|$ and $\rho_X = \sum_x p(x)|x\rangle\langle x|$, using the definition of Equation (1), we have

$$\begin{aligned} & \text{Tr} \left[\rho_{XB} \log \left(\rho_{XB}^{\frac{1}{2}} I_B \otimes \sigma_X^{-1} \rho_{XB}^{\frac{1}{2}} \right) \right] \\ = & \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_B^x \cdot \log \left(\sum_x p(x) q(x)^{-1} |x\rangle\langle x| \otimes \rho_B^x \right) \right] \\ = & \text{Tr} \left[\sum_x p(x) \log p(x) q(x)^{-1} |x\rangle\langle x| \otimes \rho_B^x \right] + \sum_x p(x) \text{Tr} [\rho_B^x \log \rho_B^x] \\ = & D(p(x) \| q(x)) - \sum_x p(x) S(\rho_B^x), \end{aligned} \tag{18}$$

where the last equality follows from the fact that

$$\text{Tr} \left[\sum_x p(x) \log p(x) q(x)^{-1} |x\rangle\langle x| \otimes \rho_B^x \right] = \text{Tr} \left[\sum_x p(x) \log p(x) q(x)^{-1} |x\rangle\langle x| \right] \cdot \text{Tr} [\rho_B^x]$$

and $\text{Tr} [\rho_B^x] = 1$.

Next, taking the minimization of σ_X for both sides of Equation (18), we have

$$\begin{aligned} & \min_{\sigma_B} \text{Tr} \left[\rho_{XB} \log \left(\rho_{XB}^{\frac{1}{2}} I_B \otimes \sigma_X^{-1} \rho_{XB}^{\frac{1}{2}} \right) \right] \\ = & \min_{\sigma_B} D(p(x) \| q(x)) - \sum_x p(x) S(\rho_B^x) \\ = & - \sum_x p(x) S(\rho_B^x). \end{aligned} \tag{19}$$

The optimization of σ_X for the first equality only depends on the first term. For all x , one takes the minimization if and only if $p(x) = q(x)$, which also implies that $D(p(x) \| q(x)) = 0$. Furthermore, combining the definitions of Equation (15) with Equation (5) to Equation (19), it holds that

$$\hat{S}_m(B|X)_\rho = \sum_x p(x) S(\rho_B^x).$$

For $\hat{S}(B|X)_\rho$, we have

$$\begin{aligned} & \text{Tr} \left[\rho_{XB} \log \left(\rho_{XB}^{\frac{1}{2}} I_B \otimes \rho_X^{-1} \rho_{XB}^{\frac{1}{2}} \right) \right] \\ = & \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_B^x \cdot \log \left(\sum_x p(x) p(x)^{-1} |x\rangle\langle x| \otimes \rho_B^x \right) \right] \\ = & - \sum_x p(x) S(\rho_B^x). \end{aligned}$$

Similarly, we have

$$\hat{S}(B|X)_\rho = \sum_x p(x) S(\rho_B^x).$$

Finally, we can obtain the same result for $S(B|X)_\rho$. We thus complete this proof. \square

Since the BS relative entropy satisfies the data processing inequality (6), for any tripartite quantum systems \mathcal{H}_{ABC} , it is easy to obtain that conditioning reduces entropy, i.e.,

$$\hat{S}(A|BC)_\rho \leq \hat{S}(A|B)_\rho. \tag{20}$$

This property also holds for the quantum conditional entropy $S(A|B)_\rho$. However, there is not always true for $S(A|B)_\rho \leq S(AC|B)_\rho$ (Problem 11.3(2), in [17]). For the BS conditional entropy, this paper provides another result in the following.

Lemma 2. For any tripartite quantum states $\rho_{ABC} \in \mathcal{S}_=(\mathcal{H}_{ABC})$, we have

$$\hat{S}(AB|C)_\rho - \log d_A \leq \hat{S}(B|C)_\rho \tag{21}$$

and

$$\hat{S}_m(AB|C)_\rho - \log d_A \leq \hat{S}_m(B|C)_\rho, \tag{22}$$

where d_A is the dimension of subsystem \mathcal{H}_A .

Proof. Since the BS relative entropy satisfies the data-processing inequality (Corollary 4.53 in [29]) and the fact that partial trace is a quantum channel [5], we have

$$\hat{D}(\rho_{ABC} \| I_{AB} \otimes \rho_C) \geq \hat{D}(\text{Tr}_A(\rho_{ABC}) \| \text{Tr}_A(I_{AB} \otimes \rho_C)).$$

Furthermore, we have

$$\hat{D}(\text{Tr}_A(\rho_{ABC}) \| \text{Tr}_A(I_{AB} \otimes \rho_C)) = \hat{D}(\rho_{BC} \| d_A I_B \otimes \rho_C),$$

where d_A is the dimension of subsystem \mathcal{H}_A . Applying the additivity of the BS relative entropy (Proposition 4.54, [29]), we then have

$$\hat{D}(\rho_{BC} \| d_A I_B \otimes \rho_C) = \hat{D}(\rho_{BC} \| I_B \otimes \rho_C) - \log d_A.$$

Recalling the definition of the BS conditional entropy $\hat{S}(A|B)_\rho$ to the above equality, we have

$$\hat{S}(AB|C)_\rho - \log d_A \leq \hat{S}(B|C)_\rho. \tag{23}$$

Similarly, for the alternative definition of the BS conditional entropy $\hat{S}_m(A|B)_\rho$, taking the minimization of σ_X , we have

$$\min_{\sigma_B} \hat{D}(\rho_{ABC} \| I_{AB} \otimes \sigma_C) \geq \min_{\sigma_B} \hat{D}(\rho_{BC} \| I_B \otimes \sigma_C) - \log d_A,$$

which implies that

$$\hat{S}_m(AB|C)_\rho - \log d_A \leq \hat{S}_m(B|C)_\rho.$$

□

The quantum conditional entropy of Equation (10) also satisfies the concavity, which plays an important role in the quantum information processing [5,17,31]. Additionally, for the BS conditional entropy of the tripartite classical-quantum state, we obtain the following result.

Theorem 1. For any tripartite classical-quantum states ρ_{XAB} , we have

$$H(X) + \sum_x p(x) \hat{S}(A_x|B)_\rho \leq \hat{S}(A|B)_\rho + \log d_X \tag{24}$$

and

$$H(X) + \sum_x p(x) \hat{S}_m(A_x|B)_\rho \leq \hat{S}_m(A|B)_\rho + \log d_X, \tag{25}$$

where $H(X)$ is the Shannon entropy, and $\hat{S}(A_x|B)_\rho$ is the BS conditional entropy for the quantum state ρ_{AB}^x .

Proof. For any tripartite classical-quantum states ρ_{XAB} , we have

$$\begin{aligned} & \text{Tr} \left[\rho_{XAB} \log \left(\rho_{XAB}^{\frac{1}{2}} I_{XA} \otimes \rho_B^{-1} \rho_{XAB}^{\frac{1}{2}} \right) \right] \\ &= \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_{AB}^x \cdot \log \left(\sum_x p(x) |x\rangle\langle x| \otimes (\rho_{AB}^x)^{\frac{1}{2}} I_A \otimes \rho_B^{-1} (\rho_{AB}^x)^{\frac{1}{2}} \right) \right] \quad (26) \\ &= \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_{AB}^x \cdot \log \left(\sum_x p(x) |x\rangle\langle x| \right) \right] \\ & \quad + \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_{AB}^x \log \left((\rho_{AB}^x)^{\frac{1}{2}} I_A \otimes \rho_B^{-1} (\rho_{AB}^x)^{\frac{1}{2}} \right) \right]. \end{aligned}$$

For the first term of the last equality, we have

$$\begin{aligned} & \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_{AB}^x \cdot \log \left(\sum_x p(x) |x\rangle\langle x| \right) \right] \\ &= \sum_x [p(x) \log p(x) \cdot \text{Tr}[\rho_{AB}^x]] \\ &= -H(X), \end{aligned}$$

where the last equality follows from the fact that $\text{Tr}[\rho_{AB}^x] = 1$. Similarly, for the second term, we further have

$$\begin{aligned} & \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_{AB}^x \log \left((\rho_{AB}^x)^{\frac{1}{2}} I_A \otimes \rho_B^{-1} (\rho_{AB}^x)^{\frac{1}{2}} \right) \right] \\ &= \sum_x p(x) \text{Tr} \left[\rho_{AB}^x \log \left((\rho_{AB}^x)^{\frac{1}{2}} I_A \otimes \rho_B^{-1} (\rho_{AB}^x)^{\frac{1}{2}} \right) \right] \\ &= \sum_x p(x) \hat{D}(\rho_{AB}^x \| I_A \otimes \rho_B) \\ &= -\sum_x p(x) \hat{S}(A_x|B)_\rho, \end{aligned}$$

where $\hat{S}(A_x|B)_\rho$ is the BS conditional entropy for the quantum state ρ_{AB}^x .

Therefore, we can obtain

$$\text{Tr} \left[\rho_{XAB} \log \left(\rho_{XAB}^{\frac{1}{2}} I_{XA} \otimes \rho_B^{-1} \rho_{XAB}^{\frac{1}{2}} \right) \right] = -H(X) - \sum_x p(x) \hat{S}(A_x|B)_\rho.$$

Using the definition of the BS conditional entropy $\hat{S}(XA|B)_\rho$ to the above equality, we have

$$\hat{S}(XA|B)_\rho = H(X) + \sum_x p(x) \hat{S}(A_x|B)_\rho. \quad (27)$$

Applying Lemma 2, we further have

$$\hat{S}(XA|B)_\rho \leq \hat{S}(A|B)_\rho + \log d_X. \quad (28)$$

Substituting Equation (27) into the above inequality (28), we can obtain the first inequality of Theorem 1.

We will replace ρ_B with σ_B to analyze the case of the alternative BS relative entropy in the same way, i.e.,

$$\text{Tr} \left[\rho_{XAB} \log \left(\rho_{XAB}^{\frac{1}{2}} I_{XA} \otimes \sigma_B^{-1} \rho_{XAB}^{\frac{1}{2}} \right) \right] = -H(X) + \sum_x p(x) \hat{D}(\rho_{AB}^x \| I_A \otimes \sigma_B),$$

taking the optimization of σ_B for the above equality and combining the definition of Equation (15). We then obtain the desired result. \square

From the above results, we know that there are two additional terms $H(X)$ and $\log d_X$ on each side of the inequality (24), which are different from the concavity of the quantum conditional entropy $S(A|B)_\rho$. To make a distinction, we call it the weak concavity of the BS conditional entropy given by Theorem 1.

Combining the above fact with Theorem 1, we can establish the relationship between $\hat{S}(A|XB)_\rho$ and $\hat{S}(XA|B)_\rho$. Using the direct sum property of the BS relative entropy (Proposition 4.54 in [29]), we have

$$\hat{S}(A|XB)_\rho = - \sum_x p(x) \hat{D}(\rho_{AB}^x \| I_A \otimes \rho_B^x). \tag{29}$$

We cannot determine the order relations between ρ_B^x with ρ_B , but one can always compare $\hat{D}(\rho_{AB}^x \| I_A \otimes \rho_B^x)$ with $\hat{D}(\rho_{AB}^x \| I_A \otimes \rho_B)$. For the case that the former is less than the latter, we have

$$\hat{S}(A|XB)_\rho \geq - \sum_x p(x) \hat{D}(\rho_{AB}^x \| I_A \otimes \rho_B). \tag{30}$$

Applying Equation (26), we can obtain

$$\hat{S}(XA|B)_\rho - \log d_X \leq \hat{S}(A|XB)_\rho. \tag{31}$$

In addition, as a special case of the inequality (20), we can easily obtain that $\hat{S}(A|XB)_\rho \leq \hat{S}(A|B)_\rho$. Therefore, it holds that

$$\hat{S}(XA|B)_\rho - \log d_X \leq \hat{S}(A|XB)_\rho \leq \hat{S}(A|B)_\rho. \tag{32}$$

Otherwise, for the case of $\hat{D}(\rho_{AB}^x \| I_A \otimes \rho_B^x) \geq \hat{D}(\rho_{AB}^x \| I_A \otimes \rho_B)$, we then obtain

$$H(X) + \hat{S}(A|XB)_\rho \leq \hat{S}(XA|B)_\rho. \tag{33}$$

For the quantum conditional entropy of Equation (10), we know that any bipartite pure states are entangled if and only if $S(A|B) < 0$. Here, we are also interested in the BS conditional entropy. Without a loss of generality, let $|\psi\rangle_{AB} = \sum_i \lambda_i |i\rangle_A |i\rangle_B$ be any bipartite pure state, where λ_i represents non-negative real numbers satisfying $\sum_i \lambda_i^2 = 1$, known as Schmidt coefficients, and $|i\rangle_A$ and $|i\rangle_B$ are orthonormal states for A and B , respectively. The number of non-zero values λ_i is called the Schmidt number for the pure state $|\psi\rangle_{AB}$ [17]. We have

$$\begin{aligned} \hat{S}(A|B)_\psi &= -\text{Tr} \left[\psi_{AB} \log \left(\psi_{AB}^{\frac{1}{2}} I_A \otimes \rho_B^{-1} \psi_{AB}^{\frac{1}{2}} \right) \right] \\ &= -\text{Tr} \left[\psi_{AB} \log \left(I_A \otimes \rho_B^{-\frac{1}{2}} \psi_{AB} I_A \otimes \rho_B^{-\frac{1}{2}} \right) \right] \\ &= -\log r, \end{aligned} \tag{34}$$

where r is the Schmidt number of $|\psi\rangle$. We remark that the bipartite pure state is entangled if the Schmidt number is greater than 1.

3.2. Belavkin–Staszewski Mutual Information

The quantum mutual information is another important measure in quantum information theory, which can describe total correlations in the bipartite quantum subsystems, and there are important applications in quantum channel capacity, quantum cryptography, and quantum thermodynamics [17,35]. Based on the property of the quantum relative entropy, there are four equal definitions for the quantum mutual information, i.e.,

$$\begin{aligned}
 I(A;B)_\rho &= D(\rho_{AB} \parallel \rho_A \otimes \rho_B) \\
 &= \min_{\sigma_B} D(\rho_{AB} \parallel \rho_A \otimes \sigma_B) \\
 &= \min_{\sigma_A} D(\rho_{AB} \parallel \sigma_A \otimes \rho_B) \\
 &= \min_{\sigma_A, \sigma_B} D(\rho_{AB} \parallel \sigma_A \otimes \sigma_B),
 \end{aligned}
 \tag{35}$$

where the minimums are taken over all density operators σ_A and σ_B on quantum systems \mathcal{H}_A and \mathcal{H}_B , respectively. However, for other general relative entropies, these equalities do not hold in general, such as max-information [24]. In this section, we will consider a new mutual information via the BS relative entropy, Similar to Equation (35), we define four different BS information terms as follows.

Definition 3. For any quantum state $\rho_{AB} \in \mathcal{S}=(\mathcal{H}_{AB})$, the BS mutual information terms are defined as

$$\hat{I}^1(A;B)_\rho = \hat{D}(\rho_{AB} \parallel \rho_A \otimes \rho_B), \tag{36}$$

$$\hat{I}^2(A;B)_\rho = \min_{\sigma_B} \hat{D}(\rho_{AB} \parallel \rho_A \otimes \sigma_B), \tag{37}$$

$$\hat{I}^{2'}(A;B)_\rho = \min_{\sigma_A} \hat{D}(\rho_{AB} \parallel \sigma_A \otimes \rho_B), \tag{38}$$

$$\hat{I}^3(A;B)_\rho = \min_{\sigma_A, \sigma_B} \hat{D}(\rho_{AB} \parallel \sigma_A \otimes \sigma_B). \tag{39}$$

Notice that, for $\hat{I}^2(A;B)_\rho$ and $\hat{I}^{2'}(A;B)_\rho$, they can be thought of as swapping the positions of the optimization operators σ_A and σ_B , so we will consider only one of them. Intuitively, the remaining three definitions of the BS mutual information $\hat{I}^i(A;B)_\rho$ decreases with i , i.e.,

$$\hat{I}^3(A;B)_\rho \leq \hat{I}^2(A;B)_\rho \leq \hat{I}^1(A;B)_\rho. \tag{40}$$

Additionally, recalling the inequality (7), we can obtain that there is not less BS mutual information than there is quantum mutual information, i.e.,

$$I(A;B)_\rho \leq \hat{I}^i(A;B)_\rho. \tag{41}$$

From the monotonicity of the BS relative entropy, it follows that discarding quantum systems does not increase the BS mutual information, i.e.,

$$\hat{I}^i(A;B)_\rho \leq \hat{I}^i(A;BC)_\rho. \tag{42}$$

Subsequently, for the quantum mutual information (35), it holds that

$$\begin{aligned}
 I(A;B)_\rho &= S(\rho_A) - S(A|B)_\rho \\
 &= S(\rho_B) - S(B|A)_\rho.
 \end{aligned}
 \tag{43}$$

The above two relations are called chain rules for the quantum mutual information. Chain rule can be regarded as a ‘bridge’ between conditional entropy with mutual information in information theory. We are also interested in exploring chain rules for the BS mutual information for bipartite classical-quantum system \mathcal{H}_{XB} (for all quantum scenarios, further discussion is needed as a remaining issue). It is well-known that the classical-quantum

state only possesses classical correlation, and there is no quantum correlation, which leads us to find some significative and interesting results. We first give the following result before discussing the chain rules for the BS mutual information.

Lemma 3. For any bipartite classical-quantum states ρ_{XB} , we have

$$\hat{I}^i(X; B)_\rho = \min_{\sigma_B} \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B), \tag{44}$$

where $i = 2, 3$.

Proof. The proof is similar to the proof of Lemma 1, thus we omit some calculation steps. We first consider the case $\hat{I}^3(X; B)_\rho$. Let $\sigma_X = \sum_x q(x) |x\rangle\langle x|$, for any quantum state σ_B , we have

$$\begin{aligned} & \text{Tr} \left[\rho_{XB} \log \left(\rho_{XB}^{\frac{1}{2}} \sigma_X^{-1} \otimes \sigma_B^{-1} \rho_{XB}^{\frac{1}{2}} \right) \right] \\ = & \text{Tr} \left[\sum_x p(x) |x\rangle\langle x| \otimes \rho_B^x \cdot \log \left(\sum_x p(x) q(x)^{-1} |x\rangle\langle x| \otimes (\rho_B^x)^{\frac{1}{2}} \sigma_B^{-1} (\rho_B^x)^{\frac{1}{2}} \right) \right] \tag{45} \\ = & \text{Tr} \left[\sum_x p(x) \log p(x) q(x)^{-1} |x\rangle\langle x| \otimes \rho_B^x \right] + \sum_x p(x) \text{Tr} \left[\rho_B^x \log (\rho_B^x)^{\frac{1}{2}} \sigma_B^{-1} (\rho_B^x)^{\frac{1}{2}} \right] \\ = & D(p(x) \| q(x)) + \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B). \end{aligned}$$

Taking the minimum optimization for both sides of Equation (45) about σ_X and σ_B , respectively, we then have,

$$\begin{aligned} \hat{I}^3(X; B)_\rho &= \min_{\sigma_X, \sigma_B} \left[D(p(x) \| q(x)) + \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B) \right] \\ &= \min_{\sigma_X} D(p(x) \| q(x)) + \min_{\sigma_B} \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B) \\ &= \min_{\sigma_B} \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B). \end{aligned}$$

The last equality follows from the fact that the relative entropy $D(p(x) \| q(x))$ is non-negative; i.e., it holds that

$$\min_{\sigma_X} D(p(x) \| q(x)) = 0,$$

for all x , if and only if $p(x) = q(x)$.

For $\hat{I}^2(X; B)_\rho$, only optimization is required for σ_B from its definition, so we directly have

$$\begin{aligned} \hat{I}^2(X; B)_\rho &= \min_{\sigma_B} \text{Tr} \left[\rho_{XB} \log \left(\rho_{XB}^{\frac{1}{2}} \rho_X^{-1} \otimes \sigma_B^{-1} \rho_{XB}^{\frac{1}{2}} \right) \right] \\ &= \min_{\sigma_B} \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B). \end{aligned}$$

□

Notice that the BS mutual information $\hat{I}^1(X; B)_\rho$ does not involve any optimizations, so we have

$$\hat{I}^1(X; B)_\rho = \sum_x p(x) \hat{D}(\rho_B^x \| \rho_B). \tag{46}$$

This result shows that the BS mutual information $\hat{I}^1(X; B)_\rho$ is identical in form with the quantum mutual information $I(X; B)_\rho$, while the latter is the well-known Holevo

information. In addition, if we consider the tripartite classical-quantum state, we can also obtain a sum form of the BS mutual information for $i = 1$, i.e.,

$$\hat{I}^1(XA; B)_\rho = \sum_x p(x) \hat{I}^1(A_x; \rho_B), \tag{47}$$

where $\hat{I}^1(A_x; \rho_B)$ is the BS mutual information between quantum states ρ_{AB}^x and $\rho_A^x \otimes \rho_B$. Other cases of the BS mutual information are similar, so we will not go into detail. Based on the above results, we obtain the chain rules for the BS mutual information for bipartite classical-quantum states as follows.

Theorem 2. For any bipartite classical-quantum state ρ_{XB} , we have

$$\hat{I}^i(X; B)_\rho + \hat{S}_m(X|B)_\rho = H(X), \tag{48}$$

where $i = 2, 3$.

Proof. The proof of this theorem is similar to the proof of Theorem 1, so we omit some of the repetition. For any bipartite classical-quantum state ρ_{XB} , we have

$$\hat{D}(\rho_{XB} \| I_X \otimes \sigma_B) = \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B) - H(X). \tag{49}$$

Employing the definition of Equation (15), we further have

$$\hat{S}_m(X|B)_\rho = H(X) - \min_{\sigma_B} \sum_x p(x) \hat{D}(\rho_B^x \| \sigma_B). \tag{50}$$

Applying Lemma 3 to Equation (50), we can then complete the proof. \square

Similarly, for $\hat{I}^1(X; B)_\rho$, we give a chain rule with respect to the definition of the BS conditional entropy $\hat{S}(X|B)_\rho$ as follows.

Corollary 1. For any bipartite classical-quantum state $\rho_{XB} \in \mathcal{S}_{\leq}(\mathcal{H}_{XB})$, we have

$$\hat{I}^1(X; B)_\rho + \hat{S}(X|B)_\rho = H(X). \tag{51}$$

Proof. From Definition 2, we can directly obtain that

$$\begin{aligned} \hat{S}(X|B)_\rho &= -\text{Tr} \left[\rho_{XB} \log \left(\rho_{XB}^{\frac{1}{2}} I_X \otimes \rho_B^{-1} \rho_{XB}^{\frac{1}{2}} \right) \right] \\ &= H(X) - \sum_x p(x) \hat{D}(\rho_B^x \| \rho_B). \end{aligned} \tag{52}$$

Combining Equation (46) with Equation (52), we then obtain the desired result. \square

Recalling Holevo information and applying the result of Lemma 1, we further have

$$\begin{aligned} I(X; B)_\rho &= S(\rho_B) - \hat{S}(B|X)_\rho \\ &= S(\rho_B) - \hat{S}_m(B|X)_\rho \\ &= S(\rho_B) - S(B|X)_\rho. \end{aligned} \tag{53}$$

This result shows that the BS conditional entropy with classical side information can be used to describe the Holevo information as well. In addition, employing the inequality (41), we have

$$S(\rho_B) - \hat{S}_m(B|X)_\rho \leq \hat{I}^3(X; B)_\rho. \tag{54}$$

Comparing this to Theorem 2, we find that, when the side information is classical, the equal sign of the chain rule for $\hat{I}^3(X; B)_\rho$ does not hold in general.

3.3. Subadditivity for the BS Relative Entropy

It is necessary to study the relationship of entropic measures between a joint system and its corresponding subsystems, which plays a vital role in estimating channel capacity bounds and analyzing error exponents. The quantum relative entropy satisfies subadditivity and superadditivity, both of which are fundamental properties of the quantum relative entropy [31]. More precisely, for any bipartite quantum states ρ_{AB} and a product state $\sigma_A \otimes \sigma_B$, the superadditivity of the quantum relative entropy is

$$D(\rho_A \parallel \sigma_A) + D(\rho_B \parallel \sigma_B) \leq D(\rho_{AB} \parallel \sigma_A \otimes \sigma_B). \tag{55}$$

This was extended to a more general setting [36]. This paper does not determine whether the BS relative entropy holds the same property. However, the following result shows an opposite relationship for the BS entropy, i.e., the subadditivity. We first give an equivalent definition of the BS relative entropy for obtaining the desired result.

Lemma 4. For any quantum state $\rho \in \mathcal{S}_=(\mathcal{H})$ and $\sigma \in \mathcal{S}_\leq(\mathcal{H})$, we have

$$\hat{D}(\rho \parallel \sigma) = \text{Tr} \left[\sigma (\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \log \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \right) \right]. \tag{56}$$

Proof. Let Π_σ be the projection onto the support of σ . One can obtain that $\rho = \rho \Pi_\sigma = \Pi_\sigma \rho$ from $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. From Equation (5), we thus have

$$\begin{aligned} \hat{D}(\rho \parallel \sigma) &= \text{Tr} \left[\rho \log(\rho^{\frac{1}{2}} \sigma^{-1} \rho^{\frac{1}{2}}) \right] \\ &= \text{Tr} \left[\rho^{\frac{1}{2}} \Pi_\sigma \rho^{\frac{1}{2}} \log(\rho^{\frac{1}{2}} \sigma^{-1} \rho^{\frac{1}{2}}) \right] \\ &= \text{Tr} \left[\rho^{\frac{1}{2}} \sigma^{\frac{1}{2}} \sigma^{-\frac{1}{2}} \rho^{\frac{1}{2}} \log(\rho^{\frac{1}{2}} \sigma^{-\frac{1}{2}} \sigma^{-\frac{1}{2}} \rho^{\frac{1}{2}}) \right], \end{aligned} \tag{57}$$

where the equalities holds from the fact that $\Pi_\sigma = \sigma^{\frac{1}{2}} \sigma^{-\frac{1}{2}}$. Employing Lemma 2.6 in [29], we have

$$\sigma^{-\frac{1}{2}} \rho^{\frac{1}{2}} \log(\rho^{\frac{1}{2}} \sigma^{-\frac{1}{2}} \sigma^{-\frac{1}{2}} \rho^{\frac{1}{2}}) = \log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \sigma^{-\frac{1}{2}} \rho^{\frac{1}{2}},$$

where $\rho^{\frac{1}{2}} \sigma^{-\frac{1}{2}} = (\sigma^{-\frac{1}{2}} \rho^{\frac{1}{2}})^\dagger$. We then have

$$\begin{aligned} \hat{D}(\rho \parallel \sigma) &= \text{Tr} \left[\rho^{\frac{1}{2}} \sigma^{\frac{1}{2}} \log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \sigma^{-\frac{1}{2}} \rho^{\frac{1}{2}} \right] \\ &= \text{Tr} \left[\log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \sigma^{-\frac{1}{2}} \rho \sigma^{\frac{1}{2}} \right]. \end{aligned} \tag{58}$$

The equality holds from the cyclic property of the trace. Since

$$\log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} = \sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}),$$

we have

$$\begin{aligned} \hat{D}(\rho \parallel \sigma) &= \text{Tr} \left[\log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \sigma \right] \\ &= \text{Tr} \left[\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \sigma \right]. \end{aligned} \tag{59}$$

Finally, we obtain the desired result by applying the cyclic property of the trace again. \square

Theorem 3. For any quantum state $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$, $\sigma_A \in \mathcal{S}_\leq(\mathcal{H}_A)$, or $\sigma_B \in \mathcal{S}_\leq(\mathcal{H}_B)$, we have

$$\hat{D}(\rho_{AB} \parallel \sigma_A \otimes \sigma_B) \leq \lambda [\log \lambda + \hat{D}(\rho_A \parallel \sigma_A) + \hat{D}(\rho_B \parallel \sigma_B)], \tag{60}$$

where $\lambda = 2^{D_{\max}(\rho_{AB} \parallel \rho_A \otimes \rho_B)}$.

Proof. Let $D_{\max}(\rho_{AB} \parallel \rho_A \otimes \rho_B) = \log \lambda$ for $\text{supp} \rho_{AB} \subseteq \text{supp}(\rho_A \otimes \rho_B)$, which implies that $\rho_{AB} \leq \lambda \rho_A \otimes \rho_B$. For any quantum state σ_A or σ_B that satisfies $\text{supp} \rho_A \subseteq \text{supp} \sigma_B$ and $\text{supp} \rho_A \subseteq \text{supp} \sigma_B$, respectively, applying Lemma 4, we have

$$\begin{aligned} & \hat{D}(\rho_{AB} \parallel \sigma_A \otimes \sigma_B) \\ = & \text{Tr} \left[(\sigma_A \otimes \sigma_B) \left(\sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_{AB} \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \right) \cdot \log \left(\sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_{AB} \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \right) \right]. \end{aligned} \tag{61}$$

Employing the basic operator inequalities of Lemma 2.13 in [29], we have

$$\sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_{AB} \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \leq \lambda \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_A \otimes \rho_B \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}}.$$

Substituting the above inequality into Equation (61), we then have

$$\begin{aligned} & \hat{D}(\rho_{AB} \parallel \sigma_A \otimes \sigma_B) \\ \leq & \text{Tr} \left[(\sigma_A \otimes \sigma_B) \left(\lambda \sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \cdot \log \left(\lambda \sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \right] \\ = & \lambda \cdot \text{Tr} \left[(\sigma_A \otimes \sigma_B) \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \right. \\ & \left. \cdot \left(\log \lambda + \log \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \right) + \log \left(\sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \right) \right] \\ \leq & \lambda \log \lambda + \lambda \hat{D}(\rho_A \parallel \sigma_A) + \lambda \hat{D}(\rho_B \parallel \sigma_B). \end{aligned}$$

The equality follows from the linearity of the trace. The last inequality holds based on

$$\text{Tr} \left[(\sigma_A \otimes \sigma_B) \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \right] \leq 1$$

and

$$\begin{aligned} & \text{Tr} \left[(\sigma_A \otimes \sigma_B) \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \log \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \right) \right] \\ = & \text{Tr} \left[\sigma_A \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \right) \log \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \right) \right] \\ = & \hat{D}(\rho_A \parallel \sigma_A). \end{aligned}$$

Similarly,

$$\text{Tr} \left[(\sigma_A \otimes \sigma_B) \left(\sigma_A^{-\frac{1}{2}} \rho_A \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \log \left(\sigma_B^{-\frac{1}{2}} \rho_B \sigma_B^{-\frac{1}{2}} \right) \right] = \hat{D}(\rho_B \parallel \sigma_B).$$

□

As mentioned above, the geometric Rényi relative entropy converges to the BS relative entropy when the limit is $\alpha \rightarrow 1$. More generally, we also provide an upper bound for the geometric Rényi relative entropy.

Theorem 4. For any quantum state $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$, $\sigma_A \in \mathcal{S}_\leq(\mathcal{H}_A)$, or $\sigma_B \in \mathcal{S}_\leq(\mathcal{H}_B)$, we have

$$\hat{D}_\alpha(\rho_{AB} \parallel \sigma_A \otimes \sigma_B) \leq \frac{\alpha}{\alpha - 1} D_{\max}(\rho_{AB} \parallel \rho_A \otimes \rho_B) + \hat{D}_\alpha(\rho_A \parallel \sigma_A) + \hat{D}_\alpha(\rho_B \parallel \sigma_B).$$

Proof. Let $D_{\max}(\rho_{AB} \parallel \rho_A \otimes \rho_B) = \log \lambda$ for $\text{supp} \rho_{AB} \subseteq \text{supp}(\rho_A \otimes \rho_B)$. This implies that $\rho_{AB} \leq \lambda \rho_A \otimes \rho_B$. For σ_A and σ_B with $\text{supp} \rho_A \subseteq \text{supp} \sigma_B$ and $\text{supp} \rho_B \subseteq \text{supp} \sigma_A$, respectively, employing the relation of Eq. (4.6.21) in [29], we have

$$\text{Tr} \left[(\sigma_A \otimes \sigma_B) \left(\sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \rho_{AB} \sigma_A^{-\frac{1}{2}} \otimes \sigma_B^{-\frac{1}{2}} \right)^\alpha \right] = \text{Tr} \left[\rho_{AB} \left(\rho_{AB}^{\frac{1}{2}} \sigma_A^{-1} \otimes \sigma_B^{-1} \rho_{AB}^{\frac{1}{2}} \right)^{\alpha-1} \right].$$

It then holds that

$$\begin{aligned} & \text{Tr} \left[\rho_{AB} \left(\rho_{AB}^{\frac{1}{2}} \sigma_A^{-1} \otimes \sigma_B^{-1} \rho_{AB}^{\frac{1}{2}} \right)^{\alpha-1} \right] \\ & \leq \text{Tr} \left[\lambda (\rho_A \otimes \rho_B) \left(\lambda \rho_A^{\frac{1}{2}} \otimes \rho_B^{\frac{1}{2}} \sigma_A^{-1} \otimes \sigma_B^{-1} \rho_A^{\frac{1}{2}} \otimes \rho_B^{\frac{1}{2}} \right)^{\alpha-1} \right] \\ & = \lambda^\alpha \text{Tr} \left[(\rho_A \otimes \rho_B) \left(\rho_A^{\frac{1}{2}} \otimes \rho_B^{\frac{1}{2}} \sigma_A^{-1} \otimes \sigma_B^{-1} \rho_A^{\frac{1}{2}} \otimes \rho_B^{\frac{1}{2}} \right)^{\alpha-1} \right] \\ & = \lambda^\alpha \text{Tr} \left[\rho_A \left(\rho_A^{\frac{1}{2}} \sigma_A^{-1} \rho_A^{\frac{1}{2}} \right)^{\alpha-1} \right] \text{Tr} \left[\rho_B \left(\rho_B^{\frac{1}{2}} \sigma_B^{-1} \rho_B^{\frac{1}{2}} \right)^{\alpha-1} \right]. \end{aligned}$$

Combining the above result with Definition 1, we have

$$\begin{aligned} \hat{D}_\alpha(\rho_{AB} \parallel \sigma_A \otimes \sigma_B) & \leq \frac{1}{\alpha - 1} \log \left[\lambda^\alpha \text{Tr} \left[\rho_A \left(\rho_A^{\frac{1}{2}} \sigma_A^{-1} \rho_A^{\frac{1}{2}} \right)^{\alpha-1} \right] \text{Tr} \left[\rho_B \left(\rho_B^{\frac{1}{2}} \sigma_B^{-1} \rho_B^{\frac{1}{2}} \right)^{\alpha-1} \right] \right] \\ & = \frac{\alpha}{\alpha - 1} D_{\max}(\rho_{AB} \parallel \rho_A \otimes \rho_B) + \hat{D}_\alpha(\rho_A \parallel \sigma_A) + \hat{D}_\alpha(\rho_B \parallel \sigma_B). \end{aligned}$$

□

Similarly, one can define a new mutual information term via the geometric Rényi relative entropy as

$$\hat{I}_\alpha(A; B)_\rho = \min_{\sigma_A, \sigma_B} \hat{D}_\alpha(\rho_{AB} \parallel \sigma_A \otimes \sigma_B). \tag{62}$$

It is then easy to draw the following conclusion.

Corollary 2. For any quantum state $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$, $\sigma_A \in \mathcal{S}_\leq(\mathcal{H}_A)$, or $\sigma_B \in \mathcal{S}_\leq(\mathcal{H}_B)$, we have

$$\hat{I}_\alpha(A; B)_\rho \leq \frac{\alpha}{\alpha - 1} D_{\max}(\rho_{AB} \parallel \rho_A \otimes \rho_B) + \hat{D}_\alpha(\rho_A \parallel \sigma_A) + \hat{D}_\alpha(\rho_B \parallel \sigma_B). \tag{63}$$

Notably, if one considers the classical-quantum state, there is no result as shown in Lemma 3 for the mutual information defined by the geometric Rényi relative entropy. Specifically, for $\alpha \in (0, 1)$, we have

$$\begin{aligned} \hat{D}_\alpha(\rho_{XB} \parallel \rho_X \otimes \rho_B) & = \frac{1}{\alpha - 1} \log \left[\sum_x p(x) \text{Tr} \left[\rho_B \left(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \right)^\alpha \right] \right] \\ & \leq \frac{1}{\alpha - 1} \sum_x p(x) \log \left[\text{Tr} \left[\rho_B \left(\sigma_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \right)^\alpha \right] \right] \\ & = \sum_x p(x) \hat{D}_\alpha(\rho_B^x \parallel \rho_B), \end{aligned} \tag{64}$$

where the inequality comes from the Jensen inequality of $-\log t$. For $\alpha \in (1, \infty)$, we obtain the opposite result, i.e.,

$$\hat{D}_\alpha(\rho_{XB} \parallel \rho_X \otimes \rho_B) \geq \sum_x p(x) \hat{D}_\alpha(\rho_B^x \parallel \rho_B). \tag{65}$$

Furthermore, if one considers the conditional entropy defined by the geometric Rényi relative entropy, for $\alpha \in (0, 1)$, we have

$$\hat{D}_\alpha(\rho_{XB} \parallel I_X \otimes \rho_B) \leq \sum_x p(x) \hat{D}_\alpha(\rho_B^x \parallel \rho_B) - H(X). \tag{66}$$

For $\alpha \in (1, \infty)$, we then have

$$\hat{D}_\alpha(\rho_{XB} \parallel I_X \otimes \rho_B) \geq \sum_x p(x) \hat{D}_\alpha(\rho_B^x \parallel \rho_B) - H(X). \tag{67}$$

4. Conclusions

This paper investigates the subadditivity of the geometric Rényi and BS relative entropies and explores the indispensable properties of the BS conditional entropy and mutual information, especially in classical-quantum settings. The subadditivity of the geometric Rényi and BS relative entropies can provide new valuable bounds to estimate channel capacity and analyze the error exponent. As mentioned above, the BS relative entropy represents a different quantum generalization of classical relative entropy. The main use of BS relative entropy is in establishing upper bounds for the rates of feedback-assisted quantum communication protocols. The primary goal of further research on BS relative entropy is to explore the intrinsic properties of its relevant conditional entropy and mutual information and to gain a better understanding of their operational relevance. We hope that the formal tools provided in this paper will be useful for this purpose.

One question worth answering is whether there is a chain rule for the mutual information in terms of the geometric Rényi relative entropy, i.e.,

$$\hat{D}_\alpha(\rho_{AB} \parallel \rho_A \otimes \rho_B) \stackrel{\geq}{\leq} \hat{S}_\alpha(\rho_A) + \hat{D}_\alpha(\rho_{AB} \parallel I_A \otimes \rho_B),$$

or the other forms, where $\hat{S}_\alpha(\rho_A)$ is the quantum Rényi entropy. Subsequently, the duality of conditional entropy is an important property for a tripartite pure state system, which can be effectively applied in random number extraction and channel coding [26]. Further research will focus on the duality of the BS conditional entropy.

Author Contributions: Conceptualization, Y.Z. and Z.X.; methodology, Y.Z. and Z.X.; formal analysis, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., B.Y. and Z.X.; project administration, B.Y. and Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Nos. 62171266, U2001205) and the Funded Projects for the Academic Leaders and Academic Backbones, Shaanxi Normal University (No. 16QNGG013).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Z.X. is supported by the National Natural Science Foundation of China (No. 62171266) and by the Funded Projects for the Academic Leaders and Academic Backbones, Shaanxi Normal University (No. 16QNGG013). B.Y. is supported by the National Natural Science Foundation of China (No. U2001205).

Conflicts of Interest: The authors declare that there is no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

- Rényi, A. On measures of information and entropy. *Proc. Symp. Math. Stat. Probab.* **1961**, *4*, 547–561.
- Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
- Cover, T. M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley Sons, Inc.: Hoboken, NJ, USA, 2006.
- Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
- Tomamichel, M. Quantum information processing with finite resources. *arXiv* **2015**, arXiv:1504.00233v1.
- Muller-Lennert, M.; Dupuis, F.; Szehr, O.; Fehr, S.; Tomamichel, M. On quantum Rényi entropies: A new generalization and some properties. *J. Math. Phys.* **2013**, *54*, 122203. [[CrossRef](#)]
- Mosonyi, M.; Ogawa, T. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. *Commun. Math. Phys.* **2015**, *334*, 1617–1648. [[CrossRef](#)]
- Audenaert, K.M.R.; Datta, N. α -z-Rényi relative entropies. *J. Math. Phys.* **2015**, *56*, 022202. [[CrossRef](#)]
- Hiai, F.; Mosonyi, M. Different quantum f -divergences and the reversibility of quantum operations. *Rev. Math. Phys.* **2017**, *29*, 1750023. [[CrossRef](#)]
- Capel, Á.; Lucia, A.; Pérez-García, D. Quantum conditional relative entropy and quasifactorization of the relative entropy. *J. Phys. A Math. Theor.* **2018**, *51*, 484001. [[CrossRef](#)]
- Petz, D. Quasi-entropies for states of a von Neumann algebra. *Publ. Res. Inst. Math. Sci.* **1985**, *4*, 787–800. [[CrossRef](#)]
- Petz, D. Quasi-entropies for finite quantum systems. *Rep. Math. Phys.* **1986**, *23*, 57–65. [[CrossRef](#)]
- Wilde, M.M.; Winter, A.; Yang, D. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Commun. Math. Phys.* **2014**, *331*, 593–622. [[CrossRef](#)]
- Petz, D.; Ruskai, M.B. Contraction of generalized relative entropy under stochastic mappings on matrices. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1998**, *1*, 83–89. [[CrossRef](#)]
- Matsumoto, K. A new quantum version of f -divergence. *arXiv* **2013**, arXiv:1311.4722.
- Umegaki, H. Conditional expectations in an operator algebra IV (entropy and information). *Kodai Math. J.* **1962**, *14*, 59–85. [[CrossRef](#)]
- Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*; Cambridge University Press: Cambridge, UK, 2000.
- Belavkin, V.P.; Staszewski, P. C^* -algebraic generalization of relative entropy and entropy. *Ann. Inst. Henri Poincaré* **1982**, *37*, 51–58.
- Katariya, V.; Wilde, M.M. Geometric distinguishability measures limit quantum channel estimation and discrimination. *arXiv* **2020**, arXiv:2004.10708.
- Bluhm, A.; Capel, A. A strengthened data processing inequality for the Belavkin-Staszewski relative entropy. *Rev. Math. Phys.* **2020**, *32*, 2050005. [[CrossRef](#)]
- Hiai, F.; Petz, D. The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.* **1991**, *143*, 99–114. [[CrossRef](#)]
- Bluhm, A.; Capel, A.; Perez-Hernandez, A. Weak quasi-factorization for the Belavkin-Staszewski relative entropy. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT), Melbourne, Australia, 12–20 July 2021; pp. 118–123.
- Fang, K.; Fawzi, H. Geometric Rényi divergence and its applications in quantum channel capacities. *Commun. Math. Phys.* **2021**, *384*, 1615–1677. [[CrossRef](#)]
- Ciganović, N.; Beaudry, N.J.; Renner, R. Smooth max-information as one-shot generalization for mutual information. *IEEE Trans. Inform. Theory* **2014**, *60*, 1573–1581. [[CrossRef](#)]
- Watanabe, Y. Randomness Extraction via a Quantum Generalization of the Conditional Collision Entropy. *IEEE Trans. Inform. Theory* **2020**, *66*, 1171–1177. [[CrossRef](#)]
- Tomamichel, M.; Colbeck, R.; Renner, R. Duality between smooth min- and max-entropies. *IEEE Trans. Inf. Theory* **2010**, *56*, 4674–4681. [[CrossRef](#)]
- König, R.; Renner, R.; Schaffner, C. The operational meaning of min- and max-entropy. *IEEE Trans. Inf. Theory* **2009**, *55*, 4337–4347. [[CrossRef](#)]
- Vitanov, A.; Dupuis, F.; Tomamichel, M.; Renner, R. Chain rules for smooth min- and max-entropies. *IEEE Trans. Inf. Theory* **2013**, *59*, 2603–2612. [[CrossRef](#)]
- Khatri, S.; Wilde, M.M. Principles of quantum communication theory: A modern approach. *arXiv* **2020**, arXiv:2011.04672v1.
- Datta, N. Min- and max-relative entropies and a new entanglement monotone. *IEEE Trans. Inf. Theory* **2009**, *55*, 2816–2826. [[CrossRef](#)]
- Wilde, M.M. *Quantum Information Theory*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2017.
- Chitambar, E.; Gour, G. Quantum resource theories. *Rev. Mod. Phys.* **2019**, *91*, 025001. [[CrossRef](#)]
- Lawson, J.D.; Lim, Y. The geometric mean, matrices, metrics, and more. *Am. Math. Mon.* **2001**, *108*, 797–812. [[CrossRef](#)]
- Renner, R. Security of Quantum Key Distribution. Ph.D. Thesis, ETH Zurich, Zurich, Switzerland, 2005.
- Sagawa, T. Entropy, divergence, and majorization in classical and quantum thermodynamics. *arXiv* **2020**, arXiv:2007.09974v1.
- Capel, Á.; Lucia, A.; Pérez-García, D. Superadditivity of quantum relative entropy for general states. *IEEE Trans. Inf. Theory* **2018**, *64*, 4758–4765. [[CrossRef](#)]