Hentila, Henri; Shkel, Yanina; Koivunen, Visa

Second-Order Converse for Rate-Limited Common Randomness Generation

Published in:
2022 IEEE International Symposium on Information Theory (ISIT)

DOI:
10.1109/ISIT50566.2022.9834737

Published: 01/01/2022

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Hentila, H., Shkel, Y., & Koivunen, V. (2022). Second-Order Converse for Rate-Limited Common Randomness Generation. In *2022 IEEE International Symposium on Information Theory (ISIT)* (pp. 2315-2320). (IEEE International Symposium on Information Theory). IEEE. https://doi.org/10.1109/ISIT50566.2022.9834737

Second-Order Converse for Rate-Limited Common Randomness Generation

Henri Hentilä*, Yanina Shkel†, Visa Koivunen*
* Department of Signal Processing and Acoustics, School of Electrical Engineering, Aalto University, Finland
† School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland

***Abstract*—We employ a recent technique based on a semigroup application of the method of types to improve on a second-order converse for the common randomness (CR) generation problem. The previously known bound lead to a correct second-order asymptotic rate, but incorrect sign on the second-order term for error rates below 1/2. The new bound has both the correct scaling and sign of the second-order term for small enough error rates.**

I. Introduction

In the common randomness (CR) generation problem, introduced in [1], two terminals with access to dependent random variables wish to agree upon a common random value. Such a random variable could be used for various cryptographic purposes, for example to form a secret key [2] with which to encrypt the communication between the terminals. The dependent random variables needed as input at the terminals can be extracted from the wireless communition channel between them [3], making this a highly relevant model in wireless communications.

The fundamental limits of CR generation have been studied and characterized in the asymptotic regime [1], [2], [4]. However, the corresponding finite-blocklength bounds, which are more relevant in practical systems subject to strict latency constraints, remain an open problem. One approach towards characterizing such bounds is the *second-order* analysis initiated in [5] and recently popularized via the information spectrum techniques of [6] and [7]. In this context, the asymptotic bounds of [1], [2], [4] describe a *first-order* term, which is augmented with a second-order term describing the rate at which we approach this first-order term as the blocklength grows. The techniques in [7], [8] have been used to derive both second-order and finite-blocklength bounds for various channel coding problems, with extensions to e.g. source coding problems [9], and wiretap channels [10].

Problems whose first-order term includes auxiliary random variables subject to Markov chain conditions – typical examples of which include coding with side information [11] and CR generation – have proven to be unamenable to second-order analysis. One possible and remarkably general approach to deriving bounds for problems of this type was recently proposed in [12]. Not only does this approach yield bounds with the correct second-order rate $\sqrt{n}$, but it also works both for discrete and continuous alphabets. However, a drawback of the method in [12] is that the sign of the second-order term is incorrect for error probabilities below $\frac{1}{2}$. Typically, the second-order term is expected to act as a penalty for working at finite blocklengths with error rates below $\frac{1}{2}$, but with the sign being wrong the effect is the opposite.

To address this shortcoming, [13] augments the technique in [12] with a semigroup application of the method of types. This augmentation allows one to modify the bounds provided by [12] to yield the correct sign on the second-order term for error rates approaching zero. In this paper, we employ the technique in [13] to tighten a converse bound on CR generation derived in [12]. Specifically, our new bound has the same first-order term and second-order rate $\sqrt{n}$ as the bound in [12], but also the correct sign on the second-order term for sufficiently small error rates. Since the bound in [12] was only tight for error rates above $\frac{1}{2}$, which is rarely the case in practice, the new bound is more useful in practical settings.

The paper is structured as follows. In Section II, we formally introduce the CR problem and discuss prior results. The new converse bound is presented and discussed in Section III. The main tools used to prove this result are developed in Section IV, and the final proof provided in Section V. The paper is then concluded in Section VI.

II. Problem Formulation

Let X and Y be random variables over discrete alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and denote their joint distribution by Q_{XY}. Two terminals, each observing one of the two random variables, wish to generate a common key K drawn uniformly from the set $\mathcal{K}$, using a message $W \in \mathcal{W}$ from the terminal observing X to the terminal observing Y, as depicted in Fig. 1. In particular, given an error threshold $\delta_1 \in (0,1)$ and uniformity threshold $\delta_2 \in (0,1)$, the goal is to construct an encoder $Q_{W|X} : \mathcal{X} \to \mathcal{W}$, and decoders $Q_{K|X} : \mathcal{X} \to \mathcal{K}$ and $Q_{\hat{K}|WY} : \mathcal{Y} \times \mathcal{W} \to \mathcal{K}$ such that

$$\mathbb{P}[K \neq \hat{K}] \leq \delta_1 \tag{1}$$

$$\|Q_K - T_K\| \leq \delta_2 \tag{2}$$

where $\|P - Q\| = \frac{1}{2}\sum_x |P(x) - Q(x)|$ denotes the variational distance between distributions P and Q, and T_K denotes the equiprobable distribution over $\mathcal{K}$.

A common setting, and the one this paper will also focus on, is when X and Y are sequences of n i.i.d. variables. In this case we denote them $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots, Y_n)$, the corresponding distribution $Q_{XY}^{\otimes n}$, and the keys and messages $K_n \in \mathcal{K}_n$ and $W_n \in \mathcal{W}_n$. In this setting, one is often interested in either maximizing the

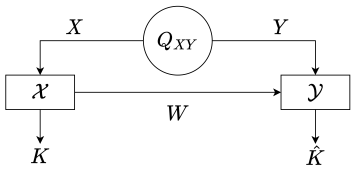

Fig. 1: CR generation with one-way communication.

key rate $R_K \triangleq \frac{1}{n}\log|\mathcal{K}_n|$, or minimizing the message rate $R_W \triangleq \frac{1}{n}\log|\mathcal{W}_n|$, or both. When $n \to \infty$, the following bound is known:

Theorem 1 ([14, Theorem 4.1]). *Every achievable rate pair (R_K, R_W) for the CR generation problem with $(X^n, Y^n) \sim Q_{XY}^{\otimes n}$, $n \to \infty$, satisfies*

$$R_K \le I(U;X), \tag{3}$$

$$R_W \ge I(U;X) - I(U;Y) \tag{4}$$

for any auxiliary rv U forming the Markov chain $U - X - Y$.

Determining the corresponding rate bounds in the nonasymptotic regime, i.e. when n is finite, remains an open problem, with some preliminary results found in e.g. [12], [15]. Given Q_{XY} and $c \in (1,\infty)$, let us define the following function, describing the first-order term of the problem:

$$d_c^\star(Q_{XY}) \triangleq \sup_{P_{U|X}} \{cI(U;Y) - I(U;X)\} \tag{5}$$

where $Q_{UXY} = P_{U|X}Q_{XY}$. In [12], it was shown that for large enough n, we have approximately

$$(c-1)\log|\mathcal{K}_n| - c\log|\mathcal{W}_n| \le nd_c^\star(Q_{XY}) + A\sqrt{n} \tag{6}$$

where $A > 0$ is a constant that depends on Q_{XY}. Dividing both sides of (6) by n and letting $n \to \infty$, one can see that this bound coincides with the ones in Theorem 1. That is, the first-order term $d_c^\star(Q_{XY})$ combines the asymptotic bounds (3) and (4) via the weight c, and the second-order term $A/\sqrt{n}$ vanishes. The rate $\sqrt{n}$ at which the second-order term decays is known to be correct [12]. However, observe that since $A > 0$, the second-order term does not act as a penalty for finite n as one would expect when $\delta_1 < \frac{1}{2}$, but rather it results in a bound that is less stringent than the one in Theorem 1.

III. New Second-Order Converse

We apply the technique introduced in [13] to derive a bound such that $(c-1)\log|\mathcal{K}_n| - c\log|\mathcal{W}_n| \le nd_c^\star(Q_{XY}) - A\sqrt{n}$ for sufficiently small error rates. That is, a bound where the second-order term indeed penalizes the fact that we are working with a finite n. More precisely, we will assume the existence of a CR scheme satisfying $(c-1)\log|\mathcal{K}_n| - c\log|\mathcal{W}_n| \ge nd_c^\star(Q_{XY}) - A\sqrt{n}$, and then lower bound the error probability as $n \to \infty$. Using a similar dispersion analysis as in [13], one can then show that this corresponds to the claim at the beginning of the paragraph.

Towards this end, define

$$\imath_{U;X}(u;x) \triangleq \log\frac{Q_{X|U}(x|u)}{Q_X(x)} \tag{7}$$

and $\imath_{U;Y}(u;y)$ similarly. Note that $I(U;X) = \mathbb{E}[\imath_{U;X}(U;X)]$. The main result of this paper is then the following.

Theorem 2. *Fix Q_{XY}, $c > 1$, and $\delta_2 \in (0,1)$. Let $(X^n, Y^n) \sim Q_{XY}^{\otimes n}$. If a CR generation scheme for (X^n, Y^n) satisfies (2) and there exists some $A \in \mathbb{R}$ such that*

$$(c-1)\log|\mathcal{K}_n| - c\log|\mathcal{W}_n| \ge nd_c^\star(Q_{XY}) - A\sqrt{n} \tag{8}$$

for every n, then the probability of error is lower bounded by

$$\liminf_{n\to\infty} \mathbb{P}[K_n \neq \hat{K}_n] \ge \sup_{\delta_1\in(0,1)} \delta_1 \Phi\left(\frac{-A - c\sqrt{\frac{8}{\min_x Q_X(x)}\ln\frac{2(1-\delta_2)}{1-(\delta_1+\delta_2)}}}{\sqrt{V}}\right) \tag{9}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{\frac{-t^2}{2}} dt$ is the standard normal CDF, and

$$V = \mathrm{Var}\{\mathbb{E}[c\imath_{U;Y}(U;Y) - \imath_{U;X}(U;X)|X,Y]\} \tag{10}$$

for any optimizer $P_{U|X}^\star$ of (5).

The detailed proof of Theorem 2 is postponed until Section V. Here we shall give a brief sketch of the proof. The idea is to first find a bound of the suboptimal form (6) for empirical distributions of realizations of $Q_{XY}^{\otimes n}$, resulting in Lemma 5 ahead. Then, by averaging over these empirical distributions, we are able to characterize under which conditions a bound of the desired form (8) holds for the original distribution $Q_{XY}^{\otimes n}$.

The (suboptimal) bound of Lemma 5 is derived using the semigroup technique of [12] integrated with the method of types as in [13]. We give a high-level overview of this technique in Section IV-B. In essence, one needs to find an alternative functional form (as opposed to their more common entropical form) of the quantities involved. Then, by using an appropriately chosen semigroup operator one is able to find tight upper and lower bounds on these functional forms.

In order to derive the lower bound on error probability in Theorem 2, we use the central limit theorem (CLT) applied to i.i.d. sequences of the gradient of the first-order term $d_c^\star$ (see Section IV-D for a discussion on this gradient). In particular, we combine the assumption (8) with the bound in Lemma 5 and a Taylor expansion of $d_c^\star$ to find an inequality describing when the error probability exceeds a certain limit. The probability of this event is then approximated via the CLT.

IV. Notation, Concepts, and Tools

In this section, we introduce and discuss notation, concepts, and tools used in the proof of Theorem 2.

A. Notation

Given an alphabet $\mathcal{Y}$, let $\mathcal{H}_+(\mathcal{Y})$ denote the set of non-negative functions on $\mathcal{Y}$, and $\mathcal{H}_{[0,1]}(\mathcal{Y})$ the subset of $\mathcal{H}_+(\mathcal{Y})$ consisting of functions whose output is restricted to $[0,1]$. Given $f \in \mathcal{H}_+(\mathcal{Y})$, define $P_Y(f) \triangleq \mathbb{E}_{P_Y}[f]$ and $P_{Y|X}(f) \triangleq \mathbb{E}_{P_{Y|X=\cdot}}[f(Y)]$ as a function on $\mathcal{H}_+(\mathcal{X})$. Given an n-type P_{XY}, let $\mathcal{T}_n(P_X)$ be the set of all x^n with type P_X, and $\mathcal{T}_{x^n}(P_{Y|X})$ the set of all y^n such that (x^n, y^n) has type P_{XY}. For $p \in (0, \infty)$, we denote the L^p-norm of $f \in \mathcal{H}_+(\mathcal{Y})$ w.r.t. a distribution P equiprobable on Ω by $\|f\|_{L^p(\Omega)} \triangleq \left(\int f^p dP\right)^{1/p} = P^{1/p}(f^p)$. When $p \to 0$, we have $\|f\|_{L^0(\Omega)} = e^{P(\ln f)}$. Given two distributions P and Q over the same alphabet, $D(P\|Q)$ denotes their relative entropy.

B. Functional Inequalities

The main ingredient of the machinery proposed in [12] is the use of functional inequalities on the quantities involved. In particular, the entropic quantities found in e.g. $d_c^\star$ have alternative functional forms (see e.g. (20)), allowing one to find upper and lower bounds on these that are generally tighter than if one had operated on the entropical form instead. The primary insight of [12] is that while directly working with functional forms where $f \in \mathcal{H}_{[0,1]}$ describes an indicator function of e.g. decoding sets generally fails, this problem can be overcome by smoothing out f via an appropriately chosen operator T. In particular, when T is a Markov semigroup operator, one is able to use something called reverse hypercontractivity to derive the desired bounds.

First, we may note that by the Donsker-Varadhan lemma (see e.g [16]), the relative entropy satisfies the following functional inequality, which can be made arbitrarily tight.

Lemma 1. *Let P and Q be distributions over the same alphabet $\mathcal{Y}$. Then, for any $f \in \mathcal{H}_+(\mathcal{Y})$*

$$D(P\|Q) \geq P(\ln f) - \ln Q(f). \tag{11}$$

As in [12] and [13], we will not directly work with a semigroup operator T, but rather with a related operator $\Lambda_{n,t} : \mathcal{H}_+(\mathcal{Y}^n) \to \mathcal{H}_+(\mathcal{Y}^n)$ which will be subject to the same reverse hypercontractivity bound as the underlying semigroup (this distinction is irrelevant to the present paper, but we point it out so as not to be imprecise). The key inequalities that arise from applying this operator to f, both of which will be used in the upcoming proof of Theorem 2, are summarized in the following lemma. Observe that the inequality (12) is a result of the reverse hypercontractivity property mentioned earlier.

Lemma 2 ([13])**.** *Let $\Lambda_{n,t}$, with $n \geq 1$, $t \geq 0$, be as defined in [13], and, for an n-type source P_{XY}, define $P_{X^nY^n}$ to be the equiprobable distribution on $\mathcal{T}_n(P_{XY})$ and $P_{Y^n|X^n}$ the induced random transformation. Then, for any $f \in \mathcal{H}_{[0,1]}(\mathcal{Y}^n)$*

$$\|\Lambda_{n,t} f\|_{L^0(\mathcal{T}_{x^n}(P_{Y|X}))} \geq P_{Y^n|X^n=x^n}^{(1+\frac{1}{t})}(f) \tag{12}$$

and for any $f \in \mathcal{H}_+(\mathcal{Y}^n)$

$$P_{Y^n}(\Lambda_{n,t} f) \leq \exp\left(\frac{nt}{\min_x P_X(x)}\right) P_{Y^n}(f). \tag{13}$$

C. Single-Letterization of First-Order Term

En route to acquiring the desired single-letter first-order term (5) in Theorem 2, we will find ourselves working with a multi-letter version of (5). In particular, for an n-type distribution P_{XY}, define $P_{X^nY^n}$ as in Lemma 2. Then

$$d_{c,n}(P_{XY}) \triangleq \sup_{S_{X^n}} \{cD(S_{Y^n}\|P_{Y^n}) - D(S_{X^n}\|P_{X^n})\} \tag{14}$$

where the supremum is over distributions S_{X^n} supported on $\mathcal{T}_n(P_X)$ and $S_{X^n} \to P_{Y^n|X^n} \to S_{Y^n}$. To relate the multi-letter first-order term $d_{c,n}(P_{XY})$ to the desired single-letter form $d_c^\star(P_{XY})$, we will use the following bound, the proof of which can be found in the appendix.

Lemma 3. *Given Q_{XY} and $c > 1$, there exists $\lambda \in (0,1)$ and $E > 0$ such that for any $n \geq 1$ and n-type P_{XY} satisfying $\|P_{XY} - Q_{XY}\| \leq \lambda$, we have*

$$d_{c,n}(P_{XY}) \leq nd_c^\star(P_{XY}) + E\ln n. \tag{15}$$

D. The Gradient of $d_c^\star$

The final key ingredient of the proof in the following section is the gradient of $d_c^\star$. To understand what is meant by the gradient in this context, observe that a distribution S_{XY} given as input to $d_c^\star$ can be interpreted as a point in $\mathbb{R}^{|\mathcal{X}||\mathcal{Y}|}$ space, with each probability $S_{XY}(x,y)$, $(x,y) \in \mathcal{X} \times \mathcal{Y}$, corresponding to the value of one of the coordinates of this point. The gradient is then the vector whose (x,y)-th coordinate is formed by taking the partial derivative of $d_c^\star$ w.r.t $S_{XY}(x,y)$. Letting $\nabla d_c^\star|_{Q_{XY}}$ denote this gradient evaluated at Q_{XY}, we have the following result, the proof of which can be found in the appendix.

Lemma 4. *For every $(x,y) \in \mathcal{X} \times \mathcal{Y}$, and optimal $P_{U|X}$*

$$\begin{aligned}&\nabla d_c^\star|_{Q_{XY}}(x,y)\\ &= \mathbb{E}[c\imath_{U;Y}(U;Y) - \imath_{U;X}(U;X)|X=x, Y=y] + (c-1)\log e.\end{aligned}$$

V. Proof of New Converse

In this section we prove Theorem 2 in two parts. In part A, we derive a converse bound of the suboptimal form (6) for n-type distributions P_{XY} within some neighborhood of Q_{XY}. Then, in part B, we convert the result back to the original distribution Q_{XY}, allowing us to characterize the error probability under the assumption that a bound of the desired form (8) holds.

A. Converse for P_{XY}

We will start by proving the following converse bound for n-types P_{XY} within some neighborhood of Q_{XY}:

Lemma 5. *Let Q_{XY} and $c \in (1, \infty)$ be given. Then, there exists $\lambda \in (0,1)$ and $E > 0$ such that the following holds: For any $n \geq 1$ and n-type P_{XY} such that $\|P_{XY} - Q_{XY}\| \leq \lambda$, let (X^n, Y^n) be equiprobable on the type class $\mathcal{T}_n(P_{XY})$. If there exists a CR generation scheme for (X^n, Y^n) satisfying (1) and (2) for $\delta_1, \delta_2 \in (0,1)$, then, for any $\delta_3 \in (\delta_1, 1)$*

$$1 - \delta_2 - \delta_3 \leq \frac{1}{|\mathcal{K}_n|} + \frac{e^{\left(\frac{nd_c^\star(P_{XY})}{c} + \frac{B\sqrt{n}}{c} + \frac{E\ln n}{c}\right)}|\mathcal{W}_n|}{\left(\frac{\delta_3 - \delta_1}{2\delta_3}\right)^{1+\frac{1}{c}} |\mathcal{K}_n|^{1-\frac{1}{c}}}$$

where $B = 2c\sqrt{(\min_x P_X(x))^{-1}\ln\{(2\delta_3)/(\delta_3 - \delta_1)\}}$.

Observe that for large enough n, the bound in Lemma 5 is essentially of the form (6) with the constant B serving as A. The proof of Lemma 5 will use the following general bound for CR generation found in [12].

Lemma 6 ([12, Lemma 4.4.4]). *Suppose that, given Q_{XY}, there exist $\delta_1, \delta_2 \in (0,1)$ a stochastic encoder $Q_{W|X}$ and deterministic decoders $Q_{K|X}$ and $Q_{\hat{K}|WY}$ such that (1) and (2) hold. Also, suppose that there exists a distribution $\tilde{Q}_X$ on X, $\delta \in [0,1)$, $\epsilon \in (0,1)$, $c \in (1,\infty)$, and $d \in (0,\infty)$ such that*

$$\|Q_X - \tilde{Q}_X\| \le \delta \tag{16}$$

$$\tilde{Q}_X(x : Q_{Y|X=x}(\mathcal{A}) \ge 1-\epsilon) \le 2^c \exp(d) Q_Y^c(\mathcal{A}) \tag{17}$$

for any $\mathcal{A} \subseteq \mathcal{Y}$. Then, for any $\delta_3, \delta_4 \in (\delta_1 + \delta, 1)$ such that $\delta_3\delta_4 = \delta_1 + \delta$, we have

$$\delta_2 \ge 1 - \delta - \delta_3 - \frac{1}{|\mathcal{K}|} - \frac{2\exp(\frac{d}{c})|\mathcal{W}|}{(\epsilon-\delta_4)^{\frac{1}{c}}|\mathcal{K}|^{1-\frac{1}{c}}}. \tag{18}$$

Proof of Lemma 5: Let P_{XY} be an n-type as in the statement of the lemma, with $P_{X^nY^n}$ the equiprobable distribution on $\mathcal{T}_n(P_{XY})$ and $P_{Y^n|X^n}$ the induced random transformation. We will start by proving the following bound, valid for any $f \in \mathcal{H}_{[0,1]}(\mathcal{Y}^n)$ (in particular, f can be thought of as the indicator function of some $\mathcal{A} \subseteq \mathcal{Y}^n$) and $\eta \in (0,1)$:

$$\begin{aligned} &\ln P_{X^n}[P_{Y^n|X^n=x^n}(f) \ge \eta] - c \ln P_{Y^n}(f) \\ &\le n d_c^\star(P_{XY}) + E\ln n - c\ln\eta + B\sqrt{n} \end{aligned} \tag{19}$$

where $B \triangleq 2c\sqrt{\frac{1}{\min_x P_X(x)}\ln\frac{1}{\eta}}$.

Towards this end, Let g be any function in $\mathcal{H}_+(\mathcal{Y}^n)$, and define an auxiliary distribution S_{X^n} via $\frac{dS_{X^n}}{dP_{X^n}} = e^{P_{Y^n|X^n}(\ln g)}(\int e^{P_{Y^n|X^n}(\ln g)}dP_{X^n})^{-1}$ and set $S_{X^n} \to P_{Y^n|X^n} \to S_{Y^n}$. Then we see that

$$\begin{aligned} &D(S_{X^n}\|P_{X^n}) \\ &= \int \ln\left(\frac{dS_{X^n}}{dP_{X^n}}\right) dS_{X^n} \\ &= \int P_{Y^n|X^n}(\ln g) dS_{X^n} - \ln\left(\int e^{P_{Y^n|X^n}(\ln g)} dP_{X^n}\right) \\ &= S_{Y^n}(\ln g) - \ln\left(\int \|g\|_{L^0(\mathcal{T}_{x^n}(P_{Y|X}))} dP_{X^n}(x^n)\right) \end{aligned} \tag{20}$$

Hence

$$\begin{aligned} &\int \|g\|_{L^0(\mathcal{T}_{x^n}(P_{Y|X}))} dP_{X^n}(x^n) && (21) \\ &= e^{-D(S_{X^n}\|P_{X^n}) + S_{Y^n}(\ln g)} \\ &\le e^{d_{c,n}(P_{XY}) - cD(S_{Y^n}\|P_{Y^n}) + cS_{Y^n}(\ln g^{1/c})} \\ &\le e^{d_{c,n}(P_{XY})} \|g\|_{L^{1/c}(\mathcal{T}_n(P_Y))} && (22) \end{aligned}$$

where the first inequality used the definition of $d_{c,n}$ and the second inequality used (11).

Now, let us choose $g = (\Lambda_{n,t}f)^c$, where f is any function in $\mathcal{H}_{[0,1]}(\mathcal{Y}^n)$, $\Lambda_{n,t}$ is as in Lemma 2, and $t > 0$ is arbitrary. Then, by (13), we have

$$\begin{aligned} \|g\|_{L^{1/c}(\mathcal{T}_n(P_Y))} &= \|(\Lambda_{n,t}f)^c\|_{L^{1/c}(\mathcal{T}_n(P_Y))} \\ &= P_{Y^n}^c(\Lambda_{n,t}f) \\ &\le \exp\left(\frac{cnt}{\min_x P_X(x)}\right) P_{Y^n}^c(f). \end{aligned} \tag{23}$$

On the other hand, for the chosen g, we can also bound (21) from below via (12) as

$$\begin{aligned} &\int \|g\|_{L^0(\mathcal{T}_{x^n}(P_{Y|X}))} dP_{X^n}(x^n) \\ &= \int \|(\Lambda_{n,t}f)\|^c_{L^0(\mathcal{T}_{x^n}(P_{Y|X}))} dP_{X^n}(x^n) \\ &\ge \int P_{Y^n|X^n=x^n}^{c(1+\frac{1}{t})}(f) dP_{X^n}(x^n) \\ &\ge \eta^{c(1+\frac{1}{t})} P_{X^n}[P_{Y^n|X^n=x^n}(f) \ge \eta]. \end{aligned} \tag{24}$$

By combining (22), (23), and (24), we thus have

$$\begin{aligned} &\ln P_{X^n}[P_{Y^n|X^n=x^n}(f) \ge \eta] - c\ln P_{Y^n}(f) \\ &\le d_{c,n}(P_{XY}) + \frac{cnt}{\min_x P_X(x)} + c\left(1+\frac{1}{t}\right)\ln\frac{1}{\eta}. \end{aligned} \tag{25}$$

Since $t > 0$ was arbitrary, we get the tightest bound by minimizing over t, yielding $t = \sqrt{\frac{\min_x P_X(x)}{n}\ln\frac{1}{\eta}}$. The bound (19) then follows from Lemma 3.

Next, we will use Lemma 6 particularized to $Q_{XY} \leftarrow P_{X^nY^n}$, and set $\tilde{Q}_X \leftarrow P_{X^n}$. Observe that in this case, (16) is trivially satisfied for any $\delta \ge 0$, and by (19) we can satisfy (17) by setting $\epsilon = 1 - \eta$ and

$$d = nd_c^\star(P_{XY}) + B\sqrt{n} + E\ln n - c\ln(1-\epsilon) - c\ln 2. \tag{26}$$

For the remaining parameters in Lemma 6, let us arbitrarily fix $\delta_3 \in (\delta_1, 1)$, and set $\delta = 0$, $\delta_4 = (\delta_1 + \delta_3)/(2\delta_3)$, and $\epsilon = (1+\delta_4)/2$. The claim then follows by direct application of the bound (18) in Lemma 6. ■

B. Proof of Theorem 2

We are now ready to prove Theorem 2. Towards this end, assume that there exists a sequence of CR generation schemes for $Q_{XY}^{\otimes n}$ such that (2) and (8) hold for every n. Now let P_{XY} be an arbitrary n-type such that $\|P_{XY} - Q_{XY}\| \le \lambda$. Then, if the error probability $\mathbb{P}[K_n \ne \hat{K}_n]$ conditioned on P_{XY} is bounded above by $\delta_1 \in (0,1)$, we have by Lemma 5 the following inequality which holds for all $\delta_3 \in (\delta_1, 1 - \frac{1}{|\mathcal{K}_n|} - \delta_2)$

$$\begin{aligned} &n(d_c^\star(P_{XY}) - d_c^\star(Q_{XY})) \\ &\ge -A\sqrt{n} - 2c\sqrt{\frac{n}{\min_x P_X(x)}\ln\frac{2\delta_3}{\delta_3 - \delta_1}} - E\ln n + D(\delta_3) \end{aligned} \tag{27}$$

where $D(\delta_3) \triangleq c\ln\left(1 - \delta_2 - \delta_3 - |\mathcal{K}_n|^{-1}\right) + (c+1)\ln\{(\delta_3 - \delta_1)/(2\delta_3)\}$. On the other hand, the Taylor expansion of $d_c^\star$ at Q_{XY} shows us that there exists some $F > 0$ such that

$$\begin{aligned} d_c^\star(P_{XY}) \le d_c^\star(Q_{XY}) &+ \langle \nabla d_c^\star|_{Q_{XY}}, P_{XY} - Q_{XY}\rangle \\ &+ F\|P_{XY} - Q_{XY}\|^2. \end{aligned} \tag{28}$$

Combining (27) and (28), we see that the error probability conditioned on P_{XY} exceeds δ_1 if

$$\begin{aligned} &n\langle \nabla d_c^\star|_{Q_{XY}}, P_{XY} - Q_{XY}\rangle \\ &< -A\sqrt{n} - 2c\sqrt{\frac{n}{\min_x P_X(x)}\ln\frac{2\delta_3}{\delta_3 - \delta_1}} \end{aligned}$$

$$- nF\|P_{XY} - Q_{XY}\|^2 - E\ln n + D(\delta_3) \qquad (29)$$

for some $\delta_3 \in (\delta_1, 1-\delta_2-\frac{1}{|\mathcal{K}_n|})$.

Let us now particularize P_{XY} to be the empirical distribution of $(X^n, Y^n) \sim Q_{XY}^{\otimes n}$, i.e. $P_{XY}(x,y) = \hat{P}_{X^nY^n}(x,y) \triangleq \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i = x, Y_i = y\}$. Then, we observe that

$$\begin{aligned} & n\langle \nabla d_c^\star|_Q, P_{XY} - Q_{XY}\rangle \\ &= \sum_{i=1}^n \sum_{x,y} \nabla d_c^\star|_Q(x,y)\mathbb{1}\{X_i = x, Y_i = y\} \\ &\quad - n\sum_{x,y}\nabla d_c^\star|_Q(x,y)Q_{XY}(x,y) \qquad (30) \\ &= \sum_{i=1}^n \left(\nabla d_c^\star|_Q(X_i,Y_i) - \mathbb{E}[\nabla d_c^\star|_Q(X,Y)]\right). \qquad (31) \end{aligned}$$

Furthermore, by Hoeffding's inequality, the following holds with probability $1 - O(e^{-n^{1/3}})$: $\|P_{XY} - Q_{XY}\| < n^{-1/3}$ and $\min_x P_X(x) > \frac{1}{2}\min_x Q_X(x)$, and (29) holds if for some $G > 0$

$$\begin{aligned} & \sum_{i=1}^n \left(\nabla d_c^\star|_Q(X_i,Y_i) - \mathbb{E}[\nabla d_c^\star|_Q(X,Y)]\right) \\ & < -A\sqrt{n} - 2c\sqrt{\frac{2n}{\min_x Q_X(x)}\ln\frac{2\delta_3}{\delta_3-\delta_1}} - Gn^{1/3} \qquad (32) \end{aligned}$$

where $\delta_3 = 1 - \delta_2 - \frac{1}{|\mathcal{K}_n|}$. The probability that the inequality (32) holds as $n \to \infty$ can be approximated via the CLT as

$$\Phi\left(\frac{-A - c\sqrt{\frac{8}{\min_x Q_X(x)}\ln\frac{2(1-\delta_2)}{1-(\delta_1+\delta_2)}}}{\sqrt{\mathrm{Var}\left(\nabla d_c^\star|_Q\right)}}\right) - o(1). \qquad (33)$$

Denoting the event that (29) holds for δ_1 by $\mathcal{E}_{\delta_1}$, we therefore have

$$\begin{aligned} & \liminf_{n\to\infty} \mathbb{P}[K_n \neq \hat{K}_n] \\ & \geq \liminf_{n\to\infty} \sum_{P_{XY}} \mathbb{P}[K_n \neq \hat{K}_n|\mathcal{E}_{\delta_1}, P_{XY}]\mathbb{P}[\mathcal{E}_{\delta_1}|P_{XY}]\mathbb{P}[P_{XY}] \\ & \geq \delta_1\Phi\left(\frac{-A - c\sqrt{\frac{8}{\min_x Q_X(x)}\ln\frac{2(1-\delta_2)}{1-(\delta_1+\delta_2)}}}{\sqrt{\mathrm{Var}\left(\nabla d_c^\star|_Q\right)}}\right) \qquad (34) \end{aligned}$$

The result now follows by supremizing over δ_1 and using Lemma 4. ∎

VI. Concluding Remarks

As mentioned in the introduction, an interesting application of the CR generation problem is to secret key (SK) generation over wireless networks. However, such keys are subject to a secrecy constraint in addition to the reliability constraint (1) and uniformity constraint (2), which the present bound does not account for. Regardless, since it is a converse bound, it also provides a converse bound for the SK generation problem.

Appendix

A. Proof of Lemma 3

As in the proof of [13, Lemma 2], the idea is to iteratively extract one random element (X_i, Y_i) from (X^n, Y^n), and bound the total divergence in (14) by one divergence term for (X_i, Y_i), and another divergence term for $(X_{\setminus i}, Y_{\setminus i})$, where $X_{\setminus i}$ denotes X^n with X_i removed. In particular, consider any S_{X^n} supported on $\mathcal{T}_n(P_X)$ and set $S_{X^n} \to P_{Y^n|X^n} \to S_{Y^n}$. Additionally, let I be equiprobable on $\{1,\ldots,n\}$ and independent of (X^n, Y^n) under P. We can then show that

$$\begin{aligned} & D(S_{Y^n}\|P_{Y^n}) \\ & = D(S_{Y_I|I}\|P_{Y_I}|S_I) + D(S_{Y_{\setminus I}|IY_I}\|P_{Y_{\setminus I}|Y_I}|S_{IY_I}) \\ & \leq D(S_{Y_I|I}\|P_{Y_I}|S_I) + D(S_{Y_{\setminus I}|IX_IY_I}\|P_{Y_{\setminus I}|X_IY_I}|S_{IX_IY_I}) \end{aligned} \qquad (35)$$

where the inequality follows from

$$\begin{aligned} & D(S_{Y_{\setminus I}|IY_I}\|P_{Y_{\setminus I}|Y_I}|S_{IY_I}) \\ & = \sum_{i,y}\{H(S_{Y_{\setminus I}|IY_I=(i,y)}, P_{Y_{\setminus I}|Y_I=y}) \\ & \qquad - H(S_{Y_{\setminus I}|IY_I=(i,y)})\}S_{IY_I}(i,y) \qquad (36) \\ & = \sum_{i,x,y}\{H(S_{Y_{\setminus I}|IX_IY_I=(i,x,y)}, P_{Y_{\setminus I}|X_IY_I=(x,y)}) \\ & \qquad - H(S_{Y_{\setminus I}|IY_I=(i,y)})\}S_{IX_IY_I}(i,x,y) \qquad (37) \\ & \leq \sum_{i,x,y}\{H(S_{Y_{\setminus I}|IX_IY_I=(i,x,y)}, P_{Y_{\setminus I}|X_IY_I=(x,y)}) \\ & \qquad - H(S_{Y_{\setminus I}|IX_IY_I=(i,x,y)})\}S_{IX_IY_I}(i,x,y) \qquad (38) \\ & = D(S_{Y_{\setminus I}|IX_IY_I}\|P_{Y_{\setminus I}|X_IY_I}|S_{IX_IY_I}) \qquad (39) \end{aligned}$$

with $H(P,Q) \triangleq \mathbb{E}_P[-\log Q]$ denoting the cross entropy. The equality (37) is due to the fact that $Y_{\setminus I} - Y_I - X_I$ under P, as well as the fact that $P_{Y_{\setminus I}|Y_I=y}$ is equiprobable over its support.

Since it was shown in the proof of [13, Lemma 2] that

$$\begin{aligned} D(S_{X^n}\|P_{X^n}) &= D(S_{X_I|I}\|P_{X_I}|S_I) \\ &\quad + D(S_{X_{\setminus I}|IX_IY_I}\|P_{X_{\setminus I}|X_IY_I}|S_{IX_IY_I}) \qquad (40) \end{aligned}$$

the rest of the proof can now be shown in the same manner as therein, by iteratively applying the (in)equalities (35) and (40) and noting that the second terms in both are martingales. ∎

B. Proof of Lemma 4

First, let us observe that, since $P_{U|X}$ in $d_c^\star(S_{XY})$ depends on S_{XY}, differentiating $d_c^\star(S_{XY})$ w.r.t. $S_{XY}(x,y)$ can be done via the total derivative

$$\frac{\partial d_c^\star(S_{XY})}{\partial S_{XY}(x,y)} + \frac{\partial d_c^\star(S_{XY})}{\partial P_{U|X}(u|x)}\frac{\partial P_{U|X}(u|x)}{\partial S_{XY}(x,y)}$$

where in the first term $P_{U|X}$ is fixed. Recalling that $d_c^\star(S_{XY})$ optimizes over $P_{U|X}$, we must have $\frac{\partial d_c^\star(S_{XY})}{\partial P_{U|X}(u|x)} = 0$, and it therefore suffices to compute the first term. In order to compute $\frac{\partial d_c^\star(S_{XY})}{\partial S_{XY}(x,y)}$, we may note that e.g. $S_{X|U}(x|u) = (\sum_{\bar{y}} S_{XY}(x,\bar{y})P_{U|X}(u,x))/(\sum_{\bar{x},\bar{y}} S_{XY}(\bar{x},\bar{y})P_{U|X}(u,\bar{x}))$ and hence by the quotient rule

$$\frac{\partial S_{X|U}(u,x)}{\partial S_{XY}(x,y)} = P_{U|X}(u|x)\left(\frac{1}{S_U(u)} - \frac{S_{X|U}(x|u)}{S_U(u)}\right).$$

Using this same strategy for other probabilities that depend on S_{XY}, the result then follows. ∎

REFERENCES

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. i. secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[2] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[3] A. A. Hassan, W. E. Stark, J. E. Hershey, and S. Chennakeshu, "Cryptographic key agreement for mobile radio," *Digital Signal Processing*, vol. 6, no. 4, pp. 207–212, Oct. 1996.

[4] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 344–366, Mar. 2000.

[5] V. Strassen, "Asymptotic estimates in Shannon's information theory," in *Proc. 3rd Trans. Prague Conf. Inf. Theory*, 1962, pp. 689–723.

[6] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4974–4966, Nov. 2009.

[7] Y. Polyanskiy, 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.

[8] Y. Polyanskiy, *Channel Coding: Non-Asymptotic Fundamental Limits*, Ph.D. dissertation, Princeton University, 2010.

[9] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.

[10] M. Hayashi, H. Tyagi, and S. Watanabe, "Strong converse for a degraded wiretap channel via active hypothesis testing," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Oct. 2014.

[11] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Non-asymptotic and second-order achievability bounds for coding with side-information," Dec. 2014, [Online]. Available: https://arxiv.org/abs/1301.6467.

[12] J. Liu, *Information Theory from A Functional Viewpoint*, Ph.D. thesis, Princeton University, 2018.

[13] J. Liu, "Dispersion bound for the Wyner-Ahlswede-Körner network via a semigroup method on types," *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 869–885, Feb. 2021.

[14] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.

[15] J. Liu, P. Cuff, and S. Verdú, "Secret key generation with one communicator and a one-shot converse via hypercontractivity," in *Proceedings of the IEEE International Symposium on Information Theory*, June 2015.

[16] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, 2 edition, 1997.