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 perhaps the first 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]:

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

in proving single-shot converses for problems involving multiple sources. Here μ, $(Q_{Y_j|X})$, (ν_j), (c_j), d are given and $\|f_j\|_{\frac{1}{c_j}} :=$

$\left(\int f_j^{1/c_j} \mathrm{d}\nu_j\right)^{c_j}$. The key tool for single-letterizing such single-shot converses to strong 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 not a good choice. On the surface, this is reminiscent of the smooth Rényi entropy [10], who showed that the infimum (resp. supremum) of the Rényi entropy of order $\alpha < 1$ (resp. $\alpha > 1$) of an auxiliary measure with a neighborhood of a product distribution behaves like the Shannon entropy. In reality, the smooth version of GBLL appears to be a much deeper problem, since structure at a finer resolution than weak typicality is involved.

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. When this principle holds, our single-shot converse proves the strong converse for the CR generation problem.

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. 2) As illustrated by the Gaussian example, the approach is applicable to some non-discrete sources where the method of types is futile. This is perhaps the first instance of a strong converse for a continuous source when 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.

Omitted proofs are given in the appendices of [13].

---

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

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 $c_j \in (0, \infty)$, $j \in \{1, \ldots, m\}$, 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 $P_X \to Q_{Y_j|X} \to P_{Y_j}$.

We shall 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 in [14], 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 shall 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 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) [15].

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

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

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 [16].

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

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

In the stationary memoryless case, define the δ-*smooth GBLL rate*[2]

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

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][16].

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]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

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

(weak) smooth property if $\mathrm{D}_{0+}(Q_X, Q_{Y_j}, c^m) = \mathrm{d}^\star(Q_X, c^m)$, and *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

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

The goal is to explore conditions for $\mathrm{D}_\delta(Q_X, Q_{Y_j}, c^m) = \mathrm{d}^\star(Q_X, c^m)$.

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.

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 that

Theorem 7. $\mathrm{D}_{0+}(Q_X, Q_{Y_j}, c^m) \leq \mathrm{d}^\star(Q_X, c^m)$ *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* $\tau_\alpha : \mathcal{X} \to \mathbb{R}$ *is measurable for each (abstract) index* $\alpha \in \mathcal{A}$. *Fix any* $\epsilon \in (0, 1)$, *and for each* $n \in \{1, \ldots\}$ *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* $\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \tau_\alpha(\hat{X}_i) \right] \leq \epsilon$, *where* $\hat{X}^n \sim P_{X^n}$ *and* $P_{UX^nY^n} := P_{UX^n} Q_{Y|X}^{\otimes n}$. *Then* $g(n) \leq g(1)$.

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[3] of $Q_X^{\otimes n}$ on such a set. Therefore the following lemma will be the key to our proofs of the smooth property:

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

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

[3]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

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

$$\leq \sup\left\{\sum c_j D(P_{Y_j|U}\|\nu_j|P_U) - D(P_{X|U}\|\mu|P_U)\right\} \quad (11)$$

where the sup on the right is over P_{UX} such that $\mathbb{E}[\tau_\alpha(\hat{X})] \leq \epsilon$.

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*

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

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

Remark 11. If $c_1 = \cdots = c_m = 0$, 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 [17, Corollary 7.4.1], 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. $\mathrm{D}_{0+}(Q_X, Q_{Y_j}, c^m) \leq \mathrm{d}^\star(Q_X, c^m)$ *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 supremums 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 $\frac{1}{n}\sum_i \tau_\alpha(\mathbf{x}_i) \leq \epsilon_1$ 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

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

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

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

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

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

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

Hence from (19)-(20) we conclude

$$\limsup_{n\to\infty}\frac{1}{n}\left[\sum_j c_j D(P_{\mathbf{Y}_j^n}\|Q_{\mathbf{Y}_j}^{\otimes n}) - D(P_{\mathbf{X}^n}\|\mu_n)\right]$$

$$\leq F((1+\epsilon_1)\mathbf{\Sigma}) + C + \epsilon_2 \quad (21)$$

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], allowing us to prove the converse of one using the achievability of the other.

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 T_1, ..., 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.

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
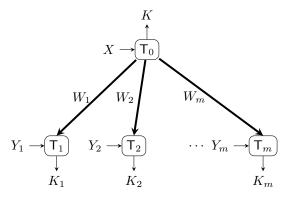
Figure 1: CR generation with one-communicator

various performance metrics (cf. [8][18]), the achievable region is the set of $(R, R_1, \ldots, R_m)$ such that

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

for all $c^m \in (0, \infty)^m$. [4]

Theorem 15 (Strong converse for one-communicator CR generation). *For finite* $|\mathcal{X}|$, $|\mathcal{Y}_1|, \ldots, |\mathcal{Y}_m|$, *suppose* $(R, R_1, \ldots, R_m)$ *fails* (22) *for some* c^m. *If* (δ_1, δ_2) *is such that*

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

$$\frac{1}{2}|Q_K - T_K| \leq \delta_2 \qquad (24)$$

can hold for some CR generation scheme at rates $(R, R_1, \ldots, R_m)$ *for sufficiently large* n *where* T_K *is the equiprobable distribution on* $\mathcal{K}$, *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 $m = 1$.[5]

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*

$$E_1(Q_X \| \mu_X) \leq \delta; \qquad (25)$$

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

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*

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

Remark 17. The relevance of the 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 $\epsilon = \epsilon' \in (0, 1)$.

[4]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 [18] with a different rate region involving $m + 1$ auxiliaries.

[5]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 the blowing-up lemma [19], 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).

Proof of Theorem 15: Again consider $m = 1$ case for simplicity. Suppose that (R, R_1) is such that (22) fails for some $c > 0$. Then, there is $\epsilon \in (0, 1)$ and $d > \mathrm{d}^\star(Q_X, c)$ such that (29) does not hold. If we choose $\delta > 0$ arbitrarily small, then δ_3 can be made arbitrarily close to δ_1, in which case δ_4 is forced to be close to 1. Pick $\epsilon' > \delta_4$. These choices combined with Remark 18, Theorem 7 and (27), show that $\delta_1 + \delta_2 \geq 1$. ∎

Another application of Lemma 16 is the following:

Theorem 19 (Weak converse for smooth GBLL).

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

Proof: For simplicity, we prove for the case of $m = 1$. For any $d > \mathrm{D}_{0+}(Q_X, Q_Y, c)$ (achievable rate for smooth GBLL) and any (R, R_1) achievable for one-communicator CR generation, we show that

$$\frac{d}{c(1-\epsilon)} + R_1 > R\left(1 - \frac{1}{c(1-\epsilon)}\right) \qquad (29)$$

for any $\epsilon \in (0, 1)$, which will establish (28) because of the achievable region formula (22).

We can choose $\delta, \delta_1, \delta_2, \delta_3, \delta_4$ such that $\delta_2 < 1 - \delta - \delta_3$ and $\delta_4 < \epsilon$. For large n, (23) and (24) can be satisfied, and by Remark 17, for $\epsilon' = \epsilon$, we can find μ_X satisfying (25) and (26) with $Q_X \leftarrow Q_X^{\otimes n}$, $Q_{Y|X} \leftarrow Q_{Y|X}^{\otimes n}$ and $d \leftarrow nd$. Thus (29) holds because the last term in (27) must vanish as $n \to \infty$. ∎

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 [18], 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 [18], 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*

$$\frac{1}{2}|Q_{K^m} - T_{K^m}| \geq 1 - \frac{1}{|\mathcal{K}|} - \frac{\prod_{l=1}^{m} |\mathcal{W}_l|^{\frac{c_l}{\sum c_i}}}{|\mathcal{K}|^{1 - \frac{1}{\sum c_i}}} \exp\left(\frac{d}{\sum c_i}\right) - \delta. \tag{30}$$

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 $\max_w Q_{K_j|W_j Y_j}(k|w, \cdot)$. However, the intuition is best explained in the case of deterministic decoders: let $\mathcal{A}_{kw_j}^j$ be the decoding set for $K_j = k$ upon receiving w_j by T_j. Then

$$\mu(K_1 = \cdots = K_m = k) \leq \mu\left(\cap_j \cup_{w_j} \mathcal{A}_{kw_j}^j\right) \tag{31}$$

$$\leq \exp(d) \prod_j Q_{Y_j}^{c_j}\left(\cup_{w_j} \mathcal{A}_{kw_j}^j\right) \tag{32}$$

where the crucial step (32), 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* $(R, R_1, \ldots, R_m)$ *fails* (22) *for some* c^m, *and there exist a coding scheme at rates* $(R, R_1, \ldots, R_m)$

$$\frac{1}{2}|Q_{K_1 \ldots K_m} - T_{K_1 \ldots K_m}| \leq \delta \tag{33}$$

for sufficiently large n. *Then* $\delta \geq 1$ *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

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

for some constant D (explicit formula given in [13]), where R_n, R_{1n}, ..., 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 (33), so a strong converse measured by (33) implies a strong converse measured by (23)-(24). On the other hand, if δ satisfies (33) 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 (33).

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 $\{R \leq \min_j H(Y_j)\}$ [18]. (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 $\{R \leq \min_j H(Y_j)\}$ 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* $\delta \in (0, 1)$,

$$D_\delta(Q_{Y^m}, Q_{Y_j}, c^m) \geq \mathrm{d}^\star(Q_{Y^m}, c^m). \tag{34}$$

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 a similar reasoning as 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 (30) 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, 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, 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, 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, "Information theoretic perspectives on Brascamp-Lieb inequalities," *draft*.

[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, 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." http://www.princeton.edu/~jingbo/preprints/ISITsmoothBL2016.pdf.

[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] 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.

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

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

[18] 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.

[19] 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.

[20] J. Liu, P. Cuff, and S. Verdú, "Secret key generation with one communicator and a one-shot converse via hypercontractivity," *arXiv:1504.05526v2*.

APPENDIX A
PROOF OF LEMMA 8

Let $I \in \{1, \ldots, n\}$ be an equiprobable random variable independent of all other random variables already defined. Observe that (9) equals

$$\sum_j c_j D(P_{Y_{jI}|UIY_j^{I-1}} \| \nu_j | P_{UIY_j^{I-1}}) - D(P_{X_I} \| \mu | P_{UIX^{I-1}})$$

$$\leq \sum_j c_j D(P_{Y_{jI}|UIX^{I-1}} \| \nu_j | P_{UIX^{I-1}}) - D(P_{X_I} \| \mu | P_{UIX^{I-1}})$$
$$\tag{35}$$

where (35) uses the Markov chain condition

$$\hat{Y}_{jI} - UI\hat{X}^{I-1} - \hat{Y}_j^{I-1}. \tag{36}$$

Also, $\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \tau_\alpha(\hat{X}_i)\right] \leq \epsilon$ implies that

$$\mathbb{E}[\tau_\alpha(\hat{X}_I)] \leq \epsilon. \tag{37}$$

Therefore, with the identification

$$P_{U,X} \leftarrow P_{UIX^{I-1},X_I} \tag{38}$$

we see $g(n) \leq g(1)$.

APPENDIX B
PROOF OF LEMMA 9

Each P_{X^n} such that $P_{X^n} \ll \mu_n$ satisfies

$$\mathbb{E}\left[\frac{1}{n} \sum_i \tau_\alpha(\hat{X}_i)\right] \leq \epsilon \tag{39}$$

since the random variable is bounded above by ϵ, P_{X^n}-almost surely. Then the result follows from Lemma 8 and the fact that μ_n and $\mu^{\otimes n}$ agree on the support of P_{X^n}.

APPENDIX C
PROOF OF LEMMA 10

Let $(\mathcal{B}_\alpha)$ be any finite partition of $\mathcal{X}$ compatible with $\mathscr{F}$. For α such that $Q_X(\mathcal{B}_\alpha) > 0$, define

$$\tau_\alpha(x) := \frac{1\{x \in \mathcal{B}_\alpha\}}{Q_X(\mathcal{B}_\alpha)} - 1, \tag{40}$$

and for α such that $Q_X(\mathcal{B}_\alpha) = 0$, put $\tau_\alpha = 0$ if $x \notin \mathcal{B}_\alpha$ and $\tau_\alpha = \infty$ otherwise. By the law of large numbers, the set $\mathcal{S}_\epsilon^n$ as defined in (10) satisfies

$$\lim_{n \to \infty} Q_X^{\otimes n}(\mathcal{S}_\epsilon^n) = 1. \tag{41}$$

Now we can invoke Lemma 9. Let μ_n be the restriction of $\mu^{\otimes n}$ on $\mathcal{S}_\epsilon^n$, and note that $D(P_{X^n} \| \mu^{\otimes n}) = D(P_{X^n} \| \mu_n)$ By the arbitrariness of $(\mathcal{B}_\alpha)$ and $\epsilon > 0$, we see the left side of (7) is upper-bounded by

$$\inf_{\mathcal{G}, \epsilon > 0} \sup_{P_X: P_{X|\mathcal{G}} \leq (1+\epsilon)Q_{X|\mathcal{G}}} \phi(P_X) \tag{42}$$

where $\mathcal{G}$ is a finitely generated σ-algebra (the σ-algebra generated by $(\mathcal{B}_\alpha)$), and $P_{X|\mathcal{G}}$ and $Q_{X|\mathcal{G}}$ are conditional distributions. Now choose any decreasing and vanishing sequence (ϵ_k) and a nested sequence $(\mathcal{G}_k)$ which contains a countable basis of $(\mathcal{X}, \mathscr{F})$. Then pick a sequence (P_X^k) such that

$$P_{X|\mathcal{G}_k}^k \leq (1+\epsilon_k)Q_{X|\mathcal{G}_k}^k \tag{43}$$

and

$$\lim_{k \to \infty} \phi(P_X^k) = \lim_{k \to \infty} \sup_{P_X: P_{X|\mathcal{G}_k} \leq (1+\epsilon_k)Q_{X|\mathcal{G}_k}} \phi(P_X) \tag{44}$$

where the limit on the right exists by monotone convergence. By (43),

$$\limsup_{k \to \infty} P_X^k(\mathcal{C}) \leq P_X(\mathcal{C}) \tag{45}$$

if $\mathcal{C} \in \mathcal{G}_l$ for some l. Since any closed subset can be constructed as the intersection of a nested sequence of such $\mathcal{C}$, it follows from the min-max inequality and the σ-continuity of probability measure that (45) actually holds for any closed $\mathcal{C}$, establishing that P_X^k converges weakly to Q_X. Thus the weak upper semi-continuity of $\phi(\cdot)$ and (44) imply that (42) is bounded above by $\phi(Q_X)$, as desired.

APPENDIX D
PROOF OF LEMMA 16

In the $m = 1$ case write $\hat{K} := K_1$. Define the joint measure

$$\mu_{XYWK\hat{K}} := \mu_X Q_{Y|X} Q_{W|X} Q_{K|X} Q_{\hat{K}|YW} \tag{46}$$

which we shall sometimes abbreviate as μ. Since $E_1(Q\|\mu) = E_1(Q_X\|\mu_X) \leq \delta$, (23) implies

$$\mu(K \neq \hat{K}) \leq \delta_1 + \delta. \tag{47}$$

Put

$$\mathcal{J} := \{k: \mu_{\hat{K}|K}(k|k) \geq 1 - \delta_4\}. \tag{48}$$

The Markov inequality implies that $\mu_K(\mathcal{J}) \geq 1 - \delta_3$. Now for each $k \in \mathcal{J}$, we have

$$(1 - \delta_4)\mu_k(k)$$
$$\leq \mu_{K\hat{K}}(k, k) \tag{49}$$
$$\leq \int_{\mathcal{F}_k} Q_{Y|X=x}\left(\bigcup_w \mathcal{A}_{wk}\right) d\mu_X(x) \tag{50}$$
$$\leq (1 - \epsilon')\mu_K(k) + \mu\left(x: Q_{Y|X=x}\left(\bigcup_{kw} \mathcal{A}_{kw} \geq 1 - \epsilon'\right)\right) \tag{51}$$
$$\leq (1 - \epsilon')\mu_K(k) + 2^c \exp(d) Q_Y^{c(1-\epsilon)}\left(\bigcup_w \mathcal{A}_{kw}\right), \tag{52}$$

where $\mathcal{F}_k \subseteq \mathcal{X}$ is the decoding set for K, and $\mathcal{A}_{kw}$ is the decoding set for $\hat{K}$ upon receiving w. Rearranging,

$$(\epsilon' - \delta_4)^{\frac{1}{c(1-\epsilon)}} \mu_k^{\frac{1}{c(1-\epsilon)}}(k) \leq 2^{\frac{1}{1-\epsilon}} \exp\left(\frac{d}{c(1-\epsilon)}\right) Q_Y\left(\bigcup_w \mathcal{A}_{kw}\right). \tag{53}$$

Now let $\tilde{\mu}$ be the restriction of μ_K on $\mathcal{J}$. Then summing both sides of (53) over $k \in \mathcal{J}$, applying the union bound, and noting that $\{\mathcal{A}_{kw}\}_k$ is a partition of $\mathcal{Y}$ for each w, we obtain

$$D_{\frac{1}{c(1-\epsilon)}}(\tilde{\mu}\|T) \leq \log|\mathcal{K}| - \frac{1}{1 - \frac{1}{c(1-\epsilon)}} \log \frac{2^{\frac{1}{1-\epsilon}}|\mathcal{W}|}{(\epsilon' - \delta_4)^{\frac{1}{c(1-\epsilon)}}}$$
$$- \frac{d}{c(1-\epsilon) - 1}. \tag{54}$$

The proof is completed invoking Proposition 24 below and noting that

$$E_1(Q_K\|\tilde\mu) \le E_1(Q_K\|\mu) + E_1(\mu\|\tilde\mu) \le \delta + \delta_3. \tag{55}$$

Proposition 24. *Suppose T is equiprobable on $\{1,\ldots,M\}$ and μ is a nonnegative measure on the same alphabet. For any $\alpha \in (0,1)$,*

$$E_1(T\|\mu) \ge 1 - \frac{1}{M} - \exp(-(1-\alpha)D_\alpha(T\|\mu)). \tag{56}$$

The special case of Proposition 24 when μ is a probability measure was used in the proof of [20, Theorem 10] (see equation (59) therein) to relate Rényi divergence and total variation distance. The extension to unnormalized μ can be easily proved in a similar way.

APPENDIX E
BOUND ON THE SECOND ORDER RATE FOR GAUSSIAN
OMNISCIENT HELPER CR GENERATION

Let

$$\mathbf{W} := \frac{\mathbf{A} + \mathbf{A}^\top}{\sqrt{2}} \tag{57}$$

be the standard Wigner matrix, where $\mathbf{A}$ is a square matrix with i.i.d. $\mathcal{N}(0,1)$ entries.

Denote by $\mathrm{Q}(\cdot)$ the tail probability of the standard Gaussian distribution and $\lambda_{\max}(\cdot)$ the largest eigenvalue of a matrix.

Theorem 25 (Bound on the second order rate for Gaussian omniscient helper CR generation)**.** *Assume that Q_{Y^m} is Gaussian with a non-degenerate covariance matrix, and there is a sequence of CR generation schemes such that*

$$\liminf_{n\to\infty} \sqrt{n}\left[\left(\sum c_j - 1\right)R_n - \sum c_j R_{jn} - \mathrm{d}^\star(Q_{Y^m}, c^m)\right]$$
$$> \frac{\log e}{2}\left(m - \sum c_j\right)D_1 + D_2 \tag{58}$$

for some $D_1, D_2 \in (0,1)$, where R_n, R_{1n}, ..., R_{mn} are the rates at blocklength n. Then

$$\liminf_{n\to\infty} \frac{1}{2}|Q_{K^m{}_n} - T_{K^m{}_n}| \ge \mathbb{P}[\lambda_{\max}(\mathbf{W}) \le D_1] - \mathrm{Q}\left(\frac{D_2}{\sqrt{V}}\right), \tag{59}$$

where

$$V := \mathrm{Var}\left(\sum_j c_j \imath_{Q_{Y_j}\|\nu_j}(Y_j) - \imath_{Q_{Y^m}\|\mu}(Y^m)\right). \tag{60}$$

Proof: First, observe that we will only need to consider the case of $\sum_j c_j \le m$, since otherwise $\mathrm{d}^\star(Q_{Y^m}, c^m) = \infty$ and Theorem 25 is vacuous. Indeed, suppose without loss of generality that $Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. For $\alpha \in (0,\infty)$ small enough, we can find U jointly Gaussian with Y^m such that $Y^m|U =$

$\mathbf{0} \sim \mathcal{N}(\mathbf{0}, \alpha\mathbf{I})$. Then we see

$$\mathrm{d}^\star(Q_{Y^m}, c^m) \ge \lim_{\alpha\downarrow 0}\sum_{j=1}^m \frac{c_j}{2}\log\frac{\sigma_{jj}}{\alpha} - \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{|\alpha\mathbf{I}|} \tag{61}$$

$$= \sum_{j=1}^m \frac{c_j}{2}\log\sigma_{jj} - \frac{1}{2}\log|\mathbf{\Sigma}|$$

$$+ \lim_{\alpha\downarrow 0}\frac{\sum_j c_j - m}{2}\log\frac{1}{\alpha} \tag{62}$$

$$= \infty \tag{63}$$

provided that $\sum_j c_j > m$ holds.

The proof is essentially based on a refinement of the achievability of smooth BLL: in the proof of Theorem 13, take $\epsilon_i \leftarrow \frac{D_i}{\sqrt{n}}$, $i = 1,2$ and $X = Y^m$. Then,

$$\lim_{n\to\infty}\mathbb{P}[\mathcal{S}^n_{\epsilon_1}] = \lim_{n\to\infty}\mathbb{P}\left[\frac{1}{n}\sum_i \mathbf{z}\mathbf{z}^\top \preceq (1+\epsilon_1)\mathbf{I}\right] \tag{64}$$

$$= \lim_{n\to\infty}\mathbb{P}\left[\frac{\sum_i \mathbf{z}\mathbf{z}^\top - \mathbf{I}}{\sqrt{n}} \preceq D_1\mathbf{I}\right] \tag{65}$$

$$= \mathbb{P}[\mathbf{W} \preceq D_1\mathbf{I}], \tag{66}$$

where $\mathbf{z} := \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and we applied multivariate CLT in (66). On the other hand, by CLT we have

$$\lim_{n\to\infty}\mathbb{P}[\mathcal{T}^n_{\epsilon_2}] = 1 - \mathrm{Q}\left(\frac{D_2}{\sqrt{V}}\right). \tag{67}$$

Also, a simple scaling argument shows that

$$F((1+\epsilon_1)\mathbf{\Sigma}) = F(\mathbf{\Sigma}) + \frac{\log(1+\epsilon_1)}{2}\left(m - \sum c_j\right) \tag{68}$$

$$\le F(\mathbf{\Sigma}) + \frac{\log e}{2\sqrt{n}}\left(m - \sum c_j\right)D_1. \tag{69}$$

Thus following the steps in the proof of Theorem 13, we can find $(\mu_n)_{n\ge 1}$ such that

$$E_1(Q_X^{\otimes n}\|\mu_n) \le \mathbb{P}[\lambda_{\max}(\mathbf{W}) \le D_1] - \mathrm{Q}\left(\frac{D_2}{\sqrt{V}}\right) + o(1) \tag{70}$$

$$\frac{1}{n}\mathrm{d}(\mu_n, Q_{Y_j}^{\otimes n}, c^m) \le \mathrm{d}^\star(Q_X, c^m) + \frac{\log e}{2\sqrt{n}}\left(m - \sum c_j\right)D_1$$
$$+ \frac{D_2}{\sqrt{n}}. \tag{71}$$

Now, invoke Theorem 20 with

$$\mu \leftarrow \mu_n; \tag{72}$$
$$\delta \leftarrow \delta_n := E_1(Q_X^{\otimes n}\|\mu_n); \tag{73}$$
$$d \leftarrow nd_n, \tag{74}$$

where d_n is defined as the right side of (71). Then

$$\frac{1}{2}|Q_{K^m{}_n} - T_{K^m{}_n}| \ge 1 - \frac{1}{|\mathcal{K}|} - \exp\left(-\frac{\tau}{\sum_j c_j}\sqrt{n}\right) - \delta_n \tag{75}$$

$$= 1 - \delta_n + o(1) \tag{76}$$

where $\tau > 0$ is defined as the difference between the left and right sides of (58). Thus (59) is established. ∎