Stability of the Gaussian Stationary Point in the Han–Kobayashi Region for Z-Interference Channels

Jingbo Liu⬛, *Member, IEEE*

Abstract— The Gaussian stationary point in an inequality motivated by the Z-interference channel was recently conjectured by Costa, Nair, Ng, and Wang to be the global optimizer, which, if true, would imply the optimality of the Han-Kobayashi region for the Gaussian Z-interference channel. This conjecture was known to be true for some parameter regimes, but the validity for all parameters, although suggested by Gaussian tensorization, was previously open. In this paper we construct several counterexamples showing that this conjecture may fail in certain regimes: A simple construction without Hermite polynomial perturbation is proposed, where distributions far from Gaussian are analytically shown to be better than the Gaussian stationary point. As alternatives, we consider perturbation along geodesics under either the L^2 or Wasserstein-2 metric, showing that the Gaussian stationary point is unstable in a certain regime. Similarity to stability of the Levy-Cramer theorem is discussed. The stability phase transition point admits a simple characterization in terms of the maximum eigenvalue of the Gaussian maximizer. Similar to the Holley-Stroock principle, we can show that in the stable regime the Gaussian stationary point is optimal in a neighborhood under the L^∞-norm with respect to the Gaussian measure. Allowing variable power control, we show that the Gaussian optimizers for the Han-Kobayashi region always lie in the stable regime. Finally, an amended conjecture is proposed.

Index Terms— Interference channels, Gaussian optimality, Han-Kobayashi bound, optimal transport, Levy-Cramer theorem, stability problems.

I. INTRODUCTION

INTERFERENCE channel is a fundamental problem in multiuser information theory, whose single-letter rate region has remained open after decades of efforts. A two-user interference channel consists of two inputs X_1 and X_2 and two outputs Y_1 and Y_2. In general, both Y_1 and Y_2 are noisy versions of functions of (X_1, X_2), and the goal is to find the maximum transmission rates (R_1, R_2) achievable by the two users. For a more precise formulation of the problem, see e.g., [9] and [20]. In the special case of Gaussian Z-interference channel, Y_1 is a noisy version only of X_1 rather than of

(X_1, X_2). With some transformations, the channel model can be expressed as (see [11])

$$Y_1 = X_1 + Z_1; \tag{1}$$

$$Y_2 = X_2 + X_1 + Z_1 + Z_2, \tag{2}$$

where $Z_i \sim \mathcal{N}(0, N_i I)$, $i = 1, 2$, and X_1, X_2, Z_1, Z_2 are independent. The Gaussian Z-interference channel is significant, partly because it is equivalent to several versions of degraded Gaussian interference channels [9].

Due to the sheer volume, we do not attempt to provide here a comprehensive review of the all the important work on the interference channel since the 70's. Interested readers may refer to the survey [47] or the more recent papers [5], [11], [26]. Below, we cite a few properties of the interference channel relevant to our purpose:

- The multi-letter Han-Kobayashi (HK) inner bound [28] is tight, whereas the single letter Han-Kobayashi bound is known to be suboptimal for some discrete channels where tensorization fails [42].
- When restricted to Gaussian inputs, the HK-region is known to satisfy the tensorization property [40].
- It is unknown whether Gaussian inputs are optimal for the HK bound for Gaussian Z-interference channels. If optimal, this would resolve the longstanding open problem about the rate region of the Gaussian Z-interference channel, because of the two itemized properties above. More or less motivated by this, the Gaussian optimality problem spurred a lot of research interests recently, e.g., [3], [5], [11], [13], [26], [37], and [43]. It was known that Gaussian inputs are optimal for computing the corners of the region [5], [9], [11], [24], [25], [26], [45], [46] but the full region or even the precise slope at Costa's corner point remains open [26].
- Gaussian inputs are known to be suboptimal for the symmetric Gaussian interference channel under constant power control [1] for some range of parameters, where the argument was based on perturbations using Hermite polynomials. Some care needs to be taken in choosing how to perturb and ensuring that the perturbed density is still a probability measure.
- Constant power control is suboptimal since they can be improved by a variable power control scheme known as "Gaussian noisebergs" [12] (see also [47]).

The second itemized above (Gaussian tensorization) seems to suggest that Gaussian inputs are optimal for the

Han-Kobayashi region. Indeed, a neat rotation-invariance argument usually succeeds in showing Gaussian optimality in entropic inequalities whenever there is a tensorization property [23] (see also similar arguments in the context of functional inequalities [31], [37]). However, the tensorization property in [40] is for Gaussian inputs rather than general inputs, therefore Gaussian optimality cannot be settled in this manner.

Since the HK bound is notoriously complicated, it is useful to formulate simpler necessary and sufficient conditions for checking its tightness. Recently, Costa, Nair, Ng and Wang [11] (see also the related [5], [26], [43]) proposed the following conjecture about linear combination of differential entropies which, if true, would imply Gaussian optimality in the Han-Kobayashi region.

Conjecture 1 [11]: For $u \geq 0$, $N_1, N_2 \geq 0$, and $\Sigma_1, A_2 \succeq 0$, the maximum

$$\max_{\substack{P_{X_1} P_{X_2} \\ \mathbb{E}[X_2 X_2^\top] \preceq A_2}} \{uh(X_1 + X_2 + Z_1 + Z_2) + h(X_1 + Z_1)$$

$$- (1+u)h(X_1 + Z_1 + Z_2) - \text{tr}(\Sigma_1 \mathbb{E}[X_1 X_1^\top])\} \quad (3)$$

where $Z_1 \sim \mathcal{N}(0, N_1 I)$, $Z_2 \sim \mathcal{N}(0, N_2 I)$ and X_i, Z_i ($i = 1, 2$) are random variables in $\mathbb{R}^d$ ($d \geq 1$), is attained by Gaussian X_1 and X_2.

It is easy to see that if (P_{X_1}, P_{X_2}) is the maximizer among all Gaussian distributions, then it is also a stationary point (i.e. vanishing gradient) in the space of all probability distributions (under suitable metric and regularity condition). Indeed, with a perturbation of a Gaussian density, the first-order change of the differential entropy depends only on the change of the covariance (see e.g. (38)), so the stationarity condition is the same as (P_{X_1}, P_{X_2}) having the optimal covariance.

Indications that the conjecture might be true include the aforementioned Gaussian tensorization property, and the fact that if either X_1 or X_2 is Gaussian then the best choice of the other random variable is also Gaussian [5], [11]. Previously, the conjecture has been proved for large enough u, which is enough to evaluate the corner points of the rate region, even though these corner points were previously established by other methods; see [26]. However, it was not known whether it is true for all u, so that one can deduce Gaussian optimality for the capacity *region*, which is the central open problem on Z-interference channels. As a partial converse, [5] observed that disproving Conjecture 1 around Costa's corner point would imply suboptimality of the HK region.

In this paper, we show that Conjecture 1 may fail for some range of parameters, by constructing counterexamples with several methods:

- We choose a set of parameters for which (3) is nonpositive with Gaussian inputs, yet strictly positive for some non-Gaussian inputs (Section II-A). This counterexample works when (fixing all other parameters) A_2 is the maximum number for which the optimal Gaussian X_1 does not have a finite variance. This simple proof is very different from the existing method of Hermite polynomial perturbation [1]. In fact, our non-Gaussian inputs have non-vanishing non-Gaussianness in the proof. Using this construction, in Section II-B we give a short alternative

proof of the fact that Gaussian inputs with constant power control is suboptimal for the HK bound, which was originally shown using a variable power control Gaussian input called "Gaussian noisebergs" [12]. We discuss an analogous geometric inequality with non-isotropic extremals (Section VI).

- Alternatively, a counterexample can be constructed by perturbing the Gaussian "vertically", i.e. along a geodesic under the $L^2(\mathbb{R})$ metric (Section III), which works whenever the variance K of the optimal Gaussian X_1 satisfies $\frac{1+u}{(K+N_1+N_2)^3} > \frac{1}{K+N_1}$. This counterexample amounts to computing the Hessian under $L^2(\mathbb{R})$, revealing that the Gaussian stationary point is unstable in a certain parameter regime. Since differentiation commutes with convolution, derivatives of the Gaussian density are useful in constructing the direction of perturbation. which is closely related the Hermite perturbation approach [1]. Previously, [41] commented that attempts of disproving Conjecture 1 using the method of [1] were unsuccessful. We note that in order to ensure nonnegativity of the density after perturbation, [1] introduced a device of adding an additional even-order Hermite polynomial (because odd-order Hermite polynomials are unbounded in both directions, whereas even-order Hermite polynomials are bounded below); here we handle the nonnegativity issue in another way which is simpler for the present problem (Remark 2). To explain the intuition we observe the similarity to the stability problem in the Levy-Cramer theorem (Remark 3).

- A counterexample can also be constructed by computing the Hessian under the Wasserstein-2 distance, based on the Otto-Villani framework of viewing the space of probability measures as an infinite dimensional Riemannian manifold with the Wasserstein-2 metric. Since this counterexample works for exactly the same set of parameters as vertical perturbation, we shall leave the calculations to Appendix B. In standard Riemannian geometry, the formula of the Hessian in a coordinate system depends on the second derivative as well as the product of the connection and the first derivative (see e.g. [22], (2.6)]), from which one can read off how the Hessian changes under a change of metric. At a general point on the manifold, semidefiniteness of the Hessian depends on the choice of metric, because of the term of the product of the connection and the first derivative. However at a stationary point where the first derivative vanishes, semidefiniteness of the Hessian does not depend on the choice of metric. In Appendix B we provide self-consistent calculations confirming that the stability phase transition point of the Hessian is consistent under either the L^2 or Wasserstein-2 distance. In contrast to the vertical perturbation method, with the Wasserstein flow approach we never need to worry about the nonnegativity constraint of the density. Moreover, recently, first and second order gradient descent methods under the Wasserstein metric have been used in numerically solving optimization problems, with convenient particle flow interpretation and computational advantages over the

traditional descent methods under the L^2 metric [2], [7], [15], [50].

For parameters of Conjecture 1 giving rise to negative-definite Hessian at the Gaussian stationary point, we show that the Gaussian stationary point is stable in the sense that it is a local maximizer under the L^∞-norm with respect to the Gaussian reference measure (Section V). The choice of L^∞-norm is in the same spirit as the celebrated Holley-Stroock principle [29, p1184] [30], which is a method of proving functional inequalities (such as log-Sobolev or Poincare) under a perturbation. For such a local optimality result, it suffices to bound $h(X_1 + X_2 + Z_1 + Z_2)$ in (3) by the surrogate $\frac{1}{2} \ln[(2\pi e)^d \det(\text{cov}(X_1 + X_2 + Z_1 + Z_2))]$. It is well-known in variational calculus that the choice of topology is critical for local optimality, and in fact this is precisely the case here: local optimality (with the surrogate objective function) is not true under the L^∞-norm with respect to the *Lebesgue measure*; a similar phenomenon was previous observed concerning the stability of the log-Sobolev inequality [21, Section 6].

From the Hessian computations we see that the stability phase transition point admits a simple characterization in terms of the maximum eigenvalue of the covariance matrix of the Gaussian maximizer in (3). On the other hand, in Section IV we establish a bound on the maximum eigenvalue of the covariance matrix of the Gaussian maximizer in the HK region with (time-varying) power control. Comparing the two results, we see that our counterexamples are not sufficient for disproving Gaussian optimality for the HK region with power control, nor do they prove [5, Hypothesis 1]. The main message of our counterexamples, however, is that we cannot expect to prove Gaussian optimality for the HK bound by proving this for any given power of input signals - a seemingly plausible approach that might have motivated Conjecture 1.

While stability of stationary points is not sufficient for global optimality, it is often the first step towards understanding many phase transition problems (e.g. [18]). Our results suggest that towards the grand goal of settling Gaussian optimality in the HK region with power control, we need to modify Conjecture 1, or find counterexamples not based on Gaussian perturbation (under the L^∞-norm). For a compact version that seems to represent much of the challenge, we may consider the following limiting special case of the optimization problem: (see Remark 6, Theorem 3, and Lemma 3 for explanations of the connection):

$$\sup\left\{ h(X+Y) - h(X) - \frac{1}{2}J(X) \right\} \tag{4}$$

where X and Y are independent one-dimensional random variables, $\mathbb{E}[Y] \le L$, and $J(\cdot)$ denotes the Fisher information [14]. The best Gaussian X (stationary point) has variance $\frac{L}{L-1}$ when $L > 1$. Results of this paper imply that the Gaussian stationary point is not optimal for $1 < L < 1.5$, but local optimal (stable) for $L > 1.5$. The true Gaussian maximizer in the HK bound with power control concerns the region of $L > 2$. If the Gaussian stationary point in (4) is not global optimal for some value of $L > 2$, we can show that the HK inner bound is not tight.

Finally, in Section VII we propose an amended conjecture that takes into account of the stability regime of parameters (more precisely, it concerns the regime where power control is not necessary for the Gaussian HK bound). The new conjecture is nontrivial, since either proving or disproving it would imply the solution to some other open questions in the literature.

Notation: We denote the differential entropy [14] of a random variable $X \sim P_X$ by $h(X)$ or $h(P_X)$. The unit of entropy is nat. γ_K denotes the centered Gaussian measure with covariance matrix K. When used as a density function, γ_K denotes the density with respect to the Lebesgue measure, unless otherwise stated. The norms with respect to a reference measure are computed using the density with respect to the reference measure; for example, $\|\gamma_K\|_{L^2(P)}^2 = \int(\frac{d\gamma_K}{dP})^2 dP$. We use standard Landau notations such as $\Theta()$, $\omega()$, and $O()$.

II. GAUSSIAN SUBOPTIMALITY WITHOUT GAUSSIAN PERTURBATION

In this section we construct a counterexample to Conjecture 1, and use it to give an alternative proof of the suboptimality of Gaussian signaling with constant power control [12].

A. Construction of a Counterexample

We first note the following about the asymptotic expansion of the differential entropy of convolution, which is similar to the calculations in the de Bruijn's inequality (see e.g. [14]).

Lemma 1: Let p and q be smooth density functions on $\mathbb{R}$. Denote the moments of q by $m_i := \int x^i q(x) dx = 0$, $i = 1, 2, \dots$, and suppose that $m_1 = 0$. Suppose that there are positive constant c and C such that for any $x \in \mathbb{R}$,

$$q(x) \le C e^{-cx^2}, \tag{5}$$

$$p(x) > c\, e^{-C|x|}, \tag{6}$$

$$\max\{|p'(x)|, |p'''(x)|, |p''''(x)|\} < Cp(x). \tag{7}$$

Let p_t be the convolution of p and $t^{-1/2}q(\frac{\cdot}{\sqrt{t}})$. Then for $t > 0$ small, we have the following estimate for the differential entropy:

$$h(p_t) - h(p)$$
$$= m_2(-\frac{1}{2}\int p'' \ln p)t + m_3(-\frac{1}{6}\int p''' \ln p)t^{3/2} + O(t^2) \tag{8}$$

where $O()$ hides constants that may depend on p and q. The proof is given in Appendix A.

Lemma 2: Suppose that $Z_2 \sim \mathcal{N}(0, N_2)$, $N_2 > 0$. Then the supremum of

$$h(X_1 + Z_2 + X_2) + h(X_1) - 2h(X_1 + Z_2) \tag{9}$$

over the distribution of (X_1, X_2), where X_1, X_2, Z_2 are independent and $\mathbb{E}[X_2^2] = N_2$, is strictly positive. Moreover if X_2 is restricted to be Gaussian then the supremum equals 0.

Proof: First, if X_2 is Gaussian, then

$$h(X_1 + Z_2 + X_2) - h(X_1 + Z_2)$$
$$= I(X_1 + Z_2 + X_2; X_2) \tag{10}$$
$$= I(X_1 + Z_2 + X_2; Z_2) \tag{11}$$

$$\leq I(X_1 + Z_2; Z_2) \tag{12}$$
$$= h(X_1 + Z_2) - h(X_1), \tag{13}$$

implying that the supremum is non-positive. Further, taking X_1 to be Gaussian with a large variance we see that the supremum equals 0.

Now consider the case where X_2 is not restricted to be Gaussian. Choose p and q satisfying the conditions in Lemma 1 and such that

$$m_3 \int p''' \ln p > 0; \tag{14}$$
$$m_1 = 0; \tag{15}$$
$$m_2 > 0. \tag{16}$$

It should be clear that the supremum of (9) is independent of the choice of N_2 (by considering a scaling of the random variables). For simplicity let us assume that $N_2 = m_2$ below. Let $X_2 \sim q$ and let X_1 be such that $\sqrt{t} X_1 \sim p$ (in particular, observe that the distribution of X_1 depends on t, but the distribution of $\sqrt{t} X_1$ does not). Note that

$$\mathbb{E}[(Z_2 + X_2)^2] = N_2 + m_2 = 2m_2; \tag{17}$$
$$\mathbb{E}[(Z_2 + X_2)^3] = \mathbb{E}[X_2^3] = m_3. \tag{18}$$

Then

$$h(X_1 + Z_2 + X_2) - h(X_1)$$
$$= h(\sqrt{t} X_1 + \sqrt{t} Z_2 + \sqrt{t} X_2) - h(\sqrt{t} X_1) \tag{19}$$
$$= 2m_2(-\frac{1}{2} \int p'' \ln p)t + m_3(-\frac{1}{6} \int p''' \ln p)t^{3/2}$$
$$+ O(t^2); \tag{20}$$

Similarly, since Z_2 is symmetric whose third moment equals zero, we have

$$h(X_1 + Z_2) - h(X_1)$$
$$= h(\sqrt{t} X_1 + \sqrt{t} Z_2) - h(\sqrt{t} X_1) \tag{21}$$
$$= m_2(-\frac{1}{2} \int p'' \ln p)t + O(t^2). \tag{22}$$

Then we note that (9) equals (20) minus twice of (22), which is $m_3(-\frac{1}{6} \int p''' \ln p)t^{3/2} + O(t^2)$. Thus the claim about positivity of the supremum follows when t is sufficiently small. □

Remark 1: Although not necessary for establishing a counterexample to Conjecture 1, let us comment that X_2 being Gaussian implies that Gaussian X_1 is optimal. Indeed, this has been shown using the doubling trick [11], or, even simpler, using Costa's entropy power inequality [5].

Clearly Lemma 2 provides a counterexample to Conjecture 1 with $u = 1$, $N_1 = 0$, $N_2 = A_2$, and $\Sigma_1 = 0$. While strictly positive N_1 sometimes plays a role in analysis related to interference channels, for example in the proof of Wasserstein continuity of smoothed entropy [45], it is not essential in the proof of Gaussian suboptimality. Below, we show that a counterexample exists for N_1 and Σ_1 strictly positive as well:

Theorem 1: Let $k = u = N_1 = 1$. There exists some $N_2 = A_2 > 0$ and $\Sigma_1 > 0$ for which (3) is strictly positive yet the supremum restricted to Gaussian X_2 is not positive.

Proof: We first provide a proof using a continuity argument. By replacing X_1 in Lemma 2 with $X_1 + Z_1$, $Z_1 \sim \mathcal{N}(0, N_1)$, and by continuity of the differential entropy in N_1 (see e.g. [3, Lemma A.3.]), we see that the supremum of

$$h(X_1 + Z_1 + Z_2 + X_2) + h(X_1 + Z_1)$$
$$- 2h(X_1 + Z_1 + Z_2) - \Sigma_1 \mathbb{E}[X_1^2] \tag{23}$$

over the distribution of (X_1, X_2), where X_1, X_2, Z_2 are independent and $\mathbb{E}[X_2^2] \leq N_2 = \text{var}(Z_2)$, is strictly positive provided that N_1 and Σ_1 are sufficiently small. Also, it is clear that the supremum in (23) does not change under the transformations $N_1' = 1$, $A_2' = N_2' = \frac{N_2}{N_1}$, and $\Sigma_1' = \Sigma_1 N_1$ (to see this consider $X_1' = \frac{1}{\sqrt{N_1}} X_1$ and $X_2' = \frac{1}{\sqrt{N_1}} X_2$). Therefore, equivalently, fixing $N_1 = 1$, we can always find $N_2 = A_2 > 0$ and $\Sigma_1 > 0$ so that the supremum is positive. In the meantime, if X_2 is restricted to be Gaussian then the supremum does not exceed 0 as in Lemma 2.

An alternative argument for positivity of (23) was suggested by Chandra Nair: for $X_1 \sim p$ as in Lemma 1, we can pick a finite $N_1 > 0$ and verify that $X_1 + Z_1$ still satisfies the assumptions on p in Lemma 1. Then treating $X_1 + Z_1$ as the new X_1, the proof of Lemma 2 still shows that (23) is positive for some $\Sigma_1 > 0$, $N_2 = A_2 > 0$ and X_2. □

B. Alternative Proof Gaussian Suboptimality Under Constant Power Control

Recall that [12] proposed a variable power control scheme called Gaussian noisebergs showing that Gaussian signaling with constant power control is suboptimal. In this section we provide a short alternative proof of this fact using Lemma 1. The variable power Gaussian signaling construction in [12] does not rule out the possibility that Gaussian signaling is optimal for the HK bound evaluated at any given power budget, which might have motivated Conjecture 1 as a plausible route for establishing the optimality of Gaussian HK region. The analysis given here rules out such a possibility.

We first recall some results in [11]. Given two distributions P_{X_1}, P_{X_2} on $\mathbb{R}^d$, define

$$\Psi(P_{X_1}, P_{X_2}) := uh(X_1 + X_2 + Z_1 + Z_2) + h(X_1 + Z_1)$$
$$- (u+1)h(X_1 + Z_1 + Z_2), \tag{24}$$

where $Z_j \sim \mathcal{N}(0, N_j I)$, $j = 1, 2$, X_1, X_2, Z_1, Z_2 are independent, and $u \geq 1$. Given P_{X_1} and P_{X_2}, define the concave envelope by

$$\mathcal{C}_{X_1}[\Psi(P_{X_1}, P_{X_2})] := \sup_{P_{X_1 U}} \{h(X_1 + X_2 + Z_1 + Z_2 | U)$$
$$+ h(X_1 + Z_1 | U) - (u+1)h(X_1 + Z_1 + Z_2 | U)\} \tag{25}$$

where the supremum is over $P_{X_1 U}$ compatible with the given P_{X_1}, and (U, X_1) is independent of (X_2, Z_1, Z_2). Define

$$F_u(P_{X_1}, P_{X_2}) := h(X_1 + X_2 + Z_1 + Z_2) - h(Z_1)$$
$$+ \mathcal{C}_{X_1}[\Psi(P_{X_1}, P_{X_2})]. \tag{26}$$

Recall that the d-letter Han-Kobayashi (HK) region for the Gaussian Z-channel is given by (see e.g. [10])

$$dR_1 \leq h(X_1 + Z_1 | Q) - h(Z_1) \tag{27}$$

$$dR_2 \leq h(X_1 + X_2 + Z_1 + Z_2 | U_1, Q)$$
$$- h(X_1 + Z_1 + Z_2 | U_1, Q) \tag{28}$$
$$d(R_1 + R_2) \leq h(X_1 + X_2 + Z_1 + Z_2 | Q)$$
$$- h(X_1 + Z_1 + Z_2 | U_1, Q)$$
$$+ h(X_1 + Z_1 | U_1, Q) - h(Z_1) \tag{29}$$

where $Z_1 \sim \mathcal{N}(0, I)$, $Z_2 \sim \mathcal{N}(0, N_2 I)$. Q is the power control random variable, and the joint distribution factors as $P_Q P_{U_1 X_1 | Q} P_{X_2 | Q} P_{Z_1} P_{Z_2}$. Following [11], we note that for $u \geq 0$, the HK bound can be expressed using the weighted form

$$d(R_1 + (1 + u)R_2)$$
$$\leq \sup_{P_{Q_1 Q_2} : \mathbb{E}[Q_j] \leq q_j} \mathbb{E}_{Q_1, Q_2} \left[\sup_{P_{X_j} : \mathbb{E}[\|X_j\|^2] \leq Q_j} F_u(P_{X_1}, P_{X_2}) \right] \tag{30}$$

where the first supremum is over distributions of (Q_1, Q_2) which are random non-negative powers satisfying a given power constraint $\mathbb{E}[Q_j] \leq q_j$, $j = 1, 2$. For any given Q_j, $j = 1, 2$, the second supremum is over (P_{X_1}, P_{X_2}) satisfying the indicated power constraints.

We remark that for $u \in (-1, 0]$, it is obvious that Gaussian variables are optimal for the weighted sum rates. Indeed, (27) is optimized by Gaussian X_1 under a power constraint. Gaussian variables are also optimal for (29) since $-h(X_1 + Z_1 + Z_2 | U_1, Q) + h(X_1 + Z_1 | U_1, Q) = -I(X_1 + Z_1 + Z_2; Z_2 | U_1, Q)$ which is maximized by Gaussian variables under a power constraint (by Gaussian saddle point; see e.g. [14]).

We now show that under constant power control (i.e. Q_1 and Q_2 are constants) and $u = 1$, Gaussian variables may be suboptimal for the weighted sum rate.

Theorem 2: There exist $N_1 = 1$, $N_2 > 0$, $u = 1$, $d = 1$, and some (deterministic) $q_1, q_2 > 0$ such that Gaussian P_{X_j}, $j = 1, 2$ are not optimal for the following optimization:

$$\sup_{P_{X_j} : \mathbb{E}[\|X_j\|^2] \leq q_j} F_u(P_{X_1}, P_{X_2}). \tag{31}$$

Proof: Using the argument around (23) we see that there exists $N_1 = 1$, $u = 1$ and $N_2 > 0$ such that

$$\sup_{P_{X_1'}, P_{X_2} : \mathbb{E}[X_2^2] \leq N_2} \Psi(P_{X_1'}, P_{X_2}) > 0. \tag{32}$$

Now we consider any X_1' and X_2 such that $\Psi(P_{X_1'}, P_{X_2}) = c > 0$ and $\mathbb{E}[X_2^2] \leq N_2$. Let $U \sim \mathcal{N}(0, A)$ independent of (X_1', X_2, Z_1, Z_2) where $A > 0$ will be chosen later. Let $X_1 = X_1' + U$. Then we have

$$\mathcal{C}_{X_1}[\Psi(P_{X_1}, P_{X_2})] \geq \Psi(P_{X_1'}, P_{X_2}) = c. \tag{33}$$

Moreover, by choosing A large enough (while other parameters are kept fixed), we have

$$h(X_1 + X_2 + Z_1 + Z_2) - h(Z_1) \tag{34}$$
$$\geq \frac{1}{2} \ln \frac{\text{var}(X_1') + A + N_1 + 2N_2}{N_1} - \frac{c}{2}. \tag{35}$$

For such A and with $q_1 := \text{var}(X_1') + A$, $q_2 := N_2$ we have

$$\sup_{P_{X_j} : \mathbb{E}[X_j^2] \leq q_j, j=1,2} F_u(P_{X_1}, P_{X_2})$$

$$\geq \frac{1}{2} \ln \frac{\text{var}(X_1') + A + N_1 + 2N_2}{N_1} + \frac{c}{2}. \tag{36}$$

On the other hand, if the supremum in (36) is restricted to Gaussian X_1, X_2 with the same variances, then it was shown in [11] that Gaussian (U, X_1) is optimal in (25). By the result of Lemma 2 we have that $\mathcal{C}_{X_1}[\Psi(P_{X_1}, P_{X_2})] \leq 0$. Moreover under $\mathbb{E}[X_j^2] \leq q_j, j = 1, 2$ it is also obvious that

$$h(X_1 + X_2 + Z_1 + Z_2) - h(Z_1)$$
$$\leq \frac{1}{2} \ln \frac{\text{var}(X_1') + A + N_1 + 2N_2}{N_1}. \tag{37}$$

The same upper bounds holds for the left side of (36) if X_1, X_2 are restricted to be Gaussian, and the claim of the theorem follows by comparing it with (36). $\square$

III. GAUSSIAN SUBOPTIMALITY VIA VERTICAL PERTURBATION

In this section we provide an alternative construction of counterexample to Conjecture 1 via "vertical perturbation"[1] of Gaussian. This amounts to showing instability of the Gaussian stationary in a certain power regime. In Appendix B, we provide an alternative proof of instability in the same regime via horizontal perturbation.

First, let us observe the following about the differential entropy under the vertical perturbation.

Proposition 1: Assume that p is probability density function on $\mathbb{R}^d$, and r is a measurable function on $\mathbb{R}^d$ satisfying $\sup_{x \in \mathbb{R}^d} |\frac{r(x)}{p(x)}| < \infty$ and $\int r = 0$. Then as $\epsilon \to 0$, we have

$$\int (p + \epsilon r) \ln(p + \epsilon r) = \int p \ln p + \epsilon \int r \ln p + \frac{\epsilon^2}{2} \int \frac{r^2}{p}$$
$$+ O(\epsilon^3). \tag{38}$$

Proof: Using the Taylor expansion of the function $t \mapsto t \ln t$ we see that for $\epsilon < \inf_{x \in \mathbb{R}^d} |\frac{p(x)}{r(x)}|$,

$$(p + \epsilon r) \ln(p + \epsilon r) = p \ln p + \epsilon r + \epsilon r \ln p + \frac{\epsilon^2 r^2}{2\xi} \tag{39}$$

where the function $\xi(x)$ is between $p(x)$ and $p(x) + \epsilon r(x)$. For $\epsilon < \frac{1}{2} \inf_{x \in \mathbb{R}^d} |\frac{p(x)}{r(x)}|$ we have

$$\left| \frac{r^2}{\xi} - \frac{r^2}{p} \right| = \frac{r^2 |p - \xi|}{\xi p} \leq \frac{2\epsilon r^3}{p^2}. \tag{40}$$

The claim then follows by integrating (39). $\square$

The key result in the construction of counterexample is the following:

Theorem 3: For any $u > 0$, $L > 0$, and $T > 0$, let $K > \frac{u}{(1+u)^{1/3} - 1}$. There exist $P_{X_1}^\epsilon$ and $P_{X_2}^\epsilon$ (indexed by ϵ in a neighborhood of 0) such that $h(P_{X_1}^\epsilon * \gamma_u * P_{X_2}^\epsilon) = h(\gamma_{K+u+L}) + O(\epsilon^{2T})$ and $h(P_{X_1}^\epsilon) - (u+1)h(P_{X_1}^\epsilon * \gamma_u) \geq h(\gamma_K) - (u+1)h(\gamma_{K+u}) + A\epsilon^2$ for some $A > 0$, as $\varepsilon \to 0$.

Proof: Define

$$P_{X_1}^\epsilon = \gamma_K - \epsilon D^3 \gamma_{K-\delta}; \tag{41}$$

---

[1] In the literature, "vertical perturbation" usually refers to a linear perturbation of the probability density, whereas a "horizontal perturbation" is along a Wasserstein geodesic (c.f. Appendix B).

$$P_{X_2}^\epsilon = \sum_{j=0}^J \epsilon^j D^{3j} \gamma_{L-j\delta}, \qquad (42)$$

where the positive integer J satisfies $J + 1 \geq T$, and $\delta > 0$ satisfies $K - \delta > 0$, $L - J\delta > 0$. Moreover $D^3 \gamma_{K-\delta}$ means differentiating the Gaussian density 3 times, which yields a finite measure on $\mathbb{R}$. Thus (41)-(42) define valid probability measures when ϵ is sufficiently close to 0.

Then for $Z_2 \sim \mathcal{N}(0, u)$ we have

$$P_{X_1}^\epsilon * \gamma_u = (\gamma_K - \epsilon D^3 \gamma_{K-\delta}) * \gamma_u \qquad (43)$$

$$= \gamma_{K+u} - \epsilon D^3 \gamma_{K+u-\delta} \qquad (44)$$

and

$$P_{X_1}^\epsilon * P_{X_2}^\epsilon$$

$$= (\gamma_K - \epsilon D^3 \gamma_{K-\delta}) * (\sum_{j=0}^J \epsilon^j D^{3j} \gamma_{L-j\delta}) \qquad (45)$$

$$= \sum_{j=0}^J \epsilon^j D^{3j} \gamma_{K+L-j\delta} - \sum_{j=0}^J \epsilon^{j+1} D^{3(j+1)} \gamma_{K+L-(j+1)\delta}$$

$$\qquad (46)$$

$$= \gamma_{K+L} - \epsilon^{J+1} D^{3(J+1)} \gamma_{K+L-(J+1)\delta}. \qquad (47)$$

Therefore

$$P_{X_1}^\epsilon * \gamma_u * P_{X_2}^\epsilon = \gamma_{K+u+L} - \epsilon^{J+1} D^{3(J+1)} \gamma_{K+u+L-(J+1)\delta}. \qquad (48)$$

Using (38) we have

$$h(P_{X_1}^\epsilon) = h(\gamma_K) - \frac{\epsilon^2}{2} \int \frac{(D^3 \gamma_{K-\delta})^2}{\gamma_K} + o(\epsilon^2); \qquad (49)$$

$$h(P_{X_1}^\epsilon * \gamma_u) = h(\gamma_{K+u}) - \frac{\epsilon^2}{2} \int \frac{(D^3 \gamma_{K+u-\delta})^2}{\gamma_{K+u}} + o(\epsilon^2); \qquad (50)$$

and

$$h(P_{X_1}^\epsilon * \gamma_u * P_{X_2}^\epsilon) = h(\gamma_{K+u+L})$$
$$- \frac{\epsilon^{2(J+1)}}{2} \int \frac{(D^{3(J+1)} \gamma_{K+u+L-(J+1)\delta})^2}{\gamma_{K+u+L}} + o(\epsilon^{2(J+1)}). \qquad (51)$$

Clearly the claim of the theorem holds if

$$- \int \frac{(D^3 \gamma_{K-\delta})^2}{\gamma_K} + (1+u) \int \frac{(D^3 \gamma_{K+u-\delta})^2}{\gamma_{K+u}} > 0. \qquad (52)$$

By the property of the Hermite polynomial (see e.g. [17]) we have

$$\int \frac{(D^k \gamma_K)^2}{\gamma_K} = \frac{1}{K^k} \int \frac{(D^k \gamma)^2}{\gamma} = \frac{1}{K^k} k! \qquad (53)$$

for any positive integer k. Thus by dominated convergence we see that as $\delta \to 0$ the left side of (52) converges to $-\frac{1}{K^3} k! + \frac{1+u}{(K+u)^3} k!$. Therefore (52) holds under $K > \frac{u}{(1+u)^{1/3}-1}$. $\qquad \square$

Remark 2: Setting $\delta > 0$ in (41) ensures that it defines a valid probability measure with nonnegative density for small ϵ. Note that if $\delta = 0$ then the density of the perturbation part (with respect to γ_K) is a Hermite polynomial of order

three. Previously, Hermite polynomial perturbation has been used in [1] for proofs of Gaussian suboptimality. In order to ensure nonnegative density, [1] would add yet another even degree Hermite polynomial (multiplied by γ_K) to (41), since an even degree Hermite polynomial is bounded from below. The ratio of the odd and even degree polynomials needs be selected with care, and a similar issue for the perturbation of X_2 would add more restrictions. Here we adopt a simpler trick without adding an even degree polynomial, but choosing a smaller Gaussian variance.

Remark 3: The intuition behind Theorem 3 is closely related to (instability of) the Levy-Cramer theorem. The Levy-Cramer theorem states that the sum of two independent non-Gaussian random variables cannot be precisely Gaussian [16]. However, the non-Gaussianness of the sum can be much smaller than that of the individual summands, and Theorem 3 essentially showed this when the non-Gaussianness is gauged by regularized relative entropy. The reverse direction (lower bounding the non-Gaussianness of the sum using non-Gaussianness of the individual summands) is the problem of stability of the Levy-Cramer theorem, the regularized entropy version of which was considered in [6].

Corollary 1: Consider any $u > 0$ and $L > 1$ satisfying $\frac{L+u}{L-1} > \frac{u}{(1+u)^{1/3}-1}$. There exists some (P_{X_1}, P_{X_2}) for which

$$uh(P_{X_1} * \gamma_u * P_{X_2}) + h(P_{X_1}) - (u+1)h(P_{X_1} * \gamma_u)$$
$$> \sup_{K>0} \{uh(\gamma_{K+u+L}) + h(\gamma_K) - (u+1)h(\gamma_{K+u})\}. \qquad (54)$$

Proof: We can check that the supremum in (54) is achieved at $K = \frac{L+u}{L-1}$ if $L > 1$. Under the assumption of the corollary the claim follows from Theorem 3. $\qquad \square$

Corollary 1 provides a counterexample to Conjecture 1 for the case of $N_1 = \Sigma_1 = 0$ and $N_2 = u$. Extension to the $N_1, \Sigma_1 > 0$ case is possible using a similar continuity argument as Theorem 1). Alternatively, we can pick finite but small enough $N_1 > 0$, define $P_{X_1}^\epsilon = \gamma_{K-N_1} - \epsilon D^3 \gamma_{K-N_1-\delta}$, and then the argument in Theorem 3 still carries through with X_1 replaced by $X_1 + Z_1$.

IV. STRUCTURE OF THE GAUSSIAN EXTREMIZERS

Although counterexamples in the previous sections disproved Conjecture 1, they do not show Gaussian suboptimality for the Han-Kobayashi (HK) region allowing power control. The reason is that in Corollary 1 the counterexample exists only when the Gaussian maximizer P_{X_1} has covariance larger than $\frac{u}{(1+u)^{1/3}-1}$. The purpose of this section is to show that the Gaussian maximizer for the HK region (allowing power control) never falls in this regime. Let us remark that Conjecture 1 is a necessary but not sufficient condition for Gaussian optimality of the HK region allowing power control, and it is not clear if Conjecture 1 can be easily amended (e.g. restricting to the case where the covariance of the Gaussian optimizer satisfies a certain bound) and remains a necessary condition.

We define the following quantities that characterize the HK region (30) restricted to Gaussian signaling: Given $u, N_1 > 0$ and positive semidefinite matrices K and L of the same

dimensions, define

$$\psi(K, L) := u \ln \det(K + N_1 I + uI + L) + \ln \det(K + N_1 I)$$
$$- (u + 1) \ln \det(K + N_1 I + uI). \quad (55)$$

For positive semidefinite matrices K and L of the same dimensions, define

$$\phi(J, L) := \sup_{K \preceq J} \psi(K, L) \quad (56)$$

where K is positive semidefinite and of the same dimensions as J. For $q_1, q_2 \geq 0$ and a positive integer d, define

$$f_d(q_1, q_2) :=$$
$$\sup_{\mathrm{tr}(J) \leq q_1, \, \mathrm{tr}(L) \leq q_2} \{\ln \det(J + N_1 I + uI + L) + \phi(J, L)\}, \quad (57)$$

where the supremum is over positive semidefinite $d \times d$ matrices J and L satisfying the indicated constraints. For $q_1, q_2 \geq 0$ define

$$g_d(q_1, q_2) = \sup_{\mathbb{E}[Q_1] \leq q_1, \, \mathbb{E}[Q_2] \leq q_2} \mathbb{E}[f_d(Q_1, Q_2)] \quad (58)$$

where the supremum is over nonnegative (possibly dependent) random variables Q_1 and Q_2 satisfying the indicated bounds. Obviously $f_d(q_1, q_2) \leq g_d(q_1, q_2)$, and the inequality may be strict for some (q_1, q_2).

A. One Dimensional Case

We first consider the case where the dimension $d = 1$.

Lemma 3: Suppose that $J, K, L > 0$, and $f_1(J, L) = g_1(J, L)$. Assume that K achieves the supremum in the definition of $\phi(J, L)$. Then $K + N_1 \leq 1 + \sqrt{1 + u}$.

Proof: We will assume, without loss of generality, that $N_1 = 0$, since otherwise either $K = 0$ or we can make the substitution $K \leftarrow K + N_1$ and the rest of the proof will carry through.

From the definition of ψ we can verify that given L and N_1,

$$\arg\max_{K \geq 0} \psi(K, L) = \begin{cases} \infty & L \leq 1; \\ \frac{u + L}{L - 1} & L > 1. \end{cases} \quad (59)$$

Therefore $K := \arg\max_{0 \leq K \leq J} \psi(K, L) = J \wedge \frac{u + L}{L - 1}$ if $L > 1$ and $K = J$ otherwise. Now consider $J_t := J + t$, $L_t := L - t$, where $t \in \mathbb{R}$.

Case 1: $L > 1$ and $J > \frac{u + L}{L - 1}$. In this case $L_t > 1$ and $J_t > \frac{u + L_t}{L_t - 1}$ for small enough t, so that

$$0 \geq \partial_t^2 f_1(J_t, L_t)|_{t=0} \quad (60)$$
$$= \partial_t^2 \phi(J_t, L_t)|_{t=0} \quad (61)$$
$$= \partial_t^2 \psi(\tfrac{u + L_t}{L_t - 1}, L_t)|_{t=0} \quad (62)$$

where the first inequality is implied by $f_1(J, L) = g_1(J, L)$. Since

$$\psi\left(\frac{u + L}{L - 1}, L\right) = (u + 1) \ln(u + L) - \ln L - (u + 1) \ln(u + 1), \quad (63)$$

by taking the second derivative we see that $L \geq \sqrt{u + 1} + 1$ and hence $K = \frac{u + L}{L - 1} \leq \sqrt{u + 1} + 1$.

Case 2: either $L \leq 1$, or $L > 1$ and $J < \frac{u + L}{L - 1}$. In this case, for small enough t,

$$0 \geq \partial_t^2 f(J_t, L_t)|_{t=0} \quad (64)$$
$$= \partial_t^2 \phi(J_t, L_t)|_{t=0} \quad (65)$$
$$= \partial_t^2 \psi(J_t, L_t)|_{t=0} \quad (66)$$

where the first inequality is implied by $f_1(J, L) = g_1(J, L)$. By computing the second derivative in (66) we see that $-\frac{1}{J^2} + \frac{1+u}{(J+u)^2} \leq 0$ and hence $K = J \leq 1 + \sqrt{u + 1}$.

Case 3: $L > 1$ and $J = \frac{u + L}{L - 1}$. In this case, let us assume (for proof by contradiction) that $K > 1 + \sqrt{u + 1}$. Then $K = J = \frac{u + L}{L - 1}$ implies that $L < 1 + \sqrt{u + 1}$. Since $\frac{d}{dl}\left(\frac{u+l}{l-1}\right)\Big|_{l=L} < \frac{d}{dl}\left(\frac{u+l}{l-1}\right)\Big|_{l=1+\sqrt{u+1}} = -1$, we see that $J_t \leq \frac{u + L_t}{L_t - 1}$ for $t \geq 0$ and $J_t > \frac{u + L_t}{L_t - 1}$ for $t < 0$ (for t in some neighborhood of 0). We argue that $\partial_t f_1(J_t, L_t)$ exists and is continuous at $t = 0$. Indeed,

$$\partial_t f_1(J_t, L_t)|_{t=0^+} = \partial_t \phi(J_t, L_t)|_{t=0^+} \quad (67)$$
$$= \partial_t \psi(J_t, L_t)|_{t=0} \quad (68)$$
$$= \partial_1 \psi(K, L) \partial_t J_t|_{t=0} + \partial_2 \psi(K, L) \partial_t L_t|_{t=0} \quad (69)$$
$$= \partial_2 \psi(K, L) \partial_t L_t|_{t=0} \quad (70)$$

where the last step used the fact that $K = \arg\max \psi(\cdot, L)$. Moreover,

$$\partial_t f_1(J_t, L_t)|_{t=0^-}$$
$$= \partial_t \phi(J_t, L_t)|_{t=0^-} \quad (71)$$
$$= \partial_t \psi\left(\frac{L_t + u}{L_t - 1}, L_t\right)\Big|_{t=0} \quad (72)$$
$$= \partial_1 \psi(K, L) \partial_t \left(\frac{L_t + u}{L_t - 1}\right)\Big|_{t=0} + \partial_2 \psi(K, L) \partial_t L_t|_{t=0} \quad (73)$$
$$= \partial_2 \psi(K, L) \partial_t L_t|_{t=0} \quad (74)$$

Therefore $\partial_t f_1(J_t, L_t)|_{t=0^+} = \partial_t f_1(J_t, L_t)|_{t=0^-}$. The assumption $f_1(J, L) = g_1(J, L)$ implies that $\partial_t^2 f_1(J_t, L_t)|_{t=0^+} \leq 0$, and by a similar calculation as Case 2, we have $K \leq 1 + \sqrt{u + 1}$, a contradiction. Thus we must have $K \leq 1 + \sqrt{u + 1}$. $\square$

B. Vector Case

The goal of this subsection is to extend Lemma 3 to the vector case and show that *any* optimal covariance matrix K, if nonzero, must have top eigenvalue upper bounded by $1 + \sqrt{1 + u} - N_1$. This essentially relies on the Gaussian tensorization property [40, Theorem 2], which says that the Gaussian HK region is exhausted by random vectors with diagonal covariance matrices. Note, however, that [40, Theorem 2] only states that there is *one* extremal solution in which the covariance matrices commute; can there be some other extremal solution in which the covariance matrices do not commute, which can be perturbed to show Gaussian suboptimality? We shall carefully revisit the steps of [40] and show it is not possible. Propositions 2 and 3 are basic facts from linear algebra, whose short proofs can be found in [40] or the arXiv version of the present paper.

Definition 1: Let K be a positive semidefinite matrix. We say a diagonal matrix $\bar{K}$ is a decreasing (resp. increasing) alignment of K if $\bar{K} = Q^\top K Q$ for some orthogonal Q and the diagonal entries of $\bar{K}$ are decreasing (resp. increasing) from top left to bottom right.

Proposition 2: Suppose that K and L are positive semidefinite matrices of the same dimensions, and let $\bar{K}$ and $\bar{L}$ be their decreasing and increasing alignments, respectively. Then $\ln\det(K + L) \leq \ln\det(\bar{K} + \bar{L})$, equality holding only if $\bar{K} = Q^\top K Q$ and $\bar{L} = Q^\top L Q$ for some orthogonal Q.

Proposition 3: Suppose that K and K are positive semidefinite matrices of the same dimensions and satisfying $K \preceq K$. Let $\bar{K}$ and $\bar{K}'$ be their decreasing alignments. Then $\bar{K}' \preceq \bar{K}$.

Proposition 4: Suppose that (J, L) achieves the supremum in the definition of $f_d(q_1, q_2)$ (for some $q_1, q_2 > 0$), and K achieves the supremum in the definition of $\phi(J, L)$. Let $\bar{K}$ and $\bar{J}$ be the decreasing alignments of K and J, and $\bar{L}$ be the increasing alignment of L. Then $(\bar{J}, \bar{L})$ achieves the supremum in the definition of $f_d(q_1, q_2)$, and $\bar{K}$ achieves the supremum in the definition of $\phi(\bar{J}, \bar{L})$.

Proof: From Proposition 3 and Proposition 2 we see that

$$\phi(\bar{J}, \bar{L}) \geq \psi(\bar{K}, \bar{L}) \geq \phi(J, L). \tag{75}$$

Using Proposition 2 again we see that $\ln\det(\bar{J} + N_1 I + uI + \bar{L}) + \phi(\bar{J}, \bar{L}) \geq \ln\det(J + N_1 I + uI + L) + \phi(J, L)$ and therefore $(\bar{J}, \bar{L})$ must achieve the supremum in the definition of $f_d(q_1, q_2)$. Next we reverse the signs of inequalities in (75): Since (J, L) achieves the supremum in the definition of $f_d(q_1, q_2)$, we have

$$\ln(J + N_1 I + uI + L) + \phi(J, L)$$
$$= f_d(q_1, q_2) \tag{76}$$
$$\geq \ln(\bar{J} + N_1 I + uI + \bar{L}) + \phi(\bar{J}, \bar{L}) \tag{77}$$

which combined with $\ln(J + N_1 I + uI + L) \leq \ln(\bar{J} + N_1 I + uI + \bar{L})$ shows that $\phi(J, L) \geq \phi(\bar{J}, \bar{L})$. This implies that equalities are achieved in (75), hence $\bar{K}$ achieves the supremum in the definition of $\phi(\bar{J}, \bar{L})$. $\qquad\square$

One consequence of Proposition 4 is the following tensorization property:

Corollary 2 [40]: Given any positive integer d and $q_1, q_2 \in (0, \infty)$, we have $g_d(dq_1, dq_2) = dg_1(q_1, q_2)$.

Remark 4: $f_d(dq_1, dq_2) = df_1(q_1, q_2)$ is not true in general.

Remark 5: Proposition 4 goes slightly further than [40]: we not only show that the value of the supremum tensorizes, but also that *all* the maximizers (the matrices achieving the supremum) must tensorize, which will be used in our Theorem 4.

Theorem 4: Suppose that (q_1, q_2) satisfies $f_d(q_1, q_2) = g_d(q_1, q_2)$, (J, L) achieves the supremum in the definition of $f_d(q_1, q_2)$, and K achieves the supremum in the definition of $\phi(J, L)$. If K is nonzero then its eigenvalues are all upper bounded by $1 + \sqrt{1 + u} - N_1$.

Proof: Let $\bar{J}$ and $\bar{K}$ be the decreasing alignments of J and K, respectively, and let $\bar{L}$ be the increasing alignment of L. We have

$$\sum_{i=1}^d g_1(\bar{J}_{ii}, \bar{L}_{ii}) \leq g_d(q_1, q_2) \tag{78}$$

$$= f_d(q_1, q_2) \tag{79}$$
$$\leq \sum_{i=1}^d f_1(\bar{J}_{ii}, \bar{L}_{ii}) \tag{80}$$
$$\leq \sum_{i=1}^d g_1(\bar{J}_{ii}, \bar{L}_{ii}) \tag{81}$$

where (78) can be seen from the definition of g_d; (79) is by the assumption of the theorem; (80) follows since Proposition 4 shows that $(\bar{J}, \bar{L})$ achieves the supremum in the definition of $f_d(q_1, q_2)$, and $\bar{K}$ achieves the supremum in the definition of $\phi(\bar{J}, \bar{L})$. Then equalities are therefore achieved throughout, and since $f_1 \leq g_1$ we must have $f_1(\bar{J}_{11}, \bar{L}_{11}) = g_1(\bar{J}_{11}, \bar{L}_{11})$. Since $\bar{K}$ achieves the supremum in the definition of $\phi(\bar{J}, \bar{L})$, $\bar{K}_{11}$ must achieve the supremum in the definition of $\phi(\bar{J}_{11}, \bar{L}_{11})$. Then by Lemma 3 we have $\bar{K}_{11} + N_1 \leq 1 + \sqrt{1 + u}$ unless $\bar{K}_{11} = 0$. $\qquad\square$

V. Local Gaussian Optimality

Results in the previous sections show that our counterexamples are not sufficient for establishing Gaussian suboptimality for the Han-Kobayashi (HK) region allowing power control. Note that the counterexample in Corollary 1 relies on perturbation along a geodesic line in $L^2(\mathbb{R})$ and uses the order-3 Hermite polynomial; one might wonder if the result can be improved by perturbing along a more sophisticated "curve" and using lower order Hermite polynomials. This is not possible, as the Hessian calculations already suggest. However, since the Hessian is not necessarily "continuous" with respect to the same metric it is calculated with, semidefiniteness of the Hessian at a single point does not provide a rigorous proof of local optimality.

In this section, we show that in the (interior of the) regime where the Hessian at the Gaussian stationary point is negative definite, local maximum is indeed achieved. The proof is based on showing that the Hessian is negative semidefinite in an $L^\infty(\gamma_K)$ neighborhood.

The $L^\infty(\gamma_K)$ metric appears to be natural for this setting; the celebrated Holley-Stroock perturbation principle (see e.g. [29, p1184] [30], [44]) provides a method of controlling the best constants in functional inequalities such as the Poincare inequality (also concerning a bound on the spectrum of a self-adjoint operator), under perturbation in the L^∞-norm. This method is simple yet avoids assumption of bounds on higher derivatives. Our argument may fail if other metrics are used (see Remark 6).

Recall the functional Ψ defined in (24), which played a role in the expression of the Han-Kobayashi region. With an abuse of notation, in this section we shall define the following functional:

$$\Psi(P_X, L) := \sup_{P_Y: \, \mathrm{cov}(P_Y) \preceq L} uh(P_X * \gamma_{uI} * P_Y) + h(P_X)$$
$$- (1 + u)h(P_X * \gamma_{uI}), \tag{82}$$

where $u > 0$ and positive integer d are given, P_X is a distribution on $\mathbb{R}^d$, and L is a $d \times d$ positive semidefinite matrix. Here we have taken $N_1 = 0$, which is without loss of

generality for the purpose proving local optimality, since the general case easily follows by taking $X_1 + Z_1$ as the new X_1. Also, we have taken the covariance of Z_2 to be uI, which is without loss of generality by a scaling argument.

We define the following as a simpler proxy for $\Psi(P_X, L)$:

$$\Theta(P_X, L) := \frac{u}{2} \ln[(2\pi e)^d \det(\text{cov}(P_X) + uI + L)] + h(P_X) - (1+u)h(P_X * \gamma_{uI}). \quad (83)$$

Note that $\Psi(P_X, L) \leq \Theta(P_X, L)$, with equality achieved only if P_X is Gaussian (the only part follows from the Levy-Cramer theorem [16]). Define

$$\Phi(L) := \sup_K \Theta(\gamma_K, L) \quad (84)$$

where the supremum is over all positive semidefinite matrices K with the same dimensions as L. Observe that the supremum in (84) can be achieved if and only if the least eigenvalue of L is strictly larger than 1; in that case the maximizer is

$$K = (L - I)^{-1}(L + uI). \quad (85)$$

Moreover, the Hessian at the maximizer K is strictly negative definite: Indeed, by rotation invariance of the log determinant function, we can assume without loss of generality that L agrees with its increasing alignment (Definition 1), in which case K agrees with its decreasing alignment. For Δ of the same dimensions as K and whose Frobenius norm $\|\Delta\|$ is sufficiently small, we have

$$\Theta(\gamma_{K+\Delta}, L) - \Theta(\gamma_K, L) \leq \Theta(\gamma_{K+\tilde{\Delta}}, L) - \Theta(\gamma_K, L) \quad (86)$$

$$\leq -\sum_{i=1}^d a_i \tilde{\Delta}_{ii}^2 + o(\|\Delta\|^2) \quad (87)$$

$$\leq -\min_{1 \leq i \leq d} a_i \|\Delta\|^2 + o(\|\Delta\|^2) \quad (88)$$

where $\tilde{\Delta}$ is a diagonal matrix whose diagonal entries are some permutation of the eigenvalues of Δ, (86) follows from Proposition 2, and $a_i > 0$ is a function of L_{ii}. Thus strict negative definiteness of the Hessian is proved.

The main result of this section is the following:

Theorem 5: Given $u > 0$ and a positive semidefinite matrix L, suppose that K is a maximizer in (84). Suppose that the top eigenvalue of K is strictly smaller than $\frac{u}{(1+u)^{1/3}-1}$. Then exist $\epsilon > 0$ such that for all P_X satisfying $\|P_X - \gamma_K\|_{L^\infty(\gamma_K)} \leq \epsilon$ and $\int x dP_X(x) = 0$, we have $\Theta(P_X, L) \leq \Phi(L)$, and consequently $\Psi(P_X, L) \leq \Phi(L)$.

The result of Theorem 5 can be extended (with the same $\frac{u}{(1+u)^{1/3}-1}$ bound on the top eigenvalue) to the case where the supremum in (84) is restricted to K satisfying $K \preceq J$, where J is a positive semidefinite matrix that commutes with L; we omit the details of the analysis. Together with Theorem 4, Theorem 5 shows that if Gaussian distributions (P_{UX_1}, P_{X_2}) is a Gaussian stationary point for the supremum in (30), then the expression to the right of the supremum in (25) cannot be improved by local (in the sense described by the Theorem 5) perturbation.

Remark 6: The Theorem may fail if $L^\infty(\gamma_K)$ is replaced by other metrics. For example, take $d = 1$ and consider the

Gaussian mixture $P_X^\epsilon = (1 - \epsilon)\gamma_K + \epsilon \mathcal{N}(0, \epsilon^{-2}K)$. Then $\lim_{\epsilon \downarrow 0} \|P_X^\epsilon - \gamma_K\|_{L^\infty(\mathbb{R})} = 0$ but $\lim_{\epsilon \downarrow 0} \|P_X^\epsilon - \gamma_K\|_{L^\infty(\gamma_K)} = \infty$, where $L^\infty(\mathbb{R})$ denotes the L^∞-norm with respect to the Lebesgue measure. Using the chain rule of entropy we can show that $|h(P_X^\epsilon) - h(P_X)| = O(\epsilon \ln \frac{1}{\epsilon})$ and $|h(P_X^\epsilon * \gamma_u) - h(P_X * \gamma_u)| = O(\epsilon \ln \frac{1}{\epsilon})$, but $\text{var}(P_X^\epsilon) = \text{var}(P_X) + \Theta(\epsilon^{-1})$. Therefore $\Theta(P_X^\epsilon, L) > \Phi(L)$ for small ϵ. We remark that a similar counterexample has been previously proposed for the stability of the log-Sobolev inequality [21]. In fact, (83) is closely related to the stability of log-Sobolev inequality, since given P_X,

$$\lim_{u \to \infty} \frac{1}{u}[h(P_X) - (1+u)h(P_X * \gamma_{uI})] = -\frac{1}{2}J(P_X) - h(P_X) \quad (89)$$

where $J(P_X)$ denotes the Fisher information (see e.g. [14]).

Proof: [Proof of Theorem 5] Pick an arbitrary r such that $\|r\|_{L^\infty(\gamma_K)} \leq \epsilon$ ($\epsilon < 1$ to be chosen later), $\int r = 0$, and $\int xr = 0$. Define $p_t := \gamma_K + tr$ and $q_t := p_t * \gamma_{uI}$, where $t \in [0, 1]$. By assumption K is the maximizer in (85) and hence a stationary point, calculations show that $\frac{d}{dt}\Theta(p_t, L)|_{t=0} = 0$ (even though p_t is not necessarily Gaussian). Thus to prove the theorem it remains show the negativity of the second derivative for $t \in [0, 1]$. We have

$$\frac{d}{dt}\text{cov}(p_t) = \frac{d}{dt}\int xx^\top p_t \quad (90)$$

$$= \int xx^\top r \quad (91)$$

and $\frac{d^2}{dt^2}\text{cov}(p_t) = -2\int xr \cdot \int x^\top r$. Therefore

$$\frac{d}{dt}\ln\det(\text{cov}(p_t) + uI + L) = \text{tr}\left((\text{cov}(p_t) + uI + L)^{-1}\frac{d}{dt}\text{cov}(p_t)\right) \quad (92)$$

and

$$\frac{d^2}{dt^2}\ln\det(\text{cov}(p_t) + uI + L) = -\text{tr}\left((\text{cov}(p_t) + uI + L)^{-1}\frac{d}{dt}\text{cov}(p_t)(\text{cov}(p_t) + uI + L)^{-1}\frac{d}{dt}\text{cov}(p_t)\right) \quad (93)$$

$$\leq -\frac{1}{(1+t\epsilon)^2}\text{tr}\left((K + uI + L)^{-1}\int xx^\top r(K + uI + L)^{-1}\int xx^\top r\right) \quad (94)$$

where we used the fact that $\text{cov}(p_t) = \int xx^\top p_t \leq \int xx^\top(1 + t\epsilon)\gamma_K \leq (1 + t\epsilon)K$. Moreover, from (38) we obtain

$$\frac{d^2}{dt^2}h(p_t) = -\int \frac{r^2}{p_t} \quad (95)$$

$$\leq -\frac{1}{1+t\epsilon}\int \frac{r^2}{\gamma_K} \quad (96)$$

and similarly

$$\frac{d^2}{dt^2}h(q_t) = -\int \frac{(r * \gamma_{uI})^2}{p_t * \gamma_{uI}} \quad (97)$$

$$\geq -\frac{1}{1-t\epsilon}\int \frac{(r * \gamma_{uI})^2}{\gamma_{K+uI}} \quad (98)$$

where we used $(1 - t\epsilon)\gamma_K(x) \le p_t(x) \le (1 + t\epsilon)\gamma_K(x)$, for all $x \in \mathbb{R}^d$.

In the rest of the proof assume without loss of generality that K is diagonal with $k_1, \ldots, k_d$ being the diagonal entries. Consider the expansion $r = \sum_{\alpha \in \{0,1,\ldots\}^d} \lambda_\alpha D^\alpha \gamma_K$, and plug it into the right sides of (94),(96),(98). It is easy to see using (53) that we have

$$\frac{d^2}{dt^2}\Theta(p_t, L) \le \sum_\alpha I_\alpha - \frac{1}{2(1+t\epsilon)^2} \operatorname{tr}\left((K+L)^{-1}\Lambda(K+L)^{-1}\Lambda\right),$$
(99)

where the notations above are explained as follows: $\Lambda := [(1 + \delta_{ij})\lambda_{e_i+e_j}]_{i,j}$, with $e_i \in \{0, 1, \ldots\}^d$ being the vector with only the i-th coordinate equal to 1 and the other $d - 1$ coordinates equal to 0; moreover for each multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \{0, 1, \ldots\}^d$ we defined

$$I_\alpha = -\frac{1}{1+t\epsilon}\lambda_\alpha^2 \prod_{i=1}^d \frac{\alpha_i!}{k_i^{\alpha_i}} + \frac{1+u}{1-t\epsilon}\lambda_\alpha^2 \prod_{i=1}^d \frac{\alpha_i!}{(k_i+u)^{\alpha_i}}. \quad (100)$$

For each multi-index α (viewed as a vector in $\mathbb{R}^d$) we define its norm $|\alpha|$ as the sum of its coordinates. Then since $\int r = 0$ and $\int xr = 0$, for $|\alpha| \le 1$ we must have $\lambda_\alpha = 0$ and hence $I_\alpha = 0$. Now we claim that if ϵ is chosen small enough (the choice depending only on u, K, L and not on (λ_α)), we have

$$\sum_{\alpha:\, |\alpha|=2} I_\alpha - \frac{1}{2(1+t\epsilon)^2} \operatorname{tr}\left((K+L)^{-1}\Lambda(K+L)^{-1}\Lambda\right) \le 0$$
(101)

and

$$I_\alpha \le 0, \quad \forall\alpha\colon |\alpha| \ge 3. \quad (102)$$

Clearly these claims would establish $\frac{d^2}{dt^2}\Theta(p_t, L) \le 0$ for $t \in [0, 1]$ and hence the conclusion of the theorem.

To see (101), it suffices to show that

$$\frac{(1+\epsilon)^2}{1-\epsilon} \le$$
$$\inf_{(\lambda_\alpha)} \frac{\sum_{|\alpha|=2}\lambda_\alpha^2 \prod_{i=1}^d \frac{\alpha_i!}{k_i^{\alpha_i}} + \frac{1}{2}\operatorname{tr}((K+L)^{-1}\Lambda(K+L)^{-1}\Lambda)}{(1+u)\sum_{|\alpha|=2}\lambda_\alpha^2 \prod_{i=1}^d \frac{\alpha_i!}{(k_i+u)^{\alpha_i}}}$$
(103)

where the infimum is over all $(\lambda_\alpha)_{\alpha:\, |\alpha|=2}$ such that the denominator in (103) is positive. Since both the numerator and the denominator in (103) are 2-homogeneous in $(\lambda_\alpha)_{\alpha:\, |\alpha|=2}$ (note that Λ is also a function of $(\lambda_\alpha)_{\alpha:\, |\alpha|=2}$), the infimum can be restricted to a compact set, and hence it can be achieved by some $(\lambda_\alpha)_{\alpha:\, |\alpha|=2}$. By simple calculations we can show that the strict negativity of the Hessian in the Gaussian optimization problem (103) is in fact equivalent to the right side of the infimum in (103) being strictly larger than 1 for each (λ_α) (to see this, consider a perturbation of K in the direction of Λ and compute the second derivatives). Therefore (103) and hence (101) is true for all $\epsilon \le \epsilon_1$ where $\epsilon_1 > 0$ is some constant depending only on u, K, and L.

To see (102), it suffices to show that

$$\frac{1-\epsilon}{1+\epsilon} \ge (1+u)\prod_{i=1}^d \left(\frac{k_i}{k_i+u}\right)^{\alpha_i}, \quad \forall\alpha\colon |\alpha|=3. \quad (104)$$

The assumption of $k_i < \frac{u}{(1+u)^{1/3}-1}$ implies that the right side of (104) is strictly less than 1, and hence (104) holds for all $\epsilon \le \epsilon_2$ where ϵ_2 depends only on u, K, L. The claim of the theorem then follows by taking $\epsilon = \min\{\epsilon_1, \epsilon_2\}$. $\square$

VI. A GEOMETRIC INEQUALITY ANALOGOUS TO LEMMA 2

In this section we discuss an analogous convex geometric inequality with a non-isotropic extremal, which may be of independent interest. The similarities between entropic and geometric inequalities are well-known (e.g. [4], [19]) and sometimes can be employed to prove new results in network information theory [34]. Intuitively, differential entropy can be seen as analogous to the logarithmic volume of a set, and the sum of random variables is analogous to the Minkowski sum of sets. Therefore, we may consider an analogue of the inequality in Conjecture 1 by replacing the entropy of independent sum with the log volume of the Minkowski sum.

Recall that the meanwidth of a convex body $C \subseteq \mathbb{R}^d$ is defined by

$$W(C) := \mathbb{E}[\sup_{u \in C} \langle u, X \rangle - \inf_{u \in C} \langle u, X \rangle] \quad (105)$$

where X has a uniform distribution on the unit sphere in $\mathbb{R}^d$. We will replace the power constraint in the entropic inequality by a meanwidth constraint for a convex body.

Theorem 6: Let B be the ball of radius $1/2$ in $\mathbb{R}^d$ (so that its meanwidth equals 1). The maximum of

$$\frac{\operatorname{vol}^{1/2}(K+B+L)\operatorname{vol}^{1/2}(K)}{\operatorname{vol}(K+B)} \quad (106)$$

over all convex body K and all convex body L satisfying the meanwidth bound $W(L) \le 1$ is strictly larger than 1. On the other hand, if we restrict $L = B$, then the supremum of the same quantity equals 1.

Proof: We first prove the claim for $L = B$. Note that by the Brunn-Minkowski inequality ($\operatorname{vol}(\alpha A + \beta B) \ge \operatorname{vol}^\alpha(A)\operatorname{vol}^\beta(B)$), we see that (106) is bounded above by 1. Moreover, (106) approaches 1 if we take $K = tK$ where K is any convex body and $t \to \infty$.

Next, we show that the supremum can be strictly larger than 1 without the restriction that $L = B$. Let us consider the case of $k = 2$ for simplicity. (See Remark 7 for more general settings.) Let K be the unit cube (similar idea actually works for any convex body K). Let L be $\frac{\pi}{4}K$ rotated by $\pi/4$. Note that the meanwidths

$$W(L) = W(B) = 1. \quad (107)$$

However, for ϵ small we have

$$\operatorname{vol}(K+\epsilon L) = 1 + 4\epsilon \cdot \frac{\pi\sqrt{2}}{4} + O(\epsilon); \quad (108)$$
$$\operatorname{vol}(K+\epsilon B) = 1 + 4\epsilon + O(\epsilon^2). \quad (109)$$

Therefore $\operatorname{vol}(K+\epsilon L) > \operatorname{vol}(K+\epsilon B)$ for small enough $\epsilon > 0$. Now let $K = tK$ and let $t \to \infty$. We have

$$\frac{\operatorname{vol}(K+B)}{\operatorname{vol}(K)} = 1 + 4t^{-1} + O(t^{-2}); \quad (110)$$

$$\frac{\text{vol}(K + B + L)}{\text{vol}(K + B)} = 1 + \pi\sqrt{2}t^{-1} + O(t^{-2}). \tag{111}$$

Therefore (106) equals

$$1 + \frac{\pi\sqrt{2} - 4}{2}t^{-1} + O(t^{-2}) \tag{112}$$

which is positive for t large enough. □

Remark 7: More generally, consider the supremum of

$$\frac{\text{vol}^u(K + B + L)\,\text{vol}(K)}{\text{vol}^{1+u}(K + B)} \tag{113}$$

where $u > 0$, and suppose that the convex bodies are in $\mathbb{R}^d$. In the above argument we can consider K to be a cube of growing size and let L be the cross-polytope with diameter of order $\sqrt{\frac{d}{\ln d}}$ (so that the mean-width is order of a constant). Then

$$\frac{\text{vol}(K + B)}{\text{vol}(K)} = 1 + \Theta(dt^{-1}); \tag{114}$$

$$\frac{\text{vol}(K + B + L)}{\text{vol}(K + B)} \approx \frac{\text{vol}(K + L)}{\text{vol}(K)} \tag{115}$$

$$= 1 + \Theta\left(dt^{-1}\sqrt{\frac{d}{\ln d}}\right). \tag{116}$$

Therefore the supremum is positive if $dt^{-1} << udt^{-1}\sqrt{\frac{d}{\ln d}}$, or equivalently $u >> \sqrt{\frac{\ln d}{d}}$. On the other hand, the supremum over ball L is 0 iff $u \le 1$. Therefore in the regime of u between 1 and $\omega(\sqrt{\frac{\ln d}{d}})$ the supremum is not achieved by balls.

VII. DISCUSSION

It seems that simple amendments of Conjecture 1, such as taking the concave envelope in X_2 or restricting to parameters in the stability regime, no longer fulfill its original purpose of implying Gaussian optimality for the HK region. However, it is still possible to imply Gaussian optimality for some parts of the region. For example, the special case of Conjecture 1 in [5] concerning Costa's corner point still appears to be a valid approach towards the slope of the corner point. Here we slightly generalize the observation therein to a larger part of the rate region.

Theorem 7: Suppose that $N_1, N_2, u, q_1, q_2 \in (0, \infty)$ are such that the maximum of

$$uh(X_1 + X_2 + Z_1 + Z_2) + h(X_1 + Z_1) - (1+u)h(X_1 + Z_1 + Z_2) \tag{117}$$

over independent random variables $X_1, X_2 \in \mathbb{R}^d$, $\mathbb{E}[\|X_i\|_2^2] \le dq_i$, $i = 1, 2$ is achieved by isotropic Gaussian distributions (possibly degenerate with zero covariance), for any positive integer d. Then the HK bound (30) with Gaussian inputs is tight for the weighted sum rate $R_1 + (1+u)R_2$.

Proof: The proof is essentially the same as [5, Lemma 4]; we can split the optimization problem into two, which achieve the supremum simultaneously:

$$R_1 + (1+u)R_2$$
$$\le \lim_{d \to \infty} \frac{1}{d} \sup\{I(X_1; Y_1) + (1+u)I(X_2; Y_2)\} \tag{118}$$

$$\le \lim_{d \to \infty} \frac{1}{d}\{\sup h(Y_2) + \sup[uh(Y_2) - (1+u)h(Y_2|X_2) + h(Y_1)]$$
$$- h(Y_1|X_1)\} \tag{119}$$

$$\le \frac{1}{2}\ln\frac{q_1 + q_2 + N_1 + N_2}{N_1} + \frac{1}{2}\sup_{K \le q_1}\psi(K, q_2) \tag{120}$$

where the suprema are over independent $X_1, X_2 \in \mathbb{R}^d$ satisfying $\mathbb{E}[\|X_i\|_2^2] \le dq_i$, $i = 1, 2$; Y_1 and Y_2 are defined in (1)-(2); $\psi(K, q_2)$ was defined in (55) (with dimension one); (118) follows by Fano's inequality; (120) follows by the assumption of Theorem 7. Note that (120) is upper bounded by the HK bound (30) (in dimension one with Gaussian inputs), therefore the claim of the theorem follows. □

The set of (N_1, N_2, u, q_1, q_2) satisfying the assumption in Theorem 7 always falls in the setting of Lemma 3, and the latter always falls in the regime of stable Gaussian stationary points. The set is not empty, as the sufficient condition in [26] shows. A more precise characterization of this set remains an open question. In particular, a concrete new conjecture can be formulated as follows:

Conjecture 2: The set of (N_1, N_2, u, q_1, q_2) satisfying the isotropic condition in Theorem 7 equals the set of (N_1, N_2, u, q_1, q_2) for which 1-letter HKGS without power control matches 1-letter HKGS with power control.

In Conjecture 2, 1-letter HKGS means the Han-Kobayashi bound (30) in dimension one with Gaussian signaling, and with/without power control means (Q_1, Q_2) in (30) is random/constant. The first mentioned set in Conjecture 2 is contained in the second, by the proof of Theorem 7. Proving Conjecture 2 would prove the conjecture about the slope at Costa's point in [5]. On the other hand, disproving Conjecture 2 would imply that HKGS with variable power control is suboptimal for parts of the capacity region. Indeed, suppose that HKGS is optimal, then by the Gaussian tensorization the best sum rate is achieved by 1-letter HKGS. If the conjecture is false then there is some (N_1, N_2, u, q_1, q_2) in the second set but not the first set in Theorem 7, and the best sum rate for this (N_1, N_2, u, q_1, q_2) is achieved by 1-letter HKGS without power control due to the membership in the second set. But this implies the membership in the first set as well, a contradiction.

VIII. ACKNOWLEDGMENT

APPENDIX A
PROOF OF LEMMA 1

Using the Taylor expansion, we have

$$p_t(y) = \int p(\sqrt{t}u + y)q(u)du \tag{121}$$

$$= \int \left[p(y) + \sqrt{t}up'(y) + \frac{1}{2}tu^2p''(y)\right.$$
$$\left. + \frac{1}{6}t^{3/2}u^3p'''(y) + \frac{1}{24}t^2u^4p''''(\xi_{y,u})\right]q(u)du \tag{122}$$

where $\xi_{y,u}$ is a value between y and $y + \sqrt{t}u$. Therefore for any $t < 1$,

$$\left| p_t(y) - p(y) - \frac{1}{2}tm_2 p''(y) - \frac{1}{6}t^{3/2}m_3 p'''(y) \right|$$

$$= \frac{1}{24}t^2 \left| \int u^4 p''''(\xi_{y,u}) q(u) du \right|. \tag{123}$$

$$\leq \frac{t^2}{24} p(y) \int u^4 \left(\sup_{v:\,|v-y| \leq \sqrt{t}|u|} \frac{|p''''(v)|}{p(y)} \right) q(u) du \tag{124}$$

$$\leq C t^2 p(y) \int u^4 e^{C\sqrt{t}|u|} q(u) du \tag{125}$$

$$\leq C^2 t^2 p(y) \int u^4 e^{C|u| - cu^2} du \tag{126}$$

$$\leq c_1 t^2 p(y) \tag{127}$$

where here and below $c_1, c_2, \ldots$ are positive constants depending on p and q but not on t. To see (125), note that (7) implies for $|(\ln p(x))'| < C$ and hence

$$\sup_{v:\,|v-y| \leq \sqrt{t}|u|} \frac{|p''''(v)|}{p(y)} \leq \sup_{v \in \mathbb{R}} \frac{|p''''(v)|}{p(v)} \cdot \sup_{v:\,|v-y| \leq \sqrt{t}|u|} \frac{p(v)}{p(y)} \tag{128}$$

$$\leq C e^{C\sqrt{t}|u|}. \tag{129}$$

Next define $\Delta_t(y) := p_t(y) - p(y)$. By the Taylor expansion of the function $x \mapsto x \ln x$ around $p(y)$, we have

$$|p_t(y) \ln p_t(y) - p(y) \ln p(y) - \Delta_t(y) - \Delta_t(y) \ln p(y)|$$

$$\leq \frac{\Delta_t^2(t)}{2\xi} \tag{130}$$

$$\leq \frac{1}{2} \cdot \frac{\Delta_t^2(y)}{p_t(y) \wedge p(y)} \tag{131}$$

where ξ denotes a number between $p(y)$ and $p_t(y)$ (from the Lagrange remainder) and $\wedge$ denotes the minimum value. Noting $\int \Delta_t = 0$, we have

$$\left| \int p_t \ln p_t - \int p \ln p - \int \Delta_t \ln p \right| \leq \int \frac{\Delta_t^2}{p_t \wedge p}. \tag{132}$$

By (127), for $t < 1$ we have

$$|\Delta_t(t)| \leq \frac{1}{2}tm_2 |p''(y)| + \frac{1}{6}t^{3/2}m_3 |p'''(y)| + c_1 t^2 p(y)$$

$$\leq c_2 t p(y). \tag{133}$$

Thus for $t < \frac{1}{2c_2} \wedge 1$ we have $\frac{t^{-2}\Delta_t^2(y)}{p_t(y) \wedge p(y)} \leq \frac{t^{-2}\Delta_t^2(y)}{\frac{1}{2}p(y)} \leq 2c_2^2 p(y)$ and

$$\int \frac{\Delta_t^2}{p_t \wedge p} \leq 2c_2^2 t^2. \tag{134}$$

Moreover, by multiplying $\ln p(y)$ to (127) and integrating, we find that for $t < \frac{1}{2c_2} \wedge 1$,

$$\left| \int \Delta_t \ln p - \frac{1}{2}tm_2 \int p'' \ln p - \frac{1}{6}t^{3/2}m_3 \int p''' \ln p \right| \leq c_3 t^2. \tag{135}$$

Comparing (132), (134) and (135) we establish (8).

APPENDIX B
GAUSSIAN SUBOPTIMALITY VIA WASSERSTEIN FLOW

In this section we provide another approach of establishing Gaussian suboptimality by showing that the Hessian under the Wasserstein-2 metric at the stationary point fails to be negative semidefinite. The choice of metric is in fact immaterial to the stability of a stationary point, as we shall explain. Nevertheless, it seems to be an interesting direction for future research to leverage the recent developments on optimization in the Wasserstein space [2], [7], [15], [50]. Let us also remark that optimal transportation has previously been used in proving Costa's corner point for the interference channels [45]. Moreover, changing the metric can sometimes simplify entropic optimization problems; see e.g. the geodesic convexity of the Brascamp-Lieb inequality [48].

A. Preparations

It is useful to first recall how Hessian is related to stability at a stationary point in the simpler setting of optimization in the Euclidean space. Suppose that the goal is to maximize $f(x)$ over $x \in \mathbb{R}^d$ subject to the constraint $g(x) \leq 0$, where f and g are both smooth functions on $\mathbb{R}^d$. Then $x^* \in \mathbb{R}^d$ is a stationary point of the constrained optimization problem if and only if there exists $\lambda \geq 0$ such that $g(x^*) \leq 0$ and

$$\nabla f(x^*) - \lambda \nabla g(x^*) = 0 \in \mathbb{R}^d. \tag{136}$$

Moreover, a necessary condition for local optimality is that the restricted Hessian satisfies

$$\text{Hess}|_{\mathcal{C}}(f - \lambda \nabla g)(x^*) \preceq 0 \in \mathbb{R}^{(d-1) \times (d-1)} \tag{137}$$

where λ is as in (136), $\mathcal{C} = \{x \in \mathbb{R}^d : (x - x^*)^\top \nabla g(x^*) = 0\}$ is a codimension one subspace, and we assumed that $\nabla g(x^*) \neq 0$. Indeed, it suffices to consider the nontrivial case of $\lambda > 0$, and if (137) is not true, then there exists a smooth curve $(-1, 1) \to \mathcal{C}$, $t \mapsto \bar{x}(t)$ satisying $\bar{x}(0) = x^*$ and

$$(f - \lambda g)(\bar{x}(t)) \geq (f - \lambda g)(x^*) + at^2 + o(t^2) \tag{138}$$

$$\geq f(x^*) + at^2 + o(t^2) \tag{139}$$

for some $a > 0$. Assuming $\nabla g(x^*) \neq 0$, we can find a smooth curve $(-1, 1) \to \{x : g(x) = 0\}$, $t \mapsto x(t)$ satisfying $\|x(t) - \bar{x}(t)\| = O(t^2)$. Then since $\nabla(f - \lambda g)(x^*) = 0$ and $\|\bar{x}(t) - x^*\| = O(t)$, we have $\|\nabla(f - \lambda g)(x(t))\| = O(t)$ and

$$f(x(t)) = (f - \lambda g)(x(t)) \tag{140}$$

$$\geq (f - \lambda g)(\bar{x}(t)) + O(t^3) \tag{141}$$

$$\geq f(x^*) + at^2 + o(t^2) \tag{142}$$

implying that x^* is not a local maximum of the constrained optimization problem.

Let us review the basics about the formal Riemannian structure associated with the optimal transport distance; more background on this topic can be found in, e.g., [2], [15], and [44]. Let $\mathcal{P}_2(\mathbb{R}^d)$ be the set of absolutely continuous probability measures with finite second moments. The Wasserstein-2 distance induces a local inner product structure, which can be viewed as generalization of the standard (finite dimensional) metric tensor in the classical Riemannian geometry. More

precisely, at any point $P \in \mathcal{P}_2(\mathbb{R}^d)$, the tangent space is given by

$$\overline{\{\nabla\phi \colon \phi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(P)}, \qquad (143)$$

i.e., the $L^2(P)$-closure of the gradients of smooth, compactly supported functions.[2] Taking the closure ensures that the tangent space is complete with respect to the inner product defined below. Given a tangent vector $\nabla\phi$, a constant speed geodesic is the curve $t \mapsto P_t$ where P_t is the push-forward of the map $x \mapsto x + t\nabla\phi(x)$. Note that $\nabla\phi$ as a vector field on $\mathbb{R}^d$ is curl-free, which ensures that such a map is an optimal transport for small enough t. We define the following inner product on the tangent space at P:

$$\langle \nabla\phi, \nabla\psi \rangle := \int_{\mathbb{R}^d} \nabla\phi(x) \cdot \nabla\psi(x) dP(x), \qquad (144)$$

which is consistent with the Wasserstein-2 distance since $\|\nabla\phi\| := \sqrt{\langle \nabla\phi, \nabla\phi \rangle} = \frac{d}{dt}W_2(P_t, P)|_{t=0}$. Furthermore, given a function f on $\mathcal{P}_2(\mathbb{R}^d)$, its gradient and Hessian at P (if exist) satisfy

$$f(P_t) = f(P) + t \langle \operatorname{grad} f(P), \nabla\phi \rangle + \frac{t^2}{2} \operatorname{Hess} f(P)(\nabla\phi, \nabla\phi) + o(t^2) \qquad (145)$$

for any $\nabla\phi$, where $\operatorname{Hess} f(P)$ is a bilinear form which sends two tangent vectors to a real number.

Under a change of the inner product on the tangent space at P, the gradient at P changes linearly. For example, if $\frac{\delta f}{\delta P}$ denotes the gradient with respect to $L^2(\mathbb{R}^d)$, then the W_2-gradient is given by $\operatorname{grad} f(P) = \nabla\frac{\delta f}{\delta P}$. On the other hand, the Hessian depends on the second order behavior of the geodesic and therefore depends not only on the metric tensor at P but also on the connection (how a tangent vector is parallel transported in a neighborhood). As such, with a change of metric, the transformation of the Hessian depends on the connection and the gradient and therefore semidefiniteness is not preserved. However, at a stationary point, semidefiniteness does not depend on the choice of metric.

In the rest of the section we shall focus on the setting of $d = 1$, which is enough for constructing a counterexample to Conjecture 1. On $\mathbb{R}$, the vector field $\nabla\phi$ can be thought of as a smooth, compactly supported function, which we denote by $U(\cdot)$ or $V(\cdot)$. The constraint $\int U(x)dx = \int \nabla\phi(x)dx = 0$ can be dropped since we will take the closure in $L^2(P)$. We shall first derive a lemma about the W_2-gradient and Hessian of the differential entropy of convolution of measures $h(p*q*r)$, where p and q are viewed as variables whereas r is a fixed probability measure. The subscripts of grad and Hess denote the variables which are differentiated in (a formal definition can be given using a formula similar to (145)).

B. Hessian Calculation

Lemma 4: Let r be a fixed probability distribution on $\mathbb{R}$, and consider $h(p*q*r)$ as a functional of a pair of probability

distributions (p, q). Denote by $\mu := p * q * r$, viewed as a probability density function on $\mathbb{R}$. Let $Z = X + Y + R$, and let $U = U(X)$ and $V = V(Y)$ be two arbitrary smooth, compactly supported functions of X and Y, respectively. Then we have

$$\operatorname{grad}_p h(p*q*r) = -\mathbb{E}[\nabla \ln \mu(Z)|X]; \qquad (146)$$
$$\operatorname{grad}_q h(p*q*r) = -\mathbb{E}[\nabla \ln \mu(Z)|Y], \qquad (147)$$

and

$$\operatorname{Hess}_{pp} h(p*q*r)(U, U)$$
$$= -\mathbb{E}[U^2 \Delta \ln \mu(Z)] - \mathbb{E}\left[\left(\frac{\nabla(\mathbb{E}[U|Z]\mu(Z))}{\mu(Z)}\right)^2\right]; \quad (148)$$

$$\operatorname{Hess}_{qq} h(p*q*r)(V, V)$$
$$= -\mathbb{E}[V^2 \Delta \ln \mu(Z)] - \mathbb{E}\left[\left(\frac{\nabla(\mathbb{E}[V|Z]\mu(Z))}{\mu(Z)}\right)^2\right]; \quad (149)$$

$$\operatorname{Hess}_{pq} h(p*q*r)(U, V)$$
$$= -\mathbb{E}[UV \Delta \ln \mu(Z)] - \mathbb{E}\left[\frac{\nabla(\mathbb{E}[U|Z]\mu(Z))\nabla(\mathbb{E}[V|Z]\mu(Z))}{\mu^2(Z)}\right]. \quad (150)$$

Proof: Let X, Y, R, U, V be as in the statement of the lemma. Define for each $s \geq 0$, $t \geq 0$,

$$Z_{st} := X + sU + Y + tV + R = Z + sU + tV. \qquad (151)$$

Denote by μ_{st} the distribution of Z_{st}. We have

$$\frac{d}{ds} h(\mu_{st}) = -\frac{d}{ds} \int \mu_{st} \ln \mu_{st} \qquad (152)$$

$$= -\int \frac{d\mu_{st}}{ds} \ln \mu_{st} \qquad (153)$$

$$= \int \nabla(\mathbb{E}[U|Z_{st}]\mu_{st}) \ln \mu_{st} \qquad (154)$$

$$= -\int \mathbb{E}[U|Z_{st}]\nabla(\ln \mu_{st})\mu_{st} \qquad (155)$$

$$= -\mathbb{E}[U\nabla \ln \mu_{st}(Z_{st})] \qquad (156)$$

where (156) used

$$\frac{d}{ds}\mu_{st} = -\nabla(\mathbb{E}[U|Z_{st}]\mu_{st}) \qquad (157)$$

which can be shown by the following method:[3] Consider an arbitrary smooth test function f; we have

$$\int f\mu_{st} = \mathbb{E}[f(Z_{st})]. \qquad (158)$$

Differentiating on both sides,

$$\int f\frac{d}{ds}\mu_{st} = \mathbb{E}[f'(Z_{st})U] \qquad (159)$$

$$= \mathbb{E}[f'(Z_{st})\mathbb{E}[U|Z_{st}]] \qquad (160)$$

$$= -\int f(Z_{st})\nabla(\mathbb{E}[U|Z_{st}]\mu_{st}) \qquad (161)$$

where the last step used integration by parts. Since f is arbitrary we have confirmed (157) and hence (146) is proved.

---

[2]Here we adopt the notations of [8], where a tangent vector is identified with the velocity field. It is worth mentioning that some authors [49] instead identified a tangent vector with the rate of change in the density which is $-\operatorname{div}(P\nabla\phi)$ by the continuity equation.

[3]The functional representation approach is of course well-known in analysis; an exploration of this viewpoint for information theory problems can be found in [32], [33], [36], [38], and [39].

Next, we have

$$\frac{d^2}{ds^2} h(\mu_{st})$$

$$= -\int \frac{d^2\mu_{st}}{ds^2} \ln \mu_{st} - \int \mu_{st} \left(\frac{d\ln\mu_{st}}{ds}\right)^2 \quad (162)$$

$$= -\int \Delta(\mathbb{E}[U^2|Z_{st}]\mu_{st}) \ln \mu_{st} - \int [\nabla(\mathbb{E}[U|Z_{st}]\mu_{st})]^2 \frac{1}{\mu_{st}} \quad (163)$$

$$= -\int \mathbb{E}[U^2|Z_{st}]\mu_{st}\Delta \ln \mu_{st} - \int [\nabla(\mathbb{E}[U|Z_{st}]\mu_{st})]^2 \frac{1}{\mu_{st}}, \quad (164)$$

where (163) used (157) and

$$\frac{d}{ds}(\mathbb{E}[U|Z_{st}]\mu_{st}) = -\nabla(\mathbb{E}[U^2|Z_{st}]\mu_{st}), \quad (165)$$

and (165) can be shown as follows: Let f be an arbitrary smooth test function; we have

$$\frac{d}{ds}\int \mathbb{E}[U|Z_{st} = z]f(z)\mu_{st}(z)dz$$

$$= \frac{d}{ds}\mathbb{E}[Uf(Z_{st})] \quad (166)$$

$$= \mathbb{E}[Uf'(Z_{st})U] \quad (167)$$

$$= \int \mathbb{E}[U^2|Z_{st} = z]f'(z)\mu_{st}(z)dz \quad (168)$$

which implies (165) via integration by parts. Now taking $s, t = 0$ in (164) establishes the formula for $\mathrm{Hess}_{pp}\, h(p*q*r)(U,U)$ claimed in the lemma, and the proofs for the other Hessian components are similar. □

Remark 8: If we take Y, V, R to be constants, we recover the following formula for the Hessian of differential entropy

$$\mathrm{Hess}\, h(\mu)(U,U) = -\int U^2\mu\Delta\ln\mu - \int [\nabla(U\mu)]^2\frac{1}{\mu} \quad (169)$$

$$= -\int U^2\mu\Delta\ln\mu + \int U\mu\nabla\left(\frac{\nabla(U\mu)}{\mu}\right) \quad (170)$$

$$= -\int U^2\mu\Delta\ln\mu + \int U\mu\nabla(\nabla U + U\nabla\ln\mu) \quad (171)$$

$$= \int U\mu(\Delta U + \nabla U\nabla\ln\mu) \quad (172)$$

$$= -\int (\nabla U)^2\mu \quad (173)$$

which is well-known in literature (e.g. [44, p13]).

Remark 9: In Lemma 4, while U denotes an arbitrary perturbation of X, it can also be thought of as a random variable jointly distributed with X, which allows us to express the gradient and the Hessian conveniently using expectation notations (e.g. (146)(148)). A similar method of expressing gradient and Hessian in terms of expectations of a random variable parameterizing an arbitrary perturbation was previously adopted in Amin Gohari's thesis [27] for the derivation of cardinality bounds in network information theory. A difference, though, is that the perturbation in [27] is vertical rather than horizontal.

Using the same method we can compute the W_2 gradient and the Hessian of the variance functional (details omitted):

Lemma 5: Let p be a probability density function on $\mathbb{R}$ with zero mean, and let U be a smooth, compactly supported function on $\mathbb{R}$.

$$\mathrm{grad}\,\mathrm{var}(p) = 2x; \quad (174)$$

$$\mathrm{Hess}\,\mathrm{var}(p)(U) = 2\,\mathrm{var}(U). \quad (175)$$

Theorem 8: Given any $u > 0$ and $\lambda \geq 0$, define the functionals

$$\Psi(p,q) = uh(p*q*\gamma_u) + h(p) - (1+u)h(p*\gamma_u); \quad (176)$$

$$\Psi_\lambda(p,q) = \Psi(p,q) - \lambda\,\mathrm{var}(q). \quad (177)$$

Suppose that $K > 0$ and $L > 0$ are such that (K, L) is a stationary point of the function $(K, L) \mapsto \Psi_\lambda(\gamma_K, \gamma_L)$. Then $\mathrm{Hess}\,\Psi_\lambda(\gamma_K, \gamma_L)$ restricted on the subspace

$$\mathcal{C} := \{(U,V)\colon \int x_2 V(x_2)\gamma_L(x_2)dx_2 = 0\} \quad (178)$$

is negative-semidefinite if and only if $K \leq \frac{u}{(1+u)^{1/3}-1}$. In particular, if $K > \frac{u}{(1+u)^{1/3}-1}$ then (γ_K, γ_L) is not a local maximum of $\Psi(p,q)$ subject to $\mathrm{var}(q) \leq L$.

Similar to Corollary 1, Theorem 8 provides a counterexample to Conjecture 1 for the case where $N_1 = \Sigma_1 = 0$ and $N_2 = u$ (extension to the $N_1, \Sigma_1 > 0$ case is possible using a similar continuity argument as Theorem 1). *Proof:* [Proof of Theorem 8] Let us assume the expansions

$$U = \sum_{\alpha\in\{0,1,2,\dots\}} A_\alpha \frac{D^\alpha\gamma_K}{\gamma_K}, \quad (179)$$

$$V = \sum_{\alpha\in\{0,1,2,\dots\}} B_\alpha \frac{D^\alpha\gamma_L}{\gamma_L}. \quad (180)$$

Although the Hessians are calculated in Lemma 4 for smooth and compactly supported $U(\cdot)$ and $V(\cdot)$, they can be extended to a larger class of functions, say $U(\cdot)$ and $V(\cdot)$ bounded in a norm defined by $\|(U,V)\|_* := \mathbb{E}^{1/2}[(1+X^2)U^2(X) + \mathbb{E}^{1/2}[|\nabla U(X)|^2] + \mathbb{E}^{1/2}[(1+Y^2)V^2(Y)] + \mathbb{E}^{1/2}[|\nabla V(Y)|^2]$. We will set A_α and B_α nonzero for only finitely many α so that $\|(U,V)\|_* < \infty$, and we can check that the Hessian as a bilinear form is continuous with respect to $\|\|_*$, so that it can be extended to all (U,V) satisfying $\|(U,V)\|_* < \infty$ by passing to a limit.

We can now explicitly compute $\mathrm{Hess}\,\Psi_\lambda(\gamma_K, \gamma_L)$ using the following facts about Hermite polynomials (see [17]):

$$\int \left(\frac{D^\alpha\gamma_K}{\gamma_K}\right)^2 \gamma_K = \frac{\alpha}{K^\alpha}, \quad \forall\alpha = 0,1,2,\dots \quad (181)$$

and

$$\mathbb{E}\left[\frac{D^\alpha\gamma_K(X_1)}{\gamma_K(X_1)}\,\Big|\,\hat{X}_1\right] = \frac{D^\alpha\gamma_{K+u}(\hat{X}_1)}{\gamma_{K+u}(\hat{X}_1)}, \quad (182)$$

where $X_1 \sim \gamma_K$, $\hat{X}_1 = X_1 + Z_2$, and $Z_2 \sim \gamma_u$ is independent of X_1. We have

$$\mathrm{Hess}_{11}\,\Psi_\lambda(\gamma_K, \gamma_L)(U,U)$$

$$= \frac{u}{K+u+L}\sum_{\alpha\geq 0} A_\alpha^2 \frac{\alpha!}{K^\alpha} - u\sum_{\alpha\geq 0} A_\alpha^2 \frac{(\alpha+1)!}{(K+u+L)^{\alpha+1}}$$

$$+ \frac{1}{K} \sum_{\alpha \geq 0} A_\alpha^2 \frac{\alpha!}{K^\alpha} - \sum_{\alpha \geq 0} A_\alpha^2 \frac{(\alpha+1)!}{K^{\alpha+1}}$$
$$- \frac{1+u}{K+u} \sum_{\alpha \geq 0} A_\alpha^2 \frac{\alpha!}{K^\alpha} + (1+u) \sum_{\alpha \geq 0} A_\alpha^2 \frac{(\alpha+1)!}{(K+u)^{\alpha+1}} \tag{183}$$

where the subscript of Hess_{11} means that the function is differentiated in the first argument twice. Since K is a stationary point, the first, third, and fifth summations cancel out, and the $\alpha = 0$ summands in the other three summations cancel out. We are left with

$$\mathrm{Hess}_{11} \Psi_\lambda(\gamma_K, \gamma_L)(U, U) = -u \sum_{\alpha \geq 1} A_\alpha^2 \frac{(\alpha+1)!}{(K+u+L)^{\alpha+1}}$$
$$- \sum_{\alpha \geq 1} A_\alpha^2 \frac{(\alpha+1)!}{K^{\alpha+1}}$$
$$+ (1+u) \sum_{\alpha \geq 1} A_\alpha^2 \frac{(\alpha+1)!}{(K+u)^{\alpha+1}}. \tag{184}$$

Next, since $B_1 = 0$ by (178), we have

$$\mathrm{Hess}_{22} \Psi_\lambda(\gamma_K, \gamma_L)(V, V)$$
$$= \frac{u}{K+u+L} \sum_{\alpha \neq 1} B_\alpha^2 \frac{\alpha!}{L^\alpha} - u \sum_{\alpha \neq 1} B_\alpha^2 \frac{(\alpha+1)!}{(K+u+L)^{\alpha+1}}$$
$$- 2\lambda \sum_{\alpha \neq 1} B_\alpha^2 \frac{\alpha!}{L^\alpha} + 2\lambda B_0^2 \tag{185}$$
$$= \frac{u}{K+u+L} \sum_{\alpha \geq 2} B_\alpha^2 \frac{\alpha!}{L^\alpha} - u \sum_{\alpha \geq 2} B_\alpha^2 \frac{(\alpha+1)!}{(K+u+L)^{\alpha+1}}$$
$$- 2\lambda \sum_{\alpha \geq 2} B_\alpha^2 \frac{\alpha!}{L^\alpha} \tag{186}$$
$$= -u \sum_{\alpha \geq 2} B_\alpha^2 \frac{(\alpha+1)!}{(K+u+L)^{\alpha+1}} \tag{187}$$

where the last step follows since $\partial_L \Psi_\lambda(\gamma_K, \gamma_L) = 0$ implies $\lambda = \frac{u}{K+u+L}$. Finally

$$\mathrm{Hess}_{12} \Psi_\lambda(\gamma_K, \gamma_L)(U, V)$$
$$= \frac{u}{K+u+L} A_0 B_0 - u \sum_{\alpha \neq 1} A_\alpha B_\alpha \frac{(\alpha+1)!}{(K+u+L)^{\alpha+1}} \tag{188}$$
$$= -u \sum_{\alpha \geq 2} A_\alpha B_\alpha \frac{(\alpha+1)!}{(K+u+L)^{\alpha+1}}. \tag{189}$$

Now we can write

$$\mathrm{Hess} \Psi_\lambda(\gamma_K, \gamma_L)(U, V)$$
$$= \mathrm{Hess}_{11} \Psi_\lambda(\gamma_K, \gamma_L)(U, U) + \mathrm{Hess}_{22} \Psi_\lambda(\gamma_K, \gamma_L)(V, V)$$
$$+ 2 \mathrm{Hess}_{12} \Psi_\lambda(\gamma_K, \gamma_L)(U, V) \tag{190}$$
$$= \sum_{\alpha \geq 1} I_\alpha \tag{191}$$

where we defined

$$I_1 := -\frac{2uA_1^2}{(K+u+L)^2} - \frac{2A_1^2}{K^2} + \frac{2(1+u)A_1^2}{(K+u)^2}; \tag{192}$$

and for $\alpha \geq 2$,

$$\frac{I_\alpha}{(\alpha+1)!} := -\frac{uA_\alpha^2}{(K+u+L)^{\alpha+1}} - \frac{A_\alpha^2}{K^{\alpha+1}} + \frac{(1+u)A_\alpha^2}{(K+u)^{\alpha+1}}$$
$$- \frac{uB_\alpha^2}{(K+u+L)^{\alpha+1}} - \frac{2uA_\alpha B_\alpha}{(K+u+L)^{\alpha+1}}. \tag{193}$$

Note that $I_1 \leq 0$, which follows since $\partial_K^2 \Psi_\lambda(\gamma_K, \gamma_L) < 0$ at the maximizer K. Moreover given A_α we have

$$\frac{I_\alpha}{(\alpha+1)!} \leq -\frac{A_\alpha^2}{K^{\alpha+1}} + \frac{(1+u)A_\alpha^2}{(K+u)^{\alpha+1}} \tag{194}$$

with equality achieved only when $B_\alpha = -A_\alpha$. Then the first claim of the theorem follows, noting that $K = \frac{u}{(1+u)^{1/3}-1}$ is the solution to $-\frac{1}{K^3} + \frac{1+u}{(K+u)^3} = 0$. The case where $\mathrm{Hess} \Psi_\lambda$ is not negative-semidefinite can be seen by choosing $B_2 = -A_2 \neq 0$ and $A_\alpha = B_\alpha = 0$ for $\alpha \neq 2$. Once the restricted Hessian is not negative-semidefinite, we can show that the stationary point does not achieve local maximum with arguments similar to (142): we consider (p_t, q_t) where p_t and q_t are Wasserstein geodesics satisfying $p_0 = \gamma_K$, $q_0 = \gamma_L$, $\frac{d}{dt} \int x^2 q_t|_{t=0} = 0$ and $\frac{d^2}{dt^2} \Psi_\lambda(p_t, q_t)|_{t=0} > 0$. We can assume without loss of generality that the corresponding tangent vectors (143) come from gradients of smooth compactly supported functions. Then let $\bar{q}_t$ be a rescaling of q_t so that $\mathrm{var}(q_t) = L$, $t > 0$. Then the same argument as in (142) shows that $(p_t, \bar{q}_t)$ beats the Gaussian stationary point for small $t > 0$, *mutatis mutandis.* $\square$

REFERENCES

[1] E. Abbe and L. Zheng, "A coordinate system for Gaussian networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 721–733, Feb. 2012.

[2] L. Ambrosio, N. Gigli, and G. Savaré, "Gradient flows in metric spaces and in the space of probability measures," in *Lectures in Mathematics ETH Zürich*. Basel, Switzerland: Birkhäuser, 2008.

[3] V. Anantharam, V. Jog, and C. Nair, "Unifying the Brascamp–Lieb inequality and the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 68, no. 12, pp. 7665–7684, Dec. 2022.

[4] E. Aras and T. A. Courtade, "Sharp maximum-entropy comparisons," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1504–1509.

[5] S. Beigi, S. Liu, C. Nair, and M. Yazdanpanah, "Some results on the scalar Gaussian interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2199–2203.

[6] S. G. Bobkov, G. P. Chistyakov, and F. Götze, "Regularized distributions and entropic stability of Cramer's characterization of the normal law," *Stochastic Processes Their Appl.*, vol. 126, no. 12, pp. 3865–3887, Dec. 2016.

[7] S. Chewi, T. Maunu, P. Rigollet, and A. J. Stromme, "Gradient descent algorithms for Bures–Wasserstein barycenters," in *Proc. Conf. Learn. Theory*, 2020, pp. 1276–1304.

[8] G. Clerc, G. Conforti, and I. Gentil, "On the variational interpretation of local logarithmic Sobolev inequalities," 2020, *arXiv:2011.05207.*

[9] M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 5, pp. 607–615, Sep. 1985.

[10] M. Costa, C. Nair, and D. Ng, "On the Gaussian Z-interference channel," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2017, pp. 1–15.

[11] M. Costa, C. Nair, D. Ng, and Y. N. Wang, "On the structure of certain non-convex functionals and the Gaussian Z-interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1522–1527.

[12] M. H. M. Costa, "Noisebergs in Z Gaussian interference channels," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2011, pp. 1–6.

[13] T. A. Courtade, "A strong entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2173–2192, Apr. 2018.

[14] T. M. Cover, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1999.

[15] K. Craig, "The exponential formula for the Wasserstein metric," *ESAIM, Control, Optim. Calculus Variat.*, vol. 22, no. 1, pp. 169–187, Jan. 2016.

[16] H. Cramér, "Über eine eigenschaft der normalen verteilungsfunktion," *Mathematische Zeitschrift*, vol. 41, no. 1, pp. 405–414, 1936.

[17] H. Cramér, *Mathematical Methods of Statistics*, vol. 43. Princeton, NJ, USA: Princeton Univ. Press, 1999.

[18] J. R. L. D. Almeida and D. J. Thouless, "Stability of the Sherrington–Kirkpatrick solution of a spin glass model," *J. Phys. A, Math. Gen.*, vol. 11, no. 5, pp. 983–990, May 1978.

[19] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Feb. 1991.

[20] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[21] R. Eldan, J. Lehec, and Y. Shenfeld, "Stability of the logarithmic Sobolev inequality via the Föllmer process," in *Annales de L'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 56. Paris, France: Institut Henri Poincaré, 2020, pp. 2253–2269.

[22] T. Gao, L.-H. Lim, and K. Ye, "Semi-Riemannian manifold optimization," 2018, *arXiv:1812.07643*.

[23] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2087–2104, Apr. 2014.

[24] A. Gohari and C. Nair, "New outer bounds for the two-receiver broadcast channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1492–1497.

[25] A. Gohari and C. Nair, "Outer bounds for multiuser settings: The auxiliary receiver approach," *IEEE Trans. Inf. Theory*, vol. 68, no. 2, pp. 701–736, Feb. 2022.

[26] A. Gohari, C. Nair, and D. Ng, "An information inequality motivated by the Gaussian Z-interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 2744–2749.

[27] A. A. Gohari, *Results and Techniques in Multiuser Information Theory*. Berkeley, CA, USA: Univ. California, 2010.

[28] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[29] R. Holley and D. Stroock, "Logarithmic sobolev inequalities and stochastic ising models," *J. Stat. Phys.*, vol. 46, nos. 5–6, pp. 1159–1194, Mar. 1987.

[30] M. Ledoux, "Logarithmic Sobolev inequalities for unbounded spin systems revisited," in *Séminaire de Probabilités XXXV*. New York, NY, USA: Springer-Verlag, 2001, pp. 167–194.

[31] E. H. Lieb, "Gaussian kernels have only Gaussian maximizers," *Inventiones Mathematicae*, vol. 102, no. 1, pp. 179–208, Dec. 1990.

[32] J. Liu, "Information theory from a functional viewpoint," Ph.D. thesis, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, 2018.

[33] J. Liu, "Dispersion bound for the Wyner–Ahlswede–Körner network via a semigroup method on types," *IEEE Trans. Inf. Theory*, vol. 67, no. 2, pp. 869–885, Aug. 2020.

[34] J. Liu, "Minoration via mixed volumes and Cover's problem for general channels," *Probab. Theory Rel. Fields*, vol. 183, pp. 1–43, Jun. 2022.

[35] J. Liu, "Stability analysis of the Gaussian Han–Kobayashi region for Z-interference channels," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Jun. 2023, pp. 1–32.

[36] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Smoothing Brascamp–Lieb inequalities and strong converses of coding theorems," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 704–721, Feb. 2020.

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

[38] J. Liu and A. Özgür, "Capacity upper bounds for the relay channel via reverse hypercontractivity," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5448–5455, Sep. 2020.

[39] J. Liu, R. van Handel, and S. Verdú, "Second-order converses via reverse hypercontractivity," *Math. Statist. Learn.*, vol. 2, no. 2, pp. 103–163, 2019.

[40] C. Nair and D. Ng, "Invariance of the Han–Kobayashi region with respect to temporally-correlated Gaussian inputs," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1372–1374, Mar. 2019.

[41] C. Nair and D. Ng, "On the scalar Gaussian interference channel," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2018, pp. 1–5.

[42] C. Nair, L. Xia, and M. Yazdanpanah, "Sub-optimality of Han–Kobayashi achievable region for interference channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 2416–2420.

[43] W. H. Ng, "On Gaussian extremizers for the capacity region of the Gaussian interference channel," Ph.D. thesis, Dept. Inf. Eng., Chin. Univ. Hong Kong, Hong Kong, 2021.

[44] F. Otto and C. Villani, "Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality," *J. Funct. Anal.*, vol. 173, no. 2, pp. 361–400, Jun. 2000.

[45] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3992–4002, Jul. 2016.

[46] H. Sato, "On degraded Gaussian two-user channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 5, pp. 637–640, Sep. 1978.

[47] X. Shang et al., "Two-user Gaussian interference channels: An information theoretic point of view," *Found. Trends Commun. Inf. Theory*, vol. 10, no. 3, pp. 247–378, 2013.

[48] S. Sra, N. K. Vishnoi, and O. Yildiz, "On geodesically convex formulations for the Brascamp–Lieb constant," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[49] C. Villani, *Optimal Transport: Old and New*, vol. 338. Berlin, Germany: Springer-Verlag, 2009.

[50] Y. Wang and W. Li, "Information Newton's flow: Second-order optimization method in probability space," 2020, *arXiv:2001.04341*.

Jingbo Liu (Member, IEEE) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2012, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA, in 2014 and 2017, respectively.

He was a Norbert Wiener Post-Doctoral Research Fellow with the MIT Institute for Data, Systems, and Society (IDSS) (2018–2020). Since 2020, he has been an Assistant Professor with the Department of Statistics and an affiliate with the Department of Electrical and Computer Engineering, University of Illinois Urbana–Champaign, IL, USA. His research interests include information theory, high dimensional statistics and probability, coding theory, and related fields. His undergraduate thesis received the Best Undergraduate Thesis Award at Tsinghua University in 2012. He gave a semi-plenary presentation at the 2015 IEEE International Symposium on Information Theory, Hong Kong, China. He was a recipient of the Princeton University Wallace Memorial Honorific Fellowship in 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 in 2018.