

## A UNIFORM RATE OF CONVERGENCE FOR THE ENTROPIC POTENTIALS IN THE QUADRATIC EUCLIDEAN SETTING

PABLO LÓPEZ-RIVERA\* 

**Abstract.** We bound the rate of uniform convergence in compact sets for both entropic potentials and their gradients towards the Brenier potential and its gradient, respectively. Both results hold in the quadratic Euclidean setting for absolutely continuous measures satisfying some convexity assumptions.

**Mathematics Subject Classification.** 49Q22, 35J96.

Received March 21, 2025. Accepted February 5, 2026.

### 1. INTRODUCTION

#### 1.1. Optimal transport

This note concerns the quadratic Euclidean optimal transport problem between absolutely continuous measures: given two probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$  which are absolutely continuous with respect to the  $d$ -dimensional Lebesgue measure, we define the quadratic optimal transport problem associated to them, in its Kantorovich formulation, as

$$\mathcal{C}_0(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2} |x - y|^2 \, d\pi(x, y), \quad (1.1)$$

where  $\Pi(\mu, \nu)$  denotes the set of probability measures on  $\mathbb{R}^d \times \mathbb{R}^d$  with first and second marginals equal to  $\mu$  and  $\nu$ , respectively. Among excellent monographs on general optimal transport theory, we can cite [1–3]. The Brenier-McCann theorem [4, 5] states that there exists a convex function  $\varphi_0: \mathbb{R}^d \rightarrow \mathbb{R}$  such that the map  $T_0: \mathbb{R}^d \rightarrow \mathbb{R}^d$  defined by  $T_0 := \nabla \varphi_0$  pushes forward the measure  $\mu$  towards  $\nu$  and induces the unique optimal coupling  $\pi_0 \in \Pi(\mu, \nu)$  for the variational problem (1.1). We say that  $T_0$  is the Brenier or the optimal transport map and  $\varphi_0$  is a Brenier potential. Moreover, if we define  $\psi_0 := \varphi_0^*$  as the convex conjugate of  $\varphi_0$ , the pair  $(f_0, g_0) := \left(\frac{1}{2} |\cdot|^2 - \varphi_0, \frac{1}{2} |\cdot|^2 - \psi_0\right)$  solves the dual problem to (1.1):

$$\mathcal{C}_0(\mu, \nu) = \sup_{\substack{(f, g) \in L^1(\mu) \times L^1(\nu), \\ f \oplus g \leq \frac{1}{2} |\cdot|^2}} \int_{\mathbb{R}^d} f \, d\mu + \int_{\mathbb{R}^d} g \, d\nu = \int_{\mathbb{R}^d} f_0 \, d\mu + \int_{\mathbb{R}^d} g_0 \, d\nu, \quad (1.2)$$

---

*Keywords and phrases:* Optimal transport, entropic optimal transport, Schrödinger bridge.

Université Paris Cité & Sorbonne Université, CNRS, Laboratoire Jacques-Louis Lions (LJLL), F-75013 Paris, France.

\* Corresponding author: [plopez@math.univ-paris-diderot.fr](mailto:plopez@math.univ-paris-diderot.fr)

where for  $(f, g) \in L^1(\mu) \times L^1(\nu)$ , the function  $f \oplus g: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  denotes their sum, which is defined for  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$  by  $(f \oplus g)(x, y) := f(x) + g(y)$ .

## 1.2. Entropic optimal transport

We can regularize Problem (1.1) by adding an entropy to the objective function: for  $\varepsilon > 0$ , we define the entropic regularization of the optimal transport problem (equivalent as well to the Schrödinger bridge [6]) as

$$\mathcal{C}_\varepsilon(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2} |x - y|^2 d\pi(x, y) + \varepsilon \mathbf{H}(\pi | \mu \otimes \nu), \quad (1.3)$$

where  $\mathbf{H}(\cdot | \mu \otimes \nu)$  denotes the relative entropy functional with respect to the measure  $\mu \otimes \nu$ : for any  $\eta \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ ,

$$\mathbf{H}(\eta | \mu \otimes \nu) := \begin{cases} \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \frac{d\eta}{d(\mu \otimes \nu)} d\eta, & \text{if } \eta \ll \mu \otimes \nu \\ +\infty, & \text{if } \eta \not\ll \mu \otimes \nu. \end{cases}$$

An excellent introductory reference to the subject is [7].

Since the entropy functional  $\mathbf{H}(\cdot | \mu \otimes \nu)$  is strictly convex, the problem (1.3) has a unique solution that we denote by  $\pi_\varepsilon \in \Pi(\mu, \nu)$ . On the other hand, there exists a pair of functions  $(f_\varepsilon, g_\varepsilon) \in L^1(\mu) \times L^1(\nu)$ , which is unique up to a constant, that solves the dual problem to (1.3):

$$\mathcal{C}_\varepsilon(\mu, \nu) = \sup_{(f, g) \in L^1(\mu) \times L^1(\nu)} \int_{\mathbb{R}^d} f d\mu + \int_{\mathbb{R}^d} g d\nu - \varepsilon \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{\frac{f \oplus g - \frac{1}{2} |\cdot - \cdot|^2}{\varepsilon}} d(\mu \otimes \nu) + \varepsilon = \int_{\mathbb{R}^d} f_\varepsilon d\mu + \int_{\mathbb{R}^d} g_\varepsilon d\nu. \quad (1.4)$$

We define for each  $\varepsilon > 0$  the entropic potentials as  $(\varphi_\varepsilon, \psi_\varepsilon) := \left( \frac{1}{2} |\cdot|^2 - f_\varepsilon, \frac{1}{2} |\cdot|^2 - g_\varepsilon \right)$ , in analogy to their unregularized counterpart  $(\varphi_0, \psi_0)$ .

## 1.3. The connection between both problems

It is known that computing  $\varphi_0$  or  $T_0 = \nabla \varphi_0$  is difficult, as one would solve the associated Monge-Ampère equation [8], a second-order nonlinear PDE. One of the advantages of the entropic problem is that the potentials  $(\varphi_\varepsilon, \psi_\varepsilon)$  are very tractable, numerically speaking, thanks to the Sinkhorn algorithm [9–11]. On the other hand, as the regularization parameter  $\varepsilon$  vanishes, the problem (1.3) converges in many senses to (1.1), a fact which allows us to approximate the optimal transport through the entropic regularization for small values of the parameter  $\varepsilon > 0$ .

More precisely, as  $\varepsilon \rightarrow 0$ , the entropic problem  $\Gamma$ -converges towards the unregularized one [12, 13], which yields the convergence of the value functions,  $\mathcal{C}_\varepsilon(\mu, \nu) \rightarrow \mathcal{C}_0(\mu, \nu)$ , and the convergence of the optimal couplings in the weak topology,  $\pi_\varepsilon \rightarrow \pi_0$ . Concerning the dual optimizers, it is known that  $\varphi_\varepsilon \rightarrow \varphi_0$  (modulo subsequence) both in  $L^1(\mu)$  and uniformly on compact sets [14, 15] and  $\nabla \varphi_\varepsilon \rightarrow \nabla \varphi_0$  (modulo subsequence) in  $L^2(\mu)$  [16]. All of the above results use  $\Gamma$ -convergence or compactness arguments, so a natural question is whether it is possible to quantify the rate at which these convergences happen.

In the case of the convergence of the value functions, several contributions have been made in the continuous setting: the first-order expansion

$$\mathcal{C}_\varepsilon(\mu, \nu) = \mathcal{C}_0(\mu, \nu) - \frac{d}{2} \varepsilon \log \varepsilon + \frac{\varepsilon}{2} (\mathbf{H}(\mu) + \mathbf{H}(\nu)) + o(\varepsilon),$$

which holds as soon as  $\mu$  has finite Fisher information, was proven to be true for the quadratic case and as a  $\Gamma$ -limit in dimension one in [17, 18] and in higher dimensions in [19]. The same expansion was established as a

pointwise limit for strictly convex cost functions in [20] (thus generalizing the quadratic case). The second-order expansion

$$\mathcal{C}_\varepsilon(\mu, \nu) = \mathcal{C}_0(\mu, \nu) - \frac{d}{2}\varepsilon \log \varepsilon + \frac{\varepsilon}{2} (\mathbb{H}(\mu) + \mathbb{H}(\nu)) + \frac{\varepsilon^2}{8} \mathbb{I}(\mu, \nu) + o(\varepsilon^2),$$

where  $\mathbb{I}(\mu, \nu)$  denotes the integrated Fisher information on the Wasserstein geodesic between  $\mu$  and  $\nu$ , was proven for the Euclidean setting in [21, 22]. For more general cost functions, the expansion

$$\mathcal{C}_\varepsilon(\mu, \nu) = \mathcal{C}_0(\mu, \nu) - \frac{d}{2}\varepsilon \log \varepsilon + O(\varepsilon)$$

was established in [23, 24]. In [25], the same formula was demonstrated. Additionally, the authors were able to identify the separate asymptotic behavior of  $\mathbb{H}(\pi_\varepsilon | \mu \otimes \nu)$  and  $\mathcal{C}_\varepsilon(\mu, \nu) - \mathbb{H}(\pi_\varepsilon | \mu \otimes \nu)$  as  $\varepsilon$  vanishes.

Concerning the convergence of the optimal plans  $\pi_\varepsilon$ , in [25] the authors were able to quantify the 2-Wasserstein distance between  $\pi_\varepsilon$  and  $\pi_0$  when the Brenier map  $T_0$  is Lipschitz:

$$W_2(\pi_\varepsilon, \pi_0) = \Theta(\sqrt{\varepsilon}).$$

Up to now, the convergence of potentials in the continuous setting was only quantified in the  $L^2(\mu)$  norm in terms of the difference of their gradients: in [26], for the quadratic setting under compactness and convexity assumptions, it was proven that

$$\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{L^2(\mu)}^2 = O(\varepsilon^2).$$

The rate

$$\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{L^2(\mu)}^2 = O(-\varepsilon \log \varepsilon)$$

was found in [23] under slightly weaker assumptions. Finally, the rate

$$\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{L^2(\mu)}^2 = O(\varepsilon)$$

was established in [25] for non-necessarily compactly supported measures.

This note aims to exhibit a bound for the rate of convergence of uniform convergence on compact sets of the potentials  $\varphi_\varepsilon$  and their gradients  $\nabla\varphi_\varepsilon$  as  $\varepsilon \rightarrow 0$ , when both measures  $\mu$  and  $\nu$  are absolutely continuous with respect to the Lebesgue measure, under the quadratic cost. The following observation is the starting point of our analysis: the entropic potentials  $\varphi_\varepsilon$  are convex functions (see Prop. 2.1 below in Sect. 2). As mentioned above, they converge (up to a subsequence) uniformly on compact sets to  $\varphi_0$ . It is a classical fact (see, for example, [27], Thm. 25.7) that convexity yields the convergence of the gradients  $\nabla\varphi_\varepsilon$  to  $\nabla\varphi_0$  uniformly on compact sets as well. This justifies to search, for a fixed compact set  $K \subseteq \mathbb{R}^d$ , an asymptotic rate of convergence as  $\varepsilon$  goes to 0 for  $\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{K, \infty} := \sup_{x \in K} |\nabla\varphi_\varepsilon(x) - \nabla\varphi_0(x)|$ .

If both  $\mu$  and  $\nu$  are Gaussian, we can explicitly bound  $\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{K, \infty}$ . We present the result of this computation as an appetizer and defer its proof to the Section 4.

**Proposition 1.1.** *Let  $\mu = \mathcal{N}(0, A)$  and  $\nu = \mathcal{N}(0, B)$  be two non-degenerate Gaussian measures with  $A, B \in \mathcal{M}_d(\mathbb{R})$  symmetric and positive-definite. For  $R > 0$ , set  $K := \overline{\mathbb{B}}(0, R) \subseteq \mathbb{R}^d$  as the Euclidean closed ball of radius*

R. Then

$$\forall \varepsilon > 0, \quad \|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{K,\infty} \leq R\varepsilon \frac{|A^{-1}|_{\text{op}}}{2} \left( \varepsilon \frac{|(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{-1}|_{\text{op}}^{\frac{1}{2}}}{4} + \varepsilon^3 \frac{|(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{-1}|_{\text{op}}^{\frac{3}{2}}}{16} + 1 \right), \quad (1.5)$$

where  $|\cdot|_{\text{op}}$  denotes the operator norm of a matrix.

Proposition 1.1 reveals that we can quantify uniform convergence on compact sets as  $O(\varepsilon)$  as  $\varepsilon \rightarrow 0$ , with a dimension-free bound depending only on the size of the compact set  $K$  and the operator norms of the matrices  $A$  and  $B$ .

#### 1.4. Main results

Our first result gives an estimate in the spirit of Proposition 1.1 for absolutely continuous measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  satisfying the following convexity assumptions (see the end of Sect. 2 for a more involved discussion on those), in order to generalize what happens in the Gaussian case:

(A1) The measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  have the form  $d\mu(x) = e^{-V(x)} dx$  and  $d\nu(y) = e^{-W(y)} dy$ , where  $V, W: \mathbb{R}^d \rightarrow \mathbb{R}$  are smooth functions and there exist  $\alpha, \beta > 0$  such that

$$\forall x \in \mathbb{R}^d, \quad \nabla^2 V(x) \preceq \alpha I_d$$

and

$$\forall y \in \mathbb{R}^d, \quad \nabla^2 W(y) \succeq \beta I_d,$$

where  $I_d$  is the identity matrix of dimension  $d$  and  $\preceq$  denotes the Löwner order on the set of positive semidefinite matrices.

(A2) The measure  $\mu$  satisfies a Poincaré inequality: there exists  $C_P(\mu) > 0$  such that for any  $h: \mathbb{R}^d \rightarrow \mathbb{R}$  smooth with  $\int_{\mathbb{R}^d} h d\mu = 0$ ,

$$\|h\|_{L^2(\mu)}^2 \leq C_P(\mu) \|\nabla h\|_{L^2(\mu)}^2.$$

(A3) The measure  $\mu$  has finite differential entropy:

$$-\infty < H(\mu) := - \int_{\mathbb{R}^d} V(x) e^{-V(x)} dx < +\infty.$$

**Theorem 1.2.** *Let  $\mu$  and  $\nu$  be two probability measures on  $\mathbb{R}^d$  that are absolutely continuous with respect to the  $d$ -dimensional Lebesgue measure and satisfy the assumptions (A1), (A2), and (A3). Then, for any  $K \subseteq \mathbb{R}^d$  compact, there exists a computable constant  $C_{\text{grad}} = C_{\text{grad}}(d, K, \mu, \nu) > 0$  such that*

$$\forall \varepsilon > 0, \quad \|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{K,\infty} \leq C_{\text{grad}} \varepsilon^{\frac{1}{d+4}}.$$

We remark that Theorem 1.2 does not provide an optimal bound; recall Proposition 1.1.

Our next result is a corollary of the previous theorem: we can also quantify the convergence in compact sets of the entropic potentials. For  $K \subseteq \mathbb{R}^d$  compact we define, doing an abuse of notation,  $\|\varphi_\varepsilon - \varphi_0\|_{K,\infty} := \sup_{x \in K} |\varphi_\varepsilon(x) - \varphi_0(x)|$ .

**Theorem 1.3.** *Let  $\mu$  and  $\nu$  be two probability measures on  $\mathbb{R}^d$  that are absolutely continuous with respect to the  $d$ -dimensional Lebesgue measure and satisfy the assumptions (A1), (A2), and (A3). In addition, suppose that the following normalization holds: for every  $\varepsilon > 0$ ,*

$$\int_{\mathbb{R}^d} \varphi_\varepsilon \, d\mu = \int_{\mathbb{R}^d} \varphi_0 \, d\mu = 0. \quad (1.6)$$

*Then, for any  $K \subseteq \mathbb{R}^d$  compact and connected, there exists a computable constant  $C_{\text{pot}} = C_{\text{pot}}(d, K, \mu, \nu) > 0$  such that*

$$\forall \varepsilon > 0, \quad \|\varphi_\varepsilon - \varphi_0\|_{K, \infty} \leq C_{\text{pot}} \left( \varepsilon^{\frac{1}{d+4}} + \varepsilon \right).$$

To the best of our knowledge, this is the first work that addresses this problem for the entropic regularization in the continuous setting. Previously in the literature, this question has been answered both in the discrete [28] and semi-discrete [29, 30] settings for the entropic regularization. On the other hand, in [31], a rate was found for the quadratic-regularized optimal transport problem in dimension one.

## 1.5. Organization

In Section 2, we will present all the necessary preliminaries on optimal transport and its entropic regularization, and detail the assumptions for our main results, Theorems 1.2 and 1.3. Then, in Section 3, we will prove both results. Finally, in Section 4, we will prove Proposition 1.1.

## 1.6. Acknowledgments

I want to thank my advisor, Max Fathi, for his guidance, corrections, and continuous support with this project. I also thank Aymeric Baradat, Joaquín Fontbona, Maxime Laborde, Flavien Léger, and Hugo Malamut for very useful conversations. I would like to thank Michael Goldman for pointing out a mistake in a previous version of this note. I am also very grateful to the referees and the editor for their comments, which have significantly improved the quality of this work.

This work has received support under the program “Investissement d’Avenir” launched by the French Government and implemented by ANR, with the reference ANR-18-IdEx-0001 as part of its program “Emergence”. It was also supported by the Agence Nationale de la Recherche (ANR) Grant ANR-23-CE40-0003 (Project CONVIVIALITY). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945332.

## 2. PRELIMINARIES, NOTATION AND ASSUMPTIONS

In this section, we state the basic notation used in this note and then review some properties of both the Brenier and entropic potentials that complement the basic statements previously introduced. Finally, we state our main assumptions and discuss their consequences and some sufficient conditions for them to hold.

### 2.1. Basic notation

For  $x, y \in \mathbb{R}^d$ , we denote the Euclidean inner product between  $x$  and  $y$  as  $x \cdot y$  and the Euclidean norm of  $x$  as  $|x| := \sqrt{x \cdot x}$ . We write the set of square matrices of dimension  $d \times d$  with real entries as  $\mathcal{M}_d(\mathbb{R})$  and for  $M \in \mathcal{M}_d(\mathbb{R})$ , we define its Hilbert-Schmidt norm, doing an abuse of notation, as  $|M| := \sqrt{\sum_{i,j=1}^d |M_{ij}|^2}$ . For a twice-differentiable function  $u: \mathbb{R}^d \rightarrow \mathbb{R}$  and  $x \in \mathbb{R}^d$ , we denote its gradient as  $\nabla u(x) \in \mathbb{R}^d$  and its Hessian

matrix as  $\nabla^2 u(x) \in \mathcal{M}_d(\mathbb{R})$ . For a function  $u: \mathbb{R}^d \rightarrow \mathbb{R}$ , we define its convex conjugate as

$$\forall y \in \mathbb{R}^d, \quad u^*(y) := \sup_{x \in \mathbb{R}^d} (x \cdot y - \varphi(x)).$$

For a Polish space  $\mathcal{X}$  equipped with its Borel  $\sigma$ -algebra  $\mathcal{B}(\mathcal{X})$ , we denote by  $\mathcal{P}(\mathcal{X})$  the set of probability measures on  $\mathcal{X}$ ; typically  $\mathcal{X} \in \{\mathbb{R}^d, \mathbb{R}^d \times \mathbb{R}^d\}$ . For a non-negative measure  $\eta$  on  $\mathcal{X}$ , we denote its support by  $\text{supp}(\eta)$ , and for  $p \in [1, +\infty]$ , we will write  $L^p(\eta)$  for the classical  $p$ -Lebesgue space, which is a Banach space if we furnish it with the  $p$ -norm,  $\|\cdot\|_{L^p(\eta)}$ . If  $\mathcal{X} = \mathbb{R}^d$  and  $\eta$  is the  $d$ -dimensional Lebesgue measure, we write the associated  $p$ -norm as  $\|\cdot\|_p$ . If we have a sequence  $(u_n)_{n \in \mathbb{N}}$  of continuous functions  $u_n: \mathcal{X} \rightarrow \mathbb{R}$ , we say that it converges uniformly on compact sets to a continuous  $u: \mathcal{X} \rightarrow \mathbb{R}$  if for each  $K \subseteq \mathcal{X}$  compact,  $\|u_n - u\|_{K, \infty} := \sup_{x \in K} |u_n(x) - u(x)| \rightarrow 0$  as  $n \rightarrow \infty$ . If  $u: \mathcal{X} \rightarrow \mathbb{R}^d$  or  $u: \mathcal{X} \rightarrow \mathcal{M}_d(\mathbb{R})$ , we can extend the definitions of both  $\|\cdot\|_p$  and  $\|\cdot\|_{K, \infty}$  to them by doing the following abuses of notation:  $\|u\|_p := \|\|u\|\|_p$  and  $\|u\|_{K, \infty} := \|\|u\|\|_{K, \infty}$ , where  $|\cdot|$  denotes the Euclidean or Hilbert-Schmidt norm depending on the codomain of  $u$ .

Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two Polish spaces. If we fix two probability measures  $\mu \in \mathcal{P}(\mathcal{X})$  and  $\nu \in \mathcal{P}(\mathcal{Y})$ , we define the associated set of coupling measures  $\Pi(\mu, \nu)$ , also named the set of transport plans between  $\mu$  and  $\nu$ , as the set containing all the measures  $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  such that for every  $A \subseteq \mathcal{X}$  Borel,  $\pi(A \times \mathcal{Y}) = \mu(A)$  and for every  $B \subseteq \mathcal{Y}$  Borel,  $\pi(\mathcal{X} \times B) = \nu(B)$ . We immediately observe that  $\mu \otimes \nu$ , the product measure between  $\mu$  and  $\nu$ , is in  $\Pi(\mu, \nu)$ , so  $\Pi(\mu, \nu)$  is a nonempty set. Given a Borel map  $T: \mathcal{X} \rightarrow \mathcal{Y}$ , we say that it pushes forward the measure  $\mu$  towards  $\nu$  if for any  $B \subseteq \mathcal{Y}$  Borel,  $\nu(B) = \mu(T^{-1}(B))$ . We remark that the map  $T$  induces a natural coupling between  $\mu$  and  $\nu$ , which is given by the pushforward of the measure  $\mu$  by the map  $x \mapsto (x, T(x))$ .

## 2.2. Further properties of the potentials

We recall the definitions of the Brenier potentials  $(\varphi_0, \psi_0)$  and their entropic counterparts  $(\varphi_\varepsilon, \psi_\varepsilon)$ . Note that they are not unique, since for any constant  $\alpha \in \mathbb{R}$ , then  $(\varphi_0 + \alpha, \psi_0 - \alpha)$  and  $(\varphi_\varepsilon + \alpha, \psi_\varepsilon - \alpha)$  are two new pairs of Brenier and entropic potentials, respectively. Therefore, if we need to enforce the uniqueness of the potentials, an additional normalization condition has to be imposed, for example, (1.6). Nevertheless, we remark that their gradients are uniquely determined.

Recall that for  $\varepsilon > 0$  we have that  $(f_\varepsilon, g_\varepsilon) = (\frac{1}{2}|\cdot| - \varphi_\varepsilon, \frac{1}{2}|\cdot| - \psi_\varepsilon)$ . The pair  $(f_\varepsilon, g_\varepsilon)$  is intimately linked with  $\pi_\varepsilon$ , the optimal coupling for (1.3), since

$$\frac{d\pi_\varepsilon}{d(\mu \otimes \nu)}(x, y) = \exp\left(\frac{f_\varepsilon(x) + g_\varepsilon(y) - \frac{1}{2}|x - y|^2}{\varepsilon}\right) \quad \mu \otimes \nu\text{-a.s.} \quad (2.1)$$

A direct consequence of (2.1), is that the pair  $(f_\varepsilon, g_\varepsilon)$  satisfies the so-called Schrödinger system: for any  $(x, y) \in \text{supp}(\mu) \times \text{supp}(\nu)$ ,

$$f_\varepsilon(x) = -\varepsilon \log\left(\int_{\mathbb{R}^d} e^{\frac{1}{\varepsilon}[g_\varepsilon(z) - \frac{1}{2}|x - z|^2]} d\nu(z)\right), \quad (2.2)$$

$$g_\varepsilon(y) = -\varepsilon \log\left(\int_{\mathbb{R}^d} e^{\frac{1}{\varepsilon}[f_\varepsilon(z) - \frac{1}{2}|y - z|^2]} d\mu(z)\right). \quad (2.3)$$

The Schrödinger system allows us to extend  $f_\varepsilon$  and  $g_\varepsilon$ , which are defined on  $\text{supp}(\mu)$  and  $\text{supp}(\nu)$ , respectively, to all of  $\mathbb{R}^d$  since the right-hand sides of equations (2.2) and (2.3) are well defined on all of  $\mathbb{R}^d$ . We also note that the regularity of the quadratic cost ensures that the potentials are continuous, hence real valued [15], Lemma 3.1. Moreover, the Schrödinger system can be exploited to derive finer properties of the pair  $(\varphi_\varepsilon, \psi_\varepsilon)$  such as convexity.

**Proposition 2.1.** *Let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ . Then, for any  $\varepsilon > 0$ , the associated potentials  $\varphi_\varepsilon$  and  $\psi_\varepsilon$  are convex.*

*Proof.* Let  $\lambda \in (0, 1)$  and  $x_1, x_2 \in \mathbb{R}^d$ . Then by (2.2) we get

$$\begin{aligned} \lambda\varphi_\varepsilon(x_1) + (1 - \lambda)\varphi_\varepsilon(x_2) &= \varepsilon \log \left( \left[ \int_{\mathbb{R}^d} e^{\frac{1}{\varepsilon}[x_1 \cdot y - \psi_\varepsilon(y)]} d\nu(y) \right]^\lambda \left[ \int_{\mathbb{R}^d} e^{\frac{1}{\varepsilon}[x_2 \cdot y - \psi_\varepsilon(y)]} d\nu(y) \right]^{1-\lambda} \right) \\ &\geq \varepsilon \log \left( \int_{\mathbb{R}^d} e^{\frac{1}{\varepsilon}[(\lambda x_1 + (1-\lambda)x_2) \cdot y - \psi_\varepsilon(y)]} d\nu(y) \right) \\ &= \varphi_\varepsilon(\lambda x_1 + (1 - \lambda)x_2), \end{aligned}$$

where we used Hölder's inequality for  $p = 1/\lambda$  and  $q = 1/(1 - \lambda)$ . The convexity of  $\psi_\varepsilon$  follows similarly from (2.3).  $\square$

### 2.3. Assumptions

Our main results are stated in terms of two probability measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  that are absolutely continuous with respect to the  $d$ -dimensional Lebesgue measure and satisfy the following assumptions, which were already stated in Section 1.4, but we repeat for convenience of the reader:

(A1) The measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  have the form  $d\mu(x) = e^{-V(x)} dx$  and  $d\nu(y) = e^{-W(y)} dy$ , where  $V, W: \mathbb{R}^d \rightarrow \mathbb{R}$  are smooth functions and there exist  $\alpha, \beta > 0$  such that

$$\forall x \in \mathbb{R}^d, \quad \nabla^2 V(x) \preceq \alpha I_d \tag{2.4}$$

and

$$\forall y \in \mathbb{R}^d, \quad \nabla^2 W(y) \succeq \beta I_d, \tag{2.5}$$

where  $I_d$  is the identity matrix of dimension  $d$  and  $\preceq$  denotes the Löwner order on the set of positive semidefinite matrices.

(A2) The measure  $\mu$  satisfies a Poincaré inequality: there exists  $C_P(\mu) > 0$  such that for any  $h: \mathbb{R}^d \rightarrow \mathbb{R}$  smooth with  $\int_{\mathbb{R}^d} h d\mu = 0$ ,

$$\|h\|_{L^2(\mu)}^2 \leq C_P(\mu) \|\nabla h\|_{L^2(\mu)}^2.$$

(A3) The measure  $\mu$  has finite differential entropy:

$$-\infty < \mathbf{H}(\mu) := - \int_{\mathbb{R}^d} V(x) e^{-V(x)} dx < +\infty.$$

The main consequence of (A1) is that we obtain quantitative control on both  $\nabla^2 \varphi_0$  and  $\nabla^2 \varphi_\varepsilon$  for every  $\varepsilon > 0$ ; that is, both  $\nabla \varphi_0$  and  $\nabla \varphi_\varepsilon$  are Lipschitz and we have explicit control on the value of their Lipschitz constants. The following generalization of Caffarelli's contraction theorem [32] provides global Lipschitz regularity for the Brenier map  $\nabla \varphi_0$ , which pushes  $\mu$  towards  $\nu$ . The statement that we use here can be found on [33], Theorem 1, for example

**Theorem 2.2.** *Suppose that both  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  satisfy (A1). Then the Brenier map  $\nabla\varphi_0$  is globally Lipschitz. Moreover, the following estimate holds:*

$$\forall x \in \mathbb{R}^d, \quad 0 \preceq \nabla^2\varphi_0(x) \preceq \sqrt{\frac{\alpha}{\beta}} I_d.$$

For the entropic counterpart of the Brenier map, namely  $\nabla\varphi_\varepsilon$ , we can also exhibit bounds on its derivative. These bounds were used in [33, 34] to give alternative proofs of Theorem 2.2 based on the entropic regularization. Here we use the ones proven in [33], Theorem 4.

**Theorem 2.3.** *Suppose that both  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  satisfy (A1). Then the entropic Brenier map  $\nabla\varphi_\varepsilon$  is globally Lipschitz. Moreover, the following estimate holds:*

$$\forall x \in \mathbb{R}^d, \quad 0 \preceq \nabla^2\varphi_\varepsilon(x) \preceq \frac{1}{2} \left( \sqrt{\frac{4\alpha}{\beta} + \varepsilon^2\alpha^2} - \varepsilon\alpha \right) I_d.$$

As we will see in the following remark, the functional inequality provided by (A2) also entails regularity properties for the measure  $\mu$  itself.

**Remark 2.4.** A Poincaré inequality implies that the measure has finite moments of all orders [35], Proposition 4.4.2, so under (A2), the measure  $\mu$  will have this property.

**Remark 2.5.** Log-concavity is a sufficient condition that entails both (A2) and (A3). We say that the measure  $d\mu(x) = e^{-V(x)} dx$  is log-concave if  $V$  is convex. Since we assumed  $V$  to be smooth, this is equivalent to

$$\forall x \in \mathbb{R}^d, \quad \nabla^2 V(x) \succeq 0.$$

First, log-concavity yields the validity of a Poincaré inequality (see, for example, [35], Thm. 4.6.2). Second, let us see that log-concavity implies (A3): indeed,

$$H(\mu) = - \int_{\mathbb{R}^d} V(x) e^{-V(x)} dx = - \int_{\mathbb{R}^d} V(x) d\mu(x) \leq -V(0) - \nabla V(0) \cdot \mathbb{E}_{X \sim \mu}[X] < +\infty,$$

where we used the convexity of  $V$ . For the lower bound, we recall that  $\mu$  has finite moments of all orders (see Rem. 2.4), so

$$-\infty < H(\mathcal{N}) \leq H(\mu),$$

where  $\mathcal{N}$  is the  $d$ -dimensional Gaussian measure with the same mean and covariance as  $\mu$ .

**Remark 2.6.** Note that under (A1), the measure  $\nu$  is log-concave (see (2.5)); in particular, it has finite differential entropy and finite moments of all orders as well.

Under the assumption (A1), we saw that Theorem 2.2 ensures the Lipschitz regularity of the Brenier map quantitatively. A direct consequence of this fact is the following: we can control the difference of the gradients of the potentials in the  $L^2(\mu)$  norm by the difference between both costs.

**Proposition 2.7.** *Suppose that the Brenier map is  $L$ -Lipschitz. Then*

$$\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{L^2(\mu)}^2 \leq 2L \langle \pi_\varepsilon - \pi_0, \frac{1}{2} |\cdot - \cdot|^2 \rangle = 2L (\mathcal{C}_\varepsilon(\mu, \nu) - \varepsilon H(\pi_\varepsilon | \mu \otimes \nu) - \mathcal{C}_0(\mu, \nu)).$$

*Proof.* First, let us remark that a direct computation gives

$$\forall x \in \mathbb{R}^d, \quad \nabla \varphi_\varepsilon(x) = \int_{\mathbb{R}^d} y \, d\pi_\varepsilon^x(y), \quad (2.6)$$

where for  $x \in \mathbb{R}^d$ , we have defined  $d\pi_\varepsilon^x(y) := \exp\left(\frac{f_\varepsilon(x) + g_\varepsilon(y) - \frac{1}{2}|x-y|^2}{\varepsilon}\right) d\nu(y)$ . In other words,  $(\pi_\varepsilon^x)_{x \in \mathbb{R}^d}$  is a disintegration of  $\pi_\varepsilon$  with respect to  $\mu$ . Then, by Jensen's inequality, we see that

$$\begin{aligned} \|\nabla \varphi_\varepsilon - \nabla \varphi_0\|_{L^2(\mu)}^2 &= \int_{\mathbb{R}^d} |\nabla \varphi_\varepsilon(x) - \nabla \varphi_0(x)|^2 \, d\mu(x) = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (y - \nabla \varphi_0(x)) \, d\pi_\varepsilon^x(y) \right|^2 \, d\mu(x) \\ &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |y - \nabla \varphi_0(x)|^2 \, d\pi_\varepsilon^x(y) \, d\mu(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - \nabla \varphi_0(x)|^2 \, d\pi_\varepsilon(x, y) \\ &= \|y - T_0(x)\|_{L^2(\pi_\varepsilon)}^2. \end{aligned}$$

Finally, we conclude by using the bound

$$\|y - T_0(x)\|_{L^2(\pi_\varepsilon)}^2 \leq 2L \langle \pi_\varepsilon - \pi_0, \frac{1}{2} |\cdot - \cdot|^2 \rangle = 2L (C_\varepsilon(\mu, \nu) - \varepsilon H(\pi_\varepsilon | \mu \otimes \nu) - C_0(\mu, \nu)),$$

which was proven in both [25], Lemma 3.8 and [23], Proposition 4.5 when the map  $T_0$  is  $L$ -Lipschitz. Let us remark that this result, in turn, is based on an unpublished result of Ambrosio reported in [36].  $\square$

That is, if we are able to control  $\langle \pi_\varepsilon - \pi_0, \frac{1}{2} |\cdot - \cdot|^2 \rangle$ , then we can control the  $L^2$  norm of the difference. In [25], Theorem 3.7, we can find this control, which can be applied as a consequence of both (A2) and (A3), recall Remarks 2.4 and 2.6.

**Theorem 2.8.** ([25], Thm. 3.7). *Suppose that both  $\mu$  and  $\nu$  have finite moments of order  $2 + \delta$ , for some  $\delta > 0$ , and that both have finite differential entropy. Then there exists  $C = C(d, \mu, \nu) > 0$  depending on the dimension  $d$  and on the moments of order  $2 + \delta$  and the differential entropies of both  $\mu$  and  $\nu$  such that for every  $\varepsilon > 0$ ,*

$$\langle \pi_\varepsilon - \pi_0, \frac{1}{2} |\cdot - \cdot|^2 \rangle \leq C\varepsilon.$$

That is, under our three assumptions, there exists a constant  $C_1 = C_1(d, \mu, \nu) > 0$  depending on  $d, \mu$  and  $\nu$  such that

$$\|\nabla \varphi_\varepsilon - \nabla \varphi_0\|_{L^2(\mu)}^2 \leq C_1 \varepsilon. \quad (2.7)$$

Now let us assume that the following normalization holds:

$$\forall \varepsilon > 0, \quad \int_{\mathbb{R}^d} \varphi_\varepsilon \, d\mu = \int_{\mathbb{R}^d} \varphi_0 \, d\mu = 0. \quad (2.8)$$

Note that if we are working with the gradients of the potentials, they are uniquely determined, so we may assume, without loss of generality, that (2.8) holds. In particular, under this normalization, we can use (A2) to control the  $L^2$  norm of the difference of the potentials:

$$\|\varphi_\varepsilon - \varphi_0\|_{L^2(\mu)}^2 \leq C_P(\mu) \|\nabla \varphi_\varepsilon - \nabla \varphi_0\|_{L^2(\mu)}^2. \quad (2.9)$$

That is, again under all our assumptions, we have that there exists a constant  $C_2 = C_2(d, \mu, \nu) > 0$  depending on  $d, \mu$  and  $\nu$  such that

$$\|\varphi_\varepsilon - \varphi_0\|_{L^2(\mu)}^2 \leq C_2 \varepsilon. \quad (2.10)$$

### 3. PROOF OF THE MAIN RESULTS

This section aims to prove Theorem 1.2 and then to obtain as a corollary Theorem 1.3. The starting point of our proof will be the Gagliardo-Nirenberg inequality, which allows us to control the  $p$ -norm of the derivatives of order  $j$  of a function by the  $r$ -norm of its derivatives of order  $k$  and the  $q$ -norm of the function itself, where the parameters  $i, j, p, q$ , and  $r$  satisfy some relations. We state the inequality in the following version, found in [37], Theorem 3.1, since it allows the critical values  $p = +\infty$  and  $r = +\infty$ . Here the ambient space is  $\mathbb{R}^d$  and we recall the notation  $\|\cdot\|_s = \|\cdot\|_{L^s(\text{Leb})}$ .

**Theorem 3.1** (Gagliardo-Nirenberg). *Let  $j, k \in \mathbb{N}$  such that  $1 \leq j < k$ . Let  $\theta \in [j/k, 1]$ ,  $q \in [1, +\infty]$ ,  $p \in (1, +\infty]$ , and  $r \in [1, +\infty]$  such that*

$$\frac{1}{p} = \frac{j}{d} + \theta \left( \frac{1}{r} - \frac{k}{d} \right) + \frac{1-\theta}{q} \quad (3.1)$$

and

$$\forall 0 \leq i \leq k - j - 1, \quad r^{(i)} \neq d, \quad (3.2)$$

where  $r^{(0)} := r$  and  $r^{(i)} := \frac{dr^{(i-1)}}{d-r^{(i-1)}}$  for  $i \geq 1$ . Then there exists a constant  $C_{\text{GN}} = C_{\text{GN}}(j, k, d, \theta, p, q, r) > 0$  such that for any  $u: \mathbb{R}^d \rightarrow \mathbb{R}$  sufficiently smooth and integrable,

$$\|\nabla^j u\|_p \leq C_{\text{GN}} \|\nabla^k u\|_r^\theta \|u\|_q^{1-\theta}. \quad (3.3)$$

We are ready to start the proof of Theorem 1.2.

*Proof (of Thm. 1.2):* The proof is divided into three consecutive steps.

**Step 1:** The first step will be to choose appropriate parameters to apply the Gagliardo-Nirenberg inequality (3.3). From Theorem 3.1 and choosing

$$j = 1, k = 2, p = +\infty, q = 2, r = +\infty,$$

the value of  $\theta$  is determined by (3.1), so we get

$$\theta = \frac{d+2}{d+4} \in [1/2, 1].$$

On the other hand, we notice that the condition (3.2) is trivially verified since  $r^{(0)} = +\infty$ . Hence, there exists a constant  $C_{\text{GN}} = C_{\text{GN}}(d) > 0$  depending on the dimension  $d$  such that for any  $u: \mathbb{R}^d \rightarrow \mathbb{R}$  sufficiently smooth and integrable,

$$\|\nabla u\|_\infty \leq C_{\text{GN}} \|\nabla^2 u\|_\infty^{\frac{d+2}{d+4}} \|u\|_2^{\frac{2}{d+4}}. \quad (3.4)$$

Now let  $K \subseteq \mathbb{R}^d$  be a compact set, take  $R > 0$  such that  $K \subseteq B(0, R)$ , where  $B(0, R)$  denotes the open ball with center 0 and radius  $R$ , and such that  $\mu(B(0, R)) > 4/5$ , and let  $w: \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a compactly supported smooth function satisfying the following properties:

- $\forall x \in \mathbb{R}^d, 0 \leq w(x) \leq 1$ ;
- $\forall x \in K, w(x) = 1$ ; and
- $\text{supp}(w) \subseteq B(0, R)$ .

Let  $\varepsilon > 0$ . Then

$$\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{K, \infty} \leq \|\nabla[(\varphi_\varepsilon - \varphi_0)w]\|_\infty.$$

Since the potentials  $\varphi_\varepsilon$  and  $\varphi_0$  are smooth enough, we may apply (3.4) to  $u := (\varphi_\varepsilon - \varphi_0)w$ , so that

$$\|\nabla[(\varphi_\varepsilon - \varphi_0)w]\|_\infty \leq C_{\text{GN}} \|\nabla^2[(\varphi_\varepsilon - \varphi_0)w]\|_\infty^{\frac{d+2}{d+4}} \|(\varphi_\varepsilon - \varphi_0)w\|_2^{\frac{2}{d+4}}. \quad (3.5)$$

**Step 2:** Now we will bound the first term on the right-hand side of (3.5):

$$\|\nabla^2[(\varphi_\varepsilon - \varphi_0)w]\|_\infty^{\frac{d+2}{d+4}}. \quad (3.6)$$

Note that

$$\begin{aligned} \|\nabla^2[(\varphi_\varepsilon - \varphi_0)w]\|_\infty &\leq \|w\nabla^2(\varphi_\varepsilon - \varphi_0)\|_\infty + \|(\varphi_\varepsilon - \varphi_0)\nabla^2w\|_\infty + 2\|\nabla(\varphi_\varepsilon - \varphi_0)\nabla w^\top\|_\infty \\ &\leq \|\nabla^2(\varphi_\varepsilon - \varphi_0)\|_\infty + \|\varphi_\varepsilon - \varphi_0\|_{B(0, R), \infty} \times \|\nabla^2w\|_\infty \\ &\quad + 2\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{B(0, R), \infty} \times \|\nabla w\|_\infty. \end{aligned}$$

For the first term on the right-hand side, (A1) plays a key role: note that for every  $\varepsilon > 0$ , the upper bound in Theorem 2.3 is such that

$$\frac{1}{2} \left( \sqrt{\frac{4\alpha}{\beta} + \varepsilon^2\alpha^2} - \varepsilon\alpha \right) \leq \sqrt{\frac{\alpha}{\beta}}.$$

In particular, using also the estimate granted by Theorem 2.2, we get

$$\|\nabla^2\varphi_\varepsilon - \nabla^2\varphi_0\|_\infty \leq \|\nabla^2\varphi_\varepsilon\|_\infty + \|\nabla^2\varphi_0\|_\infty \leq 2\sqrt{\frac{\alpha}{\beta}}.$$

In order to bound the two remaining terms on the right-hand side, let us state and prove the following auxiliary lemma.

**Lemma 3.2.** *In the above context, there exists  $M = M(R, \mu, \nu) > 0$  depending on  $R$ ,  $\mu$  and  $\nu$  such that for any  $\varepsilon > 0$ , there exists  $x_\varepsilon^* \in B(0, R)$  satisfying both*

$$|\nabla\varphi_\varepsilon(x_\varepsilon^*) - \nabla\varphi_0(x_\varepsilon^*)| < M \quad (3.7)$$

and

$$|\varphi_\varepsilon(x_\varepsilon^*) - \varphi_0(x_\varepsilon^*)| < M. \quad (3.8)$$

*Proof.* To start, let us prove that

$$\sup_{\varepsilon > 0} \int_{\mathbb{R}^d} |\nabla \varphi_\varepsilon - \nabla \varphi_0|^2 d\mu < +\infty,$$

so let  $\varepsilon > 0$ . First, recall equation (2.6):

$$\nabla \varphi_\varepsilon(x) = \int_{\mathbb{R}^d} y d\pi_\varepsilon^x(y).$$

By Jensen's inequality and the fact that  $\pi_\varepsilon \in \Pi(\mu, \nu)$ , we have that

$$\int_{\mathbb{R}^d} |\nabla \varphi_\varepsilon(x)|^2 d\mu(x) \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |y|^2 d\pi_\varepsilon^x(y) d\mu(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2 d\pi_\varepsilon(x, y) = \int_{\mathbb{R}^d} |y|^2 d\nu(y).$$

On the other hand, note that

$$\int_{\mathbb{R}^d} |\nabla \varphi_0(x)|^2 d\mu(x) = \int_{\mathbb{R}^d} |y|^2 d\nu(y)$$

since the Brenier map  $x \mapsto \nabla \varphi_0(x)$  pushes forward  $\mu$  towards  $\nu$ . Then

$$\int_{\mathbb{R}^d} |\nabla \varphi_\varepsilon - \nabla \varphi_0|^2 d\mu \leq 2 \int_{\mathbb{R}^d} |\nabla \varphi_\varepsilon|^2 d\mu + 2 \int_{\mathbb{R}^d} |\nabla \varphi_0|^2 d\mu \leq 4 \int_{\mathbb{R}^d} |y|^2 d\nu(y),$$

so  $\sup_{\varepsilon > 0} \int_{\mathbb{R}^d} |\nabla \varphi_\varepsilon - \nabla \varphi_0|^2 d\mu < +\infty$  since  $\nu$  has finite moments of all orders; recall Remarks 2.4 and 2.6.

Let  $M_1 > 0$  to be fixed. By Chebyshev's inequality, we have

$$\mu(\{x \in \mathbb{R}^d : |\nabla \varphi_\varepsilon(x) - \nabla \varphi_0(x)| \geq M_1\}) \leq \frac{1}{M_1^2} \int_{\mathbb{R}^d} |\nabla \varphi_\varepsilon - \nabla \varphi_0|^2 d\mu \leq \frac{4}{M_1^2} \int_{\mathbb{R}^d} |y|^2 d\nu(y),$$

so we may fix  $M_1$  big enough such that  $\frac{4}{M_1^2} \int_{\mathbb{R}^d} |y|^2 d\nu(y) \leq \frac{1}{5}$ . Hence,

$$\mu(\{x \in \mathbb{R}^d : |\nabla \varphi_\varepsilon(x) - \nabla \varphi_0(x)| < M_1\}) > \frac{4}{5}, \quad (3.9)$$

so

$$\mu(\{x \in B(0, R) : |\nabla \varphi_\varepsilon(x) - \nabla \varphi_0(x)| < M_1\}) > \frac{3}{5}. \quad (3.10)$$

Without loss of generality, we may assume that the normalization (2.8) holds so that (2.9) is verified. Similarly to the previous argument, we deduce that there exists  $M_2 > 0$  such that

$$\mu(\{x \in B(0, R) : |\varphi_\varepsilon(x) - \varphi_0(x)| < M_2\}) > \frac{3}{5}. \quad (3.11)$$

Let  $M := \max\{M_1, M_2\}$ , which depends on  $R$ ,  $\mu$ , and  $\nu$ . If we put together (3.10) and (3.11), we deduce that

$$\mu(\{x \in B(0, R) : |\varphi_\varepsilon(x) - \varphi_0(x)| < M, |\nabla \varphi_\varepsilon(x) - \nabla \varphi_0(x)| < M\}) > 0,$$

which yields the existence of  $x_\varepsilon^* \in B(0, R)$  such that both (3.7) and (3.8) hold.  $\square$

Recall that we wanted to bound

$$\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{\mathbb{B}(0,R),\infty} \quad (3.12)$$

and

$$\|\varphi_\varepsilon - \varphi_0\|_{\mathbb{B}(0,R),\infty}. \quad (3.13)$$

By Lemma 3.2, there exists  $M = M(R, \mu, \nu) > 0$  and  $x_\varepsilon^* \in \mathbb{B}(0, R)$  verifying both (3.7) and (3.8). For (3.12), note that

$$\begin{aligned} \sup_{x \in \mathbb{B}(0,R)} |\nabla\varphi_\varepsilon(x) - \nabla\varphi_0(x)| &\leq \sup_{x \in \mathbb{B}(0,R)} |\nabla\varphi_\varepsilon(x) - \nabla\varphi_\varepsilon(x_\varepsilon^*)| + \sup_{x \in \mathbb{B}(0,R)} |\nabla\varphi_0(x) - \nabla\varphi_0(x_\varepsilon^*)| \\ &\quad + |\nabla\varphi_\varepsilon(x_\varepsilon^*) - \nabla\varphi_0(x_\varepsilon^*)| \\ &\leq 4\sqrt{\frac{\alpha}{\beta}}R + M, \end{aligned}$$

where we used the second-order estimates holding under (A1) and (3.7). For the remaining term (3.13), observe that a Taylor expansion combined with the aforementioned second-order estimates yields

$$\begin{aligned} \sup_{x \in \mathbb{B}(0,R)} |\varphi_\varepsilon(x) - \varphi_0(x)| &\leq |\varphi_\varepsilon(x_\varepsilon^*) - \varphi_0(x_\varepsilon^*)| + 2R|\nabla\varphi_\varepsilon(x_\varepsilon^*) - \nabla\varphi_0(x_\varepsilon^*)| + 4R^2\sqrt{\frac{\alpha}{\beta}} \\ &\leq M + 2RM + 4R^2\sqrt{\frac{\alpha}{\beta}}. \end{aligned}$$

In conclusion, if we combine the three bounds, we know that there exists a constant  $C = C(K, \mu, \nu) > 0$  depending on  $K$  (since the constant depends on  $R$ , which in turn depends on  $K$ ),  $\mu$ , and  $\nu$  such that for every  $\varepsilon > 0$ ,

$$\|\nabla^2[(\varphi_\varepsilon - \varphi_0)w]\|_{\infty}^{\frac{d+2}{d+4}} \leq C. \quad (3.14)$$

**Step 3:** Now we aim to bound the remaining term in the right-hand side of (3.5) in an asymptotic way in order to obtain a quantity that converges to 0 as  $\varepsilon$  vanishes; that is, we want to bound

$$\|(\varphi_\varepsilon - \varphi_0)w\|_2^{\frac{2}{d+4}}.$$

To control this term, the bound (2.10) will be crucial. Indeed, since  $\text{supp}(w) \subseteq \overline{\mathbb{B}(0, R)}$ , then we have that

$$\begin{aligned} \|(\varphi_\varepsilon - \varphi_0)w\|_2^2 &= \int_{\overline{\mathbb{B}(0,R)}} |\varphi_\varepsilon(x) - \varphi_0(x)|^2 w(x) \, dx = \int_{\overline{\mathbb{B}(0,R)}} |\varphi_\varepsilon(x) - \varphi_0(x)|^2 e^{V(x)} w(x) e^{-V(x)} \, dx \\ &\leq \int_{\overline{\mathbb{B}(0,R)}} |\varphi_\varepsilon(x) - \varphi_0(x)|^2 e^{V(x)} \, d\mu(x) \\ &\leq \|e^V\|_{\overline{\mathbb{B}(0,R),\infty}}^2 \|\varphi_\varepsilon - \varphi_0\|_{L^2(\mu)}^2, \end{aligned}$$

so if we combine this with (2.10), we get

$$\|(\varphi_\varepsilon - \varphi_0)w\|_2^{\frac{2}{d+4}} \leq C' \varepsilon^{\frac{1}{d+4}}, \quad (3.15)$$

where  $C' = C(d, K, \mu, \nu) > 0$  is a constant depending on  $d, K$  (since it depends on  $R$  and the cutoff function  $w$ ),  $\mu$  and  $\nu$ . Finally, if we put together (3.5), (3.14), and (3.15), we deduce the existence of a constant  $C_{\text{grad}} = C_{\text{grad}}(d, K, \mu, \nu) > 0$  depending on the dimension  $d$ , the compact set  $K$ , and the measures  $\mu$  and  $\nu$  such that, uniformly on  $\varepsilon > 0$ ,

$$\|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{K,\infty} \leq C_{\text{grad}} \varepsilon^{\frac{1}{d+4}}.$$

□

We end by proving Theorem 1.3.

*Proof (of Thm. 1.3):* Let  $K \subseteq \mathbb{R}^d$  be a compact and connected set. Since we have assumed that the potentials are normalized as  $\int_{\mathbb{R}^d} \varphi_\varepsilon d\mu = \int_{\mathbb{R}^d} \varphi_0 d\mu = 0$ , we can use [38], Theorem 1, p. 290 and the bound (3.15) to get

$$\|\varphi_\varepsilon - \varphi_0\|_{K,\infty} \leq C'' \left( \|\nabla\varphi_\varepsilon - \nabla\varphi_0\|_{K,\infty} + \varepsilon \right),$$

where  $C'' = C''(d, K, \mu, \nu) > 0$  is a constant depending on  $d, K, \mu$  and  $\nu$ . The conclusion follows from the estimate given by Theorem 1.2. □

#### 4. THE GAUSSIAN CASE

In this section, we provide the proof of Proposition 1.1.

*Proof (of Prop. 1.1):* For Gaussian measures, we may compute explicitly  $\nabla\varphi_0$  and  $\nabla\varphi_\varepsilon$  (see [39–42]): both are linear maps given by

$$\forall x \in \mathbb{R}^d, \quad \nabla\varphi_0(x) = A^{-\frac{1}{2}}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}A^{-\frac{1}{2}}x$$

and

$$\forall x \in \mathbb{R}^d, \quad \nabla\varphi_\varepsilon(x) = \left( A^{-\frac{1}{2}} \left( A^{\frac{1}{2}}BA^{\frac{1}{2}} + \frac{\varepsilon^2}{4}I_d \right)^{\frac{1}{2}} A^{-\frac{1}{2}} - \frac{\varepsilon}{2}A^{-1} \right) x.$$

Now fix  $R > 0$  and let  $K = \overline{B}(0, R) \subseteq \mathbb{R}^d$ , so that

$$\sup_{x \in K} |\nabla\varphi_\varepsilon(x) - \nabla\varphi_0(x)| = R \left| \left( A^{-\frac{1}{2}} \left( A^{\frac{1}{2}}BA^{\frac{1}{2}} + \frac{\varepsilon^2}{4}I_d \right)^{\frac{1}{2}} A^{-\frac{1}{2}} - \frac{\varepsilon}{2}A^{-1} \right) - A^{-\frac{1}{2}}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}A^{-\frac{1}{2}} \right|_{\text{op}},$$

where for a matrix  $M \in \mathcal{M}_d(\mathbb{R})$  we write  $|M|_{\text{op}} := \sup_{|x|=1} |Mx|$  for its operator norm.

Let us note that we may expand the matrix  $\left( A^{\frac{1}{2}}BA^{\frac{1}{2}} + \frac{\varepsilon^2}{4}I_d \right)^{\frac{1}{2}}$ , using a Taylor expansion of order one for the matrix square root function around the point  $A^{\frac{1}{2}}BA^{\frac{1}{2}}$  (see [43], Thm. 1.1):

$$\left( A^{\frac{1}{2}}BA^{\frac{1}{2}} + \frac{\varepsilon^2}{4}I_d \right)^{\frac{1}{2}} = (A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}} + \frac{\varepsilon^2}{4} \int_0^{+\infty} e^{-2t(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}} dt + R(A, B, \varepsilon),$$

where  $R(A, B, \varepsilon)$  is a matrix such that

$$|R(A, B, \varepsilon)|_{\text{op}} \leq \frac{\varepsilon^4}{32} \lambda_{\min}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{-\frac{3}{2}} = \frac{\varepsilon^4}{32} \left| (A^{\frac{1}{2}}BA^{\frac{1}{2}})^{-1} \right|_{\text{op}}^{\frac{3}{2}},$$

where  $\lambda_{\min}(\cdot)$  denotes the smallest eigenvalue of a positive-definite matrix. On the other hand, note that

$$\left| \int_0^{+\infty} e^{-2t(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}} dt \right|_{\text{op}} \leq \int_0^{+\infty} e^{-2t\lambda_{\min}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}} dt = \frac{1}{2\lambda_{\min}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}} = \frac{1}{2} \left| (A^{\frac{1}{2}}BA^{\frac{1}{2}})^{-1} \right|_{\text{op}}^{\frac{1}{2}}.$$

In consequence, we have that

$$\begin{aligned} \sup_{x \in K} |\nabla \varphi_\varepsilon(x) - \nabla \varphi_0(x)| &= R \left| \left( A^{-\frac{1}{2}} \left( A^{\frac{1}{2}}BA^{\frac{1}{2}} + \frac{\varepsilon^2}{4}I_d \right)^{\frac{1}{2}} A^{-\frac{1}{2}} - \frac{\varepsilon}{2}A^{-1} \right) - A^{-\frac{1}{2}}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}A^{-\frac{1}{2}} \right|_{\text{op}} \\ &\leq R \left| A^{-\frac{1}{2}} \left( A^{\frac{1}{2}}BA^{\frac{1}{2}} + \frac{\varepsilon^2}{4}I_d \right)^{\frac{1}{2}} A^{-\frac{1}{2}} - A^{-\frac{1}{2}}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}A^{-\frac{1}{2}} \right|_{\text{op}} + R \left| \frac{\varepsilon}{2}A^{-1} \right|_{\text{op}} \\ &\leq \varepsilon R \frac{|A^{-1}|_{\text{op}}}{2} \left( \varepsilon \frac{|(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{-1}|_{\text{op}}^{\frac{1}{2}}}{4} + \varepsilon^3 \frac{|(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{-1}|_{\text{op}}^{\frac{3}{2}}}{16} + 1 \right), \end{aligned}$$

so  $\sup_{x \in K} |\nabla \varphi_\varepsilon(x) - \nabla \varphi_0(x)| = O(\varepsilon)$  as  $\varepsilon \rightarrow 0$ , with a constant which depends only on  $R$  and the matrices  $A$  and  $B$ . □

#### DATA AVAILABILITY STATEMENT

No datasets were generated or analyzed during this study.

#### REFERENCES

- [1] C. Villani, *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI (2003).
- [2] C. Villani, *Optimal transport*, vol. 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin (2009).
- [3] F. Santambrogio, *Optimal transport for applied mathematicians*, vol. 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham (2015).
- [4] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **44** (1991) 375–417.
- [5] R.J. McCann, Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** (1995) 309–323.
- [6] C. Léonard, A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** (2014) 1533–1574.
- [7] M. Nutz, Introduction to entropic optimal transport in Lecture notes, Columbia University (2021).
- [8] A. Figalli, On the Monge-Ampère equation. *Astérisque* **1148** (2019) 477–503.
- [9] M. Cuturi, Sinkhorn distances: lightspeed computation of optimal transport, *Advances in Neural Information Processing Systems*, vol. 26, edited by C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Weinberger. Curran Associates, Inc. (2013).
- [10] J. Altschuler, J. Niles-Weed and P. Rigollet, Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. *Adv. Neural Inform. Process. Syst.* **30** (2017) 1961–1971.
- [11] G. Peyré and M. Cuturi, Computational optimal transport. *Found. Trends Mach. Learn.* **11** (2019) 355–607.
- [12] C. Léonard, From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262** (2012) 1879–1920.
- [13] G. Carlier, V. Duval, G. Peyré and B. Schmitzer, Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.* **49** (2017) 1385–1418.

- [14] N. Gigli and L. Tamanini, Second order differentiation formula on  $\text{RCD}^*(K, N)$  spaces. *J. Eur. Math. Soc.* **23** (2021) 1727–1795.
- [15] M. Nutz and J. Wiesel, Entropic optimal transport: convergence of potentials. *Probab. Theory Related Fields* **184** (2022) 401–424.
- [16] A. Chiarini, G. Conforti, G. Greco and L. Tamanini, Gradient estimates for the Schrödinger potentials: convergence to the Brenier map and quantitative stability. *Commun. Part. Differ. Equ.* **48** (2023) 895–943.
- [17] S. Adams, N. Dirr, M.A. Peletier and J. Zimmer, From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. *Commun. Math. Phys.* **307** (2011) 791–815.
- [18] M.H. Duong, V. Laschos and M. Renger, Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM Control Optim. Calc. Var.* **19** (2013) 1166–1188.
- [19] M. Erbar, J. Maas and D.R.M. Renger, From large deviations to Wasserstein gradient flows in multiple dimensions. *Electron. Commun. Probab.* **20** (2015) 12.
- [20] S. Pal, On the difference between entropic cost and the optimal transport cost. *Ann. Appl. Probab.* **34** (2024) 1003–1028.
- [21] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard and G. Peyré, Faster Wasserstein distance estimation with the Sinkhorn divergence. *Adv. Neural Inform. Process. Syst.* **33** (2020) 2257–2269.
- [22] G. Conforti and L. Tamanini, A formula for the time derivative of the entropic cost and applications. *J. Funct. Anal.* **280** (2021) Paper No. 108964.
- [23] G. Carlier, P. Pegon and L. Tamanini, Convergence rate of general entropic optimal transport costs. *Calc. Var. Part. Differ. Equ.* **62** (2023) Paper No. 116.
- [24] S. Eckstein and M. Nutz, Convergence rates for regularized optimal transport via quantization. *Math. Oper. Res.* **49** (2024) 1223–1240.
- [25] H. Malamut and M. Sylvestre, Convergence rates of the regularized optimal transport: disentangling suboptimality and entropy. *SIAM J. Math. Anal.* **57** (2025) 2533–2558.
- [26] A.-A. Pooladian and J. Niles-Weed, Entropic estimation of optimal transport maps. arXiv preprint arXiv:2109.12004 (2021).
- [27] R.T. Rockafellar, *Convex analysis*, vol. No. 28 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ (1970).
- [28] R. Cominetti and J. San Martín, Asymptotic analysis of the exponential penalty trajectory in linear programming. *Math. Program.* **67** (1994) 169–187.
- [29] A. Delalande, Nearly tight convergence bounds for semi-discrete entropic optimal Transport, in *International Conference on Artificial Intelligence and Statistics*. PMLR (2022) 1619–1642.
- [30] R. Sadhu, Z. Goldfeld and K. Kato, Approximation rates of entropic maps in semidiscrete optimal transport. *Electron. Commun. Probab.* **30** (2025) Paper No. 36.
- [31] A. González-Sanz and M. Nutz, Sparsity of quadratically regularized optimal transport: Scalar case. arXiv preprint arXiv:2410.03353 (2024).
- [32] L.A. Caffarelli, Monotonicity properties of optimal transportation and the FKG and related inequalities. *Commun. Math. Phys.* **214** (2000) 547–563.
- [33] S. Chewi and A.-A. Pooladian, An entropic generalization of Caffarelli’s contraction theorem via covariance inequalities. *C. R. Math. Acad. Sci. Paris* **361** (2023) 1471–1482.
- [34] M. Fathi, N. Gozlan and M. Prod’homme, A proof of the Caffarelli contraction theorem via entropic regularization. *Calc. Var. Part. Differ. Equ.* **59** (2020) Paper No. 96.
- [35] D. Bakry, I. Gentil and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, vol. 348 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham (2014).
- [36] N. Gigli, On Hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proc. Edinb. Math. Soc.* **54** (2011) 401–409.
- [37] F. Soudský, A. Molchanova and T. Roskovec, Interpolation between Hölder and Lebesgue spaces with applications. *J. Math. Anal. Appl.* **466** (2018) 160–168.
- [38] L.C. Evans, *Partial differential equations*, vol. 19 of *Graduate Studies in Mathematics*, 2nd edn. American Mathematical Society, Providence, RI (2010).
- [39] M. Gelbrich, On a formula for the  $L^2$  Wasserstein metric between measures on Euclidean and Hilbert spaces. *Math. Nachr.* **147** (1990) 185–203.

- [40] H. Janati, B. Muzellec, G. Peyré and M. Cuturi. Entropic optimal transport between unbalanced gaussian measures has a closed form. *Adv. Neural Inform. Process. Syst.* **33** (2020) 10468–10479.
- [41] E. del Barrio and J.-M. Loubes, The statistical effect of entropic regularization in optimal transportation. arXiv preprint arXiv:2006.05199 (2020).
- [42] A. Mallasto, A. Gerolin and H.Q. Minh, Entropy-regularized 2-Wasserstein distance between Gaussian measures. *Inf. Geom.* **5** (2022) 289–323.
- [43] P. Del Moral and A. Niclas, A Taylor expansion of the square root matrix function. *J. Math. Anal. Appl.* **465** (2018) 259–266.



**Please help to maintain this journal in open access!**

This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting [subscribers@edpsciences.org](mailto:subscribers@edpsciences.org).

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.