Smoothing Brascamp-Lieb Inequalities and Strong Converses for Common Randomness Generation

Jingbo Liu*, Thomas A. Courtade†, Paul Cuff* and Sergio Verdú*

*Department of Electrical Engineering, Princeton University
†Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
Email: {jingbo,cuff,verdu}@princeton.edu, courtade@eecs.berkeley.edu

Abstract—**We study the infimum of the best constant in a functional inequality, the Brascamp-Lieb-like inequality, over auxiliary measures within a neighborhood of a product distribution. In the finite alphabet and the Gaussian cases, such an infimum converges to the best constant in a mutual information inequality. Implications for strong converse properties of two common randomness (CR) generation problems are discussed. In particular, we prove the strong converse property of the rate region for the omniscient helper CR generation problem in the discrete and the Gaussian cases. The latter case is a rare instance of a strong converse for a continuous source when the rate region involves auxiliary random variables.**

I. Introduction

In the last few years, information theory has seen vibrant developments in the study of the non-vanishing error probability regime, and in particular, the successes in applying normal approximations to gauge the back-off from the asymptotic limits as a function of delay. Extending the achievements for point-to-point communication systems in [1][2][3] to network information theory problems usually requires new ideas for proving tight non-asymptotic bounds. For achievability, single-shot covering lemmas and packing lemmas [4][5] supply convenient tools for distilling single-shot achievability bounds from the classical asymptotic achievability proofs. These single-shot bounds are easy to analyze in the stationary memoryless case by choosing the auxiliary random variables to be i.i.d. and applying the law of large numbers or the central limit theorem.

In contrast, there are few examples of single-shot converse bounds in the network setting. Indeed, unlike their achievability counterparts, single-shot converses are often non-trivial to single-letterize to a strong converse. In fact, there are few methods for obtaining strong converses for network information theory problems whose single-letter solutions involve auxiliaries; see e.g. [6, Section 9.2 "Open problems and challenges ahead"]. Exceptions include the strong converses for select source networks [7] where the method of types plays a pivotal role.

In this paper, through the example of a common randomness (CR) generation problem [8, Theorem 4.2], we demonstrate the power of a functional inequality, the *Generalized Brascamp-Lieb-like (GBLL) inequality* [9]:

$$\int \exp\left(\sum_{j=1}^{m} \mathbb{E}[\log f_j(Y_j)|X=x] - d\right) \mathrm{d}\mu(x) \le \prod_{j=1}^{m} \|f_j\|_{\frac{1}{c_j}}, \quad (1)$$

in proving single-shot converses for problems involving multiple sources. Here μ, $(Q_{Y_j|X})$, (ν_j), (c_j), d are given and $\|f_j\|_{\frac{1}{c_j}} := \left(\int f_j^{1/c_j} \mathrm{d}\nu_j\right)^{c_j}$. For the CR generation problem, a single-shot converse bound is given in terms of any (c_j) and d that satisfy (1) (see Theorem 20 and the succeeding paragraph for an outline). Then, the key tool for showing strong converses from such single-shot converses is the "achievability" of the following problem: infimize the best constant d in (1) with the substitutions $\mu \leftarrow \mu_n$, $\nu_j \leftarrow \nu_j^{\otimes n}$ and $Q_{Y_j|X} \leftarrow Q_{Y_j|X}^{\otimes n}$, where the auxiliary measure μ_n is within a neighborhood (say in total variation) of $\mu^{\otimes n}$. Interestingly, a product μ_n is generally good. On the surface, this is reminiscent of the smooth Rényi entropy [10], for which the infimum (for order $\alpha < 1$) or supremum (for $\alpha > 1$) over an auxiliary measure within a neighborhood of a product distribution behaves like the Shannon entropy. However, smooth GBLL appears to be a much deeper problem, whose resolution requires far finer tools than weak typicality.

The general philosophy appears to be that under certain regularity conditions, $\frac{d}{n}$ (where d is the best constant in the setting of product measures and smoothing above) converges to the best constant in a mutual information inequality. We provide a general approach for verifying this principle, and apply it to the discrete memoryless and the Gaussian source (and more general sources in a future paper). When this principle holds, the single-shot converse establishes the strong converse property.

The proposed approach to strong converses has two main advantages compared with the method of types approach in [7], which are nicely illustrated by the example of CR generation: 1) The argument covers possibly stochastic decoders. Note that for the omniscient helper CR generation problem, there is no indication that deterministic encoders are non-asymptotically optimal. 2) As illustrated by the Gaussian example, the approach is applicable to non-discrete sources where the method of types is futile. This is perhaps the first instance of a strong converse for a continuous source where the rate region involves auxiliaries. We also refine the analysis to bound the second-order rate.

In addition, we discuss the "converse" part of smooth BLL, which generally follows from the achievability of CR generation problems. In fact, smooth BLL and CR generation may be considered as dual problems where the achievability of one implies the converse of the other, and vice versa.[1]

It is also interesting to note that for hypercontractivity, which is a special case of the BLL inequality with the best constant being zero, Anantharam et al. [12] showed the equivalence between a relative entropy inequality and a mutual information inequality. This equivalence is lost for positive best constants. Thus smooth BLL is a conceptually satisfying way to regain the connection between these two inequalities.

This work was supported in part by NSF Grants CCF-1528132, CCF-0939370 (Center for Science of Information), CCF-1319299, CCF-1319304, CCF-1350595 and AFOSR FA9550-15-1-0180.

[1] Another example of such "dual problems" is channel resolvability and identification coding [11].

Omitted proofs can be found in [13].

II. Preliminaries

Definition 1. Given a nonnegative μ on $\mathcal{X}$, ν_j on $\mathcal{Y}_j$, and random transformations $Q_{Y_j|X}$, and $c_j \in (0,\infty)$, $j \in \{1,\dots,m\}$, define

$$\mathrm{d}(\mu,(Q_{Y_j|X}),(\nu_j),c^m) := \sup\left\{\sum_{l=1}^m c_l D(P_{Y_l}\|\nu_j) - D(P_X\|\mu)\right\}$$

where the sup is over $P_X \ll \mu$ and $P_X \to Q_{Y_j|X} \to P_{Y_j}$[2].

We abbreviate the notation in Definition 1 as $\mathrm{d}(\mu,\nu_j,c^m)$ when there is no confusion.

Note that μ and ν_j are not necessarily probability measures, and $\mu \to Q_{Y_j|X} \to \nu_j$ need not hold. These liberties are useful, e.g., in the proof of Theorem 13. Generalizing an approach of Carlen and Cordero-Erausquin [14] (see also the independent work of Nair [15]), we established the following [9]:

Proposition 2. *Under the assumptions of Definition 1,* $\mathrm{d}(\cdot)$ *is the minimum* d *such that* (1) *holds for all nonnegative measurable functions* f_j.

We call (1) a *generalized Brascamp-Lieb-like inequality* (GBLL). The case of deterministic $Q_{Y_j|X}$ was considered in [14], which we call a *Brascamp-Lieb-like inequality* (BLL). In the special case where $Q_{Y_j|X}$'s are a linear projections and μ and ν_j are either Gaussian or Lebesgue, (1) is called a Brascamp-Lieb inequality; it is well-known that a Brascamp-Lieb inequality holds for a specific value of d if and only if it holds for all Gaussian functions (f_j) [16].

Definition 3. For nonnegative measures ν and μ on the same measurable space $(\mathcal{X},\mathscr{F})$ and $\gamma \in [1,\infty)$, define

$$E_\gamma(\nu\|\mu) := \sup_{\mathcal{A}\in\mathscr{F}}\{\nu(\mathcal{A}) - \gamma\mu(\mathcal{A})\}. \tag{2}$$

Note that under this definition $E_1(P\|\mu)$ does not equal $\frac{1}{2}|P-\mu|$ if μ is not a probability measure. Properties of E_γ used in this paper can be found in [17].

Definition 4. For $\delta \in [0,1)$, Q_X, $(Q_{Y_j|X})$ and (ν_j), define

$$\mathrm{d}_\delta(Q_X,\nu_j,c^m) := \inf_{\mu:\, E_1(Q_X\|\mu)\le\delta} \mathrm{d}(\mu,\nu_j,c^m). \tag{3}$$

In the stationary memoryless case, define the *δ-smooth GBLL rate*[3]

$$\mathrm{D}_\delta(Q_X,\nu_j,c^m) := \limsup_{n\to\infty}\frac{1}{n}\mathrm{d}_\delta(Q_X^{\otimes n},\nu_j^{\otimes n},c^m), \tag{4}$$

and the *smooth GBLL rate* is the limit

$$\mathrm{D}_{0^+}(Q_X,\nu_j,c^m) := \lim_{\delta\downarrow 0}\mathrm{D}_\delta(Q_X,\nu_j,c^m). \tag{5}$$

Remark 5. Allowing unnormalized measures avoids the unnecessary step of normalization in the proof, and is in accordance with the literature on smooth Rényi entropy, where such a relaxation generally gives rise to nicer properties and tighter non-asymptotic bounds, cf. [10][17].

Definition 6. Given Q_X, $(Q_{Y_j|X})$ and $c^m \in (0,\infty)^m$, define

$$\mathrm{d}^\star(Q_X,c^m) := \sup_{P_{U|X}}\left\{\sum_{l=1}^m c_l I(U;Y_l) - I(U;X)\right\}. \tag{6}$$

[2]This notation means that P_{Y_j} is the output of the random transformation $Q_{Y_j|X}$ when the input is P_X.

[3]As is clear from the context, the random transformations implicit on the right side of (4) are $(Q_{Y_j|X}^{\otimes n})$.

We say Q_X, $(Q_{Y_j|X})$ and (c_j) satisfy the *δ-smooth property* if

$$\mathrm{D}_\delta(Q_X,Q_{Y_j},c^m) = \mathrm{d}^\star(Q_X,c^m), \tag{7}$$

and the *strong smooth property* if (7) holds for all $\delta \in (0,1)$.

From these definitions and a tensorization property of $\mathrm{d}(\cdot)$ [9] we clearly have

$$\mathrm{d}(Q_X,Q_{Y_j},c^m) = \mathrm{D}_0(Q_X,Q_{Y_j},c^m) \ge \mathrm{D}_\delta(Q_X,Q_{Y_j},c^m). \tag{8}$$

Generally, the relationship between $\mathrm{D}_\delta(Q_X,Q_{Y_j},c^m)$ and $\mathrm{d}^\star(Q_X,c^m)$ is not clear. In the rest of the paper, however, we show their relations under various conditions to make the point that $\mathrm{D}_\delta(Q_X,Q_{Y_j},c^m) = \mathrm{d}^\star(Q_X,c^m)$ in "sufficiently regular" cases.

III. Achievabilities for Smooth GBLL

Under various conditions, we provide upper bounds on $\mathrm{D}_\delta(Q_X,Q_{Y_j},c^m)$, establishing the achievability part of the strong smooth property. (As alluded before, achievability of smooth GBLL implies converses of operational problems like CR generation.)

A. Hypercontractivity

If $\mathrm{d}^\star(Q_X,c^m) = 0$, by an extension of the proof of equivalent formulations of hypercontractivity [12] we also have $\mathrm{d}(Q_X,Q_{Y_j},c^m) = 0$, establishing that $\mathrm{D}_0(Q_X,Q_{Y_j},c^m) = \mathrm{d}^\star(Q_X,c^m)$.

B. Finite $|\mathcal{X}|$, and Beyond

The main objective of this section is to show:

Theorem 7. $\mathrm{D}_{0^+}(Q_X,Q_{Y_j},c^m) \le \mathrm{d}^\star(Q_X,c^m)$ *if* $\mathcal{X}$ *is finite.*

We present a general method of proving achievability of smooth GBLL which, although not intuitive at the first sight, turns out to be successful for the distinct cases of the discrete and Gaussian sources. The following tensorization result is useful:

Lemma 8. *Suppose* $\tau_\alpha\colon \mathcal{X}\to\mathbb{R}$ *is measurable for each (abstract) index* $\alpha \in \mathcal{A}$. *Fix any* $\epsilon \in (0,1)$, *and for each* $n \in \{1,\dots\}$ *define* $g(n)$ *as the supremum of*

$$\frac{1}{n}\left[\sum_j c_j D(P_{Y^n|U}\|\nu_j^{\otimes n}|P_U) - D(P_{X^n|U}\|\mu^{\otimes n}|P_U)\right] \tag{9}$$

over P_{UX^n} *such that* $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \tau_\alpha(\hat{X}_i)\right] \le \epsilon$, *where* $\hat{X}^n \sim P_{X^n}$ *and* $P_{UX^nY^n} := P_{UX^n}Q_{Y|X}^{\otimes n}$. *Then* $g(n) \le g(1)$.

The functions $\tau_\alpha(\cdot)$ can be thought of as (possibly negative) cost functions that enforce the P_{UX} maximizing (9) to satisfy $P_X \approx Q_X$. If the probability that an i.i.d. sequence induces a small cost is large, then one can choose the μ in the definition of the smooth property to be the restriction[4] of $Q_X^{\otimes n}$ on such a set. Therefore the following consequence of Lemma 8, will be the key to our proofs of the smooth property:

Lemma 9. *Suppose* τ_α *is as in Lemma 8 and define*

$$\mathcal{S}_\epsilon^n := \left\{x^n\colon \frac{1}{n}\sum_{i=1}^n \tau_\alpha(x_i) \le \epsilon\right\}. \tag{10}$$

[4]In this paper, by restriction of a measure on a set we mean the result of cutting off the measure outside that set (without renormalizing).

If P_{X^n} is supported on $\mathcal{S}_\epsilon^n$ for each n, then

$$\limsup_{n\to\infty}\frac{1}{n}\left[\sum_j c_j D(P_{Y_j^n}\|\nu_j^{\otimes n})-D(P_{X^n}\|\mu^{\otimes n})\right]$$
$$\le \sup\left\{\sum c_j D(P_{Y_j|U}\|\nu_j|P_U)-D(P_{X|U}\|\mu|P_U)\right\} \quad (11)$$

where the sup on the right side is over P_{UX} such that $\mathbb{E}[\tau_\alpha(\hat{X})]\le\epsilon$, and $\hat{X}\sim P_X$.

A remarkable aspect of Lemma 9 is that the left side of (11), which is a multi-letter quantity from the definition of $\mathrm{d}(\cdot)$, is upper bounded by a single-letter quantity.

Lemma 10. *Suppose $(\mathcal{X},\mathscr{F})$ is a second countable topological space and Q_X is a Borel measure. Define*

$$\sigma\colon P_X\mapsto\sum c_j D(P_{Y_j}\|Q_{Y_j})-D(P_X\|Q_X). \quad (12)$$

If ϕ, the concave envelope of σ, is upper semicontinuous at Q_X, then $\mathrm{D}_{0^+}(Q_X,Q_{Y_j},c^m)\le\mathrm{d}^\star(Q_X,c^m)$.

Remark 11. If $c_1=\dots=c_m=0$, then $\phi(P_X)=-D(P_X\|Q_X)$ always satisfies the upper semicontinuity in Lemma 10 because of the weak semicontinuity of the relative entropy. On the other hand, taking $m=1$, $c_1=2$, Q_X any distribution on a countably infinite alphabet with $H(Q_X)<\infty$, and $Q_{Y_1|X}$ the identity transformation, we see $\sigma(P_X)=H(P_X)+D(P_X\|Q_X)$ and the upper semicontinuity condition in Lemma 10 fails.

Proof of Theorem 7. Assume w.l.o.g. that $Q_X(x)>0,\forall x$ since otherwise we can delete x from $\mathcal{X}$. Then Q_X is in the interior of the probability simplex. Moreover $\phi(\cdot)$ in Lemma 10 is clearly bounded. Thus by a property of convex functions [18, Theorem 7.4], the weak semicontinuity in Lemma 10 is fulfilled. □

Remark 12. For general $\mathcal{X}$, one cannot use the property of convex functions to conclude the semicontinuity as in the proof of Theorem 7. In fact, whenever $|\mathcal{X}|=\infty$, there are points in $\mathcal{X}$ with arbitrarily small probability, thus Q_X cannot be in the interior of the probability simplex even under the stronger topology of total variation.

C. Gaussian Case

The semicontinuity assumption in Lemma 10 appears too strong for the case of the Gaussian distribution, which has a non-compact support. Nevertheless, we can proceed by picking a different $\tau_\alpha(\cdot)$ in Lemma 9.

Theorem 13. $\mathrm{D}_{0^+}(Q_X,Q_{Y_j},c^m)\le\mathrm{d}^\star(Q_X,c^m)$ *if $Q_{\mathbf{X}}$ and $(Q_{\mathbf{Y}_j|\mathbf{X}})$ are Gaussian.*

The proof hinges on our prior result [9] about the Gaussian optimality in an optimization under a covariance constraint: suppose μ and ν_j are the Lebesgue measures. Define

$$F(\mathbf{M}):=\sup\left\{-\sum c_j h(\mathbf{Y}_j|U)+h(\mathbf{X}|U)\right\} \quad (13)$$
$$=\sup\left\{\sum c_j D(P_{\mathbf{Y}_j|U}\|\nu_j|P_U)-D(P_{\mathbf{X}|U}\|\mu|P_U)\right\} \quad (14)$$

where the suprema are over $P_{U\mathbf{X}}$ such that $\mathbf{\Sigma}_{\mathbf{X}}\preceq\mathbf{M}$. Also suppose w.l.o.g. that $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$ under $Q_{\mathbf{X}}$.

Proposition 14 ([9]). *$F(\mathbf{M})$ equals the sup in (14) restricted to constant U and Gaussian $\mathbf{X}$, which implies that*

$$F(\mathbf{\Sigma})+C=\mathrm{d}^\star(Q_{\mathbf{X}},Q_{\mathbf{Y}_j},c^m) \quad (15)$$

where

$$C:=\sum_j c_j h(\mathbf{Y}_j)-h(\mathbf{X}_j). \quad (16)$$

Proof of Theorem 13. Put $\mathcal{A}$ as the set of unit length vectors in $\mathcal{X}$ (a Euclidean space), and for each $\alpha\in\mathcal{A}$ define $\tau_\alpha(\mathbf{x}):=(\alpha^\top\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{x})^2-1$. Now, observe that for $\mathbf{x}^n\in\mathcal{X}^n$,

$$\frac{1}{n}\sum_i\tau_\alpha(\mathbf{x}_i):=\alpha^\top\mathbf{\Sigma}^{-\frac{1}{2}}\left(\frac{1}{n}\sum_i\mathbf{x}\mathbf{x}^\top\right)\mathbf{\Sigma}^{-\frac{1}{2}}\alpha-1, \quad (17)$$

so $\frac{1}{n}\sum_i\tau_\alpha(\mathbf{x}_i)\le\epsilon_1$ for all $\alpha\in\mathcal{A}$ is equivalent to the bound on the empirical covariance: $\frac{1}{n}\sum_i\mathbf{x}\mathbf{x}^\top\preceq(1+\epsilon_1)\mathbf{\Sigma}$. Consider also the "weakly typical set" $\mathcal{T}_{\epsilon_2}^n$, defined as the set of sequences $\mathbf{x}^n$ such that

$$\frac{1}{n}\sum_i\left[\imath_{Q_{\mathbf{X}}\|\mu}(\mathbf{x}_i)-\sum_j c_j\mathbb{E}[\imath_{Q_{\mathbf{Y}_j}\|\nu_j}(\mathbf{Y}_j)|\mathbf{X}=\mathbf{x}_i]\right]\le C+\epsilon_2 \quad (18)$$

where C was defined in (16). Now set μ_n as the restriction of $Q_{\mathbf{X}}^{\otimes n}$ on $\mathcal{S}_{\epsilon_1}^n\cap\mathcal{T}_{\epsilon_2}^n$. If $P_{\mathbf{X}^n}\ll\mu_n$, by Lemma 9 we have

$$\limsup_{n\to\infty}\frac{1}{n}\left[\sum_j c_j D(P_{\mathbf{Y}_j^n}\|\nu_j^{\otimes n})-D(P_{\mathbf{X}^n}\|\mu^{\otimes n})\right]\le F((1+\epsilon_1)\mathbf{\Sigma}). \quad (19)$$

Since $P_{\mathbf{X}^n}$ is supported on $\mathcal{T}_{\epsilon_2}^n$, we also have

$$\frac{1}{n}\left[\sum_j c_j D(P_{\mathbf{Y}_j^n}\|\nu_j^{\otimes n})-D(P_{\mathbf{X}^n}\|\mu^{\otimes n})\right]+C$$
$$\ge\frac{1}{n}\left[\sum_j c_j D(P_{\mathbf{Y}_j^n}\|Q_{\mathbf{Y}_j}^{\otimes n})-D(P_{\mathbf{X}^n}\|Q_{\mathbf{X}}^{\otimes n})\right]-\epsilon_2 \quad (20)$$

Hence from (19)-(20) we conclude

$$\limsup_{n\to\infty}\frac{1}{n}\left[\sum_j c_j D(P_{\mathbf{Y}_j^n}\|Q_{\mathbf{Y}_j}^{\otimes n})-D(P_{\mathbf{X}^n}\|\mu_n)\right]$$
$$\le F((1+\epsilon_1)\mathbf{\Sigma})+C+\epsilon_2 \quad (21)$$

where we used $D(P_{\mathbf{X}^n}\|Q_{\mathbf{X}}^{\otimes n})=D(P_{\mathbf{X}^n}\|\mu_n)$. Also, by the law of large numbers, $\lim_{n\to\infty}Q_{\mathbf{X}}^{\otimes n}(\mathcal{S}_{\epsilon_1}^n\cap\mathcal{T}_{\epsilon_2}^n)=1$ so $\lim_{n\to\infty}E_1(Q_{\mathbf{X}}^{\otimes n}\|\mu_n)=1$. Thus (21), Proposition 14 and the continuity of F (which can be verified since (13) is essentially a matrix optimization problem) imply the desired result. □

IV. Converse for the One-Communicator Problem

We prove a single-shot bound connecting smooth GBLL and one-communicator CR generation [8, Theorem 4.2][19], allowing us to prove the converse of one using the achievability of the other. Note that the one-communicator CR generation model is different from other models of CR generation where strong converse properties have been shown using the method of type or meta-converse arguments (and the rate regions of which do not involve auxiliary random variables); see e.g. the two party model with unlimited interactive communication over noisy channels [20] and the multiparty model with unlimited noiseless communications [21] [22].

Let Q_{XY^m} be the joint distribution of sources X, Y_1, …, Y_m, observed by terminals T_0, …, T_m as shown in Figure 1. The communicator T_0 computes the integers $W_1(X)$, …, $W_m(X)$ and sends them to $\mathsf{T}_1,\dots,\mathsf{T}_m$, respectively. Then, terminals T_0, …, T_m compute integers $K(X)$, $K_1(Y_1,W_1)$,…,

$K_m(Y_m, W_m)$. The goal is to produce $K = K_1 = \cdots = K_m$ with high probability with K almost equiprobable.

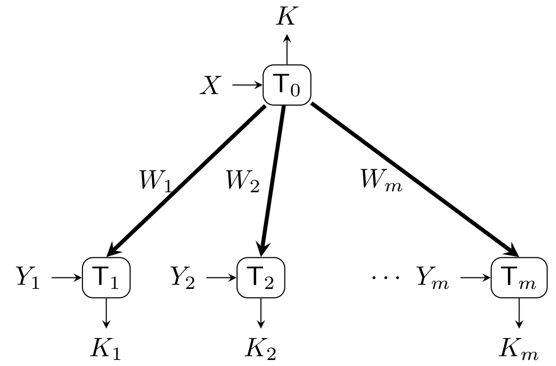

Figure 1: CR generation with one-communicator

In the stationary memoryless case, put $X \leftarrow X^n$, $Y_j \leftarrow Y_j^n$. Denote by R and R_j the rates of K and W_j, respectively. Under various performance metrics (cf. [8][19]), the achievable region[5] is the set of $(R, R_1, \ldots, R_m)$ such that

$$\mathrm{d}^\star(Q_X, c^m) + \sum_j c_j R_j \geq \left(\sum_j c_j - 1\right) R \tag{22}$$

for all $c^m \in (0,\infty)^m$.

Theorem 15 (Strong converse for one-communicator CR generation)**.** *Let $|\mathcal{X}|$, $|\mathcal{Y}_1|, \ldots, |\mathcal{Y}_m|$ be finite, and let T_K denote the equiprobable distribution on $\mathcal{K}$. Suppose $(R, R_1, \ldots, R_m)$ fails* (22) *for some c^m, and that there exists a CR generation scheme at these rates for which*

$$\mathbb{P}[K = K_1 = \cdots = K_m] \geq 1 - \delta_1; \tag{23}$$

$$\frac{1}{2}|Q_K - T_K| \leq \delta_2 \tag{24}$$

for sufficiently large n, then $\delta_1 + \delta_2 \geq 1$.

The following lemma establishes a *single-shot* connection between one-communicator CR generation and smooth GBLL, which allows us to prove the converse of one problem from the achievability of the other. For simplicity of presentation, we state it in the case of $m = 1$.[6]

Lemma 16. *Suppose that 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* (23) *and* (24) *hold. Also, suppose that there exist μ_X, $\delta, \epsilon, \epsilon' \in (0,1)$ and $c, d \in (0,\infty)$ such that*

$$E_1(Q_X \| \mu_X) \leq \delta; \tag{25}$$

$$\mu_X\left(x \colon Q_{Y|X=x}(\mathcal{A}) \geq 1 - \epsilon'\right) \leq 2^c \exp(d) Q_Y^{c(1-\epsilon)}(\mathcal{A}) \tag{26}$$

for any $\mathcal{A} \subseteq \mathcal{Y}$. Then, for any $\delta_3, \delta_4 \in (0,1)$ such that $\delta_3\delta_4 = \delta_1 + \delta$, we have

$$\delta_2 \geq 1 - \delta - \delta_3 - \frac{1}{|\mathcal{K}|} - \frac{2^{\frac{1}{1-\epsilon}} \exp\left(\frac{d}{c(1-\epsilon)}\right) |\mathcal{W}|}{(\epsilon' - \delta_4)^{\frac{1}{c(1-\epsilon)}} |\mathcal{K}|^{1 - \frac{1}{c(1-\epsilon)}}}. \tag{27}$$

Remark 17. The relevance of Lemma 16 to smooth GBLL is seen by setting $f(y) := (1_{\mathcal{A}}(y) + Q_Y(\mathcal{A}) 1_{\bar{\mathcal{A}}}(y))^c$ in (1). We then see (26) holds for any $\epsilon = \epsilon' \in (0,1)$.

[5]We remark in passing that the corresponding *key* generation problem, which places the additional constraint that $W_j \perp K$ asymptotically for each j, is solved in [19] with a different rate region involving $m+1$ auxiliaries.

[6]Note that this problem is unlike the usual "image-size characterization" [7, Chapter 15] which is difficult to generalize to $m \geq 3$ case.

Remark 18. In the stationary memoryless case $Q_X \leftarrow Q_X^{\otimes n}$, $Q_{Y|X} \leftarrow Q_{Y|X}^{\otimes n}$, suppose $|\mathcal{X}|, |\mathcal{Y}| < \infty$. Using Theorem 7 and the blowing-up lemma [23], we can show that for any $\delta, \epsilon, \epsilon' \in (0,1)$ and $d > \mathrm{d}^\star(Q_X, c)$, there exists n large enough such that (26) is satisfied with $d \leftarrow nd$ for some μ_X (more precisely, the restriction of $Q_X^{\otimes n}$ on a strongly typical set) satisfying (25).

Another application of Lemma 16 is the following:

Theorem 19 (Weak converse for smooth GBLL)**.**

$$\mathrm{D}_{0^+}(Q_X, Q_{Y_j}, c^m) \geq \mathrm{d}^\star(Q_X, c^m) \tag{28}$$

V. Converse for the Omniscient Helper Problem

Note that Theorem 19 only establishes a weak converse for smooth GBLL and Theorem 15 is only for finite alphabets and deterministic decoders, because of the use of the blowing-up lemma. In this section we improve these results in a special case where $X = (Y_1, \ldots, Y_m)$, that is, in the special case of smooth BLL and *omniscient helper* CR generation.

To see why the problem becomes simpler in this special case, note that the set $\{x \colon Q_{Y|X=x}(\mathcal{A}) \geq 1 - \epsilon'\}$ in (26) can be regarded as the "preimage" of the set $\mathcal{A}$ under the random transformation. In the case of deterministic $Q_{Y_j|X}$, there is no difference regarding the choice of $\epsilon' \in (0,1)$. However, in general a large ϵ' may imply a large ϵ on the right side of (26). Nevertheless, under the conditions for the blowing-up lemma, ϵ' and ϵ can be chosen independently (Remark 18).

In our prior work [19], a single-shot bound was derived via hypercontractivity which shows the strong converse property of the secret key (or CR) per unit cost. From the current perspective, no smoothing is needed for that particular c^m (which can be viewed as the orientation of the supporting hyperplane) for the reason explained in Section III-A. Straightforward extensions of the analysis from hypercontractivity to BLL inequality yields only a loose outer bound for the rate region when $\mathrm{d}(Q_X, Q_{Y_j}, c^m) > \mathrm{d}^\star(Q_X, c^m)$. However, following the philosophy in the present paper, we may choose μ which is E_1-close to Q_X and expect that $\mathrm{d}(\mu, Q_{Y_j}, c^m) \approx \mathrm{d}^\star(Q_X, c^m)$. Thus by a slight change of the analysis in [19], we can show the following.

Theorem 20 (single-shot converse for omniscient helper CR generation)**.** *If $d \geq \mathrm{d}(\mu, Q_{Y_j}, c^m)$ for some μ satisfying $E_1(Q_{Y^m} \| \mu) \leq \delta$, then*

$$\frac{1}{2}|Q_{K^m} - T_{K^m}| \geq 1 - \frac{1}{|\mathcal{K}|} - \frac{\prod_{l=1}^m |\mathcal{W}_l|^{\frac{c_l}{\sum c_i}}}{|\mathcal{K}|^{1 - \frac{1}{\sum c_i}}} \exp\left(\frac{d}{\sum c_i}\right) - \delta. \tag{29}$$

where $T_{K^m}(k^m) := \frac{1}{|\mathcal{K}|} 1\{k_1 = \cdots = k_m\}$.

Note that Theorem 20 applies for stochastic encoders and decoders, and in its proof, the function $f_j(\cdot)$ in (1) will take the role of $\max_w Q_{K_j|W_jY_j}(k|w,\cdot)$. However, the intuition is best explained in the case of deterministic decoders: let $\mathcal{A}^j_{kw_j}$ be the decoding set for $K_j = k$ upon receiving w_j by T_j. Then

$$\mu(K_1 = \cdots = K_m = k) \leq \mu\left(\cap_j \cup_{w_j} \mathcal{A}^j_{kw_j}\right) \tag{30}$$

$$\leq \exp(d) \prod_j Q_{Y_j}^{c_j}\left(\cup_{w_j} \mathcal{A}^j_{kw_j}\right) \tag{31}$$

where the crucial step (31), which may be viewed as a change-of-measure from a joint distribution to uncorrelated distributions (with powers), follows by choosing indicator functions in the BLL inequality. After some manipulations, one can bound the total variation between μ_{K^m} (consequently Q_{K^m}) and T_{K^m}.

Corollary 21 (Strong converse for omniscient helper CR generation)**.** *Suppose $(R, R_1, \dots, R_m)$ fails* (22) *for some c^m, and there exist a coding scheme at rates $(R, R_1, \dots, R_m)$*

$$\frac{1}{2}|Q_{K_1 \dots K_m} - T_{K_1 \dots K_m}| \le \delta \tag{32}$$

for sufficiently large n. Then $\delta \ge 1$ if Q_{Y^m}, $(Q_{Y_j|Y^m})$ and c^m satisfy the smooth property (as in the case of discrete/Gaussian Q_{Y^m}).

In the Gaussian case, refining the analysis in Theorem 13, we can derive a second-order achievability bound for smooth BLL, which, in view of Theorem 20, implies a second-order converse bound for CR generation: for any sequence of CR generation schemes with non-vanishing error probability, we have

$$\liminf_{n\to\infty} \sqrt{n} \left[\left(\sum c_j - 1 \right) R_n - \sum c_j R_{jn} - \mathrm{d}^\star(Q_{Y^m}, c^m) \right] \le D$$

for some constant D (explicitly given in [13]), where R_n, R_{1n}, $\dots$, R_{mn} are rates at blocklength n.

Remark 22. We used slightly different performance measures for the one-communicator problem and the omniscient helper problem. If δ_1 and δ_2 satisfy (23)-(24) then $\delta \leftarrow \delta_1 + \delta_2$ satisfies (32), so a strong converse measured by (32) implies a strong converse measured by (23)-(24). On the other hand, if δ satisfies (32) then $\delta_1 \leftarrow \delta$ and $\delta_2 \leftarrow \delta$ satisfy (23)-(24). Thus the strong converse in the sense of (23)-(24) only implies a "$\frac{1}{2}$-converse" in the sense of (32).

Unlike the more general one-communicator case, the rate region for omniscient helper *key* generation can be obtained as the intersection of the region for omniscient helper CR generation and $\{R \le \min_j H(Y_j)\}$ [19]. (Though, the misleading similarities between the rate regions for the omniscient helper CR and key generation is only a coincidence from optimizing of the rate regions.) As a consequence, the strong converse for the omniscient helper key generation is also proved, since the key generation counterpart obviously places more constraints, and the strong converse property of the outer-bound $\{R \le \min_j H(Y_j)\}$ is comparatively trivial.

As alluded before, the achievability for the omniscient helper CR generation implies the strong converse for smooth BLL:

Corollary 23. *For any Q_{Y^m}, c^m, and $\delta \in (0, 1)$,*

$$\mathrm{D}_\delta(Q_{Y^m}, Q_{Y_j}, c^m) \ge \mathrm{d}^\star(Q_{Y^m}, c^m). \tag{33}$$

Theorem 20 essentially establishes a single-shot connection between the smooth BLL and omniscient helper CR generation. Thus the proof of Corollary 23 follows easily by reasoning as in the proof of Theorem 19. In fact, for a general sequence (not necessarily stationary memoryless) of sources, if the δ-smooth BLL rate is strictly smaller than the supremum of $(\sum_j c_j - 1)R - \sum_j c_j R_j$ over achievable rates, then the second and third terms on the right side of (29) can be made to vanish exponentially in the blocklength. Thus $(1-\delta)$-achievability of CR generation implies δ-converse for smooth BLL.

References

[1] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

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

[3] 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.

[4] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *50th Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL), pp. 1-8, 2012.

[5] J. Liu, P. Cuff, and S. Verdú, "One-shot mutual covering lemma and Marton's inner bound with a common message," in *Proceedings of 2015 IEEE International Symposium on Information Theory*, (Hong Kong, China), pp. 1457–1461, June 2015.

[6] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 1-2, pp. 1–184, 2014.

[7] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems (second edition)*. Cambridge University Press, 2011.

[8] 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.

[9] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Brascamp-Lieb inequality and its reverse: An information theoretic view," in *ISIT 2016*.

[10] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," in *Advances in Cryptology-ASIACRYPT 2005*, pp. 199–216, Springer, 2005.

[11] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[12] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover," *http://arxiv.org/pdf/1304.6133v1.pdf*.

[13] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Smoothing Brascamp-Lieb inequalities and strong converses for CR generation," *arXiv:1602.02216*.

[14] E. A. Carlen and D. Cordero-Erausquin, "Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities," *Geometric and Functional Analysis*, vol. 19, no. 2, pp. 373–405, 2009.

[15] C. Nair, "Equivalent formulations of hypercontractivity using information measures," *IZS Workshop*, Feb. 2014.

[16] H. J. Brascamp and E. H. Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," *Advances in Mathematics*, vol. 20, no. 2, pp. 151–173, 1976.

[17] J. Liu, P. Cuff, and S. Verdú, "E_γ-Resolvability," *arXiv:1511.07829*.

[18] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[19] J. Liu, P. Cuff, and S. Verdú, "Secret key generation with one communicator and a one-shot converse via hypercontractivity," in *Proceedings of 2015 IEEE International Symposium on Information Theory*, (Hong Kong, China), pp. 710–714, June 2015.

[20] S. Venkatesan and V. Anantharam, "The common randomness capacity of a pair of independent discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 215–224, Jan. 1998.

[21] H. Tyagi and S. Watanabe, "A bound for multiparty secret key agreement and implications for a problem of secure computing," in *Advances in Cryptology–EUROCRYPT 2014*, pp. 369–386, Springer, 2014.

[22] H. Tyagi and P. Narayan, "How many queries will resolve common randomness?," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5363–5378, Sept. 2013.

[23] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Probability Theory and Related Fields*, vol. 34, no. 2, pp. 157–177, 1976.