Comptes Rendus

Mathématique

Sinho Chewi and Aram-Alexandre Pooladian

An entropic generalization of Caffarelli's contraction theorem via covariance inequalities

Functional analysis / *Analyse fonctionnelle*

An entropic generalization of Caffarelli's contraction theorem via covariance inequalities

Une généralisation entropique du théorème de contraction de Caffarelli à l'aide d'inégalités de covariance

Sinho Chewi [a] **and Aram-Alexandre Pooladian** [*, b]

[a] School of Mathematics, Institute for Advanced Study, Princeton, USA

[b] Center for Data Science, New York University, New York, USA

E-mails: schewi@ias.edu (S. Chewi), aram-alexandre.pooladian@nyu.edu
(A-A. Pooladian)

Abstract. The optimal transport map between the standard Gaussian measure and an α-strongly log-concave probability measure is $\alpha^{-1/2}$-Lipschitz, as first observed in a celebrated theorem of Caffarelli. In this paper, we apply two classical covariance inequalities (the Brascamp–Lieb and Cramér–Rao inequalities) to prove a sharp bound on the Lipschitz constant of the map that arises from *entropically regularized* optimal transport. In the limit as the regularization tends to zero, we obtain an elegant and short proof of Caffarelli's original result. We also extend Caffarelli's theorem to the setting in which the Hessians of the log-densities of the measures are bounded by arbitrary positive definite commuting matrices.

Résumé. La fonction de transport optimale entre la mesure gaussienne standardisée et une mesure de probabilité α-fortement log-concave est $\alpha^{-1/2}$-Lipschitz, comme l'a noté Caffarelli dans le célèbre théorème qui porte désormais son nom. Dans ce travail, nous utilisons deux inégalités de covariance classiques (l'inégalité de Brascamp–Lieb ainsi de celle de Cramèr–Rao) pour établir une borne optimale sur la constante de Lipschitz de la fonction de transport associée au transport optimal avec *régularisation entropique*. En étudiant le cas limite où l'effet de la régularisation disparait, nous obtenons une démonstration courte et élégante du théorème de Caffarelli. De surcroît, cette approche nous permet d'étendre la validité du théoreme de Caffarelli au cas de log-densités dont les hessiens sont contrôlés par des matrices positives définies qui peuvent être choisies arbitrairement tant qu'elles commutent entre elles.

* Corresponding author.

1. Introduction

In [6], Caffarelli proved the following seminal result.

Theorem 1 (Caffarelli's contraction theorem). *Let $P = \exp(-V)$ and $Q = \exp(-W)$ have smooth densities on $\mathbb{R}^d$, with $\nabla^2 V \preceq \beta_V I$ and $\nabla^2 W \succeq \alpha_W I > 0$. Then, the optimal transport map $\nabla\varphi_0$ from P to Q is $\sqrt{\beta_V/\alpha_W}$-Lipschitz.*

Here, $\varphi_0 : \mathbb{R}^d \to \mathbb{R}$ is a convex function, known as a *Brenier potential*. The optimal transport map $\nabla\varphi_0 : \mathbb{R}^d \to \mathbb{R}^d$ pushes forward P to Q, in the sense that if X is a random variable with law P, then $\nabla\varphi_0(X)$ is a random variable with law Q. See Section 2.2 and the textbook [36] for background on optimal transport.

Caffarelli's contraction theorem can be used to transfer functional inequalities, such as a Poincaré inequality, from the standard Gaussian measure on $\mathbb{R}^d$ to other probability measures [3]. Towards this end, recent works have also constructed and studied alternative Lipschitz transport maps (e.g. [18, 26–28]), but still the properties of the original optimal transport map remain of fundamental interest, with many questions unresolved [8, 35].

Indeed, besides the application to functional inequalities, the structural properties of optimal transport maps play a fundamental role in theoretical and methodological advances in optimal transport, such as the control of the curvature of the Wasserstein space through the notion of extendible geodesics [1, 21], the stability of Wasserstein barycenters [7], and the statistical estimation of optimal transport maps [16].

In applied domains, however, the inauspicious computational and statistical burden of solving the original optimal transport problem has instead led practitioners to consider *entropically regularized* optimal transport, as pioneered by Cuturi in [11]. In addition to its practical merits, entropic optimal transport enjoys a rich mathematical theory, rooted in its connection to the classical Schrödinger bridge problem [23], which has led to powerful applications to high-dimensional probability [12, 14, 22]. As such, it is natural to study the properties of the entropic analogue of the optimal transport map.

In this paper, we prove a generalization of Caffarelli's contraction theorem to the setting of entropic optimal transport. Namely, we study the Hessian of the *entropic Brenier potential* (see Section 2.3), which admits a representation as a covariance matrix (Lemma 6). By applying two well-known inequalities for covariance matrices (the Brascamp–Lieb inequality and the Cramér–Rao inequality), we quickly deduce a sharp upper bound on the operator norm of the Hessian which holds for any value $\varepsilon > 0$ of the regularization parameter.

As a byproduct of our analysis, by sending $\varepsilon \searrow 0$ and appealing to recent convergence results for the entropic Brenier potentials [4], we obtain the shortest proof of Caffarelli's contraction theorem to date. Notably, our argument allows us to sidestep the regularity of the optimal transport map, which is a key obstacle in Caffarelli's original proof and many others in the literature (see, e.g., [20]).

Recently, in [12] (see also [32]), Fathi, Gozlan, and Prod'homme gave a proof of Caffarelli's theorem using a surprising equivalence between Theorem 1 and a statement about Wasserstein projections, which was discovered through the theory of weak optimal transport [15]. In order to verify the latter, their proof also used ideas from entropic optimal transport.[1] In comparison, we note that our argument is much more direct.

To further demonstrate the applicability of our technique, in Section 4 we prove a generalization of Caffarelli's result which reveals a remarkable extremal property of optimal transport maps

---

[1] In particular, with some effort, a bound on the Hessian of the entropic Brenier potential can also be read off from their proof.

between Gaussians. Namely, if $\nabla^2 V \preceq A^{-1}$ and $\nabla^2 W \succeq B^{-1}$, where A and B are arbitrary commuting positive definite matrices, then the Hessian of the Brenier potential from P to Q is pointwise upper bounded (in the PSD ordering) by $A^{-1/2} B^{1/2}$, the Hessian of the Brenier potential from $\mathcal{N}(0, A)$ to $\mathcal{N}(0, B)$. To the best of our knowledge, this result is new.

2. Background

2.1. *Assumptions*

We study probability measures P, Q on $\mathbb{R}^d$ satisfying the following mild regularity assumptions.

Assumption 2 (Regularity conditions). *We henceforth refer to the* source measure *as P and the* target measure *as Q. We say that (P, Q) satisfies our regularity conditions if:*

(1) *P has full support on $\mathbb{R}^d$ and Q is supported on a convex subset of $\mathbb{R}^d$. Let Ω_Q denote the interior of the support of Q, so that Ω_Q is a convex open set.*

(2) *P and Q admit positive Lebesgue densities on $\mathbb{R}^d$ and Ω_Q, which we can therefore be written $\exp(-V)$ and $\exp(-W)$ respectively for functions $V, W : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$. We abuse notation and identify the measures with their densities, thus writing $P = \exp(-V)$ and $Q = \exp(-W)$.*

(3) *We assume that V and W are twice continuously differentiable on $\mathbb{R}^d$ and Ω_Q respectively.*

Some of these assumptions can be eventually relaxed, but they suffice for the purposes of this work. Throughout the rest of the paper and for the sake of simplicity, these regularity assumptions are assumed to hold for the probability measures under consideration.

2.2. *Optimal transport without regularization*

Let P and Q be probability measures with finite second moment. The *optimal transport problem* is the following optimization problem:

$$\underset{\pi \in \Pi(P, Q)}{\text{minimize}} \quad \int \tfrac{1}{2} \|x - y\|^2 \, \mathrm{d}\pi(x, y) \tag{1}$$

where $\Pi(P, Q)$ is the set of joint probability measures with marginals P and Q. The following fundamental result characterizes the optimal solution to (1).

Theorem 3 (Brenier's theorem). *Suppose that P admits a density with respect to Lebesgue measure. Then, there exists a proper, convex, lower semicontinuous function $\varphi_0 : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ such that the optimal transport plan in (1) can be written $\pi_0 = (\mathrm{id}, \nabla\varphi_0)_\sharp P$. The function φ_0 is called the* Brenier potential, *and the mapping $\nabla\varphi_0$ is called the* optimal transport map *from P to Q. Moreover, the optimal transport map $\nabla\varphi_0$ is unique up to P-almost everywhere equality.*

The Brenier potential φ_0 is obtained as the solution to the dual problem

$$\underset{\varphi \in \Gamma_0}{\text{maximize}} \quad \int \left(\frac{\|\cdot\|^2}{2} - \varphi \right) \mathrm{d}P + \int \left(\frac{\|\cdot\|^2}{2} - \varphi^* \right) \mathrm{d}Q, \tag{2}$$

where φ^* is the convex conjugate to φ, and Γ_0 is the set of proper, convex, lower semicontinuous functions on $\mathbb{R}^d$.

We refer to [36] for further background.

2.3. *Optimal transport with entropic regularization*

We recall that *entropic optimal transport* is the problem that arises when we add the Kullback–Leibler (KL) divergence, $D_{\mathrm{KL}}(\cdot \,\|\, \cdot)$, as a regularizer to (1):

$$\underset{\pi \in \Pi(P,Q)}{\text{minimize}} \quad \int \tfrac{1}{2} \|x - y\|^2 \, \mathrm{d}\pi(x, y) + \varepsilon \, D_{\mathrm{KL}}(\pi \,\|\, P \otimes Q) \,. \tag{3}$$

The following statement characterizes the solution to (3) [4, 10, 29].

Theorem 4 (Entropic optimal transport). *Let P and Q be probability measures on $\mathbb{R}^d$ and fix $\varepsilon > 0$. Then there exists a unique solution $\pi_\varepsilon \in \Pi(P,Q)$ to (3). Moreover, π_ε has the form*

$$\pi_\varepsilon(\mathrm{d}x, \mathrm{d}y) = \exp\left(\frac{f_\varepsilon(x) + g_\varepsilon(y) - \tfrac{1}{2}\|x - y\|^2}{\varepsilon} \right) P(\mathrm{d}x)\, Q(\mathrm{d}y), \tag{4}$$

where $(f_\varepsilon, g_\varepsilon)$ are maximizers for the dual problem

$$\underset{(f,g) \in L^1(P) \times L^1(Q)}{\text{maximize}} \quad \int f \, \mathrm{d}P + \int g \, \mathrm{d}Q - \varepsilon \iint e^{(f(x) + g(y) - \tfrac{1}{2}\|x - y\|^2)/\varepsilon} \, \mathrm{d}P(x)\, \mathrm{d}Q(y) + \varepsilon \,. \tag{5}$$

The constraint that π_ε has marginals P and Q implies the following dual optimality conditions for $(f_\varepsilon, g_\varepsilon)$ (see [4, 25] for more details):

$$f_\varepsilon(x) = -\varepsilon \log \int e^{(g_\varepsilon(y) - \tfrac{1}{2}\|x - y\|^2)/\varepsilon} \, \mathrm{d}Q(y) \qquad \left(x \in \mathbb{R}^d\right), \tag{6}$$

$$g_\varepsilon(y) = -\varepsilon \log \int e^{(f_\varepsilon(x) - \tfrac{1}{2}\|x - y\|^2)/\varepsilon} \, \mathrm{d}P(x) \qquad \left(y \in \mathbb{R}^d\right). \tag{7}$$

In particular, f_ε and g_ε are smooth. In this work, it is more convenient to work with the *entropic Brenier potentials*, defined as

$$(\varphi_\varepsilon, \psi_\varepsilon) := \left(\frac{1}{2}\|\cdot\|^2 - f_\varepsilon, \ \frac{1}{2}\|\cdot\|^2 - g_\varepsilon \right). \tag{8}$$

Since $(f_\varepsilon, g_\varepsilon)$ are only unique up to adding a constant to f_ε and subtracting the same constant from g_ε, we fix the normalization convention $\int f_\varepsilon \, \mathrm{d}P = \int g_\varepsilon \, \mathrm{d}Q$. Under this condition, it was shown by Nutz and Wiesel in [4] that we have convergence to the Brenier potential $\varphi_\varepsilon \to \varphi_0$ as $\varepsilon \searrow 0$; we recall an abbreviated version of the statement for the convenience of the reader:

Theorem 5. *For any choice of regularization parameter $\varepsilon > 0$, let $(\varphi_\varepsilon, \psi_\varepsilon)$ be the unique entropic Brenier potentials with the normalization condition*

$$\int \left(\frac{1}{2}\|\cdot\|^2 - \varphi_\varepsilon \right) \mathrm{d}P = \int \left(\frac{1}{2}\|\cdot\|^2 - \psi_\varepsilon \right) \mathrm{d}Q \,.$$

If (φ_0, φ_0^*) are unique, it holds that $\lim_{\varepsilon \searrow 0} \varphi_\varepsilon = \varphi_0$ in $L^1(P)$ and $\lim_{\varepsilon \searrow 0} \psi_\varepsilon = \varphi_0^*$ in $L^1(Q)$.

Adopting this new notation, with $P = \exp(-V)$ and $Q = \exp(-W)$, we can rewrite the entropic optimal plan as

$$\pi_\varepsilon(\mathrm{d}x, \mathrm{d}y) = \exp\left(-\frac{\varphi_\varepsilon(x) + \psi_\varepsilon(y) - \langle x, y \rangle}{\varepsilon} - V(x) - W(y) \right) \mathrm{d}x \, \mathrm{d}y \,.$$

The entropic Brenier potentials were first introduced to develop a computationally tractable estimator of the optimal transport map $\nabla \varphi_0$ [30, 31, 34]. Indeed, this is motivated by the following observation, which acts as an entropic version of Brenier's theorem. Write $\pi_\varepsilon^{Y|X=x}$ for the conditional distribution of Y given $X = x$ for $(X, Y) \sim \pi_\varepsilon$, and similarly define $\pi_\varepsilon^{X|Y=y}$. Then, by [31, Proposition 1], $\nabla \varphi_\varepsilon$ is the barycentric projection

$$\nabla \varphi_\varepsilon(x) = \int y \, \mathrm{d}\pi_\varepsilon^{Y|X=x}(y) \,. \tag{9}$$

For clarity, we abuse notation and abbreviate $\pi_\varepsilon^{Y|X=x}$ by π_ε^x and $\pi_\varepsilon^{X|Y=y}$ by π_ε^y when there is no danger of confusion.

The following lemma is a straightforward computation using (4), (6), and (7).

Lemma 6. *It holds that*

$$\nabla^2 \varphi_\varepsilon(x) = \varepsilon^{-1} \operatorname{Cov}_{Y \sim \pi_\varepsilon^x}(Y), \qquad \text{and} \qquad \nabla^2 \psi_\varepsilon(y) = \varepsilon^{-1} \operatorname{Cov}_{X \sim \pi_\varepsilon^y}(X).$$

In particular, both φ_ε and ψ_ε are convex. Moreover, under our regularity conditions,

$$\nabla_y^2 \log\left(1/\pi_\varepsilon^x\right)(y) = \varepsilon^{-1} \nabla^2 \psi_\varepsilon(y) + \nabla^2 W(y),$$
$$\nabla_x^2 \log\left(1/\pi_\varepsilon^y\right)(x) = \varepsilon^{-1} \nabla^2 \varphi_\varepsilon(x) + \nabla^2 V(x).$$

2.4. *Covariance inequalities*

In our proofs, we make use of the following key inequalities.

Lemma 7. *Let $P = \exp(-V)$ be a probability measure on $\mathbb{R}^d$ and assume that V is twice continuously differentiable on the interior of its domain. Then, the following hold.*

(1) *(Brascamp–Lieb inequality) If in addition we assume that P is strictly log-concave, then it holds that*

$$\operatorname{Cov}_{X \sim P}(X) \preceq \mathbb{E}_{X \sim P}\left[\left(\nabla^2 V(X)\right)^{-1}\right].$$

(2) *(Cramér–Rao inequality)*

$$\operatorname{Cov}_{X \sim P}(X) \succeq \left(\mathbb{E}_{X \sim P}\left[\nabla^2 V(X)\right]\right)^{-1}.$$

The Brascamp–Lieb inequality is classical, and we refer readers to [3, 5, 9] for several proofs. To make our exposition more self-contained, we provide a proof of the Cramér–Rao inequality in the appendix.

3. Main theorem

We now state and prove our main theorem.

Theorem 8. *Let $P = \exp(-V)$ and $Q = \exp(-W)$.*

(1) *Suppose that (P, Q) satisfy our regularity assumptions, as well as*

$$\nabla^2 V \preceq \beta_V I, \qquad \text{and} \qquad \nabla^2 W \succeq \alpha_W I > 0.$$

Then, for every $\varepsilon > 0$ and all $x \in \mathbb{R}^d$, the Hessian of the entropic Brenier potential satisfies

$$\nabla^2 \varphi_\varepsilon(x) \preceq \frac{1}{2}\left(\sqrt{4\beta_V/\alpha_W + \varepsilon^2 \beta_V^2} - \varepsilon\beta_V\right) I.$$

(2) *Suppose that (Q, P) satisfy our regularity assumptions, as well as*

$$\nabla^2 V \succeq \alpha_V I > 0, \qquad \text{and} \qquad \nabla^2 W \preceq \beta_W I.$$

Then, for every $\varepsilon > 0$ and all $x \in \Omega_P := \operatorname{int}(\operatorname{supp}(P))$, the Hessian of the entropic Brenier potential satisfies

$$\nabla^2 \varphi_\varepsilon(x) \succeq \frac{1}{2}\left(\sqrt{4\alpha_V/\beta_W + \varepsilon^2 \alpha_V^2} - \varepsilon\alpha_V\right) I.$$

Observe that as $\varepsilon \searrow 0$, we formally expect the following bounds on the Brenier potential:

$$\sqrt{\alpha_V/\beta_W}\, I \preceq \nabla^2 \varphi_0(x) \preceq \sqrt{\beta_V/\alpha_W}\, I.$$

In particular, this recovers Caffarelli's contraction theorem (Theorem 1). We make this intuition rigorous below by appealing to convergence results for the entropic potentials as the regularization parameter ε tends to zero.

Proof of Theorem 8 - Upper bound. Fix $x \in \mathbb{R}^d$. Recall from Lemma 6 that

$$\nabla^2 \varphi_\varepsilon(x) = \varepsilon^{-1} \operatorname{Cov}_{Y \sim \pi_\varepsilon^x}(Y).$$

By an application of the Brascamp–Lieb inequality, this results in the upper bound

$$
\begin{aligned}
\nabla^2 \varphi_\varepsilon(x) &= \varepsilon^{-1} \operatorname{Cov}_{Y \sim \pi_\varepsilon^x}(Y) \\
&\leq \varepsilon^{-1} \mathbb{E}_{Y \sim \pi_\varepsilon^x} \left[\left(\varepsilon^{-1} \nabla^2 \psi_\varepsilon(Y) + \nabla^2 W(Y) \right)^{-1} \right] \\
&\leq \mathbb{E}_{Y \sim \pi_\varepsilon^x} \left[\left(\nabla^2 \psi_\varepsilon(Y) + \varepsilon \alpha_W I \right)^{-1} \right],
\end{aligned}
\tag{10}
$$

where in the last inequality we also used the lower bound on the spectrum of $\nabla^2 W$. Next, using Lemma 6 and the Cramér–Rao inequality (Lemma 7), we obtain the lower bound

$$
\begin{aligned}
\nabla^2 \psi_\varepsilon(Y) &= \varepsilon^{-1} \operatorname{Cov}_{X \sim \pi_\varepsilon^Y}(X) \\
&\geq \varepsilon^{-1} \left(\mathbb{E}_{X \sim \pi_\varepsilon^Y} \left[\varepsilon^{-1} \nabla^2 \varphi_\varepsilon(X) + \nabla^2 V(X) \right] \right)^{-1} \\
&\geq \left(\mathbb{E}_{X \sim \pi_\varepsilon^Y} \left[\nabla^2 \varphi_\varepsilon(X) + \varepsilon \beta_V I \right] \right)^{-1},
\end{aligned}
$$

where we used the upper bound on the spectrum of $\nabla^2 V$. Combining these inequalities,

$$\nabla^2 \varphi_\varepsilon(x) \leq \mathbb{E}_{Y \sim \pi_\varepsilon^x} \left[\left(\left(\mathbb{E}_{X \sim \pi_\varepsilon^Y} \left[\nabla^2 \varphi_\varepsilon(X) + \varepsilon \beta_V I \right] \right)^{-1} + \varepsilon \alpha_W I \right)^{-1} \right].$$

Now, define the quantity

$$L_\varepsilon := \sup_{x \in \mathbb{R}^d} \lambda_{\max} \left(\nabla^2 \varphi_\varepsilon(x) \right).$$

From (10) and the fact that ψ_ε is convex (Lemma 6), it follows that L_ε is finite: $L_\varepsilon \leq (\varepsilon \alpha_W)^{-1}$. Then, we have shown

$$\lambda_{\max} \left(\nabla^2 \varphi_\varepsilon(x) \right) \leq \left(\left(L_\varepsilon + \varepsilon \beta_V \right)^{-1} + \varepsilon \alpha_W \right)^{-1}.$$

Taking the supremum over $x \in \mathbb{R}^d$,

$$L_\varepsilon \leq \left(\left(L_\varepsilon + \varepsilon \beta_V \right)^{-1} + \varepsilon \alpha_W \right)^{-1}.$$

Solving the inequality yields

$$L_\varepsilon \leq \frac{1}{2} \left(\sqrt{4 \beta_V / \alpha_W + \varepsilon^2 \beta_V^2} - \varepsilon \beta_V \right).
\tag{11}$$

Lower bound. The lower bound argument is symmetric, but we give the details for completeness. Using Lemma 6 and the Cramér–Rao inequality (Lemma 7),

$$
\begin{aligned}
\nabla^2 \varphi_\varepsilon(x) &= \varepsilon^{-1} \operatorname{Cov}_{Y \sim \pi_\varepsilon^x}(Y) \\
&\geq \varepsilon^{-1} \left(\mathbb{E}_{Y \sim \pi_\varepsilon^x} \left[\varepsilon^{-1} \nabla^2 \psi_\varepsilon(Y) + \nabla^2 W(Y) \right] \right)^{-1} \\
&\geq \left(\mathbb{E}_{Y \sim \pi_\varepsilon^x} \left[\nabla^2 \psi_\varepsilon(Y) + \varepsilon \beta_W I \right] \right)^{-1}.
\end{aligned}
$$

Applying Lemma 6 and the Brascamp–Lieb inequality (Lemma 7),

$$
\begin{aligned}
\nabla^2 \psi_\varepsilon(Y) &= \varepsilon^{-1} \operatorname{Cov}_{X \sim \pi_\varepsilon^Y}(X) \\
&\leq \varepsilon^{-1} \mathbb{E}_{X \sim \pi_\varepsilon^Y} \left[\left(\varepsilon^{-1} \nabla^2 \varphi_\varepsilon(X) + \nabla^2 V(X) \right)^{-1} \right] \\
&\leq \mathbb{E}_{X \sim \pi_\varepsilon^Y} \left[\left(\nabla^2 \varphi_\varepsilon(X) + \varepsilon \alpha_V I \right)^{-1} \right].
\end{aligned}
$$

Combining the two inequalities and setting

$$\ell_\varepsilon := \inf_{x \in \Omega_P} \lambda_{\min} \left(\nabla^2 \varphi_\varepsilon(x) \right),$$

we deduce that

$$\ell_\varepsilon \geq \left((\ell_\varepsilon + \varepsilon \alpha_V)^{-1} + \varepsilon \beta_W\right)^{-1}.$$

On the other hand, from Lemma 6, we know that $\ell_\varepsilon \geq 0$. Solving the inequality then yields

$$\ell_\varepsilon \geq \frac{1}{2}\left(\sqrt{4\alpha_V/\beta_W + \varepsilon^2 \alpha_V^2} - \varepsilon \alpha_V\right). \qquad \square$$

Next, we rigorously deduce Caffarelli's contraction theorem from Theorem 8.

Proof of Caffarelli's contraction (Theorem 1). For every $\varepsilon > 0$, by Theorem 8, we have proven that $\nabla^2 \varphi_\varepsilon \preceq L_\varepsilon I$, with L_ε as in (11). Equivalently, this can be reformulated as saying that $\frac{L_\varepsilon \|\cdot\|^2}{2} - \varphi_\varepsilon$ is convex. Fix some $\delta > 0$; in particular, for ε sufficiently small, $\frac{(\sqrt{\beta_V/\alpha_W}+\delta)\|\cdot\|^2}{2} - \varphi_\varepsilon$ is convex.

Upon passing to a sequence $\varepsilon_k \searrow 0$, existing results on the convergence of entropic optimal transport potentials show that $\varphi_{\varepsilon_k} \to \varphi_0$ in $L^1(P)$ (see Theorem 5). Passing to a further subsequence, we obtain $\varphi_{\varepsilon_k} \to \varphi_0$ (P-almost surely). It follows that $\frac{(\sqrt{\beta_V/\alpha_W}+\delta)\|\cdot\|^2}{2} - \varphi_0$ is convex for every $\delta > 0$ (see the remark after [33, Theorem 25.7]), and thus for $\delta = 0$. $\qquad \square$

Remark 9. Our main theorem provides both upper and lower bounds for $\nabla^2 \varphi_\varepsilon$. In the case when $\varepsilon = 0$, the lower bound follows from the upper bound. Indeed, if φ_0 is the Brenier potential for the optimal transport from P to Q, then the convex conjugate φ_0^* is the Brenier potential for the optimal transport from Q to P. By applying Caffarelli's contraction theorem to φ_0^* and appealing to convex duality, it yields a lower bound on $\nabla^2 \varphi_0$. However, we are not aware of a method of deducing the lower bound from the upper bound for positive values of ε.

Remark 10. In Appendix B, by inspecting the Gaussian case, we show that Theorem 8 is sharp for every $\varepsilon > 0$.

Remark 11. In the proof of Theorem 8, we do not use the full force of the Brascamp–Lieb inequality. Rather, we use the covariance inequality in Lemma 7 which is a corollary of the usual Brascamp–Lieb inequality obtained by applying it to linear test functions.

An inspection of the proof of the upper bound in Theorem 8 reveals the following more general pair of inequalities.

Proposition 12. *Let (P, Q) be probability measures satisfying our regularity conditions. Then, for all $x \in \mathbb{R}^d$, $y \in \Omega_Q$,*

$$\nabla^2 \varphi_\varepsilon(x) \preceq \mathbb{E}_{Y \sim \pi_\varepsilon^x}\left[\left(\nabla^2 \psi_\varepsilon(Y) + \varepsilon \nabla^2 W(Y)\right)^{-1}\right],$$

$$\nabla^2 \psi_\varepsilon(y) \succeq \left(\mathbb{E}_{X \sim \pi_\varepsilon^y}\left[\nabla^2 \varphi_\varepsilon(X) + \varepsilon \nabla^2 V(X)\right]\right)^{-1}.$$

In the next section, we use these inequalities to prove a generalization of Caffarelli's theorem.

4. A generalization to commuting positive definite matrices

In the next result, we replace the main assumptions of Caffarelli's theorem, namely $\nabla^2 V \preceq \beta_V I$ and $\nabla^2 W \succeq \alpha_W I$, by the conditions

$$\nabla^2 V \preceq A^{-1} \quad \text{and} \quad \nabla^2 W \succeq B^{-1}, \tag{12}$$

where A and B are commuting positive definite matrices. Recall that the Hessian of the Brenier potential between the Gaussian distributions $\mathcal{N}(0, A)$ and $\mathcal{N}(0, B)$ is the matrix $A^{-1/2} B^{1/2}$ [13]. In light of this observation, the following theorem is sharp for every pair of commuting positive definite (A, B), and shows that the Brenier potential between Gaussians achieves the largest possible Hessian among all source and target measures obeying the constraint (12).

Theorem 13. *Let (P, Q) satisfy our regularity conditions as well as the condition* (12). *Then, the Hessian of the Brenier potential satisfies the uniform bound: for all $x \in \mathbb{R}^d$, it holds that*

$$\nabla^2 \varphi_0(x) \preceq A^{-1/2} B^{1/2}.$$

As in Theorem 8, the proof technique also yields a lower bound on $\nabla^2 \varphi_0$ under appropriate assumptions. We omit this result because it is straightforward.

Proof. Let C_ε be the smallest constant $C \geq 0$ such that $\nabla^2 \varphi_\varepsilon(x) \preceq A^{-1/2} B^{1/2} + CI$ for all $x \in \mathbb{R}^d$. In light of Theorem 8, C_ε is well-defined and finite. Equivalently,

$$C_\varepsilon = \sup_{x \in \mathbb{R}^d} \sup_{e \in \mathbb{R}^d, \, \|e\|=1} \left\langle e, \left[\nabla^2 \varphi_\varepsilon(x) - A^{-1/2} B^{1/2}\right] e \right\rangle.$$

Let (x, e) achieve the above supremum. Using our assumptions and Proposition 12, we obtain

$$
\begin{aligned}
C_\varepsilon &= \left\langle e, \left[\nabla^2 \varphi_\varepsilon(x) - A^{-1/2} B^{1/2}\right] e \right\rangle \\
&\leq \left\langle e, \left[\left(\mathbb{E}_{Y \sim \pi_\varepsilon^x} \nabla^2 \psi_\varepsilon(Y) + \varepsilon B^{-1}\right)^{-1} - A^{-1/2} B^{1/2}\right] e \right\rangle \\
&\leq \left\langle e, \left[\left(\left(A^{-1/2} B^{1/2} + C_\varepsilon I + \varepsilon A^{-1}\right)^{-1} + \varepsilon B^{-1}\right)^{-1} - A^{-1/2} B^{1/2}\right] e \right\rangle.
\end{aligned}
$$

From our assumptions and Theorem 8, we know that the spectrum of $M_\varepsilon := A^{-1/2} B^{1/2} + C_\varepsilon I$ is bounded away from zero and infinity as $\varepsilon \searrow 0$, which justifies the Taylor expansion

$$
\begin{aligned}
\left(\left(M_\varepsilon + \varepsilon A^{-1}\right)^{-1} + \varepsilon B^{-1}\right)^{-1} &= \left(M_\varepsilon^{-1} - \varepsilon M_\varepsilon^{-1} A^{-1} M_\varepsilon^{-1} + \varepsilon B^{-1} + O(\varepsilon^2)\right)^{-1} \\
&= M_\varepsilon + \varepsilon A^{-1} - \varepsilon M_\varepsilon B^{-1} M_\varepsilon + O(\varepsilon^2) I.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
C_\varepsilon &\leq \left\langle e, \left[M_\varepsilon + \varepsilon A^{-1} - \varepsilon M_\varepsilon B^{-1} M_\varepsilon + O(\varepsilon^2) I - A^{-1/2} B^{1/2}\right] e \right\rangle \\
&\leq C_\varepsilon + \varepsilon \left\langle e, \left[A^{-1} - M_\varepsilon B^{-1} M_\varepsilon\right] e \right\rangle + O(\varepsilon^2) \\
&= C_\varepsilon - \varepsilon \left\langle e, \left[2 C_\varepsilon A^{-1/2} B^{1/2} + C_\varepsilon^2 B^{-1}\right] e \right\rangle + O(\varepsilon^2).
\end{aligned}
$$

This shows that $\lim_{\varepsilon \searrow 0} C_\varepsilon = 0$ (otherwise $(C_\varepsilon)_{\varepsilon > 0}$ would have a strictly positive cluster point which would contradict the above inequality for small enough $\varepsilon > 0$).

By combining this fact with convergence of the entropic Brenier potentials as in the proof of Theorem 1, we deduce the result. $\square$

Next, we show how our theorem recovers and extends a result of Valdimarsson [35], which was used to derive new forms of the Brascamp–Lieb inequality.[2]

Theorem 14. *Suppose that*

- $\bar{A}, \bar{B},$ *and G are positive definite matrices;*
- $\bar{A} \preceq G$ *and $\bar{B}$ commutes with G;*
- $P = \exp(-\widetilde{V}) * \mu$, *where $\nabla^2 \widetilde{V} \preceq \bar{B}^{-1} G$, $*$ denotes convolution, and μ is an arbitrary probability measure on $\mathbb{R}^d$;*
- $Q = \exp(-W)$ *with $\nabla^2 W \succeq \bar{B}^{-1/2} \bar{A}^{-1} \bar{B}^{-1/2}$.*

Then, the Brenier potential satisfies $\nabla^2 \varphi_0 \preceq G$.

Remark 15. Valdimarsson's original result required that $P = \mathcal{N}(0, \bar{B} G^{-1}) * \mu$.

To prove this result, we check that convolution with any probability measure only makes the density more log-smooth.

---

[2]This is a different Brascamp–Lieb inequality than the one in Lemma 7.

Lemma 16. *Let $\widetilde{P} \propto \exp(-\widetilde{V})$ be a probability measure, where $\widetilde{V} : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable. Let $P := \widetilde{P} * \mu = \exp(-V)$ where μ is any probability measure on $\mathbb{R}^d$. Suppose that for some positive definite matrix A^{-1}, we have $\nabla^2 \widetilde{V} \preceq A^{-1}$. Then, $\nabla^2 V \preceq A^{-1}$ as well.*

Proof. An elementary computation shows that if we define the probability measure

$$\nu_y(\mathrm{d}x) := \frac{\exp\left(-\widetilde{V}(y-x)\right)\mu(\mathrm{d}x)}{\int \exp\left(-\widetilde{V}(y-x')\right)\mu(\mathrm{d}x')}$$

then

$$\nabla^2 V(y) = \mathbb{E}_{X \sim \nu_y}\left[\nabla^2 \widetilde{V}(y-X)\right] - \mathrm{Cov}_{X \sim \nu_y}\left(\nabla \widetilde{V}(y-X)\right),$$

from which the result follows. $\qquad\square$

Proof of Theorem 14. Under Lemma 16 and the third assumption, it holds that $P \propto \exp(-V)$ with $\nabla^2 V \preceq \bar{B}^{-1}G$. The other assumptions imply that $Q \propto \exp(-W)$ with

$$\nabla^2 W \succeq \bar{B}^{-1/2}\bar{A}^{-1}\bar{B}^{-1/2} \succeq \bar{B}^{-1/2}G^{-1}\bar{B}^{-1/2} = \bar{B}^{-1}G^{-1}.$$

By Theorem 13, it holds that $\nabla^2 \varphi_0 \preceq G$. $\qquad\square$

Remark 17. It is natural to ask whether Theorem 13 can be obtained by first applying Caffarelli's contraction theorem to show that the optimal transport map $\widetilde{T}_0$ between the measures $(A^{-1/2})_\sharp P$ and $(B^{-1/2})_\sharp Q$ is 1-Lipschitz, and then considering the mapping $T_0(x) := B^{1/2}\widetilde{T}_0(A^{-1/2}x)$. Although T_0 is indeed a valid transport mapping from P to Q, under our assumptions ∇T_0 is not guaranteed to be symmetric, so it does not make sense to ask that $\nabla T_0 \preceq A^{-1/2}B^{1/2}$.

In Valdimarsson's application to Brascamp–Lieb inequalities, it is crucial that the transport map T_0 is chosen so that ∇T_0 is a symmetric positive definite matrix. Symmetry of ∇T_0 implies that T_0 is the gradient $\nabla \varphi_0$ of a function $\varphi_0 : \mathbb{R}^d \to \mathbb{R}$, and positive definiteness implies that φ_0 is convex. By Brenier's theorem, the unique gradient of a convex function that pushes forward P to Q is the optimal transport map. Thus, it is crucial that we consider the *optimal* transport map here; in particular, alternative maps such as the ones in [18, 26] cannot be applied.

5. Discussion

We have proven a generalization of Caffarelli's celebrated theorem on the Lipschitz properties of the optimal transport map to the setting of entropic optimal transport using two complementary covariance inequalities (the Brascamp–Lieb inequality and the Cramér–Rao inequality).

We conjecture that our proof technique can also be used to recover the bounds on the moment measure mapping in [19], provided that the existence of an "entropic moment measure" can be established (with convergence towards the true moment measure as the regularization tends to zero). As this is outside the scope of this work, we do not pursue this question here.

Appendix A. Proof of the Cramér–Rao lower bound

Proof of Lemma 7, Cramér–Rao inequality. For any smooth and compactly supported test function $h : \mathbb{R}^d \to \mathbb{R}$, integration by parts yields

$$\mathbb{E}_P \nabla h = \int \nabla h \, \mathrm{d}P = -\int (h \nabla \ln P) \, \mathrm{d}P = \int (h - \mathbb{E}_P h) \, \nabla V \, \mathrm{d}P$$

where we used the fact that $\mathbb{E}_P \nabla \ln P = 0$. Therefore,

$$\left\langle \mathbb{E}_P \nabla h, \left(\mathbb{E}_P \nabla^2 V\right)^{-1} \mathbb{E}_P \nabla h \right\rangle = \int (h - \mathbb{E}_P h) \left\langle \nabla V, \left(\mathbb{E}_P \nabla^2 V\right)^{-1} \mathbb{E}_P \nabla h \right\rangle \, \mathrm{d}P. \tag{13}$$

Applying the Cauchy–Schwarz inequality,

$$(13) \le \sqrt{(\operatorname{Var}_P h) \int \left\langle \mathbb{E}_P \nabla h, \left(\mathbb{E}_P \nabla^2 V\right)^{-1} (\nabla V)^{\otimes 2} \left(\mathbb{E}_P \nabla^2 V\right)^{-1} \mathbb{E}_P \nabla h \right\rangle \, \mathrm{d}P}.$$

Integration by parts shows that $\int \nabla V^{\otimes 2} \, \mathrm{d}P = \int \nabla^2 V \, \mathrm{d}P$, and upon rearranging we deduce that

$$\operatorname{Var}_P h \ge \left\langle \mathbb{E}_P \nabla h, \left(\mathbb{E}_P \nabla^2 V\right)^{-1} \mathbb{E}_P \nabla h \right\rangle. \tag{14}$$

By approximation, this continues to hold for any locally Lipschitz $h : \mathbb{R}^d \to \mathbb{R}$ with $\mathbb{E}_P \lVert \nabla h \rVert < \infty$.

Specializing the inequality (14) to $h := \langle e, \cdot \rangle$ for a unit vector $e \in \mathbb{R}^d$ then recovers the Cramér–Rao inequality of Lemma 7. $\qquad\square$

Appendix B. Gaussian case

Suppose $P = \mathcal{N}(0, A)$ and $Q = \mathcal{N}(0, B)$ are Gaussians. Then, it is known that the Hessian of the Brenier potential is given by [13]

$$\nabla^2 \varphi_0(x) = A^{-1/2} \left(A^{1/2} B A^{1/2}\right)^{1/2} A^{-1/2}.$$

If we have

$$A^{-1} \preceq \beta I \qquad \text{and} \qquad B^{-1} \succeq \alpha I > 0,$$

then Caffarelli's contraction theorem (Theorem 1) implies

$$\left\lVert \nabla^2 \varphi_0 \right\rVert_{\mathrm{op}} \le \sqrt{\beta/\alpha}.$$

This matches the bound of [2, Lemma 2].

For $\varepsilon > 0$, the upper bound from Theorem 8 implies

$$\left\lVert \nabla^2 \varphi_\varepsilon \right\rVert_{\mathrm{op}} \le \frac{1}{2} \left(\sqrt{4\beta/\alpha + \varepsilon^2 \beta^2} - \varepsilon\beta \right). \tag{15}$$

On the other hand, from [17, 24], it is known that

$$\nabla^2 \varphi_\varepsilon(x) = A^{-1/2} \left(A^{1/2} B A^{1/2} + \frac{\varepsilon^2}{4} I\right)^{1/2} A^{-1/2} - \frac{\varepsilon}{2} A^{-1}.$$

In particular, if we take $A = \beta^{-1} I$ and $B = \alpha^{-1} I$, then (15) is an equality. Hence, Theorem 8 is sharp for every $\varepsilon > 0$.

References

[1] A. Ahidar-Coutrix, T. Le Gouic, Q. Paris, "Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics", *Probab. Theory Relat. Fields* **177** (2020), no. 1-2, p. 323-368.

[2] J. Altschuler, S. Chewi, P. Gerber, A. Stromme, "Averaging on the Bures–Wasserstein manifold: dimension-free convergence of gradient descent", Advances in Neural Information Processing Systems, vol. 34, Curran Associates, Inc., 2021.

[3] D. Bakry, I. Gentil, M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften, vol. 348, Springer, 2014.

[4] E. Bernton, P. Ghosal, M. Nutz, "Entropic optimal transport: geometry and large deviations", *Duke Math. J.* **171** (2022), no. 16, p. 3363-3400.

[5] S. G. Bobkov, M. Ledoux, "From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities", *Geom. Funct. Anal.* **10** (2000), no. 5, p. 1028-1052.

[6] L. A. Caffarelli, "Monotonicity properties of optimal transportation and the FKG and related inequalities", *Commun. Math. Phys.* **214** (2000), no. 3, p. 547-563.

[7] S. Chewi, T. Maunu, P. Rigollet, A. J. Stromme, "Gradient descent algorithms for Bures–Wasserstein barycenters", in *Proceedings of Thirty Third Conference on Learning Theory* (J. Abernethy, S. Agarwal, eds.), Proceedings of Machine Learning Research, vol. 125, PMLR, 2020, p. 1276-1304.

[8] M. Colombo, A. Figalli, Y. Jhaveri, "Lipschitz changes of variables between perturbations of log-concave measures", *Ann. Sc. Norm. Super. Pisa, Cl. Sci.* **17** (2017), no. 4, p. 1491-1519.

[9] D. Cordero-Erausquin, "Transport inequalities for log-concave measures, quantitative forms, and applications", *Can. J. Math.* **69** (2017), no. 3, p. 481-501.

[10] I. Csiszár, "*I*-divergence geometry of probability distributions and minimization problems", *Ann. Probab.* **3** (1975), p. 146-158.

[11] M. Cuturi, "Sinkhorn distances: lightspeed computation of optimal transport", Advances in Neural Information Processing Systems, vol. 26, Curran Associates, Inc., 2013.

[12] M. Fathi, N. Gozlan, M. Prod'homme, "A proof of the Caffarelli contraction theorem via entropic regularization", *Calc. Var. Partial Differ. Equ.* **59** (2020), no. 3, article no. 96.

[13] M. Gelbrich, "On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces", *Math. Nachr.* **147** (1990), no. 1, p. 185-203.

[14] I. Gentil, C. Léonard, L. Ripani, L. Tamanini, "An entropic interpolation proof of the HWI inequality", *Stochastic Processes Appl.* **130** (2020), no. 2, p. 907-923.

[15] N. Gozlan, N. Juillet, "On a mixture of Brenier and Strassen theorems", *Proc. Lond. Math. Soc.* **120** (2020), no. 3, p. 434-463.

[16] J.-C. Hütter, P. Rigollet, "Minimax estimation of smooth optimal transport maps", *Ann. Stat.* **49** (2021), no. 2, p. 1166-1194.

[17] H. Janati, B. Muzellec, G. Peyré, M. Cuturi, "Entropic optimal transport between unbalanced Gaussian measures has a closed form", Advances in Neural Information Processing Systems, vol. 33, Curran Associates, Inc., 2020.

[18] Y.-H. Kim, E. Milman, "A generalization of Caffarelli's contraction theorem via (reverse) heat flow", *Math. Ann.* **354** (2012), no. 3, p. 827-862.

[19] B. Klartag, "Logarithmically-concave moment measures I", in *Geometric aspects of functional analysis*, Lecture Notes in Mathematics, vol. 2116, Springer, 2014, p. 231-260.

[20] A. V. Kolesnikov, "Mass transportation and contractions", (2011), https://arxiv.org/abs/1103.1479.

[21] T. Le Gouic, Q. Paris, P. Rigollet, A. J. Stromme, "Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space", *J. Eur. Math. Soc.* **25** (2023), no. 6, p. 2229-2250.

[22] M. Ledoux, "Remarks on some transportation cost inequalities", 2018.

[23] C. Léonard, "A survey of the Schrödinger problem and some of its connections with optimal transport", *Discrete Contin. Dyn. Syst.* **34** (2014), no. 4, p. 1533-1574.

[24] A. Mallasto, A. Gerolin, H. Q. Minh, "Entropy-regularized 2-Wasserstein distance between Gaussian measures", *Inf. Geom.* **5** (2022), no. 1, p. 289-323.

[25] G. Mena, J. Niles-Weed, "Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem", Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc., 2019.

[26] D. Mikulincer, Y. Shenfeld, "The Brownian transport map", 2021, https://arxiv.org/abs/2111.11521.

[27] ———, "On the Lipschitz properties of transportation along heat flows", 2022, https://arxiv.org/abs/2201.01382.

[28] J. Neeman, "Lipschitz changes of variables via heat flow", 2022, https://arxiv.org/abs/2201.03403.

[29] G. Peyré, M. Cuturi, "Computational optimal transport", *Found. Trends Mach. Learn.* **11** (2019), no. 5-6, p. 355-607.

[30] A.-A. Pooladian, M. Cuturi, J. Niles-Weed, "Debiaser beware: pitfalls of centering regularized transport maps" (K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, S. Sabato, eds.), Proceedings of Machine Learning Research, vol. 162, Curran Associates, Inc., 2022, p. 17830-17847.

[31] A.-A. Pooladian, J. Niles-Weed, "Entropic estimation of optimal transport maps", 2021, https://arxiv.org/abs/2109.12004v1.

[32] M. Prod'homme, "Contributions au problème du transport optimal et à sa régularité", PhD Thesis, 2021, Thèse de doctorat dirigée par Max Fathi et Felix Otto, Mathématiques et Applications Toulouse 3 2021.

[33] R. T. Rockafellar, *Convex analysis*, Princeton Landmarks in Mathematics, Princeton University Press, 1997, reprint of the 1970 original, Princeton Paperbacks.

[34] V. Seguy, B. B. Damodaran, R. Flamary, N. Courty, A. Rolet, M. Blondel, "Large-scale optimal transport and mapping estimation", in *International Conference on Learning Representations*, Curran Associates, Inc., 2018.

[35] S. I. Valdimarsson, "On the Hessian of the optimal transport potential", *Ann. Sc. Norm. Super. Pisa, Cl. Sci.* **6** (2007), no. 3, p. 441-456.

[36] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, 2003.