Quantitative Stability of the Entropy Power Inequality

Thomas A. Courtade, *Member, IEEE*, Max Fathi, and Ashwin Pananjady

***Abstract*— We establish quantitative stability results for the entropy power inequality (EPI). Specifically, we show that if uniformly log-concave densities nearly saturate the EPI, then they must be close to Gaussian densities in the quadratic Kantorovich-Wasserstein distance. Furthermore, if one of the densities is Gaussian and the other is log-concave, or more generally has positive spectral gap, then the deficit in the EPI can be controlled in terms of the L^1-Kantorovich-Wasserstein distance or relative entropy, respectively. As a counterpoint, an example shows that the EPI can be unstable with respect to the quadratic Kantorovich-Wasserstein distance when densities are uniformly log-concave on sets of measure arbitrarily close to one. Our stability results can be extended to non-log-concave densities, provided certain regularity conditions are met. The proofs are based on mass transportation.**

***Index Terms*— Entropy power inequality, optimal transport, stability.**

I. Introduction

LET X and Y be independent random vectors on $\mathbb{R}^n$ with corresponding laws μ and ν, each absolutely continuous with respect to Lebesgue measure. The celebrated entropy power inequality (EPI) proposed by Shannon [2] and proved by Stam [3] asserts that

$$N(X+Y) \geq N(X) + N(Y), \tag{1}$$

where $N(X) \equiv N(\mu) := \frac{1}{2\pi e} e^{2\, h(X)/n}$ denotes the entropy power of X, and $h(X) \equiv h(\mu) = -\int_{\mathbb{R}^n} f \log f$ is the entropy of X having density f. For a parameter $t \in (0,1)$, let us define

$$\delta_{\mathsf{EPI},t}(\mu,\nu) := h(\sqrt{t}X + \sqrt{1-t}Y) - \Big(th(X) + (1-t)h(Y)\Big). \tag{2}$$

Unaware of the works by Shannon, Stam and Blachman [4], Lieb [5] rediscovered the EPI by establishing $\delta_{\mathsf{EPI},t}(\mu,\nu) \geq 0$ and noting its formal equivalence to (1). The equivalence of these inequalities is reviewed in Section III-B. Due to the formal equivalence of the Shannon-Stam and Lieb inequalities, we shall generally refer to both as the EPI.

It is well known that equality is achieved in the Shannon-Stam EPI if and only if X and Y are Gaussian vectors with proportional covariances. Equivalently, $\delta_{\mathsf{EPI},t}(\mu,\nu)$ vanishes if and only if μ, ν are Gaussian measures that are identical up to translation.[1] However, despite the fundamental role the EPI plays in information theory and crisp characterization of equality cases, few stability estimates are known. Specifically, our motivating question is the following quantitative reinforcement of equality conditions for the EPI:

If $\delta_{\mathsf{EPI},t}(\mu,\nu)$ is small, must μ and ν be 'close' to Gaussian measures, which are themselves 'close' to each other, in a precise and quantitative sense?

Toward answering this question, our main result is a dimension-free, quantitative stability estimate for the EPI. More specifically, we show that if the measures μ, ν have uniformly log-concave densities and nearly saturate either form of the EPI, then they must also be close to Gaussian measures in the quadratic Kantorovich-Wasserstein distance. We also show that the EPI is *not* stable (with respect to the same criterion) in situations where the densities *nearly* satisfy the same regularity conditions. Other quantitative deficit estimates are obtained when one of the two variables is Gaussian and the other is log-concave or has positive spectral gap. Dimension-dependent estimates are obtained in certain more general situations.

Before stating the main results, let us first introduce some notation. We let $\Gamma \equiv \Gamma(\mathbb{R}^n)$ denote the set of centered Gaussian probability measures on $\mathbb{R}^n$, and let γ denote the standard Gaussian measure on $\mathbb{R}^n$. That is,[2]

$$d\gamma(x) = d\gamma^n(x) = e^{-|x|^2/2} \frac{dx}{(2\pi)^{n/2}}.$$

Next, we recall that the L^2-Kantorovich-Wasserstein distance between probability measures μ, ν is defined according to

$$W_2(\mu,\nu) = \inf \left(\mathbb{E}|X-Y|^2\right)^{1/2},$$

where $|\cdot|$ denotes the Euclidean metric on $\mathbb{R}^n$ and the infimum is over all couplings on X, Y with marginal laws $X \sim \mu$ and $Y \sim \nu$. If $X \sim \mu$ is a centered random vector, then we

Manuscript received June 12, 2017; revised January 3, 2018; accepted February 1, 2018. Date of publication February 20, 2018; date of current version July 12, 2018. This paper was presented in part at the 2017 International Symposium on Information Theory [1]. T. A. Courtade and A. Pananjady were supported by NSF (Center for Science of Information) under Grant CCF-1528132 and Grant CCF-0939370. M. Fathi was supported by NSF FRG under Grant DMS-1361122. T. A. Courtade and M. Fathi acknowledge support from the France–Berkeley Fund.

T. A. Courtade and A. Pananjady are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: courtade@berkeley.edu).

M. Fathi is with CNRS, 75016 Paris, France, and also with the Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse, France.

Communicated by I. Sason, Associate Editor at Large.

Digital Object Identifier 10.1109/TIT.2018.2808161

[1] Lieb did not settle the cases of equality; this was done later by Carlen and Soffer [6].

[2] Explicit dependence of quantities on the ambient dimension n will be suppressed in situations where our arguments are the same in all dimensions.

write $\Sigma_\mu = \mathbb{E}XX^\top$ to denote the covariance matrix of X. For a symmetric positive semidefinite matrix Σ, we write $\Sigma^{1/2}$ to denote the unique symmetric positive semidefinite matrix satisfying $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$.

Remark 1: Both forms of the EPI are invariant to translation of the measures μ, ν. Thus, our persistent assumption of centered probability measures is for convenience and comes without loss of generality.

Remark 2: Without explicit mention, we assume throughout that all entropies exist in the usual Lebesgue sense. When dealing with the Shannon-Stam or Lieb inequalities, one generally needs to make this assumption. Indeed, it is possible for $h(X+Y)$ to not exist in the usual Lebesgue sense even though $h(X)$ and $h(Y)$ exist and are finite [7, Proposition 1]. That being said, our main results primarily concern log-concave distributions. As such, all involved entropies are guaranteed to exist (e.g., [8]).

Organization

The rest of this paper is organized as follows: Sections II-A and II-B describe our main stability results for log-concave densities and the relationship to previous work, respectively. Section II-C gives an example where the EPI is not stable with respect to the quadratic Kantorovich-Wasserstein distance when regularity conditions are not met. Section III gives proofs of our main results and a brief discussion of techniques, which are based on optimal mass transportation. We conclude in Section IV with extensions of our results to more general settings.

II. Main Results

This section describes our main results, and also compares to previously known stability estimates. Proofs are generally deferred until Section III.

A. Stability of the EPI for Log-Concave Densities

Our main result is the following:

Theorem 3: Let $\mu = e^{-\varphi}\gamma$ and $\nu = e^{-\psi}\gamma$ be centered probability measures, where φ and ψ are convex. Then

$$\delta_{\mathsf{EPI},t}(\mu,\nu) \geq \frac{t(1-t)}{2} \inf_{\gamma_1,\gamma_2\in\Gamma} \Big(W_2^2(\mu,\gamma_1) + W_2^2(\nu,\gamma_2) + W_2^2(\gamma_1,\gamma_2)\Big). \tag{3}$$

Remark 4: The notation $\mu = e^{-\varphi}\gamma$ is shorthand for $\frac{d\mu}{d\gamma} = e^{-\varphi}$. Measures of the form $\mu = e^{-\varphi}\gamma$ for convex φ have several names in the literature. Such names include 'strongly log-concave densities', 'log-concave perturbation of Gaussian', 'uniformly convex potential' and 'strongly convex potential' (see [9, pp. 50–51]). This situation also corresponds to the Bakry-Émery condition $CD(1,\infty)$ when the space is $\mathbb{R}^n$.

Under the assumptions of the theorem, the three terms in the RHS of (3) explicitly give necessary conditions for the deficit $\delta_{\mathsf{EPI},t}(\mu,\nu)$ to be small. In particular, μ,ν must each be quantitatively close to Gaussian measures, which are themselves quantitatively close to one another. Additionally, W_2^2 is additive on product measures, so the estimate (3) is dimension-free, which is compatible with the additivity of $\delta_{\mathsf{EPI},t}$ on product measures.

Theorem 3 may be readily adapted to the setting of uniformly log-concave densities. Toward this end, let $\eta > 0$ and recall that $h(\eta^{1/2}X) = h(X) + \frac{1}{2}\log\eta$, so that $\delta_{\mathsf{EPI},t}(\mu,\nu)$ is invariant under the rescaling $(X,Y) \to (\eta^{1/2}X, \eta^{1/2}Y)$. Similarly, if $X \sim \mu$ has density f that is uniformly log-concave in the sense that

$$-\nabla^2 \log f \geq \eta \mathrm{I}, \tag{4}$$

then a change of variables reveals that the density f_η associated with the rescaled random variable $\eta^{1/2}X$ satisfies $-\nabla^2 \log f_\eta \geq \mathrm{I}$. In particular, $f_\eta dx = e^{-\varphi}d\gamma$ for some convex function φ. Thus, Theorem 3 is equivalent to the following:

Corollary 5: If μ and ν are centered probability measures with densities satisfying (4)*, then*

$$\delta_{\mathsf{EPI},t}(\mu,\nu) \geq \eta\frac{t(1-t)}{2} \inf_{\gamma_1,\gamma_2\in\Gamma} \Big(W_2^2(\mu,\gamma_1) + W_2^2(\nu,\gamma_2) + W_2^2(\gamma_1,\gamma_2)\Big).$$

This result will also apply to certain families of non log-concave measures, see Remark 17.

For convenience, let $\mathsf{d}^2_{W_2}(\mu) := \inf_{\gamma_0\in\Gamma} W_2^2(\mu,\gamma_0)$ denote the squared W_2-distance from μ to the set of centered Gaussian measures. Using the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ and the triangle inequality for W_2, we may conclude a weaker, but potentially more convenient variant of Corollary 5.

Corollary 6: If μ and ν are centered probability measures with densities satisfying (4)*, then*

$$\delta_{\mathsf{EPI},t}(\mu,\nu) \geq \eta\frac{t(1-t)}{8}\Big(\mathsf{d}^2_{W_2}(\mu) + \mathsf{d}^2_{W_2}(\nu) + W_2^2(\mu,\nu)\Big). \tag{5}$$

The Shannon-Stam form of the entropy power inequality (1) is oftentimes preferred to Lieb's inequality for applications in information theory. Starting with Corollary 6, we may establish an analogous estimate for the Shannon-Stam EPI. Recall first that (1) attains equality if and only if μ,ν are Gaussian with proportional covariance matrices. Motivated by this, we define the quantity

$$\mathsf{d}_F^2(\mu,\nu) := \inf_{\theta\in(0,1)} \left\| \sqrt{\theta}\,\Sigma_\mu^{1/2} - \sqrt{1-\theta}\,\Sigma_\nu^{1/2}\right\|_F^2, \tag{6}$$

where $\|A\|_F := \sqrt{\mathrm{Tr}(A^\top A)}$ denotes Frobenius (or Hilbert-Schmidt) norm of a real matrix A, to provide a convenient measure of distance between the second order statistics of μ,ν. In particular, $\mathsf{d}_F^2(\mu,\nu) = 0$ if and only if Σ_μ and Σ_ν are proportional. With this notation established, we have the following quantitative reinforcement of equality conditions in the Shannon-Stam EPI:

Corollary 7: Let μ and ν be centered probability measures on $\mathbb{R}^n$ satisfying (4) *with parameters η_μ and η_ν, respectively. Then,*

$$N(\mu * \nu) \geq (N(\mu) + N(\nu))\,\Delta_{\mathsf{EPI}}(\mu,\nu),$$

where $\Delta_{\mathsf{EPI}}(\mu,\nu)$ *is defined as the quantity*

$$\exp\left(\frac{\min\{\theta\eta_\mu,(1-\theta)\eta_\nu\}}{4\,n}\Big((1-\theta)\mathsf{d}^2_{W_2}(\mu) + \theta\mathsf{d}^2_{W_2}(\nu)+\mathsf{d}^2_F(\mu,\nu)\Big)\right), \quad (7)$$

and θ *is chosen to satisfy* $\theta/(1-\theta)=N(\mu)/N(\nu)$.

As remarked above, equality is attained in (1) if and only if μ,ν are Gaussian with proportional covariances. Under the stated assumptions of log-concavity, these conditions are explicitly captured by the last three terms in (7).

We also derive a stability estimate when one vector is simply log-concave (but not uniformly so) and the other vector is Gaussian, involving the L^1-Kantorovich-Wasserstein distance with respect to the ℓ_1 metric on $\mathbb{R}^n$: That is,

$$W_{1,1}(\mu,\nu) := \inf \mathbb{E}\|X-Y\|_1 = \inf \mathbb{E}\sum_{i=1}^n |X_i - Y_i|,$$

where the infimum is over all couplings of X, Y with marginal laws $X\sim\mu$ and $Y\sim\nu$.

Theorem 8: For any log-concave centered random vector in $\mathbb{R}^n$ *with law* μ,

$$\delta_{\mathsf{EPI},t}(\mu,\gamma) \ge Ct(1-t)\min(n^{-1}W^2_{1,1}(\mu,\gamma),1), \quad (8)$$

with C *an absolute constant that does not depend on* μ *or dimension* n.

This estimate is reminiscent of the deficit estimates on Talagrand's inequality of [10] and [11], with a remainder term that stays bounded when the distance becomes large. Note that $W_{1,1}$ grows linearly with dimension for product measures, so the term $n^{-1}W^2_{1,1}(\mu,\gamma)$ in the RHS of (8) has the correct dependence on dimension.

Finally, in a more general direction, a measure μ is said to have spectral gap λ if, for all smooth $s:\mathbb{R}^n\to\mathbb{R}$ with $\int s\,d\mu = 0$,

$$\lambda\int s^2\,d\mu \le \int |\nabla s|^2\,d\mu.$$

Theorem 9: If μ *is a centered probability measure on* $\mathbb{R}^n$ *with spectral gap* λ*, then*

$$\delta_{\mathsf{EPI},t}(\mu,\gamma) \ge \min(\lambda,1)t(1-t)D(\mu\|\gamma),$$

where $D(\mu\|\gamma)=\int d\mu\log\frac{d\mu}{d\gamma}$ *is the relative entropy between* μ *and* γ.

All log-concave distributions have positive spectral gap [12], so the hypothesis of Theorem 9 is weaker than that of Theorem 8. However, the advantage of (8) is that it does not rely on any quantitative information on μ, only that it is log-concave.

B. Relation to Prior Work

As remarked above, a few stability estimates are known for the EPI. Here, we review those that are most relevant and comment on the relationship to our results. To begin, we mention a stability result due to Toscani [13], which asserts for probability measures μ,ν with log-concave densities, there is an explicit function R such that

$$N(\mu*\nu) \ge (N(\mu)+N(\nu))\,R(\mu,\nu),$$

where $R(\mu,\nu)\ge 1$ with equality only if μ,ν are Gaussian measures. However, this result is not directly comparable to ours since the function $R(\mu,\nu)$ is quite complicated,[3] and does not explicitly control the distance of μ,ν to the space of Gaussian measures. Toscani leaves this as an open problem [13, Remark 7]. Corollary 7 provides a satisfactory answer to his problem when μ,ν satisfy (4) for some parameter $\eta>0$. Similarly, Theorems 8 and 9 provide an answer when one of the measures is Gaussian and the other satisfies regularity conditions.

Next, we compare Corollary 5 to the main result of Ball and Nguyen [14], which states that if μ is a centered isotropic probability measure (i.e., $\Sigma_\mu = \mathrm{I}$) with spectral gap λ and log-concave density (not necessarily uniformly), then

$$\delta_{\mathsf{EPI},1/2}(\mu,\mu) \ge \frac{\lambda}{4(1+\lambda)}D(\mu\|\gamma) \ge \frac{\lambda}{8(1+\lambda)}W_2^2(\mu,\gamma), \quad (9)$$

where the second inequality is Talagrand's information-transportation inequality. Now, if μ satisfies (4), then Corollary 5 yields the similar bound

$$\delta_{\mathsf{EPI},1/2}(\mu,\mu) \ge \frac{\eta}{4}\inf_{\gamma_0\in\Gamma} W_2^2(\mu,\gamma_0).$$

Given the similarity, Corollary 5 may be viewed as a close relative of (9), which holds for non-identical measures and all parameters $t\in(0,1)$. However, two points should be mentioned: (i) a stability estimate with respect to W_2 is weaker than one involving relative entropy; and (ii) η-uniform log-concavity in the sense of (4) ensures a spectral gap of at least η, but not vice versa. Thus, it is interesting to ask whether the hypothesis of Corollary 5 can be weakened to require only a spectral gap; the result of Ball and Nguyen and a similar earlier result by Ball *et al.* [15] in dimension one provides some grounds for cautious optimism. In Section 4, we shall obtain a one-dimensional result for non log-concave measures under a stronger assumption than [15] (namely a Cheeger isoperimetric inequality), but with the advantage of being valid for non-identical measures.

It should be noted that Theorem 9 has weaker hypotheses than (9) since log-concavity is not assumed, however it applies to the EPI deficit $\delta_{\mathsf{EPI},t}(\mu,\gamma)$, rather than convolution of identical measures μ which may be of interest in applications to the central limit theorem.

The results referenced above assume log-concave densities, as do we (for the most part). In contrast, the refined EPI established in [16] provides a qualitative stability estimate for the EPI when μ is arbitrary and ν is Gaussian. However, the deficit is quantified in terms of the so-called *strong data processing function*, and is therefore not directly comparable to the present results. Nevertheless, a noteworthy consequence is a *reverse* entropy power inequality, which does bear some resemblance to the result of Corollary 7. In particular, for

[3] $R(\mu,\nu)$ is expressed in terms of integrals of nonlinear functionals evaluated along the evolutes of μ,ν under the heat semigroup.

arbitrary probability measures μ, ν on $\mathbb{R}^n$ with finite second moments, it was shown in [17] that

$$N(\mu * \nu) \leq (N(\mu) + N(\nu))\left((1-\theta)\mathsf{p}(\mu) + \theta \mathsf{p}(\nu)\right), \quad (10)$$

where θ is the same as in the definition of $\Delta_{\mathsf{EPI}}(\mu, \nu)$ and $\mathsf{p}(\mu) := \frac{1}{n}N(\mu)J(\mu) \geq 1$ is the *Stam defect*, with $J(\mu)$ denoting Fisher information. We have $\mathsf{p}(\mu) = 1$ only if μ is Gaussian, and thus $\mathsf{p}(\mu)$ may reasonably be interpreted as a measure of how far μ is from the set of Gaussian measures. Thus, the deficit term $(1-\theta)\mathsf{p}(\mu) + \theta\mathsf{p}(\nu)$ in (10) bears a pleasant resemblance to the deficit term $(1-\theta)\mathsf{d}^2_{W_2}(\mu) + \theta \mathsf{d}^2_{W_2}(\nu)$ in Corollary 7. Importantly, though, the former is an upper bound on $N(\mu * \nu)$, while the latter yields a lower bound.

We would be remiss to not mention that the inequality $\mathsf{p}(\mu) \geq 1$ mentioned above is known as Stam's inequality, and is equivalent to Gross' celebrated logarithmic Sobolev inequality for Gaussian measure [18], [19]. Taking $\mu = \nu$ in (10) gives the sharpening $\mathsf{p}(\mu) \geq \exp\left(\frac{2}{n}\delta_{\mathsf{EPI},t}(\mu,\mu)\right)$, holding for any probability measure μ with finite second moment. Equivalently, if μ is centered, then

$$\frac{1}{2}I(\mu\|\gamma) \geq D(\mu\|\gamma) + \delta_{\mathsf{EPI},t}(\mu,\mu),$$

where $I(\mu\|\gamma) := \int |\nabla \log \frac{d\mu}{d\gamma}|^2\, d\mu$ denotes the relative Fisher information of μ with respect to γ. When μ satisfies (4), we have a dimension-free quantitative stability result for the logarithmic Sobolev inequality $\frac{1}{2}I(\mu\|\gamma) \geq D(\mu\|\gamma)$. This is an improvement upon the main result of Indrei and Marcon [20], who consider the subset of densities satisfying (4) for some parameter $\eta > 0$, whose Hessians are also uniformly upper bounded. Unfortunately, this improvement is already obsolete, as Fathi *et al.* [10] have recently shown that a similar result holds for all probability measures with positive spectral gap. Interestingly though, (10) does imply a general upper bound on $\delta_{\mathsf{EPI},t}(\mu,\nu)$ involving Fisher informations. Specifically, for arbitrary probability measures μ, ν with finite second moments,

$$\delta_{\mathsf{EPI},t}(\mu,\nu) \leq (1-t)\left(\frac{1}{2}I(\mu\|\gamma) - D(\mu\|\gamma)\right) + t\left(\frac{1}{2}I(\nu\|\gamma) - D(\nu\|\gamma)\right).$$

See [17] for details. Other deficit estimates for the logarithmic Sobolev inequality have been obtained in [21] and [22].

If $X \sim \mu$ is a radially symmetric random vector on $\mathbb{R}^n$, $n \geq 2$, satisfying modest regularity conditions (e.g,. convolution with a Gaussian measure of small variance is sufficient), then it was recently established in [23] that, for any $\varepsilon > 0$

$$\delta_{\mathsf{EPI},1/2}(\mu,\mu) \geq C_\varepsilon(\mu) n^\varepsilon D^{1+\varepsilon}(\mu\|\gamma_\mu), \quad (11)$$

where γ_μ denotes the Gaussian measure with the same covariance as μ, and $C_\varepsilon(\mu)$ is an explicit function that depends only on ε, a finite number of moments of μ, and its regularity. This closely parallels quantitative estimates on entropy production in the Boltzmann equation [24], [25]. Neither (3) nor (11) imply the other since the hypotheses required are quite different (strong log-concavity vs. radial symmetry). However, both results do give quantitative bounds on entropy production under convolution in terms of a distance from Gaussian measures. In general, the constants in (3) will be much better than those in (11) which, although numerical, can be quite small. We return to the setting of radially symmetric measures in Section IV.

Finally, we mention Carlen and Soffer's qualitative stability estimate for the EPI that holds under general conditions [6]. Roughly speaking, their result is the following: if probability measures μ, ν on $\mathbb{R}^n$ are isotropic with Fisher informations upper bounded by J_0, then there is a nonlinear function $\Theta\colon \mathbb{R} \to [0,\infty)$, strictly increasing from 0, depending only on the dimension n, the parameter t, the number J_0 and smoothness and decay properties of μ, ν that satisfies

$$\delta_{\mathsf{EPI},t}(\mu,\nu) \geq \Theta(D(\mu\|\gamma)).$$

The construction of the function Θ relies on a compactness argument, and therefore is non-explicit. As such, it is again not directly comparable to our results. However, it does settle cases of equality.

C. Instability of the EPI: An Example

As a counterpoint to Theorem 3 and to provide justification for the regularity assumptions therein, we observe that there are probability measures that satisfy the hypotheses required in Theorem 3 on sets of measure arbitrarily close to one, but severely violate its conclusion.

Proposition 10: There is a family of probability measures $(\mu_\epsilon)_{\epsilon>0}$ on $\mathbb{R}$ with finite and uniformly bounded entropies and second moments such that

1) *The measures μ_ϵ essentially satisfy* (4) *for $\eta = 1$. That is, $\lim_{\epsilon\downarrow 0}\mu_\epsilon(\Omega_\epsilon) = 1$, where $\Omega_\epsilon := \{x \mid -\frac{d^2}{dx^2}\log f_\epsilon(x) \geq 1\}$ with $d\mu_\epsilon = f_\epsilon dx$.*
2) *The measures μ_ϵ saturate the EPI as ϵ approaches zero. That is, $\lim_{\epsilon\downarrow 0}\delta_{\mathsf{EPI},t}(\mu_\epsilon,\mu_\epsilon) = 0$ for all $t \in (0,1)$.*
3) *The measures μ_ϵ are bounded away from Gaussians in the W_2 metric; specifically, $\liminf_{\epsilon\downarrow 0}\inf_{\gamma_0\in\Gamma} W_2^2(\mu_\epsilon,\gamma_0) > 1/3$.*

We remark that the measures $(\mu_\epsilon)_{\epsilon>0}$ in the proposition are not necessarily pathological. In fact, it suffices to consider simple Gaussian mixtures that approximate a Gaussian measure, albeit with heavy tails. Moreover, the result can be trivially extended to arbitrary dimension by considering product measures.

Proof: Define the density f_ϵ as

$$f_\epsilon(x) = \epsilon\frac{\sqrt{\epsilon}}{\sqrt{\pi}}e^{-\epsilon x^2} + (1-\epsilon)\frac{\sqrt{1-\epsilon}}{\sqrt{\pi}}e^{-(1-\epsilon)x^2}. \quad (12)$$

Evidently, f_ϵ is a Gaussian mixture having unit variance; the mixture components have variance $(2\epsilon)^{-1}$ and $(2(1-\epsilon))^{-1}$, respectively.

Proof of 1: On any interval, as $\epsilon \downarrow 0$, the densities $(f_\epsilon)_{\epsilon>0}$ and their derivatives converge uniformly to those of the Gaussian density with variance $1/2$. Therefore,

$$-\lim_{\epsilon\downarrow 0} f''_\epsilon(x) = 2 \quad \forall x \in \mathbb{R}. \quad (13)$$

Since the measures μ_ϵ converge weakly to a Gaussian measure with variance 1/2, it is straightforward to conclude that $\lim_{\epsilon\downarrow 0}\mu_\epsilon(\Omega_\epsilon) = 1$, where Ω_ϵ is defined as in the statement of the proposition.

Proof of 2: This follows immediately from [26, Th. 1] due to pointwise convergence of uniformly bounded densities and uniformly bounded second moments.

Proof of 3: Let $m_p(\mu)$ denote the p^{th} absolute moment associated with μ. For conjugate exponents $p, q \geq 1$, Hölder's inequality implies that

$$W_2^2(\mu_\epsilon, \gamma_s) \geq s + 1 - 2m_p^{1/p}(\mu_\epsilon)m_q^{1/q}(\gamma_s), \tag{14}$$

where γ_s is the Gaussian measure with variance s. Indeed, for $X \sim \mu_\epsilon$ and $Z \sim \gamma_s$, there is a coupling between X and Z such that

$$\begin{aligned} W_2^2(\mu_\epsilon, \gamma_s) &= \mathbb{E}|X - Z|^2 = \mathbb{E}[Z^2] + \mathbb{E}[X^2] - 2\mathbb{E}[XZ] \\ &= 1 + s - 2\mathbb{E}[XZ], \end{aligned}$$

where the second equality follows since X and Z have second moments 1 and s, respectively. Now, Hölder's inequality yields $\mathbb{E}[XZ] \leq \mathbb{E}[|X|^p]^{1/p}\mathbb{E}[|Z|^q]^{1/q} \equiv m_p^{1/p}(\mu_\epsilon)m_q^{1/q}(\gamma_s)$, giving (14).

Recall the characterization of Gaussian moments

$$m_q(\gamma_s) = \sqrt{\frac{(2s)^q}{\pi}}\Gamma\left(\frac{q+1}{2}\right)$$

for $s, q > 0$. Let us now specialize (14) by choosing $p = 3/2$ and $q = 3$. In this case,

$$\begin{aligned} W_2^2(\mu_\epsilon, \gamma_s) &\geq 1 + s - 2m_{3/2}^{2/3}(\mu_\epsilon)\sqrt{\frac{2\ s}{\pi^{1/3}}} \\ &\geq 1 - \frac{2}{\pi^{1/3}}m_{3/2}^{4/3}(\mu_\epsilon), \end{aligned}$$

where the first inequality is (14) and the second inequality follows by minimizing over $s > 0$.

By the dominated convergence theorem, we have

$$\lim_{\epsilon\downarrow 0} m_{3/2}(\mu_\epsilon) = m_{3/2}(\gamma_{1/2}) = \sqrt{\frac{1}{\pi}}\Gamma(5/4).$$

So, putting everything together, we may conclude

$$\begin{aligned} \liminf_{\epsilon\downarrow 0} \inf_{\gamma_0\in\Gamma} W_2^2(\mu_\epsilon, \gamma_0) &= \liminf_{\epsilon\downarrow 0} \inf_{s>0} W_2^2(\mu_\epsilon, \gamma_s) \\ &\geq 1 - \frac{2}{\pi^{1/3}}\left(\sqrt{\frac{1}{\pi}}\Gamma(5/4)\right)^{4/3} \\ &= 1 - \frac{2}{\pi}\left(\Gamma(5/4)\right)^{4/3} \\ &\approx 0.441562 > 1/3. \end{aligned}$$

Note that the first line follows since μ_ϵ has mean zero, and therefore the minimization over all Gaussian measures may be restricted to minimization over those with mean zero. □

Remark 11: Our construction of f_ϵ is closely related to the counterexamples proposed by Bobylev and Cercignani in their disproof of Cercignani's conjecture on entropy production in the Boltzmann equation [27]. This construction also appeared in the context of the Boltzmann equation in [28, Proposition 23].

III. Discussion and Proofs

The remainder of this paper makes use of ideas from optimal mass transport. All necessary ingredients are highlighted as they are needed. The unfamiliar reader is directed to the references, as well as the comprehensive introductions [29], [30]. Although not related to the present paper, we remark that optimal transport and information theory together play an important role in establishing concentration of measure; see, e.g., [31]–[33] and references therein. In probabilistic terms, the theory of optimal transport systematically investigates the problem of *transporting* one probability measure μ to another probability measure ν. To this end, a map $T: \mathbb{R}^n \to \mathbb{R}^n$ is said to transport a probability measure μ to another probability measure ν (both on $\mathbb{R}^n$) if the pushforward of μ under T is ν (i.e., $\nu = T\#\mu$). That is,

$$\int f \circ T d\mu = \int f d\nu$$

for all continuous bounded functions $f: \mathbb{R}^n \to \mathbb{R}$. In the language of probability, if X has law μ and T transports μ to ν, then the random variable $T(X)$ has law ν. Under general conditions (e.g., μ is absolutely continuous w.r.t. the Lebesgue measure), valid transportation maps are guaranteed to exist. More remarkably, we are often ensured the existence of transportation maps with useful structure. For example, in dimension 1, if F_μ, F_ν denote the cumulative distribution functions for μ and ν, respectively, then $T = F_\nu^{-1} \circ F_\mu$ transports μ to ν, provided F_μ, F_ν are strictly increasing. This construction results in a transportation map T that is monotone increasing; in other words, it is realized as the derivative of a convex function. The spirit of this construction generalizes to any dimension, and is a cornerstone result of optimal transport theory:

Theorem 12 (Brenier-McCann, [29], [34], [35]): Consider two probability measures μ, ν on $\mathbb{R}^n$, and assume that μ is absolutely continuous with respect to the Lebesgue measure. There exists a unique map T (which we shall call the **Brenier map***) transporting μ onto ν that arises as the gradient of a convex lower semicontinuous function. Moreover, this map is such that*

$$W_2^2(\mu, \nu) = \mathbb{E}[|X - T(X)|^2],$$

where X has law μ, and therefore $T(X)$ has law ν. In other words, $(X, T(X))$ is an optimal coupling for the distance W_2.

The starting point for the proofs of our main results comes from a recent paper of Rioul [36]. Through an impressively short sequence of direct but carefully chosen steps, Rioul recently gave a new proof of the EPI based on transportation of measures. From his proof, we may readily distill the following:

Lemma 13: Let $T_1: \mathbb{R}^n \to \mathbb{R}^n$ and $T_2: \mathbb{R}^n \to \mathbb{R}^n$ be diffeomorphisms satisfying $\mu = T_1\#\gamma$ and $\nu = T_2\#\gamma$. If μ and ν have finite entropies, then

$$\delta_{\mathsf{EPI},t}(\mu, \nu) \geq \mathbb{E}\log\frac{|\det(t\nabla T_1(X^*) + (1-t)\nabla T_2(Y^*))|}{|\det(\nabla T_1(X^*))|^t|\det(\nabla T_2(Y^*))|^{1-t}}, \tag{15}$$

where $X^* \sim \gamma$ and $Y^* \sim \gamma$ are independent.

Remark 14: For a vector-valued map $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n): \mathbb{R}^n \to \mathbb{R}^n$, we write $\nabla\Phi$ to denote its Jacobian. That is, $(\nabla\Phi(x))_{ij} = \frac{\partial}{\partial x_i}\Phi_j(x)$.

In words, (15) shows that the deficit in the EPI can always be bounded from below by a function of the Jacobians ∇T_1 and ∇T_2, where T_1 and T_2 are are invertible and differentiable maps that transport measures γ to μ and γ to ν, respectively.

When T_1 and T_2 are Knöthe maps (see [30], [37], [38]), the Jacobians ∇T_1 and ∇T_2 are upper triangular matrices with positive diagonal entries. Using this property, Rioul concludes $\delta_{\mathsf{EPI},t}(\mu, \nu) \geq 0$ using concavity of the logarithm applied to the eigenvalues (diagonal entries) of ∇T_1 and ∇T_2. By strict concavity of the logarithm, saturation of this inequality implies that the diagonal entries of ∇T_1 and ∇T_2 must be equal almost everywhere. Combining this information with the fact that a relative entropy term (omitted above) must vanish, Rioul recovers the well known necessary and sufficient conditions for $\delta_{\mathsf{EPI},t}(\mu, \nu)$ to vanish. Specifically, μ and ν must be Gaussian measures, equal up to translation.

Instead of the Knöthe map, our proofs will use the Brenier map (defined in Theorem 12), which is always realized as the gradient of a convex function. Toward this end, departing from Rioul's argument based on Knöthe maps, if T_1 and T_2 are taken instead to be Brenier maps (again, transporting γ to μ and γ to ν, respectively), then the Jacobians ∇T_1 and ∇T_2 are symmetric positive definite by the Brenier-McCann Theorem. Thus, concavity of the log-determinant function on the positive semidefinite cone immediately gives the EPI from (15). Moreover, by strict concavity of the log-determinant function, equality in the EPI implies that $\nabla T_1(X^*) = \nabla T_2(Y^*)$ almost everywhere, and are thus constant. Hence, T_1 and T_2 are necessarily affine functions, identical up to translation. This immediately implies that $\delta_{\mathsf{EPI},t}(\mu, \nu) = 0$ only if μ, ν are Gaussian measures with identical covariances.

Unfortunately, while both arguments easily settle cases of equality in the EPI, neither yield quantitative stability estimates. However, we note that the Brenier map is generally better suited for establishing quantitative stability in functional inequalities. Indeed, it was remarked by Figalli, Maggi and Pratelli in their comparison to Gromov's proof of the isoperimetric inequality that the Brenier map is generally more efficient than the Knöthe map in establishing quantitative stability estimates due to its rigid structure [39]. We shall fruitfully exploit the properties of the Brenier map in our proof of Theorem 3.

A. Proof of Theorem 3

The proof of Theorem 3 is short, but makes use of several foundational results from the theory of optimal transport. We will also need the following lemma; a proof can be found in the Appendix.

Lemma 15: For positive definite matrices A, B and $t \in [0, 1]$, we have

$$\log\det(tA + (1-t)B) \geq t\log\det(A) + (1-t)\log\det(B) + \frac{t(1-t)}{2\max\{\lambda^2_{\max}(A), \lambda^2_{\max}(B)\}} \times \|A - B\|_F^2,$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue.

In addition, we remind the reader that a random vector X having a log-concave density enjoys (i) finite second moment (in fact, finite moments of all orders); and (ii) finite entropy $h(X)$. Since Theorem 3 requires log-concave densities, these conditions will be implicitly assumed throughout the proof.

Proof of Theorem 3: Assume first that the densities $e^{-\varphi}$ and $e^{-\psi}$ are smooth and strictly positive on $\mathbb{R}^n$. Also, let $X^* \sim \gamma$ and $Y^* \sim \gamma$ be independent. Define T_1 to be the Brenier map transporting γ to μ, and let T_2 denote the Brenier map transporting γ to ν. We recall here that a Brenier map is always the gradient of a convex function by the Brenier-McCann theorem, and therefore ∇T_2 and ∇T_1 are positive semidefinite since they coincide with Hessians of convex functions. In fact, since all densities involved are non-vanishing, they are positive definite. Moreover, when the densities are strictly positive on the whole space, we know by results of Caffarelli [40], [41] that the maps T_1 and T_2 are C^1-smooth.

Using the assumed smoothness and convexity of the potentials φ and ψ, Caffarelli's contraction theorem (see [42] and, e.g., [29, Th. 9.14]) implies that T_1 and T_2 are 1-Lipschitz, so that $\lambda_{\max}(\nabla T_1) \leq 1$ and $\lambda_{\max}(\nabla T_2) \leq 1$. Therefore, since ∇T_2 and ∇T_1 are positive definite, Lemma 15 yields the following (pointwise) estimate

$$\begin{aligned}
&\log\det(t\nabla T_1 + (1-t)\nabla T_2) \\
&\quad \geq t\log\det(\nabla T_1) + (1-t)\log\det(\nabla T_2) \\
&\qquad + \frac{t(1-t)}{2}\|\nabla T_1 - \nabla T_2\|_F^2.
\end{aligned}$$

Combined with (15) we obtain:

$$\begin{aligned}
&\delta_{\mathsf{EPI},t}(\mu, \nu) \\
&\quad \geq \frac{t(1-t)}{2}\mathbb{E}\|\nabla(T_1(X^*) - X^*) - \nabla(T_2(Y^*) - Y^*)\|_F^2.
\end{aligned}$$

Now, define matrices $A = \mathbb{E}[\nabla(T_1(X^*) - X^*)]$ and $B = \mathbb{E}[\nabla(T_2(Y^*) - Y^*)]$. By orthogonality, we have

$$\begin{aligned}
&\mathbb{E}\|\nabla(T_1(X^*) - X^*) - \nabla(T_2(Y^*) - Y^*)\|_F^2 \\
&= \mathbb{E}\|\nabla(T_1(X^*) - (I+A)X^*) - \nabla(T_2(Y^*) - (I+B)Y^*)\|_F^2 \\
&\quad + \|A - B\|_F^2 \\
&= \mathbb{E}\|\nabla(T_1(X^*) - (I+A)X^*)\|_F^2 \\
&\quad + \mathbb{E}\|\nabla(T_2(Y^*) - (I+B)Y^*)\|_F^2 + \|A - B\|_F^2 \\
&\geq \mathbb{E}|T_1(X^*) - (I+A)X^*|^2 \\
&\quad + \mathbb{E}|T_2(Y^*) - (I+B)Y^*|^2 + \|(I+A) - (I+B)\|_F^2 \\
&= \mathbb{E}|T_1(X^*) - (I+A)X^*|^2 \\
&\quad + \mathbb{E}|T_2(Y^*) - (I+B)Y^*|^2 + \mathbb{E}|(I+A)X^* - (I+B)X^*|^2
\end{aligned}$$

The final inequality is due to the Gaussian Poincaré inequality $\int |f|^2\, d\gamma \leq \int |\nabla f|^2\, d\gamma$, holding for every C^1-smooth $f: \mathbb{R}^n \to \mathbb{R}$ with mean zero. Indeed, its application is justified by C^1-smoothness of the Brenier maps among log-concave distributions, and the identity

$$\mathbb{E}[T_1(X^*) - (I+A)X^*] = \int x\, d\mu(x) - (I+A)\int x\, d\gamma(x) = 0,$$

which holds similarly for Y^*. The desired inequality now follows from the definition of W_2. Indeed, let γ_A and γ_B denote the laws of $(I+A)X^*$ and $(I+B)Y^*$, respectively. Evidently, γ_A and γ_B are both Gaussian measures. Since $T_1(X^*) \sim \mu$ and $T_2(Y^*) \sim \nu$ by construction,

$$\mathbb{E}|T_1(X^*) - (I+A)X^*|^2 \geq W_2^2(\mu, \gamma_A)$$

and similarly,

$$\mathbb{E}|T_2(Y^*) - (I+B)Y^*|^2 \geq W_2^2(\nu, \gamma_B).$$

Since X^*, Y^* are equal in distribution, $(I+B)X^*$ also has law γ_B, so that

$$\mathbb{E}|(I+A)X^* - (I+B)X^*|^2 \geq W_2^2(\gamma_A, \gamma_B).$$

Summarizing above, we may conclude

$$\begin{aligned}
&\delta_{\mathsf{EPI},t}(\mu, \nu) \\
&\geq \frac{t(1-t)}{2}\mathbb{E}\|\nabla(T_1(X^*) - X^*) - \nabla(T_2(Y^*) - Y^*)\|_F^2 \\
&\geq \frac{t(1-t)}{2}\Big(W_2^2(\mu, \gamma_A) + W_2^2(\nu, \gamma_B) + W_2^2(\gamma_A, \gamma_B)\Big) \\
&\geq \frac{t(1-t)}{2}\inf_{\gamma_1,\gamma_2\in\Gamma}\Big(W_2^2(\mu, \gamma_1) + W_2^2(\nu, \gamma_2) + W_2^2(\gamma_1, \gamma_2)\Big).
\end{aligned}$$

Thus, Theorem 3 holds when the densities are smooth and positive everywhere. If this is not the case, we may first convolve μ, ν with a Gaussian having small variance so that the resulting densities will be both smooth and positive. The general result then follows by considering arbitrarily small perturbations, which is justified by Proposition 16, found below. □

Proposition 16: For a probability measure μ and $s > 0$, let μ_s denote the probability measure corresponding to $X + \sqrt{s}Z$, where $X \sim \mu$ and $Z \sim \gamma$ are independent. If probability measures μ, ν have finite second moments, and

$$\Delta(\mu, \nu) := \inf_{\gamma_1,\gamma_2\in\Gamma}\Big(W_2^2(\mu, \gamma_1) + W_2^2(\nu, \gamma_2) + W_2^2(\gamma_1, \gamma_2)\Big),$$

then $\lim_{s\downarrow 0}\Delta(\mu_s, \nu_s) = \Delta(\mu, \nu)$. Moreover, if μ, ν have densities, then $\lim_{s\downarrow 0}\delta_{\mathsf{EPI},t}(\mu_s, \nu_s) = \delta_{\mathsf{EPI},t}(\mu, \nu)$. Finally, if μ satisfies (4) with parameter $\eta > 0$, then μ_s satisfies (4) with parameter $\eta/(1 + s\eta)$.

Proof: Evidently, if γ_s denotes the Gaussian measure with covariance matrix $s\mathrm{I}$, then $\mu_s = \mu * \gamma_s$. Observe that for any fixed Gaussian measures γ_1, γ_2 (not to be confused with γ_s),

$$\begin{aligned}
&W_2^2(\mu, \gamma_1) + W_2^2(\nu, \gamma_2) + W_2^2(\gamma_1, \gamma_2) \\
&\quad\geq W_2^2(\mu * \gamma_s, \gamma_1 * \gamma_s) + W_2^2(\nu * \gamma_s, \gamma_2 * \gamma_s) \\
&\qquad + W_2^2(\gamma_1 * \gamma_s, \gamma_2 * \gamma_s) \\
&\quad\geq \Delta(\mu_s, \nu_s),
\end{aligned}$$

so that $\Delta(\mu, \nu) \geq \Delta(\mu_s, \nu_s)$. It remains to prove an inequality in the reverse direction.

Since W_2 is a metric on the space of probability measures with finite second moments, the reverse triangle inequality implies that $W_2(\mu, \nu) \geq |W_2(\mu, \delta_0) - W_2(\nu, \delta_0)|$, where δ_0 is the Dirac measure at 0. Squaring both sides, it follows that $W_2^2(\mu, \nu) \geq (\sqrt{\mathbb{E}|X|^2} - \sqrt{\mathbb{E}|Y|^2})^2$ for $X \sim \mu$ and $Y \sim \nu$. Therefore, there is $K = K(\mu, \nu) < \infty$ depending only on the second moments of μ and ν such that

$$\Delta(\mu, \nu) = \inf_{\gamma_1,\gamma_2\in\Gamma_K}\Big(W_2^2(\mu, \gamma_1) + W_2^2(\nu, \gamma_2) + W_2^2(\gamma_1, \gamma_2)\Big),$$

where $\Gamma_K \subset \Gamma$ denotes the set of centered Gaussian measures with second moments bounded by K. From this, it is straightforward to argue that for s sufficiently small,

$$\begin{aligned}
&\Delta(\mu_s, \nu_s) \\
&\geq \inf_{\gamma_1,\gamma_2\in\Gamma}\Big(W_2^2(\mu, \gamma_1) + W_2^2(\nu, \gamma_2) + W_2^2(\gamma_1, \gamma_2)\Big) - \sqrt{s}C,
\end{aligned} \tag{16}$$

where $C = C(\mu, \nu) < \infty$ depends only on the second moments of μ and ν. Thus, the first part of the claim follows.

The second claim follows immediately from the fact that $\lim_{s\downarrow 0} h(\mu_s) = h(\mu)$ for any μ having density and finite second moment. See, e.g., [6, Lemma 1.2].

The third claim can be found in [9, Th. 3.7(b)]. □

Remark 17: The assumption of log-concavity is mainly used to ensure that the optimal transport map is Lipschitz. This can sometimes still be the case in other situations. For example, the recent work [43] shows that this property holds for certain families of bounded perturbations of the Gaussian measure, including the radially symmetric case.

B. Proof of Corollary 7

Before proving Corollary 7, let us briefly review the formal equivalence between the different forms of the EPI. To this end, let X, Y be independent random vectors in $\mathbb{R}^n$ whose entropies exist. Let $t \in (0, 1)$ and observe that the Shannon-Stam inequality (1) together with convexity of $x \mapsto e^x$ gives

$$\begin{aligned}
&\exp(2\, h(\sqrt{t}X + \sqrt{1-t}Y)/n) \\
&\quad\geq \exp(2\, h(\sqrt{t}X)/n) + \exp(2\, h(\sqrt{1-t}Y)/n) \\
&\quad= t\exp(2\, h(X)/n) + (1-t)\exp(2\, h(Y)/n) \\
&\quad\geq \exp(2(th(X) + (1-t)h(Y))/n).
\end{aligned}$$

Taking logarithms reveals that the Shannon-Stam inequality implies Lieb's inequality:

$$h(\sqrt{t}X + \sqrt{1-t}Y) \geq th(X) + (1-t)h(Y).$$

However, Lieb argued that a simple rescaling allows one to make the reverse deduction. In particular, define the random vectors $\tilde{X} = t^{-1/2}X$ and $\tilde{Y} = (1-t)^{-1/2}Y$. In this case, we have by Lieb's inequality

$$h(X+Y) = h(\sqrt{t}\tilde{X} + \sqrt{1-t}\tilde{Y}) \geq th(\tilde{X}) + (1-t)h(\tilde{Y}) \tag{17}$$

for all $t \in [0, 1]$. In particular, we may choose t to satisfy $t/(1-t) = N(X)/N(Y)$, in which case

$$\begin{aligned}
&th(\tilde{X}) + (1-t)h(\tilde{Y}) \\
&\quad= th(X) + (1-t)h(Y) - \frac{n}{2}(t\log t + (1-t)\log(1-t)) \\
&\quad= \frac{n}{2}(\exp(2h(X)/n) + \exp(2h(Y)/n)).
\end{aligned} \tag{18}$$

Thus, substituting into (17), we obtain

$$h(X+Y) \geq \frac{n}{2}(\exp(2\, h(X)/n) + \exp(2\, h(Y)/n)),$$

which is precisely Shannon-Stam inequality (1).

Proof: Let X, Y and $\tilde{X}, \tilde{Y}$ be as above. If $X \sim \mu$ satisfies (4) with parameter η_μ, then by a change of variables the density of $\tilde{X}$ satisfies (4) with parameter $t\eta_\mu$. A similar statement holds for $\tilde{Y}$. Thus, both $\tilde{X}$ and $\tilde{Y}$ satisfy (4) with parameter $\min\{t\eta_\mu, (1-t)\eta_\nu\}$.

Next, let $\tilde{\mu}$ and $\tilde{\nu}$ denote the laws of $\tilde{X}$ and $\tilde{Y}$, respectively. We note that $t(1-t)\mathsf{d}^2_{W_2}(\tilde{\mu}) = (1-t)\mathsf{d}^2_{W_2}(\mu)$ and $t(1-t)\mathsf{d}^2_{W_2}(\tilde{\nu}) = t\mathsf{d}^2_{W_2}(\nu)$ by a simple rescaling. Also, we have the following lower bound on $W_2^2(\tilde{\mu}, \tilde{\nu})$:

$$t(1-t)W_2^2(\tilde{\mu}, \tilde{\nu}) \geq \mathsf{d}^2_F(\mu, \nu).$$

This follows from rescaling, the fact that W_2 is non-increasing under rescaled convolution, the central limit theorem, weak continuity of W_2, the identity for W_2 distance between Gaussian measures mentioned in the proof of Theorem 3, and finally the definition of d^2_F in (6).

Now, using the above identities and applying (5), we have

$$\begin{aligned}
&h(X+Y)\\
&= h(\sqrt{t}\tilde{X} + \sqrt{1-t}\tilde{Y})\\
&\geq th(\tilde{X}) + (1-t)h(\tilde{Y})\\
&\quad + \min\{t\eta_\mu, (1-t)\eta_\nu\}\frac{t(1-t)}{8}\Big(\mathsf{d}^2_{W_2}(\tilde{\mu})\\
&\qquad\qquad + \mathsf{d}^2_{W_2}(\tilde{\nu}) + W_2^2(\tilde{\mu}, \tilde{\nu})\Big)\\
&\geq th(\tilde{X}) + (1-t)h(\tilde{Y})\\
&\quad + \frac{\min\{t\eta_\mu, (1-t)\eta_\nu\}}{8}\Big((1-t)\mathsf{d}^2_{W_2}(\mu)\\
&\qquad\qquad + t\mathsf{d}^2_{W_2}(\nu) + \mathsf{d}^2_F(\mu, \nu)\Big).
\end{aligned}$$

Now, we imitate Lieb's argument and choose t to satisfy $t/(1-t) = N(X)/N(Y)$. In this case, the identity (18) holds, giving

$$\begin{aligned}
&h(X+Y)\\
&\geq \frac{n}{2}\left(\exp(2\ h(X)/n) + \exp(2\ h(Y)/n)\right)\\
&\quad + \frac{\min\{t\eta_\mu, (1-t)\eta_\nu\}}{8}\Big((1-t)\mathsf{d}^2_{W_2}(\mu) + t\mathsf{d}^2_{W_2}(\nu)\\
&\qquad\qquad + \mathsf{d}^2_F(\mu, \nu)\Big)
\end{aligned}$$

Multiplying through by $2/n$, taking exponents and relabeling $t \leftarrow \theta$ completes the proof. □

C. Proof of Theorem 8

Proof of Theorem 8: Assume that $X^* \sim \gamma$, and let T be the Brenier map sending γ onto μ. For convenience, let us define random variables $(\lambda_i)_{1\leq i\leq n}$ as the eigenvalues of $\nabla T(X^*)$ in increasing order (so that $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ a.s.). The combination of (15) and Lemma 15 yields in this case

$$\begin{aligned}
\delta_{\mathsf{EPI},t}(\mu, \gamma) &\geq \frac{t(1-t)}{2}\mathbb{E}\left[\frac{\|\nabla T(X^*) - \mathrm{I}\|_F^2}{1 + \lambda_{\max}(\nabla T(X^*))^2}\right]\\
&= \frac{t(1-t)}{2}\mathbb{E}\left[\frac{\sum_{i=1}^n(\lambda_i - 1)^2}{1 + \lambda_n^2}\right]. \qquad (19)
\end{aligned}$$

The L^1 Poincaré inequality for the Gaussian measure (cf. [44, Proposition 1.8]) states that

$$\int_{\mathbb{R}^n} |f| d\gamma \leq 2\int_{\mathbb{R}^n} |\nabla f| d\gamma \qquad (20)$$

for every smooth mean zero $f : \mathbb{R}^n \to \mathbb{R}$. See the discussion in Section IV for more information about L^1 Poincaré inequalities and their relation to Cheeger inequalities.

Since $(T(X^*), X^*)$ is a valid coupling between μ and γ, we have

$$W_{1,1}(\mu, \gamma) \leq \mathbb{E}\|T(X^*) - X^*\|_1 = \sum_{i=1}^n \mathbb{E}|T_i(X^*) - X_i^*|.$$

Smoothness of the Brenier map among log-concave densities and the fact that $T(X^*) - X^*$ is zero mean justifies application of (20). This in conjunction with the $\ell_1 - \ell_2$ bound $\|x\|_1 \leq \sqrt{n}\|x\|_2$ for $x \in \mathbb{R}^n$ gives the estimate

$$\begin{aligned}
W_{1,1}(\mu, \gamma) &\leq 2\sum_{i=1}^n \mathbb{E}|\nabla(T_i(X^*) - X_i^*)|\\
&\leq \sqrt{n}2\mathbb{E}\|\nabla T(X^*) - I)\|_F\\
&= \sqrt{n}2\mathbb{E}\left[\sqrt{\sum(\lambda_i - 1)^2}\right].
\end{aligned}$$

Hence in this situation, if we have an L^2 bound on the largest eigenvalue of ∇T, we can deduce a $W_{1,1}$ estimate on the deficit (in contrast to using a uniform bound as in the proof of Theorem 3).

A result of Kolesnikov asserts that $\mathbb{E}[\lambda_n^2] \leq \frac{3}{2}(\mathbb{E}[\lambda_n])^2$ (see [45, Th. 6.1] and the discussion at the top of page 1526). Moreover,

$$\mathbb{E}[\lambda_n] \leq 1 + \mathbb{E}[|\lambda_n - 1|] \leq 1 + \mathbb{E}\left[\sqrt{\sum(\lambda_i - 1)^2}\right].$$

From this estimate we deduce

$$\begin{aligned}
&\mathbb{E}\left[\sqrt{\sum(\lambda_i - 1)^2}\right]\\
&\quad\leq \sqrt{4 + 3\mathbb{E}\left[\sqrt{\sum(\lambda_i - 1)^2}\right]^2}\sqrt{\frac{2}{t(1-t)}\delta_{\mathsf{EPI},t}(\mu, \gamma)}.
\end{aligned}$$

Since $r/\sqrt{1+r^2} \geq c\min(r, 1)$ and $2\mathbb{E}\left[\sqrt{\sum(\lambda_i - 1)^2}\right] \geq n^{-1/2}W_{1,1}(\mu, \gamma)$, we conclude

$$\sqrt{2\delta_{\mathsf{EPI},t}(\mu, \gamma)/(t(1-t))} \geq C\min(n^{-1/2}W_{1,1}(\mu, \gamma), 1),$$

and the result follows.

D. Proof of Theorem 9

For a probability measure ν, let $I(\nu\|\gamma)$ and $D(\nu\|\gamma)$ denote the relative Fisher information and relative entropy, respectively, of ν with respect to the standard Gaussian measure γ on $\mathbb{R}^n$. Since we shall only be concerned with γ as reference measure, we opt henceforth for the more compact notation $I(\nu) := I(\nu\|\gamma)$ and $D(\nu) := D(\nu\|\gamma)$ to simplify the expressions in this section. Thus, if h is the density of ν with respect to γ, then

$$D(\nu) := D(\nu\|\gamma) = \int h \log h \, d\gamma$$

and

$$I(\nu) := I(\nu\|\gamma) = \int \frac{|\nabla h|^2}{h} \quad d\gamma,$$

provided h is sufficiently smooth.

If $X \sim \nu$ and $Z \sim \gamma$ are independent, we let ν_t denote the law of $e^{-t}X + (1 - e^{-2t})^{1/2}Z$. In other words, ν_t is the evolution of ν at time t along the Ornstein-Uhlenbeck process. The following integral form of de Bruijn identity is classical [6], [46], [47]: If ν is centered and has finite second moment, then

$$D(\nu) = \int_0^\infty I(\nu_s) ds. \tag{21}$$

In the proof of [10, Th. 1], Fathi, Indrei and Ledoux established the following inequality:

Theorem 18: If ν is a centered probability measure with spectral gap $\lambda > 0$, then for all $t \geq 0$

$$I(\nu_t) \leq e^{-2t} I(\nu) \frac{1}{1 + \lambda(e^{2t} - 1)}.$$

We also note that the convolution inequality for Fisher information implies that $I(\nu_t) \leq e^{-2t} I(\nu)$, so Theorem 18 yields a stability result for the exponential decay of Fisher information along the Ornstein-Uhlenbeck process, provided the starting measure has positive spectral gap.

We now aim to establish the following Corollary of (18), from which Theorem 9 will follow.

Corollary 19: If ν is a centered probability measure with spectral gap λ, then for all $t \geq 0$

$$D(\nu_t) \leq e^{-2t} D(\nu) \max\left(\frac{1}{1 + \lambda(e^{2t} - 1)}, e^{-2t}\right)$$

Proof: In their proof of [10, Th. 1], Fathi, Indrei and Ledoux also noted that if ν has spectral gap λ, then ν_s has spectral gap at least

$$\lambda_s = \frac{\lambda e^{2s}}{1 + \lambda(e^{2s} - 1)}$$

So, as a consequence of this and Theorem 18, we have

$$\begin{aligned} I(\nu_{t+s}) &\leq e^{-2t} I(\nu_s) \frac{1}{1 + \lambda_s(e^{2t} - 1)} \\ &= e^{-2t} I(\nu_s) \frac{1 + \lambda(e^{2s} - 1)}{1 + \lambda(e^{2(t+s)} - 1)}. \end{aligned} \tag{22}$$

By de Bruijn's identity[4]:

$$\int_0^\infty I(\nu_{t+s}) ds = D(\nu_t),$$

so we need only integrate the RHS of (22). Using the differential form of de Bruijn's identity $\frac{d}{ds} I(\nu_s) = -D(\nu_s)$, integration by parts gives

$$\begin{aligned} &\int_0^\infty I(\nu_s) \frac{1 + \lambda(e^{2s} - 1)}{1 + \lambda(e^{2(t+s)} - 1)} ds \\ &\quad = -\left[D(\nu_s) \frac{1 + \lambda(e^{2s} - 1)}{1 + \lambda(e^{2(t+s)} - 1)}\right]_0^\infty + \int_0^\infty D(\nu_s) u'(s) ds \\ &\quad = D(\nu) \frac{1}{1 + \lambda(e^{2t} - 1)} + \int_0^\infty D(\nu_s) u'(s) ds, \end{aligned} \tag{23}$$

where, for convenience, we have defined

$$u(s) := \frac{1 + \lambda(e^{2s} - 1)}{1 + \lambda(e^{2(t+s)} - 1)}.$$

Now, we note that

$$u'(s) = 2 \frac{e^{2s}(e^{2t} - 1)(\lambda - 1)\lambda}{(1 + \lambda(e^{2(s+t)} - 1))^2}.$$

Thus, if $\lambda \leq 1$, then $u'(s) \leq 0$ and therefore $\int_0^\infty D(\nu_s) u'(s) ds \leq 0$. Hence, (22) together with (23) and de Bruijn's identity yields

$$D(\nu_t) \leq D(\nu) \frac{1}{1 + \lambda(e^{2t} - 1)}.$$

On the other hand, if $\lambda \geq 1$, then $u'(s) \geq 0$, and we may bound

$$\begin{aligned} &\int_0^\infty D(\nu_s) u'(s) ds \\ &\leq D(\nu) \int_0^\infty e^{-2s} u'(s) ds \\ &= D(\nu) 2(e^{2\ t} - 1)(\lambda - 1)\lambda \int_0^\infty \frac{1}{(1 + \lambda(e^{2(s+t)} - 1))^2} ds \\ &= D(\nu) \frac{\lambda(e^{2\ t} - 1)}{(\lambda - 1)} \left(\frac{\lambda - 1}{1 + \lambda(e^{2\ t} - 1)} + \log\left(1 - \frac{\lambda - 1}{\lambda e^{2t}}\right)\right) \end{aligned}$$

to conclude

$$\begin{aligned} &\int_0^\infty I(\nu_s) \frac{1 + \lambda(e^{2s} - 1)}{1 + \lambda(e^{2(t+s)} - 1)} ds \\ &\quad = D(\nu) \frac{1}{1 + \lambda(e^{2t} - 1)} + \int_0^\infty D(\nu_s) u'|(s) ds \\ &\quad \leq D(\nu) + D(\nu) \frac{\lambda(e^{2\ t} - 1)}{(\lambda - 1)} \log\left(1 - \frac{\lambda - 1}{\lambda e^{2t}}\right) \\ &\quad \leq e^{-2t} D(\nu), \end{aligned}$$

where the last inequality is due to $\log(1 - x) \leq -x$ for $x < 1$. So, as before, (22) together with (23) and de Bruijn's identity yields

$$D(\nu_t) \leq e^{-2t} D(\nu),$$

completing the proof. □

Remark 20: In principle, Corollary 19 should follow via an application of Gronwall's lemma to Theorem 18. However, the integration involved there appears no less tedious than the computations above, so we have opted for the direct proof given here.

[4]Positive spectral gap implies finite moments of all orders, so an explicit assumption of finite second moment is not necessary here.

Now, Theorem 9 follows immediately. In particular if ν is centered and has spectral gap λ, then definitions and Corollary 19 yield the desired inequality

$$\begin{aligned}
&\delta_{\mathsf{EPI},e^{-2t}}(\nu,\gamma) \\
&\quad = e^{-2t}D(\nu) - D(\nu_t) \\
&\quad \geq e^{-2t}(1-e^{-2t})D(\nu)\min\left(\frac{\lambda}{e^{-2t}+\lambda(1-e^{-2t})},1\right) \\
&\quad \geq e^{-2t}(1-e^{-2t})\min(\lambda,1)\,D(\nu).
\end{aligned}$$

IV. Extensions

The proof of Theorem 3 uses the fact that, under the assumptions of uniform log-concavity, the Brenier maps are Lipschitz to bound the square of the largest eigenvalue $\lambda_{\max}^2$ in the deficit estimate for the log-concavity of the determinant. A natural question is whether we can use weaker assumptions on the map and still obtain a deficit estimate for the EPI. It turns out that we can get an estimate, provided the largest eigenvalue of $\nabla T(x)$ grows at most linearly at infinity. We shall later see that in dimension 1, as well as for multi-dimensional radially symmetric measures, this assumption of eigenvalue growth holds as soon as the law of the random variable satisfies a Cheeger isoperimetric inequality, which is a stronger assumption than the spectral gap assumption used for the one-dimensional result of [15], but equivalent in (non-uniformly) log-concave situations.

A first case in which we establish a deficit estimate is when one of the two variables is Gaussian:

Proposition 21: Let μ be a centered probability measure on $\mathbb{R}^n$, and let T be the Brenier map sending the standard Gaussian measure γ onto μ. If T is C^1 and satisfies the pointwise bound $\lambda_{\max}(\nabla T(x)) \leq c\sqrt{1+|x|^2}$ for all x, for some $c > 1$, then

$$\delta_{\mathsf{EPI},t}(\mu,\gamma) \geq \frac{t(1-t)}{8c^2n}W_2^2(\mu,\gamma).$$

This estimate can be compared to Theorem 8. Its advantage is that it involves the more natural W_2 distance, and that its assumptions may be satisfied in certain non-log-concave situations (as we shall later see for one-dimensional random variables). However, it has the downside of sometimes being worse, even for product measures in high dimension, and of requiring more quantitative information on the measure μ, via the constant c in the eigenvalue bound.

Proof: Using the assumption, the combination of (15) and Lemma 15 gives

$$\delta_{\mathsf{EPI},t}(\mu,\gamma) \geq \frac{t(1-t)}{2c^2}\mathbb{E}\left[\frac{\|\nabla T(X^*)-\mathrm{I}\|_F^2}{1+\ |X^*|^2}\right].$$

According to [48, Corollary 5.6] (see also [49] for the one-dimensional case), the standard Gaussian measure in dimension n satisfies the weighted Poincaré inequality

$$\mathbb{E}\left[\frac{|\nabla f(X^*)|^2}{1+|X^*|^2}\right] \geq \frac{1}{4n}\mathrm{Var}(f(X^*)).$$

Applying this result and following the same steps as in the proof of Theorem 3 yields

$$\delta_{\mathsf{EPI},t}(\mu,\gamma) \geq \frac{t(1-t)}{8c^2n}W_2^2(\mu,\gamma).$$

which concludes the proof. ☐

We also prove a lower bound when neither of the two measures are Gaussian, but with an even worse dependence on the dimension:

Proposition 22: Let T_1 and T_2 be the Brenier maps sending Gaussians to X and Y, respectively. Assume that X and Y are centered and that the maps T_i are C^1 and satisfy the pointwise bound $\lambda_{\max}(\nabla T_i(x)) \leq c\sqrt{1+|x|^2}$ for all x, for some $c > 1$. Then there is a universal constant $C > 0$ such that

$$\begin{aligned}
&\delta_{\mathsf{EPI},t}(\mu,\nu) \\
&\geq \frac{Ct(1-t)}{(cn)^2}\inf_{\gamma_1,\gamma_2\in\Gamma}(W_2^2(\mu,\gamma_1)+W_2^2(\nu,\gamma_2)+W_2^2(\gamma_1,\gamma_2)).
\end{aligned}$$

Proof: Let us define

$$\begin{aligned}
c_n &:= \mathbb{E}[(1+|X^*|^2)^{-1}], \\
A &:= c_n^{-1}\mathbb{E}\left[\frac{\nabla T_1(X^*)-\mathrm{I}}{1+|X^*|^2}\right], \\
B &:= c_n^{-1}\mathbb{E}\left[\frac{\nabla T_2(Y^*)-\mathrm{I}}{1+|Y^*|^2}\right].
\end{aligned}$$

As in the proof of Theorem 3, we have

$$\delta_{\mathsf{EPI},t}(\mu,\nu) \geq \frac{t(1-t)}{2c^2}\mathbb{E}\left[\frac{\|\nabla T_1(X^*)-\mathrm{I}-(\nabla T_2(Y^*)-\mathrm{I})\|_F^2}{(1+|X^*|^2)(1+|Y^*|^2)}\right],$$

where we have used the naïve bound $\lambda_{\max}^2 \leq c^2(1+|x|^2)(1+|y|^2)$, which ultimately leads to the worsening of the dependence on the dimension. We then have

$$\begin{aligned}
&\mathbb{E}\left[\frac{\|\nabla T_1(X^*)-\mathrm{I}-(\nabla T_2(Y^*)-\mathrm{I})\|_F^2}{(1+|X^*|^2)(1+|Y^*|^2)}\right] \\
&= \mathbb{E}\left[\frac{\|\nabla T_1(X^*)-\mathrm{I}-c_nA\|_F^2}{(1+|X^*|^2)(1+|Y^*|^2)}\right] \\
&\quad + \mathbb{E}\left[\frac{\|\nabla T_2(Y^*)-\mathrm{I}-c_nB\|_F^2}{(1+|X^*|^2)(1+|Y^*|^2)}\right] + c_n^2\|c_n(A-B)\|_F^2 \\
&\quad - 2\mathbb{E}\left\langle\frac{\nabla T_1(X^*)-\mathrm{I}-c_nA}{1+|X^*|^2},\frac{\nabla T_2(Y^*)-\mathrm{I}-c_nB}{1+|Y^*|^2}\right\rangle \\
&\quad + 2\mathbb{E}\left[\left\langle\frac{c_n(A-B)}{1+|Y^*|^2},\frac{\nabla T_1(X^*)-\mathrm{I}-c_nA}{1+|X^*|^2}\right\rangle\right] \\
&\quad - 2\mathbb{E}\left[\left\langle\frac{c_n(A-B)}{1+|X^*|^2},\frac{\nabla T_2(Y^*)-\mathrm{I}-c_nB}{1+|Y^*|^2}\right\rangle\right] \\
&= c_n\mathbb{E}\left[\frac{\|\nabla T_1(X^*)-(\mathrm{I}+c_nA)\|_F^2}{(1+|X^*|^2)}\right] \\
&\quad + c_n\mathbb{E}\left[\frac{\|\nabla T_2(Y^*)-(\mathrm{I}+c_nB)\|_F^2}{(1+|Y^*|^2)}\right] + c_n^2\|c_nA-c_nB\|_F^2
\end{aligned}$$

and then the proof continues in the same way as for the previous proposition. The constant c_n above is the expectation of $(1+|X^*|^2)^{-1}$, which is of order n^{-1}. ☐

To apply these results, we want to know when does the Brenier map satisfy the eigenvalue bound. We shall prove that for one-dimensional measures and for radially symmetric log concave measures, such an assumption holds when the measure satisfies a certain isoperimetric inequality.

Proposition 23: If the law of X is given by an exponential measure $\mu(dx) = \frac{1}{2}\exp(-|x|)dx$, then the Brenier map T transporting a standard Gaussian random variable onto X satisfies the bound

$$T'(x) \leq c\sqrt{1+x^2}$$

for some $c > 0$

Proof: The optimal map from a measure μ onto a measure ν with positive densities in dimension one is given by $x \longrightarrow F_\nu^{-1}(F_\mu(x))$, where $F_\mu(x) = \mu(]-\infty, x])$ is the distribution function associated with μ. For the exponential measure the distribution function can be explicitly computed as $F_{\exp}(x) = 1 - e^{-x}/2$ for $x \geq 0$ and $e^x/2$ for $x < 0$.

Consider $x > 0$. A direct computation shows that

$$T'(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{2}{1 - F_\gamma(x)}$$

where F_γ is the distribution function of the standard Gaussian measure. There exists a constant c such that $1 - F_\gamma(x) \geq \frac{e^{-x^2/2}}{c\sqrt{1+x^2}}$, and the bound on T' immediately follows. By symmetry, the same bound applies when $x < 0$.

To prove the lower bound on $1 - F_\gamma(x)$, we just use the fact that

$$\begin{aligned} 1 - F_\gamma(x) &= \frac{e^{-x^2/2}}{\sqrt{2\pi}x} - \frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{e^{-t^2/2}}{t^2}dt \\ &\geq \frac{e^{-x^2/2}}{\sqrt{2\pi}x} - \frac{1 - F_\gamma(x)}{x^2}, \end{aligned}$$

and the existence of a suitable constant easily follows. □

Definition 24: A probability measure μ is said to satisfy a Cheeger isoperimetric inequality with constant $\lambda > 0$ if for any measureable set A we have

$$\mu^+(\partial A) \geq \lambda\mu(A)(1 - \mu(A))$$

where $\mu^+(\partial A) := \limsup_{\epsilon\downarrow 0} \frac{\mu(A+B_\epsilon)-\mu(A)}{\epsilon}$, with B_ϵ the ball with center 0 and radius ϵ.

For log-concave measures (i.e., those having a log-concave density), Buser's theorem [50] states that satisfying a Cheeger inequality and a Poincaré inequality is equivalent, up to universal constants (or more precisely, the extension of Buser's theorem to weighted spaces [51]). In general, the Cheeger inequality is stronger than the Poincaré inequality. It is equivalent (up to universal constants) to the L^1 Poincaré inequality $\int |\nabla f| d\mu \geq c \int |f| d\mu$ for every function with average zero. More generally, for log-concave measures the isoperimetric inequality and the exponential concentration property are equivalent [52].

Theorem 25: Assume X is a one-dimensional random variable with positive density and median 0, that satisfies a Cheeger inequality with constant α. Then the optimal map transporting the exponential measure onto X is α^{-1}-Lipschitz.

As a consequence, the optimal map T transporting the Gaussian measure onto the law of X satisfies $T'(x) \leq \frac{c}{\alpha}(1 + |x|^2)$.

This result can be extended to the non-centered case just by translating.

Note that a measure that is the image of the exponential measure by a Lipschitz map necessarily satisfies Cheeger's inequality, so the first part of this statement is actually an equivalence.

Proof: Showing an upper bound on the derivative of the map is the same as proving a lower bound on the derivative of the inverse map $\tilde{T}$ (which is the optimal map sending μ onto the exponential measure). A computation shows that, if we denote by f the density of the law of X, we have for x positive

$$\tilde{T}'(x) = \frac{2f(x)}{1 - F_\mu(x)}.$$

Since μ has median 0, the Cheeger inequality implies that for $x \geq 0$ we get $f(x) \geq \frac{\alpha}{2}(1 - F_\mu(x))$ and the first part of the result immediately follows, after applying the same reasoning for negative x.

The second part can be deduced just by using the fact that in dimension 1 the optimal map from the Gaussian onto μ is the composition of the map from the Gaussian onto the exponential measure with T. □

The same argument can be generalized to radially symmetric random vectors with log-concave density:

Proposition 26: Assume that X is a radially symmetric random vector in $\mathbb{R}^n$, whose law is log-concave and satisfies a Cheeger inequality with constant α. There exists a constant c_n (depending on n, but otherwise independent of X) such that the optimal map T transporting the Gaussian measure onto X satisfies $\lambda_{\max}(\nabla T(x)) \leq c_n\alpha^{-1}\sqrt{1+|x|^2}$.

Remark 27: Bobkov [53] showed that the optimal constant λ in the Poincaré inequality for a radially symmetric log-concave random variable satisfies $n\mathbb{E}[|X|^2]^{-1}/12 \leq \lambda \leq n\mathbb{E}[|X|^2]^{-1}$. Since for log-concave measures the square of the Cheeger constant and the Poincaré constant are equivalent [51], the constant α in the Proposition always exists and is comparable to $\sqrt{n}\mathbb{E}[|X|^2]^{-1/2}$, up to universal constants.

Proof: Let μ_{rad} be the law of $|X|$, and $\tilde{T}$ be the optimal transport sending $\gamma_{\text{rad}} := \frac{r^{n-1}}{\sqrt{2\pi}}e^{-r^2/2}dr$ onto μ_{rad}. The Brenier optimal map sending the Gaussian onto the law of X is then given by $x \longrightarrow \tilde{T}(|x|)x/|x|$. This can be checked by verifying that it sends the Gaussian measure onto the law of X (which is a simple change of variable argument) and that it is the gradient of the convex function $H(|x|)$ with $H' = \tilde{T}$. Since the Brenier map is the only transport map that arises as the gradient of a convex function, T is necessarily the Brenier map.

We then compute the gradient of the map T, which is given by $\nabla T(x)v = \left(\frac{v}{|x|} - \frac{x}{|x|^3}\langle v, x\rangle\right)\tilde{T}(|x|) + \frac{x}{|x|^2}\tilde{T}'(|x|)\langle x, v\rangle$. Since $T(0) = 0$, using the mean value theorem, we therefore have $\nabla T(x)v = \left(v - \frac{x}{|x|^2}\langle v, x\rangle\right)\tilde{T}'(t) + \frac{x}{|x|^2}\tilde{T}'(|x|)\langle x, v\rangle$ for some $t \in (0, |x|)$. From this, we see that to prove the desired upper bound on the eigenvalues of ∇T, it is enough to show that $\tilde{T}'(r) \leq c\sqrt{1+r^2}$.

To prove this bound, we consider the symmetrized versions of μ_{rad} and γ_{rad} by extending them by symmetry to $\mathbb{R}$, and dividing the density by 2 so that they are still probability measures. We denote these measures by $\hat{\mu}_{rad}$ and $\hat{\gamma}_{rad}$. These measures are still log-concave, and their Cheeger constants are comparable to those of the original measures, up to a universal constant, via Bobkov's theorem we mentioned in Remark 27. Moreover, their median is located at 0. We also extend $\tilde{T}$ to $\mathbb{R}$ by antisymmetry, and denote the function we obtain by $\hat{T}$. It is easy to check that $\hat{T}$ is the optimal map sending $\hat{\gamma}_{\text{rad}}$ onto $\hat{\mu}_{\text{rad}}$.

Following the same arguments as for the one-dimensional case, denoting by p the density of $\hat{\mu}$, we have for $r \geq 0$

$$\begin{aligned}\hat{T}'(r) &\leq Cr^{n-1}e^{-r^2/2}/p(F^{-1}_{\hat{\mu}_{rad}}(F_{\hat{\gamma}_{rad}}(r)) \\ &\leq C\alpha r^{n-1}e^{-r^2/2}F_{\hat{\gamma}_{rad}}(r)^{-1}(1-F_{\hat{\gamma}_{rad}}(r))^{-1} \\ &\leq C\alpha r^{n-1}e^{-r^2/2}(1-F_{\gamma_{rad}}(r))^{-1} \\ &\leq C\alpha\sqrt{1+r^2}\end{aligned}$$

where we have used the estimate $1 - F_{\gamma_{rad}}(r) \geq Ce^{-r^2/2}r^{n-2}$ for large enough r, and C was some positive constant that changed from line to line. Note that the constant is dimension-dependent in the final inequality. □

Remark 28: In this proof, the log-concavity is only used to ensure that $\hat{\mu}_{rad}$ satisfies a Cheeger inequality with a constant comparable to α. This is not necessarily the case for non log-concave radially symmetric measures (for example, the uniform measure on a two-dimensional annulus).

Appendix
An Estimate for the Log-Determinant Function

*Definition 29: A twice differentiable function f : **dom** $f \to \mathbb{R}$ is said to be $\mathsf{m}(x, y)$-strongly convex between $x, y \in$ **dom** f if $\nabla^2 f(tx + (1-t)y) \geq \mathsf{m}(x, y)\mathrm{I}$, for all $t \in [0, 1]$.*

Lemma 30: For all $t \in [0, 1]$, a $\mathsf{m}(x, y)$-strongly convex function f between x and y satisfies

$$\begin{aligned}&tf(x) + (1-t)f(y) \\ &\qquad \geq f(tx + (1-t)y) + t(1-t)\frac{\mathsf{m}(x, y)}{2}|x-y|^2.\end{aligned}$$

Proof: The Taylor series expansion of f for any two points $a, b \in$ **dom** f yields

$$\begin{aligned}f(a) &= f(b) + \langle\nabla f(b), b-a\rangle \\ &\quad + \frac{1}{2}\langle b-a, \nabla^2 f(t_0a + (1-t_0)b)(b-a)\rangle \qquad (24) \\ &\geq f(b) + \langle\nabla f(b), b-a\rangle + \frac{\mathsf{m}(a, b)}{2}|a-b|^2, \qquad (25)\end{aligned}$$

where equation (24) holds for some $t_0 \in [0, 1]$, and inequality (25) follows from Definition 29.

Denote $w \triangleq tx + (1-t)y$. Applying inequality (25) twice yields the two inequalities

$$f(x) \geq f(w) + \langle\nabla f(w), w-x\rangle + \frac{\mathsf{m}(x, w)}{2}|w-x|^2 \qquad (26a)$$

$$f(y) \geq f(w) + \langle\nabla f(w), w-y\rangle + \frac{\mathsf{m}(y, w)}{2}|w-y|^2. \qquad (26b)$$

We now multiply equation (26a) by t and (26b) by $(1-t)$ and add them to obtain

$$\begin{aligned}&tf(x) + (1-t)f(y) \\ &\quad \geq f(w) + \frac{t(1-t)^2\mathsf{m}(x, w) + t^2(1-t)\mathsf{m}(y, w)}{2}|y-x|^2.\end{aligned}$$

By definition, we have $\mathsf{m}(x, w) \geq \mathsf{m}(x, y)$ and $\mathsf{m}(y, w) \geq \mathsf{m}(x, y)$, and this proves the lemma. □

Proof of Lemma 15: The function $f(\cdot) = -\log\det(\cdot)$ is known to be strictly convex and twice differentiable in the interior of the positive semidefinite cone. Substituting the definition of function f into Lemma 30 yields

$$\begin{aligned}\log\det(tA + (1-t)B) &\geq t\log\det(A) + (1-t)\log\det(B) \\ &\quad + t(1-t)\frac{\mathsf{m}(A, B)}{2}\|A-B\|_F^2.\end{aligned}$$

It is a standard fact that $\nabla^2 f(M) = M^{-1} \otimes M^{-1}$, with $\otimes$ denoting the Kronecker product. The minimum eigenvalue of this Kronecker product is given by $1/\lambda^2_{\max}(M)$, where $\lambda_{\max}(M)$ is the largest eigenvalue of M. We therefore have

$$\mathsf{m}(A, B) \geq \min_{t\in[0,1]} \frac{1}{\lambda^2_{\max}(tA + (1-t)B)}.$$

Using the convexity of the maximum eigenvalue then yields the desired result. □

References

[1] T. A. Courtade, M. Fathi, and A. Pananjady, "Wasserstein stability of the entropy power inequality for log-concave random vectors," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 659–663.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 623–656, Jul./Oct. 1948.

[3] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. Control*, vol. 2, no. 2, pp. 101–112, Jun. 1959.

[4] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 2, pp. 267–271, Apr. 1965.

[5] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," *Commun. Math. Phys.*, vol. 62, no. 1, pp. 35–41, 1978.

[6] E. Carlen and A. Soffer, "Entropy production by block variable summation and central limit theorems," *Commun. Math. Phys.*, vol. 140, no. 2, pp. 339–371, 1991.

[7] S. G. Bobkov and G. P. Chistyakov, "Entropy power inequality for the Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 708–714, Feb. 2015.

[8] S. Bobkov and M. Madiman, "The entropy per coordinate of a random vector is highly constrained under convexity conditions," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4940–4954, Aug. 2011.

[9] A. Saumard and J. A. Wellner, "Log-concavity and strong log-concavity: A review," *Statist. Surveys*, vol. 8, p. 45, Apr. 2014.

[10] M. Fathi, E. Indrei, and M. Ledoux, "Quantitative logarithmic Sobolev inequalities and stability estimates," *Discrete Continuous Dyn. Syst.*, vol. 36, no. 12, pp. 6835–6853, 2016.

[11] D. Cordero-Erausquin, "Transport inequalities for log-concave measures, quantitative forms and applications," *Can. J. Math.*, vol. 69, no. 3, pp. 481–501, 2017.

[12] S. G. Bobkov, "Isoperimetric and analytic inequalities for log-concave probability measures," *Ann. Probab.*, vol. 27, no. 4, pp. 1903–1921, Oct. 1999.

[13] G. Toscani, "A strengthened entropy power inequality for log-concave densities," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6550–6559, Dec. 2015.

[14] K. Ball and V. H. Nguyen, "Entropy jumps for isotropic log-concave random vectors and spectral gap," *Studia Math.*, vol. 213, no. 1, pp. 81–96, 2012.

[15] K. Ball, F. Barthe, and A. Naor, "Entropy jumps in the presence of a spectral gap," *Duke Math. J.*, vol. 119, no. 1, pp. 41–63, 2003.

[16] T. A. Courtade, "Strengthening the entropy power inequality," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2294–2298.
[17] T. A. Courtade, "Concavity of entropy power: Equivalent formulations and generalizations," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2017, pp. 56–60.
[18] L. Gross, "Logarithmic Sobolev inequalities," *Amer. J. Math.*, vol. 97, no. 4, pp. 1061–1083, 1975.
[19] E. A. Carlen, "Superadditivity of Fisher's information and logarithmic Sobolev inequalities," *J. Funct. Anal.*, vol. 101, no. 1, pp. 194–211, 1991.
[20] E. Indrei and D. Marcon, "A quantitative Log–Sobolev inequality for a two parameter family of functions," *Int. Math. Res. Notices*, vol. 2014, no. 20, pp. 5563–5580, Jan. 2013. [Online]. Available: https://doi.org/10.1093/imrn/rnt138
[21] S. G. Bobkov, N. Gozlan, C. Roberto, and P.-M. Samson, "Bounds on the deficit in the logarithmic Sobolev inequality," *J. Funct. Anal.*, vol. 267, no. 11, pp. 4110–4138, 2014.
[22] F. Bolley, I. Gentil, and A. Guillin, "Dimensional improvements of the logarithmic Sobolev, Talagrand and Brascamp–Lieb inequalities," *Ann. Probab.*, vol. 46, no. 1, pp. 261–301, 2018.
[23] T. A. Courtade. (2016). "Entropy jumps for radially symmetric random vectors." [Online]. Available: https://arxiv.org/abs/1608.05430
[24] G. Toscani and C. Villani, "Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation," *Commun. Math. Phys.*, vol. 203, no. 3, pp. 667–706, Jun. 1999.
[25] C. Villani, "Cercignani's conjecture is sometimes true and always almost true," *Commun. Math. Phys.*, vol. 234, no. 3, pp. 455–490, Mar. 2003.
[26] M. Godavarti and A. Hero, "Convergence of differential entropies," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 171–176, Jan. 2004.
[27] A. V. Bobylev and C. Cercignani, "On the rate of entropy production for the Boltzmann equation," *J. statist. Phys.*, vol. 94, nos. 3–4, pp. 603–618, Feb. 1999.
[28] E. A. Carlen, M. C. Carvalho, J. L. Roux, M. Loss, and C. Villani, "Entropy and chaos in the Kac model," *Kinetic Rel. Models*, vol. 3, no. 1, pp. 85–122, 2010.
[29] C. Villani, *Topics in Optimal Transportation*, vol. 58. Providence, RI, USA: AMS, 2003.
[30] C. Villani, *Optimal Transport: Old and New*, vol. 338. Berlin, Germany: Springer-Verlag, 2008.
[31] M. Ledoux, *The Concentration of Measure Phenomenon*, vol. 89. Providence, RI, USA: AMS, 2005.
[32] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. London, U.K.: Oxford Univ. Press, 2013.
[33] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Found. Trends Commun. Inf. Theory*, vol. 10, nos. 1–2, pp. 1–246, 2013.
[34] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Commun. Pure Appl. Math.*, vol. 44, no. 4, pp. 375–417, Jun. 1991.
[35] R. J. McCann, "Existence and uniqueness of monotone measure-preserving maps," *Duke Math. J.*, vol. 80, no. 2, pp. 309–324, 1995.
[36] O. Rioul, "Yet another proof of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3595–3599, Jun. 2017.
[37] H. Knöthe, "Contributions to the theory of convex bodies," *Michigan Math. J.*, vol. 4, no. 1, pp. 39–52, 1957.
[38] M. Rosenblatt, "Remarks on a multivariate transformation," *Ann. Math. Statist.*, vol. 23, no. 3, pp. 470–472, 1952.
[39] A. Figalli, F. Maggi, and A. Pratelli, "A mass transportation approach to quantitative isoperimetric inequalities," *Inventiones Math.*, vol. 182, no. 1, pp. 167–211, Oct. 2010.
[40] L. A. Caffarelli, "A localization property of viscosity solutions to the Monge–Ampere equation and their strict convexity," *Ann. Math.*, vol. 131, no. 1, pp. 129–134, 1990.
[41] L. A. Caffarelli, "The regularity of mappings with a convex potential," *J. Amer. Math. Soc.*, vol. 5, no. 1, pp. 99–104, 1992.
[42] L. A. Caffarelli, "Monotonicity properties of optimal transportation and the FKG and related inequalities," *Commun. Math. Phys.*, vol. 214, no. 3, pp. 547–563, Nov. 2000.
[43] M. Colombo, A. Figalli, and Y. Jhaveri, "Lipschitz changes of variables between perturbations of log-concave measures," *Annali della Scuola Normale Superiore*, vol. 17, no. 4, pp. 1491–1519, Dec. 2017.
[44] M. Ledoux, "Logarithmic Sobolev inequalities for unbounded spin systems revisited," in *Séminaire de Probabilités XXXV* (Lecture Notes in Mathematics), vol. 1755, J. Azéma, M. Émery, M. Ledoux, and M. Yor, Eds. Berlin, Germany: Springer, 2001. [Online]. Available: https://doi.org/10.1007/978-3-540-44671-2_13
[45] A. V. Kolesnikov, "Hessian metrics, CD(K,N)-spaces, and optimal transportation of log-concave measures," *Discrete Continuous Dyn. Syst.*, vol. 34, no. 4, pp. 1511–1532, 2014.
[46] A. R. Barron, "Entropy and the central limit theorem," *Ann. Probab.*, vol. 14, no. 1, pp. 336–342, 1986.
[47] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, vol. 348. Cham, Switzerland: Springer, 2013.
[48] M. Bonnefont, A. Joulin, and Y. Ma, "Spectral gap for spherically symmetric log-concave probability measures, and beyond," *J. Funct. Anal.*, vol. 270, no. 7, pp. 2456–2482, 2016.
[49] M. Bonnefont, A. Joulin, and Y. Ma, "A note on spectral gap and weighted Poincaré inequalities for some one-dimensional diffusions," *ESAIM, Probab. Statist.*, vol. 20, pp. 18–29, Jun. 2016.
[50] P. Buser, "A note on the isoperimetric constant," *Ann. Sci. École Norm. Supérieure (4)*, vol. 15, no. 2, pp. 213–230, 1982.
[51] M. Ledoux, "Spectral gap, logarithmic Sobolev constant, and geometric bounds," *Surv. Differ. Geometry*, vol. 9, pp. 219–240, 2004. [Online]. Available: http://dx.doi.org/10.4310/SDG.2004.v9.n1.a6
[52] E. Milman, "On the role of convexity in isoperimetry, spectral gap and concentration," *Inventiones Math.*, vol. 177, no. 1, pp. 1–43, 2009.
[53] S. G. Bobkov, "Spectral gap and concentration for some spherically symmetric probability measures," *Geometric Aspects of Functional Analysis* (Lecture Notes in Mathematics), vol. 1807, V. D. Milman and G. Schechtman, Eds. Berlin, Germany: Springer, 2003. [Online]. Available: https://doi.org/10.1007/978-3-540-36428-3_4

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

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

Max Fathi received his PhD from the Université Pierre et Marie Curie, Paris in 2014. He is currently a CNRS researcher at the Institut de Mathématiques de Toulouse.

Ashwin Pananjady is a Ph.D. candidate in the Department of Electrical Engineering and Computer Sciences (EECS) at the University of California, Berkeley. He received the B.Tech. degree (with honors) in electrical engineering from IIT Madras in 2014. His research interests include statistics, optimization, and information theory. He is a recipient of the Governor's Gold Medal from IIT Madras.