Quantitative Stability of Barycenters in the Wasserstein Space

Guillaume Carlier, Alex Delalande, Quentin Merigot

QUANTITATIVE STABILITY OF BARYCENTERS IN THE WASSERSTEIN SPACE

GUILLAUME CARLIER, ALEX DELALANDE, AND QUENTIN MÉRIGOT

Abstract. Wasserstein barycenters define averages of probability measures in a geometrically meaningful way. Their use is increasingly popular in applied fields, such as image, geometry or language processing. In these fields however, the probability measures of interest are often not accessible in their entirety and the practitioner may have to deal with statistical or computational approximations instead. In this article, we quantify the effect of such approximations on the corresponding barycenters. We show that Wasserstein barycenters depend in a Hölder-continuous way on their marginals under relatively mild assumptions. Our proof relies on recent estimates that allow to quantify the strong convexity of the barycenter functional. Consequences regarding the statistical estimation of Wasserstein barycenters and the convergence of regularized Wasserstein barycenters towards their non-regularized counterparts are explored.

Keywords: Optimal transport, Barycenters, Quantitative stability.

2020 Mathematics Subject Classification: 49Q22, 49K40.

1. Introduction

Wasserstein barycenters are Fréchet means in Wasserstein spaces: they define averages of families of probability measures that are consistent with the optimal transport geometry and generalize to more than two measures the fundamental notion of displacement interpolation due to McCann [28]. As such, they average out probability measures in a geometrically meaningful way and appear as a relevant tool to interpolate or summarize measure data. This notion of barycenter have indeed found many successful applications, for instance in image processing [32], geometry processing [36], language processing [19, 14, 27], statistics [37] or machine learning [15, 22]. We refer the readers to existing surveys [31, 29] for further applications. In such applications however, the probability measures of interest are often not accessible in their entirety. They may be accessible for instance only through noisy samples in a statistical context, or they may be approximated in order to use existing computational methods that estimate Wasserstein barycenters (see e.g. [12, 4, 15, 3]) while paying an affordable computational cost. Thus, in addition to the computational error induced by the algorithm used to calculate the barycenter, the practitioner may be subject to an extra statistical or approximation error that corresponds to the approximation of the marginal measures of interest. While works focusing on the computation of Wasserstein barycenters may now come with guarantees on the first type of error (see e.g. [3]), very little is known on the second type of error, which corresponds broadly speaking to a stability error since it quantifies the effect of a perturbation of the marginals on the corresponding barycenters. In this work, we focus on this type of error and show that the Wasserstein barycenter depends in an Hölder-continuous way on its marginal measures under regularity assumptions on (some of) the latter. In the remainder of this section, we define Wasserstein barycenters and the setting we focus on. We then show that mild regularity assumptions are necessary in order to hope for any stability result. Next, we give the dual formulation of the Wasserstein barycenter problem in our context, that is necessary to present our main assumption. This assumption and our main result are then

stated and we conclude this section by giving some immediate but useful consequences of our main result.

1.1. **Wasserstein barycenters.** Introduced in [1] for finite families of probability measures supported over a Euclidean space, the definition of Wasserstein barycenters have been extended to infinite families of probability measures in [7, 30], possibly supported over a Riemannian manifold in [23, 25]. In this work, we focus on families of probability measures supported over a compact Euclidean domain. Let $\Omega = B(0, R) \subset \mathbb{R}^d$ be the ball of $\mathbb{R}^d$ centered at zero and of radius $R > 0$ and denote $\mathcal{P}(\Omega)$ the set of Borel probability measures over Ω. We endow $\mathcal{P}(\Omega)$ with the 2-Wasserstein distance W_2 defined for any $\rho, \mu \in \mathcal{P}(\Omega)$ by

$$\mathrm{W}_2(\rho, \mu) = \left(\min_{\gamma \in \Gamma(\rho,\mu)} \int_{\Omega \times \Omega} \|x - y\|^2 \, \mathrm{d}\gamma(x, y) \right)^{1/2},$$

where the minimum is taken over the set $\Gamma(\rho, \mu)$ of transport plans between ρ and μ. We equip $\mathcal{P}(\Omega)$ with the the topology induced by W_2 (i.e. the weak topology) and denote $\mathcal{P}(\mathcal{P}(\Omega))$ the set of corresponding Borel probability measures over $\mathcal{P}(\Omega)$. For a measure $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$, we introduce its *variance functional* $F_{\mathbb{P}}$ defined from $\mathcal{P}(\Omega)$ to $\mathbb{R}$ by:

$$F_{\mathbb{P}} : \mu \mapsto \frac{1}{2} \int_{\mathcal{P}(\Omega)} \mathrm{W}_2^2(\rho, \mu) \mathrm{d}\mathbb{P}(\rho).$$

A Wasserstein barycenter of $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ is then defined as a minimizer $\mu_{\mathbb{P}}$ of the variance functional $F_{\mathbb{P}}$:

$$\mu_{\mathbb{P}} \in \arg \min_{\mu \in \mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu).$$

Such a minimizer always exists, and it is uniquely defined whenever $\mathbb{P}(\mathcal{P}_{a.c.}(\Omega)) > 0$, where $\mathcal{P}_{a.c.}(\Omega)$ denotes the set of probability measures over Ω that are absolutely continuous with respect to the Lebesgue measure [23, 25].

1.2. **Stability of Wasserstein barycenters.** As mentioned above, the population of interest $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ may not always be accessible in practice, and one may have to deal with another measure $\mathbb{Q} \in \mathcal{P}(\mathcal{P}(\Omega))$ instead. The stability question that then comes up is the following: can we bound a distance between minimizers $\mu_{\mathbb{P}}$ of $F_{\mathbb{P}}$ and $\mu_{\mathbb{Q}}$ of $F_{\mathbb{Q}}$ in terms of a distance between $\mathbb{P}$ and $\mathbb{Q}$? While the above-defined 2-Wasserstein distance gives a natural metric to compare $\mu_{\mathbb{P}}$ and $\mu_{\mathbb{Q}}$, there remains to choose a metric in order to compare $\mathbb{P}$ and $\mathbb{Q}$. For this, we will use the following 1-Wasserstein distance over $\mathcal{P}(\mathcal{P}(\Omega))$, defined for any $\mathbb{P}, \mathbb{Q}$ in $\mathcal{P}(\mathcal{P}(\Omega))$ by

$$\mathcal{W}_1(\mathbb{P}, \mathbb{Q}) = \min_{\gamma \in \Gamma(\mathbb{P},\mathbb{Q})} \int_{\mathcal{P}(\Omega) \times \mathcal{P}(\Omega)} \mathrm{W}_2(\rho, \tilde{\rho}) \mathrm{d}\gamma(\rho, \tilde{\rho}).$$

This choice of distance is justified by the fact that Wasserstein distances are naturally defined for probability measures on the compact metric space $(\mathcal{P}(\Omega), \mathrm{W}_2)$ and that they allow to compare measures that have incomparable support. The 1-Wasserstein distance being the weakest of the Wasserstein distances, our bounds are ensured to be the sharpest in terms of this optimal transport geometry. We are thus interested in bounding $\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{Q}})$ in terms of $\mathcal{W}_1(\mathbb{P}, \mathbb{Q})$ for $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{P}(\Omega))$.

1.2.1. *Consistency of Wasserstein barycenters.* Before looking for any quantitative stability result, one may first wonder if the Wasserstein barycenters depend at least in a continuous way on their marginals. This question, framed under the notion of *consistency* of Wasserstein barycenters, has been answered positively in [7, 8] in some specific settings and in [25] in the most general setting. Theorem 3 of [25] ensures in particular the following:

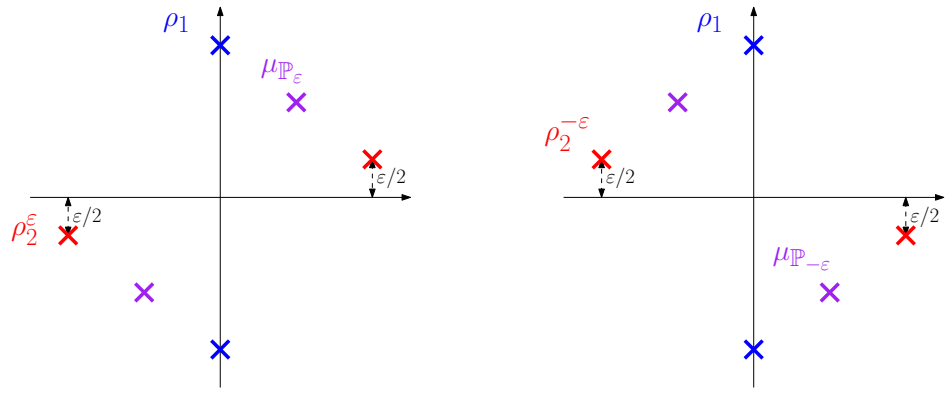

FIGURE 1. Let $\rho_1 = \frac{1}{2}(\delta_{(0;1)} + \delta_{(0;-1)})$. For $\varepsilon > 0$ and $x_\varepsilon = (1;\varepsilon/2) \in \mathbb{R}^2$, let $\rho_2^\varepsilon = \frac{1}{2}(\delta_{x^\varepsilon} + \delta_{-x^\varepsilon})$. Introduce $\mathbb{P}_\varepsilon = \frac{1}{2}(\delta_{\rho_1} + \delta_{\rho_2^\varepsilon})$. Then for $\varepsilon \leq \frac{1}{2}$, $\mathrm{W}_2(\mu_{\mathbb{P}_\varepsilon}, \mu_{\mathbb{P}_{-\varepsilon}}) = 1$ while $\mathcal{W}_1(\mathbb{P}_\varepsilon, \mathbb{P}_{-\varepsilon}) \leq \varepsilon$.

Theorem (Le Gouic, Loubes)**.** *Let $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ and a sequence $(\mathbb{P}_n)_{n\geq 1} \in \mathcal{P}(\mathcal{P}(\Omega))$ be such that*

$$\mathcal{W}_1(\mathbb{P}_n, \mathbb{P}) \xrightarrow[n\to+\infty]{} 0.$$

For all $n \geq 1$, denote $\mu_{\mathbb{P}_n}$ a barycenter of $\mathbb{P}_n$. Then the sequence $(\mu_{\mathbb{P}_n})_{n\geq 1}$ is precompact in $(\mathcal{P}(\Omega), \mathrm{W}_2)$ and any limit is a barycenter of $\mathbb{P}$.

This result ensures the continuity of Wasserstein barycenters with respect to the marginal measures, at least in our setting, so that we can now legitimately look for bounds that quantify this continuity.

1.2.2. *Quantitative stability in dimension $d = 1$.* In dimension $d = 1$, the derivation of quantitative stability bounds for Wasserstein barycenters is straightforward. Indeed, in this context W_2 is Hilbertian, which ensures a Lipschitz behavior of the barycenters with respect to their marginals. More precisely, denoting Q_ρ the *quantile function* of a measure $\rho \in \mathcal{P}(\Omega)$ (i.e. the generalized inverse of its cumulative distribution function), one has for any measures $\rho, \mu \in \mathcal{P}(\Omega)$ that $\mathrm{W}_2(\rho, \mu) = \|Q_\rho - Q_\mu\|_{\mathrm{L}^2([0,1])}$. This leads for any $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ to a simple formula for the unique barycenter:

$$\mu_{\mathbb{P}} = \left(\int_{\mathcal{P}(\Omega)} Q_\rho \mathrm{d}\mathbb{P}(\rho) \right)_{\#} \lambda_{[0,1]},$$

where $\lambda_{[0,1]}$ denotes the Lebesgue measure over $[0,1]$. Using this fact and the triangle inequality, one immediately obtains the following Lipschitz stability result, that actually holds for any families of measures in the set $\mathcal{P}_2(\mathbb{R})$ of probability measures supported over $\mathbb{R}$ that admit a finite second-order moment:

Proposition. *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{P}_2(\mathbb{R}))$ and denote $\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}$ their respective barycenters. Then*

$$\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}) \leq \mathcal{W}_1(\mathbb{P}, \mathbb{Q}).$$

This fact was exploited in [6] to characterize the statistical rate of convergence of empirical Wasserstein barycenters towards their population counterpart in an asymptotic setting for probability measures supported over the real line.

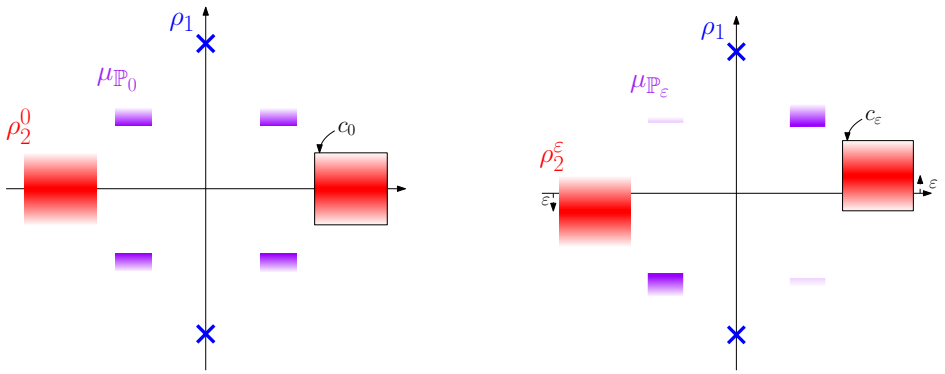

FIGURE 2. Let $\rho_1 = \frac{1}{2}(\delta_{(0;1)} + \delta_{(0;-1)})$. For $a \in (0,1)$ and $\varepsilon > 0$, let $c_\varepsilon = [1 - \frac{a}{2}; 1 + \frac{a}{2}] \times [-\frac{a}{2} + \varepsilon; \frac{a}{2} + \varepsilon]$ and ρ_2^ε the probability measure with density $\rho_2^\varepsilon(x,y) = \frac{\alpha}{2^{1-2\alpha}a^{1+2\alpha}} \left(|y-\varepsilon|^{2\alpha-1} \mathbb{1}_{c_\varepsilon}(x,y) + |y+\varepsilon|^{2\alpha-1} \mathbb{1}_{-c_\varepsilon}(x,y) \right)$ for some $\alpha > 0$. Introduce $\mathbb{P}_\varepsilon = \frac{1}{2}(\delta_{\rho_1} + \delta_{\rho_2^\varepsilon})$. Then for $\varepsilon \leq \frac{a}{2}$, $\mathrm{W}_2(\mu_{\mathbb{P}_0}, \mu_{\mathbb{P}_\varepsilon}) \sim \varepsilon^\alpha$ while $\mathcal{W}_1(\mathbb{P}_0, \mathbb{P}_\varepsilon) \leq \varepsilon$.

1.2.3. *Quantitative stability in dimension $d \geq 2$.* In dimension $d \geq 2$, the derivation of any quantitative stability bound turns out to be much more difficult. This first comes from the fact that without assumption on $\mathbb{P}$ and $\mathbb{Q}$, the barycenters $\mu_{\mathbb{P}}$ and $\mu_{\mathbb{Q}}$ may not be uniquely defined, which makes hopeless the derivation of any stability result. Even when uniqueness of the barycenters is ensured, one can easily build examples where no quantitative stability bound holds, see for instance the setting illustrated in Figure 1. This example relies on barycenters with only discrete marginals, and recovers in the limit $\varepsilon = 0$ the pathological case where the barycenter is not uniquely defined. One may circumvent this issue by ensuring, even in the limit $\varepsilon = 0$, uniqueness of the barycenter. As mentioned above, this can be done by imposing that some of the marginal measures are absolutely continuous. Nevertheless, even under such an assumption on the marginals, one can easily build an example where the barycenter achieves an Hölder behavior with respect to its marginal, but with an Hölder exponent that can be chosen arbitrarily small, see Figure 2. These negative results show that, even in dimension $d = 2$, regularity assumptions on the marginals $\mathbb{P}, \mathbb{Q}$ that go beyond sole absolute continuity are necessary in order to hope to derive stability estimates for their barycenters.

1.2.4. *Previous works.* Consistently with the above remarks, previous works having dealt with the stability of Wasserstein barycenters have either worked under stringent assumptions on the marginal measures or regularized the barycenter problem in order to ensure more regular solutions. In [2, 26] for instance, the question of the rate of convergence of the empirical barycenter in a Wasserstein space towards its population counterpart has been answered at the cost of assumptions that require in particular to have guarantees on the regularity of the (unknown) population barycenter (see sub-section 1.5.2 for more details). In [5, 11], a regularization of the barycenter problem has been considered and stability bounds and central limit theorems were deduced for the solutions to this regularized problem. In this work, we do not regularize the variance functional and work under less restrictive assumptions on the marginal measures than previous works having dealt with the stability of Wasserstein barycenters. In order to state these assumptions, we first need to introduce the dual problem to the Wasserstein barycenter problem.

1.3. **Dual formulation.** Building from [1], we show that the Wasserstein barycenter problem admits the following dual formulation with strong duality. The proof of this proposition is deferred to the appendix, Section A.

Proposition 1.1 (Dual formulation)**.** *For any $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$, one has*

$$\min_{\mu \in \mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu) = \frac{1}{2} \int_{\mathcal{P}(\Omega)} M_2(\rho) \mathrm{d}\mathbb{P}(\rho) - (\mathrm{D})_{\mathbb{P}},$$

where $M_2(\rho) = \langle \|\cdot\|^2 | \rho \rangle$ is the second-order moment of ρ and where $(\mathrm{D})_{\mathbb{P}}$ corresponds to the dual value

$$(\mathrm{D})_{\mathbb{P}} = \min \left\{ \int_{\mathcal{P}(\Omega)} \langle \psi_\rho^* | \rho \rangle \mathrm{d}\mathbb{P}(\rho) \mid (\psi_\rho)_\rho \in \mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega)), \quad \int_{\mathcal{P}(\Omega)} \psi_\rho(\cdot) \mathrm{d}\mathbb{P}(\rho) = \frac{\|\cdot\|^2}{2} \right\}.$$

In the expression above, $\psi_\rho^*(\cdot) = \sup_{y \in \Omega} \{ \langle \cdot | y \rangle - \psi_\rho(y) \}$ corresponds to the convex conjugate of ψ_ρ and $\mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ denotes the set of essentially bounded $\mathbb{P}$-measurable mappings from $\mathcal{P}(\Omega)$ to the Sobolev space $W^{1,\infty}(\Omega)$ of bounded Lipschitz continuous functions from Ω to $\mathbb{R}$.

Remark 1.1. Note that in the above minimization problem, $(\psi_\rho)_\rho$ is to be understood as the following mapping, defined $\mathbb{P}$-almost everywhere:

$$(\psi_\rho)_\rho : \left\{ \begin{array}{ll} \mathcal{P}(\Omega) & \to W^{1,\infty}(\Omega), \\ \rho & \mapsto \psi_\rho. \end{array} \right.$$

Remark 1.2. By Kantorovich duality [42], for $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$, the *collection of functions* $(\psi_\rho)_\rho$ solving $(\mathrm{D})_{\mathbb{P}}$ gives solutions to the optimal transport problems between $\mathbb{P}$-a.e. $\rho \in \mathcal{P}(\Omega)$ and any barycenter $\mu_{\mathbb{P}} \in \arg\min (\mathrm{P})_{\mathbb{P}}$:

$$\begin{aligned} \frac{1}{2} \mathrm{W}_2^2(\rho, \mu_{\mathbb{P}}) &= \frac{1}{2} M_2(\rho) + \frac{1}{2} M_2(\mu_{\mathbb{P}}) - \left(\langle \psi_\rho^* | \rho \rangle + \langle \psi_\rho | \mu_{\mathbb{P}} \rangle \right) \\ &= \frac{1}{2} M_2(\rho) + \frac{1}{2} M_2(\mu_{\mathbb{P}}) - \left(\min_{\psi \in \mathcal{C}(\Omega)} \langle \psi^* | \rho \rangle + \langle \psi | \mu_{\mathbb{P}} \rangle \right). \end{aligned} \tag{1}$$

As such, $\psi_\rho = \psi_\rho^{**}$ for $\mathbb{P}$-a.e. ρ, so that this function – that we call later on a (Kantorovich) potential – is convex and Lipschitz continuous with Lipschitz constant smaller than R. When $\mathbb{P}(\mathcal{P}_{a.c.}(\Omega)) > 0$ and $\rho \in \mathrm{spt}(\mathbb{P}) \cap \mathcal{P}_{a.c.}(\Omega)$, the convex function ψ_ρ^* is the Brenier potential [10] and its gradients achieves the optimal transport from ρ to the unique barycenter $\mu_{\mathbb{P}}$:

$$\left(\nabla \psi_\rho^*\right)_{\#} \rho = \mu_{\mathbb{P}}, \quad \text{and} \quad \mathrm{W}_2^2(\rho, \mu_{\mathbb{P}}) = \left\| \nabla \psi_\rho^* - \mathrm{id} \right\|_{\mathrm{L}^2(\rho; \mathbb{R}^d)}^2.$$

1.4. **Assumptions.** For any $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$, the variance functional $F_{\mathbb{P}}$ is convex. Stability estimates for the minimizers of $F_{\mathbb{P}}$ (which are the Wasserstein barycenters of $\mathbb{P}$) may thus be obtained from estimates on the *strong convexity* or *curvature* of $F_{\mathbb{P}}$. However, without any assumption on $\mathbb{P}$, the variance functional $F_{\mathbb{P}}$ is in general not *strongly-convex* in any sense. In fact, it is easy to construct examples where $F_{\mathbb{P}}$ showcases an affine behavior with respect to the linear structure of $\mathcal{P}(\Omega)$:

Example 1.2. For any $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ of the form $\mathbb{P} = \sum_i \lambda_i \delta_{\delta_{x_i}}$, one has for any $y, z \in \Omega$ and $t \in [0,1]$ the relation

$$F_{\mathbb{P}}((1-t)\delta_y + t\delta_z) = (1-t) F_{\mathbb{P}}(\delta_y) + t F_{\mathbb{P}}(\delta_z).$$

Our main stability applies to measures $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ such that the variance functional $F_{\mathbb{P}}$ satisfies a strong convexity estimate, also called a *variance inequality* in the language of

[38, 13]. Because for any $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ we have

$$F_{\mathbb{P}}(\cdot) = \frac{1}{2}\int_{\mathcal{P}(\Omega)} \mathrm{W}_2^2(\rho, \cdot)\mathrm{d}\mathbb{P}(\rho),$$

it suffices to choose a $\mathbb{P}$ whcih gives positive mass to probability measures $\rho \in \mathcal{P}(\Omega)$ for which the squared Wasserstein distance $\frac{1}{2}\mathrm{W}_2^2(\rho, \cdot)$ satisfies itself a strong convexity estimate. In turn, relying on Kantorovich's dual formulation displayed in (1), this can be obtained from the assumption that the minimized functional in the dual problem presents a form of local strong convexity. We will denote $\mathcal{K}_\rho : \psi \mapsto \langle \psi^* | \rho \rangle$ the *Kantorovich functional* associated to $\rho \in \mathcal{P}(\Omega)$. This convex functional appears in the minimization problem (1); its gradient formally reads $\nabla \mathcal{K}_\rho(\psi) = -(\nabla \psi^*)_{\#}\rho$. We will make the following assumption:

Assumption 1.3. *There exists constants $\alpha_{\mathbb{P}} \in (0, 1]$, $c_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}}, m_{\mathbb{P}}, M_{\mathbb{P}} \in (0, +\infty)$ and a measurable set $S_{\mathbb{P}} \subset \mathcal{P}(\Omega)$ verifying $\mathbb{P}(S_{\mathbb{P}}) = \alpha_{\mathbb{P}}$ and such that for all $\rho \in S_{\mathbb{P}}$,*

1. *$\rho \in \mathcal{P}_{a.c.}(\Omega)$,*
2. *$m_{\mathbb{P}} \leq \rho_{|\mathrm{spt}(\rho)} \leq M_{\mathbb{P}}$,*
3. *$\mathrm{spt}(\rho)$ has a $\mathcal{H}^{d-1}$-rectifiable boundary and $\mathcal{H}^{d-1}(\partial \mathrm{spt}(\rho)) \leq \mathrm{per}_{\mathbb{P}}$,*
4. *$\forall \psi, \tilde{\psi} \in \mathcal{C}(\Omega), \quad c_{\mathbb{P}} \mathbb{V}\mathrm{ar}_\rho(\tilde{\psi}^* - \psi^*) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi) - \langle \psi - \tilde{\psi} | (\nabla \psi^*)_{\#}\rho \rangle$,*

where $\mathrm{spt}(\rho)$ denotes the support of ρ, $\partial \mathrm{spt}(\rho)$ denotes the topological boundary of this support and $\mathcal{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure.

While conditions (1), (2) and (3) speak for themselves, condition (4) might seem ad hoc and difficult to verify. However, conditions under which a measure $\rho \in \mathcal{P}_{a.c.}(\Omega)$ verifies the local strong convexity estimate (4) of Assumption 1.3 are given in [18] as a consequence of the Brascamp-Lieb concentration inequality [9]. In particular, this estimate holds for an absolutely continuous measure ρ, supported on a compact convex set, and whose density is bounded away from zero and infinity. In the appendix, we slightly extend this result to measures supported on a connected union of convex sets, thus showing that the convexity of the support of ρ is not absolutely necessary to get strong convexity of $\mathcal{K}_\rho$.

Proposition 1.4. *Let $\rho \in \mathcal{P}_{a.c.}(\Omega)$ and assume that there exists $m_\rho, M_\rho \in (0, +\infty)$ such that $m_\rho \leq \rho \leq M_\rho$ on $\mathrm{spt}(\rho)$. Assume in addition that ρ satisfies a Poincaré-Wirtinger inequality and that $\mathrm{spt}(\rho)$ is a connected finite union of convex sets. Then there exists $c_\rho > 0$ such that for all $\psi, \tilde{\psi} \in \mathcal{C}(\Omega)$,*

$$c_\rho \mathbb{V}\mathrm{ar}_\rho(\tilde{\psi}^* - \psi^*) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi) - \langle \tilde{\psi} - \psi | \nabla \mathcal{K}_\rho(\psi) \rangle.$$

We refer to Proposition B.2 of the appendix for a precise statement and a proof. We conjecture that such a strong convexity estimate actually holds for any absolutely continuous measure satisfying the Poincaré-Wirtinger inequality, maybe with mild additional assumptions on the density and its support. However, this is not the focus of the present article and we leave this for future work. On a more technical side, we note that the Borel measurability of a set $S_{\mathbb{P}} \subset \mathcal{P}(\Omega)$ as defined in Assumption 1.3 needs to be checked depending on the application. Obviously, measurability holds when the number of marginals is finite ($\mathbb{P}$ is discrete) and $S_{\mathbb{P}}$ is a (finite) subset of these marginals. When the number of marginals is not finite, we note that if $S_{\mathbb{P}}$ is made of *all* the measures $\rho \in \mathcal{P}(\Omega)$ that satisfy conditions (1)–(3) and that have convex supports, then the measures of $S_{\mathbb{P}}$ satisfy condition (4) by [18] and the set $S_{\mathbb{P}}$ is closed for the weak topology.

1.5. **Main result and consequences.** Under Assumption 1.3, we prove that the Wasserstein barycenters depend in a Hölder-continuous way on their marginals:

Theorem 1.5. *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{P}(\Omega))$ and assume that $\mathbb{P}$ satisfies Assumption 1.3. Let $\mu_{\mathbb{P}}$ be the barycenter of $\mathbb{P}$ and $\mu_{\mathbb{Q}}$ be a barycenter of $\mathbb{Q}$. Then*

$$\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}) \leq \left(\frac{C_{d, m_{\mathbb{P}}, M_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}}, c_{\mathbb{P}}}}{\alpha_{\mathbb{P}}} \right)^{1/6} \mathcal{W}_1(\mathbb{P}, \mathbb{Q})^{1/6},$$

where $C_{d, m_{\mathbb{P}}, M_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}}, c_{\mathbb{P}}} = C_d \frac{M_{\mathbb{P}}^3}{m_{\mathbb{P}}} \frac{\mathrm{per}_{\mathbb{P}}^2}{c_{\mathbb{P}}} R^5$ and where C_d is a dimensional constant. It also holds, with the same constant:

$$\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}) \leq \left(\frac{C_{d, m_{\mathbb{P}}, M_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}}, c_{\mathbb{P}}}}{\alpha_{\mathbb{P}}} \right)^{1/5} \|\mathbb{P} - \mathbb{Q}\|_{\mathrm{TV}}^{1/5}.$$

In this result, the Hölder exponents might not be optimal. However, our structure of proof does not leave space to much improvement of these exponents (see Remark 2.3). Theorem 1.5 is essentially a corollary of the fact that whenever a measure $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ satisfies Assumption 1.3, its variance functional $F_{\mathbb{P}}$ satisfies a strong convexity estimate (Theorem 2.1). Note that a strong convexity estimate for the variance functional $F_{\mathbb{P}}$ may have an interest beyond the stability of Wasserstein barycenters with respect to their marginals, e.g. to control the *bias* induced by the entropic penalization of the variance functional as introduced in [5] (see Corollary 2.2). We defer the detailed proof of Theorem 1.5 to Section 2. Let us now mention some consequences of Theorem 1.5 in applications.

1.5.1. *Statistical estimation of barycenter with a finite number of marginals.* For a probability measure $\rho \in \mathcal{P}(\Omega)$ and an i.i.d. sequence $(x_j)_{j=1,\dots,n}$ sampled from ρ, it is well-known that the empirical measure $\hat{\rho}^n = \frac{1}{n} \sum_{j=1}^n \delta_{x_j}$ converges weakly to ρ almost-surely as $n \to \infty$ [41]. By Theorem 1 of [21], the rate of this convergence can be controlled in expected Wasserstein distance: there exists a constant C_d depending only on d such that

$$\mathbb{E}\mathrm{W}_2^2(\hat{\rho}^n, \rho) \leq C_d R^2 \begin{cases} n^{-1/2} & \text{if } d < 4, \\ n^{-1/2} \log(n) & \text{if } d = 4, \\ n^{-2/d} & \text{else,} \end{cases}$$

where the expectation is taken with respect to $(x_j)_{j=1,\dots,n} \sim \rho^{\otimes n}$. Theorem 1.5 together with a double use of Jensen's inequality allows to translate these rates to the statistical estimation of a Wasserstein barycenter with a finite number of marginals:

Corollary 1.6. *Let $\mathbb{P}_m = \sum_{i=1}^m \lambda_i \delta_{\rho_i} \in \mathcal{P}(\mathcal{P}(\Omega))$ satisfying Assumption 1.3. For all $i \in \{1, \dots, m\}$, denote $\hat{\rho}_i^n = \frac{1}{n} \sum_{j=1}^n \delta_{x_{i,j}}$ an empirical measure built from an i.i.d. sequence $(x_{i,j})_{1 \leq j \leq n}$ sampled from ρ_i. Then the barycenters $\mu_{\mathbb{P}_m}$ of $\mathbb{P}_m$ and $\mu_{\widehat{\mathbb{P}}_m^n}$ of $\widehat{\mathbb{P}}_m^n = \sum_{i=1}^m \lambda_i \delta_{\hat{\rho}_i^n}$ verify*

$$\mathbb{E}\mathrm{W}_2^2(\mu_{\widehat{\mathbb{P}}_m^n}, \mu_{\mathbb{P}_m}) \lesssim \frac{1}{\alpha_{\mathbb{P}_m}^{1/3}} \begin{cases} n^{-1/12} & \text{if } d < 4, \\ n^{-1/12} \log(n)^{1/6} & \text{if } d = 4, \\ n^{-1/(3d)} & \text{else,} \end{cases}$$

where $\lesssim$ hides a multiplicative constant that depends on $d, R, m_{\mathbb{P}_m}, M_{\mathbb{P}_m}, \mathrm{per}_{\mathbb{P}_m}$ and $c_{\mathbb{P}_m}$.

1.5.2. *Convergence rate of empirical barycenters in the Wasserstein space.* Another statistical question occurs in the setting where the population of marginals $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ is only known through samples $(\rho_i)_{1 \leq i \leq m} \sim \mathbb{P}^{\otimes m}$. Introducing the plug-in estimator $\mathbb{P}_m = \frac{1}{m} \sum_{i=1}^m \delta_{\rho_i}$, it is natural to wonder how well $\mu_{\mathbb{P}_m}$ approaches $\mu_{\mathbb{P}}$ in terms of m. This question, asked in the more general framework of barycenters in Alexandrov spaces, has been the object of recent research [2, 26]. In the Wasserstein space, the authors of [26] show in particular that $\mathbb{E}\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{P}_m})$ converges at the parametric rate $m^{-1/2}$ under the assumption that $\mathbb{P}$ admits a barycenter $\mu_{\mathbb{P}}$ that it is such that there exists a bi-Lipschitz optimal transport map between

any $\rho \in \mathrm{spt}(\mathbb{P})$ and $\mu_{\mathbb{P}}$, and that the Lipschitz constants of these maps and their inverses do not differ by a value more than 1. Under similar assumptions, the authors of [13] derive a strong convexity estimate for the variance functional at its minimum which helps them derive rates of convergence of gradient descent algorithms for the (stochastic) estimation of barycenters. Such assumptions however require to have guarantees on the regularity of a barycenter of $\mathbb{P}$, which can be obtained when restricted to specific families of probability measures (e.g. Gaussian measures), but are difficult to get in general (for instance, barycenters of measures with convex support may not have a convex support [34], which hampers a straightforward use of Caffarelli's regularity theory). In contrast, our stability result entails that for barycenters $\mu_{\mathbb{P}}$ of $\mathbb{P}$ and $\mu_{\mathbb{P}_m}$ of $\mathbb{P}_m$,

$$\mathbb{E}\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{P}_m}) \lesssim \frac{1}{\alpha_{\mathbb{P}}^{1/6}} \mathbb{E}\mathcal{W}_1(\mathbb{P}, \mathbb{P}_m)^{1/6},$$

whenever $\mathbb{P}$ satisfies Assumption 1.3. This implies that any rate of convergence of $\mathbb{E}\mathcal{W}_1(\mathbb{P}, \mathbb{P}_m)$ with respect to m is readily transferred to $\mathbb{E}\mathrm{W}_2(\mu_{\mathbb{P}}, \mu_{\mathbb{P}_m})$, up to an exponent. However, $\mathcal{P}(\Omega)$ is an infinite dimensional space and there is no general convergence rate for $\mathbb{E}\mathcal{W}_1(\mathbb{P}, \mathbb{P}_m)$. Nonetheless, assuming some structure on the population $\mathbb{P}$ may help derive convergence bounds. One may use for instance the notion of upper Wasserstein dimension of $\mathbb{P}$ introduced in [43] (Definition 4), defined from quantities that depend on the covering numbers of (subsets of) the support of $\mathbb{P}$. Assuming that this dimension is strictly upper bounded by $s > 0$, the authors of [43] show that

$$\mathbb{E}\mathcal{W}_1(\mathbb{P}, \mathbb{P}_m) \lesssim m^{-1/s},$$

where $\lesssim$ hides a multiplicative constant that depends on R and s. We note finally that if we assume that $\mathbb{P}$ satisfies Assumption 1.3 with $\alpha_{\mathbb{P}} = 1$, our results allow to get the following finite-sample guarantee for the empirical estimation of barycenters in the Wasserstein space:

Theorem 1.7. *Let $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ satisfying Assumption 1.3 with $\alpha_{\mathbb{P}} = 1$. For $m \geq 1$, introduce the plug-in estimator $\mathbb{P}_m = \frac{1}{m}\sum_{i=1}^m \delta_{\rho_i}$ built from an m-sample $(\rho_i)_{1\leq i\leq m} \sim \mathbb{P}^{\otimes m}$. Then the barycenters $\mu_{\mathbb{P}}$ of $\mathbb{P}$ and $\mu_{\mathbb{P}_m}$ of $\mathbb{P}_m$ satisfy*

$$\mathbb{E}\mathrm{W}_2(\mu_{\mathbb{P}_m}, \mu_{\mathbb{P}}) \lesssim m^{-1/30},$$

where $\lesssim$ hides a multiplicative constant that depends on $d, R, m_{\mathbb{P}}, M_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}}$ and $c_{\mathbb{P}}$.

The main idea to prove this result is to see the minimization of $F_{\mathbb{P}}$ through the lens of risk minimization, so that the minimization of $F_{\mathbb{P}_m}$ corresponds to a problem of empirical risk minimization (ERM) [40]. Under the assumptions of Theorem 1.7, the empirical risk $F_{\mathbb{P}_m}$ is ensured to be *strongly-convex* almost surely, which allows to derive stability bounds for the empirical risk minimizer $\mu_{\mathbb{P}_m}$ with respect to its population counterpart $\mu_{\mathbb{P}}$ using classical ideas from the ERM litterature [35]. The detailed proof of Theorem 1.7 is deferred to Section 3.

1.5.3. *Error induced by a discretization of the marginals.* Let $\rho \in \mathcal{P}(\Omega)$ and let $h > 0$ be a discretization parameter. Denoting $(x_i^h)_{1\leq i\leq N_h}$ an h-net of Ω and $(V_i^h)_{1\leq i\leq N_h}$ the corresponding Voronoi tessellation of Ω, it is trivial to verify that the discretization $\rho^h = \sum_{i=1}^{N_h} \rho(V_i^h)\delta_{x_i^h}$ verifies

$$\mathrm{W}_2(\rho, \rho^h) \leq h.$$

Such a type of discretization, with controlled error bound, may be useful in practice for computational purposes. The stability result of Theorem 1.5 allows to translate the error bound made when discretizing the marginals to the corresponding barycenter:

Corollary 1.8. *Let $\mathbb{P}_m = \sum_{i=1}^m \lambda_i \delta_{\rho_i} \in \mathcal{P}(\mathcal{P}(\Omega))$ satisfying Assumption 1.3. Let $h > 0$ and for all $i \in \{1, \dots, m\}$, denote $\rho_i^h = \sum_{j=1}^{N_h} \rho_i(V_j^h)\delta_{x_j^h}$ a discretization of ρ_i built from the h-net $(x_j^h)_{1\leq j\leq N_h}$. Then the barycenters $\mu_{\mathbb{P}_m}$ of $\mathbb{P}_m$ and $\mu_{\mathbb{P}_m^h}$ of $\mathbb{P}_m^h = \sum_{i=1}^m \lambda_i \delta_{\rho_i^h}$ verify*

$$\mathrm{W}_2(\mu_{\mathbb{P}_m^h}, \mu_{\mathbb{P}_m}) \lesssim \frac{1}{\alpha_{\mathbb{P}_m}^{1/6}} h^{1/6},$$

where $\lesssim$ hides a multiplicative constant that depends on $d, R, m_{\mathbb{P}_m}, M_{\mathbb{P}_m}, \mathrm{per}_{\mathbb{P}_m}$ and $c_{\mathbb{P}_m}$.

2. Strong convexity of the variance functional

Let us recall the definition of the variance functional associated to some $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$:

$$F_{\mathbb{P}} : \left\{ \begin{array}{ll} \mathcal{P}(\Omega) & \to \mathbb{R}, \\ \mu & \mapsto \frac{1}{2}\int_{\mathcal{P}(\Omega)} \mathrm{W}_2^2(\rho, \mu)\mathrm{d}\mathbb{P}(\rho). \end{array} \right.$$

Our objective in this section is to get a *strong convexity* estimate for the variance functional $F_{\mathbb{P}}$ when $\mathbb{P}$ satisfies Assumption 1.3. More precisely, we wish to quantify to much extent the graph of the convex functional $F_{\mathbb{P}}$ lies above its tangents. For any measure $\mu \in \mathcal{P}(\Omega)$, the directions of the tangents of the graph of $F_{\mathbb{P}}$ at μ are given by the subdifferential of $F_{\mathbb{P}}$ evaluated at μ and denoted $\partial F_{\mathbb{P}}(\mu)$. This subdifferential may be described using Kantorovich's duality formula [42], already mentioned in equation (1), that ensures that for any $\rho, \mu \in \mathcal{P}(\Omega)$, one has

$$\frac{1}{2}\mathrm{W}_2^2(\rho, \mu) = \langle \frac{1}{2}\left\|\cdot\right\|^2 |\rho\rangle + \langle \frac{1}{2}\left\|\cdot\right\|^2 |\mu\rangle - \left(\min_{\psi \in \mathcal{C}(\Omega)} \langle \psi^* | \rho\rangle + \langle \psi | \mu\rangle \right). \tag{2}$$

From this formula, one easily has the following description of the subdifferential of the half-squared Wasserstein distance to a fixed measure $\rho \in \mathcal{P}(\Omega)$ (Proposition 7.17 of [33]):

$$\partial \left[\frac{1}{2}\mathrm{W}_2^2(\rho, \cdot)\right](\mu) = \left\{ \frac{1}{2}\left\|\cdot\right\|^2 - \psi_{\rho\to\mu} \mid \psi_{\rho\to\mu} \in \arg\min_{\psi\in\mathcal{C}(\Omega)} \langle \psi^*|\rho\rangle + \langle \psi|\mu\rangle \right\}.$$

This allows to directly characterize the subdifferential of $F_{\mathbb{P}}$ at any $\mu \in \mathcal{P}(\Omega)$ as follows:

$$\partial F_{\mathbb{P}}(\mu) = \left\{ \int_{\mathcal{P}(\Omega)} \left(\frac{1}{2}\left\|\cdot\right\|^2 - \psi_{\rho\to\mu} \right) \mathrm{d}\mathbb{P}(\rho) \mid \text{for } \mathbb{P}\text{-a.e. } \rho,\ \psi_{\rho\to\mu} \in \arg\min_{\psi\in\mathcal{C}(\Omega)} \langle \psi^*|\rho\rangle + \langle \psi|\mu\rangle \right\}.$$

Thus for any μ and ν in $\mathcal{P}(\Omega)$, for any *collection* $(\psi_{\rho\to\mu})_\rho \in \mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ of Kantorovich potentials in the transport between $\mathbb{P}$-almost every $\rho \in \mathcal{P}(\Omega)$ and μ, we have by definition of the subdifferential:

$$F_{\mathbb{P}}(\mu) + \langle \int_{\mathcal{P}(\Omega)} \left(\frac{1}{2}\left\|\cdot\right\|^2 - \psi_{\rho\to\mu} \right) \mathrm{d}\mathbb{P}(\rho) | \nu - \mu\rangle \leq F_{\mathbb{P}}(\nu).$$

Our strong convexity estimate quantifies the gap in this subdifferential inequality under the hypothesis that $\mathbb{P}$ satisfies Assumption 1.3:

Theorem 2.1. *Let $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ satisfying Assumption 1.3. Let $\mu, \nu \in \mathcal{P}(\Omega)$ and let $(\psi_{\rho\to\mu})_\rho \in \mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ be a* collection *of Kantorovich potentials in the transport between $\mathbb{P}$-almost every $\rho \in \mathcal{P}(\Omega)$ and μ. Then it holds:*

$$\alpha_{\mathbb{P}}\mathrm{W}_2^6(\mu, \nu) \lesssim F_{\mathbb{P}}(\nu) - F_{\mathbb{P}}(\mu) - \langle \int_{\mathcal{P}(\Omega)} \left(\frac{1}{2}\left\|\cdot\right\|^2 - \psi_{\rho\to\mu} \right) \mathrm{d}\mathbb{P}(\rho) | \nu - \mu\rangle,$$

where $\lesssim$ hides on the right-hand side the multiplicative constant

$$C_{d, m_{\mathbb{P}}, M_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}}} = C_d \frac{M_{\mathbb{P}}^3}{m_{\mathbb{P}}} \frac{\mathrm{per}_{\mathbb{P}}^2}{c_{\mathbb{P}}} R^4,$$

where C_d is a dimensional constant.

Remark 2.1. This estimate holds without any regularization of the variance functional. As such, it may be used directly to study the stability of *smoothed* notions of Wasserstein barycenters defined from a regularization of the variance functional [5, 11], yielding stability bounds that may not depend on the regularization parameter(s). Other versions of *smoothed* Wasserstein barycenters have also been obtained from a regularization of the Wasserstein distance itself, such as the celebrated entropic regularization [15]. The stability of these barycenters may also be obtained from similar strong convexity estimates found in the context of entropic optimal transport [16, 17]. Finally, the estimate of Theorem 2.1 can be used to study the convergence of regularized Wasserstein barycenters towards their non-regularized counterparts, as indicated by the following corollary.

Corollary 2.2. *Let $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\Omega))$ satisfying Assumption 1.3. For $\lambda \geq 0$, denote*

$$\mu_{\mathbb{P}}^{\lambda} = \arg\min_{\mu \in \mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu) + \lambda G(\mu),$$

where $G : \mathcal{P}(\Omega) \to \mathbb{R}$ is either the entropy $G(\mu) = \int_\Omega \mu \log \mu$ or $G(\mu) = \int_\Omega \mu^p$ for some $p \geq 1$. Then for any $\lambda > 0$,

$$\mathrm{W}_2(\mu_{\mathbb{P}}^{\lambda}, \mu_{\mathbb{P}}^{0}) \lesssim \lambda^{1/6},$$

where $\lesssim$ hides a multiplicative constant that depends on $d, R, m_{\mathbb{P}}, M_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}}, c_{\mathbb{P}}$ and $\alpha_{\mathbb{P}}$.

Proof. Theorem 2.1 together with the positiveness of G and the definition of $\mu_{\mathbb{P}}^{\lambda}$ yield

$$\alpha_{\mathbb{P}} \mathrm{W}_2^6(\mu_{\mathbb{P}}^{\lambda}, \mu_{\mathbb{P}}^{0}) \lesssim F_{\mathbb{P}}(\mu_{\mathbb{P}}^{\lambda}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}^{0}) \leq F_{\mathbb{P}}(\mu_{\mathbb{P}}^{\lambda}) + \lambda G(\mu_{\mathbb{P}}^{\lambda}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}^{0}) \leq F_{\mathbb{P}}(\mu_{\mathbb{P}}^{0}) + \lambda G(\mu_{\mathbb{P}}^{0}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}^{0}).$$

The conclusion follows from the boundedness of $G(\mu_{\mathbb{P}}^{0})$ induced by the maximum principle followed by $\mu_{\mathbb{P}}^{0} = \mu_{\mathbb{P}} \in \mathcal{P}_{a.c.}(\Omega)$ when $\mathbb{P}$ satisfies Assumption 1.3 (Proposition 4.7 and Remark 4.8 of [11]):

$$\left\| \mu_{\mathbb{P}}^{0} \right\|_{\mathrm{L}^\infty} \leq M_{\mathbb{P}} / \alpha_{\mathbb{P}}^{d}. \qquad \square$$

Before proving Theorem 2.1, let us use it to prove the stability estimate of Theorem 1.5.

Proof of Theorem 1.5. Let $(\psi_\rho)_\rho = (\psi_{\rho \to \mu_{\mathbb{P}}})_\rho \in \mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ be a collection of potentials that give a dual solution to the barycenter problem with population $\mathbb{P}$ (Proposition 1.1). We have in particular

$$\int_{\mathcal{P}(\Omega)} \psi_\rho(\cdot) \mathrm{d}\mathbb{P}(\rho) = \frac{1}{2} \left\| \cdot \right\|^2.$$

Applying Theorem 2.1 with $\mu = \mu_{\mathbb{P}}$, $\nu = \mu_{\mathbb{Q}}$ and with the collection of potentials $(\psi_\rho)_\rho$, we have the bound

$$\alpha_{\mathbb{P}} \mathrm{W}_2^6(\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}) \lesssim F_{\mathbb{P}}(\mu_{\mathbb{Q}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}).$$

By definition of $\mu_{\mathbb{Q}}$ as a minimizer of $F_{\mathbb{Q}}$, we have

$$F_{\mathbb{Q}}(\mu_{\mathbb{P}}) - F_{\mathbb{Q}}(\mu_{\mathbb{Q}}) \geq 0.$$

Thus the following bound holds:

$$\begin{aligned} \alpha_{\mathbb{P}} \mathrm{W}_2^6(\mu_{\mathbb{P}}, \mu_{\mathbb{Q}}) &\lesssim F_{\mathbb{P}}(\mu_{\mathbb{Q}}) - F_{\mathbb{Q}}(\mu_{\mathbb{Q}}) + F_{\mathbb{Q}}(\mu_{\mathbb{P}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}) \\ &= \langle \frac{1}{2} \mathrm{W}_2^2(\cdot, \mu_{\mathbb{Q}}) | \mathbb{P} - \mathbb{Q} \rangle + \langle \frac{1}{2} \mathrm{W}_2^2(\cdot, \mu_{\mathbb{P}}) | \mathbb{Q} - \mathbb{P} \rangle \\ &= \langle \frac{1}{2} (\mathrm{W}_2^2(\cdot, \mu_{\mathbb{Q}}) - \mathrm{W}_2^2(\cdot, \mu_{\mathbb{P}})) | \mathbb{P} - \mathbb{Q} \rangle. \end{aligned} \tag{3}$$

The mapping $\rho \mapsto \frac{1}{2}(\mathrm{W}_2^2(\rho,\mu_{\mathbb{Q}}) - \mathrm{W}_2^2(\rho,\mu_{\mathbb{P}}))$ being $4R$-Lipschitz continuous with respect to W_2, we finally have with the Kantorovich-Rubinstein duality result the bound

$$\langle \frac{1}{2}(\mathrm{W}_2^2(\cdot,\mu_{\mathbb{Q}}) - \mathrm{W}_2^2(\cdot,\mu_{\mathbb{P}}))|\mathbb{P}-\mathbb{Q}\rangle \leq 4R\mathcal{W}_1(\mathbb{P},\mathbb{Q}),$$

which gives the first estimate of the statement. Now remark that for any $\rho \in \mathcal{P}(\Omega)$, the triangle inequality yields

$$\left|\mathrm{W}_2^2(\rho,\mu_{\mathbb{Q}}) - \mathrm{W}_2^2(\rho,\mu_{\mathbb{P}})\right| \leq (\mathrm{W}_2(\rho,\mu_{\mathbb{Q}}) + \mathrm{W}_2(\rho,\mu_{\mathbb{P}}))\mathrm{W}_2(\mu_{\mathbb{P}},\mu_{\mathbb{Q}}) \leq 4R\mathrm{W}_2(\mu_{\mathbb{P}},\mu_{\mathbb{Q}}).$$

Injecting this last bound into (3) gives the second bound of the statment:

$$\alpha_{\mathbb{P}}\mathrm{W}_2^6(\mu_{\mathbb{P}},\mu_{\mathbb{Q}}) \lesssim \mathrm{W}_2(\mu_{\mathbb{P}},\mu_{\mathbb{Q}})\left\|\mathbb{P}-\mathbb{Q}\right\|_{\mathrm{TV}}. \qquad \square$$

Let us now prove Theorem 2.1. This result simply relies on the fact that for any $\rho \in \mathcal{P}(\Omega)$ that belongs to the set $S_{\mathbb{P}}$ from Assumption 1.3, the function $\frac{1}{2}\mathrm{W}_2^2(\rho,\cdot)$ satisfies a strong convexity estimate given in the following proposition. Theorem 2.1 then immediately follows from this proposition after summing over $\rho \sim \mathbb{P}$. In the following statement, we recall that the *Kantorovich functional* $\mathcal{K}_\rho$ associated to a measure $\rho \in \mathcal{P}(\Omega)$ corresponds to the map

$$\mathcal{K}_\rho : \left\{ \begin{array}{ll} \mathcal{C}(\Omega) & \to \mathbb{R}, \\ \psi & \mapsto \int_\Omega \psi^* \mathrm{d}\rho. \end{array} \right.$$

Proposition 2.3. *Let $\rho \in \mathcal{P}_{a.c.}(\Omega)$ be absolutely continuous and such that there exists $m_\rho, M_\rho, \mathrm{per}_\rho, c_\rho \in (0,+\infty)$ verifying*

1. *$m_\rho \leq \rho_{|\mathrm{spt}(\rho)} \leq M_\rho$,*
2. *$\mathrm{spt}(\rho)$ has a $\mathcal{H}^{d-1}$-rectifiable boundary and $\mathcal{H}^{d-1}(\partial\mathrm{spt}(\rho)) \leq \mathrm{per}_\rho$,*
3. *$\forall \psi, \tilde{\psi} \in \mathcal{C}(\Omega), \quad c_\rho \mathbb{V}\mathrm{ar}_\rho(\tilde{\psi}^* - \psi^*) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi) - \langle \psi - \tilde{\psi} | (\nabla\psi^*)_\# \rho\rangle$.*

Then for any $\mu,\nu \in \mathcal{P}(\Omega)$ and any Kantorovich potential $\psi_{\rho\to\mu} \in \mathcal{C}(\Omega)$ in the optimal transport between ρ and μ, one has

$$\forall \mu,\nu, \quad \mathrm{W}_2^6(\mu,\nu) \lesssim \frac{1}{2}\mathrm{W}_2^2(\nu,\rho) - \frac{1}{2}\mathrm{W}_2^2(\mu,\rho) - \langle \frac{1}{2}\left\|\cdot\right\|^2 - \psi_{\rho\to\mu} | \nu - \mu\rangle,$$

where $\lesssim$ hides on the right-hand side the multiplicative constant

$$C_{d,m_\rho,M_\rho,\mathrm{per}_\rho} = C_d \frac{M_\rho^3}{m_\rho}\frac{\mathrm{per}_\rho^2}{c_\rho}R^4,$$

where C_d is a dimensional constant.

Remark 2.2 (*Linear* convexity vs. *displacement* convexity)**.** We emphasize on the fact that Proposition 2.3 gives a strong convexity estimate for $\frac{1}{2}\mathrm{W}_2^2(\rho,\cdot)$ with respect to the *linear* structure on $\mathcal{P}(\Omega)$, and not with respect to the metric structure of $(\mathcal{P}(\Omega),\mathrm{W}_2)$. Convexity of a functional with respect to this structure is referred to the notion of *displacement* convexity. Strong convexity of $\frac{1}{2}\mathrm{W}_2^2(\rho,\cdot)$ with respect to the metric structure of $(\mathcal{P}(\Omega),\mathrm{W}_2)$ is trivial to get in dimension $d=1$ because of the Hilbertian nature of W_2 in this context (see Section 1.2.2). However, this is limited to the unidimensional setting and $\frac{1}{2}\mathrm{W}_2^2(\rho,\cdot)$ is notoriously *not* displacement convex in dimension $d \geq 2$ (see for instance Section 7.3.3 of [33]).

Remark 2.3 (Exponent)**.** We note that the value 6 of the exponent on the left-hand side term of the estimate of Proposition 2.3 might not be optimal. However, 4 should be a lower-bound on the value of this exponent. This may be seen from the following example: in dimension

$d = 1$ and for $\varepsilon > 0$, set $\mu^\varepsilon = (\frac{1}{2} - \frac{\varepsilon}{2})(\delta_{-1} + \delta_1) + \varepsilon\delta_0$. Then we have $\mathrm{W}_2^2(\mu^0, \mu^\varepsilon) = \varepsilon$. For $\rho = \lambda_{|[-\frac{1}{2},\frac{1}{2}]}$, the following computation holds:

$$\frac{1}{2}\mathrm{W}_2^2(\mu^\varepsilon, \rho) - \frac{1}{2}\mathrm{W}_2^2(\mu^0, \rho) = \int_0^{\varepsilon/2} (|0-x|^2 - |1-x|^2)\mathrm{d}x = \frac{\varepsilon^2}{4} - \frac{\varepsilon}{2}.$$

Finally, one can choose a Kantorovich potential in the transport from ρ to μ^0 to be $\psi_{\rho\to\mu^0} = \iota_{[-1,1]}$ (i.e. valued 0 on $[-1,1]$ and $+\infty$ outside this segment), so that

$$\langle \frac{\|\cdot\|^2}{2} - \psi_{\rho\to\mu^0} | \mu^\varepsilon - \mu^0 \rangle = \langle \frac{\|\cdot\|^2}{2} | \mu^\varepsilon - \mu^0 \rangle = -\frac{\varepsilon}{2}.$$

Hence:

$$\frac{1}{2}\mathrm{W}_2^2(\mu^\varepsilon, \rho) - \frac{1}{2}\mathrm{W}_2^2(\mu^0, \rho) - \langle \frac{\|\cdot\|^2}{2} - \psi_{\rho\to\mu^0} | \mu^\varepsilon - \mu^0 \rangle = \frac{\varepsilon^2}{4} = \frac{1}{4}\mathrm{W}_2^4(\mu^0, \mu^\varepsilon).$$

Proof of Proposition 2.3. Let $\psi_{\rho\to\nu} \in \mathcal{C}(\Omega)$ be a Kantorovich potential in the optimal transport between ρ and ν. Then the conjugates $\psi^*_{\rho\to\mu}, \psi^*_{\rho\to\nu}$ are both convex Brenier potentials [10] in the optimal transport between the absolutely continuous source ρ and the targets μ, ν, in the sense that:

$$(\nabla\psi^*_{\rho\to\mu})_\#\rho = \mu \quad \text{and} \quad (\nabla\psi^*_{\rho\to\nu})_\#\rho = \nu.$$

Therefore, the coupling $(\nabla\psi^*_{\rho\to\mu}, \nabla\psi^*_{\rho\to\nu})_\#\rho$ is an admissible transport plan between μ and ν and as such:

$$\mathrm{W}_2^2(\mu, \nu) \leq \left\| \nabla\psi^*_{\rho\to\mu} - \nabla\psi^*_{\rho\to\nu} \right\|^2_{\mathrm{L}^2(\rho;\mathbb{R}^d)}. \tag{4}$$

We now quote a Gagliardo-Nirenberg type inequality, extracted from Proposition 4.1 in [18], that ensures that for any compact domain K of $\mathbb{R}^d$ with $\mathcal{H}^{d-1}$-rectifiable boundary and $u, v : K \to \mathbb{R}$ two Lipschitz functions on K that are convex on any segment included in K, there exists a constant C_d depending only on d such that

$$\|\nabla u - \nabla v\|^6_{\mathrm{L}^2(K)} \leq C_d \mathcal{H}^{d-1}(\partial K)^2 (\|\nabla u\|_{\mathrm{L}^\infty(K)} + \|\nabla v\|_{\mathrm{L}^\infty(K)})^4 \|u - v\|^2_{\mathrm{L}^2(K)}.$$

We note from [18] that the exponents in this inequality are optimal. Using that the Brenier potentials $\psi^*_{\rho\to\mu}, \psi^*_{\rho\to\nu}$ are both convex and R-Lipschitz continuous and leveraging assumptions (1) and (2) made on ρ, we can apply this inequality to get that for any constant $c \in \mathbb{R}$:

$$\frac{1}{M_\rho^3} \left\| \nabla\psi^*_{\rho\to\mu} - \nabla\psi^*_{\rho\to\nu} \right\|^6_{\mathrm{L}^2(\rho;\mathbb{R}^d)} \leq C_d (\mathrm{per}_\rho)^2 (2R)^4 \frac{1}{m_\rho} \left\| \psi^*_{\rho\to\mu} - \psi^*_{\rho\to\nu} - c \right\|^2_{\mathrm{L}^2(\rho)}.$$

Minimizing over $c \in \mathbb{R}$ in the last inequality yields:

$$\left\| \nabla\psi^*_{\rho\to\mu} - \nabla\psi^*_{\rho\to\nu} \right\|^6_{\mathrm{L}^2(\rho;\mathbb{R}^d)} \lesssim \mathbb{V}\mathrm{ar}_\rho(\psi^*_{\rho\to\mu} - \psi^*_{\rho\to\nu}). \tag{5}$$

But assumption (3) on ρ ensures:

$$c_\rho \mathbb{V}\mathrm{ar}_\rho(\psi^*_{\rho\to\mu} - \psi^*_{\rho\to\nu}) \leq \mathcal{K}_\rho(\psi_{\rho\to\mu}) - \mathcal{K}_\rho(\psi_{\rho\to\nu}) + \langle \psi_{\rho\to\mu} - \psi_{\rho\to\nu} | \nu \rangle. \tag{6}$$

Finally, notice that by Kantorovich's duality formula (2) and by definition of $\psi_{\rho\to\mu}, \psi_{\rho\to\mu}$ as Kantorovich potentials, one has:

$$\frac{1}{2}\mathrm{W}_2^2(\rho, \mu) = \langle \frac{1}{2}\|\cdot\|^2 | \rho \rangle + \langle \frac{1}{2}\|\cdot\|^2 | \mu \rangle - \mathcal{K}_\rho(\psi_{\rho\to\mu}) - \langle \psi_{\rho\to\mu} | \mu \rangle,$$
$$\frac{1}{2}\mathrm{W}_2^2(\rho, \nu) = \langle \frac{1}{2}\|\cdot\|^2 | \rho \rangle + \langle \frac{1}{2}\|\cdot\|^2 | \nu \rangle - \mathcal{K}_\rho(\psi_{\rho\to\nu}) - \langle \psi_{\rho\to\nu} | \nu \rangle.$$

This yields:

$$\mathcal{K}_\rho(\psi_{\rho\to\mu}) - \mathcal{K}_\rho(\psi_{\rho\to\nu}) + \langle \psi_{\rho\to\mu} - \psi_{\rho\to\nu} | \nu \rangle = \frac{1}{2}\mathrm{W}_2^2(\nu, \rho) - \frac{1}{2}\mathrm{W}_2^2(\mu, \rho) - \langle \frac{1}{2}\|\cdot\|^2 - \psi_{\rho\to\mu} | \nu - \mu \rangle. \tag{7}$$

The conclusion follows after combining (4), (5), (6) and (7) together. □

3. Convergence of empirical barycenters in the Wasserstein space

This section is devoted to the proof of Theorem 1.7. This proof relies on a classical symmetrization technique used in the study of empirical processes [39], already employed in the context of strongly-convex empirical risk minimization (see e.g. the proof of Theorem 2 in [35]).

Proof of Theorem 1.7. Applying the strong convexity estimate of Theorem 2.1 to $F_{\mathbb{P}}$ at the minimizer $\mu = \mu_{\mathbb{P}}$ and with $\nu = \mu_{\mathbb{P}_m}$, we have the bound

$$\begin{aligned} \mathrm{W}_2^6(\mu_{\mathbb{P}}, \mu_{\mathbb{P}_m}) &\lesssim F_{\mathbb{P}}(\mu_{\mathbb{P}_m}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}) \\ &\leq F_{\mathbb{P}}(\mu_{\mathbb{P}_m}) - F_{\mathbb{P}_m}(\mu_{\mathbb{P}_m}) + F_{\mathbb{P}_m}(\mu_{\mathbb{P}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}) \end{aligned} \tag{8}$$

We now proceed to the control of the expectation with respect to $(\rho_i)_{1\leq i\leq m} \sim \mathbb{P}^{\otimes m}$ of the above two differences.

Control of $\mathbb{E}(F_{\mathbb{P}}(\mu_{\mathbb{P}_m}) - F_{\mathbb{P}_m}(\mu_{\mathbb{P}_m}))$. Notice that

$$F_{\mathbb{P}}(\mu_{\mathbb{P}_m}) - F_{\mathbb{P}_m}(\mu_{\mathbb{P}_m}) = \frac{1}{2}\int_{\mathcal{P}(\Omega)} \mathrm{W}_2^2(\rho, \mu_{\mathbb{P}_m})\mathrm{d}\mathbb{P}(\rho) - \frac{1}{2m}\sum_{i=1}^m \mathrm{W}_2^2(\rho_i, \mu_{\mathbb{P}_m}).$$

In order to control the expectation of this difference, we introduce another i.i.d. m-sample of $\mathbb{P}$: $(\rho_i')_{1\leq i\leq m} \sim \mathbb{P}^{\otimes m}$. One can then notice that

$$\begin{aligned} \mathbb{E}\frac{1}{2}\int_{\mathcal{P}(\Omega)} \mathrm{W}_2^2(\rho, \mu_{\mathbb{P}_m})\mathrm{d}\mathbb{P}(\rho) &= \mathbb{E}_{(\rho_i)_i\sim\mathbb{P}^{\otimes m}}\mathbb{E}_{\rho\sim\mathbb{P}}\frac{1}{2}\mathrm{W}_2^2(\rho, \mu_{\mathbb{P}_m}) \\ &= \mathbb{E}_{(\rho_i)_i\sim\mathbb{P}^{\otimes mN}}\mathbb{E}_{(\rho_i')_i\sim\mathbb{P}^{\otimes m}}\frac{1}{2m}\sum_{i=1}^m \mathrm{W}_2^2(\rho_i', \mu_{\mathbb{P}_m}) \\ &= \mathbb{E}\frac{1}{2m}\sum_{i=1}^m \mathrm{W}_2^2(\rho_i', \mu_{\mathbb{P}_m}), \end{aligned} \tag{9}$$

where the last expectation is against all the random variables $(\rho_i)_{1\leq i\leq m} \sim \mathbb{P}^{\otimes m}$ and $(\rho_i')_{1\leq i\leq m} \sim \mathbb{P}^{\otimes m}$. Now for any $i \in \{1, \dots, m\}$, introduce the empirical measure

$$\mathbb{P}_m^{(i)} = \frac{1}{m}\sum_{j=1, j\neq i}^m \delta_{\rho_j} + \frac{1}{m}\delta_{\rho_i'}.$$

Then for any $i \in \{1, \dots, m\}$, taking again the expectation against all the random variables $(\rho_i)_{1\leq i\leq m} \sim \mathbb{P}^{\otimes m}$ and $(\rho_i')_{1\leq i\leq m} \sim \mathbb{P}^{\otimes m}$ one has

$$\mathbb{E}\mathrm{W}_2^2(\rho_i, \mu_{\mathbb{P}_m}) = \mathbb{E}\mathrm{W}_2^2(\rho_i', \mu_{\mathbb{P}_m^{(i)}}).$$

This ensures the equality

$$\mathbb{E}\frac{1}{2m}\sum_{i=1}^m \mathrm{W}_2^2(\rho_i, \mu_{\mathbb{P}_m}) = \mathbb{E}\frac{1}{2m}\sum_{i=1}^m \mathrm{W}_2^2(\rho_i', \mu_{\mathbb{P}_m^{(i)}}). \tag{10}$$

From equations (9) and (10), the expectation of the first difference appearing in (8) reads:

$$\mathbb{E}(F_{\mathbb{P}}(\mu_{\mathbb{P}_m}) - F_{\mathbb{P}_m}(\mu_{\mathbb{P}_m})) = \frac{1}{2m}\sum_{i=1}^m \mathbb{E}(\mathrm{W}_2^2(\rho_i', \mu_{\mathbb{P}_m}) - \mathrm{W}_2^2(\rho_i', \mu_{\mathbb{P}_m^{(i)}})).$$

Using that $\Omega = B(0,R)$ is bounded and the triangle inequality, we have the bound

$$\frac{1}{2m}\sum_{i=1}^{m}\mathbb{E}(\mathrm{W}_2^2(\rho_i',\mu_{\mathbb{P}_m}) - \mathrm{W}_2^2(\rho_i',\mu_{\mathbb{P}_m^{(i)}})) \leq \frac{(2R+2R)}{2m}\sum_{i=1}^{m}\mathbb{E}\mathrm{W}_2(\mu_{\mathbb{P}_m},\mu_{\mathbb{P}_m^{(i)}}) \\ = 2R\mathbb{E}\mathrm{W}_2(\mu_{\mathbb{P}_m},\mu_{\mathbb{P}_m^{(i)}}).$$

Using that $\mathbb{P}$ satisfies Assumption 1.3 with $\alpha_{\mathbb{P}} = 1$, we have that $\mathbb{P}_m$ (or $\mathbb{P}_m^{(i)}$) almost surely satisfies Assumption 1.3 with $\alpha_{\mathbb{P}_m} = 1$ (and with the same other constants as $\mathbb{P}$ in this assumption). Thus Theorem 1.5 ensures almost surely the bound:

$$\mathrm{W}_2(\mu_{\mathbb{P}_m},\mu_{\mathbb{P}_m^{(i)}}) \lesssim \left\|\mathbb{P}_m - \mathbb{P}_m^{(i)}\right\|_{\mathrm{TV}}^{1/5} \lesssim \frac{1}{m^{1/5}}.$$

We thus have the following bound on the expectation of the first difference appearing in (8):

$$\mathbb{E}(F_{\mathbb{P}}(\mu_{\mathbb{P}_m}) - F_{\mathbb{P}_m}(\mu_{\mathbb{P}_m})) \lesssim \frac{1}{m^{1/5}}. \tag{11}$$

Control of $\mathbb{E}(F_{\mathbb{P}_m}(\mu_{\mathbb{P}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}))$. Notice that

$$F_{\mathbb{P}_m}(\mu_{\mathbb{P}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}) = \frac{1}{2m}\sum_{i=1}^{m}\mathrm{W}_2^2(\rho_i,\mu_{\mathbb{P}}) - \frac{1}{2}\int_{\mathcal{P}(\Omega)}\mathrm{W}_2^2(\rho,\mu_{\mathbb{P}})\mathrm{d}\mathbb{P}(\rho)$$

Bounding the expectation of this second difference term is much more straightforward. For any $i \in \{1,\dots,m\}$, denote $X_i = \frac{1}{2}\mathrm{W}_2^2(\rho_i,\mu_{\mathbb{P}})$ the scalar random variable built from the random sample $\rho_i \sim \mathbb{P}$. Denote $\mathbb{E}X$ the expectation of this random variable (independent of i). Using this notation we can write the expectation of the second difference term of (8) as follows:

$$\frac{1}{2m}\sum_{i=1}^{m}\mathrm{W}_2^2(\rho_i,\mu_{\mathbb{P}}) - \frac{1}{2}\int_{\mathcal{P}(\Omega)}\mathrm{W}_2^2(\rho,\mu_{\mathbb{P}})\mathrm{d}\mathbb{P}(\rho) = \frac{1}{m}\sum_{i=1}^{m}X_i - \mathbb{E}X.$$

Using Jensen's inequality, we thus have:

$$\mathbb{E}(F_{\mathbb{P}_m}(\mu_{\mathbb{P}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}})) = \mathbb{E}\left(\frac{1}{m}\sum_{i=1}^{m}X_i - \mathbb{E}X\right) \leq \left(\mathbb{V}\mathrm{ar}\left(\frac{1}{m}\sum_{i=1}^{m}X_i\right)\right)^{1/2} \lesssim \frac{1}{m^{1/2}}. \tag{12}$$

Conclusion. Injecting the bounds (11) and (12) in the expectation of (8) thus yields

$$\mathbb{E}\mathrm{W}_2^6(\mu_{\mathbb{P}},\mu_{\mathbb{P}_m}) \lesssim \frac{1}{m^{1/5}} + \frac{1}{m^{1/2}} \lesssim \frac{1}{m^{1/5}}.$$

Jensen's inequality used in the above bound finally yields the statement. □

Acknowledgement. The authors acknowledge the support of the Lagrange Mathematics and Computing Research Center and of the ANR (MAGA, ANR-16-CE40-0014). We thank Blanche Buet for interesting discussions related to this work.

Appendix A. Dual formulation for the Wasserstein barycenter problem

Proof of Proposition 1.1. Instead of showing directly the formulation of Proposition 1.1, we will rather show

$$\min_{\mu\in\mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu) = \max\left\{\int_{\mathcal{P}(\Omega)}\langle\phi_\rho^c|\rho\rangle\mathrm{d}\mathbb{P}(\rho) \mid (\phi_\rho)_\rho \in \mathrm{L}^\infty(\mathbb{P};W^{1,\infty}(\Omega)),\quad \int_{\mathcal{P}(\Omega)}\phi_\rho(\cdot)\mathrm{d}\mathbb{P}(\rho) = 0\right\},$$

where for any $\rho \in \mathcal{P}(\Omega)$, ϕ_ρ^c denotes the following c-transform of ϕ_ρ: $\phi_\rho^c(\cdot) = \inf_{y\in\Omega}\frac{1}{2}\|\cdot - y\|^2 - \phi_\rho(y)$. Such a formulation entails the result of Proposition 1.1 by the change of variable $(\psi_\rho)_\rho = \frac{\|\cdot\|^2}{2} - (\phi_\rho)_\rho \in \mathrm{L}^\infty(\mathbb{P};W^{1,\infty}(\Omega))$.

Duality. Let's first show that the value of $\min_{\mu\in\mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu)$ is equal to the value of the following supremum

$$(\mathrm{D})_{\mathbb{P}}{}' := \sup\left\{ \int_{\mathcal{P}(\Omega)} \langle \phi_\rho^c | \rho\rangle \mathrm{d}\mathbb{P}(\rho) \mid (\phi_\rho)_\rho \in \mathrm{L}^1(\mathbb{P};\mathcal{C}(\Omega)), \quad \int_{\mathcal{P}(\Omega)} \phi_\rho(\cdot)\mathrm{d}\mathbb{P}(\rho) = 0 \right\},$$

where $\mathrm{L}^1(\mathbb{P};\mathcal{C}(\Omega))$ denotes the set of $\mathbb{P}$-measurable and Bochner integrable mappings from $\mathcal{P}(\Omega)$ to the space $(\mathcal{C}(\Omega), \|\cdot\|_\infty)$ of continuous function from Ω to $\mathbb{R}$ equipped with the supremum norm. Introduce the functional $H : \mathcal{C}(\Omega) \to \mathbb{R}$ defined for all $\varphi \in \mathcal{C}(\Omega)$ by

$$H(\varphi) = \inf\left\{ -\int_{\mathcal{P}(\Omega)} \langle \phi_\rho^c | \rho\rangle \mathrm{d}\mathbb{P}(\rho) \mid (\phi_\rho)_\rho \in \mathrm{L}^1(\mathbb{P};\mathcal{C}(\Omega)), \quad \int_{\mathcal{P}(\Omega)} \phi_\rho(\cdot)\mathrm{d}\mathbb{P}(\rho) = \varphi(\cdot) \right\}.$$

Notice then that $(\mathrm{D})_{\mathbb{P}}{}' = -H(0)$. On the other hand, notice that H has the following convex conjugate: for $\mu \in \mathcal{P}(\Omega)$,

$$\begin{aligned}
H^*(\mu) &= \sup\left\{\langle\varphi|\mu\rangle - H(\varphi) \mid \varphi \in \mathcal{C}(\Omega)\right\} \\
&= \sup\left\{ \langle\varphi|\mu\rangle + \int_{\mathcal{P}(\Omega)} \langle \phi_\rho^c|\rho\rangle \mathrm{d}\mathbb{P}(\rho) \mid \varphi \in \mathcal{C}(\Omega), (\phi_\rho)_\rho \in \mathrm{L}^1(\mathbb{P};\mathcal{C}(\Omega)), \int_{\mathcal{P}(\Omega)} \phi_\rho(\cdot)\mathrm{d}\mathbb{P}(\rho) = \varphi(\cdot) \right\} \\
&= \sup\left\{ \int_{\mathcal{P}(\Omega)} \langle\phi_\rho|\mu\rangle \mathrm{d}\mathbb{P}(\rho) + \int_{\mathcal{P}(\Omega)} \langle\phi_\rho^c|\rho\rangle \mathrm{d}\mathbb{P}(\rho), \quad (\phi_\rho)_\rho \in \mathrm{L}^1(\mathbb{P};\mathcal{C}(\Omega)) \right\} \\
&= \int_{\mathcal{P}(\Omega)} \left(\sup_{\phi_\rho\in\mathcal{C}(\Omega)} \langle\phi_\rho|\mu\rangle + \langle\phi_\rho^c|\rho\rangle \right) \mathrm{d}\mathbb{P}(\rho) \\
&= \int_{\mathcal{P}(\Omega)} \frac{1}{2}\mathrm{W}_2^2(\mu,\rho)\mathrm{d}\mathbb{P}(\rho),
\end{aligned}$$

where we used the Kantorovich duality formula (see for instance [42]) to get to the last line. We thus have

$$\min_{\mu\in\mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu) = \inf_{\mu\in\mathcal{P}(\Omega)} H^*(\mu) = -H^{**}(0).$$

Therefore, showing that $(\mathrm{D})_{\mathbb{P}}{}' = \min_{\mu\in\mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu)$ corresponds to show that $H(0) = H^{**}(0)$. Since H is convex (by concavity of the c-transform operation), this will follow from the continuity of H at 0 for the supremum-norm over $\mathcal{C}(\Omega)$ (Proposition 4.1 of [20]). For this, we can first notice that H never takes the value $-\infty$: for any $\varphi \in \mathcal{C}(\Omega)$ and $(\phi_\rho)_\rho \in \mathrm{L}^1(\mathbb{P};\mathcal{C}(\Omega))$ such that $\int_{\mathcal{P}(\Omega)} \phi_\rho(\cdot)\mathrm{d}\mathbb{P}(\rho) = \varphi(\cdot)$, one has

$$\forall \rho \in \mathcal{P}(\Omega), \quad -\phi_\rho^c(x) = \sup_{y\in\mathbb{R}^d} \phi_\rho(y) - \frac{1}{2}\|x-y\|^2 \geq \phi_\rho(0) - \frac{1}{2}\|x\|^2.$$

If follows that

$$H(\varphi) \geq \varphi(0) - \int_{\mathcal{P}(\Omega)} \frac{M_2(\rho)}{2}\mathrm{d}\mathbb{P}(\rho) > -\infty.$$

On the other hand, notice that H is bounded from above in a neighborhood of 0 in $\mathcal{C}(\Omega)$: for any $\varphi \in \mathcal{C}(\Omega)$ such that $\|\varphi\|_\infty \leq 1$, one has $-\varphi^c(x) \leq 1$ for any $x \in \mathbb{R}^d$ so that

$$H(\varphi) \leq -\int_{\mathcal{P}(\Omega)} \langle(\varphi)^c|\rho\rangle \mathrm{d}\mathbb{P}(\rho) \leq 1.$$

A standard convex analysis result (Proposition 2.5 in [20]) then ensures that H is continuous at 0, so that $H(0) = H^{**}(0)$ and $(\mathrm{D})_{\mathbb{P}}{}' = \min_{\mu\in\mathcal{P}(\Omega)} F_{\mathbb{P}}(\mu)$.

Restriction to $\mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$. We show here that we can run the supremum $(\mathrm{D})_{\mathbb{P}}{}'$ only over $\mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ instead of $\mathrm{L}^1(\mathbb{P}; \mathcal{C}(\Omega))$, that is

$$(\mathrm{D})_{\mathbb{P}}{}' = \sup\left\{ \int_{\mathcal{P}(\Omega)} \langle \phi_\rho^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho) \mid (\phi_\rho)_\rho \in \mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega)), \quad \int_{\mathcal{P}(\Omega)} \phi_\rho(\cdot) \mathrm{d}\mathbb{P}(\rho) = 0 \right\}.$$

Let $(\phi_\rho)_\rho \in \mathrm{L}^1(\mathbb{P}; \mathcal{C}(\Omega))$ be an admissible solution to $(\mathrm{D})_{\mathbb{P}}{}'$, i.e. $(\phi_\rho)_\rho$ satisfies

$$\int_{\mathcal{P}(\Omega)} \phi_\rho(\cdot) \mathrm{d}\mathbb{P}(\rho) = 0. \tag{13}$$

Then we can build from $(\phi_\rho)_\rho$ another admissible solution $(\tilde{\phi}_\rho)_\rho$ that belongs to $\mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ and that performs better at $(\mathrm{D})_{\mathbb{P}}{}'$, i.e. that verifies

$$\int_{\mathcal{P}(\Omega)} \langle \tilde{\phi}_\rho^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho) \geq \int_{\mathcal{P}(\Omega)} \langle \phi_\rho^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho). \tag{14}$$

Indeed, introduce $(\hat{\phi}_\rho)_\rho := (\phi_\rho^{cc})_\rho$. Then for all $\rho \in \mathcal{P}(\Omega)$, $\hat{\phi}_\rho = \phi_\rho^{cc}$ is obviously $2R$-Lipschitz (as a c-transform) and satisfies $\hat{\phi}_\rho^c = \phi_\rho^c$ and $\hat{\phi}_\rho \geq \phi_\rho$ (as a double c-transform). Using then (13), one has that

$$\alpha(\cdot) := \int_{\mathcal{P}(\Omega)} \hat{\phi}_\rho(\cdot) \mathrm{d}\mathbb{P}(\rho) \geq 0,$$

where α is also $2R$-Lipschitz. Now denoting $\tilde{\phi}_\rho = \hat{\phi}_\rho - \alpha$ for all $\rho \in \mathcal{P}(\Omega)$, the mapping $(\tilde{\phi}_\rho)_\rho \in \mathrm{L}^1(\mathbb{P}; \mathcal{C}(\Omega))$ is admissible to $(\mathrm{D})_{\mathbb{P}}{}'$ by construction and satisfies $\tilde{\phi}_\rho \leq \hat{\phi}_\rho$ for all $\rho \in \mathcal{P}(\Omega)$, so that $\tilde{\phi}_\rho^c \geq \hat{\phi}_\rho^c = \phi_\rho^c$ (using that the c-transform is order-reversing). For each $\rho \in \mathcal{P}(\Omega)$, up to subtracting $\tilde{\phi}_\rho(0)$ to $\tilde{\phi}_\rho$ (this operation leaves $(\tilde{\phi}_\rho)_\rho$ admissible to $(\mathrm{D})_{\mathbb{P}}{}'$ and does not change its value), one can assume that $\tilde{\phi}_\rho(0) = 0$. Noticing that $\tilde{\phi}_\rho$ is $4R$-Lipschitz by construction, we have the bound $\left\|\tilde{\phi}_\rho\right\|_{W^{1,\infty}(\Omega)} \leq 4R(1+R)$. We thus have built an admissible $(\tilde{\phi}_\rho)_\rho \in \mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ from an admissible $(\phi_\rho)_\rho \in \mathrm{L}^1(\mathbb{P}; \mathcal{C}(\Omega))$ that satisfies (14), which shows that we can run the supremum $(\mathrm{D})_{\mathbb{P}}{}'$ only over $\mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ instead of $\mathrm{L}^1(\mathbb{P}; \mathcal{C}(\Omega))$

Existence of a maximizer. There now remains to show that the supremum in $(\mathrm{D})_{\mathbb{P}}{}'$ can be replaced by a maximum. Let $\left((\phi_\rho^n)_\rho\right)_{n\geq 0}$ be a maximizing sequence to $(\mathrm{D})_{\mathbb{P}}{}'$, and assume from what precedes that this sequence belongs to $\mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ and satisfies for all $n \geq 0$ and $\rho \in \mathcal{P}(\Omega)$, $\left\|\phi_\rho^n\right\|_{W^{1,\infty}(\Omega)} \leq 4R(1+R)$. Further assume that this sequence verifies for all $n \geq 1$,

$$\int_{\mathcal{P}(\Omega)} \langle (\phi_\rho^n)^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho) \geq (\mathrm{D})_{\mathbb{P}}{}' - \frac{1}{n}. \tag{15}$$

For any $n \geq 0$, the mapping $(\rho, x) \mapsto \phi_\rho^n(x)$ is bounded in $\mathrm{L}^2(\mathbb{P} \otimes \lambda)$ where λ denotes the Lebesgue measure over Ω. Therefore, by Banach-Alaoglu theorem, the sequence $\left((\phi_\rho^n)_\rho\right)_{n\geq 0}$ (seen as a sequence in $\mathrm{L}^2(\mathbb{P} \otimes \lambda)$) admits a weakly converging subsequence in $\mathrm{L}^2(\mathbb{P} \otimes \lambda)$, that we do not relabel and for which we denote $(\phi_\rho^\infty)_\rho$ the weak limit in $\mathrm{L}^2(\mathbb{P} \otimes \lambda)$. Using now Mazur's lemma, we know that there exists a sequence of integers $(N_n)_{n\geq 0}$ and coefficients $((\lambda_{n,k})_{n\leq k\leq N_n})_{n\geq 0} \geq 0$ satisfying for all $n \geq 0$, $\sum_{k=n}^{N_n} \lambda_{n,k} = 1$ such that the sequence $\left((\bar{\phi}_\rho^n)_\rho\right)_{n\geq 0}$ defined for all $n \geq 0$ and $\rho \in \mathcal{P}(\Omega)$ by $\bar{\phi}_\rho^n := \sum_{k=n}^{N_n} \lambda_{n,k} \phi_\rho^k$ converges strongly to $(\phi_\rho^\infty)_\rho$ in $\mathrm{L}^2(\mathbb{P} \otimes \lambda)$. By concavity of the c-transform operation and equation (15), we then

have the bound

$$\begin{aligned}\int_{\mathcal{P}(\Omega)} \langle (\bar{\phi}_\rho^n)^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho) &\geq \sum_{k=n}^{N_n} \lambda_{n,k} \int_{\mathcal{P}(\Omega)} \langle (\phi_\rho^k)^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho) \\ &\geq \sum_{k=n}^{N_n} \lambda_{n,k} \left((\mathrm{D})_{\mathbb{P}}{}' - \frac{1}{k} \right) \\ &\geq (\mathrm{D})_{\mathbb{P}}{}' - \frac{1}{n}. \end{aligned} \tag{16}$$

The sequence $\left((\bar{\phi}_\rho^n)_\rho\right)_{n\geq 0}$ is therefore also a maximizing sequence of $(\mathrm{D})_{\mathbb{P}}{}'$ and it also satisfies for any $n \geq 0$ and $\rho \in \overline{\mathcal{P}}(\Omega)$ the bound

$$\left\| \bar{\phi}_\rho^n \right\|_{W^{1,\infty}(\Omega)} \leq 4R(1+R). \tag{17}$$

Since the sequence $\left((\bar{\phi}_\rho^n)_\rho\right)_{n\geq 0}$ strongly converges to $(\phi_\rho^\infty)_\rho$ in $\mathrm{L}^2(\mathbb{P}\otimes\lambda)$, one can extract a subsequence (that we do not relabel) such that for $\mathbb{P}$-almost-every $\rho \in \mathcal{P}(\Omega)$, the sequence $(\bar{\phi}_\rho^n)_{n\geq 0}$ converges to ϕ_ρ^∞ in $\mathrm{L}^2(\lambda)$. Using (17) and Arzelà-Ascoli theorem, we deduce that for $\mathbb{P}$-almost-every $\rho \in \mathcal{P}(\Omega)$, the sequence $(\bar{\phi}_\rho^n)_{n\geq 0}$ converges uniformly to ϕ_ρ^∞ in $\mathcal{C}(\Omega)$ and that

$$\left\| \phi_\rho^\infty \right\|_{W^{1,\infty}(\Omega)} \leq 4R(1+R).$$

In particular, $(\phi_\rho^\infty)_\rho$ belongs to $\mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ and we have the limit

$$0 = \int_{\mathcal{P}(\Omega)} \bar{\phi}_\rho^n(\cdot) \mathrm{d}\mathbb{P}(\rho) \xrightarrow[n\to\infty]{} \int_{\mathcal{P}(\Omega)} \phi_\rho^\infty(\cdot) \mathrm{d}\mathbb{P}(\rho),$$

so that $(\phi_\rho^\infty)_\rho$ is admissible to $(\mathrm{D})_{\mathbb{P}}{}'$. Eventually, for $\mathbb{P}$-almost-every $\rho \in \mathcal{P}(\Omega)$, we have the limit

$$\langle (\bar{\phi}_\rho^n)^c | \rho \rangle \xrightarrow[n\to\infty]{} \langle (\phi_\rho^\infty)^c | \rho \rangle, \tag{18}$$

so that by Lebesgue's dominated convergence theorem and the bound (16),

$$\int_{\mathcal{P}(\Omega)} \langle (\phi_\rho^\infty)^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho) = \lim_{n\to+\infty} \int_{\mathcal{P}(\Omega)} \langle (\bar{\phi}_\rho^n)^c | \rho \rangle \mathrm{d}\mathbb{P}(\rho) = (\mathrm{D})_{\mathbb{P}}{}',$$

which proves that $(\phi_\rho^\infty)_\rho \in \mathrm{L}^\infty(\mathbb{P}; W^{1,\infty}(\Omega))$ is a maximizer for $(\mathrm{D})_{\mathbb{P}}{}'$. □

Appendix B. Strong-convexity of $\mathcal{K}_\rho$ for measures with non-convex support

This section gathers occurrences of measures ρ where the strong convexity estimate (4) of Assumption 1.3 is verified.

B.1. **Measures with convex support.** This result is mostly extracted from [18].

Proposition B.1. *Let $\rho \in \mathcal{P}_{a.c.}(\Omega)$. Assume that $\mathrm{spt}(\rho)$ is convex and that there exists $m_\rho, M_\rho \in (0,+\infty)$ such that $m_\rho \leq \rho \leq M_\rho$ on $\mathrm{spt}(\rho)$. Let $\psi, \tilde{\psi} \in \mathcal{C}(\Omega)$. Then*

$$\langle \psi - \tilde{\psi} | (\nabla \psi^*)_\# \rho \rangle + C_{d,R,m_\rho,M_\rho} \mathbb{V}\mathrm{ar}_\rho(\tilde{\psi}^* - \psi^*) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi),$$

where $C_{d,R,m_\rho,M_\rho} = \left(e(d+1)2^{d+1} R \mathrm{diam}(\mathrm{spt}(\rho)) \left(\frac{M_\rho}{m_\rho}\right)^2 \right)^{-1}$.

Proof. We only present here a formal sketch of the proof, which heavily relies on computations done in Section 2 of [18]. Assuming that ψ and $\tilde{\psi}$ are smooth enough (see Proposition 2.4 of [18]) and introducing for $t \in [0,1], \psi^t = (1-t)\psi + t\tilde{\psi}$, Proposition 2.2 of [18] allows to differentiate $\mathcal{K}_\rho(\psi^t)$ with respect to t and to obtain:

$$\begin{aligned}\mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\rho(\psi^t)\Big|_{t=0} + \int_0^1\int_0^s \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}_\rho(\psi^t)\mathrm{d}t\mathrm{d}s \\ &= \langle \psi - \tilde{\psi} | (\nabla\psi^*)_\# \rho\rangle + \int_0^1\int_0^s\int_\Omega \langle \nabla v(\nabla(\psi^t)^*) | \mathrm{D}^2(\psi^t)^* \cdot \nabla v(\nabla(\psi^t)^*)\rangle \mathrm{d}\rho\mathrm{d}t\mathrm{d}s, \quad (19)\end{aligned}$$

were $v = \tilde{\psi} - \psi$. Reasoning as in the proof of Proposition 2.4 of [18], the Brascamp-Lieb concentration inequality [9] and the log-concavity of the determinant seen as an application on the set of s.d.p. matrices ensure the following bound:

$$C_{R,m_\rho,M_\rho} \min(t,1-t)^d 2\mathbb{V}\mathrm{ar}_{\frac{1}{2}(\mu+\tilde{\mu})}(\tilde{\psi}-\psi) \leq \int_\Omega \langle \nabla v(\nabla(\psi^t)^*) | \mathrm{D}^2(\psi^t)^* \cdot \nabla v(\nabla(\psi^t)^*)\rangle \mathrm{d}\rho,$$

where $C_{R,m_\rho,M_\rho} = \left(eR\mathrm{diam}(\mathrm{spt}(\rho)) \left(\frac{M_\rho}{m_\rho}\right)^2 \right)^{-1}$, $\mu = (\nabla\psi^*)_\#\rho$ and $\tilde{\mu} = (\nabla\tilde{\psi})_\#\rho$. Back to (19), this leads to

$$\langle \psi - \tilde{\psi} | (\nabla\psi^*)_\#\rho\rangle + C_{d,R,m_\rho,M_\rho} 2\mathbb{V}\mathrm{ar}_{\frac{1}{2}(\mu+\tilde{\mu})}(\tilde{\psi}-\psi) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi),$$

where $C_{d,R,m_\rho,M_\rho} = \left(e(d+1)2^{d+1}R\mathrm{diam}(\mathrm{spt}(\rho)) \left(\frac{M_\rho}{m_\rho}\right)^2 \right)^{-1}$. We conclude using the convex analysis argument of Proposition 3.1 from [18], which directly ensures

$$\mathbb{V}\mathrm{ar}_\rho(\tilde{\psi}^* - \psi^*) \leq 2\mathbb{V}\mathrm{ar}_{\frac{1}{2}(\mu+\tilde{\mu})}(\tilde{\psi}-\psi).$$

We get the general case (without the smoothness assumptions on ψ and $\tilde{\psi}$) using approximation arguments presented in Proposition 2.5 and 2.7 of [18]. □

B.2. **Measures with connected union of convex sets as support.** We extend Proposition B.1 to the case of a source measure ρ with a possibly non-convex support. We will assume that $\mathrm{spt}(\rho)$ can be written as a connected finite union of convex sets.

Proposition B.2. *Let $\rho \in \mathcal{P}_{a.c.}(\Omega)$ such that there exists $m_\rho, M_\rho \in (0,+\infty)$ verifying $m_\rho \leq \rho \leq M_\rho$ on $\mathrm{spt}(\rho)$. Assume that $\mathrm{spt}(\rho)$ is connected and that there exists $N \geq 1$ convex sets $(C_i)_{1\leq i\leq N}$ in Ω such that $\mathrm{spt}(\rho) = \bigcup_{i=1}^N C_i$. Also assume that for any $i \neq j$ such that $C_i \cap C_j \neq \emptyset$, one has $\rho(C_i \cap C_j) > 0$. Then there exists a constant c_ρ depending on ρ such that for any $\psi, \tilde{\psi} \in \mathcal{C}(\Omega)$,*

$$\langle \psi - \tilde{\psi} | (\nabla\psi^*)_\#\rho\rangle + c_\rho \mathbb{V}\mathrm{ar}_\rho(\tilde{\psi}^* - \psi^*) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi).$$

Remark B.1 (Constant c_ρ and Poincaré-Wirtinger constant of ρ)**.** The constant c_ρ of Proposition B.2 is not made precise in the statement. A look at the proof of this proposition only allows to bound c_ρ in terms of the second smallest eigenvalue $\lambda_2(L)$ of a weighted graph Laplacian L, that is built from the graph whose vertices are the convex sets C_i and whose edge weights are the masses $\rho(C_i \cap C_j)$ that ρ grants to the intersection of the convex sets C_i and C_j. The constant c_ρ then reads:

$$c_\rho = \left(e(d+1)2^{d+1}R^2 \left(\frac{M_\rho}{m_\rho}\right)^2 \left(N^2 + \frac{2N^3}{\lambda_2(L)}\right)\right)^{-1}.$$

The quantity $\lambda_2(L)$ is not explicit, but it can be linked to the *weighted Cheeger constant* of ρ, defined by

$$h(\rho) = \inf_{A \subset \mathrm{spt}(\rho)} \frac{|\partial A|_\rho}{\min(\rho(A), \rho(\mathrm{spt}(\rho) \setminus A))},$$

where $|\partial A|_\rho = \int_{\partial A \cap \mathrm{int}(\mathrm{spt}(\rho))} \rho(x) \mathrm{d}\mathcal{H}^{d-1}(x)$ and where the infimum is taken over Lipschitz domains $A \subset \mathrm{int}(\mathrm{spt}(\rho))$ with boundary of finite $\mathcal{H}^{d-1}$-measure. Quoting [24] (Lemma 5.3), this constant can in turn be linked to the L^1 Poincaré-Wirtinger constant $C_{PW}(\rho)$ of ρ. Indeed, $h(\rho)$ is positive whenever ρ satisfies an L^1 Poincaré-Wirtinger inequality, i.e. whenever there exists a finite $C_{PW}(\rho) > 0$ such that for all smooth function f on Ω,

$$\|f - \mathbb{E}_\rho f\|_{\mathrm{L}^1(\rho)} \leq C_{PW}(\rho) \|\nabla f\|_{\mathrm{L}^1(\rho;\mathbb{R}^d)}.$$

The Poincaré-Wirtinger constant $C_{PW}(\rho)$ and the Cheeger constant $h(\rho)$ are then related by the inequality

$$h(\rho) \geq \frac{2}{C_{PW}(\rho)}.$$

Using ideas similar to the ones found in Section 5.2 of [24], the eigenvalue $\lambda_2(L)$ can be bounded in terms of the Cheeger constant of ρ, and thus in terms of $C_{PW}(\rho)$. We do not detail this comparison here but only report that c_ρ may be written

$$c_\rho = \left(e(d+1)2^{d+1} R^2 \left(\frac{M_\rho}{m_\rho} \right)^2 N \left(N + \frac{1}{2} \left(\frac{M_\rho s_{d-1} R^{d-1} N^2 C_{PW}(\rho)}{\varepsilon^2} \right)^3 \right) \right)^{-1},$$

where s_{d-1} denotes the surface area of the unit sphere in $\mathbb{R}^d$ and

$$\varepsilon = \min \left(\min_{i,j | C_i \cap C_j \neq \emptyset} \rho(C_i \cap C_j), \min_i \rho\left(C_i \setminus \cup_{j \neq i} C_j\right) \right) > 0.$$

Proof of Proposition B.2. Let's denote for now $f = \tilde{\psi}^* - \psi^*$. We will first exploit a discrete Laplacian over $\mathcal{X} = \mathrm{spt}(\rho)$ in order to upper bound $\mathbb{V}\mathrm{ar}_\rho(f)$ by a sum of variances of f w.r.t. probability measures supported over the convex sets $(C_i)_i$. We will then use Proposition B.1 to conclude.

For any $i \in \{1, \dots, N\}$, we denote $\rho_i = \frac{1}{\rho(C_i)} \rho_{|C_i}$ and $m_i = \int_{C_i} f \mathrm{d}\rho_i$. Then one has the following bound:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}_\rho(f) &= \frac{1}{2} \int_{\mathcal{X} \times \mathcal{X}} (f(x) - f(y))^2 \mathrm{d}\rho(x) \mathrm{d}\rho(y) \\
&\leq \frac{1}{2} \sum_{i,j} \int_{C_i \times C_j} (f(x) - f(y))^2 \mathrm{d}\rho(x) \mathrm{d}\rho(y) \\
&= \frac{1}{2} \sum_{i,j} \int_{C_i \times C_j} (f(x) - m_i + m_i - m_j + m_j - f(y))^2 \mathrm{d}\rho(x) \mathrm{d}\rho(y) \\
&= \left(\sum_i \rho(C_i) \right) \sum_i \int_{C_i} (f(x) - m_i)^2 \mathrm{d}\rho(x) + \frac{1}{2} \sum_{i,j} (m_i - m_j)^2 \rho(C_i) \rho(C_j) \\
&= \left(\sum_i \rho(C_i) \right) \sum_i \rho(C_i) \mathbb{V}\mathrm{ar}_{\rho_i}(f) + \frac{1}{2} \sum_{i,j} (m_i - m_j)^2 \rho(C_i) \rho(C_j). \qquad (20)
\end{aligned}$$

We now consider the graph $G = (\{C_i\}_{1 \leq i \leq N}, \{w_{ij}\}_{1 \leq i,j \leq N})$ with vertices $\{C_i\}_{1 \leq i \leq N}$ and weighted edges $\{w_{ij}\}_{1 \leq i,j \leq N}$ defined by

$$\forall i, j \in \{1, \dots, N\}, \quad w_{ij} = \rho(C_i \cap C_j).$$

By construction, this graph has a single connected component. We introduce the weighted Laplacian matrix $L \in \mathbb{R}^{N\times N}$ of G as follows:

$$\forall i,j \in \{1,\dots,N\}, \quad L_{ij} = \begin{cases} \sum_k w_{ik} & \text{if } i=j, \\ -w_{ij} & \text{else.} \end{cases}$$

Then L is a symmetric and positive semi-definite matrix. Its null space is made of constant vectors and we denote $\lambda_2(L)$ its second smallest eigenvalue, which is non-zero. Denoting $m = (m_i)_{1\le i\le N} \in \mathbb{R}^N$, we introduce $\bar{m} = \left(\frac{1}{N}\sum_i m_i\right)\mathbb{1}_N \in \mathbb{R}^N$ the constant vector whose coordinates equal the mean of m (we use $\mathbb{1}_N = (1)_{1\le i\le N} \in \mathbb{R}^N$). Notice that $m-\bar{m}$ is in the orthogonal to the null space of L, ensuring the following bound:

$$\begin{aligned}
\frac{1}{2}\sum_{i,j}(m_i-m_j)^2\rho(C_i)\rho(C_j) &\le N^2\frac{1}{2}\sum_{i,j}(m_i-m_j)^2\frac{1}{N^2} \\
&= N\left\|m-\bar{m}\right\|^2 \\
&\le \frac{N}{\lambda_2(L)}\langle m-\bar{m}|L\left(m-\bar{m}\right)\rangle \\
&= \frac{N}{\lambda_2(L)}\sum_{i,j} w_{ij}(m_i^2-m_im_j) \\
&= \frac{N}{\lambda_2(L)}\sum_{i,j}\frac{w_{ij}}{2}(m_i-m_j)^2. \qquad (21)
\end{aligned}$$

But for any i,j such that $w_{ij}>0$, denoting $m_{i\cap j} = \frac{1}{\rho(C_i\cap C_j)}\int_{C_i\cap C_j} f\mathrm{d}\rho$, one has

$$\frac{1}{2}(m_i-m_j)^2 \le (m_{i\cap j}-m_i)^2 + (m_{i\cap j}-m_j)^2.$$

And for such i,j,

$$\begin{aligned}
(m_{i\cap j}-m_i)^2 &= \left(\frac{1}{\rho(C_i\cap C_j)}\int_{C_i\cap C_j}(f-m_i)\mathrm{d}\rho\right)^2 \\
&\le \frac{1}{\rho(C_i\cap C_j)}\int_{C_i}(f-m_i)^2\mathrm{d}\rho \\
&= \frac{\rho(C_i)}{w_{ij}}\mathbb{V}\mathrm{ar}_{\rho_i}(f),
\end{aligned}$$

where we used Jensen's inequality and the fact that $C_i\cap C_j \subset C_i$. A similar bound can be shown for $(m_{i\cap j}-m_j)^2$, and plugging these into (21) yields

$$\begin{aligned}
\frac{1}{2}\sum_{i,j}(m_i-m_j)^2\rho(C_i)\rho(C_j) &\le \frac{N}{\lambda_2(L)}\sum_i\sum_{j|C_i\cap C_j\neq\emptyset}\left(\rho(C_i)\mathbb{V}\mathrm{ar}_{\rho_i}(f)+\rho(C_j)\mathbb{V}\mathrm{ar}_{\rho_j}(f)\right) \\
&\le \frac{2N^2}{\lambda_2(L)}\sum_i \rho(C_i)\mathbb{V}\mathrm{ar}_{\rho_i}(f).
\end{aligned}$$

Injecting this into (20) yields

$$\mathbb{V}\mathrm{ar}_\rho(f) \le \left(N+\frac{2N^2}{\lambda_2(L)}\right)\sum_i \rho(C_i)\mathbb{V}\mathrm{ar}_{\rho_i}(f). \qquad (22)$$

Now recalling that $f=\psi-\tilde{\psi}$, we have by Proposition B.1 for any $i\in\{1,\dots,N\}$ that

$$\langle\psi-\tilde{\psi}|(\nabla\psi^*)_\#\rho_i\rangle + C_{d,R,m_\rho,M_\rho}\mathbb{V}\mathrm{ar}_{\rho_i}(\tilde{\psi}^*-\psi^*) \le \mathcal{K}_{\rho_i}(\tilde{\psi}) - \mathcal{K}_{\rho_i}(\psi),$$

where $C_{d,R,m_\rho,M_\rho} = \left(e(d+1)2^{d+1}R^2\left(\frac{M_\rho}{m_\rho}\right)^2\right)^{-1}$. Weighting this last inequality with $\rho(C_i)$ and summing over $i \in \{1,\dots,N\}$, this raises

$$\langle \psi - \tilde{\psi} | (\nabla \psi^*)_{\#}\rho \rangle + \frac{C_{d,R,m_\rho,M_\rho}}{N} \sum_{i=1}^{N} \rho(C_i) \mathbb{V}\mathrm{ar}_{\rho_i}(\tilde{\psi}^* - \psi^*) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi).$$

Using (22) eventually gives

$$\langle \psi - \tilde{\psi} | (\nabla \psi^*)_{\#}\rho \rangle + c_\rho \mathbb{V}\mathrm{ar}_{\rho}(\tilde{\psi}^* - \psi^*) \leq \mathcal{K}_\rho(\tilde{\psi}) - \mathcal{K}_\rho(\psi),$$

where $c_\rho = \left(e(d+1)2^{d+1}R^2\left(\frac{M_\rho}{m_\rho}\right)^2\left(N^2 + \frac{2N^3}{\lambda_2(L)}\right)\right)^{-1}$. □

References

[1] Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[2] A. Ahidar-Coutrix, T. Le Gouic, and Q. Paris. Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics. *Probability Theory and Related Fields*, 177(1):323–368, Jun 2020.

[3] Jason M Altschuler and Enric Boix-Adsera. Wasserstein barycenters can be computed in polynomial time in fixed dimension. *Journal of Machine Learning Research*, 22(44):1–19, 2021.

[4] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[5] Jérémie Bigot, Elsa Cazelles, and Nicolas Papadakis. Penalization of Barycenters in the Wasserstein Space. *SIAM Journal on Mathematical Analysis*, 51(3):2261–2285, 2019.

[6] Jérémie Bigot, Raúl Gouet, Thierry Klein, and Alfredo López. Upper and lower risk bounds for estimating the Wasserstein barycenter of random measures on the real line. *Electronic Journal of Statistics*, 12(2):2253 – 2289, 2018.

[7] Bigot, Jérémie and Klein, Thierry. Characterization of barycenters in the wasserstein space by averaging optimal transport maps. *ESAIM: PS*, 22:35–57, 2018.

[8] Emmanuel Boissard, Thibaut Le Gouic, and Jean-Michel Loubes. Distribution's template estimate with Wasserstein metrics. *Bernoulli*, 21(2):740 – 759, 2015.

[9] Herm Jan Brascamp and Elliott H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, August 1976.

[10] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[11] Guillaume Carlier, Katharina Eichinger, and Alexey Kroshnin. Entropic-wasserstein barycenters: Pde characterization, regularity, and clt. *SIAM Journal on Mathematical Analysis*, 53(5):5880–5914, 2021.

[12] Carlier, Guillaume, Oberman, Adam, and Oudet, Edouard. Numerical methods for matching for teams and wasserstein barycenters. *ESAIM: M2AN*, 49(6):1621–1642, 2015.

[13] Sinho Chewi, Tyler Maunu, Philippe Rigollet, and Austin J. Stromme. Gradient descent algorithms for Bures-Wasserstein barycenters. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 1276–1304. PMLR, 09–12 Jul 2020.

[14] Pierre Colombo, Guillaume Staerman, Pablo Piantanida, and Chloé Clavel. Automatic Text Evaluation through the Lens of Wasserstein Barycenters. In *EMNLP 2021*, Punta Cana, Dominica, November 2021.

[15] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32(2) of *Proceedings of Machine Learning Research*, pages 685–693, Bejing, China, 22–24 Jun 2014. PMLR.

[16] Alex Delalande. Nearly tight convergence bounds for semi-discrete entropic optimal transport. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 1619–1642. PMLR, 28–30 Mar 2022.

[17] Alex Delalande. *Quantitative Stability in Quadratic Optimal Transport*. Theses, Université Paris-Saclay, December 2022.
[18] Alex Delalande and Quentin Mérigot. Quantitative stability of optimal transport maps under variations of the target measure. *Duke Mathematical Journal*, 2022.
[19] Pierre Dognin, Igor Melnyk, Youssef Mroueh, Jarret Ross, Cicero Dos Santos, and Tom Sercu. Wasserstein barycenter model ensembling. In *International Conference on Learning Representations*, 2019.
[20] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems*. Society for Industrial and Applied Mathematics, 1999.
[21] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, Aug 2015.
[22] Nhat Ho, XuanLong Nguyen, Mikhail Yurochkin, Hung Hai Bui, Viet Huynh, and Dinh Phung. Multilevel clustering via Wasserstein means. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1501–1509. PMLR, 06–11 Aug 2017.
[23] Young-Heon Kim and Brendan Pass. Wasserstein barycenters over riemannian manifolds. *Advances in Mathematics*, 307:640–683, 2017.
[24] Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc.*, 21(9):2603–2651, 2019.
[25] Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of wasserstein barycenters. *Probability Theory and Related Fields*, 168(3):901–917, Aug 2017.
[26] Thibaut Le Gouic, Quentin Paris, Philippe Rigollet, and Austin Stromme. Fast convergence of empirical barycenters in alexandrov spaces and the wasserstein space. *J. Eur. Math. Soc.*, 2022.
[27] Xin Lian, Kshitij Jain, Jakub Truszkowski, Pascal Poupart, and Yaoliang Yu. Unsupervised multilingual alignment using wasserstein barycenter. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 3702–3708. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Main track.
[28] Robert J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.
[29] Victor M. Panaretos and Yoav Zemel. *An Invitation to Statistics in Wasserstein Space*. SpringerBriefs in Probability and Mathematical Statistics. Springer Cham, 2020.
[30] Brendan Pass. Optimal transportation with infinitely many marginals. *Journal of Functional Analysis*, 264(4):947–963, 2013.
[31] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
[32] Julien Rabin, Gabriel Peyré, Julie Delon, and Bernot Marc. Wasserstein Barycenter and its Application to Texture Mixing. In *SSVM'11*, pages 435–446, Israel, 2011. Springer.
[33] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55:58–63, 2015.
[34] Filippo Santambrogio and Xu-Jia Wang. Convexity of the support of the displacement interpolation: Counterexamples. *Applied Mathematics Letters*, 58:152–158, 2016.
[35] Shai Shalev-Shwartz, Ohad Shamir, Nathan Srebro, and Karthik Sridharan. Learnability, stability and uniform convergence. *Journal of Machine Learning Research*, 11(90):2635–2670, 2010.
[36] Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.*, 34(4), jul 2015.
[37] Sanvesh Srivastava, Cheng Li, and David B. Dunson. Scalable bayes via barycenter in wasserstein space. *Journal of Machine Learning Research*, 19(8):1–35, 2018.
[38] Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive curvature. *Contemp. Math.*, 338, 01 2003.
[39] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996.
[40] V. Vapnik. Principles of risk minimization for learning theory. In J. Moody, S. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems*, volume 4. Morgan-Kaufmann, 1991.
[41] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 19(1/2):23–26, 1958.
[42] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
[43] Jonathan Weed and Francis R. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 2019.

Ceremade, Univ. Paris-Dauphine PSL, 75775 Paris and Mokaplan, Inria Paris
Email address: carlier@ceremade.dauphine.fr

Lagrange Mathematics and Computing Research Center, 75007, Paris, France
Email address: delalande.alex@gmail.com

Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France and Institut universitaire de France
Email address: quentin.merigot@universite-paris-saclay.fr