A Correlation Measure Based on Vector-Valued $L_p$ Norms

M.M. Mojahedian*, S. Beigi†, A. Gohari*, M.H. Yassaee† and M.R. Aref*

*Information Systems and Security Lab. (ISSL), Sharif University of Technology, Tehran, Iran
†School of Mathematics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Abstract—In this paper, a new measure of correlation is introduced. This measure depends on a parameter $\alpha$, and is defined in terms of vector-valued $L_p$ norms. The measure is within a constant of the exponential of $\alpha$-Rényi mutual information, and reduces to the trace norm (total variation distance) for $\alpha = 1$. We provide some properties and applications of this measure of correlation. In particular, we establish a bound on the secrecy exponent of the wiretap channel (under the total variation metric) in terms of the $\alpha$-Rényi mutual information according to Csiszár’s proposal.

A full version of this paper is accessible at Arxiv [1]

I. INTRODUCTION

Mutual information, Rényi mutual information of order $\alpha$, and the total variation distance $\|p_{AB} - p_{AB}\|_1$ are well-known examples of correlation measures. The Rényi mutual information of order $\alpha$ (Sibson’s proposal) for $\alpha \in (0, 1) \cup (1, \infty)$ is defined as follows:

$$I_\alpha(A; B) = \frac{\alpha}{\alpha - 1} \log \left( \sum_b \left( \sum_a p_A(a) p_B|A|(b|a)^\alpha \right)^{1/\alpha} \right).$$

(1)

In this paper, we introduce a new measure of correlation $V_\alpha(A; B)$ for a given bipartite distribution $p_{AB}$ as follows:

$$V_\alpha(A; B) = \sum_b \left( \sum_a p_A(a) p_B|A|(b|a) - p_B(b)\right)^{1/\alpha}. $$

(2)

The full version of this paper in [1] defines this measure for quantum bipartite states based on vector-valued $L_p$ norms. The point is that $V_\alpha(A; B)$ equals the 1-norm of the function $g(b) = \|p_A^{1/\alpha} p_{B|A} - p_B p_A^{1/\alpha}\|_\alpha$. This observation is used in the full version of this paper to generalize the definition of $V_\alpha$ to the non-commutative case, and to study its properties based on the theory of vector-valued $L_p$ norms. In this paper, for simplicity of exposition we restrict to the classical case in which case the desired properties of $V_\alpha$ can be proven directly without referring to this theory.

A main motivation for introducing the new measure of correlation is its application in decoupling and privacy enhancement theorems. We show that the average of $V_\alpha(A_0; B)$, when $p_{A_0,B}$ is the outcome of a certain random map $f(A) = A_0$, can be bounded by $c V_\alpha(A; B)$ where $c < 1$ is a constant.

The outline of this paper is as follows: Section II presents some of the properties of our new measure of correlation. In Section III, we study the special case of $\alpha = 2$. The application of the new measure in bounding the random coding exponents is discussed in Section IV.

II. PROPERTIES OF THE NEW CORRELATION METRIC

$V_\alpha(A; B)$ as defined in (2) for a joint pmf $p_{AB}$ has the following properties (see [1] for the proofs):

- For any $p_{AB}$ the function $\alpha \mapsto V_\alpha(A; B)$ is non-decreasing. In particular, for any $\alpha \geq 1$ we have

$$V_\alpha(A; B) \geq V_1(A; B) = \|p_{AB} - p_{AB}\|_1. $$

(3)

- For any $p_{AB}$ and all Markov chains $X - A - B - Y$ we have

$$V_\alpha(X; Y) \leq V_\alpha(A; B). $$

- For any bipartite density matrix $p_{AB}$ we have

$$2^{-\frac{1}{\alpha}I_\alpha(A; B)} - 1 \leq V_\alpha(A; B) \leq 2^{-\frac{1}{\alpha}I_\alpha(A; B)} + 1, $$

(4)

where $\alpha'$ is the H"{o}lder conjugate of $\alpha$.

The above relation between $V_\alpha$ and Rényi mutual information shows that $V_\alpha(A; B)$ is approximately $2^{-\frac{1}{\alpha}I_\alpha(A; B)}$ for any $\alpha > 1$. On the other hand $V_1(A; B) = \|p_{AB} - p_{AB}\|_1$ is in terms of a total variation distance. Therefore, $V_\alpha$ interpolates between the total variation distance and the exponent of the Rényi mutual information.

The conditional version of the correlation can be also defined as

$$V_\alpha(A; B|C) = \sum_c p(c) V_\alpha(A; B|C = c).$$

Then, we have the following result

**Theorem 1.** Let $p_{ABC}$ be a tripartite pmf and assume that $p_{AC}$ is uniform distribution. Then for any $1 \leq \alpha \leq 2$ we have

$$V_\alpha(A; B|C) \leq 2^{\frac{1}{\alpha} - 1} V_\alpha(AC; B).$$

The proof can be found in [1].

Finally, we have the following desirable property of $V_\alpha$ which is a key property we use in our application. We say that a function is $k$-to-1 if the pre-image of each output symbol is a set of size $k$.

---

This work was partially supported by Iran National Science Foundation (INSF) under contract No. 92/32575.

1See [2] for different definitions and properties of the Rényi mutual information.
Theorem 2. Let $A = A_0 \times C$ and $B$ be arbitrary finite sets. Suppose that $p(a) = 1/|A|$ is the uniform distribution, and $p_{B|A}$ be any arbitrary conditional distribution. Then we have

$$
\mathbb{E}_F[V_0(A; B)] \leq 2^{\frac{k}{2} - 1}|C|^{-\frac{k}{2}} V_0(A; B),
$$

where the expectation is taken with respect to the uniform distribution over all $|C|$-to-1 functions $f : A \to A_0$. Here, we prove the theorem for the special case $\alpha = 2$. Let $F$ be the set of all $|C|$-to-1 functions $f : A \to A_0$. Let $F \in F$ be uniformly chosen at random. The conditional pmf of $p_{B|A}$ is itself random and depends on the choice of random mapping as follows:

$$
P(b|a) = \frac{p(b|a)}{p(a)} = \frac{1}{|C|} \sum_a p(b|a) \mathbb{1}[F(a) = a_0],
$$

where $\mathbb{1}[:]$ is the indicator function and $(a)$ comes from this fact that both $A$ and $A_0$ are uniform in their alphabet sets. For simplicity, assume that $A_0 = \{1, 2, \ldots, |A_0|\}$.

Now, we bound $V_2(A_0; B)$ as follows

$$
\mathbb{E}_F V_2(A_0; B) = \mathbb{E}_F \sum_b \sqrt{\sum_a \frac{1}{|A_0|} (p(b|a) - p(b))^2} \leq \sum_b \sqrt{\mathbb{E}_F (p(b|a_0) - p(b))^2} = \sum_b \sqrt{\mathbb{E}_F (p(b|a_0) - p(b))^2},
$$

where (6) is due to the linearity of expectation, and Jensen’s inequality for concave function $\sqrt{x}$. Equation (7) follows from symmetry of the random mapping $F$. Equation (8) is obtained by substituting equation (5) into equation (7). We now compute the expression in (8):

$$
\mathbb{E}_F \left( \sum_a (p(b|a) - p(b)) \mathbb{1}[F(a) = 1] \right)^2 = \mathbb{E}_F \sum_{a, \tilde{a}} (p(b|a) - p(b)) \cdot (p(b|\tilde{a}) - p(b)) \times \mathbb{1}[F(a) = 1] \cdot \mathbb{1}[F(\tilde{a}) = 1]
$$

$$
= \mathbb{E}_F \sum_a (p(b|a) - p(b))^2 \cdot \mathbb{1}[F(a) = 1] + \mathbb{E}_F \sum_{a \neq \tilde{a}} (p(b|a) - p(b)) \cdot (p(b|\tilde{a}) - p(b))
$$

$$
= \sum_a (p(b|a) - p(b))^2 \cdot \mathbb{E}_F \mathbb{1}[F(a) = 1] + \sum_{a \neq \tilde{a}} (p(b|a) - p(b)) \cdot (p(b|\tilde{a}) - p(b))
$$

$$
\leq \frac{|C|}{|A|} \sum_a (p(b|a) - p(b))^2,
$$

where (9) follows from the fact that

$$
\mathbb{E}_F \mathbb{1}[F(a) = 1] = \mathbb{P}[F(a) = 1] = \frac{1}{|A_0|} = \frac{|C|}{|A|},
$$

and

$$
\sum_a (p(b|a) - p(b)) \cdot (p(b|\tilde{a}) - p(b)) = \sum_{a, \tilde{a}} (p(b|a) - p(b)) \cdot (p(b|\tilde{a}) - p(b))
$$

$$
- \sum_a (p(b|a) - p(b))^2
$$

$$
\leq \left( \sum_a (p(b|a) - p(b)) \right)^2 = 0,
$$

as from the uniform distribution on $A$ we have $\sum_a (p(b|a) - p(b)) = |A| \cdot \sum_a (p(a)p(b|a) - p(a)p(b)) = 0$. Substituting (9) in (8), we have:

$$
\mathbb{E}_F V_2(A_0; B) \leq \frac{1}{\sqrt{|C|}} \sum_b \sqrt{\frac{1}{|A|} \sum_a (p(b|a) - p(b))^2} = \frac{1}{\sqrt{|C|}} V_2(A; B).
$$

The ratio $|C|^{-\frac{k}{2}}$ in Theorem 2 is asymptotically tight up to a constant as the following example shows:

**Example 1.** Let $p_A$ be uniform on $A$, and $p_{B|A}$ be a classical erasure channel, i.e., the alphabet set of $B$ is $B = \{e\} \cup A$ and for all $a \in A$,

$$
p_{B|A}(e|a) = \epsilon, \quad p_{B|A}(a'|a) = 1 - \epsilon,
$$

and $p_{B|A}(a'|a) = 0$ if $a' \neq a$. Then a direct calculation shows that

$$
V_\alpha(A; B) = |A|^{\frac{1}{\alpha}} \cdot (1 - \epsilon) \left( 1 - \frac{1}{|A_0|} \right)^\alpha + \left( |A| - 1 \right) \frac{1}{|A_0|^\alpha}.
$$

Furthermore, for any $|C|$-to-1 function $f : A \to A_0$ with $A_0 = f(A)$ we have

$$
V_\alpha(A_0; B) = |A_0|^{\frac{1}{\alpha}} \cdot (1 - \epsilon) \left( 1 - \frac{1}{|A_0|} \right)^\alpha + \left( |A_0| - 1 \right) \frac{1}{|A_0|^\alpha}.
$$

1133
Hence, for fixed $|C| = \frac{|A|}{|A_0|}$ when $|A|$ tends to infinity we have
\[
\lim_{|A| \to \infty} \min_{f:|C|\to1} V_\alpha(A_0; B) \leq |C|^\frac{\alpha}{\alpha-1},
\]
for any $\alpha > 1$. Thus Theorem 2 is asymptotically tight up to a constant.

From Theorem 2 and utilizing (3) and (4), we obtain that
\[
\mathbb{E}[\|p_{AB} - p_{A_0B}\|_1] \leq \frac{2}{2^{1/\alpha} - 1} |C|^{1/\alpha} \left(2^{1/\alpha} I_{\alpha}(A_0; B) + 1\right). \tag{10}
\]

For the special case of $\alpha = 2$, inequality (10) can also be obtained by the result of Renner on privacy amplification [3, Theorem 5.5.1] for classical-quantum systems (see also [4]). Nevertheless, (10) is stronger than Renner’s result, at least in the fully classical case. While Renner’s result works only for $\alpha = 2$, equation (10) allows for all orders $\alpha \in (1, 2]$. On the other hand, Renner’s result is more general because it does not assume uniform distribution on the random variable $A$.

Another related result is in Hayashi’s work on privacy amplification [5, Theorem 1]. Hayashi uses a different definition of conditional Rényi entropy than the one used in this paper and obtains a different result. A detailed discussion can be found in [1].

III. Special Case of $\alpha = 2$

We focus on the special case $\alpha = 2$ here, and find equivalent expressions for $V_2(A; B)$. From (2) we have
\[
V_2(A; B) = \sum_b \sqrt{\sum_a p(a) \left(p(b|a) - p(b)\right)^2}. \tag{11}
\]

Given any realization $b \in B$, we can view $p_{0|A}$ as a random variable (a function of the random variable $A$ with $p_{0|A}(a) = p(b|a)$). Thus, (11) is equivalent with
\[
V_2(A; B) = \sum_b \sqrt{\text{Var}_A \left[p_{0|A}\right]}.
\]

Furthermore, using the Bayes’ rule one obtains
\[
V_2(A; B) = \sum_b p(b) \sqrt{\sum_a p^2(a|b) / p(a) - 1} = \mathbb{E}_B \left[\sqrt{\chi^2 \left(p_{A|B} \parallel p_A\right)}\right]. \tag{12}
\]

where $\chi^2(\cdot)$ is the chi-square distance.

A. $V_2(A; B)$ as the Tsallis mutual information

$V_2(A; B)$ can be also understood in terms of f-information. Given a convex function $f$ satisfying $f(1) = 0$ and two distributions $p(x)$ and $q(x)$ on a discrete space $X$, the f-divergence between $p$ and $q$ is defined as
\[
D_f(p || q) = \sum_x q(x) f \left(\frac{p(x)}{q(x)}\right).
\]

There are two proposals for defining a mutual information in terms of such a divergence. The first one given in [6, Eq. 3.10.1] is
\[
I^KZ_t(A; B) := D_f(p_{AB} || p_A \times p_B),
\]
and has been studied in the literature (e.g. see [7, Theorem 5.2], [8]). Another definition is given in [9, Eq. 79]:
\[
I^KZ_t(A; B) := \min_{q_B} D_f(p_{AB} \parallel p_A \times q_B).
\]

Herein, we propose yet a new definition of mutual f-information. Given a convex function $f$, we define its mutual f-information by
\[
I_f(A; B) := \min_{q_B} D_f(p_{AB} \parallel p_A \times q_B) = \min_{q_B} \sum_a p(a) D_f(p_{B|a} \parallel q_B) - D_f(p_B \parallel q_B). \tag{13}
\]

Observe that our mutual f-information is smaller than the previous ones:
\[
I_f(A; B) \leq I^{PV}_f(A; B) \leq I^{CZ}_t(A; B).
\]

Moreover, when $D_f(\cdot || \cdot)$ is the KL divergence, $I_f(A; B)$ reduces to Shannon’s mutual information.

Since we expect mutual f-information to satisfy the data processing inequality, we impose a further assumption on the convex function $f$. Interestingly, this assumption is the same as the one that gives the subadditivity of the $\Phi$-entropy (see [10][Exercise 14.2]).

Definition 1. Define $\mathcal{F}$ be the class of convex functions $f(t)$ on $[0, \infty)$ that are not affine (not of the form $t \mapsto at + b$ for some constants $a$ and $b$), $f(1) = 0$, and $1/f''$ is concave.

Examples of functions in $\mathcal{F}$ include $f(t) = t \log t$ and $f(t) = \frac{1}{\alpha-1}(t^{\alpha} - 1)$ for $\alpha \in (1, 2]$.

Theorem 3. For any function $f \in \mathcal{F}$, the mutual f-information $I_f(A; B)$ satisfies the followings:
(i) $I_f(A; B) = 0$ if and only if $A$ and $B$ are independent.
(ii) If $C = A - B - D$ forms a Markov chain, then $I_f(A; B) \geq I_f(C; D)$.

The proof can be found in [1].

Let $f_\alpha(t) = \frac{1}{\alpha-1}(t^{\alpha} - 1)$ for $\alpha \in (1, 2]$. As mentioned above, this function belongs to $\mathcal{F}$. Then, following (13) we can define the Tsallis mutual information of order $\alpha$ by
\[
I_\alpha(A; B) = \frac{1}{\alpha-1} \min_{q_B} \left\{ \sum_a p(a) \left(\sum_b q(b)^{1-\alpha} p(b|a)^{\alpha}\right) - \sum_b q(b)^{1-\alpha} p(b)^{\alpha}\right\}. \tag{14}
\]

The reason that we call it Tsallis mutual information is that the Tsallis relative entropy can be defined in terms of the function $f_\alpha$.

Theorem 4. The Tsallis mutual information defined in (14) equals
\[
I_\alpha(A; B) = \frac{1}{\alpha - 1} \left(\sum_b \left(\sum_a p(a)p(b|a)^\alpha - p(b)^\alpha\right)^{\frac{1}{\alpha}}\right)^{\frac{1}{\alpha}}.
\]

In particular, we have
\[
\sqrt{I_\alpha(A; B)} = V_2(A; B).
\]

The proof can be found in [1].
IV. AN APPLICATION

Herein, we only present one application, of finding a secrecy exponent for the wiretap channel (other applications can be found in [1]). In order to address the wiretap channel in Section IV-B, we first develop a tool in Section IV-A.

A. Statistics of random binning

Let \((A^n, B^n)\) be i.i.d. classical random variables distributed according to \(p_{AB}\):

\[
p(a^n b^n) = \prod_{i=1}^{n} p(a_i b_i).
\]

Suppose that we randomly (and uniformly) bin the set \(A^n\) into \(2^n R\) bins and let \(A_0\) denote the bin index. Finding the correlation between the bin index \(A_0\) and \(B^n\) (averaged over all random bin mappings) is of interest, see [11]. It is known that if the binning rate \(R\) is below the Slepian-Wolf rate, i.e., \(R < H(A|B)\), the average total variation distance \(\|p_{A^n B^n} - p_{A_0} \times p_{B^n}\|\) vanishes asymptotically as \(n\) tends to infinity.

Here we are interested in the same question as above when we replace \(V_1(A_0; B^n)\) with the correlation measure \(V_n(A_0; B^n)\) for some \(\alpha \in [1,2]\). Our tool for answering this question is Theorem 2, yet this theorem is applicable only if the first variable is distributed uniformly. For this reason, we do not assume that \(A^n\) is i.i.d., but is completely uniform on a type set.

Let \(p_{AB}\) be a bipartite distribution such that \(p_a(a)\) is a rational number for all \(a \in A\). In the following, let \(n\) be a natural number such that \(np(a)\) is an integer for all \(a \in A\). For each symbol \(a \in A\), let \(T_n(p_A) \subseteq A^n\) be the set of all sequences \(a^n\) of length \(n\) whose empirical distribution (type) is equal to \(p_a, i.e.,\) each symbol \(a' \in A\) occurs \(np(a')\) times in sequence \(a^n\). Instead of the i.i.d. distribution on \(A^n\), let \(a^n\) be uniformly distributed over \(T_n(p_A)\). The conditional distribution of \(B^n\) given \(A^n\) is still assumed to be

\[
p(b^n|a^n) = \prod_{i=1}^{n} p(b_i|a_i).
\]

For random binning, we use a randomly chosen \(k\)-to-1 function \(f\) on \(T_n(p_A) \subseteq A^n\) and let \(A_0 = f(A^n)\). We call this a regular random binning. This corresponds to a binning procedure with rate

\[
R = \frac{1}{n} \log \left( \frac{|T_n(p_A)|}{k} \right). \tag{15}
\]

**Theorem 5.** Let \(A^n\) be uniformly distributed over \(T_n(p_A)\) and \(p_{B^n|A^n} = \prod_{i=1}^{n} p_{b_i|a_i}\). Also let \(k\) be an integer that divides \(|T_n(p_A)|\) and define \(R\) by (15). Then for every \(\alpha \in (1,2]\) we have

\[
\mathbb{E}[V_\alpha(A_0; B^n)] \leq 2^{-n \left( H(A) - I_\alpha(A;B) - R + o(n) \right)}, \tag{16}
\]

where \(A_0 = f(A^n)\), the average is taken over all \(k\)-to-1 functions \(f: T_n(p_A) \rightarrow A_0\) (i.e., over all regular random bin mappings \(f\)) and \(I_\alpha(A;B)\) is the \(\alpha\)-Rényi mutual information according to Csiszár’s proposal [2, Eq. 29] defined by

\[
I_\alpha(A;B) = \min_{q_B} \sum_a p(a) D_{\alpha}(p_{B|a} \parallel q_B).
\]

In particular, the average correlation \(\mathbb{E}[V_n(A_0; B^n)]\) vanishes as \(n\) tends to infinity if

\[
R < H(A) - I_\alpha(A;B).
\]

**Furthermore, we have**

\[
\mathbb{E}[\|p_{A^n B^n} - p_{A_0} \times p_{B^n}\|_1] \leq 2^{-n \max_{0 \leq \alpha \leq 1} \left( \mathbb{D}(H(A) - I_\alpha(A;B) - R + o(n)) \right)}
\]

\[
= 2^{-n \left( \min_{\alpha \in (0,1]} \mathbb{D}(H(A) - I_\alpha(A;B)) + \mathbb{D}(H(A) - R + o(n)) \right)}.
\]

From [12, Eq. 24], we have \(H(A) \geq I_\alpha(A;B)\) with equality when \(B = A\). Thus, the above bound \(H(A) - I_\alpha(A;B)\) on the binning rate is always non-negative. Moreover, since \(I_\alpha(A;B) \geq I(A;B)\), we have \(H(A) - I_\alpha(A;B) \leq H(A) - I(A;B) = H(A|B)\). Hence, the bound given in the statement of the theorem on \(R\) does not exceed \(H(A|B)\), the conditional Slepian-Wolf rate, as expected.

**Remark 1.** To the best of our knowledge, the generalized cut-off rates of Csiszar for the dependencies of random bin indices are not defined or studied in the literature. However, we point out that resolvability exponents are studied in [13]–[16]. In particular, [16] finds the following resolvability exponent for i.i.d. codewords:

\[
\alpha(R', P_A, P_B|A) = \max_{\alpha \in [0,1]} \frac{1}{\alpha} \left( R' - I_\alpha(A;B) \right), \tag{17}
\]

where \(I_\alpha(A;B)\) is the \(\alpha\)-Rényi mutual information according to Sibson’s proposal. To relate the resolvability problem and our problem, let \(R' = H(A) - R\). Then, we see that the exponent of [16] has the same form as our exponent, except that our \(\alpha\)-Rényi mutual information is computed according to Csiszar’s proposal which result in stronger bounds.²

B. The wiretap channel

A wiretap channel is determined by a bipartite conditional distribution \(p_{Y|Z,X}\) in which \(X\) is the input of the channel, output \(Y\) is received by the legitimate receiver and output \(Z\) is received by an eavesdropper. The goal of communication over a wiretap channel is to securely send information to the legitimate receiver. It is well-known that for any input distribution \(p_X\), the rate \(I(X;Y) - I(X;Z)\) is achievable. Our goal here is to establish a bound on the secrecy exponent of random coding over a wiretap channel.

**Theorem 6.** Let \(p_{Y|Z,X}\) be an arbitrary wiretap channel and take \(\alpha \in (1,2]\). Then for any input distribution \(p_X\) there exists

²The very recent revision of [16] shows a similar result for the constant composition code, in which the Sibson’s \(\alpha\)-Rényi mutual information is replaced by Csiszar’s \(\alpha\)-Rényi mutual information.
a code for reliably sending message $M$ of rate $R$ over the channel (with asymptotically vanishing error) such that
\[ V_\alpha(M; Z^n) \leq 2^{-n \alpha \left( I(X; Y) - H_\alpha^c(X; Z) - R + o(n) \right)}. \]  
(18)

In particular, for such a code we have
\[ \left\| p_M Z^n - p_M X Z^n \right\|_1 \leq 2^{-n \alpha \left( I(X; Y) - H_\alpha^c(X; Z) - R + o(n) \right)}. \]  
(19)

Proof. By a continuity type argument we can assume with no loss of generality that $p(x)$ for any $x \in X$ is a rational number, and in the following, we take $n$ to be a sufficiently large number such that $n p(x)$ is a natural number for all $x$. Let $T_n(p_X) \subseteq X^n$ be the set of sequences of type $p_X$, and let $X^n$ be uniformly distributed over $T_n(p_X)$.

Choose positive reals $R_1, R_2, R_3$, which may depend on $n$, such that
- $R_1 = R + o(n)$,
- $R_2 > H(X|Y)$,
- $R_1 + R_3 < H(X) - I^n_\alpha(X; Z)$
- $2^n R_i$ is an integer for $i = 1, 2, 3$ and $|T_n(p_X)| = \prod_{i=1}^3 2^n R_i$.

Observe that if $R < I(X; Y) - I^n_\alpha(X; Z)$ such a triple $(R_1, R_2, R_3)$ exists.

Let $f = (m, g, u) : T_n(p_X) \rightarrow [2^n R_1] \times [2^n R_2] \times [2^n R_3]$ be a random 1-to-1 function (relabelling), and define $M = m(X^n)$, $G = g(X^n)$, $U = u(X^n)$. Note that, for example, $(m, g) : T_n(p_X) \rightarrow [2^n R_1] \times [2^n R_2]$ is a random $2^{n R_1}$-to-1 function. Moreover, since $X^n$ is distributed uniformly over $T_n(p_X)$, random variables $M, G$ and $U$ will be uniform and mutually independent.

If $R_3 > H(X|Y)$, having access to $(U, Y^n)$, the legitimate receiver can decode $X^n$ with a vanishing average error probability:
\[ \mathbb{E}[\Pr(\text{error})] \rightarrow 0, \]  
(20)
as $n$ goes to infinity, where the average is taken over the random choice of $f$. Next, by Theorem 5 since $R_1 + R_3 < H(X) - I^n_\alpha(X; Z)$, we have
\[ \mathbb{E}[V_\alpha(M; U : Z^n)] \leq 2^{-n \alpha \left( H(X) - I^n_\alpha(X; Z) - R_1 - R_3 + o(n) \right)}. \]  
(21)

On the other hand, by Theorem 1 we obtain
\[ \mathbb{E}[V_\alpha(M; Z^n|U)] \leq 2^{-n \alpha - 1} \mathbb{E}[V_\alpha(M; U : Z^n)] \leq 2^{-n \alpha \left( H(X) - I^n_\alpha(X; Z) - R_1 - R_3 + o(n) \right)}. \]  
(22)

Therefore, using (20) and (22), and Markov’s inequality together with a union bound, for any $\epsilon > 0$ and sufficiently large $n$, there exists $u \in [2^n R_2]$ and a random labeling $f_0$ such that
\[ \Pr(\text{error}|f_0, U = u) \leq \epsilon, \]  
(23)

and
\[ V_\alpha(M; Z^n|f_0, U = u) \leq 2^{-n \alpha \left( H(X) - I^n_\alpha(X; Z) - R_1 - R_3 + o(n) \right)}. \]  
(24)

Now, as in [11], the code can be constructed as follows. We treat $M$ as the message (which is distributed uniformly), select $G$ uniformly at random and independent of $M$ and transmit the codeword $X^n = f_0^{-1}(M, G, u)$. The legitimate receiver can decode $M$ with an asymptotically vanishing error because of (23), and the eavesdropper would gain no information about $M$ due to (24).

\[ \square \]

References

[1] M. M. Mojadidian, S. Beigi, A. Gohari, M. H. Yassae, and M. R. Aref, “A correlation measure based on vector-valued $I_p$-norms,” 2018.

[2] S. Verdù, “Mutual information,” in Information Theory and Applications Workshop (ITA), 2015. IEEE, 2015, pp. 1–6.

[3] R. Renner, “Security of quantum key distribution,” International Journal of Quantum Information, vol. 6, no. 01, pp. 1–127, 2008. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0219749908003256.

[4] R. Renner and R. König, “Universally composable privacy amplification against quantum adversaries,” in Second Theory of Cryptography Conference TCC, volume 3578 of Lecture Notes in Computer Science. Springer, 2005.

[5] M. Hayashi, “Exponential decreasing rate of leaked information in universal random privacy amplification,” IEEE Transactions on Information Theory, vol. 57, no. 6, pp. 3989–4001, 2011.

[6] J. Cohen, J. Kempermann, and G. Zbaganu, Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population. Springer Science & Business Media, 1998.

[7] M. R. Aref, “Strong data processing inequalities and $\phi$-sobolev inequalities for discrete channels,” IEEE Transactions on Information Theory, vol. 62, no. 6, pp. 3355–3389, 2016.

[8] H. Hsu, S. Assouad, S. Salamatian, and F. P. Calmon, “Generalizing bottleneck problems,” arXiv preprint arXiv:1802.05861, 2018.

[9] Y. Polyanskiy and S. Verdù, “Ariomato channel coding converse and rényi divergence,” in Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on. IEEE, 2010, pp. 1327–1333.

[10] S. Boucheron, G. Lugosi, and O. Bousquet, “Concentration inequalities,” in Advanced Lectures on Machine Learning. Springer, 2004, pp. 208–240.

[11] M. H. Yassae, M. R. Aref, and A. Gohari, “Achievability proof via output statistics of random binning,” IEEE Transactions on Information Theory, vol. 60, no. 11, pp. 6760–6786, 2014.

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

[13] M. B. Parizi, E. Telatar, and N. Merhav, “Exact random coding secrecy exponents for the wiretap channel,” IEEE Transactions on Information Theory, vol. 63, no. 1, pp. 509–531, 2017.

[14] M. Hayashi, “General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1562–1575, 2006.

[15] T. S. Han, H. Endo, and M. Sasaki, “Reliability and secrecy functions of the wiretap channel under cost constraint,” IEEE Transactions on Information Theory, vol. 60, no. 11, pp. 6819–6843, 2014.

[16] S. Yagli and P. Cuff, “Exact soft-covering exponent,” arXiv preprint arXiv:1801.00714, 2018.