Smoothing Brascamp-Lieb Inequalities and Strong Converses of Coding Theorems

Jingbo Liu, *Member, IEEE*, Thomas A. Courtade, *Senior Member, IEEE*, Paul Cuff, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*— The Brascamp-Lieb inequality in functional analysis can be viewed as a measure of the "uncorrelatedness" of a joint probability distribution. We define the smooth Brascamp-Lieb (BL) divergence as the infimum of the best constant in the Brascamp-Lieb inequality under a perturbation of the joint probability distribution. An information spectrum upper bound on the smooth BL divergence is proved, using properties of the subgradient of a certain convex functional. In particular, in the i.i.d. setting, such an infimum converges to the best constant in a certain mutual information inequality. We then derive new single-shot converse bounds for the omniscient helper common randomness generation problem and the Gray-Wyner source coding problem in terms of the smooth BL divergence, where the proof relies on the functional formulation of the Brascamp-Lieb inequality. Exact second-order rates are thus obtained in the stationary memoryless and nonvanishing error setting. These offer rare instances of strong converses/second-order converses for continuous sources when the rate region involves auxiliary random variables.**

***Index Terms*— Shannon theory, coding theorems, strong converse, finite blocklength, Brascamp-Lieb inequality, hypercontractivity, common randomness, Gray-Wyner network.**

I. Introduction

In THE last few years, information theory has witnessed 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 [3]–[5] to network information theory problems usually requires new ideas for proving tight non-asymptotic bounds. For achievability, single-shot covering lemmas and packing lemmas [6], [7] supply convenient tools for distilling single-shot achievability bounds from the classical asymptotic achievability proofs. While these single-shot bounds hold regardless of the finiteness of the alphabets or the memory, their asymptotics are easy to evaluate 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. Other single-shot achievability proof techniques for network information theory include stochastic likelihood encoder/decoder [8] and approximation of output statistics [9].

In contrast, although the binary hypothesis testing approach (and the related information spectrum approach) has been successfully applied to the single-user settings [3], [4], [10], progress on its extensions to network information problems has been modest. There are relatively few examples of single-shot converse bounds in the network setting. Moreover, unlike their achievability counterparts, it usually requires more effort to single-letterize a single-shot converse to a strong converse or a second-order converse, partly because it is not obvious that a product auxiliary distribution is optimal in the evaluation of the single-shot converse bounds (consider for example [4, Theorem 48] for point-to-point channel coding, which relies on the reduction to fixed composition). Several researchers have also noted the dearth of methods for obtaining strong converses for network information theory problems whose single-letter solutions involve auxiliaries; see e.g. [11, Section 6.3] [12, Section 9.2]. Although the method of types has proven to be applicable for the strong converses of some problems of this type (including selected source and channel networks [13], Gelfand-Pinsker coding [14], and Gray-Wyner coding [15]–[17]), the method of types crucially relies on the finite alphabet assumption. To our knowledge, no previous methods exist for establishing a strong converse for nonfinite distributions when the rate region involves an auxiliary (with the exception of certain Gaussian cases where the converse part can be reduced to a single-user problem, such as dirty paper coding [18]).

In this paper, we demonstrate the power of a functional inequality, the *Brascamp-Lieb inequality* [19]–[22], in proving single-shot converses for problems involving multiple sources and an "omniscient helper". For recent discussions on the connection between the Brascamp-Lieb inequality and information measures, see [23]–[26]. For recent studies of the computational aspects of the Brascamp-Lieb inequalities or applications in computer science, see [27]–[30].

To be concrete, consider $c_1, \ldots, c_m \in (0, \infty)$, $d \in \mathbb{R}$, a nonnegative finite measure μ on $\mathcal{Y}^m := \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$, and σ-finite measures $\nu_1, \ldots, \nu_m$ on $\mathcal{Y}_1, \ldots, \mathcal{Y}_m$. Then, an inequality of the following form is sometimes referred to

Manuscript received April 25, 2017; revised September 11, 2019; accepted October 21, 2019. Date of publication November 12, 2019; date of current version January 20, 2020. This work was supported in part by the National Science Foundation (NSF) under Grant CCF-1750430 (CAREER), Grant CCF-1704967, Grant CCF-0939370 (Center for Science of Information), Grant CCF-1116013, Grant CCF-1319299, Grant CCF-1319304, Grant CCF-1350595, and Grant AFOSR FA9550-15-1-0180. This work was presented in part at the 2015 IEEE International Symposium on Information Theory and at the 2016 IEEE International Symposium on Information Theory.

J. Liu is with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: jingbo@mit.edu).

T. A. Courtade is with the Department of Electrical Engineering and Computer Sciences, UC Berkeley, CA 94720-1770 USA (e-mail: courtade@eecs.berkeley.edu).

P. Cuff was with the Department of Electrical Engineering, Princeton University, New Jersey, NJ 08544 USA. He is now with Renaissance Technologies LLC, Long Island, NY 11733 USA (e-mail: cuff@princeton.edu).

S. Verdú was with the Department of Electrical Engineering, Princeton University, New Jersey, NJ 08544 USA (e-mail: verdu@informationtheory.org).

Communicated by N. Merhav, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2019.2953151

as a *Brascamp-Lieb type inequality* (see e.g. [22])

$$\int \prod_{j=1}^{m} f_j(y_j)\mathrm{d}\mu(y^m) \leqslant \exp(d) \prod_{j=1}^{m} \|f_j\|_{\frac{1}{c_j}}, \quad \forall f_1, \ldots, f_m \geqslant 0, \tag{1}$$

where $\|f_j\|_{\frac{1}{c_j}} := \left(\int f_j^{1/c_j} \mathrm{d}\nu_j\right)^{c_j}$. Traditionally, a *Brascamp-Lieb* (BL) inequality refers to the special case of (1) where μ and (ν_j) are Gaussian or Lebesgue measures. In that case, it is known that (1) holds if and only if it holds for all Gaussian functions (f_j) [19], [20]. In the present paper however, we do not focus on the Gaussian case, and the measures considered are not necessarily Gaussian or Lebesgue.

If we define the *Brascamp-Lieb (BL) divergence* $\mathrm{d}(\mu, (\nu_j), c^m)$ as the best (i.e. smallest possible) constant d for (1) to hold, then several well-known information measures can be recovered as special cases, such as the Rényi divergence (taking $m = 1$) or hypercontractivity (taking $m = 2$, $\mu = Q_{Y_1Y_2}$, and $\nu_j = Q_{Y_j}$, $j = 1, 2$). A key fact that we invoke is the *entropic representation* of the BL divergence: for any joint distribution Q_{Y^m},

$$\begin{aligned} &\mathrm{d}(Q_{Y^m}, (Q_{Y_j}), c^m) \\ &= \sup_{P_{Y^m} \ll Q_{Y^m}} \left\{ \sum_{j=1}^{m} c_j D(P_{Y_j} \| Q_{Y_j}) - D(P_{Y^m} \| Q_{Y^m}) \right\}. \end{aligned} \tag{2}$$

which can be derived from (1) using convex duality theory [22] or large deviation arguments [31]. Here $D(P_{Y^m} \| Q_{Y^m})$ denotes the relative entropy (we review the definitions of various information theoretic quantities in Section II). While similar objects such as hypercontractivity have recently seen applications in various converse results [1], [32], [33], one obstacle preventing them from becoming a canonical tool for network information theory is that in general, (2) can be strictly larger than

$$\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) = \sup_{Q_{U|Y^m}} \left\{ \sum_{j=1}^{m} c_j I(U; Y_j) - I(U; Y^m) \right\}, \tag{3}$$

where $I(U; Y^m)$ denotes mutual information, and the supremum is understood as over standard probability Q_{UY^m} whose Y^m-marginal is the given Q_{Y^m}. More specifically, single-shot converse bounds derived from (1) involve $\mathrm{d}(Q_{Y^m}, (Q_{Y_j}), c^m)$, whereas single-letter rate regions involving mutual information or conditional entropy have supporting hyperplane characterizations in terms of $\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m)$. For example, [1], [33] derived single-shot converse bounds for common randomness generation problems using hypercontractivity. These bounds are only first-order tight in the regime of *vanishing* communication rate, which is essentially due to the fact (observed by Anantharam *et al.* [34, Theorem 4]) that $\mathrm{d}(Q_{Y^m}, (Q_{Y_j}), c^m) = 0$ if and only if $\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) = 0$.

In order to bridge the gap between $\mathrm{d}(Q_{Y^m}, (Q_{Y_j}), c^m)$ and $\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m)$, we draw insight from the notion of smooth Rényi divergence in non-asymptotic information theory, introduced by Renner and coauthors [35]–[37]. This naturally leads us to introduce the *smooth BL divergence*: for $\delta \in (0, 1)$,

$$\mathrm{d}_\delta(Q_{Y^m}, (Q_{Y_j}), c^m) := \inf_{\mu:\ \int |\mathrm{d}\mu - \mathrm{d}Q_{Y^m}|_+ \leqslant \delta} \mathrm{d}(\mu, (Q_{Y_j}), c^m) \tag{4}$$

where the infimum is over nonnegative finite measures μ such that the positive part of the measure $\mu - Q_{Y^m}$ is at most δ. Recall that when proving the strong converse using the smooth Rényi divergence, we need to show that in the stationary memoryless setting (i.e., $Q_Y \leftarrow Q_Y^{\otimes n}$), the smooth Rényi divergence grows linearly at the rate of the relative entropy (regardless of the Rényi order). This can be done by simply taking μ to be supported on the weakly typical set, hence obviating the need for the finite alphabet assumption.

The asymptotic analysis of the smooth BL divergence, in contrast, is more elusive. A classical strong converse technique called *image-size characterization* [13] bounds the cardinalities of subsets of the strongly typical set and their images. In the setting of (1), the corresponding image-size inequality is of the form

$$|\mathcal{A}| \leqslant D \prod_{j=1}^{m} |\mathcal{A}_j|^{c_j} \tag{5}$$

for any subset $\mathcal{A}$ of the strongly typical set (w.r.t. Q_{Y^m}), while $\mathcal{A}_j$ denotes the projection of $\mathcal{A}$ to $\mathcal{Y}_j$. Inspired by this, it is natural to try μ in (4) with the conditional measure on the strongly typical set. The restriction to the strongly typical set ensures that the empirical distribution is close to Q_{Y^m}, which is reflected by the fact that Q_{UY^m} and Q_{Y^m} are consistent in (3) and the mutual information terms arise from single-letterizing the relative entropy between multi-letter distributions. In the case of finite $\mathcal{Y}^m$, this successfully shows that (see e.g. [38, Chapter 3])

$$\mathrm{d}_\delta(Q_{Y^m}^{\otimes n}, (Q_{Y_j}^{\otimes n}), c^m) = n\,\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) + O(\sqrt{n}). \tag{6}$$

However, the strong typicality approach has no hope of obtaining the exact prefactor in the $O(\sqrt{n})$ term, or even getting the sign correct. Moreover, the strong typicality approach requires an assumption of finite alphabets.

In the present paper, we adopt a different, typicality-free approach. With a simple, yet non-obvious, argument capitalizing on the property of the subgradient of a certain convex functional, we show the following single-shot bound: for any $\delta \in (0, 1)$,

$$\delta \leqslant \mathbb{P}\left[\sum_{j=1}^{m} c_j \imath_{U;Y_j}(u; Y_j) - \imath_{U;Y^m}(u; Y^m) > \mathrm{d}_\delta \right]. \tag{7}$$

Here $\mathrm{d}_\delta := \mathrm{d}_\delta(Q_{Y^m}, (Q_{Y_j}), c^m)$, the information density is defined as $\imath_{U;Y^m}(u; y^m) := \frac{\mathrm{d}Q^\star_{Y^m U}}{\mathrm{d}Q_U \times Q_{Y^m}}(u, y^m)$, with $Q^\star_{UY^m}$ being any maximizer in (3) (assuming it exists), $Y^m \sim Q_{Y^m}$, and u is any element in $\mathcal{U}$ (we will show that the term to the left of $>$ in (7) is independent of u almost surely). This indeed recovers the exact prefactor in $O(\sqrt{n})$ in (3), and does not require finite alphabets. For example, if Q_{Y^m} is a Gaussian distribution, then there exists an optimal $Q^\star_{UY^m}$ which is a jointly Gaussian distribution [39].

We apply the smooth BL divergence to the converses of two network information theory problems: omniscient helper common randomness generation [1], [40, Theorem 4.2], and Gray-Wyner source coding (including the almost lossless case with finite alphabets and squared distortion case with jointly Gaussian sources). In both cases, we first prove new single-shot converse bounds in terms of smooth BL divergence, and then perform an asymptotic analysis to obtain the exact second-order rate. The exact second-order rates for the Gray-Wyner source coding in the *discrete memoryless* cases were previously derived by Watanabe [16] and Zhou *et al.* [17] using the method of types and Fano's inequality, relying crucially on i.i.d. and finite-alphabet assumptions.

The proposed smooth BL divergence approach to non-asymptotic converses has several advantages compared with existing approaches such as the method of types, as nicely illustrated by its applications to common randomness generation and source coding:

1) In the discrete memoryless case, while the classical image-size characterization (based on strong typicality) shows that the second-order term scales as $O(\sqrt{n})$, there is no hope of obtaining the exact prefactor. In fact, the sign of the prefactor is invariably wrong when the error probability is less than $1/2$. In contrast, the smooth BL approach recovers the exact prefactor.
2) While the method of types is capable of obtaining the exact second-order prefactor in the discrete memoryless case, it is incapable of handling infinite alphabets. In contrast, our approach leads to rare instances of second-order converses for continuous sources.
3) In the omniscient helper CR generation problem, our approach has the desirable feature of allowing possibly stochastic encoders and decoders.[1] Stochastic encoders and decoders are tricky to handle with the image-size technique as it only concerns the cardinalities of the encoding and decoding sets.

In addition, we discuss the converse[2] part of smooth BL divergence, which generally follows from the achievability of CR generation problems. In fact, smooth BL divergence and CR generation may be considered as dual problems where the achievability of one implies the converse of the other.[3] Such converse proofs based on the achievability of another usually have certain advantages, partly because the achievability is constructive.

Let us remark that our application examples (common randomness generation and source coding) concern the setting in which one terminal observes the entire source realization Y^m. In other settings where such an omniscient terminal is absent (e.g. the Wyner-Alhswede-Körner source problem [42], [43]), although the definition of d_δ extends and a counterpart of (7) follows by the same proof, it requires additional efforts to connect d_δ with operational quantities (e.g. error probability). In [44] (see also [38] and [45]), this connection was achieved through a novel reverse hypercontractivity approach. Moreover, a very different approach for handling auxiliary random variables with Markov constraints (as in the Wyner-Alhswede-Körner problem) by introducing soft constraints was recently proposed by Tyagi and Watanabe [46]. This work was presented in part at ISIT 2015 [1] and ISIT 2016 [2]. The proofs in this paper differ significantly from the conference version [2].

[1]The (asymptotic) rate region with stochastic encoders can be strictly larger than with deterministic encoders, since in the former case the CR rate is unbounded whereas in the latter case it is bounded by the entropy of the sources. Regarding the decoders, we argue in Remark 9 that allowing stochasticity can strictly decrease the (single-shot) error, but within a constant factor.

[2]Since the smooth BL divergence is defined as an infimum over an auxiliary distribution, we take the liberty of referring to lower/upper bounds on the smooth BL divergence as converse/achievability results.

[3]Another example of such "dual problems" in information theory is channel resolvability and identification coding [41].

II. Preliminaries

We start by introducing the notation and the formal definitions of quantities of interest. Probability measures and random transformations are denoted by capital Latin letters, such as P and $P_{Y|X}$. Unnormalized nonnegative measures are denoted by lowercase Greek letters, and the Lebesgue measure is denoted by λ. Random variables are written in capital letters. For finite alphabets we sometimes use the notation of inner product in Euclidean space $\langle f, P\rangle := \int f \mathrm{d}P$ to denote an integral. A vector $(a_m, a_{m+1}, \ldots, a_n)$ is sometimes abbreviated as a_m^n, or a^n in the case of $m = 1$, or more simply, the boldface letter $\mathbf{a}$ if the range of the indices is clear from the context. The closure of a set $\mathcal{A}$ is denoted as $\mathrm{cl}(\mathcal{A})$.

The *relative information* between two nonnegative σ-finite measures $\mu \ll \nu$ on the same measurable space $(\mathcal{X}, \mathscr{F})$ is defined as the logarithm of the Radon-Nikodym derivative:

$$\imath_{\mu\|\nu}(x) := \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x), \quad \forall x \in \mathcal{X}. \tag{8}$$

The relative entropy and the conditional relative entropy are defined as:

$$D(P_X\|\mu_X) := \mathbb{E}[\imath_{P_X\|\mu_X}(X)]; \tag{9}$$

$$D(P_{Y|X}\|\mu_Y|P_X) := D(P_{Y|X}P_X\|\mu_Y \times P_X). \tag{10}$$

where $X \sim P_X$. Given P_{XY}, the mutual information is defined as

$$I(X;Y) := D(P_{Y|X}\|P_Y|P_X). \tag{11}$$

We use λ to denote the Lebesgue measure on a Euclidean space. Then the differential entropy and the conditional differential entropy are defined as

$$h(P_X) := -D(P_X\|\lambda); \tag{12}$$

$$h(P_{X|U}|P_U) := -D(P_{X|U}\|\lambda|P_U). \tag{13}$$

We now give formal definitions of the key quantities of interest.

Definition 1. Given a finite measure μ on $\mathcal{Y}^m$, nonnegative σ-finite measures $\nu_1, \ldots, \nu_m$ on $\mathcal{Y}_1, \ldots, \mathcal{Y}_m$, and $c_1, \ldots, c_m \in (0, \infty)$, define the *Brascamp-Lieb (BL) divergence*

$$\mathrm{d}(\mu, (\nu_j), c^m) := \sup_{P_{Y^m}} \left\{ \sum_{j=1}^m c_j D(P_{Y_j}\|\nu_j) - D(P_{Y^m}\|\mu) \right\}. \tag{14}$$

As a convention, the supremum in (14) is over $P_{Y^m} \ll \mu$ such that each term in (14) is finite, and the supremum is set to $-\infty$ if there is no such P_{Y^m}.

We remark that the choice of the collection of measures (ν_j) will depend on our applications: in converses of common randomness generation problems, ν_j will be the marginal distribution at one terminal; in source coding problems, ν_j will be the counting measure or the Lebesgue measure.

By convex duality [22] or large deviation arguments [31], the following equivalent formulation of the smooth BL divergence can be shown:

Proposition 1.

$$\mathrm{d}(\mu,(\nu_j),c^m)=\sup_{f_1,\ldots,f_m\geqslant 0}\left\{\log\int\prod_{j=1}^m f_j\,\mathrm{d}\,\mu-\sum_{j=1}^m\log\|f_j\|_{\frac{1}{c_j}}\right\}, \tag{15}$$

where $\|f_j\|_{\frac{1}{c_j}}:=\left(\int f_j^{1/c_j}\mathrm{d}\nu_j\right)^{c_j}$.

For nonnegative measures ν and μ on the same measurable space $(\mathcal{X},\mathscr{F})$ where $\nu(\mathcal{X})<\infty$, one can define the following measure of their distance (see e.g. [47])

$$E_\gamma(\nu\|\mu):=\sup_{\mathcal{A}\in\mathscr{F}}\{\nu(\mathcal{A})-\gamma\mu(\mathcal{A})\}, \tag{16}$$

for any choice of $\gamma\in[1,\infty)$. In the present paper, we will always take $\gamma=1$ and use E_1 to measure the perturbation in the definition of the smooth divergences. Note that $E_1(P\|\mu)=\int|\mathrm{d}P-\mathrm{d}\mu|^+$ in general and is *not* equal to the total variation $\frac{1}{2}|P-\mu|$ if μ is not a probability measure. In fact, if we restrict μ to be a probability measure and use the total variation in the definition of the smooth divergence instead, we would not be able to obtain the exact dispersion in the later applications in converse proofs.

Definition 2. Given a probability measure Q_{Y^m}, nonnegative σ-finite measures $(\nu_j)_{j=1}^m$ on $\mathcal{Y}_1,\ldots,\mathcal{Y}_m$, $\delta\in[0,1]$, and $c^m\in(0,\infty)^m$,

$$\mathrm{d}_\delta(Q_{Y^m},(\nu_j),c^m):=\inf_{\mu:\ E_1(Q_{Y^m}\|\mu)\leqslant\delta}\mathrm{d}(\mu,(\nu_j),c^m). \tag{17}$$

Remark 1. The Brascamp-Lieb divergence is a generalization of several information measures, including the strong data processing constant, hypercontractivity, and Rényi divergence; see a summary in [23]. For example, for $\alpha\in(1,\infty)$, the Rényi divergence between two probability measures P and Q on the same alphabet can be expressed in terms of the BL divergence:

$$D_\alpha(P\|Q)=\frac{\alpha}{\alpha-1}\,\mathrm{d}\left(P,Q,\frac{\alpha-1}{\alpha}\right). \tag{18}$$

This can be seen either from (15) and the variational formula of Rényi divergence (see e.g. [48] [49, (7)]), or (14) and the entropic representation of the Rényi divergence (see e.g. [39], [50]). Consequently, the smooth Rényi divergence [36] can be expressed in terms of a smooth BL divergence:

$$D_\alpha^\delta(P\|Q):=\inf_{\mu:\ E_1(P\|\mu)\leqslant\delta}D_\alpha(\mu\|Q) \tag{19}$$

$$=\frac{\alpha}{\alpha-1}\mathrm{d}_\delta\left(P,Q,\frac{\alpha-1}{\alpha}\right) \tag{20}$$

for $\delta\in(0,1)$.

We now introduce a quantity which plays a central role in the asymptotic characterizations of the smooth BL divergence. We first give its definition in terms of auxiliary random variables. It is well-known in information theory that auxiliary random variables take the role of convexifying sets [51], [52]. An equivalent concave envelope formulation will be given later in Remark 5.

Definition 3. Given Q_{Y^m}, (ν_j) and c^m as in Definition 2,

$$\mathrm{d}^\star(Q_{Y^m},(\nu_j),c^m):=\sup_{P_{UY^m}:\ P_{Y^m}=Q_{Y^m}}\left\{\sum_{j=1}^m c_jD(P_{Y_j|U}\|\nu_j|P_U)-I(U;Y^m)\right\}, \tag{21}$$

where $(U,Y^m)\sim P_{UY^m}$.

Remark 2. In the supremum in (21), we do not need to impose any cardinality constraint on U (we can assume that $(\mathcal{U}\times\mathcal{Y}^m,P_{UY^m})$ is any *standard probability space* such that $P_{Y^m}=Q_{Y^m}$). On the other hand, the supremum does not change if U is restricted to be finite. This follows from the same reasoning in the proof of the proverbial fact that the mutual information equals the supremum over finite partitions: first by Gelfand-Yaglom-Perez (see [53, Theorem 2.1.2]) where the relative entropy is approximated by its conditional over finite partitions consisting of subsets of $\mathcal{U}\times\mathcal{Y}^m$; second by Dobrushin (see [53, Theorem 2.1.1]), further approximation is made by finite partitions consisting of rectangle sets in $\mathcal{U}\times\mathcal{Y}^m$.

Remark 3. In the special case of $\nu_j=Q_{Y_j}$, we have

$$\mathrm{d}^\star(Q_{Y^m},(Q_{Y_j}),c^m)=\sup_{P_{U|Y^m}}\left\{\sum_{j=1}^m c_jI(U;Y_j)-I(U;Y^m)\right\}. \tag{22}$$

Remark 4. Since

$$\mathrm{d}^\star(Q_{Y^m},(\nu_j),c^m)$$

$$=\sup_{\substack{P_{UY^m}:\\ P_{Y^m}=Q_{Y^m}}}\left[\sum_{j=1}^m c_jD(P_{Y_j|U}\|\nu_j|P_U)-D(P_{Y^m|U}\|Q_{Y^m}|P_U)\right] \tag{23}$$

$$\leqslant\sup_{P_{UY^m}}\left[\sum_{j=1}^m c_jD(P_{Y_j|U}\|\nu_j|P_U)-D(P_{Y^m|U}\|Q_{Y^m}|P_U)\right] \tag{24}$$

$$\leqslant\sup_{P_{UY^m}}\left[\sup_u\left[\sum_{j=1}^m c_jD(P_{Y_j|U=u}\|\nu_j)-D(P_{Y^m|U=u}\|Q_{Y^m})\right]\right] \tag{25}$$

$$=\sup_{P_{Y^m}}\left[\sum_{j=1}^m c_jD(P_{Y_j}\|\nu_j)-D(P_{Y^m}\|Q_{Y^m})\right] \tag{26}$$

$$=\mathrm{d}(Q_{Y^m},(\nu_j),c^m), \tag{27}$$

we see that, in general,

$$\mathrm{d}^\star(Q_{Y^m},(\nu_j),c^m)\leqslant \mathrm{d}(Q_{Y^m},(\nu_j),c^m), \tag{28}$$

and the inequality can be strict (e.g. consider examples of Gaussian distributions). However, if $\mathrm{d}(Q_{Y^m},(Q_{Y_j}),c^m)>0$, then one can still show that $\mathrm{d}^\star(Q_{Y^m},(Q_{Y_j}),c^m)>0$, by taking $\mathcal{U}=\{0,1\}$, $P_U(1)=t$, $P_{Y^m|U=1}=P^\star_{Y^m}$, $P_{Y^m|U=0}=\frac{1}{1-t}(Q_{Y^m}-tP^\star_{Y^m})$, and letting $t\downarrow 0$. Here $P^\star_{Y^m}$ is chosen such that $\sum_{j=1}^m c_j D(P^\star_{Y_j}\|Q_{Y_j})-D(P^\star_{Y^m}\|Q_{Y^m})>0$ and $\frac{\mathrm{d}P^\star_{Y^m}}{\mathrm{d}Q_{Y^m}}$ is bounded.

We list a few basic tensorization properties and include the short proofs.

Proposition 2 (Tensorization)**.** *Given Q_{Y^m}, (ν_j) and c^m as in Definition 3, and any $n\geqslant 1$, we have*

1) $\mathrm{d}(Q_{Y^m}^{\otimes n},(\nu_j^{\otimes n}),c^m)=n\,\mathrm{d}(Q_{Y^m},(\nu_j),c^m)$.
2) $\mathrm{d}^\star(Q_{Y^m}^{\otimes n},(\nu_j^{\otimes n}),c^m)=n\,\mathrm{d}^\star(Q_{Y^m},(\nu_j),c^m)$.
3) *If P_{UY^m} achieves the supremum in the definition of $\mathrm{d}^\star(Q_{Y^m},(\nu_j),c^m)$, then $P_{UY^m}^{\otimes n}$ achieves the supremum in the definition of $\mathrm{d}^\star(Q_{Y^m}^{\otimes n},(\nu_j^{\otimes n}),c^m)$.*

Proof. The $\geqslant$ parts of 1) and 2) are immediate from the definitions. For the $\leqslant$ part, given any $P_{UY^{mn}}$, let $I\in\{1,\dots,n\}$ be equiprobable and independent of (U,Y^{mn}) under P. Then[4]

$$\begin{aligned}
&D(P_{Y_j{}^n|U}\|\nu_j^{\otimes n}|P_U)\\
&=\sum_{i=1}^n D(P_{Y_{ji}|UY_j{}^{i-1}}\|\nu_j|P_{UY_j{}^{i-1}}) &(29)\\
&\leqslant \sum_{i=1}^n D(P_{Y_{ji}|UY^{m,i-1}}\|\nu_j|P_{UY^{m,i-1}}) &(30)\\
&= nD(P_{Y_{jI}|IUY^{m,I-1}}\|\nu_j|P_{IUY^{m,I-1}}), &(31)
\end{aligned}$$

and

$$\begin{aligned}
&D(P_{Y^{mn}|U}\|Q_{Y^m}^{\otimes n}|P_U)\\
&=\sum_{i=1}^n D(P_{Y^m{}_i|UY^{m,i-1}}\|Q_{Y^m}|P_{UY^{m,i-1}}) &(32)\\
&= nD(P_{Y^m{}_I|IUY^{m,I-1}}\|Q_{Y^m}|P_{IUY^{m,I-1}}). &(33)
\end{aligned}$$

Identifying $(I,U,Y^{m,I-1})$ as U and $Y^m{}_I$ as Y^m, the claim of 1) (resp. 2)) follows noting that $\sup_{P_{UY^m}}\{\sum_{j=1}^m c_j D(P_{Y_j|U}\|\nu_j|P_U)-D(P_{Y^m|U}\|Q_{Y^m}|P_U)\}$ where the supremum is without (resp. with) the constraint $P_{Y^m}=Q_{Y^m}$ equals $\mathrm{d}(Q_{Y^m},(\nu_j),c^m)$ (resp. $\mathrm{d}^\star(Q_{Y^m},(\nu_j),c^m)$). Claim 3) follows from Claim 2). □

III. A Single-Shot Upper Bound on the Smooth BL Divergence

A principal goal of this paper is to upper-bound the smooth BL divergence in terms of $\mathrm{d}^\star$. Then by proving single-shot converses in terms of the smooth BL divergence for common randomness generation and the Gray-Wyner source coding, we obtain sharp second-order converses.

[4]Note that $Y_j{}^i:=(Y_{j1},\dots,Y_{ji})$ which is note to be confused with $Y_j^i:=(Y_j,Y_{j+1},\dots,Y_i)$.

In this section we prove an estimate of the smooth BL divergence mentioned in (7). This relies crucially on the properties of a convex functional ϕ, defined below in (34).

Proposition 3. *Given a probability measure Q_{Y^m}, nonnegative σ-finite measures $(\nu_j)_{j=1}^m$ on $\mathcal{Y}_1,\dots,\mathcal{Y}_m$, and $(c_j)_{j=1}^m\in(0,\infty)^m$, define the function of joint probability measures $P_{Y^m}\ll Q_{Y^m}$,*

$$\begin{aligned}
&\phi(P_{Y^m})\\
&:=\sup_{P_{U|Y^m}}\left\{\sum_{j=1}^m c_j D(P_{Y_j|U}\|\nu_j|P_U)-D(P_{Y^m|U}\|Q_{Y^m}|P_U)\right\},
\end{aligned} \tag{34}$$

where the supremum is understood as over standard probability space P_{UY^m} with the given marginal P_{Y^m}. Then

1) *ϕ is concave.*
2) *Suppose that $P^\star_{UY^m}$ achieves the supremum in (34), and $\sum_{j=1}^m c_j\imath_{P^\star_{UY_j}\|P^\star_U\times\nu_j}-\imath_{P^\star_{UY^m}\|P^\star_U\times Q_{Y^m}}$ is absolutely integrable with respect to $P^\star_{UY^m}$. Then*[5]

$$\begin{aligned}
&\nabla\phi|_{P_{Y^m}}(y^m)\\
&:=\sum_{j=1}^m c_j\imath_{P^\star_{UY_j}\|P^\star_U\times\nu_j}(u,y_j)-\imath_{P^\star_{UY^m}\|P^\star_U\times Q_{Y^m}}(u,y^m),
\end{aligned} \tag{35}$$

   where the right side is independent of u, $P^\star_{UY^m}$-a.s., defines a subgradient[6] of ϕ at P_{Y^m}.
3) *If $\mathcal{Y}^m$ is finite, then $\nabla_{P_{Y^m}}\mathrm{d}^\star(P_{Y^m},(\nu_j),c^m)|_{Q_{Y^m}}=\nabla\phi|_{Q_{Y^m}}$, where the left side denotes the conventional gradient over a finite dimensional space (assuming that it exists at Q_{Y^m}).*

Remark 5. The relation between $\mathrm{d}^\star(\cdot)$ and $\phi(\cdot)$ may be a bit confusing; let us clarify as follows: fix any Q_{Y^m}, (ν_j) and $c^m\in(0,\infty)^m$. In Remark 2 we commented that the supremums in the definitions of $\mathrm{d}^\star(\cdot)$ (and also $\phi(\cdot)$, for the same reason) can be restricted to finite. Thus upon defining the functional $\varphi\colon P_{Y^m}\mapsto\sum_{j=1}^m c_j D(P_{Y_j}\|\nu_j)-D(P_{Y^m}\|Q_{Y^m})$, we can write

$$\mathrm{d}(Q_{Y^m},(Q_{Y_j}),c^m)=\sup_{P_{Y^m}\ll Q_{Y^m}}\varphi(P_{Y^m}), \tag{36}$$

$$\mathrm{d}^\star(Q_{Y^m},(Q_{Y_j}),c^m)=(\mathrm{conc}\,\varphi)(Q_{Y^m}), \tag{37}$$

$$\phi(P_{Y^m})=(\mathrm{conc}\,\varphi)(P_{Y^m}),\quad \forall P_{Y^m}\ll Q_{Y^m}, \tag{38}$$

where conc denotes the concave envelope operator.

Proof of Proposition 3. 1) Consider arbitrary $P^{(i)}_{Y^m}$, $i=0,1$. Suppose that $P^{(i)}_{U|Y^m}$ achieves the supremum in (34) when $P_{Y^m}=P^{(i)}_{Y^m}$ (if the supremum is not achieved,

[5]Note that the right side of (35) may be written as $\sum_{j=1}^m c_j\imath_{P^\star_{Y_j|U=u}\|\nu_j}(y_j)-\imath_{P^\star_{Y^m|U=u}\|Q_{Y^m}}(y^m)$. Though on the first sight this only depends on $P^\star_{Y^m|U}$, it is actually dependent on P_{Y^m} since $P^\star_{Y^m|U}$ is computed from P_{Y^m}.

[6]We do not say the "sup-gradient" of a concave function since it is unconventional.

the claim still holds by an approximation argument). Then for any $\alpha \in (0,1)$, and $P_{Y^m}^{(\alpha)} = (1-\alpha)P_{Y^m}^{(0)} + \alpha P_{Y^m}^{(1)}$, let $\mathcal{U}^{(\alpha)}$ be the disjoint union of $\mathcal{U}^{(0)}$ and $\mathcal{U}^{(1)}$, and set $P_{UY^m}^{(\alpha)}$ as the convex combination of $P_{UY^m}^{(0)}$ and $P_{UY^m}^{(1)}$. This induces a $P_{U|Y^m}^{(\alpha)}$ for which $\sum_{j=1}^m c_j D(P_{Y_j|U}^{(\alpha)}\|\nu_j|P_U) - D(P_{Y^m|U}^{(\alpha)}\|Q_{Y^m}|P_U) = (1-\alpha)\phi(P_{Y^m}^{(0)}) + \alpha\phi(P_{Y^m}^{(1)})$. This shows that $\phi(P_{Y^m}^{(\alpha)}) \geqslant (1-\alpha)\phi(P_{Y^m}^{(0)}) + \alpha\phi(P_{Y^m}^{(1)})$.

2) Let $f(u,y^m)$ be the right side of (35). We first argue that the right side of (35) equals a function only of y^m, $P_{UY^m}^\star$-a.s. The intuition is easy to obtain in the case of finite $\mathcal{U}\times\mathcal{Y}^m$: Suppose that $(u,y^m)\neq(u',y^m)$ are both on the support of P_{UY^m} and $f(u,y^m) > f(u',y^m)$. We can define $P_{UY^m}^t := P_{UY^m}^\star + t\cdot\delta_{u,y^m} - t\cdot\delta_{u',y^m}$, where δ_{u,y^m} denotes a point mass at (u,y^m). Then as $t\to 0$ we have $\phi(P_{UY^m}^t) = \phi(P_{UY^m}) + t(f(u,y^m) - f(u',y^m)) + o(t)$, which contradicts the assumption that $P_{UY^m}^\star$ achieves the supremum. Next we give a measure theoretic proof for the general case; this is usually done using the hyperplane separation theorems. By assumption, f belongs to $L^1(P_{UY^m})$. Let $\mathcal{V}$ be the set of $L^1(P_{Y^m})$ functions, viewed a subspace of $L^1(P_{UY^m})$. Define its dual

$$\mathcal{V}^* := \left\{ g\in L^\infty(P_{UY^m}) \colon \int hg\,\mathrm{d}P_{UY^m} = 0,\ \forall h\in\mathcal{V}\right\}. \tag{39}$$

Now for any $g\in\mathcal{V}^*$, define $P_{UY^m}^t$ by

$$\frac{\mathrm{d}P_{UY^m}^t}{\mathrm{d}P_{UY^m}^\star} = 1 + tg \tag{40}$$

which is well-defined probability measure for $|t|$ small enough, and has marginal P_{Y^m} in view of the definition of $\mathcal{V}^*$. Using the dominated convergence theorem to bring the differentiation inside the integrals in computing the relative entropy terms in the definition of ϕ, we find

$$\phi(P_{UY^m}^t) = \phi(P_{UY^m}^\star) + t\int gf\,\mathrm{d}P_{UY^m}^\star + o(t) \tag{41}$$

as $t\to 0$. Then

$$\int fg\,\mathrm{d}P_{UY^m}^\star = 0 \tag{42}$$

since $P_{UY^m}^\star$ is a maximizer. It remains to show that $f\in\mathcal{V}$ (i.e., the "double dual" of $\mathcal{V}$ is itself). Suppose that $f\notin\mathcal{V}$. Since $\mathcal{V}$ is closed in $L^1(P_{UY^m})$ and the singleton $\{f\}$ is compact, by the Hahn-Banach theorem (see [54, P106]), there exists $g\in L^\infty(P_{UY^m})$ such that

$$\int fg\,\mathrm{d}P_{UY^m}^\star < \inf_{h\in\mathcal{V}}\int hg\,\mathrm{d}P_{UY^m}. \tag{43}$$

Since $\mathcal{V}$ is a linear subspace, we see that the right side of (43) can either be 0 or $-\infty$. The latter case is ruled out because of the strict inequality of (43), hence the right side of (43) is 0 and $h\in\mathcal{V}^*$. But then the left side of (43) must also be 0 as we have shown in (42), a contradiction. Hence we proved that the right side of (35) must lie in $L^1(P_{Y^m})$.

Next we show that $\nabla\phi|_{P_{Y^m}}$ as defined in (35) is a subgradient, that is, for any probability measure $S_{Y^m}\ll P_{Y^m}$,

$$\phi(S_{Y^m}) - \phi(P_{Y^m}) \leqslant \int \nabla\phi|_{P_{Y^m}}\,\mathrm{d}(S_{Y^m} - P_{Y^m}). \tag{44}$$

It suffices to prove (44) when $\frac{\mathrm{d}S_{Y^m}}{\mathrm{d}Q_{Y^m}}$ is bounded, as the general claim will then follow with an approximation argument. In that case,

$$S_{Y^m}^t := (1+t)P_{Y^m} - tS_{Y^m} \tag{45}$$

is a probability measure when $t\in(0,t_0)$ for some $t_0 > 0$. Then $P_{Y^m} = \frac{1}{1+t}S_{Y^m}^t + \frac{t}{1+t}S_{Y^m}$. By the concavity of ϕ we have $\phi(P_{Y^m}) \geqslant \frac{1}{1+t}\phi(S_{Y^m}^t) + \frac{t}{1+t}\phi(S_{Y^m})$, and upon rearrangement,

$$\phi(S_{Y^m}^t) - \phi(P_{Y^m}) \leqslant t\left(\phi(P_{Y^m}) - \phi(S_{Y^m})\right). \tag{46}$$

Hence we establish (44) by

$$\begin{aligned}
&\phi(P_{Y^m}) - \phi(S_{Y^m})\\
&\geqslant \lim_{t\downarrow 0}\frac{\phi(S_{Y^m}^t) - \phi(P_{Y^m})}{t}\\
&\geqslant \lim_{t\downarrow 0}\frac{\psi(S_{Y^m}^t, P_{U|Y^m}^\star) - \psi(P_{Y^m}, P_{U|Y^m}^\star)}{t} &(47)\\
&= \int \nabla\phi|_{P_{Y^m}}\,\mathrm{d}(P_{Y^m} - S_{Y^m}) &(48)
\end{aligned}$$

where

- In (47), we defined $\psi\colon (P_{Y^m}, P_{U|Y^m}) \mapsto \sum_{j=1}^m c_j D(P_{Y_j|U}\|\nu_j|P_U) - D(P_{Y^m|U}\|Q_{Y^m}|P_U)$, and defined $P_{U|Y^m}^\star$ as the regular conditional probability induced by the maximizer $P_{UY^m}^\star$. Note that equality does not necessarily hold in (47) because $P_{U|Y^m}^\star$ maximizes $\psi(P_{Y^m},\cdot)$ but not necessarily $\psi(S_{Y^m}^t,\cdot)$.
- (48) follows by using the dominated convergence theorem to bring the derivative into the integrals in the definition of the relative entropies.

3) From the definitions we have

$$\mathrm{d}^\star(Q_{Y^m},(\nu_j),c^m) = \phi(Q_{Y^m}). \tag{49}$$

On the other hand since $D(P_{Y^m|U}\|Q_{Y^m}|P_U) = D(P_{Y^m|U}\|P_{Y^m}|P_U) + D(P_{Y^m}\|Q_{Y^m})$, we also have

$$\begin{aligned}
\mathrm{d}^\star(P_{Y^m},(\nu_j),c^m) &= \phi(P_{Y^m}) + D(P_{Y^m}\|Q_{Y^m}) &(50)\\
&= \phi(P_{Y^m}) + o(|P_{Y^m} - Q_{Y^m}|) &(51)
\end{aligned}$$

when P_{Y^m} and Q_{Y^m} are close. From (49) and (51) we see that the gradients of $\mathrm{d}^\star(\cdot,(\nu_j),c^m)$ and $\phi(\cdot)$ are equal.

□

Remark 6. Note that by the definition (44), the subgradient $\nabla\phi|_{P_{Y^m}}$ can be thought of as a measurable function on $\mathcal{Y}^m$

modulo an additive constant. However, it is convenient to normalize it so that

$$\int \nabla\phi|_{Q_{Y^m}} \mathrm{d}Q_{Y^m} = \mathrm{d}^\star(Q_{Y^m}, (\nu_j), c^m) \tag{52}$$

which is consistent with (35) and will also be convenient later.

The main result of this section is the following upper bound on the smooth BL divergence.

Theorem 4. *Let Q_{Y^m} be a probability measure on $\mathcal{Y}^m$ and ν_j be a nonnegative σ-finite measure on $\mathcal{Y}_j$, $j = 1, \ldots, m$. Suppose that $P^\star_{U|Y^m}$ achieves the supremum in the definition of $\mathrm{d}^\star(Q_{Y^m}, (\nu_j), c^m)$, and define $\nabla\phi$ by* (35)*. Then for any $\delta \in (0,1)$ and $c^m \in (0,\infty)^m$, we have*

$$\delta \leqslant \mathbb{P}[\nabla\phi|_{Q_{Y^m}}(Y^m) > \mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m)]. \tag{53}$$

Alternatively, for any $\lambda \in \mathbb{R}$,

$$\mathrm{d}_{\mathbb{P}[\nabla\phi|_{Q_{Y^m}}(Y^m)>\lambda]} \leqslant \lambda. \tag{54}$$

Proof. Let $\gamma := \mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m) - \mathrm{d}^\star(Q_{Y^m}, (\nu_j), c^m)$. Define

$$\mathcal{C} := \{y^m \colon \nabla\phi|_{Q_{Y^m}}(y^m) \leqslant \mathrm{d}^\star(Q_{Y^m}, (\nu_j), c^m) + \gamma\}, \tag{55}$$

then our goal is to show that $Q_{Y^m}(\mathcal{C}) \geqslant 1-\delta$, or equivalently,

$$\mathrm{d}_{1-Q_{Y^m}(\mathcal{C})}(Q_{Y^m}, (\nu_j), c^m) \leqslant \mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m). \tag{56}$$

In the definition of $\mathrm{d}_{1-Q_{Y^m}(\mathcal{C})}(Q_{Y^m}, (\nu_j), c^m)$, take μ to be the restriction of Q_{Y^m} on $\mathcal{C}$, i.e., $\frac{\mathrm{d}\mu}{\mathrm{d}Q_{Y^m}}(y^m) = 1\{y^m \in \mathcal{C}\}$. Then obviously $E_1(Q_{Y^m}\|\mu) \leqslant 1 - Q_{Y^m}(\mathcal{C})$. Supposing that $P_{Y^m} \ll \mu$ achieves the supremum in the definition of $\mathrm{d}(\mu, (\nu_j), c^m)$ (if the supremum is not achievable we apply an approximation argument and the proof carries through), we have

$$\begin{aligned}
\mathrm{d}(\mu, (\nu_j), c^m) &= \sum_{j=1}^m c_j D(P_{Y_j}\|\nu_j) - D(P_{Y^m}\|Q_{Y^m}) && (57)\\
&\leqslant \phi(P_{Y^m}) && (58)\\
&\leqslant \phi(Q_{Y^m}) + \int \nabla\phi|_{Q_{Y^m}} \, \mathrm{d}(P_{Y^m} - Q_{Y^m}) && (59)\\
&\leqslant \phi(Q_{Y^m}) + \gamma && (60)\\
&= \mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m). && (61)
\end{aligned}$$

where

- (57) is because $D(P_{Y^m}\|\mu) = D(P_{Y^m}\|Q_{Y^m})$, by the definition of μ.
- (59) follows from the definition of the subgradient.
- (60) follows since $P_{Y^m} \ll \mu$ implies that P_{Y^m} is supported on $\mathcal{C}$, which in turn implies that

$$\begin{aligned}
\int \nabla\phi|_{Q_{Y^m}} \, \mathrm{d}P_{Y^m} &\leqslant \phi(Q_{Y^m}) + \gamma && (62)\\
&= \int \nabla\phi|_{Q_{Y^m}} \, \mathrm{d}Q_{Y^m} + \gamma. && (63)
\end{aligned}$$

Thus we have established (56). The proof of (54) is similar. □

Of particular interest is the case of the Gaussian or Lebesgue measures, where we are able to express $\nabla\phi|_{Q_{Y^m}}$ in more explicit forms. Define the following function of any for any positive semidefinite matrix $\mathbf{\Sigma}$:

$$\begin{aligned}
\psi(\mathbf{\Sigma}) &:= \sup_{R_{Y^m}=\mathcal{N}(\mathbf{0},\mathbf{\Sigma}_R),\, \mathbf{\Sigma}_R \preceq \mathbf{\Sigma}} \left\{ -\sum_{j=1}^m c_j h(Y_j) + h(Y^m) \right\} \\
&\quad + \mathrm{Tr}[\mathbf{M}_Q \mathbf{\Sigma}] && (64)\\
&= \sup_{\mathbf{\Sigma}_R \preceq \mathbf{\Sigma}} \left\{ -\sum_{j=1}^m \frac{c_j}{2} \log(2\pi e [\mathbf{\Sigma}_R]_{ii}) + \frac{m}{2} \log(2\pi e |\mathbf{\Sigma}_R|) \right\} \\
&\quad + \mathrm{Tr}[\mathbf{M}_Q \mathbf{\Sigma}] && (65)
\end{aligned}$$

where $\mathbf{M}_Q$ is the symmetric matrix such that $y^{m\top}\mathbf{M}_Q y^m = \sum_{j=1}^m c_j \log \frac{\mathrm{d}\lambda}{\mathrm{d}\nu_j}(y_j) - \log \frac{\mathrm{d}\lambda^m}{\mathrm{d}Q_{Y^m}}(y^m)$, $\forall y^m$, $Y^m \sim R_{Y^m}$, and $\mathbf{\Sigma}_R \preceq \mathbf{\Sigma}$ means that $\mathbf{\Sigma} - \mathbf{\Sigma}_R$ is a positive-semidefinite matrix. Note that the computation of $\psi(\mathbf{\Sigma})$ is simply a matrix optimization problem. The next proposition shows how the computation of ϕ and its subgradient is reduced to that of ψ.

Proposition 5. *Let $Q_{Y^m} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_Q)$ be a centered Gaussian distribution, and let $\nu_1, \ldots, \nu_m$ each be either a centered Gaussian or the Lebesgue measure. Let $c^m \in (0,\infty)^m$. Then*

1) *$\phi(P_{Y^m}) = \psi(\mathbf{\Sigma}_P)$, for any centered Gaussian distribution $P_{Y^m} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_P)$.*
2) *Let $\nabla\phi|_{Q_{Y^m}}$ be defined by* (35)*. Then*

$$\nabla\phi|_{Q_{Y^m}}(y^m) = y^{m\top}(\nabla\psi|_{\mathbf{\Sigma}_Q}) y^m, \quad y^m \in \mathbb{R}^m, \tag{66}$$

where $\nabla\psi|_{\mathbf{\Sigma}_Q}$ is a subgradient of the concave function ψ.

Proof. 1) From the definition we can deduce that

$$\begin{aligned}
&\phi(P_{Y^m}) \\
&:= \sup_{R_{UY^m}\colon R_{Y^m}=P_{Y^m}} \left\{ -\sum_{j=1}^m c_j h(Y_j|U) + h(Y^m|U) \right\} \\
&\quad + \mathrm{Tr}[\mathbf{M}_Q \mathbf{\Sigma}_P] && (67)\\
&\leqslant \sup_{R_{UY^m}\colon \mathbb{E}_R[Y^{m\top}Y^m] \preceq \mathbf{\Sigma}_P} \left\{ -\sum_{j=1}^m c_j h(Y_j|U) + h(Y^m|U) \right\} \\
&\quad + \mathrm{Tr}[\mathbf{M}_Q \mathbf{\Sigma}_P] && (68)\\
&= \psi(\mathbf{\Sigma}_P). && (69)
\end{aligned}$$

The last step used the fact that the Gaussian measure achieved the supremum in (68), which is shown in the exact form in [23, Theorem 14]; see also the references therein. To show the reverse direction $\phi(P_{Y^m}) \geqslant \psi(\mathbf{\Sigma}_P)$, for any $\mathbf{\Sigma}_R$ in the definition of $\psi(\mathbf{\Sigma}_P)$, construct the P_{UY^m} in the definition of $\phi(P_{Y^m})$ by letting $U \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_P - \mathbf{\Sigma}_R)$ and $(Y^m - U) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_R)$ be independent.

2) First we observe that ψ is indeed a concave function: for any $\mathbf{\Sigma}_P^0$, $\mathbf{\Sigma}_P^1$ and $p \in (0,1)$,

$$\begin{aligned}
&(1-p)\psi(\mathbf{\Sigma}_P^0) + p\psi(\mathbf{\Sigma}_P^1) \\
&\leqslant \sup_{R_{UY^m}\colon \mathbb{E}_R[Y^{m\top}Y^m] \preceq \mathbf{\Sigma}_P} \left\{ -\sum_{j=1}^m c_j h(Y_j|U) + h(Y^m|U) \right\} \\
&\quad + \mathrm{Tr}[\mathbf{M}_Q \mathbf{\Sigma}_P] && (70)\\
&= \psi(\mathbf{\Sigma}_P) && (71)
\end{aligned}$$

where $\mathbf{\Sigma}_P := (1-p)\mathbf{\Sigma}_P^0 + p\mathbf{\Sigma}_P^1$ and (70) is shown by choosing U to be binary. Next, consider any $P_{Y^m} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_P)$. From the proof of [23, Theorem 14] we know that the supremum in (67) can be achieved (importantly, this relies on the bounded second moment constraint in the supremum in (67)) by some constant U and centered Gaussian R_{Y^m}. Then by choosing $U \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_P - \mathbf{\Sigma}_R)$ and $(Y^m - U) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_R)$ to be independent, we see that there exists a centered Gaussian distribution $P^\star_{UY^m}$ achieving the supremum in the definition of $\phi(P_{Y^m})$. In particular, there exists some $\mathbf{A} \in \mathbb{R}^{m\times m}$ and $c_0 \in \mathbb{R}$ such that $\nabla\phi|_{Q_{Y^m}}(y^m) = y^{m\top}\mathbf{A}y^m + c_0$ for any $y^m \in \mathbb{R}^m$, and hence

$$\int \nabla\phi|_{Q_{Y^m}} \,\mathrm{d}(P_{Y^m} - Q_{Y^m}) = \mathrm{Tr}[\mathbf{A}(\mathbf{\Sigma}_P - \mathbf{\Sigma}_Q)]. \tag{72}$$

We remark that from the proof of [23, Theorem 14], $P^\star_{UY^m}$ is unique up to a transformation of U, so $\nabla\phi|_{Q_{Y^m}}$ defined in (35) is also unique. The subgradient property gives

$$\phi(P_{Y^m}) \leqslant \phi(Q_{Y^m}) + \int \nabla\phi|_{Q_{Y^m}} \,\mathrm{d}(P_{Y^m} - Q_{Y^m}). \tag{73}$$

These combined with the first part of the proposition show that

$$\psi(\mathbf{\Sigma}_P) \leqslant \psi(\mathbf{\Sigma}_Q) + \mathrm{Tr}[\mathbf{A}(\mathbf{\Sigma}_P - \mathbf{\Sigma}_Q)]. \tag{74}$$

By the definition of the subgradient of ψ, we see that $\mathbf{A}$ is a subgradient $\nabla\psi|_{\mathbf{\Sigma}_Q}$. The constant c_0 is not important if we view the subgradient as the equivalent class modulo an additive constant. Alternatively, under the normalization

$$\int \nabla\phi|_{Q_{Y^m}} \mathrm{d}Q = \phi(Q_{Y^m}) = \psi(\mathbf{\Sigma}_Q) = \mathrm{Tr}[\nabla\psi|_{\mathbf{\Sigma}_Q}\mathbf{\Sigma}_Q]$$

we can argue that $c_0 = 0$.

□

IV. Application: Omniscient-Helper Common Randomness Generation

In this section we prove a single-shot converse bound for the omniscient helper common randomness (CR) generation problem [40, Theorem 4.2] in terms of the smooth BL divergence. This allows us to prove not only the exact second-order converse for common randomness generation, but also asymptotic lower bounds on the smooth BL divergence.

A. A Single-Shot Converse for Common Randomness Generation

Figure 1 shows the setup of the common randomness generation problem, in the single-shot version. Let Q_{Y^m} be the joint distribution of sources Y_1, …, Y_m, observed by terminals T_1, …, T_m. Terminal T_0 which observes Y^m is called an omniscient helper. Terminal T_0 computes the integers $W_1(Y^m)$, …, $W_m(Y^m)$ and sends them to T_1, …, T_m, respectively. Then, terminals T_0, …, T_m compute integers $K(Y^m)$, $K_1(Y_1, W_1)$,…, $K_m(Y_m, W_m)$. The goal is to make $K = K_1 = \cdots = K_m$ with high probability and K almost equiprobable. In this paper we primarily focus on the case where the computation at the terminals can be stochastic (i.e., there exists infinite private randomness at each terminal); we will clarify when there is potential confusion.

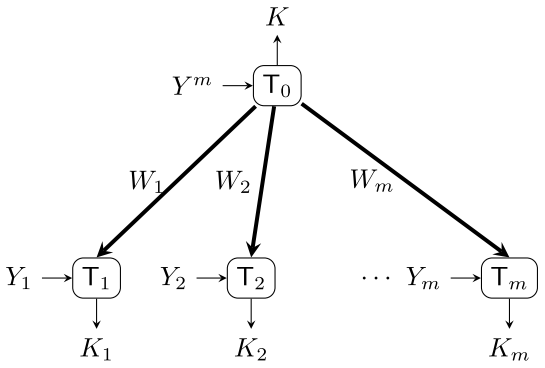

Fig. 1. Common randomness generation with an omniscient helper.

Let us recall previous results on this problem. In the stationary memoryless case where the sources have the per-letter distribution Q_{Y^m}, take $Y_j \leftarrow {Y_j}^n$ in the above single-shot formulation. Define

$$R = \liminf_{n\to\infty} \frac{1}{n} \log |\mathcal{K}|; \tag{75}$$

$$R_j = \limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{W}_j|, \quad j = 1, \ldots, m. \tag{76}$$

In [40], Ahlswede and Csiszár used the *entropy characterization* method [13] to obtain a single-letter expression of the achievable rate region for CR generation under the performance constraints

$$\liminf_{n\to\infty} \frac{1}{n} H(K) \geqslant R; \tag{77}$$

$$\lim_{n\to\infty} \mathbb{P}[K = K_1 = K_2 = \cdots = K_m] = 1. \tag{78}$$

However, it is known that the achievable region does not change when some other performance metrics are adopted [40]. Let us also remark that the corresponding *key* generation problem, which places the additional constraint that $W_j \perp K$ asymptotically for each j, in fact has the same achievable region as the common randomness generation problem without a secrecy constraint [1, Theorem 9].[7] However, the present paper is not concerned with the secrecy constraint.

Let us recall the single-letter region characterized by Ahlswede and Csiszár [40, Theorem 4.2]. If T_0 has no private randomness (i.e., $(W_j)_{j=1}^m$ cannot be computed stochastically), then the achievable region is the closure of

$$\mathcal{R}_0 := \bigcup_{P_{U|Y^m}} \left\{ \begin{array}{c} (R, R_1, \ldots, R_m) \in [0,\infty)^{m+1}: \\ R \leqslant I(U; Y^m), \\ R_j \geqslant I(U; Y^m) - I(U; Y_j), \\ j = 1, \ldots, m. \end{array} \right\}. \tag{79}$$

[7] In general, the secrecy constraint can strictly decrease the region when the transmitter does not see all other terminals [1, Theorem 2].

If T_0 has private randomness, the achievable region is obtained by replacing the inequalities in (79) with

$$R \leqslant I(U;VY^m), \tag{80}$$

$$R_j \geqslant I(U;VY^m) - I(U;Y_j), \tag{81}$$

and union over all V independent of Y^m and $P_{U|VY^m}$.[8] We now present an equivalent, more compact representation of the region in (80) and (81).

Proposition 6. *If T_0 has private randomness, $(R, R_1, \ldots, R_m) \in [0,\infty)^{m+1}$ is achievable if any only if*

$$\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) + \sum_{j=1}^m c_j R_j \geqslant \left(\sum_{j=1}^m c_j - 1\right) R \tag{82}$$

for all $c^m \in (0,\infty)^m$ (equivalently, for all $c^m \in (0,\infty)^m$ such that $\sum c_j > 1$, since (82) is trivially true otherwise, by the fact that $\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) \geqslant 0$).

Proof. We first show that when T_0 has private randomness, the achievable region is the closure of

$$\mathcal{R} := \bigcup_{P_{U|Y^m}} \left\{ \begin{array}{c} (R, R_1, \ldots, R_m): \\ R = I(U;Y^m) + r_0, \\ R_j = I(U;Y^m) - I(U;Y_j) + r_j, \\ r_0 \geqslant 0, \quad r_1, \ldots, r_m \geqslant r_0. \end{array} \right\}. \tag{83}$$

Of course, $r_1, \ldots, r_m$ can be interpreted as the additional communication rates used for sending the private randomness which can be added on top of the common randomness generated by a deterministic protocol, thus (83) is obviously an inner bounded of the region characterized (80) and (81). On the other hand, if R and R_j satisfies (80) and (81) then applying the chain rule of conditional mutual information we have

$$R \leqslant I(U;VY^m) \tag{84}$$

$$= I(U;Y^m) + I(U;V|Y^m) \tag{85}$$

$$R_j \geqslant I(U;VY^m) - I(U;Y_j) \tag{86}$$

$$= I(U;Y^m) - I(U;Y_j) + I(U;V|Y^m). \tag{87}$$

If $R \geqslant I(U;Y^m)$, set $r_0 := R - I(U;Y^m)$ and $r_j := R_j - [I(U;Y^m) - I(U;Y_j)]$, we see $0 \leqslant r_0 \leqslant I(U;V|Y^m) \leqslant r_j$ so that $(R, R_1, \ldots, R_m)$ is in (83). If $R < I(U;Y^m)$, let $p := \frac{R}{I(U;Y^m)} \in [0,1]$ and let $B \sim \mathrm{Bernoulli}(p)$ be independent of (U, Y^m). Let U' be the random variable which equals U when $B = 1$ and void when $B = 0$. Then

$$\begin{aligned} I(U'B;Y^m) &= I(U';Y^m|B) \\ &= pI(U';Y^m|B=1) \\ &= pI(U;Y^m|B=1) \\ &= pI(U;Y^m); \end{aligned} \tag{88}$$

similarly

$$I(U'B;Y_j) = pI(U;Y_j). \tag{89}$$

Now we treat (U', B) as the auxiliary U in (83), and put $r_0 := R - I(U'B;Y^m)$ and $r_j := R_j - [I(U'B;Y^m) - I(U'B;Y_j)]$. From (87), (88) and (89) we have $0 = r_0 \leqslant r_j$, $j = 1, \ldots, m$, so again $(R, R_1, \ldots, R_m)$ is in $\mathcal{R}$, and we have proved that the closure of $\mathcal{R}$ is the achievable region.

Next, by applying a linear transform we see that $(R, R_1, \ldots, R_m) \in \mathrm{cl}(\mathcal{R})$ if and only if $(R, R - R_1, \ldots, R - R_m) \in \mathrm{cl}(\mathcal{S})$ where

$$\mathcal{S} := \bigcup_{P_{U|Y^m}} \left\{ \begin{array}{c} (R, S_1, \ldots, S_m): \\ R = I(U;Y^m) + r_0, \\ S_j = I(U;Y_j) - s_j, \\ r_0, s_1, \ldots, s_m \geqslant 0. \end{array} \right\}. \tag{90}$$

Since $\mathrm{cl}(\mathcal{S})$ is a closed convex set, from convex analysis [55] we know that it is the intersection of closed half spaces one side of the supporting hyperplane. Moreover, since increasing the first coordinate or decreasing any of the other coordinates of a point in $\mathrm{cl}(\mathcal{S})$ will leave it in $\mathrm{cl}(\mathcal{S})$, the outward normal of at any boundary point of $\mathrm{cl}(\mathcal{S})$ is of the form $(c_0, -c_1, \ldots, -c_m)$ where $c_0, c_1, \ldots, c_m \in [0,\infty)^m$. Hence

$$\mathrm{cl}(\mathcal{S}) = \bigcap_{\substack{(c_0,c_1,\ldots,c_m) \\ \in [0,\infty)^{m+1}}} \left\{ \begin{array}{c} (R, S_1, \ldots, S_m) \in \mathbb{R}^{m+1}: \\ \sum_{j=1}^m c_j S_j - c_0 R \leqslant \\ \sup_{P_{U|Y^m}} \left\{ \sum_{j=1}^m c_j I(U;Y_j) - c_0 I(U;Y^m) \right\} \end{array} \right\}. \tag{91}$$

By a limiting argument it suffices to take $c_j > 0$, $j = 0, \ldots, m$ in (91), and then by homogeneity it suffices to take $c_0 = 1$. Substituting $S_j = R - R_j$ and the claim follows. □

Note that the entropy characterization approach of Ahlswede and Csiszár [40] is only sufficient for proving a weak converse (i.e. assuming a vanishing error probability in (78)). Our goal here is to prove sharp second-order converse results. Previously in our conference paper [1], a single-shot bound was derived via hypercontractivity which shows the strong converse property of common randomness per unit cost. More precisely, [1] showed that for any nonvanishing error probability the achievable rates must satisfy

$$\sum_{j=1}^m c_j R_j \geqslant \left(\sum_{j=1}^m c_j - 1\right) R \tag{92}$$

for any $c^m \in (0,\infty)^m$ such that $\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) = 0$. Note that (92) only characterizes the ratio of the CR to communication rates, rather than the entire region. Extending the proof in [1] directly will show an outer bound similar to (82) but with $\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m)$ replaced by $\mathrm{d}(Q_{Y^m}, (Q_{Y_j}), c^m)$, hence it is strictly suboptimal even in terms of the first-order region. A similar issue appeared in the single-shot converse for another common randomness generation problem between two terminals [2], [33], and in fact, more broadly in many other problems in network information theory. Here we complete the picture by bridging $\mathrm{d}^\star(\cdot)$ and $\mathrm{d}(\cdot)$ with the "smoothing" machinery.

Theorem 7 (Single-shot converse for omniscient helper CR generation)**.** *Fix Q_{Y^m}, $\delta \in [0,1)$, and $c^m \in (0,\infty)^m$ such that $\sum_{j=1}^m c_j > 1$. Let Q_{K^m} be the actual CR distribution in a coding scheme for omniscient helper CR generation, using*

[8] In the statement of [40, Theorem 4.2] the region is presented in a less transparent form, but from its proof it is clear that the region is given by (80), where V takes the role of the private randomness.

stochastic encoders and deterministic decoders (or stochastic decoders, if $c_j \leqslant 1$, $j = 1, \ldots, m$). Then

$$\frac{1}{2}|Q_{K^m} - T_{K^m}| \geqslant 1 - \frac{1}{|\mathcal{K}|} - \frac{\prod_{j=1}^m |\mathcal{W}_j|^{\frac{c_j}{\sum c_i}}}{|\mathcal{K}|^{1-\frac{1}{\sum c_i}}} \exp\left(\frac{\mathrm{d}_\delta(Q_{Y^m}, (Q_{Y_j}), c^m)}{\sum c_i}\right) - \delta, \tag{93}$$

where

$$T_{K^m}(k^m) := \frac{1}{|\mathcal{K}|} 1\{k_1 = \cdots = k_m\} \tag{94}$$

is the target CR distribution and $\mathcal{K}$ and $(\mathcal{W}_j)_{j=1}^m$ denote the CR and message alphabets.

Remark 7. The performance metric (93) takes into account both the uniformity and the agreement of the common randomness generated. We remark that the use of the total variation distance as a performance metric for common randomness or key generation has previously appeared in other places, such as [56], [57].

Proof. Suppose μ achieves the infimum in the definition of $\mathrm{d}_\delta(Q_{Y^m}, (Q_{Y_j}), c^m)$. For any $k \in \mathcal{K}$,

$$\mu\left(\bigcap_{j=1}^m \{K_j = k\}\right) = \int_{\mathcal{Y}^m} \sum_{w^m} \prod_{j=1}^m P_{K_j|Y_j W_j = w_j}(k) P_{W^m|Y^m}(w^m) \mathrm{d}\mu \tag{95}$$

$$\leqslant \int_{\mathcal{Y}^m} \max_{w^m} \prod_{j=1}^m P_{K_j|Y_j W_j = w_j}(k) \mathrm{d}\mu \tag{96}$$

$$= \int_{\mathcal{Y}^m} \prod_{j=1}^m \max_{w_j} P_{K_j|Y_j W_j = w_j}(k) \mathrm{d}\mu \tag{97}$$

$$\leqslant \exp(d) \prod_{j=1}^m \left[\int_{\mathcal{Y}_j} \max_{w_j} P_{K_j|Y_j W_j = w_j}^{\frac{1}{c_j}}(k) \mathrm{d}P_{Y_j}\right]^{c_j} \tag{98}$$

$$\leqslant \exp(d) \prod_{j=1}^m \left[\int_{\mathcal{Y}_j} \max_{w_j} P_{K_j|Y_j W_j = w_j}(k) \mathrm{d}P_{Y_j}\right]^{c_j} \tag{99}$$

$$\leqslant \exp(d) \prod_{j=1}^m \left[\sum_{w_j} \int_{\mathcal{Y}_j} P_{K_j|Y_j W_j = w_j}(k) \mathrm{d}P_{Y_j}\right]^{c_j} \tag{100}$$

where

- In (98) we defined $d := \mathrm{d}(\mu, (Q_{Y_j}), c^m)$.
- (99) uses $\max_{w_j} P_{K_j = k|Y_j W_j = w_j} \leqslant 1$ and the assumption of $0 < c_j \leqslant 1$ or deterministic decoders.

Raising both sides of (100) to the power of $\frac{1}{\sum_{i=1}^m c_i}$, we obtain

$$\mu^{\frac{1}{\sum_i c_i}}\left(\bigcap_{j=1}^m \{K_j = k\}\right) \leqslant \exp\left(\frac{d}{\sum_{i=1}^m c_i}\right) \prod_{j=1}^m \left[\sum_{w_j} \int_{\mathcal{Y}_j} P_{K_j = k|Y_j W_j = w_j} \mathrm{d}P_{Y_j}\right]^{\frac{c_j}{\sum_i c_i}} \tag{101}$$

But the function $t^m \mapsto \prod_{j=1}^m t_j^{\frac{c_j}{\sum_i c_i}}$ is a concave function on $[0, \infty)^m$; one way to see the concavity is that $\prod_{j=1}^m t_j^{\frac{c_j}{\sum_i c_i}} = \lim_{p \downarrow 0} \left(\sum_j \frac{c_j t_j^p}{\sum_i c_i}\right)^{\frac{1}{p}}$ which is the 0-norm of the random variable t_J where $\mathbb{P}[J = j] = \frac{c_j}{\sum_i c_i}$. Therefore by Jensen's inequality,

$$\frac{1}{|\mathcal{K}|} \sum_{k=1}^{|\mathcal{K}|} \prod_{j=1}^m \left[\sum_{w_j} \int_{\mathcal{Y}_j} P_{K_j|Y_j W_j = w_j}(k) \mathrm{d}P_{Y_j}\right]^{\frac{c_j}{\sum_i c_i}} \tag{102}$$

$$\leqslant \prod_{j=1}^m \left[\sum_{w_j} \int_{\mathcal{Y}_j} \frac{1}{|\mathcal{K}|} \sum_{k=1}^{|\mathcal{K}|} P_{K_j|Y_j W_j = w_j}(k) \mathrm{d}P_{Y_j}\right]^{\frac{c_j}{\sum_i c_i}} \tag{103}$$

$$= \prod_{j=1}^m \left[\sum_{w_j} \int_{\mathcal{Y}_j} \frac{1}{|\mathcal{K}|} \mathrm{d}P_{Y_j}\right]^{\frac{c_j}{\sum_i c_i}} \tag{104}$$

$$= \prod_{j=1}^m \left(\frac{|\mathcal{W}_j|}{|\mathcal{K}|}\right)^{\frac{c_j}{\sum_i c_i}} \tag{105}$$

Combining (101) and (105) we obtain

$$\frac{1}{|\mathcal{K}|} \sum_{k=1}^{|\mathcal{K}|} \mu^{\frac{1}{\sum_i c_i}}\left(\bigcap_{j=1}^m \{K_j = k\}\right) \leqslant \exp\left(\frac{d}{\sum_{i=1}^m c_i}\right) \prod_{j=1}^m \left(\frac{|\mathcal{W}_j|}{|\mathcal{K}|}\right)^{\frac{c_j}{\sum_i c_i}}. \tag{106}$$

Now, let $\bar{T}_{K^m}$ and $\bar{\mu}_{K^m}$ be the restrictions of T_{K^m} and μ_{K^m} on the event $\{K_1 = \cdots = K_m\}$. Then $\bar{T}_{K^m}$ is the equiprobable distribution on a set of cardinality $|\mathcal{K}|$, and $\bar{\mu}_{K^m}(k) = \mu\left(\bigcap_{j=1}^m \{K_j = k\}\right)$, $k = 1, \ldots, |\mathcal{K}|$. Invoking Lemma 8 which we show after the current proof with $\alpha = \frac{1}{\sum_i c_i}$, we obtain

$$E_1(T_{K^m} \| \mu_{K^m}) = E_1(\bar{T}_{K^m} \| \bar{\mu}_{K^m}) \tag{107}$$

$$\geqslant 1 - \frac{1}{|\mathcal{K}|} - |\mathcal{K}|^{\frac{1}{\sum_j c_j} - 1} \sum_{k=1}^{|\mathcal{K}|} \mu\left(\bigcap_{j=1}^m \{K_j = k\}\right)^{\frac{1}{\sum_j c_j}}. \tag{108}$$

Combining (106) and (108), we have

$$E_1(T_{K^m} \| \mu_{K^m}) \geqslant 1 - \frac{1}{|\mathcal{K}|} - \frac{\prod_{j=1}^m |\mathcal{W}_j|^{\frac{c_j}{\sum_{i=1}^m c_i}}}{|\mathcal{K}|^{1-\frac{1}{\sum_{i=1}^m c_i}}} \exp\left(\frac{\mathrm{d}(\mu_{Y^m}, (Q_{Y_j}), c^m)}{\sum_{i=1}^m c_i}\right). \tag{109}$$

Then, Theorem 7 follows since

$$\frac{1}{2}|Q_{K^m} - T_{K^m}| = E_1(T_{K^m} \| Q_{K^m}) \tag{110}$$

$$\geqslant E_1(T_{K^m} \| \mu_{K^m}) - E_1(Q_{K^m} \| \mu_{K^m}) \tag{111}$$

$$\geqslant E_1(T_{K^m} \| \mu_{K^m}) - E_1(Q_{Y^m} \| \mu_{Y^m}) \tag{112}$$

$$\geqslant E_1(T_{K^m} \| \mu_{K^m}) - \delta. \tag{113}$$

□

Lemma 8. *Suppose T is equiprobable on $\{1,\dots,M\}$ and μ is a nonnegative finite measure on the same alphabet. For any $\alpha \in (0,1)$,*

$$E_1(T\|\mu) \geqslant 1 - \frac{1}{M} - \exp(-(1-\alpha)D_\alpha(\mu\|T)), \tag{114}$$

where $E_1(T\|\mu)$ was defined in (16) *and the Rényi divergence is defined as $D_\alpha(\mu\|T) := \frac{1}{\alpha-1}\log\sum_{x=1}^M \mu^\alpha(x)T^{1-\alpha}(x)$.*

Proof. Consider the optimization problem over nonnegative vector $a^M = (a_1,\dots,a_M)$:

$$\text{minimize } f(a^M) := \sum_{m=1}^M \left|\frac{1}{M} - a_m\right|^+ \tag{115}$$

$$\text{subject to } g(a^M) := \frac{1}{M}\sum_{m=1}^M a_m^\alpha \leqslant \lambda \tag{116}$$

where $\lambda > 0$ is some constant. If μ has probability masses $a_1,\dots,a_M$, then $E_1(T\|\mu) = f(a^M)$ and $D_\alpha(\mu\|T)$ is a monotonically decreasing function of $g(a^M)$. We claim that (115)-(116) have an optimal solution $\hat{a}^M$ satisfying the property:

$$\left|\left\{m : \hat{a}_m \in \left(0,\frac{1}{M}\right)\right\}\right| \leqslant 1. \tag{117}$$

Indeed, if otherwise, and there are i and j for which $0 < \hat{a}_i \leqslant \hat{a}_j < \frac{1}{M}$, then we can choose $t := \min\{\hat{a}_i, \frac{1}{M} - \hat{a}_j\}$, such that either $\tilde{a}_i := \hat{a}_i - t = 0$ or $\tilde{a}_j := \hat{a}_j + t = \frac{1}{M}$. For $m \in \{1,\dots,M\}\backslash\{i,j\}$, put $\tilde{a}_m = a_m$. By convexity, we can check that $\tilde{a}^M$ is still an optimal solution to (115)-(116). However, the count on the left side of (117) decreases by 1 when we modify $\hat{a}$ to $\tilde{a}$. We can continue this process until (117) is satisfied. Next, notice that for any a^M, if $b_m := a_m 1\{a_m < \frac{1}{M}\} + \frac{1}{M}1\{a_m \geqslant \frac{1}{M}\}$, then

$$f(b^M) = f(a^M), \tag{118}$$

$$g(b^M) \leqslant (g(a^M)). \tag{119}$$

Thus an optimizer $\hat{a}^M$ of (115)-(116) can further be assumed to be of the following form:

$$\hat{a}^M = \left[\frac{1}{M},\frac{1}{M},\dots,\frac{1}{M},\eta,0,0,\dots,0\right] \tag{120}$$

for some $0 < \eta \leqslant \frac{1}{M}$. Suppose that the number of zeros on the right side of (120) is k. Then from $f(\hat{a}^M) \geqslant \frac{k}{M}$ and $\lambda \geqslant g(\hat{a}^M) \geqslant \frac{1}{M}\cdot\frac{M-k-1}{M^\alpha}$, we deduce that the optimal value for (115)-(116) satisfies

$$f \geqslant 1 - \lambda M^\alpha - \frac{1}{M}. \tag{121}$$

Thus we have shown that

$$E_1(T\|\mu) \geqslant 1 - \frac{1}{M} - M^{\alpha-1}\sum_{m=1}^M \mu^\alpha(m) \tag{122}$$

which is the desired inequality. □

Remark 8. Let Q_{KK^m} and

$$T_{KK^m}(k,k^m) := \frac{1}{|\mathcal{K}|}1\{k = k_1 = \cdots = k_m\} \tag{123}$$

denote the actual and the target distributions of the CR generated by T_0, T_1,...,T_m, respectively. Since

$$|Q_{KK^m} - T_{KK^m}| \geqslant |Q_{K^m} - T_{K^m}|, \tag{124}$$

Theorem 7 also provides a lower bound on $|Q_{KK^m} - T_{KK^m}|$. Actually, if the decoders are deterministic, T_0 can always produce K such that the two total variations are equal, because T_0 is aware of the CR produced by the other terminals.

Remark 9. Allowing stochastic decoders can strictly lower $\frac{1}{2}|Q_{K^m} - T_{K^m}|$: consider the special case where $m = 2$, Y_1 and Y_2 are constant, and there are no messages sent. Then the minimum $\frac{1}{2}|Q_{K^m} - T_{K^m}|$ achieved by deterministic decoders is $1 - \frac{1}{|\mathcal{K}|}$. On the other hand, T_1 and T_2 can each independently output an integer in $\{1,\dots,\sqrt{|\mathcal{K}|}\}$ equiprobably, achieving $\frac{1}{2}|Q_{K^m} - T_{K^m}| = 1 - \frac{1}{\sqrt{|\mathcal{K}|}}$. Nevertheless, we can argue that allowing stochastic decoders can at most reduce the error by a factor of 4: Suppose $\frac{1}{2}|Q_{K^m} - T_{K^m}| \leqslant \delta$ for some stochastic decoders, then $\frac{1}{2}|Q_{K_1} - T_{K_1}| \leqslant \delta$ and $Q(K_1 = K_2 =,\dots,= K_m) \leqslant \delta$. We can then remove the stochasticity of decoders at $\mathsf{T}_2\dots\mathsf{T}_m$ but retain the last two inequalities. Indeed, let $f_k(Y_m,W_m)$ denote the probability of producing k upon observing (Y_m,W_m) at T_m. Since $K_m - (Y_m,W_m) - (K_1,\dots,K_{m-1})$, we have $Q(K_1 = \cdots = K_m) = \mathbb{E}\left[\sum_{k=1}^{|\mathcal{K}|} f_k(Y_m,W_m)\mathbb{P}\left[K_1 = \cdots = K_{m-1} = k|Y_m,W_m\right]\right]$; this probability cannot decrease if T_m switches to the deterministic protocol that selects a k that maximizes $\mathbb{P}[K_1 = \cdots = K_{m-1} = k|Y_m,W_m]$. By iterating this argument for T_{m-1},...,T_2, we see that $\frac{1}{2}|Q_{K^m} - T_{K^m}| \leqslant 2\delta$ is achievable with deterministic decoders at T_2,...,T_m; this pays a factor of 2 for $\frac{1}{2}|Q_{K^m} - T_{K^m}|$. Applying the similar argument again, but starting with $\frac{1}{2}|Q_{K_m} - T_{K_m}| \leqslant 2\delta$ and $Q(K_1 = K_2 = \cdots = K_m) \leqslant 2\delta$, we can further remove the stochasticity of the decoder at T_1, at the cost of another factor of 2.

B. Second-Order Converse for Common Randomness Generation

Corollary 9. *Consider any stationary memoryless source with per-letter distribution Q_{Y^m}, any $c^m \in (0,\infty)^m$, and a sequence of omniscient helper CR generation schemes allowing stochastic encoders (indexed by n). Define*

$$A := \limsup_{n\to\infty}\frac{1}{\sqrt{n}}\left[\sum_{j=1}^m c_j\log|\mathcal{W}_j| + n\,\mathrm{d}^\star - (\sum_{j=1}^m c_j - 1)\log|\mathcal{K}|\right] \tag{125}$$

where $\mathrm{d}^\star := \mathrm{d}^\star(Q_{Y^m},(Q_{Y_j}),c^m)$. Also assume that $|\mathcal{K}| \to \infty$ and $P^\star_{UY^m}$ is a maximizer in the definition of $\mathrm{d}^\star(Q_{Y^m},(Q_{Y_j}),c^m)$, and define $\phi(\cdot)$ as in Section III. Then[9]

$$\liminf_{n\to\infty}\frac{1}{2}|Q_{K^m} - T_{K^m}| \geqslant \mathrm{Q}\left(\frac{A}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}})}}\right) \tag{126}$$

[9] $\mathrm{Q}(\cdot)$ denotes the standard Gaussian tail probability function.

where Q_{K^m} denotes the actual CR distribution and T_{K^m} is the target distribution as defined in (94)*.*

Proof. Using the bound on the smooth BL divergence (54), we obtain from (93) that

$$\frac{1}{2}|Q_{K^m} - T_{K^m}| \geqslant 1 - \frac{1}{|\mathcal{K}|} - \inf_{\gamma\in\mathbb{R}} \left\{ \frac{\prod_{j=1}^m |\mathcal{W}_j|^{\frac{c_j}{\sum c_i}}}{|\mathcal{K}|^{1-\frac{1}{\sum c_i}}} \exp\left(\frac{\gamma}{\sum c_i}\right) + \mathbb{P} \right\} \tag{127}$$

where $\mathbb{P} := \mathbb{P}\left[\sum_{i=1}^n \nabla\phi|_{Q_{Y^m}}(Y^m{}_i) > \gamma\right]$. Here we used Proposition 2-3) to show that

$$\nabla\phi|_{Q_{Y^{mn}}}(Y^{mn}) = \sum_{i=1}^n \nabla\phi|_{Q_{Y^m}}(Y^m{}_i). \tag{128}$$

Taking $\gamma = n\,\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) - \sqrt{n}A'$ for any $A' > A$ shows that

$$\liminf_{n\to\infty} \frac{1}{2}|Q_{K^m} - T_{K^m}| \geqslant \mathrm{Q}\left(\frac{A'}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}})}}\right). \tag{129}$$

Then take $A' \downarrow A$. □

C. Second-Order Achievability for Common Randomness Generation

In this section we show that the second-order converse (126) is tight for the discrete memoryless sources. The proof uses standard method of types analysis. First, consider the following encoding and decoding rules designed for a specific type $\hat{P}_{Y^{mn}}$.

Encoding at T_0: For any given $c^m \in (0,\infty)^m$, let $P^\star_{U|Y^m}$ be a maximizer in the definition of $\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m)$. Let $P^\star_U$ be the output distribution of $\hat{P}_{Y^{mn}}$ through $P^\star_{U|Y^m}$. Construct a codebook of size $|\tilde{\mathcal{K}}|$, where

$$\log|\tilde{\mathcal{K}}| := nI(\hat{P}_{Y^{mn}}, P^\star_{U|Y^m}) + n^{0.01} \tag{130}$$

and the codewords are i.i.d. according to the equiprobable distribution on the $P^\star_U$ type. Upon observing Y^{mn}, T_0 sends the empirical distribution $\hat{P}_{Y^{mn}}$ using $O(\log n)$ bits to all other terminals. Then T_0 equiprobably selects among codewords (if any) u_K such that (u_K, Y^{mn}) has the joint type $(P^\star_{U|Y^m}\hat{P}_{Y^{mn}})$. Random binning is used to send this selected index i to other terminals. For Terminal T_j, each u_i codeword is mapped randomly to one of $|\mathcal{W}_j|$ bins where

$$\log|\mathcal{W}_j| = nI(\hat{P}_{Y^{mn}}, P^\star_{U|Y^m}) - nI(\hat{P}_{Y_j{}^n}, P^\star_{U|Y_j}) + n^{0.02}. \tag{131}$$

Decoding at T_j: Note that $|\tilde{\mathcal{K}}|$ and $|\mathcal{W}_j|$ defined above depend on $\hat{P}_{Y^{mn}}$, which is known by all terminals as a part of the messages from T_0. Terminal T_j decodes a codeword having $P^\star_{U|Y_j}\hat{P}_{Y_j{}^n}$ joint type with $Y_j{}^n$ and also in the right bin.

Error and rate analysis: Conditioned on the type $\hat{P}_{Y^{mn}}$, the error probability of incorrectly decoding the codeword u_K is $O(n^{-100})$, by the standard covering and random binning analysis (see e.g. Slepian-Wolf coding [58]). Moreover for any fixed codebook, the probability that any given codeword index K is selected is $\exp(nI(\hat{P}_{Y^{mn}}, P^\star_{U|Y^m}) + O(\log n))$. Therefore using the hash lemma [40, Lemma 3.1], there exists a mapping $f\colon \tilde{\mathcal{K}} \to \mathcal{K}$ where

$$\log|\mathcal{K}| = nI(\hat{P}_{Y^{mn}}, P^\star_{U|Y^m}) - n^{0.01} \tag{132}$$

such that $f(K)$ is close to the equiprobable distribution on $\mathcal{K}$ with error $O(n^{-100})$ in the total variation. Moreover, let $\mathcal{E}_0$ be the event that $|\hat{P}_{Y^{mn}} - Q_{Y^m}| > n^{-0.49}$. Using large deviations,

$$\mathbb{P}[\mathcal{E}_0] = O(n^{-100}). \tag{133}$$

Under the complement of $\mathcal{E}_0$, by the Taylor expansion we have

$$\sum_{j=1}^m c_j(\log|\mathcal{W}_j| - \log|\mathcal{K}|) + \log|\mathcal{K}| = n\sum_{j=1}^m c_j I(\hat{P}_{Y_j}, P^\star_{U|Y_j{}^n}) - nI(\hat{P}_{Y^{mn}}, \hat{P}^\star_{U|Y^m}) + O(n^{0.02}) \tag{134}$$

$$= n\,\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) + n\left\langle \nabla\phi|_{Q_{Y^m}},\, \hat{P}_{Y^{mn}} - Q_{Y^m}\right\rangle + O(n^{0.02}) \tag{135}$$

$$= \sum_{i=1}^n \nabla\phi|_{Q_{Y^m}}(Y^m{}_i) + O(n^{0.02}). \tag{136}$$

We showed that (136) is the cost for the error to be $O(n^{-100})$. If we want the error to converge to $\mathrm{Q}\left(\frac{A}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}})}}\right)$, the left side of (134) needs to exceed $n\,\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) - \sqrt{n}A$.

A tweak: We have obtained the correct second-order upper bound on the left side of (134), but the proof is not finished yet since each term therein vary with $\hat{P}_{Y^{mn}}$ even though the sum is bounded. To finish, pick any $P_{U|Y^m} \neq P^\star_{U|Y^m}$ and set $P^t_{U|Y^m} := tP_{U|Y^m} + (1-t)P^\star_{U|Y^m}$. Under the complement of $\mathcal{E}_0$, for each $\hat{P}_{Y^{mn}}$ find $t = O(n^{-1/2})$ such that

$$\log|\mathcal{W}_j| = nI(\hat{P}_{Y^{mn}}, P^t_{U|Y^m}) - nI(\hat{P}_{Y_j{}^n}, P^t_{U|Y_j}) + n^{0.02} \tag{137}$$

always equals $nI(Q_{Y^m}, P^\star_{U|Y^m}) - nI(Q_{Y_j}, P^\star_{U|Y_j})$ which is independent of $\hat{P}_{Y^{mn}}$. The new codebook size

$$\log|\mathcal{K}| = nI(\hat{P}_{Y^{mn}}, P^t_{U|Y^m}) - n^{0.01} \tag{138}$$

will change according to this new encoding rule $P^t_{U|Y^m}$, but one can verify using the first order optimality of $P^\star_{U|Y^m}$ that

$$\sum_{j=1}^m c_j(\log|\mathcal{W}_j| - \log|\mathcal{K}|) + \log|\mathcal{K}| = \sum_{i=1}^n \nabla\phi|_{Q_{Y^m}}(Y^m{}_i) + o(n^{1/2}) \tag{139}$$

still holds. We thus obtain:

Theorem 10. *Consider any discrete memoryless source with per-letter distribution Q_{Y^m}, any $c^m \in (0,\infty)^m$, and $A \in \mathbb{R}$.*

There exists a sequence of omniscient helper CR generation schemes allowing stochastic encoders (indexed by n) such that

$$(\sum_{j=1}^{m} c_j - 1)\log|\mathcal{K}| - \sum_{j=1}^{m} c_j \log|\mathcal{W}_j| \leqslant n\,\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) - A\sqrt{n} \tag{140}$$

and

$$\limsup_{n\to\infty} \frac{1}{2}|Q_{K^m} - T_{K^m}| \leqslant \mathrm{Q}\left(\frac{A}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}})}}\right). \tag{141}$$

Here $\phi(\cdot)$ is as in Section III, Q_{K^m} denotes the actual CR distribution, and T_{K^m} is the target distribution as defined in (94)*.*

D. Second-Order Converse for Smooth BL Divergence

Theorem 7 essentially establishes a single-shot connection between the smooth BL divergence and omniscient helper CR generation: the achievability of one implies the converse of the other. The second-order achievability of common randomness generation implies the following second-order converse for smooth BL divergence, which is tight in view of the upper bound in Theorem 4.

Corollary 11. *Fix any discrete memoryless source Q_{Y^m}, $c^m \in (0,\infty)^m$, and $\delta \in (0,1)$.*

$$\mathrm{d}_\delta(Q_{Y^m}^{\otimes n}, (Q_{Y_j}^{\otimes n}), c^m) \geqslant n\,\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) + \sqrt{n\,\mathrm{Var}(\nabla\phi|_{Q_{Y^m}})}\mathrm{Q}^{-1}(\delta) - o(\sqrt{n}). \tag{142}$$

Proof. If the claim were not true, then by (93) one could prove a second-order converse for common randomness generation that contradicts the achievability (Theorem 10). □

Remark 10. For non-discrete alphabets, it is possible to prove the achievability of common randomness generation using the likelihood encoder [1]. The total variation error vanishes for rates in the interior of the same rate region. Thus by the same reasoning as Corollary 11, we have strong converse of smooth BL divergence for any stationary (not necessarily discrete) memoryless source:

$$\mathrm{d}_\delta(Q_{Y^m}^{\otimes n}, (Q_{Y_j}^{\otimes n}), c^m) \geqslant n\,\mathrm{d}^\star(Q_{Y^m}, (Q_{Y_j}), c^m) - o(n). \tag{143}$$

for any $\delta \in (0,1)$.

V. Application: Almost Lossless Gray-Wyner Network

In this section we prove a single-shot converse bound for the lossless Gray-Wyner source coding problem [59] using the smooth BL divergence. This will imply the exact second-order converse, previously also obtained by [16] using the method of types analysis.

Figure 2 shows the (single-shot) formulation of the problem. The sources are discrete random variables $Y_1, \ldots, Y_m$ with the joint distribution Q_{Y^m}. Terminal T_0 observes Y^m while

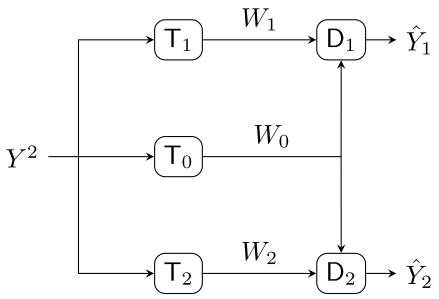

Fig. 2. Gray-Wyner network.

Terminal T_j observes Y_j, $j = 1,\ldots,m$. For each $j = 0,\ldots,m$, Terminal T_j computes integer $W_j(Y^m)$. For $j = 1,\ldots,m$, the decoder D_j receives (W_0, W_j) and computes $\hat{Y}_j(W_0, W_j) \in \mathcal{Y}_j$. The goal is that $\hat{Y}^m = Y^m$ with high probability. In the literature, the Gray-Wyner network usually refers to the $m = 2$ case of this model.

Gray and Wyner [59] computed the exact first order rate region in the discrete memoryless case. Take $Q_{Y^m} \leftarrow Q_{Y^m}^{\otimes n}$ and $Y_j \leftarrow {Y_j}^n$ in the above single-shot formulation. Define

$$R_j = \limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{W}_j| \tag{144}$$

for $j = 0,\ldots,m$. The achievable rate region, defined as the set of $(R_0,\ldots,R_m)$ for which there exist a sequence of coding schemes (indexed by n) such that

$$\limsup_{n\to\infty} \mathbb{P}[\hat{Y}^m \neq Y^m] = 0, \tag{145}$$

is the closure of the set of $(R_0,\ldots,R_m) \in [0,\infty)^{m+1}$ such that

$$R_0 + \sum_{j=1}^{m} c_j R_j \geqslant \inf_{P_{U|Y^m}} \left\{ \sum_{j=1}^{m} c_j H(Y_j|U) + I(U;Y^m) \right\} \tag{146}$$

$$=: -\mathrm{d}^\star(Q_{Y^m}, (\nu_j), c^m) \tag{147}$$

for all $c^m \in (0,\infty)^m$, where $(U, Y^m) \sim P_{U|Y^m}Q_{Y^m}$ and ν_j is the counting measure on $\mathcal{Y}_j$.

Theorem 12. *Fix Q_{Y^m}, $\delta \in (0,1)$, $c^m \in (0,\infty)^m$ and let ν_j be the counting measure on $\mathcal{Y}_j$ where $|\mathcal{Y}_j| < \infty$, $j = 1,\ldots,m$. Then any coding scheme for Gray-Wyner satisfies*

$$\mathbb{P}[\hat{Y}^m \neq Y^m] \geqslant 1 - \exp(\mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m)) \prod_{j=0}^{m} |\mathcal{W}_j|^{c_j} - \delta \tag{148}$$

where $c_0 := 1$.

Remark 11. Notice that the $\mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m)$ in (148) can be negative since ν_j is the counting measure rather than a probability measure.

Proof. Note that $\hat{Y}^m$ can be viewed as a function of y^m, since all the messages W_0,…,W_m are functions of y^m. Define the *correctly decodable set*

$$\mathcal{D} := \{y^m \colon \hat{Y}^m(y^m) = y^m\}. \tag{149}$$

Let μ be a minimizer in (17) (if the minimum is not achieved, the proof can still proceed by approximation). Define $\mu|_{\mathcal{D}}$ as the restriction of μ on $\mathcal{D}$, that is,

$$\frac{\mathrm{d}\,\mu|_{\mathcal{D}}}{\mathrm{d}\,\mu}(y^m) = 1\{y^m \in \mathcal{D}\}. \tag{150}$$

Let ϵ be the error probability. By the triangle inequality for E_γ-distance,

$$E_1(Q_{Y^m} \| \mu|_{\mathcal{D}}) \leqslant E_1(Q_{Y^m} \| Q_{Y^m}|_{\mathcal{D}}) + E_1(Q_{Y^m}|_{\mathcal{D}} \| \mu|_{\mathcal{D}}) \tag{151}$$

$$\leqslant \epsilon + E_1(Q_{Y^m} \| \mu) \tag{152}$$

$$\leqslant \epsilon + \delta. \tag{153}$$

For each $u \in \mathcal{W}_0$, set $\mathcal{A}_u := W_0^{-1}(u) \cap \mathcal{D}$, and denote by $\mathcal{A}_{uj}$ its projection onto the j-th coordinate. Then by Proposition 1, we have

$$\mu(\mathcal{A}_u) \leqslant \exp(\mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m)) \prod_{j=1}^m \nu_j(\mathcal{A}_{uj})^{c_j} \tag{154}$$

$$\leqslant \exp(\mathrm{d}_\delta(Q_{Y^m}, (\nu_j), c^m)) \prod_{j=1}^m |\mathcal{W}_j|^{c_j} \tag{155}$$

for each u, where (155) used the estimate $\nu_j(\mathcal{A}_{uj}) \leqslant |\mathcal{W}_j|$ which follows from the fact that $\mathcal{A}_{uj} \subseteq \hat{Y}_j(u, \mathcal{W}_j)$ by the definition of correctly decodable sets. The desired result then follows noting that

$$|\mathcal{W}_0| \max_u \mu(\mathcal{A}_u) \geqslant \sum_u \mu(\mathcal{A}_u) \tag{156}$$

$$= \mu(\mathcal{D}) \tag{157}$$

$$\geqslant 1 - E_1(Q_{Y^m} \| \mu|_{\mathcal{D}}) \tag{158}$$

$$\geqslant 1 - (\epsilon + \delta). \tag{159}$$

□

The exact second-order asymptotics for the Gray-Wyner coding [59] was derived in [16] using the method of types. In [16], a weighted distribution on $\mathcal{Y}^m$ is defined where the probability of the set of correctly decodable y^m is amplified, so that a nonvanishing error is turned into a polynomially vanishing error. Then Fano's inequality is applied to the weighted distribution. Here we show that such a second-order converse also follows from the smooth BL divergence bound (Theorem 12), which was proved via a fundamentally different approach.

Proposition 13 ([16])**.** *Consider a stationary memoryless source with per-letter distribution Q_{Y^m}, where $|\mathcal{Y}_1|, \ldots, |\mathcal{Y}_m| < \infty$. Let $c_1, \ldots, c_m \in (0, \infty)$, $c_0 := 1$ and $A \in \mathbb{R}$. For an arbitrary sequence of Gray-Wyner coding schemes (indexed by n), define*

$$A := \limsup_{n\to\infty} \sqrt{n} \left\{ \mathrm{d}^\star(Q_{Y^m}, (\nu_j), c^m) + \frac{1}{n} \sum_{j=0}^m c_j \log |\mathcal{W}_j| \right\} \tag{160}$$

where ν_j denotes the counting measure on $\mathcal{Y}_j$. Then

$$\liminf_{n\to\infty} \mathbb{P}[\hat{Y}^{mn} \neq Y^{mn}] \geqslant \mathrm{Q}\left(\frac{A}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}}(Y^m))}} \right) \tag{161}$$

where $\phi(\cdot)$ is as in Section III.

Proof. Using the bound on the smooth BL divergence (Theorem 4), we obtain from (148) that

$$\mathbb{P}[\hat{Y}^{mn} \neq Y^{mn}] \geqslant$$
$$1 - \inf_{\gamma\in\mathbb{R}} \left\{ \exp(\gamma) \prod_{j=0}^m |\mathcal{W}_j|^{c_j} + \mathbb{P}\left[\sum_{i=1}^n \nabla\phi|_{Q_{Y^m}}(Y^m{}_i) > \gamma \right] \right\} \tag{162}$$

where $Y^m \sim Q_{Y^m}$. Here we used Proposition 2-3) to show that

$$\nabla\phi|_{Q_{Y^{mn}}}(Y^{mn}) = \sum_{i=1}^n \nabla\phi|_{Q_{Y^m}}(Y^m{}_i). \tag{163}$$

Taking $\gamma = n\,\mathrm{d}^\star(Q_{Y^m}, (\nu_j), c^m) - \sqrt{n}A'$ for any $A' > A$ shows that

$$\liminf_{n\to\infty} \mathbb{P}[\hat{Y}^{mn} \neq Y^{mn}] \geqslant \mathrm{Q}\left(\frac{A'}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}}(Y^m))}} \right). \tag{164}$$

Taking $A' \downarrow A$ establishes the claim. □

Remark 12. For the discrete memoryless case, [16] showed that the bound (161) is tight.

VI. Application: Gaussian Lossy Gray-Wyner Network With Square Distortion

We now consider the lossy version of the Gray-Wyner problem. The setup is still as in Figure 2. But in contrast to the lossless version, we are given distortion functions $\Delta_j \colon \mathcal{Y} \times \hat{\mathcal{Y}} \to \mathbb{R}$ and the goal is to minimize $\mathbb{P}[\exists j \colon \Delta_j(Y_j, \hat{Y}_j) > D_j]$ where D_j is a given distortion level, $j = 1, \ldots, m$.

In the stationary memoryless case with per-letter source distribution Q_{Y^m}, take $Y_j \leftarrow Y_j{}^n$ and $\Delta_j(y_j, \hat{y}_j) \leftarrow \frac{1}{n}\sum_{i=1}^n \Delta_j(y_{ji}, \hat{y}_{ji})$. Denote by

$$R_j := \limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{W}_j| \tag{165}$$

the rate of the j-th message as before, for $j = 0, \ldots, m$. The achievable rate region is defined as the closure of the set of $(R_0, \ldots, R_m) \in [0, \infty)^{m+1}$ such that

$$\lim_{n\to\infty} \mathbb{P}[\exists j \colon \Delta_j(Y_j{}^n, \hat{Y}_j^n) > D_j] = 0. \tag{166}$$

In [59], Gray and Wyner showed that the achievable region is the closure of the set of $(R_0, \ldots, R_m)$ such that

$$R_0 \geqslant I(U; Y^m); \tag{167}$$

$$R_j \geqslant \mathrm{R}_{Y_j|U}(D_j), \quad j = 1, \ldots, m, \tag{168}$$

for an auxiliary random variable U [59]. Here $\mathrm{R}_{Y_j|U}(\cdot)$ denotes the conditional rate-distortion function. That is,

$$\mathrm{R}_{Y_j|U}(D) := \inf_{d(\cdot)} \int \mathrm{R}_{Y_j|U=u}(d(u))\,\mathrm{d}\,Q_U(u) \tag{169}$$

where the infimum is over nonnegative measurable functions $d\colon \mathcal{U} \to \mathbb{R}$ such that $\mathbb{E}[d(U)] \leqslant D$, and $\mathrm{R}_{Y_j|U=u}$ denotes the conventional (single-letter) rate-distortion function for a single source with per-letter distribution $Q_{Y_j|U=u}$ and per-letter distortion function Δ_j (see e.g. [58]). The second-order converse for discrete lossy Gray-Wyner coding was proved in [17], which relies heavily on the method of types and does not appear to be applicable to continuous sources such as the Gaussian sources.

We proceed to derive a single-shot converse for general distortion measures, and then we particularize it to quadratic distortions to derive a second-order converse for Gaussian sources.

Theorem 14. *Suppose $\delta \in (0,1)$, $Y^m \sim Q_{Y^m}$, $\Delta_j\colon \mathcal{Y}_j \times \hat{\mathcal{Y}}_j \to [0,\infty)$, and ν_j is an arbitrary nonnegative σ-finite measure on $\hat{\mathcal{Y}}_j$, $j = 1,\ldots,m$. Then any coding scheme for the Gray-Wyner network satisfies*

$$\begin{aligned}&\mathbb{P}[\exists j\colon \Delta_j(Y_j,\hat{Y}_j) > D_j]\\ \geqslant& 1 - \exp(\mathrm{d}_\delta(Q_{Y^m},(\nu_j),c^m))|\mathcal{W}_0|\prod_{j=1}^m |\mathcal{W}_j|^{c_j} L_j^{c_j} - \delta\end{aligned} \tag{170}$$

where $L_j := \sup_{y\in\hat{\mathcal{Y}}_j} \nu_j(\Delta_j(\cdot,y) \leqslant D_j)$, $j = 1,\ldots,m$.

Proof. The proof is similar to the proof of Theorem 12. Define the correctly decodable set

$$\mathcal{D} := \{y^m\colon \Delta(y_j,\hat{Y}_j(W_0(y^m),W_j(y^m))) \leqslant D_j, \forall j\}. \tag{171}$$

Steps (151)-(153) still follow. The bound (155) will be replaced by

$$\mu(\mathcal{A}_u) \leqslant \exp(d)\prod_{j=1}^m |\mathcal{W}_j|^{c_j} L_j^{c_j} \tag{172}$$

because for each j, $\mathcal{A}_{uj} \subseteq \{y_j\colon \exists w_j,\, \Delta(y_j,\hat{Y}_j(u,w_j)) \leqslant D_j\}$ (that is, $\hat{Y}_j(u,\mathcal{W}_j)$ is a D_j-covering of $\mathcal{A}_{uj}$), which implies that $\nu_j(\mathcal{A}_{uj}) \leqslant |\mathcal{W}_j| L_j$. The rest of the proof follows verbatim. □

Remark 13. Clearly Theorem 14 recovers Theorem 12 as a special case when $\Delta(\cdot,\cdot)$ is the Hamming distortion. We presented Theorem 12 first since it is simpler and contains the main ingredients of Theorem 14.

Next, we particularize Theorem 14 to the case of stationary memoryless Gaussian source and square distortion, and prove a second-order converse. Let us first simplify the first-order region. Let $\Delta_j\colon (y,\hat{y}) \in \mathbb{R}^2 \mapsto (y-\hat{y})^2$ be the per-letter distortion function, $j = 1,\ldots,m$. By (167) and (168), $(R_0,\ldots,R_m) \in [0,\infty)^{m+1}$ is achievable if and only if for any $c^m \in (0,\infty)^m$ and $c_0 := 1$,

$$\sum_{j=0}^m c_j R_j \geqslant \inf_{P_{UV^m|Y^m}} \left\{ I(U;Y^m) + \sum_{j=1}^m c_j I(V_j;Y_j|U) \right\}, \tag{173}$$

where the infimum is over $P_{UV^m|Y^m}$ such that V_j is a real valued random variable satisfying $\mathbb{E}[(V_j - Y_j)^2] \leqslant D_j$, $j = 1,\ldots,m$, and $(U,V^m,Y^m) \sim P_{UV^m|Y^m}Q_{Y^m}$. The following provides a supporting hyperplane characterization of this achievable region.

Proposition 15. *Let Q_{Y^m} be an m-dimensional Gaussian distribution with a non-degenerate covariance matrix. Fix $D^m \in (0,+\infty)^m$. Define*

$$\mathcal{R} := \bigcup_{P_{U|Y^m}} \left\{ \begin{array}{c} (R,R_1,\ldots,R_m) \in \mathbb{R}^{m+1}\colon \\ R \geqslant I(U;Y^m), \\ R_j \geqslant h(Y_j|U) - \frac{1}{2}\log 2\pi e D_j, \\ j = 1,\ldots,m. \end{array} \right\}. \tag{174}$$

1) *$\mathcal{R}$ is convex. The (inward pointing) normal at every boundary point R_0^m can be chosen as $(1,c_1,\ldots,c_m)$ with $c^m \in [0,1]^m$. If such $c^m \in (0,1)^m$, then R_0^m is the unique intersection of the supporting hyperplane and $\mathcal{R}$, and there exists a Gaussian $P_{U|Y^m}$ such that*

$$R_0 = I(U;Y); \tag{175}$$

$$R_j = h(Y_j|U) - \frac{1}{2}\log 2\pi e D_j, \quad j = 1,\ldots,m. \tag{176}$$

2)

$$\mathrm{cl}(\mathcal{R}) = \bigcap_{c^m\in(0,1)^m} \left\{ \begin{array}{c} (R,R_1,\ldots,R_m) \in \mathbb{R}^{m+1}\colon \\ R + \sum_{j=1}^m c_j R_j \geqslant \mathrm{d}^\star \end{array} \right\} \tag{177}$$

where

$$\mathrm{d}^\star := \inf_{P_{U|Y^m}} \left\{ I(U;Y^m) + \sum_{j=1}^m c_j \left[h(Y_j|U) - \frac{1}{2}\log 2\pi e D_j \right] \right\}.$$

3) *$\mathrm{cl}(\mathcal{R}) \cap [0,+\infty)^{m+1}$ is the achievable rate region.*

Proof. 1) The convexity is standard using the chain rules of the information quantities (similar to the proof of the convexity of a rate region). To see that each supporting hyperplane has a normal vector (pointing towards $\mathcal{R}$) of the form $(1,c_1,\ldots,c_m)$, $c^m \in [0,1]^m$: first choose c_0^m orthogonal to a supporting hyperplane of $\mathcal{R}$ at a boundary point R_0^m and pointing into $\mathcal{R}$. From the form of (174) we can see that $c_j \geqslant 0$, $j = 0,\ldots,m$. We also see from (174) that for any (no matter how small) $\hat{R}_1,\ldots,\hat{R}_m$ there exists $\hat{R}$ large enough such that $(\hat{R},\hat{R}_1,\ldots,\hat{R}_m) \in \mathcal{R}$, which implies that $c_0 \neq 0$. Thus by re-normalization we can assume without loss of generality that $c_0 = 1$. Then $c_j \leqslant 1$, $j = 1,\ldots,m$, by Proposition 16 which is given after the present proof. The claim for the case of $c^m \in (0,1)^m$ also follows from Proposition 16.

2) From convex analysis [55] the closed convex set $\mathrm{cl}(\mathcal{R})$ is the intersection of closed half spaces on one side

of the supporting hyperplanes. As argued in the proof of Part 1), the normal (pointing inward) vector of each supporting hyperplane can be chosen as $(1, c_1, \ldots, c_m)$ where $c^m \in [0,1]^m$. Moreover, since $(0,1)^m$ is dense in $[0,1]^m$, we can verify the geometric fact that such an intersection can be restricted to supporting hyperplanes whose normal vector has the form $(1, c_1, \ldots, c_m)$ where $c^m \in (0,1)^m$.

3) To see the achievable region contains $\mathrm{cl}(\mathcal{R}) \cap [0,+\infty)^{m+1}$, it suffices to show the achievability of an arbitrary $(\max\{R_j, 0\})_{j=0}^m$ where R_0^m is on the boundary of $\mathcal{R} \subseteq \mathbb{R}^{m+1}$. Choose a Gaussian $P_{U|Y^m}$ according to Part 1). Now for each $j = 1, \ldots, m$, if $\sigma^2_{Y_j|U} > D_j$, then for each u, we can construct $P_{N_jV_j|U=u}$ under which N_j and V_j are independent Gaussian with means 0 and $\mathbb{E}[Y_j|U=u]$ and variances D_j and $\sigma^2_{Y_j|U} - D_j \geqslant 0$ respectively, such that their sum has the distribution of $P_{Y_j|\mathsf{U}=u}$. Then we may as well put V_j, N_j, Y_j and U in the same probability space so that $Y_j = N_j + V_j$. Otherwise, $\sigma^2_{Y_j|U} \leqslant D_j$, we let V_j be constant. As such, in both cases we have

$$\max\{R_j, 0\} = h(Y_j|U) - h(Y_j|UV_j) \geqslant 0, \tag{178}$$

$$\mathbb{E}[|Y_j - V_j|^2] = D_j, \tag{179}$$

for $j = 1, \ldots, m$. Then $(\max\{R_j, 0\})_{j=0}^m$ is achievable in view of (167) and (168).

For the converse, consider any $R_0^m \in [0,\infty)^{m+1}$ that satisfies

$$R_0 \geqslant I(U; Y^m); \tag{180}$$

$$R_j \geqslant I(V_j; Y_j|U), \quad j = 1, \ldots, m; \tag{181}$$

for some $P_{U|Y^m}$ and $(P_{V_j|UY^m})_{j=1}^m$ such that $\mathbb{E}[|V_j - Y_j|^2] \leqslant D_j$, $j = 1, \ldots, m$. In view of (167) and (168) and the equivalent formulation of $\mathcal{R}$ in (177), it suffices to show that

$$R + \sum_{j=1}^m c_j R_j \geqslant \inf_{P_{U|Y^m}} \left\{ I(U; Y^m) + \sum_{j=1}^m c_j \left[h(Y_j|U) - \frac{1}{2} \log 2\pi e D_j \right] \right\} \tag{182}$$

for any $c^m \in (0,1)^m$. This follows because

$$I(V_j; Y_j|U) = h(Y_j|U) - h(Y_j|V_jU) \tag{183}$$

$$= h(Y_j|U) - h(Y_j - V_j|V_jU) \tag{184}$$

$$\geqslant h(Y_j|U) - h(Y_j - V_j) \tag{185}$$

$$\geqslant h(Y_j|U) - \frac{1}{2} \log 2\pi e D_j. \tag{186}$$

□

Proposition 16. *Fix Q_{Y^m} m-dimensional Gaussian with a non-degenerate covariance matrix $\mathbf{\Sigma}$. If $c^m \in [0,+\infty)^m$ and $c_j > 1$ for some j, then*

$$\inf_{P_{U|Y^m}} \left\{ -h(Y^m|U) + \sum_{j=1}^m c_j h(Y_j|U) \right\} = -\infty. \tag{187}$$

If $c^m \in [0,1)^m$, then the infimum is finite and achieved, and there exist a $\tilde{\mathbf{\Sigma}}$: $\mathbf{0} \preceq \tilde{\mathbf{\Sigma}} \preceq \mathbf{\Sigma}$ such that for any minimizer $P_{U|Y^m}$, $Y^m|U = u$ is Gaussian with covariance matrix $\tilde{\mathbf{\Sigma}}$ (for almost all u and under $P_{U|Y^m}Q_{Y^m}$).

Proof. If $P_{U|Y^m}$ is Gaussian and the covariance matrix of Y^m given U is $\mathbf{\Sigma}$ under $P_{U|Y^m}Q_{Y^m}$, then

$$- h(Y^m|U) + \sum_{j=1}^m c_j h(Y_j|U) = \frac{\sum c_j - m}{2} \log 2\pi e - \frac{1}{2} \log |\tilde{\mathbf{\Sigma}}| + \sum \frac{c_j}{2} \log \tilde{\Sigma}_{jj} \tag{188}$$

where $\tilde{\Sigma}_{jj}$ is the j-th diagonal entry of the matrix $\tilde{\mathbf{\Sigma}}$. The first claim for the case of $c_j > 1$ follows by taking $\tilde{\mathbf{\Sigma}}$ to be diagonal with $\tilde{\Sigma}_{jj} \downarrow 0$.

Next, suppose that $c^m \in [0,1)^m$. In [23] it is shown that the value of the left side in (187) does not change if the infimum is restricted to Gaussian $P_{U|Y^m}$. Choose a sequence of Gaussian $P^i_{U|Y^m}$, $i = 1, \ldots$, for which $-h(Y^m|U) + \sum_{j=1}^m c_j h(Y_j|U)$ converges to the left side of (187). Let $\mathbf{\Sigma}^i$ be the covariance matrix of Y^m given U under $P^i_{U|Y^m}Q_{Y^m}$. Since $\mathbf{\Sigma}^i \preceq \mathbf{\Sigma}$ for each i, by passing to a convergent subsequence we can assume that $\mathbf{\Sigma}^i \to \mathbf{\Sigma}^\star$ for some $\mathbf{\Sigma}^\star \preceq \mathbf{\Sigma}$. Observe that (188) is bounded below by

$$- h(Y^m|U) + \sum_{j=1}^m c_j h(Y_j|U) \geqslant \frac{\sum c_j - m}{2} \log 2\pi e - \sum \frac{1 - c_j}{2} \log \tilde{\Sigma}_{jj}, \tag{189}$$

hence $\tilde{\Sigma}^i_{jj}$ must be bounded away from 0, for large enough i. Thus $\mathbf{\Sigma}^\star$ has strictly positive diagonals. By the continuity of the right side of (188) in $\tilde{\mathbf{\Sigma}}$, we see that $\mathbf{\Sigma}^\star$ is in fact a minimizer of the right side of (188) under the constraint $\tilde{\mathbf{\Sigma}} \preceq \mathbf{\Sigma}$. Now let $U^\star$ and N^m independent m-dimensional Gaussian vectors whose means sum to $\mathbb{E}[Y^m]$ and whose variances are $\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}}$ and $\tilde{\mathbf{\Sigma}}$, respectively. Put $Y^m = U^\star + N^m$, and the corresponding $P_{U|Y^m}$ is a minimizer for the left side of (187). □

The constraint in (177) can be rewritten as

$$R + \sum_{j=1}^m c_j R_j \geqslant -\mathrm{d}^\star(Q_{Y^m}, (\lambda), c^m) - \sum_{j=1}^m \frac{c_j}{2} \log 2\pi e D_j, \tag{190}$$

for any $c^m \in (0,1)$. We now prove a second-order converse.

Theorem 17. *Let Q_{Y^m} be an m-dimensional Gaussian distribution with a non-degenerate covariance matrix, and λ be the Lebesgue measure on $\mathbb{R}$. Let $c^m \in (0,\infty)^m$, and define $c_0 := 1$. Consider a sequence of Gray-Wyner coding schemes (indexed by n) for the stationary memoryless source with per-letter distribution Q_{Y^m}, and define*

$$A := \limsup_{n\to\infty} \sqrt{n} \left[\frac{1}{n} \sum_{j=0}^m c_j \log |\mathcal{W}_j| + \mathrm{d}^\star + \sum_{j=0}^m \frac{c_j}{2} \log 2\pi e D_j \right] \tag{191}$$

where $\mathrm{d}^\star := \mathrm{d}^\star(Q_{Y^m}, (\lambda), c^m)$. *Then for any* $D_j \in (0, \infty)$, $j = 1, \ldots, m$,

$$\liminf_{n\to\infty} \mathbb{P}[\exists j : \|Y_j{}^n - \hat{Y}_j^n\|^2 > nD_Pj] \geqslant \mathrm{Q}\left(\frac{A}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}}(Y^m))}}\right) \tag{192}$$

where $\phi(\cdot)$ *is as in Section III.*

Proof. First, observe that we will only need to consider the case of $\sum_j c_j \leqslant m$, since otherwise $\mathrm{d}^\star(Q_{Y^m}, (\lambda), c^m) = \infty$ by Proposition 16, in which case the claim is vacuous. Using the bound on the smooth BL divergence (Theorem 4), we take $\nu_j = \lambda$ in (170) and obtain that

$$\mathbb{P}[\exists j \colon \|Y_j{}^n - \hat{Y}_j^n\|^2 > nD_j] \geqslant 1 - \inf_{\gamma\in\mathbb{R}} \left\{ \exp(\gamma) \prod_{j=0}^{m} |\mathcal{W}_j|^{c_j} L_j^{c_j} + \mathbb{P}\left[\sum_{i=1}^{n} \nabla\phi|_{Q_{Y^m}}(Y^m{}_i) > \gamma\right] \right\} \tag{193}$$

where $Y^m \sim Q_{Y^m}$. Here we used Proposition 2-3) to show that

$$\nabla\phi|_{Q_{Y^{mn}}}(Y^{mn}) = \sum_{i=1}^{n} \nabla\phi|_{Q_{Y^m}}(Y^m{}_i). \tag{194}$$

Note that in Theorem 14,

$$L_j = |B_n(\sqrt{nD_j})| \tag{195}$$

$$= \frac{1 + O(n^{-1})}{\sqrt{n\pi}} (2\pi e D_j)^{\frac{n}{2}} \tag{196}$$

is the volume of an n-dimensional ball of radius $\sqrt{nD_j}$. Taking $\gamma = n\,\mathrm{d}^\star(Q_{Y^m}, (\lambda), c^m) - \sqrt{n}A'$ for any $A' > A$ shows that

$$\mathbb{P}[\exists j \colon \|Y_j{}^n - \hat{Y}_j^n\|^2 > nD_j] \geqslant \mathrm{Q}\left(\frac{A'}{\sqrt{\mathrm{Var}(\nabla\phi|_{Q_{Y^m}}(Y^m))}}\right). \tag{197}$$

Taking $A' \downarrow A$ establishes the claim. □

ACKNOWLEDGMENT

The proofs in this article differ significantly from the conference version [2]

REFERENCES

[1] J. Liu, P. Cuff, and S. Verdú, "Secret key generation with one communicator and a one-shot converse via hypercontractivity," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 710–714.

[2] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Smoothing Brascamp-Lieb inequalities and strong converses for common randomness generation," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 1043–1047.

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

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

[5] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[6] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2012, pp. 1–8.

[7] J. Liu, P. Cuff, and S. Verdú, "One-shot mutual covering lemma and Marton's inner bound with a common message," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 1457–1461.

[8] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1287–1291.

[9] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1574–1605, Apr. 2015.

[10] P. Moulin, "Asymptotic Neyman-Pearson games for converse to the channel coding theorem," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1541–1545.

[11] V. Kostina, "Lossy data compression: Nonasymptotic fundamental limits," Ph.D. dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, 2013.

[12] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Found. Trends Commun. Inf. Theory*, vol. 11, nos. 1–2, pp. 1–184, Sep. 2014.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[14] H. Tyagi and P. Narayan, "The Gelfand-Pinsker channel: Strong converse and upper bound for the reliability function," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2009, pp. 1954–1957.

[15] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2009, pp. 2316–2320.

[16] S. Watanabe, "Second-order region for Gray–Wyner network," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 1006–1018, Feb. 2017.

[17] L. Zhou, V. Y. F. Tan, and M. Motani, "Discrete lossy Gray–Wyner revisited: Second-order asymptotics, large and moderate deviations," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1766–1791, Mar. 2017.

[18] J. Scarlett, "On the dispersions of the Gel'fand–Pinsker channel and dirty paper coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4569–4586, Sep. 2015.

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

[20] E. H. Lieb, "Gaussian kernels have only Gaussian maximizers," *Invent. Math.*, vol. 102, no. 1, pp. 179–208, 1990.

[21] F. Barthe, "Optimal Young's inequality and its converse: A simple proof," *Geometric Funct. Anal.*, vol. 8, no. 2, pp. 234–242, 1998.

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

[23] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Information-theoretic perspectives on Brascamp-Lieb inequality and its reverse," 2017, *arXiv:1702.06260*. [Online]. Available: https://arxiv.org/abs/1702.06260

[24] S. Beigi and C. Nair, "Equivalent characterization of reverse Brascamp-Lieb-type inequalities using information measures," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1038–1042.

[25] J. Liu, T. A. Courtade, P. W. Cuff, and S. Verdú, "A forward-reverse Brascamp-Lieb inequality: Entropic duality and Gaussian optimality," *Entropy*, vol. 20, no. 6, p. 418, 2018.

[26] V. Anantharam, V. Jog, and C. Nair, "Unifying the Brascamp-Lieb inequality and the entropy power inequality," 2019, *arXiv:1901.06619*. [Online]. Available: https://arxiv.org/abs/1901.06619

[27] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson, "Algorithmic and optimization aspects of Brascamp-Lieb inequalities, via operator scaling," *Geometric Funct. Anal.*, vol. 28, no. 1, pp. 100–145, 2018.

[28] Z. Allen-Zhu, A. Garg, Y. Li, R. Oliveira, and A. Wigderson, "Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing," in *Proc. 50th Annu. ACM SIGACT Symp. Theory Comput.*, Jun. 2018, pp. 172–181.

[29] S. Sra, N. K. Vishnoi, and O. Yildiz, "On geodesically convex formulations for the Brascamp-Lieb constant," in *Proc. Approximation, Randomization, Combinat. Optim. Algorithms Techn. (APPROX/RANDOM)*, 2018, pp. 1–25.

[30] M. Hardt and A. Moitra, "Algorithms and hardness for robust subspace recovery," in *Proc. Conf. Learn. Theory*, Jun. 2013, pp. 354–375.

[31] C. Nair, *Equivalent Formulations of Hypercontractivity Using Information Measures*. Zürich, Switzerland: ETH Zürich, 2014.

[32] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6351–6355, Oct. 2011.
[33] V. Guruswami and J. Radhakrishnan, "Tight bounds for communication-assisted agreement distillation," in *Proc. 31st Conf. Comput. Complex.*, vol. 50, Jun. 2016, Art. no. 6.
[34] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover," 2013, *arXiv:1304.6133*. [Online]. Available: https://arxiv.org/abs/1304.6133
[35] R. Renner and S. Wolf, "Smooth Renyi entropy and applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2004, p. 233.
[36] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," in *Advances in Cryptology—ASIACRYPT 2005*. Berlin, Germany: Springer, 2005, pp. 199–216.
[37] T. Holenstein and R. Renner, "On the randomness of independent experiments," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1865–1871, Apr. 2011.
[38] J. Liu, "Information theory from a functional viewpoint," Ph.D. dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, 2018.
[39] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Brascamp-Lieb inequality and its reverse: An information theoretic view," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 1048–1052.
[40] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.
[41] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[42] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 3, pp. 294–300, May 1975.
[43] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.
[44] J. Liu, R. van Handel, and S. Verdú, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 943–947.
[45] J. Liu, "Dispersion bound for the Wyner-Ahlswede-Körner network via reverse hypercontractivity on types," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1854–1858.
[46] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2018, pp. 1849–1853.
[47] J. Liu, P. Cuff, and S. Verdú, "E_γ-resolvability," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2629–2658, May 2017.
[48] R. Atar, K. Chowdhary, and P. Dupuis, "Robust bounds on risk-sensitive functionals via Rényi divergence," *SIAM/ASA J. Uncertain. Quantif.*, vol. 3, no. 1, pp. 18–33, 2015.
[49] R. Atar and N. Merhav, "Information–theoretic applications of the logarithmic probability comparison bound," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5366–5386, Oct. 2015.
[50] V. Anantharam, "A variational characterization of Rényi divergences," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 893–897.
[51] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[52] C. Nair, "Upper concave envelopes and auxiliary random variables," *Int. J. Adv. Eng. Sci. Appl. Math.*, vol. 5, no. 1, pp. 12–20, 2013.
[53] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, A. Feinstein, Ed. San Francisco, CA, USA: Holden-Day, 1964.
[54] K. Yosida, *Functional Analysis*. Berlin, Germany: Springer-Verlag, 1965.
[55] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.
[56] M. Hayashi, H. Tyagi, and S. Watanabe, "Secret key agreement: General capacity and second-order asymptotics," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3796–3810, May 2016.
[57] H. Tyagi and S. Watanabe, "Converses for secret key agreement and secure computing," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4809–4827, Sep. 2015.
[58] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.
[59] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, pp. 1681–1721, Nov. 1974.

Jingbo Liu (S'14–M'18) received the B.E. degree from Tsinghua University, Beijing, China in 2012 and the Ph.D. degree from Princeton University, Princeton, NJ, USA in 2018, both in electrical engineering. He is currently a Postdoctoral Fellow at the MIT Institute for Data, Systems, and Society. His research interests include information theory, statistical inference, high dimensional probability, and the related fields. His undergraduate thesis received the best undergraduate thesis award at Tsinghua University (2012). He gave a semi-plenary presentation at the 2015 IEEE Int. Symposium on Information Theory, Hong-Kong, China. He was a recipient of the Princeton University Wallace Memorial Fellowship (2016). His Ph.D. thesis received the Bede Liu Best Dissertation Award of Princeton and the Thomas M. Cover Dissertation Award of the IEEE Information Theory Society (2018).

Thomas A. Courtade (S'06–M'13–SM'18) received the B.Sc. degree *(summa cum laude)* in electrical engineering from Michigan Technological University in 2007, and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA) in 2008 and 2012, respectively. He is currently an Assistant Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. Prior to joining UC Berkeley in 2014, he was a Postdoctoral Fellow supported by the NSF Center for Science of Information.

Prof. Courtade's honors include an NSF CAREER award and a Hellman Fellowship. He also received the Distinguished Ph.D. Dissertation Award and an Excellence in Teaching Award from the UCLA Department of Electrical Engineering, and a Jack Keil Wolf Student Paper Award for the 2012 International Symposium on Information Theory.

Paul Cuff (S'08–M'10) received the B.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 2006 and 2009. From 2009 to 2017 he was an assistant professor of electrical engineering at Princeton University. Since 2017 he has been a member of the general research group at Renaissance Technologies.

As a graduate student, Dr. Cuff was awarded the ISIT 2008 Student Paper Award for his work titled Communication Requirements for Generating Correlated Random Variables and was a recipient of the National Defense Science and Engineering Graduate Fellowship and the Numerical Technologies Fellowship. As faculty he received the NSF Career Award in 2014 and the AFOSR Young Investigator Program Award in 2015.

Sergio Verdú (S'80–M'84–SM'88–F'93) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona in 1980, and the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1984. He was on the faculty of Princeton University from 1984 to 2018.

He is the recipient of the 2007 Claude E. Shannon Award, and the 2008 IEEE Richard W. Hamming Medal. He is a member of both the National Academy of Engineering and the National Academy of Sciences. In 2016, he received the National Academy of Sciences Award for Scientific Reviewing. He is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 and 2012 Information Theory Paper Awards, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from the IEEE Communications Society. In 1998, Cambridge University Press published his book *Multiuser Detection*, for which he received the 2000 Frederick E. Terman Award from the American Society for Engineering Education. He was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005, and was elected corresponding member of the Real Academia de Ingeniería of Spain in 2013.

Dr. Verdú served as President of the IEEE Information Theory Society in 1997, and on its Board of Governors (1988–1999, 2009–2014). He has also served in various editorial capacities for the IEEE TRANSACTIONS ON INFORMATION THEORY: Associate Editor (*Shannon Theory*, 1990-1993; *Book Reviews*, 2002–2006), Guest Editor of the Special *50th Anniversary Commemorative Issue* (published by IEEE Press as Information Theory: Fifty years of discovery), and member of the Executive Editorial Board (2010–2013). He served as Creative Producer of the film *The Bit Player*, a documentary on Claude Shannon's life and legacy. He cochaired the Europe-United States *Frontiers of Engineering* program, of the National Academy of Engineering during 2009–2013. He served as the founding Editor-in-Chief of *Foundations and Trends in Communications and Information Theory*. He served as co-chair of the 2000 and 2016 IEEE International Symposia on Information Theory. He has held visiting appointments at the Australian National University, the Technion-Israel Institute of Technology, the University of Tokyo, the University of California, Berkeley, the Mathematical Sciences Research Institute, Berkeley, Stanford University, and the Massachusetts Institute of Technology.