Wasserstein Stability of the Entropy Power Inequality for Log-Concave Random Vectors

Thomas A. Courtade*, Max Fathi† and Ashwin Pananjady*
*University of California, Berkeley, Department of Electrical Engineering and Computer Sciences
†CNRS & Université Paul Sabatier, Institut de Mathématiques de Toulouse

***Abstract*—We establish quantitative stability results for the entropy power inequality (EPI) in arbitrary dimension. Specifically, we show that if uniformly log-concave densities nearly saturate the EPI, then they must be close to Gaussian densities in the quadratic Wasserstein distance. Further, if one of the densities is log-concave and the other is Gaussian, then the deficit in the EPI can be controlled in terms of the L^1-Wasserstein distance.**

As a counterpoint, an example shows that the EPI can be unstable with respect to the quadratic Wasserstein distance even if densities are uniformly log-concave on sets of measure arbitrarily close to one. The proofs are based on optimal transportation.

I. Introduction

Let X and Y be independent random vectors on $\mathbb{R}^n$ with respective laws μ and ν, each absolutely continuous with respect to Lebesgue measure. The celebrated EPI proposed by Shannon and proved by Stam [1] asserts that

$$N(\mu * \nu) \geq N(\mu) + N(\nu), \tag{1}$$

where $N(\mu) := \frac{1}{2\pi e} e^{2h(\mu)/n}$ denotes the entropy power of μ, and $h(\mu) = h(X) = -\int f \log f$ is entropy associated to the density f of X (all logarithms have base e). For $t \in (0,1)$, let us define the deficit

$$\delta_t(\mu,\nu) := h(\sqrt{t}X + \sqrt{\bar{t}}Y) - \Big(t\, h(X) + \bar{t}\, h(Y)\Big),$$

where $\bar{t} := 1-t$. Unaware of the works by Shannon, Stam and Blachman [2], Lieb [3] rediscovered the EPI by establishing $\delta_t(\mu,\nu) \geq 0$ and noting its equivalence to (1). Due to the equivalence of the Shannon-Stam and Lieb inequalities, we shall generally refer to both as the EPI.

It is well known that $\delta_t(\mu,\nu)$ vanishes if and only if μ,ν are Gaussian measures that are identical up to translation [4]. However, despite the fundamental role the EPI plays in information theory, few stability estimates are known. Specifically, *if $\delta_t(\mu,\nu)$ is small, must μ and ν be 'close' to Gaussian measures, which are themselves 'close' to each other, in a precise and quantitative sense?* This is our motivating question.

Toward answering this question, our main result is a dimension-free, quantitative stability estimate for the EPI. More specifically, we show that if 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 quadratic 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. A weaker deficit estimate is obtained involving the L^1-Wasserstein distance for log-concave measures when one of the two variables is Gaussian. A further estimate is obtained when one of the variables is Gaussian and the other has nonzero spectral gap.

Before stating the main results, let us introduce notation. Throughout, $|\cdot|$ denotes Euclidean length on $\mathbb{R}^n$. 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[1],

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

Next, we recall that for $p \geq 1$, the L^p-Wasserstein distance between probability measures μ,ν is defined according to

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

where 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 write $\Sigma_\mu = \mathbb{E}[XX^\top]$ to denote the covariance matrix of X. We remark here that 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.

Organization: The rest of this paper is organized as follows: Sections II and III describe our main results and the relationship to previous work, respectively. Section IV gives an example where the EPI is not stable with respect to quadratic Wasserstein distance when regularity conditions are not met. Section V gives proofs of our main results and a brief discussion of techniques, which are primarily based on optimal mass transportation.

II. Main Results

Our main result is the following:

Theorem 1. *Let $\mu = e^{-\varphi}\gamma$ and $\nu = e^{-\psi}\gamma$ be centered probability measures, where φ and ψ are convex. Then*

$$\delta_t(\mu,\nu) \geq \frac{t\bar{t}}{2}\inf_{\gamma_1,\gamma_2\in\Gamma}\Big(W_2^2(\mu,\gamma_1) + W_2^2(\nu,\gamma_2) + W_2^2(\gamma_1,\gamma_2)\Big). \tag{2}$$

Under the assumptions of the theorem, the three terms in the RHS of (2) explicitly give necessary conditions for the deficit $\delta_t(\mu,\nu)$ 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

T. C. and A. P. were supported in part by NSF grants CCF-1528132 and CCF-0939370. M. F. was supported by NSF FRG grant DMS-1361122.

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

978-1-5090-4096-4/17/$31.00 ©2017 IEEE

additive on product measures, so (2) is dimension-free, which is compatible with the additivity of δ_t on product measures.

Theorem 1 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 $\delta_t(\mu,\nu)$ is invariant under the rescaling $(X,Y) \to (\eta^{1/2}X, \eta^{1/2}Y)$. Similarly, if $X \sim \mu$ has density f which is η-uniformly log-concave, i.e.,

$$-\nabla^2 \log f \geq \eta \mathrm{I}, \tag{3}$$

then a change of variables reveals that the density f_η associated to 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 1 is equivalent to the following:

Corollary 1. *If μ and ν are centered probability measures with densities satisfying* (3)*, then*

$$\delta_t(\mu,\nu) \geq \eta \frac{t\bar{t}}{2} \inf_{\gamma_1,\gamma_2\in\Gamma} \Big(W_2^2(\mu,\gamma_1) + W_2^2(\nu,\gamma_2) + W_2^2(\gamma_1,\gamma_2)\Big).$$

Variations on this result also apply to certain families of non log-concave measures, see Remark 2.

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 1.

Corollary 2. *If μ and ν are centered probability measures with densities satisfying* (3)*, then*

$$\delta_t(\mu,\nu) \geq \eta \frac{t\bar{t}}{8} \Big(\mathsf{d}^2_{W_2}(\mu) + \mathsf{d}^2_{W_2}(\nu) + W_2^2(\mu,\nu)\Big). \tag{4}$$

Shannon's form of the entropy power inequality (1) is oftentimes preferred to Lieb's inequality for applications in information theory. In this form, it is only necessary that μ,ν be Gaussian with *proportional* covariances in order to achieve equality. Motivated by this, we define for two centered probability measures μ,ν the quantity

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

where $\|\cdot\|_F$ denotes Frobenius norm and $\Sigma^{1/2}$ denotes the unique positive semidefinite matrix such that $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. In particular, $\mathsf{d}_F(\mu,\nu) = 0$ if and only if the covariance matrices Σ_μ and Σ_ν are proportional.

Starting with Corollary 2, straightforward computations give an analogous estimate for Shannon's EPI:

Corollary 3. *Let μ,ν be centered probability measures on $\mathbb{R}^n$ satisfying* (3) *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 equal to

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

with θ being chosen to satisfy $\theta/\bar{\theta} = N(\mu)/N(\nu)$.

We note that, under the stated assumptions of log-concavity, equality conditions for (1) are explicitly captured by the three main terms defining $\Delta_{\mathsf{EPI}}(\mu,\nu)$.

Finally, we give stability estimates when one variable is Gaussian.

Theorem 2. *If μ is centered and has log-concave density,*

$$\delta_t(\mu,\gamma) \geq C\, t\,\bar{t}\, \min(W_1^2(\mu,\gamma), 1), \tag{5}$$

with C a numerical constant that does not depend on μ.

This estimate is reminiscent of the deficit estimates on Talagrand's inequality of [5], [6], with a remainder term that stays bounded when the distance becomes large.

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 $\int s d\mu = 0$,

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

Theorem 3. *If μ is centered and has spectral gap λ, then*

$$\delta_t(\mu,\gamma) \geq \min(\lambda,1)\frac{t\,\bar{t}}{2} D(\mu\|\gamma).$$

All log-concave distributions have positive spectral gap [7], so the hypothesis of Theorem 3 is weaker than that of Theorem 2. However, the advantage of (5) is that it does not rely on any quantitative information on μ, only that it is log-concave.

III. Relation to Prior Work

Chronologically speaking, the first stability result for the EPI is a *qualitative* estimate due to Carlen and Soffer [4], which holds under general conditions. Roughly speaking, they show that there is a function $\Theta : \mathbb{R} \to [0,\infty)$, strictly increasing from 0, depending only on the dimension n, the parameter t and the smoothness and decay properties of μ,ν that satisfies $\delta_t(\mu,\nu) \geq \Theta(\mathcal{D}(\mu))$, where $\mathcal{D}$ denotes the *nongaussianness* of μ. The construction of the function Θ relies on a compactness argument, and is therefore non-explicit so is not directly comparable to our results. However, it did allow Carlen and Soffer to rigorously settle the cases of equality.

A few *quantitative* stability estimates for the EPI have been developed in recent years. We review them below and comment on the relationship to our results; a more detailed survey may be found in [8]. To begin, we mention a stability result due to Toscani [9], which asserts for probability measures μ,ν with log-concave densities, there is a function R such that

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

where $R(\mu,\nu) \geq 1$ with equality only if μ,ν are Gaussian measures. Although it can be written explicitly in terms of integrals of nonlinear functionals evaluated along the evolutes of μ,ν under the heat semigroup, the function $R(\mu,\nu)$ is complicated and does not explicitly control the distance of μ,ν to the space of Gaussian measures. Toscani leaves this as an open problem [9, Remark 7]. Corollary 3 provides a satisfactory answer to his problem when μ,ν are uniformly log-concave. Similarly, Theorem 2 provides an answer when one of the measures is log-concave and the other Gaussian.

Next, we compare to the main result of Ball and Nguyen [10], 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, then

$$\delta_{1/2}(\mu,\mu) \geq \tfrac{\lambda}{4(1+\lambda)} D(\mu\|\gamma) \geq \tfrac{\lambda}{8(1+\lambda)} W_2^2(\mu,\gamma), \qquad (6)$$

where the second inequality is due to Talagrand's Gaussian transportation cost inequality. Theorem 3 gives a similar bound under weaker hypotheses, with the caveat that one measure is required to be Gaussian. Likewise, if μ is uniformly log-concave then Corollary 1 yields a similar bound. Indeed, Corollary 1 may be viewed as an extension of (6) to 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) uniform log-concavity implies a positive spectral gap, but not vice versa. It is interesting to ask whether the hypothesis of Corollary 1 can be weakened to require only a spectral gap; Theorem 3 along with earlier results by Ball, Barthe and Naor [11] and Johnson and Barron [12] in dimension one provides some grounds for cautious optimism.

The latter two results mentioned above assume log-concave densities, as do we (for the most part). In contrast, the refined EPI established in [13] 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 3. In particular, for arbitrary probability measures μ, ν on $\mathbb{R}^n$ with finite second moments, it was shown in [14] that

$$N(\mu * \nu) \leq (N(\mu) + N(\nu)) \left(\bar\theta\, \mathsf{p}(\mu) + \theta\, \mathsf{p}(\nu)\right), \qquad (7)$$

where θ is the same as in the definition of $\Delta_{\mathsf{EPI}}(\mu,\nu)$ and $\mathsf{p}(\mu) := \frac{1}{n} N(\mu) J(\mu)$, with $J(\mu)$ denoting Fisher information. Stam's inequality asserts $\mathsf{p}(\mu) \geq 1$ with equality only if μ is Gaussian, so that $\mathsf{p}(\mu)$ may be interpreted as a measure of how far μ is from the set of Gaussian measures. Thus, the deficit term $\bar\theta\, \mathsf{p}(\mu) + \theta\, \mathsf{p}(\nu)$ in (7) bears a pleasant resemblance to the deficit term $\bar\theta\, \mathsf{d}^2_{W_2}(\mu) + \theta\, \mathsf{d}^2_{W_2}(\nu)$ in Corollary 3. Importantly, though, the former is an upper bound on $N(\mu*\nu)$, while the latter yields a lower bound.

Finally, 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 [15] that, for any $\varepsilon > 0$

$$\delta_{1/2}(\mu,\mu) \geq C_\varepsilon(\mu) n^\varepsilon D^{1+\varepsilon}(\mu\|\gamma_\mu), \qquad (8)$$

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 [16], [17]. Neither (2) nor (8) imply the other since the hypotheses required are quite different (uniform 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 (2) will be much better than those in (8) which, although numerical, can be quite small.

IV. Instability of the EPI: An Example

As a counterpoint to Theorem 1 and to provide justification for the regularity assumptions therein, we observe that there are probability measures that satisfy the hypotheses required in Theorem 1 on sets of measure arbitrarily close to one, but severely violate its conclusion.

Proposition 1. *There is a sequence 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 satisfy* (3) *for $\eta = 1$ with high probability. 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, $\lim_{\epsilon\downarrow 0} \delta_t(\mu_\epsilon,\mu_\epsilon) = 0$ for all $t \in (0,1)$.*
3) *$(\mu_\epsilon)_{\epsilon>0}$ are bounded away from Gaussians in W_2. 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. Indeed, the proposition follows by choosing $d\mu_\epsilon = f_\epsilon dx$, with

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

The computations can be found in [18].

V. Discussion and Proofs

The remainder of this paper makes use of ideas from optimal transport, and reader familiarity assumed. The unfamiliar reader is directed to the comprehensive introduction [19]. We recall that a map $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to transport a measure μ to ν if the pushforward of μ under T is ν (i.e., $\nu = T\#\mu$).

Our starting point comes from a recent paper of Rioul [20]. 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 1. *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*

$$\delta_t(\mu,\nu) \geq \mathbb{E}\log \frac{|\det(t\nabla T_1(X^*) + \bar t\,\nabla T_2(Y^*))|}{|\det(\nabla T_1(X^*))|^t\, |\det(\nabla T_2(Y^*))|^{\bar t}}, \qquad (9)$$

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

Remark 1. *For a vector-valued map $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) : \mathbb{R}^n \to \mathbb{R}^n$, we write $\nabla\Phi$ to denote its Jacobian. That is, $(\nabla\Phi(x))_{ij} = \frac{\partial}{\partial x_i}\Phi_j(x)$.*

In words, (9) 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 [21], [22]), the Jacobians ∇T_1 and ∇T_2 are upper triangular matrices with positive diagonal entries. Using this property, Rioul concludes $\delta_t(\mu,\nu) \geq 0$ 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 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 $\delta_t(\mu,\nu)$ to vanish. Specifically, μ and ν must be Gaussian measures, equal up to translation.

In our proof, instead of the Knöthe map, we shall use the Brenier map from optimal transport theory, which has a useful rigid structure:

Theorem 4 (Brenier-McCann [23], [24])**.** *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 (called 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 is a random variable with law μ, and therefore $T(X)$ has law ν. In other words, $(X,T(X))$ is an optimal coupling for the Wasserstein distance W_2.

In contrast to 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 (9). Moreover, by strict concavity of the log-determinant function, equality in the EPI implies $\nabla T_1(X^*)=\nabla T_2(Y^*)$ almost everywhere, and are thus constant. Hence, T_1 and T_2 are necessarily affine functions, identical up to translation. This immediately implies $\delta_t(\mu,\nu)=0$ 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 [25]. We crucially use the properties of the Brenier map in proving Theorem 1.

A. Proof of Theorem 1

The proof of Theorem 1 is short, but makes use of several foundational results from the theory of optimal transport. We will need the following lemma; see [18] for a proof.

Lemma 2. *For positive definite matrices A,B and $t\in[0,1]$,*

$$\log\det(tA+\bar{t}B)\geq t\log\det(A)+\bar{t}\,\log\det(B)+\frac{t\,\bar{t}}{2\max\{\lambda_{\max}^2(A),\lambda_{\max}^2(B)\}}\|A-B\|_F^2,$$

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

In addition, we remind the reader that a random vector X having log-concave density enjoys (i) finite second moment (in fact, finite moments of all orders); and (ii) finite entropy $h(X)$. Since Theorem 1 requires log-concave densities, these conditions will be implicitly assumed throughout the proof.

Proof of Theorem 1. 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 [26], [27] 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 [28] and, e.g., [19, Theorem 9.14]) implies 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 2 yields the following (pointwise) estimate

$$\log\det(t\nabla T_1+\bar{t}\nabla T_2)\geq t\log\det(\nabla T_1)+\bar{t}\log\det(\nabla T_2)+\frac{t\bar{t}}{2}\|\nabla T_1-\nabla T_2\|_F^2.$$

Combined with (9) we obtain:

$$\delta_t(\mu,\nu)\geq\frac{t\bar{t}}{2}\mathbb{E}\|\nabla(T_1(X^*)-X^*)-\nabla(T_2(Y^*)-Y^*)\|_F^2.$$

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

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

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

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

which holds similarly for Y^*. The desired inequality now follows from the definition of W_2.

Thus, Theorem 1 holds when the densities are smooth and positive everywhere. If this is not the case, we may regularize μ,ν by first convolving with a Gaussian measure. The general result then follows by considering arbitrarily small perturbations. See [18] for details. □

Remark 2. *Assuming uniform log-concavity ensures that the optimal transport map is Lipschitz. This can still be the case* *in other situations. For example, this holds for families of bounded perturbations of the Gaussian measure, including the radially symmetric case [29]. Extensions to settings where eigenvalues grow at most linearly (e.g,. as in the case of the exponential measure) can be found in [18].*

B. Proof of Theorems 2 and 3

Proof of Theorem 2. Assume that $X^* \sim \gamma$, and let T be the Brenier map sending γ onto μ. For convenience, let us write λ_i for the eigenvalues of $\nabla T(x)$, in increasing order (so that $\lambda_n = \lambda_{\max}(\nabla T(x))$). Since the Brenier map sending γ onto γ is the identity, the combination of (9) and Lemma 2 yields in this case

$$\delta_t(\mu,\gamma) \geq \frac{t\bar{t}}{2}\mathbb{E}\left[\frac{\|\nabla T(X^*) - \mathrm{I}\|_F^2}{1+\lambda_{\max}(\nabla T(X^*))^2}\right] = \frac{t\bar{t}}{2}\mathbb{E}\left[\frac{\sum_{i=1}^n (\lambda_i - 1)^2}{1+\lambda_n^2}\right].$$

By the Cauchy-Schwarz inequality and the lower bound on $\delta_t(\mu,\gamma)$ noted above, we have

$$\begin{aligned}\mathbb{E}\left[\sqrt{\sum(\lambda_i-1)^2}\right] &\leq \sqrt{\mathbb{E}[1+\lambda_n^2]}\sqrt{\mathbb{E}\left[\frac{\sum(\lambda_i-1)^2}{1+\lambda_n^2}\right]}\\ &\leq \sqrt{\mathbb{E}[1+\lambda_n^2]}\sqrt{\frac{2}{t(1-t)}\delta_t(\mu,\gamma)}.\end{aligned}$$

The L^1 Poincaré inequality for the Gaussian measure implies

$$W_1(\mu,\gamma) \leq 2\mathbb{E}\left[\sqrt{\sum(\lambda_i-1)^2}\right].$$

Hence, if we have an L^2 bound on the largest eigenvalue of ∇T, we can deduce a W_1 estimate on the deficit (in contrast to using a uniform bound as in the proof of Theorem 1). To this end, a result of Kolesnikov asserts that $\mathbb{E}[\lambda_n^2] \leq \frac{3}{2}\mathbb{E}[\lambda_n]^2$ (see [30], Theorem 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

$$\mathbb{E}\left[\sqrt{\sum(\lambda_i-1)^2}\right] \leq \sqrt{4+3\mathbb{E}\left[\sqrt{\sum(\lambda_i-1)^2}\right]^2}\sqrt{\frac{2}{t\bar{t}}\delta_t(\mu,\gamma)}.$$

Since $r/\sqrt{1+r^2} \geq c\min(r,1)$ and $2\mathbb{E}\left[\sqrt{\sum(\lambda_i-1)^2}\right] \geq W_1(\mu,\gamma)$, we have $\sqrt{2\delta_t(\mu,\gamma)/(t\bar{t})} \geq C\min(W_1(\mu,\gamma),1)$, so the result follows. □

Proof of Theorem 3. Let $I(\nu\|\gamma)$ denote the Fisher information of ν relative to γ. It was shown in [5, Proof of Theorem 1] that if ν is a centered probability measure with spectral gap λ, then for all $t \geq 0$

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

where ν_t denotes the evolute of ν along the Ornstein-Uhlenbeck process at time t. The claim follows by identifying $\nu \leftarrow \mu_\tau$, integrating with respect to the time variable τ, and applying de Bruijn's identity. Detailed computations are omitted due to space constraint. □

REFERENCES

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

[2] N. M. Blachman, "The convolution inequality for entropy powers," *Info. Theory, IEEE Transactions on*, vol. 11, no. 2, pp. 267–271, 1965.

[3] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," *Communications in Mathematical Physics*, vol. 62, no. 1, pp. 35–41, 1978.

[4] E. A. Carlen and A. Soffer, "Entropy production by block variable summation and central limit theorems," *Communications in mathematical physics*, vol. 140, no. 2, pp. 339–371, 1991.

[5] M. Fathi, E. Indrei, and M. Ledoux, "Quantitative logarithmic Sobolev inequalities and stability estimates," *Dyn. Systems*, vol. 36, no. 12, 2016.

[6] D. Cordero-Erausquin, "Transport inequalities for log-concave measures, quantitative forms and applications," *to appear in Canad. J. Math.*, 2015.

[7] S. G. Bobkov, "Isoperimetric and analytic inequalities for log-concave probability measures," *The Annals of Probability*, vol. 27, no. 4, pp. 1903–1921, 1999.

[8] M. Madiman, J. Melbourne, and P. Xu, "Forward and reverse entropy power inequalities in convex geometry," *arXiv preprint arXiv:1604.04225*, 2016.

[9] G. Toscani, "A strengthened entropy power inequality for log-concave densities," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6550–6559, 2015.

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

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

[12] O. Johnson and A. Barron, "Fisher information inequalities and the central limit theorem," *Probability Theory and Related Fields*, vol. 129, no. 3, pp. 391–409, 2004.

[13] T. A. Courtade, "Strengthening the entropy power inequality," in *2016 IEEE Intl. Symp. on Information Theory*, July 2016, pp. 2294–2298.

[14] ——, "Links between the logarithmic Sobolev inequality and the convolution inequalities for entropy and fisher information," *arXiv preprint arXiv:1608.05431*, 2016.

[15] ——, "Entropy jumps for radially symmetric random vectors," *arXiv preprint arXiv:1608.05430*, 2016.

[16] G. Toscani and C. Villani, "Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation," *Comm. in math. physics*, vol. 203, no. 3, pp. 667–706, 1999.

[17] C. Villani, "Cercignani's conjecture is sometimes true and always almost true," *Comm. in math. physics*, vol. 234, no. 3, pp. 455–490, 2003.

[18] T. A. Courtade, M. Fathi, and A. Pananjady, "Wasserstein stability of the entropy power inequality for log-concave densities," *arXiv preprint arXiv:1610.07969*, 2016.

[19] C. Villani, *Topics in optimal transportation*. American Mathematical Soc., 2003, no. 58.

[20] O. Rioul, "Yet another proof of the entropy power inequality," *arXiv preprint arXiv:1606.05969*, 2016.

[21] H. Knothe, "Contributions to the theory of convex bodies." *The Michigan Mathematical Journal*, vol. 4, no. 1, pp. 39–52, 1957.

[22] M. Rosenblatt, "Remarks on a multivariate transformation," *The annals of mathematical statistics*, vol. 23, no. 3, pp. 470–472, 1952.

[23] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on pure and applied mathematics*, vol. 44, no. 4, pp. 375–417, 1991.

[24] R. J. McCann, "Existence and uniqueness of monotone measure-preserving maps," *Duke Mathematical Journal*, vol. 80, no. 2, pp. 309–324, 1995.

[25] A. Figalli, F. Maggi, and A. Pratelli, "A mass transportation approach to quantitative isoperimetric inequalities," *Inventiones mathematicae*, vol. 182, no. 1, pp. 167–211, 2010.

[26] L. A. Caffarelli, "A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity," *Ann. of Math. (2)*, vol. 131, no. 1, pp. 129–134, 1990.

[27] ——, "The regularity of mappings with a convex potential," *Journal of the American Mathematical Society*, vol. 5, no. 1, pp. 99–104, 1992.

[28] ——, "Monotonicity properties of optimal transportation and the FKG and related inequalities," *Communications in Mathematical Physics*, vol. 214, no. 3, pp. 547–563, 2000.

[29] M. Colombo, A. Figalli, and Y. Jhaveri, "Lipschitz changes of variables between perturbations of log-concave measures," *preprint*, 2016.

[30] A. V. Kolesnikov, "Hessian metrics, $CD(K,N)$-spaces, and optimal transportation of log-concave measures," *Discrete Contin. Dyn. Syst.*, vol. 34, no. 4, pp. 1511–1532, 2014.