Sparsity of Quadratically Regularized Optimal Transport: Scalar Case

Alberto González-Sanz*   Marcel Nutz†

October 4, 2024

Abstract

The quadratically regularized optimal transport problem is empirically known to have sparse solutions: its optimal coupling π_ε has sparse support for small regularization parameter ε, in contrast to entropic regularization whose solutions have full support for any $\varepsilon > 0$. Focusing on continuous and scalar marginals, we provide the first precise description of this sparsity. Namely, we show that the support of π_ε shrinks to the Monge graph at the sharp rate $\varepsilon^{1/3}$. This result is based on a detailed analysis of the dual potential f_ε for small ε. In particular, we prove that f_ε is twice differentiable a.s. and bound the second derivative uniformly in ε, showing that f_ε is uniformly strongly convex. Convergence rates for f_ε and its derivative are also obtained.

Keywords Optimal Transport; Quadratic Regularization; Sparsity
AMS 2020 Subject Classification 49N10; 49N05; 90C25

1 Introduction

For probability measures μ_0, μ_1 on $\mathbb{R}^d$, the optimal transport problem with quadratic cost is

$$\mathsf{OT} := \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int \frac{1}{2} \|x - y\|^2 d\pi(x, y) \tag{1}$$

where $\Pi(\mu_0, \mu_1)$ denotes the set of couplings (or transport plans) of μ_0, μ_1. Given a regularization parameter $\varepsilon > 0$, we study the quadratically regularized transport problem where couplings are penalized by the squared L^2-norm of their density,

$$\mathsf{QOT}_\varepsilon := \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int \frac{1}{2} \|x - y\|^2 d\pi(x, y) + \frac{\varepsilon}{2} \left\| \frac{d\pi}{d(\mu_0 \otimes \mu_1)} \right\|^2_{L^2(\mu_0 \otimes \mu_1)}. \tag{2}$$

The basic idea of regularized optimal transport is to approximate the linear problem (1) by a strictly convex one. Since the influential work of [14] proposing entropic regularization to enable Sinkhorn's algorithm, applications of optimal transport have exploded in machine learning, statistics, image processing, and other domains where distributions or data sets need to be compared (see, e.g., [40]). Regularization has several purposes in this context:

---

*Department of Statistics, Columbia University, ag4855@columbia.edu

facilitating computation as in [14], improving sampling complexity [22, 33], obtaining stable couplings [2, 24], etc. Two regularizations are primarily used; entropic regularization penalizes couplings by the Kullback–Leibler divergence while quadratic regularization penalizes by the L^2-norm of the density. Entropic regularization (EOT) is the most frequent choice and has the strongest smoothness properties (e.g., [36, 40]). This smoothness is linked to the full support property: the support of the optimal coupling equals the one of the product $\mu_0 \otimes \mu_1$ for any value of the regularization parameter, even though the unregularized optimal transport is concentrated on a graph. This can be undesirable ("overspreading") depending on the application—it may correspond to blurrier images in an image processing task as shown in [5] or bias in a manifold learning task as in [47]. By contrast, quadratic regularization (QOT) is empirically known to give rise to couplings with *sparse support* for small ε, and can therefore be useful in such situations. While EOT has been studied in hundreds of works in the last few years, QOT is traditionally considered less analytically tractable and theoretical results are scarce.

Focusing on the scalar case $d = 1$ and continuous (cf. Assumption 2.1) marginals μ_0, μ_1, the present work is the first to describe the sparsity of the optimal coupling π_ε for (2). More generally, we provide a fairly complete quantitative picture of QOT for small regularization parameter ε. Our main result shows that the support $\mathcal{S}_\varepsilon$ of the optimal coupling π_ε shrinks at rate $\varepsilon^{\frac{1}{3}}$. More precisely, this holds uniformly for all sections of the support: there is $C > 0$ such that for small ε,

$$C^{-1} \varepsilon^{\frac{1}{3}} \leq |\mathcal{S}_{x,\varepsilon}| \leq C \varepsilon^{\frac{1}{3}}$$

for all x, where $\mathcal{S}_{x,\varepsilon}$ is the x-section of $\mathcal{S}_\varepsilon$ and $|\mathcal{S}_{x,\varepsilon}|$ is its diameter; see Theorem 2.5. Moreover, $\mathcal{S}_\varepsilon$ approaches the support of the limiting optimal transport π_0 at the same rate. Note that the support of π_0 is the graph of the optimal transport map (Monge map) T_0 inducing π_0. Denoting by d the Hausdorff distance, the convergence can be stated as

$$C^{-1} \varepsilon^{\frac{1}{3}} \leq d\left(\mathcal{S}_{x,\varepsilon}, T_0(x)\right) \leq C \varepsilon^{\frac{1}{3}}.$$

More precisely, we show this in L^2-norm; see Theorem 2.6. Our result has a direct interpretation: it guarantees that the solution of the quadratically regularized problem does not place mass far from the optimal transport map, unlike the solution of EOT. We conjecture that the rate generalizes to d-dimensional space as $\varepsilon^{\frac{1}{d+2}}$, however at present we can show this only in special cases (such as the self-transport case $\mu_0 = \mu_1$). Our convergence rate confirms the heuristic link between QOT and the porous media equation recently emphasized in [21] (see also Section 1.1 below): in the small-time limit $t \to 0$, the support of the solution to that equation tends to the support of the initial condition at the same rate $t^{\frac{1}{3}}$ (for $d = 1$, and $t^{\frac{1}{d+2}}$ in general). Separately, we establish an interesting geometric fact that has not been documented before (Proposition 4.2): like the limiting optimal transport $T_0(x)$, the support $\mathcal{S}_{x,\varepsilon}$ is increasing in x (i.e., for $x < x'$ there is a bijection $\phi : \mathcal{S}_{x,\varepsilon} \to \mathcal{S}_{x',\varepsilon}$ with $\phi(y) \geq y$).

Our results on the support are based on a detailed analysis of the dual potentials $(f_\varepsilon, g_\varepsilon)$ associated with π_ε via

$$\frac{d\pi_\varepsilon}{d(\mu_0 \otimes \mu_1)}(x, y) = \left(\frac{xy - f_\varepsilon(x) - g_\varepsilon(y)}{\varepsilon} \right)_+.$$

While we refer to Section 2.1 for detailed definitions, one may say that f_ε is the analogue of the convex Kantorovich potential f_0 inducing Monge's map via its derivative $f_0' = T_0$. We

2

show that f_ε is $\mathcal{C}^1$ and twice differentiable a.s., and provide expressions for the derivatives. The main technical result, obtained in tandem with sparsity, is uniform strong convexity. More precisely,

$$C^{-1} \leq f''_\varepsilon(x) \leq C$$

for small ε; see Theorem 2.2. We also obtain rates for $f_\varepsilon \to f_0$ and $f'_\varepsilon \to f'_0 = T_0$; see Theorem 2.6 and its corollary.

The proof of our main results is based on a geometric approach that, to the best of our knowledge, is novel. It is easy to bound the following geometric quantity: the product of the diameter $|\mathcal{S}_{x,\varepsilon}|$ and the maximum value of the density $\frac{d\pi_\varepsilon}{d(\mu_0 \otimes \mu_1)}(x, \cdot)$. Thus, we study the relation between these two, which is closely related to the second derivative of the potential. Specifically, we bound f''_ε in terms of a ratio of diameters of x-sections and y-sections of the support. At the same time, we bound these diameters in term of the second derivatives. Analyzing the powers in those relations yields the desired uniform bounds for all quantities.

We remark that a generalization of our results to the multivariate case would have substantial consequences for the regularity theory of (unregularized) optimal transport because the Kantorovich potential inherits the properties of f_ε. Specifically, this approach may enable boundary regularity results along the lines of [9] without the α-Hölder conditions on the marginals that seems to be essential for the Monge–Ampère approach. (That may be a hint that the generalization to d dimensions is not straightforward.) We remark that EOT has been used to obtain regularity results in a similar spirit [11]. However, that approach using covariance inequalities is valid only for marginals with full support on $\mathbb{R}^d$, hence cannot yield boundary regularity.

While completing this paper, we learned about concurrent and independent ongoing research by Johannes Wiesel and Xingyu Xu who kindly shared their results. Their work has the same goal of bounding the size of the optimal support and distance to the Monge map for small ε, but their approach does not involve studying the shape of the potentials in detail. Instead, their arguments start with a relatively direct bound on the maximum of the density based on equations (14), (15). This bound is very general, allowing for much less restrictive assumptions than in our work (multivariate setting, non-convex marginal supports, unbounded marginal density, etc.) and leading to the first quantitative results in such a setting. On the other hand, it is not accurate enough to deliver the sharp rate. Assuming that the Monge map is Lipschitz, the size of the support is bounded from above by $C\varepsilon^{\frac{1}{2d+2}}$ and the distance to the Monge map by $C\varepsilon^{\frac{1}{4(d+1)^2}}$. The sharp rate $\varepsilon^{\frac{1}{d+2}}$ is obtained only when the Monge map is the identity.

1.1 Related literature

Quadratically regularized optimal transport was first explored by [5] in the discrete setting, with experiments highlighting the sparsity and theoretical results including convergence for small regularization parameter. On the other hand, [19] studied quadratic regularization for a minimum-cost flow problem on a graph, including discrete optimal transport as a special case. Sparsity is discussed in several examples. Subsequently, [16] considered optimal transport with convex regularization, again with quadratic regularization as a special case. The asymptotic questions considered in the present work do not have an exact counterpart in the discrete case. Indeed, it is known from [32] that the optimal coupling π_ε converges

stationarily to the minimum-norm optimal transport π_0; that is, there exists a threshold $\varepsilon_0 > 0$ such that $\pi_\varepsilon = \pi_0$ for $\varepsilon \leq \varepsilon_0$. While one cannot speak of a convergence rate, the recent work [25] can be seen as an analogue, providing the value of the threshold ε_0. Also in the discrete case, [26] investigates whether the support of π_ε is monotone in ε before reaching ε_0, and finds a negative answer.

In the continuous setting, quadratically regularized optimal transport was first explored from a computational point of view. Several works including [17, 23, 27, 29, 43] approach the dual problem by optimization techniques. For instance, [29] computes regularized Wasserstein barycenters using neural networks and finds that the quadratic penalty produces sharper results than the entropic. More recently, [47] uses quadratic regularization in a manifold learning task related to single cell RNA sequencing and notes that sparsity is crucial to avoid biasing the affinity matrix.

The first work rigorously addressing a continuous setting is [31], deriving duality results and presenting two algorithms. Their problem formulation differs slightly from (2) in that the L^2-norm is taken with respect to the Lebesgue measure rather than $\mu_0 \otimes \mu_1$. We prefer the latter as it is generally applicable and behaves naturally under approximation by sampling or discretization. More recently, [37] derives the basic results of [31] in a more general setting and studies the uniqueness of the dual potentials. Noting that the potential must converge to the Kantorovich potential for $\varepsilon \to 0$ by the Arzelà–Ascoli theorem, [37] concludes that the support of π_ε converges to the one of π_0. This qualitative convergence result is stated in Proposition 4.1 below.

Prior to the present work, quantitative results were available only for the convergence $\mathsf{QOT}_\varepsilon \to \mathsf{OT}$ of the optimal transport *cost*. Indeed, for a more general class of regularizations, a rate is provided in [18] based on a quantization argument; the qualitative (Gamma) convergence was previously shown in [30]. Specializing the result of [18] to quadratic regularization yields $C^{-1}\varepsilon^{\frac{2}{d+2}} \leq \mathsf{QOT}_\varepsilon - \mathsf{OT} \leq C\varepsilon^{\frac{2}{d+2}}$. The very recent analysis of [21] shows that this power is the exact leading term and identifies the constant:

$$\lim_{\varepsilon \to 0} \frac{\mathsf{QOT}_\varepsilon - \mathsf{OT}}{\varepsilon^{\frac{2}{d+2}}} = C_d \int_{\mathbb{R}^d} \left(u_0(x)u_1(T_0(x))\right)^{-\frac{1}{(d+2)}} d\mu_0(x)$$

where C_d is an explicit dimensional constant and u_i is the density of μ_i. This result is obtained by establishing a link to the porous media equation $\partial_t u = \Delta u^2$. Namely, an approximate solution of QOT is obtained by modifying the Barenblatt profile which is a fundamental solution of the porous media equation (see [45]). Based on the shape of the approximate solution in [21] one can conjecture that support of the true solution also shrinks at rate $\varepsilon^{\frac{1}{d+2}}$, which is confirmed by our results.

Finally, we mention that the limit $\varepsilon \to 0$ has been studied in detail for the EOT problem, starting with [12] for discrete and [34, 35] (see also [10, 28]) for continuous marginals. See [1, 3, 4, 7, 13, 38, 39], among others. Of course, the optimal support has not been of much interest in EOT, as it is simply the support of $\mu_0 \otimes \mu_1$. There is a direct analogy between the role of the heat equation in [39] for EOT and the porous media equation in [21] for QOT. In the present work, our arguments follow a novel route.

The remainder of this paper is organized as follows. Section 2 details notation and background, then states the main results. Section 3 establishes convexity and determines

4

the first derivative of the potential f_ε, whereas Section 4 is devoted to the second derivative. The main part is Section 5, bounding the second derivative and the diameter of the support. Finally, Section 6 details how those results imply the convergence rate to unregularized optimal transport.

2 Main Results

2.1 Background and Notation

Let $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R})$ be probability measures on $\mathbb{R}$ with finite second moments and $\Pi(\mu_0, \mu_1)$ their set of couplings; i.e., $\pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ such that $\pi(\cdot \times \mathbb{R}) = \mu_0(\cdot)$ and $\pi(\mathbb{R} \times \cdot) = \mu_1(\cdot)$. Given $\varepsilon > 0$, the quadratically regularized optimal transport problem between μ_0 and μ_1 is

$$\inf_{\pi \in \Pi(\mu_0, \mu_1)} \int \frac{1}{2}(x-y)^2 d\pi(x,y) + \frac{\varepsilon}{2} \int \left(\frac{d\pi}{d(\mu_0 \otimes \mu_1)}(x,y) \right)^2 d(\mu_0 \otimes \mu_1)(x,y). \qquad (3)$$

We refer to [37] for the following facts. The problem (3) has a unique minimizer, denoted $\pi_\varepsilon \in \Pi(\mu_0, \mu_1)$. The dual problem of (3) is

$$\sup_{(f,g) \in L^2(\mu_0) \times L^2(\mu_1)} \int f(x) d\mu_0(x) + \int g(y) d\mu_1(y)$$
$$- \frac{1}{2\varepsilon} \int \left(f(x) + g(y) - \frac{(x-y)^2}{2} \right)_+^2 d(\mu_0 \otimes \mu_1)(x,y). \qquad (4)$$

Given any solution $(\tilde{f}, \tilde{g})$ of (4), we have the "optimality condition"

$$\frac{d\pi_\varepsilon}{d(\mu_0 \otimes \mu_1)}(x,y) = \left(\frac{\tilde{f}(x) + \tilde{g}(y) - \frac{1}{2}(x-y)^2}{\varepsilon} \right)_+. \qquad (5)$$

To be in line with convex analysis, it will be convenient to reparametrize the solution as

$$\tilde{f}(x) = \tfrac{1}{2}x^2 - f_\varepsilon(x), \qquad \tilde{g}(x) = \tfrac{1}{2}y^2 - g_\varepsilon(x), \qquad (6)$$

so that (5) becomes

$$\frac{d\pi_\varepsilon}{d(\mu_0 \otimes \mu_1)}(x,y) = \left(\frac{xy - f_\varepsilon(x) - g_\varepsilon(y)}{\varepsilon} \right)_+. \qquad (7)$$

Under Assumption 2.1 below, the "potentials" $(f_\varepsilon, g_\varepsilon)$ are unique (a.s.) up to a parallel shift; that is, up to changing $(f_\varepsilon, g_\varepsilon)$ to $(f_\varepsilon + \alpha, g_\varepsilon - \alpha)$ for some $\alpha \in \mathbb{R}$. Moreover, we may choose $(f_\varepsilon, g_\varepsilon)$ to be (everywhere defined and) Lipschitz continuous on $\mathbb{R}$. See [37, Theorem 3.7, Lemma 2.5].

Next, we introduce standard notation and results for the (unregularized) optimal transport problem (see [46]),

$$\inf_{\pi \in \Pi(\mu_0, \mu_1)} \int \frac{1}{2}(x-y)^2 d\pi(x,y). \qquad (8)$$

5

For absolutely continuous marginal μ_0, this problem has a unique solution $\pi_0 \in \Pi(\mu_0, \mu_1)$. Moreover, $\pi_0 = (\mathrm{Id} \times T_0)_\# \mu_0$ where $T_0 : \mathbb{R} \to \mathbb{R}$ is the optimal transport map, often called Monge map. Here $\#$ denotes push-forward, $f_\#(\mu) := \mu(f^{-1}(\cdot))$. In the present scalar case, it holds that $T_0 = F_1^{-1} \circ F_0$, where $F_i(x) = \mu_i((-\infty, x])$ is the cumulative distribution function of μ_i. On the other hand, $T_0 = f_0'$ is the derivative of a convex function f_0 called the Kantorovich potential. Namely, (f_0, g_0) solves the dual problem

$$\inf_{(f,g):\ f(x)+g(x) \geq xy} \int f(x) d\mu_0(x) + \int g(y) d\mu_1(y). \tag{9}$$

Here we have applied the same reparametrization as in (6) to work directly with convex functions—indeed, (f_0, g_0) is the limit of $(f_\varepsilon, g_\varepsilon)$ from (7). Again, (f_0, g_0) is unique up to parallel shift under Assumption 2.1 below.

Some more notation will be useful. For brevity,

$$|B| \text{ denotes the Lebesgue measure of a Borel set } B \subset \mathbb{R}.$$

We will use this primarily when B is an interval, so that $|B|$ is also the length. As usual, the essential supremum of $f : \Omega_0 \to \mathbb{R}$ is $\operatorname{ess\,sup}_{\Omega_0} f = \inf\{M \in \mathbb{R} : |\{x \in \Omega_0 : f(x) > M\}| = 0\}$, and similarly $\operatorname{ess\,inf}_{\Omega_0} f = -\operatorname{ess\,sup}_{\Omega_0}(-f)$. Finally, χ_A denotes the (measure-theoretic) indicator function, $\chi_A(x) = 1$ for $x \in A$ and $\chi_A(x) = 0$ for $x \notin A$.

2.2 Results

All results are stated under the following standard condition on the marginals.

Assumption 2.1. The marginals $\mu_i \in \mathcal{P}(\mathbb{R})$ are compactly supported and have continuous densities bounded from above and below. That is, for $i = 0, 1$, there are compact intervals $\Omega_i = [a_i, b_i] \subset \mathbb{R}$ and $u_i \in \mathcal{C}(\Omega_i)$ such that $d\mu_i = u_i dx \in \mathcal{P}(\Omega_i)$ and

$$0 < \lambda \leq u_i \leq \Lambda < \infty \quad \text{on } \Omega_i \tag{10}$$

for some constants λ, Λ.

The results below are stated for the potentials f_ε and the x-sections of the support. As Assumption 2.1 is symmetric, the same results hold for g_ε and the y-sections. Our first result establishes the regularity of the potential f_ε and most importantly its *strong convexity* uniformly in ε.

Theorem 2.2 (Regularity)**.** *There exist $C, \varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$,*

  (i) *f_ε is convex and in $\mathcal{C}^1(\Omega_0)$ (see Proposition 3.2 for a formula for f_ε');*

 (ii) *f_ε is twice differentiable a.e.[1] (see Proposition 4.3 for a formula for f_ε'');*

(iii) *f_ε is uniformly strongly convex with[2]*

$$\operatorname*{ess\,sup}_{\Omega_0} f_\varepsilon'' \leq C \quad and \quad \operatorname*{ess\,inf}_{\Omega_0} f_\varepsilon'' \geq C^{-1}.$$

---

[1] In fact, twice differentiable except at two particular points $x^{(m)}, x^{(M)}$; see Section 4.

[2] Similarly, one can take the pointwise sup and inf over $\Omega_0 \setminus \{x^{(m)}, x^{(M)}\}$.

The constant C depends only on the constants and the moduli of continuity of u_i in Assumption 2.1.

The condition that ε be small is essential for strong convexity. In fact, f_ε is affine for large ε; see Remark 4.4.

Remark 2.3 (Necessity of Assumption 2.1)**.** Given the uniform convergence $f_\varepsilon \to f_0$ (see [37] or below), the assertion of Theorem 2.2 implies that the transport map $T_0 = f_0'$ is Lipschitz with (a.s. defined) derivative bounded away from 0 and infinity. Recall that $T_0 = F_1^{-1} \circ F_0$ in the present scalar case, whence the marginals μ_i have densities bounded above and below. In that sense, (10) is necessary for Theorem 2.2 to hold. (The continuity of the densities could potentially be relaxed.)

Remark 2.4 (Multivariate case)**.** We can extend Theorem 2.2 (i),(ii) to the multivariate case. However, the key result (iii) is much more delicate, which is why we leave the multivariate case for future work. We observe that (iii) implies in particular a similar regularity result for the Kantorovich potential. It is known that the (boundary) regularity of the Kantorovich potential is a very deep result in the multivariate case [6, 44, 9].

Our second and main result quantifies the sparsity of the optimal coupling π_ε. Specifically, we describe the diameter of the sections of the support,

$$\mathcal{S}_{x,\varepsilon} := \{y \in \Omega_1 : \ xy - f_\varepsilon(x) - g_\varepsilon(y) \geq 0\}.$$

Note that each section $\mathcal{S}_{x,\varepsilon}$ is an interval due to the convexity of g_ε, hence the diameter is equal to the Lebesgue measure $|\mathcal{S}_{x,\varepsilon}|$. Our result shows that the diameter is (exactly) of order $\varepsilon^{\frac{1}{3}}$, uniformly in x.

Theorem 2.5 (Sparsity)**.** *There exist $C, \varepsilon_0 > 0$ such that*

$$C^{-1}\varepsilon^{\frac{1}{3}} \leq |\mathcal{S}_{x,\varepsilon}| \leq C\varepsilon^{\frac{1}{3}}$$

for all $x \in \Omega_0$ and $\varepsilon < \varepsilon_0$.

While Theorem 2.5 gives the sharp rate at which the support of π_ε shrinks, this alone does not imply that the support is close to the unregularized optimal transport. The next result provides the sharp convergence rate for the Hausdorff distance

$$d\left(\mathcal{S}_{x,\varepsilon}, T_0(x)\right) = \sup_{y \in \mathcal{S}_{x,\varepsilon}} |y - T_0(x)|$$

between $\mathcal{S}_{x,\varepsilon}$ and the image $\{T_0(x)\}$ of Monge's map; note that the latter is the x-section of the support of π_0. The convergence is measured in $L^2(\Omega_0)$-norm $\|h\|_{L^2(\Omega_0)} := \int_{\Omega_0} h(x)^2 dx$. As a consequence, we also obtain the rate of convergence of f_ε' towards T_0 in terms of the same norm.

Theorem 2.6 (Convergence)**.** *There exist $C, \varepsilon_0 > 0$ such that for every $\varepsilon < \varepsilon_0$,*

(i) the Hausdorff distance between the optimal supports satisfies

$$C^{-1}\varepsilon^{\frac{1}{3}} \leq \|d\left(\mathcal{S}_{x,\varepsilon}, T_0(x)\right)\|_{L^2(\Omega_0, dx)} \leq C\varepsilon^{\frac{1}{3}},$$

(ii) the derivatives f'_ε converge to Monge's map $T_0 = f'_0$ according to

$$\|f'_\varepsilon - T_0\|_{L^2(\Omega_0, dx)} \le C\varepsilon^{\frac{1}{3}}. \tag{11}$$

The rate (11) for the derivatives of the potentials implies a rate for the potentials themselves. As the potentials are only determined up to an additive constant, convergence can only hold if the constant is chosen in a suitable way. For instance, we can normalize the potentials symmetrically; i.e.,

$$\int f_\varepsilon(x)\mu_0(x) = \int g_\varepsilon(y)\mu_1(y), \qquad \int f_0(x)\mu_0(x) = \int g_0(y)\mu_1(y). \tag{12}$$

Another popular normalization is $\int g_\varepsilon(y)\mu_1(y) = \int g_0(y)\mu_1(y) = 0$; the subsequent result also holds for that choice. We use the Hölder norm

$$\|h\|_{C^{0,\frac{1}{2}}(\Omega_0)} := \sup_{x\in\Omega_0} |h(x)| + \sup_{x,y\in\Omega_0, x\ne y} \frac{|h(x)-h(y)|}{|x-y|^{\frac{1}{2}}}.$$

Corollary 2.7. *Let the potentials be normalized according to (12). There exist $C, \varepsilon_0 > 0$ such that for every $\varepsilon < \varepsilon_0$,*

$$\|f_\varepsilon - f_0\|_{C^{0,\frac{1}{2}}(\Omega_0)} \le C\varepsilon^{\frac{1}{3}}.$$

Remark 2.8. We conjecture that the rate (11) is sharp, whereas the rate for $\|f_\varepsilon - f_0\|$ in Corollary 2.7 is not. The sharp rate for $\|f_\varepsilon - f_0\|$ is conjectured to be $\varepsilon^{\frac{2}{3}}$.

2.3 Outline of the proofs

Theorems 2.2 and 2.5 are proved simultaneously though several steps. First, in Section 3, we prove that f_ε is differentiable with a monotone increasing derivative $f'_\varepsilon(x) = \frac{\int_{\mathcal{S}_{x,\varepsilon}} y d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})}$, which implies that f_ε is convex. In Section 4, we show that f_ε is twice differentiable (except possibly at two points) and provide an explicit formula for $f''_\varepsilon(x)$. In Section 5, this formula is used to bound $f''_\varepsilon(x)$. First, in Section 5.1, we provide a relationship between the diameter $|\mathcal{S}_{x,\varepsilon}|$ of the support and the maximum (over y) of the density of π_ε, whereas Section 5.2 shows that $f'_\varepsilon(x)$ is approximately at the center of the support $\mathcal{S}_{x,\varepsilon}$. In Section 5.3, we give an initial bound on $f''_\varepsilon(x)$ in terms of the ratio between the diameters of different supports. Roughly, the bound is of the form

$$C^{-1} \inf_{y\in\partial\mathcal{S}_{x,\varepsilon}} \frac{|\mathcal{S}_{x,\varepsilon}|}{|\mathcal{S}^{(0)}_{y,\varepsilon}|} \le f''_\varepsilon(x) \le C \sup_{y\in\partial\mathcal{S}_{x,\varepsilon}} \frac{|\mathcal{S}_{x,\varepsilon}|}{|\mathcal{S}^{(0)}_{y,\varepsilon}|},$$

where $\mathcal{S}^{(0)}_{y,\varepsilon}$ is the y-section of the support (the symmetric analogue of $\mathcal{S}_{x,\varepsilon}$ with reversed roles for x, y). We proceed by relating and simultaneously bounding the four objects $\mathcal{S}_{x,\varepsilon}$, f''_ε and $\mathcal{S}^{(0)}_{y,\varepsilon}$, g''_ε. A key intermediate result is a bound for $|\mathcal{S}_{x,\varepsilon}|$ in terms of $\sigma_M(f_\varepsilon) = \operatorname{ess\,sup}_{\Omega_0} f''_\varepsilon$ and $\sigma_m(f_\varepsilon) = \operatorname{ess\,inf}_{\Omega_0} f''_\varepsilon$, namely

$$\left(\frac{\varepsilon}{\sigma_M(f_\varepsilon)}\right)^{\frac{1}{3}} \le |\mathcal{S}_{x,\varepsilon}| \le \left(\frac{\varepsilon}{\sigma_m(f_\varepsilon)}\right)^{\frac{1}{3}}, \tag{13}$$

as well as an analogous bound for $|\mathcal{S}_{y,\varepsilon}^{(0)}|$ in terms of $\sigma_M(g_\varepsilon) = \operatorname{ess\,sup}_{\Omega_1} g_\varepsilon''$ and $\sigma_m(g_\varepsilon) = \operatorname{ess\,inf}_{\Omega_1} g_\varepsilon''$. Using these bounds, we are able to relate the derivatives of f_ε and g_ε by

$$\sigma_M(f_\varepsilon) \leq C \left(\frac{\sigma_M(g_\varepsilon)}{\sigma_m(f_\varepsilon)} \right)^{\frac{1}{3}}, \qquad \sigma_m(f_\varepsilon) \geq \frac{1}{C} \left(\frac{\sigma_m(g_\varepsilon)}{\sigma_M(f_\varepsilon)} \right)^{\frac{1}{3}},$$

$$\sigma_M(g_\varepsilon) \leq C \left(\frac{\sigma_M(f_\varepsilon)}{\sigma_m(g_\varepsilon)} \right)^{\frac{1}{3}}, \qquad \sigma_m(g_\varepsilon) \geq \frac{1}{C} \left(\frac{\sigma_m(f_\varepsilon)}{\sigma_M(g_\varepsilon)} \right)^{\frac{1}{3}}.$$

After some algebra, this implies uniform bounds for all four quantities $\sigma_m(f_\varepsilon), \sigma_M(f_\varepsilon)$, $\sigma_m(g_\varepsilon)$, $\sigma_M(g_\varepsilon)$, establishing the results on $f_\varepsilon''(x)$ in Theorem 2.2. Now, (13) also yields the rate of $|\mathcal{S}_{x,\varepsilon}|$ in Theorem 2.5.

The proof of Theorem 2.6 is obtained from the rate of $|\mathcal{S}_{x,\varepsilon}|$, the rate for the regularized transport cost from [18] and an argument that we adapt from [8].

3   Convexity and first order regularity

In this section we prove that the potential f_ε is a $\mathcal{C}^1$, convex function. Recall the notation introduced in Assumption 2.1 (which is in force for the remainder of the paper) and the continuous functions f_ε and g_ε defined by (7), which satisfy (see [37])

$$\int (xy - f_\varepsilon(x) - g_\varepsilon(y))_+ d\mu_1(y) = \varepsilon \quad \text{for all } x \in \Omega_0, \tag{14}$$

$$\int (xy - f_\varepsilon(x) - g_\varepsilon(y))_+ d\mu_0(x) = \varepsilon \quad \text{for all } y \in \Omega_1. \tag{15}$$

These equations express the fact that the measure π_ε of (7) has marginals μ_0 and μ_1. Recall also the sections of the optimal support,

$$\mathcal{S}_{x,\varepsilon} = \{ y \in \Omega_1 : \ xy - f_\varepsilon(x) - g_\varepsilon(y) \geq 0 \},$$

and note that (14) implies $\mu_1(\mathcal{S}_{x,\varepsilon}) > 0$ for any $x \in \Omega_0$.

Lemma 3.1. *Let $x \in \Omega_0$ and $t > 0$ be such that $x + t \in \Omega_0$. Then*

$$T_\varepsilon(x) := \frac{\int_{\mathcal{S}_{x,\varepsilon}} y\, d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})} \quad \text{satisfies} \quad T_\varepsilon(x+t) \geq \frac{f_\varepsilon(x+t) - f_\varepsilon(x)}{t} \geq T_\varepsilon(x).$$

In particular, T_ε is monotone.

Proof. The inequality

$$(a)_+ - (b)_+ \leq \chi_{a \geq 0}(a - b), \quad a, b \in \mathbb{R}, \tag{16}$$

where $\chi_{a \geq 0} := 1$ if $a \geq 0$ and $\chi_{a \geq 0} := 0$ otherwise, yields

$$(xy - f_\varepsilon(x) - g_\varepsilon(y))_+ \leq ((x+t)y - f_\varepsilon(x+t) - g_\varepsilon(y))_+ + \chi_{\mathcal{S}_{x,\varepsilon}} (f_\varepsilon(x+t) - f_\varepsilon(x) - ty) \tag{17}$$

as well as

$$(xy - f_\varepsilon(x) - g_\varepsilon(y))_+ \geq ((x+t)y - f_\varepsilon(x+t) - g_\varepsilon(y))_+$$
$$+ \chi_{\mathcal{S}_{x+t,\varepsilon}} (f_\varepsilon(x+t) - f_\varepsilon(x) - ty). \tag{18}$$

Integrating (17) with respect to μ_1 and using (14) yields $f_\varepsilon(x+t) - f_\varepsilon(x) \geq tT_\varepsilon(x)$. Similarly, integrating (18) yields the second inequality in the claim. $\qquad \square$

9

Noting that $T_\varepsilon(x) \in \Omega_1$ is uniformly bounded, Lemma 3.1 implies in particular that f_ε is Lipschitz. Next, we prove that in fact $f_\varepsilon \in \mathcal{C}^1(\Omega_0)$. We write int A for the interior of a set A.

Proposition 3.2. *For all $\varepsilon > 0$, the function f_ε is convex on Ω_0. Moreover, $f_\varepsilon \in \mathcal{C}^1(\Omega_0)$ with derivative*

$$f_\varepsilon'(x) = \frac{\int_{\mathcal{S}_{x,\varepsilon}} y d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})} \in \text{int } \mathcal{S}_{x,\varepsilon} \quad \text{for all} \quad x \in \Omega_0. \tag{19}$$

Proof. As T_ε is monotone by Lemma 3.1, there is a l.s.c. convex function $\phi : \Omega_0 \to \mathbb{R}$ such that

$$T_\varepsilon(x) \in \partial \phi(x) := \{y \in \mathbb{R} : \phi(z) \geq \phi(x) + y(z - x) \text{ for all } z \in \mathbb{R}\}$$

for all $x \in \Omega_0$; cf. [42, Theorem 12.17]. Since $\partial\phi(x)$ is the singleton $\{\phi'(x)\}$ for a.e. $x \in \Omega_0$, it follows that $T_\varepsilon(x) = \phi'(x)$ for a.e. $x \in \Omega_0$. Call Ω_0' the set of all $x \in \Omega_0$ where the latter holds. Then T_ε is continuous on Ω_0'; cf. [41, Theorem 25.5.]. On the other hand, since f_ε is Lipschitz, Rademacher's theorem yields a set $\Omega_0'' \subset \Omega_0$ with $|\Omega_0 \setminus \Omega_0''| = 0$ on which f_ε is differentiable. In view of Lemma 3.1 we conclude that $f_\varepsilon' = T_\varepsilon = \phi'$ on $\Omega_0' \cap \Omega_0''$. This implies that $f_\varepsilon = \phi + \alpha$ on Ω_0, for some constant α (cf. [15, Lemma 2.5]), and completes the proof that f_ε is convex. Symmetrically, g_ε is convex, which in particular implies that $\mathcal{S}_{x,\varepsilon}$ is convex; i.e., an interval. (The convexity of the support Ω_1 is crucial here.)

Let $\{x_n\} \subset \Omega_0$ satisfy $x_n \to x \in \Omega_0$. As f_ε is continuous, the boundary points of the (nonempty) intervals $\mathcal{S}_{x_n,\varepsilon}$ converge to the ones of $\mathcal{S}_{x,\varepsilon}$. Since μ is atomless, this implies

$$\mu_1(\mathcal{S}_{x_n,\varepsilon}) \to \mu_1(\mathcal{S}_{x,\varepsilon}) \quad \text{and} \quad \int_{\mathcal{S}_{x_n,\varepsilon}} y d\mu_1(y) \to \int_{\mathcal{S}_{x,\varepsilon}} y d\mu_1(y).$$

As a consequence, T_ε is continuous, which by Lemma 3.1 implies that $f_\varepsilon \in \mathcal{C}^1(\Omega_0)$ with $f_\varepsilon'(x) = T_\varepsilon(x) = \frac{\int_{\mathcal{S}_{x,\varepsilon}} y d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})}$. This weighted average over $\mathcal{S}_{x,\varepsilon}$ is an element of $\mathcal{S}_{x,\varepsilon}$ by convexity, and it cannot lie on the boundary as μ_1 is absolutely continuous. $\square$

4   Second order differentiability

This section investigates the sign and the differentiability properties of the first derivative f_ε'. For brevity, define

$$\xi(x,y) := -f_\varepsilon(x) - g_\varepsilon(y) + xy.$$

For each x, the function $y \mapsto \xi(x,y)$ is concave, so that the set $\mathcal{S}_{x,\varepsilon} = \{y \in \Omega_1 : \xi(x,y) \geq 0\}$ is a compact interval

$$\mathcal{S}_{x,\varepsilon} = [y_m(x), y_M(x)].$$

(The dependence on ε is suppressed in some of our notation). For ease of reference, we state the qualitative result of [37, Theorem 4.1] as follows.

Proposition 4.1. *The support of π_ε converges to the support of π_0 in Hausdorff distance d as $\varepsilon \to 0$. Since the support of π_0 is the graph of the Lipschitz map T_0, it follows that*

$$\lim_{\varepsilon \to 0} \sup_{x \in \Omega_0} d(\mathcal{S}_{x,\varepsilon}, T_0(x)) = 0 \quad \text{and} \quad \lim_{\varepsilon \to 0} \sup_{x \in \Omega_0} |\mathcal{S}_{x,\varepsilon}| = 0.$$

10

We already know that

$$f_\varepsilon'(x) = \frac{\int_{\mathcal{S}_{x,\varepsilon}} y u_1(y) dy}{\mu_1(\mathcal{S}_{x,\varepsilon})}$$

is monotone, hence differentiable a.e. We shall derive an expression for the derivative $f_\varepsilon''(x)$. Note that numerator and denominator are both of the form

$$\Xi(x) = \int_{y_m(x)}^{y_M(x)} h(y) dy \tag{20}$$

for a continuous function h on Ω_1. We first establish that $y_m(x)$ and $y_M(x)$ are continuous functions of x which are differentiable except at those points x where the boundary of $\mathcal{S}_{x,\varepsilon}$ overlaps with the boundary of Ω_1. The continuity follows from the continuity of f_ε and convexity of g_ε (recall that $|\mathcal{S}_{x,\varepsilon}| > 0$ for all $x \in \Omega_0$); cf. Proposition 3.2. Turning to the differentiability, recall from Assumption 2.1 that $\Omega_1 = [a_1, b_1]$. We start by claiming that

$$\{x \in \Omega_0 : y_M(x) = b_1 \text{ and } \xi(x, y_M(x)) = 0\} \quad \text{is at most a singleton, denoted } \{x^{(M)}\},$$
$$\{x \in \Omega_0 : y_m(x) = a_1 \text{ and } \xi(x, y_m(x)) = 0\} \quad \text{is at most a singleton, denoted } \{x^{(m)}\}.$$

By (19), for any $(x, y) \in \Omega_0 \times \Omega_1$,

$$y = f_\varepsilon'(x) \quad \text{implies} \quad y \in (y_m(x), y_M(x)) \subset (a_1, b_1). \tag{21}$$

Fix $y \in \Omega_1$ and suppose there are two points $x_1 < x_2$ in Ω_0 such that $0 = \xi(x_1, y) = \xi(x_2, y)$. Then as $x \mapsto \xi(x, y) = xy - g_\varepsilon(y) - f_\varepsilon(x)$ is differentiable, there exists $x \in (x_1, x_2)$ with

$$0 = \partial_x \xi(x, y) = y - f_\varepsilon'(x).$$

Hence (21) shows $y \neq a_1, b_1$ and both claims follow.

The qualitative convergence result Proposition 4.1 shows $\sup_{x \in \Omega_0} |y_M(x) - y_m(x)| \to 0$ as $\varepsilon \to 0$. In particular, we can pick $\varepsilon_0 > 0$ such that $\sup_{x \in \Omega_0} |y_M(x) - y_m(x)| \leq \frac{1}{3}|\Omega_1|$ for $\varepsilon < \varepsilon_0$, which implies that $\mathcal{S}_{x,\varepsilon}$ can include *at most one* of the two boundary points of Ω_1. Fix $\varepsilon < \varepsilon_0$. We partition $\Omega_0 \setminus \{x^{(m)}, x^{(M)}\}$ into the three (relatively open) sets where $\mathcal{S}_{x,\varepsilon}$ includes the lower, no, or upper boundary point:

$$\Omega_0^{(1)} = \{x \in \Omega_0 : y_M(x) < b_1 \text{ and } \xi(x, y_m(x)) > 0\},$$
$$\Omega_0^{(2)} = \{x \in \Omega_0 : a_1 < y_m(x) < y_M(x) < b_1\},$$
$$\Omega_0^{(3)} = \{x \in \Omega_0 : \xi(x, y_M(x)) > 0 \text{ and } y_m(x) > a_1\}.$$

If $x \in \Omega_0^{(2)} \cup \Omega_0^{(3)}$, then $y_m(x)$ is a zero of

$$y \mapsto \xi(x, y) = xy - g_\varepsilon(y) - f_\varepsilon(x).$$

Since $\partial_y \xi(x, y) = x - g_\varepsilon'(y)$ cannot be zero at $y_m(x)$ by (21), the implicit function theorem yields

$$y_m'(x) = -\frac{y_m(x) - f_\varepsilon'(x)}{x - g_\varepsilon'(y_m(x))}. \tag{22}$$

11

(For $x = x^{(m)}$, this holds only if $y'_m(x)$ is interpreted as the right derivative, whereas the left derivative is zero.) Note that $y_m(x) - f'_\varepsilon(x) < 0$ by (21) whereas $x - g'_\varepsilon(y_m(x)) = \partial_y \xi(x, y_m(x)) > 0$ by concavity. Hence $y'_m(x) > 0$, showing that y_m is strictly increasing on $\Omega_0^{(2)} \cup \Omega_0^{(3)}$. For $x \in \Omega_0^{(1)} \cup \Omega_0^{(2)}$, the same argument yields

$$y'_M(x) = -\frac{y_M(x) - f'_\varepsilon(x)}{x - g'_\varepsilon(y_M(x))} > 0 \tag{23}$$

and strict monotonicity of $x \mapsto y'_M(x)$ on $\Omega_0^{(1)} \cup \Omega_0^{(2)}$. In particular, we have established that y_m, y_M are increasing continuous functions. The definition of the sets $\Omega_0^{(i)}$ then implies that each $\Omega_0^{(i)}$ is an interval.

Next, we check that $\Omega_0^{(i)} \neq \emptyset$ for all i. This is clear for $\Omega_0^{(2)}$ (otherwise, by the choice of ε_0, no mass is transported to the middle third of Ω_1). We claim that $g'_\varepsilon(a_1) \in \Omega_0^{(1)}$. Note first that, by the symmetric counterpart of (19),

$$g'_\varepsilon(a_1) = \frac{\int_{\mathcal{S}_{a_1,\varepsilon}^{(0)}} x \, d\mu_0(x)}{\mu_0(\mathcal{S}_{a_1,\varepsilon}^{(0)})} \in \mathcal{S}_{a_1,\varepsilon}^{(0)} \subset \Omega_0,$$

where $\mathcal{S}_{y,\varepsilon}^{(0)} := \{x \in \Omega_0 : f_\varepsilon(x) + g_\varepsilon(y) - xy \le 0\}$. The concave function $\xi(g'_\varepsilon(a_1), \cdot)$ attains its maximum at any $y \in \Omega_1$ solving $g'_\varepsilon(a_1) - g'_\varepsilon(y) = 0$. Clearly $y = a_1$ satisfies the latter, so that $a_1 \in \mathrm{argmax}_{y \in \Omega_1} \xi(g'_\varepsilon(a_1), y)$. Hence, we must have $\xi(g'_\varepsilon(a_1), a_1) > 0$, for otherwise $\mu_1(\mathcal{S}_{a_1,\varepsilon}^{(0)}) = 0$ contradicting (14). It follows that $y_m(g'_\varepsilon(a_1)) = a_1$ and, as $|\mathcal{S}_{g'_\varepsilon(a_1),\varepsilon}| \le \frac{1}{3}|\Omega_1|$, that $y_M(g'_\varepsilon(a_1)) < b_1$. In summary, $g'_\varepsilon(a_1) \in \Omega_0^{(1)}$. Similarly, $g'_\varepsilon(b_1) \in \Omega_0^{(3)}$.

As $\Omega_0^{(i)} \neq \emptyset$ for all i, it follows that the points $x^{(m)}$ and $x^{(M)}$ indeed exist and lie in the interior of $\Omega_0 = [a_0, b_0]$. Since the union of the three intervals is $\Omega_0 \setminus \{x^{(m)}, x^{(M)}\}$, necessarily

$$\Omega_0^{(1)} = [a_0, x^{(m)}), \qquad \Omega_0^{(2)} = (x^{(m)}, x^{(M)}), \qquad \Omega_0^{(3)} = (x^{(M)}, b_0].$$

The following summarizes our discussion.

Proposition 4.2. *Let $\varepsilon > 0$ be small enough such that $\sup_{x \in \Omega_0} |\mathcal{S}_{x,\varepsilon}| \le \frac{1}{3}|\Omega_1|$. There is exactly one $x \in (a_0, b_0)$ with $y_M(x) = b_1$ and $\xi(x, y_M(x)) = 0$, denoted $x^{(M)}$, and exactly one $x \in (a_0, b_0)$ with $y_m(x) = a_1$ and $\xi(x, y_m(x)) = 0$, denoted $x^{(m)}$. The interval-valued map $x \mapsto \mathcal{S}_{x,\varepsilon} = [y_m(x), y_M(x)]$ is increasing in the sense that*

$$x \mapsto y_m(x) \quad and \quad x \mapsto y_M(x) \quad are\ increasing.$$

More precisely, $y_m(x) = a_1$ for $x \in [a_0, x^{(m)}]$ whereas $x \mapsto y_m(x)$ is strictly increasing on $(x^{(m)}, b_0]$ with derivative (22). Similarly, $y_M(x) = b_1$ for $x \in [x^{(M)}, b_0]$ whereas $x \mapsto y_M(x)$ is strictly increasing on $[a_0, x^{(M)})$ with derivative (23).

We can now take the derivative in (20): Leibniz' rule yields that f' is differentiable on $\Omega_0 \setminus \{x^{(m)}, x^{(M)}\}$ with derivative

$$\Xi'(x) = h(y_m(x)) \frac{f'_\varepsilon(x) - y_m(x)}{x - g'_\varepsilon(y_m(x))} \chi_{\Omega_0^{(2)} \cup \Omega_0^{(3)}}(x) + h(y_M(x)) \frac{f'_\varepsilon(x) - y_M(x)}{x - g'_\varepsilon(y_M(x))} \chi_{\Omega_0^{(1)} \cup \Omega_0^{(2)}}(x).$$

After some calculations that we omit, the following formula is obtained.

Proposition 4.3. *Let $\varepsilon > 0$ be small enough such that $\sup_{x \in \Omega_0} |\mathcal{S}_{x,\varepsilon}| \leq \frac{1}{3}|\Omega_1|$. Then using the notation introduced above,*

$$f''_\varepsilon(x) = u_1(y_m(x)) \frac{(f'_\varepsilon(x) - y_m(x))^2}{(x - g'_\varepsilon(y_m(x)))(y_M(x) - y_m(x))} \chi_{\Omega_0^{(2)} \cup \Omega_0^{(3)}}(x)$$
$$+ u_1(y_M(x)) \frac{(f'_\varepsilon(x) - y_M(x))^2}{(g'_\varepsilon(y_M(x)) - x)(y_M(x) - y_m(x))} \chi_{\Omega_0^{(1)} \cup \Omega_0^{(2)}}(x). \quad (24)$$

Remark 4.4. Of course there is also a regime (for large ε, due to $\pi_\varepsilon \to \mu_0 \otimes \mu_1$ [25, Remark 2.9]) where $\xi(x, \cdot)$ is strictly positive on Ω_1 and in particular $\mathcal{S}_{x',\varepsilon} = \Omega_1$ for x' close to x. Then it is immediate from (19) that $f''_\varepsilon(x) = 0$. (In fact, when π_ε has full support, the functions f_ε and g_ε are affine and can be found in closed form.) In particular, the condition $\varepsilon \leq \varepsilon_0$ is essential to guarantee the strong convexity in Theorem 2.2.

5 Bounds for the second derivative

Recall our notation

$$\xi(x,y) = -f_\varepsilon(x) - g_\varepsilon(y) + xy, \qquad \mathcal{S}_{x,\varepsilon} = \{y \in \Omega_1 : \xi(x,y) \geq 0\} = [y_m(x), y_M(x)].$$

In this section, we bound f''_ε by geometric arguments. First, in Section 5.1, we relate $|\mathcal{S}_{x,\varepsilon}|$ to the maximum of $\xi(x, \cdot)$; i.e., to the maximum of the density of π_ε. In Section 5.2, we show that $f'_\varepsilon(x)$ is approximately at the center of the support $\mathcal{S}_{x,\varepsilon}$. Section 5.3 uses these two auxiliary results to prove a crucial intermediate bound relating $|\mathcal{S}_{x,\varepsilon}|$ with the minimum and the maximum of f''_ε; cf. (30). Finally, we deduce the desired bounds on f''_ε.

5.1 Geometric approach

For each $x \in \Omega_0$, the function $\xi(x, \cdot)$ is concave, behaving roughly like a parabola open to the bottom. Our starting point is the following simple observation, relating the height $\max_{y \in \Omega_1} \xi(x, y)$ to the size $|\mathcal{S}_{x,\varepsilon}|$ of the support of $(\xi(x, \cdot))_+$.

Lemma 5.1. *There exists $C > 0$ such that*

$$C^{-1}\varepsilon \leq \max_{y \in \Omega_1} \xi(x, y)|\mathcal{S}_{x,\varepsilon}| \leq C\varepsilon$$

for all $x \in \Omega_0$ and $\varepsilon > 0$.

Proof. Using (14) and Assumption 2.1,

$$\varepsilon = \int (\xi(x,y))_+ d\mu_1(y) \leq \Lambda \int_{\Omega_1} (\xi(x,y))_+ dy \leq \Lambda |\mathcal{S}_{x,\varepsilon}| \max_{y \in \Omega_1} \xi(x, y),$$

giving the lower bound. On the other hand, as $y \mapsto \xi(x,y)$ is concave, its graph lies above the graph of the piecewise affine function $\ell(y)$ interpolating the three points

$$(y_m(x), \xi(x, y_m(x))), \qquad \left(\operatorname*{argmax}_{y \in \Omega_1} \xi(x, y), \max_{y \in \Omega_1} \xi(x, y_m(x)) \right), \qquad (y_M(x), \xi(x, y_M(x))),$$

which, in turn, lies above the "tent" interpolating the points

$$(y_m(x), 0), \qquad \left(\operatorname*{argmax}_{y \in \Omega_1} \xi(x, y), \max_{y \in \Omega_1} \xi(x, y_m(x)) \right), \qquad (y_M(x), 0),$$

by the definition of $y_m(x), y_M(x)$. The integral of the latter tent corresponds to the area of a triangle, giving

$$\varepsilon = \int (\xi(x, y))_+ d\mu_1(y) \geq \lambda \int_{\Omega_1} (\xi(x, y))_+ dy \geq \lambda \int_{\Omega_1} (\ell(y))_+ dy \geq \frac{\lambda}{2} |\mathcal{S}_{x,\varepsilon}| \max_{y \in \Omega_1} \xi(x, y)$$

after recalling that $|\mathcal{S}_{x,\varepsilon}| = y_M(x) - y_m(x)$. This establishes the upper bound. $\qquad\square$

In view of Lemma 5.1, our goal is to control the ratio between $\max_{y \in \Omega_1} \xi(x, y)$ and $\mu_1(\mathcal{S}_{x,\varepsilon})$ (or equivalently $|\mathcal{S}_{x,\varepsilon}|$). Clearly this ratio is closely related to the second derivative of $\xi(x, \cdot)$, or equivalently g_ε''. The following result shows the equivalence of $\max_{y \in \Omega_1} \xi(x, y)$ and $\max_{x' \in \Omega_0} \xi(x', f_\varepsilon'(x))$, which will be important for controlling f_ε'' and g_ε'' later on.

Lemma 5.2. *There exists $C > 0$ such that for every $x \in \Omega_0$,*

$$C^{-1} \max_{y \in \Omega_1} \xi(x, y) \leq \xi(x, f_\varepsilon'(x)) = \max_{x' \in \Omega_0} \xi(x', f_\varepsilon'(x)) \leq \max_{y \in \Omega_1} \xi(x, y).$$

Proof. We can write (14) as

$$\int_{\mathcal{S}_{x,\varepsilon}} xy d\mu_1(y) - f_\varepsilon(x) \mu_1(\mathcal{S}_{x,\varepsilon}) - \int_{\mathcal{S}_{x,\varepsilon}} g_\varepsilon(y) d\mu_1(y) = \varepsilon.$$

Therefore,

$$\frac{\int_{\mathcal{S}_{x,\varepsilon}} xy d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})} - f_\varepsilon(x) - \frac{\int_{\mathcal{S}_{x,\varepsilon}} g_\varepsilon(y) d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})} = \frac{\varepsilon}{\mu_1(\mathcal{S}_{x,\varepsilon})}.$$

Recalling from Proposition 3.2 the formula $f_\varepsilon'(x) = \frac{\int_{\mathcal{S}_{x,\varepsilon}} y d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})}$, we deduce

$$x f_\varepsilon'(x) - f_\varepsilon(x) - \frac{\int_{\mathcal{S}_{x,\varepsilon}} g_\varepsilon(y) d\mu_1(y)}{\mu_1(\mathcal{S}_{x,\varepsilon})} = \frac{\varepsilon}{\mu_1(\mathcal{S}_{x,\varepsilon})}.$$

As g_ε is convex, Jensen's inequality and the aforementioned formula yield

$$\frac{\varepsilon}{\mu_1(\mathcal{S}_{x,\varepsilon})} \leq -f_\varepsilon(x) - g_\varepsilon(f_\varepsilon'(x)) + x f_\varepsilon'(x) = \xi(x, f_\varepsilon'(x)),$$

and now the first inequality follows from Lemma 5.1. The second inequality follows trivially, whereas the equality holds by the first-order condition: $y := f_\varepsilon'(x)$ satisfies $\frac{d}{dx} \xi(x, y) = 0$. $\quad\square$

5.2 Distance of $f_\varepsilon'(x)$ to the boundary of $\mathcal{S}_{x,\varepsilon}$

Proposition 3.2 shows that $f_\varepsilon'(x) \in \mathcal{S}_{x,\varepsilon}$. The following result states that the distance between $f_\varepsilon'(x)$ and either of the two boundary points $y_m(x), y_M(x)$ of $\mathcal{S}_{x,\varepsilon}$ is of the same order as $|\mathcal{S}_{x,\varepsilon}| = y_M(x) - y_m(x)$, and more precisely that $f_\varepsilon'(x)$ is located in the center half of $\mathcal{S}_{x,\varepsilon}$ for small ε.

Lemma 5.3. *There exists $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$ and $x \in \Omega_0$,*

$$\frac{1}{4}|\mathcal{S}_{x,\varepsilon}| \leq y_M(x) - f_\varepsilon'(x) \leq |\mathcal{S}_{x,\varepsilon}|,$$

$$\frac{1}{4}|\mathcal{S}_{x,\varepsilon}| \leq f_\varepsilon'(x) - y_m(x) \leq |\mathcal{S}_{x,\varepsilon}|.$$

Proof. The upper bounds are trivial due to $f_\varepsilon'(x) \in \mathcal{S}_{x,\varepsilon}$. Fix $x \in \Omega_0$ and define the mid point of $\mathcal{S}_{x,\varepsilon}$ as

$$m(x) := \frac{\int_{\mathcal{S}_{x,\varepsilon}} y \, dy}{|\mathcal{S}_{x,\varepsilon}|} = \frac{y_M(x) + y_m(x)}{2}.$$

Let $0 < \eta < 1$. As u_1 is uniformly continuous on Ω_1 and $u_1 \geq \lambda > 0$, and as $\sup_{x \in \Omega_0} |\mathcal{S}_{x,\varepsilon}| \to 0$ by Proposition 4.1, there exists ε_0 (independent of x) such that $\varepsilon \leq \varepsilon_0$ implies

$$1 - \eta \leq \frac{u_1(y)}{u_1(y')} \leq 1 + \eta \quad \text{for all} \quad y, y' \in \mathcal{S}_{x,\varepsilon}. \tag{25}$$

For any bounded function $h(y)$, it follows that

$$\left| \frac{1}{\mu_1(\mathcal{S}_{x,\varepsilon})} \int_{\mathcal{S}_{x,\varepsilon}} h(y) d\mu_1(y) - \frac{1}{|\mathcal{S}_{x,\varepsilon}|} \int_{\mathcal{S}_{x,\varepsilon}} h(y) dy \right| \leq \frac{\eta}{|\mathcal{S}_{x,\varepsilon}|} \int_{\mathcal{S}_{x,\varepsilon}} |h(y)| dy.$$

Using this with $h(y) := y - m(x)$, the formula $f_\varepsilon'(x) = \frac{1}{\mu_1(\mathcal{S}_{x,\varepsilon})} \int_{\mathcal{S}_{x,\varepsilon}} y \, d\mu_1(y)$ of Proposition 3.2 yields

$$|f_\varepsilon'(x) - m(x)| = \left| \frac{1}{\mu_1(\mathcal{S}_{x,\varepsilon})} \int_{\mathcal{S}_{x,\varepsilon}} (y - m(x)) d\mu_1(y) - \frac{1}{|\mathcal{S}_{x,\varepsilon}|} \int_{\mathcal{S}_{x,\varepsilon}} (y - m(x)) dy \right|$$

$$\leq \frac{\eta}{|\mathcal{S}_{x,\varepsilon}|} \int_{\mathcal{S}_{x,\varepsilon}} |y - m(x)| dy \leq \frac{\eta}{2|\mathcal{S}_{x,\varepsilon}|} \int_{\mathcal{S}_{x,\varepsilon}} |\mathcal{S}_{x,\varepsilon}| dy = \frac{\eta}{2} |\mathcal{S}_{x,\varepsilon}|.$$

(We remark that this conclusion holds for all $\varepsilon > 0$, that is with $\varepsilon_0 = \infty$, if (25) holds everywhere on Ω_1.) The stated lower bounds are derived via the reverse triangle inequality

$$f_\varepsilon'(x) - y_m(x) \geq |m(x) - y_m(x)| - |f_\varepsilon'(x) - m(x)|$$

$$= \frac{|y_M(x) - y_m(x)|}{2} - |f_\varepsilon'(x) - m(x)| \geq \frac{1 - \eta}{2} |y_M(x) - y_m(x)|$$

by taking $\eta = 1/2$, and similarly for $y_M(x)$. $\qquad\square$

5.3 Bounds for the second derivative

We now proceed to bound the second derivative f_ε'' in terms of the size of the support. Recall that $\mathcal{S}_{y,\varepsilon}^{(0)} := \{x \in \Omega_0 : f_\varepsilon(x) + g_\varepsilon(y) - xy \leq 0\}$. Next, we use the symmetric version of Lemma 5.3 with exchanged roles of the marginals: for $\varepsilon < \varepsilon_0$ and $y \in \Omega_1$,

$$\frac{1}{4}|\mathcal{S}_{y,\varepsilon}^{(0)}| \leq |g_\varepsilon'(y) - x_M(y)| \leq |\mathcal{S}_{y,\varepsilon}^{(0)}|.$$

15

Specifically, we fix $x \in \Omega_0$ and use this for $y = y_M(x)$. Noting that $x_M(y_M(x)) = x$, we obtain

$$\frac{1}{4}|\mathcal{S}^{(0)}_{y_M(x),\varepsilon}| \leq |g'_\varepsilon(y_M(x)) - x| \leq |\mathcal{S}^{(0)}_{y_M(x),\varepsilon}|.$$

Similarly,

$$\frac{1}{4}|\mathcal{S}^{(0)}_{y_m(x),\varepsilon}| \leq |g'_\varepsilon(y_m(x)) - x| \leq |\mathcal{S}^{(0)}_{y_m(x),\varepsilon}|.$$

We now plug these estimates, as well as the estimates stated in Lemma 5.3, into the formula (24) for $f''_\varepsilon(x)$. Recalling from Section 4 that

$$g'_\varepsilon(y_M(x)) - x > 0 \text{ for all } x \in \Omega_0^{(1)} \cup \Omega_0^{(2)}, \qquad g'_\varepsilon(y_m(x)) - x < 0 \text{ for all } x \in \Omega_0^{(3)} \cup \Omega_0^{(2)},$$

we obtain for all $x \in \Omega_0 \setminus \{x^{(m)}, x^{(M)}\}$ that

$$f''_\varepsilon(x) \leq C\Lambda \left(\frac{|\mathcal{S}_{x,\varepsilon}|}{\left|\mathcal{S}^{(0)}_{y_m(x),\varepsilon}\right|} \chi_{\Omega_0^{(3)}\cup\Omega_0^{(2)}}(x) + \frac{|\mathcal{S}_{x,\varepsilon}|}{\left|\mathcal{S}^{(0)}_{y_M(x),\varepsilon}\right|} \chi_{\Omega_0^{(3)}\cup\Omega_0^{(2)}}(x) \right) \tag{26}$$

and

$$f''_\varepsilon(x) \geq \frac{\lambda}{C} \left(\frac{|\mathcal{S}_{x,\varepsilon}|}{\left|\mathcal{S}^{(0)}_{y_m(x),\varepsilon}\right|} \chi_{\Omega_0^{(3)}\cup\Omega_0^{(2)}}(x) + \frac{|\mathcal{S}_{x,\varepsilon}|}{\left|\mathcal{S}^{(0)}_{y_M(x),\varepsilon}\right|} \chi_{\Omega_0^{(3)}\cup\Omega_0^{(2)}}(x) \right). \tag{27}$$

From these two bounds and Lemma 5.1, we see that

$$\sigma_m(f_\varepsilon) := \inf_{x \in \Omega_0 \setminus \{x^{(m)}, x^{(M)}\}} f''_\varepsilon(x) > 0,$$

$$\sigma_M(f_\varepsilon) := \sup_{x \in \Omega_0 \setminus \{x^{(m)}, x^{(M)}\}} f''_\varepsilon(x) < +\infty.$$

We define $\sigma_m(g_\varepsilon)$ and $\sigma_M(g_\varepsilon)$ in a symmetric manner. Next, we prove that these four quantities are upper and lower bounded uniformly in ε. Recall from (15) that

$$\varepsilon = \int (\xi(x', f'_\varepsilon(x)))_+ d\mu_0(x') \geq \lambda \int_{a_0}^{b_0} (\xi(x', f'_\varepsilon(x)))_+ dx'. \tag{28}$$

Define the convex conjugate $f^*_\varepsilon(y) := \sup_{x \in \Omega_0} \{xy - f_\varepsilon(x)\}$. A second order Taylor development of $f_\varepsilon(x')$ around x gives

$$\begin{aligned}
\xi(x', f'_\varepsilon(x)) &= -f_\varepsilon(x') - g_\varepsilon(f'_\varepsilon(x)) + x'f'_\varepsilon(x) \\
&= [f^*_\varepsilon(f'_\varepsilon(x)) - g_\varepsilon(f'_\varepsilon(x))] - [f_\varepsilon(x') + f^*_\varepsilon(f'_\varepsilon(x)) - x'f'_\varepsilon(x)] \\
&\geq [f^*_\varepsilon(f'_\varepsilon(x)) - g_\varepsilon(f'_\varepsilon(x))] - \frac{\sigma_M(f_\varepsilon)(x - x')^2}{2}.
\end{aligned} \tag{29}$$

Call $t = [f^*_\varepsilon(f'_\varepsilon(x)) - g_\varepsilon(f'_\varepsilon(x))]$. As above, the qualitative convergence result Proposition 4.1 yields $\varepsilon_0 > 0$ such that $\sup_{y \in \Omega_1} |\mathcal{S}^{(0)}_{y,\varepsilon}| \leq \frac{1}{2}|\Omega_1|$ for $\varepsilon < \varepsilon_0$, and hence at least one of the

conditions $a_0 \notin \mathcal{S}_{y,\varepsilon}^{(0)}$ or $b_0 \notin \mathcal{S}_{y,\varepsilon}^{(0)}$ must hold for any y. Assume w.l.o.g. that $b_0 \notin \mathcal{S}_{y,\varepsilon}^{(0)}$. Returning to (28) and applying (29) to the integrand, we deduce

$$
\begin{aligned}
\varepsilon &\geq \lambda \int_x^\infty (\xi(x', f_\varepsilon'(x)))_+ dx' \\
&\geq \lambda \int_x^\infty \left(t - \frac{\sigma_M(f_\varepsilon)(x-x')^2}{2} \right)_+ dx' \\
&= \lambda \sigma_M(f_\varepsilon) \int_0^\infty \left(\frac{t}{\sigma_M(f_\varepsilon)} - \frac{u^2}{2} \right)_+ du \\
&= \frac{2\sqrt{2}}{3} \lambda \sigma_M(f_\varepsilon) \left(\frac{t}{\sigma_M(f_\varepsilon)} \right)^{3/2} = \frac{2\sqrt{2}}{3} \lambda (\sigma_M(f_\varepsilon))^{-1/2} t^{3/2}.
\end{aligned}
$$

By the same means, $\varepsilon \leq \frac{4\sqrt{2}}{3}\Lambda(\sigma_m(f_\varepsilon))^{-1/2}t^{3/2}$. In summary, for all $\varepsilon < \varepsilon_0$ and $x \in \Omega_0$,

$$
\frac{2\sqrt{2}}{3}\lambda(\sigma_M(f_\varepsilon))^{-1/2}t^{3/2} \leq \varepsilon \leq \frac{4\sqrt{2}}{3}\Lambda(\sigma_m(f_\varepsilon))^{-1/2}t^{3/2}.
$$

Here $t = f_\varepsilon^*(f_\varepsilon'(x)) - g_\varepsilon(f_\varepsilon'(x)) = \max_{x' \in \Omega_0} \xi(x', f_\varepsilon'(x))$ due to $\partial_{x'}\xi(x', f_\varepsilon'(x)) = f_\varepsilon'(x) - f_\varepsilon'(x')$. Recalling also from Lemma 5.2 that

$$
C^{-1} \max_{y \in \Omega_1} \xi(x,y) \leq \max_{x' \in \Omega_0} \xi(x', f_\varepsilon'(x)) \leq \max_{y \in \Omega_1} \xi(x,y) \quad \text{for all} \quad x \in \Omega_0,
$$

and allowing C to vary from line to line, we deduce

$$
C^{-1}(\sigma_M(f_\varepsilon))^{-1/2} \max_{y \in \Omega_1}(\xi(x,y))_+^{3/2} \leq \varepsilon \leq C(\sigma_m(f_\varepsilon))^{-1/2} \max_{y \in \Omega_1}(\xi(x,y))_+^{3/2}.
$$

In view of Lemma 5.1, this implies

$$
C^{-1}(\sigma_M(f_\varepsilon))^{-1/2}|\mathcal{S}_{x,\varepsilon}|^{-3/2} \leq \varepsilon^{-1/2} \leq C(\sigma_m(f_\varepsilon))^{-1/2}|\mathcal{S}_{x,\varepsilon}|^{-3/2},
$$

and after rearranging we conclude

$$
C^{-1}\left(\frac{\varepsilon}{\sigma_M(f_\varepsilon)} \right)^{\frac{1}{3}} \leq |\mathcal{S}_{x,\varepsilon}| \leq C\left(\frac{\varepsilon}{\sigma_m(f_\varepsilon)} \right)^{\frac{1}{3}}. \tag{30}
$$

We observe that Theorem 2.5 will follow from (30) once uniform bounds for $\sigma_M(f_\varepsilon)$ and $\sigma_m(f_\varepsilon)$ are derived. To established the latter, we plug (30) into (26) and (27) to obtain the following bounds for all $x \notin \{x^{(m)}, x^{(M)}\}$:

$$
f_\varepsilon''(x) \leq C\left(\frac{\sigma_M(g_\varepsilon)}{\sigma_m(f_\varepsilon)} \right)^{\frac{1}{3}} \quad \text{and} \quad f_\varepsilon''(x) \geq \frac{1}{C}\left(\frac{\sigma_m(g_\varepsilon)}{\sigma_M(f_\varepsilon)} \right)^{\frac{1}{3}}.
$$

By taking the supremum in the first expression and the infimum in the second, we deduce

$$
\sigma_M(f_\varepsilon) \leq C\left(\frac{\sigma_M(g_\varepsilon)}{\sigma_m(f_\varepsilon)} \right)^{\frac{1}{3}} \quad \text{and} \quad \sigma_m(f_\varepsilon) \geq \frac{1}{C}\left(\frac{\sigma_m(g_\varepsilon)}{\sigma_M(f_\varepsilon)} \right)^{\frac{1}{3}}. \tag{31}
$$

17

Symmetrically,

$$\sigma_M(g_\varepsilon) \leq C \left(\frac{\sigma_M(f_\varepsilon)}{\sigma_m(g_\varepsilon)} \right)^{\frac{1}{3}} \quad \text{and} \quad \sigma_m(g_\varepsilon) \geq \frac{1}{C} \left(\frac{\sigma_m(f_\varepsilon)}{\sigma_M(g_\varepsilon)} \right)^{\frac{1}{3}}. \tag{32}$$

Plugging (32) into (31) yields

$$\sigma_M(f_\varepsilon) \leq C \left(\frac{\frac{\sigma_M(f_\varepsilon)}{\sigma_m(g_\varepsilon)}}{\frac{\sigma_m(g_\varepsilon)}{\sigma_M(f_\varepsilon)}} \right)^{\frac{1}{9}} = C \left(\frac{\sigma_M(f_\varepsilon)}{\sigma_m(g_\varepsilon)} \right)^{\frac{2}{9}}$$

and

$$\sigma_m(g_\varepsilon) \geq \frac{1}{C} \left(\frac{\frac{\sigma_m(g_\varepsilon)}{\sigma_M(f_\varepsilon)}}{\frac{\sigma_M(f_\varepsilon)}{\sigma_m(g_\varepsilon)}} \right)^{\frac{1}{9}} = \frac{1}{C} \left(\frac{\sigma_m(g_\varepsilon)}{\sigma_M(f_\varepsilon)} \right)^{\frac{2}{9}},$$

which implies

$$\sigma_M(f_\varepsilon) \leq C \left(\frac{1}{\sigma_m(g_\varepsilon)} \right)^{\frac{2}{7}} \quad \text{and} \quad \sigma_m(g_\varepsilon) \geq \frac{1}{C} \left(\frac{1}{\sigma_M(f_\varepsilon)} \right)^{\frac{2}{7}}$$

and finally

$$\sigma_M(f_\varepsilon) \leq C \sigma_M(f_\varepsilon)^{\frac{4}{49}} \quad \text{and} \quad \sigma_m(g_\varepsilon) \geq C^{-1} \sigma_m(g_\varepsilon)^{\frac{4}{49}}$$

for some $C > 0$. It follows that $\sigma_M(f_\varepsilon)$ and $\sigma_m(g_\varepsilon)$ are uniformly upper and lower bounded, respectively. Symmetric statements hold for $\sigma_M(g_\varepsilon)$ and $\sigma_m(f_\varepsilon)$, completing the proof of Theorem 2.2. Moreover, Theorem 2.5 now follows from (30).

6 Convergence to unregularized optimal transport

The goal of this section is to prove Theorem 2.6 and its corollary. We first prove a rate for the barycentric projection of π_ε.

Lemma 6.1. *Let $S_\varepsilon(x)$ be the barycentric projection of π_ε; that is,*

$$S_\varepsilon(x) := \int y \left(\frac{\xi(x,y)}{\varepsilon} \right)_+ d\mu_1(y).$$

There exists $C > 0$ such that for all $\varepsilon \in (0,1]$,

$$\|T_0 - S_\varepsilon\|_{L^2(\Omega_0, dx)} \leq C \varepsilon^{\frac{1}{3}}.$$

The argument follows [8, Section 4.2] where a similar result is shown in the context of entropic optimal transport.

Proof. Recall that (f_0, g_0) denote the potentials for the unregularized optimal transport problem, which are convex conjugates of one another. Since g_0 is convex and $\varepsilon^{-1}(\xi(x,y))_+ d\mu_1(y)$ is a probability measure for each $x \in \Omega_0$ by (14), Jensen's inequality implies that

$$\int g_0(y) \left(\frac{\xi(x,y)}{\varepsilon} \right)_+ d\mu_1(y) \geq g_0 \left(\int y \left(\frac{\xi(x,y)}{\varepsilon} \right)_+ d\mu_1(y) \right) = g_0(S_\varepsilon(x)).$$

18

This yields the inequality in

$$0 = \int g_0(y)d\pi_\varepsilon(x,y) - \int g_0(y)d\mu_1(y)$$

$$= \int \left(\int g_0(y) \left(\frac{\xi(x,y)}{\varepsilon} \right)_+ d\mu_1(y) \right) d\mu_0(x) - \int g_0(T_0(x))d\mu_0(x)$$

$$\geq \int g_0\left(S_\varepsilon(x)\right)d\mu_0(x) - \int g_0(T_0(x))d\mu_0(x).$$

As g_0 is strongly convex,

$$g_0\left(S_\varepsilon(x)\right) - g_0(T_0(x)) \geq (S_\varepsilon(x) - T_0(x))g_0'(T_0(x)) + C^{-1}(S_\varepsilon(x) - T_0(x))^2$$

for some $C > 0$ and we deduce

$$0 \geq \int g_0'(T_0(x))\left(S_\varepsilon(x) - T_0(x)\right)d\mu_0(x) + \lambda C^{-1} \|S_\varepsilon - T_0\|^2_{L^2(\Omega_0,dx)}.$$

In view of $g_0'(T_0(x)) = x$, it follows that

$$\lambda C^{-1}\|S_\varepsilon(x) - T_0\|^2_{L^2(\Omega_0,dx)} \leq \int (xT_0(x) - xS_\varepsilon(x))\,d\mu_0(x)$$

$$= \int xT_0(x)d\mu_0(x) - \int x \int y \left(\frac{\xi(x,y)}{\varepsilon} \right)_+ d\mu_1(y)d\mu_0(x)$$

$$= \int xT_0(x)d\mu_0(x) - \int xy\,d\pi_\varepsilon(x,y)$$

$$= \frac{1}{2}\int(x-y)^2 d\pi_\varepsilon(x,y) - \frac{1}{2}\int(x-T_0(x))^2 d\mu_0(x).$$

The last line is the difference between the transport cost of π_ε and the optimal transport cost. As the quadratic penalty is nonnegative, that difference is dominated by the difference between the optimal regularized transport cost and the optimal transport cost. By [18, Corollary 3.14], the latter satisfies

$$\frac{1}{2}\int(x-y)^2 d\pi_\varepsilon(x,y) - \frac{1}{2}\int(x-T_0(x))^2 d\mu_0(x) + \varepsilon \left\| \frac{d\pi_\varepsilon}{d(\mu_0 \otimes \mu_1)} \right\|^2_{L^2(\mu_0 \otimes \mu_1)} \leq C\varepsilon^{\frac{2}{3}}$$

for all $\varepsilon \in (0,1]$, for some $C > 0$. (This rate is sharp by [18, Proposition 4.4].) As a result, $\|S_\varepsilon(x) - T_0\|^2_{L^2(\Omega_0,dx)} \leq C\varepsilon^{\frac{2}{3}}$, which was the claim. $\qquad\square$

Proof of Theorem 2.6. By Proposition 3.2 we have $f_\varepsilon'(x) \in \mathcal{S}_{x,\varepsilon}$ for all $x \in \Omega_0$. Thus

$$\|f_\varepsilon'(x) - T_0(x)\|_{L^2(\Omega_0,dx)} \leq \|d\left(\mathcal{S}_{x,\varepsilon}, T_0(x)\right)\|_{L^2(\Omega_0,dx)}$$

and hence the second claim in Theorem 2.6 follows from the first claim, namely that

$$C^{-1}\varepsilon^{\frac{1}{3}} \leq \|d\left(\mathcal{S}_{x,\varepsilon}, T_0(x)\right)\|_{L^2(\Omega_0,dx)} \leq C\varepsilon^{\frac{1}{3}}.$$

Note that the lower bound already follows from Theorem 2.5. Whereas for the upper bound, in view of Theorem 2.5, it suffices to exhibit measurable functions $\{S_\varepsilon\}_{\varepsilon > 0}$ such that $S_\varepsilon(x) \in \mathcal{S}_{x,\varepsilon}$ for all $x \in \Omega_0$ and $\|S_\varepsilon - T_0\|_{L^2(\Omega_0,dx)} \leq C\varepsilon^{\frac{1}{3}}$ for all $\varepsilon \leq \varepsilon_0$. By Lemma 6.1, the barycentric projection has these properties. $\qquad\square$

19

Proof of Corollary 2.7. By (11) we have

$$\|f'_\varepsilon - f'_0\|_{L^2(\Omega_0, dx)} \leq C\varepsilon^{\frac{1}{3}}. \tag{33}$$

Let

$$\alpha_\varepsilon := \frac{1}{|\Omega_0|} \int_{\Omega_0} (f_\varepsilon(x) - f_0(x)) dx.$$

Poincaré's inequality [20, Theorem 1, p. 275] applied to (33) shows

$$\|f_\varepsilon - f_0 - \alpha_\varepsilon\|_{L^2(\Omega_0, dx)} \leq C\|f'_\varepsilon - f'_0\|_{L^2(\Omega_0, dx)} \leq C\varepsilon^{\frac{1}{3}} \tag{34}$$

where the constant C is allowed to change but is independent of ε. Let

$$\beta_\varepsilon := \int (f_\varepsilon(x) - f_0(x))\mu_0(x).$$

The mean α_ε under the Lebesgue measure is not directly comparable to the mean β_ε under μ_0. However, using the variational characterization of the mean and Assumption 2.1,

$$\|f_\varepsilon - f_0 - \beta_\varepsilon\|_{L^2(\mu_0)} = \inf_{\alpha \in \mathbb{R}} \|f_\varepsilon - f_0 - \alpha\|_{L^2(\mu_0)} \leq C \inf_{\alpha \in \mathbb{R}} \|f_\varepsilon - f_0 - \alpha\|_{L^2(\Omega_0, dx)}$$

which in conjunction with (34) yields

$$\|f_\varepsilon - f_0 - \beta_\varepsilon\|_{L^2(\mu_0)} \leq C\varepsilon^{\frac{1}{3}}. \tag{35}$$

We argue below that

$$|\beta_\varepsilon| \leq C\varepsilon^{\frac{1}{3}}. \tag{36}$$

Then, (35) implies $\|f_\varepsilon - f_0\|_{L^2(\mu_0)} \leq C\varepsilon^{\frac{1}{3}}$, and thus $\|f_\varepsilon - f_0\|_{L^2(\Omega_0, dx)} \leq C\varepsilon^{\frac{1}{3}}$ by Assumption 2.1. Together with (33), we now have

$$\|f_\varepsilon - f_0\|_{W^{1,2}(\Omega_0)} \leq C\varepsilon^{\frac{1}{3}},$$

so that Morrey's inequality [20, Theorem 4, p. 266] yields the claim.

It remains to show (36). In fact, we prove the stronger claim

$$2|\beta_\varepsilon| = \left| \int (f_\varepsilon(x) - f_0(x))\mu_0(x) + \int (g_\varepsilon(y) - g_0(y))\mu_1(y) \right| \leq C\varepsilon^{\frac{2}{3}} \tag{37}$$

where the equality is due to (12). Since π_0 is optimal for the transport problem, (2) implies

$$\frac{\varepsilon}{2} \left\| \frac{d\pi_\varepsilon}{d(\mu_0 \otimes \mu_1)} \right\|^2_{L^2(\mu_0 \otimes \mu_1)} \leq \mathsf{QOT}_\varepsilon - \mathsf{OT}. \tag{38}$$

Next, we use the strong duality of QOT_ε (e.g., [37, Theorem 2.2]). The dual expression (4) for QOT_ε (taking into account the change of variables (6)) and the analogous dual of OT yield

$$\mathsf{QOT}_\varepsilon - \mathsf{OT} = \int (f_0(x) - f_\varepsilon(x))\mu_0(x) + \int (g_0(y) - g_\varepsilon(y))\mu_1(y) - \frac{\varepsilon}{2} \left\| \frac{d\pi_\varepsilon}{d(\mu_0 \otimes \mu_1)} \right\|^2_{L^2(\mu_0 \otimes \mu_1)}.$$

In view of (38), this implies

$$\mathsf{QOT}_\varepsilon - \mathsf{OT} \leq \int (f_0(x) - f_\varepsilon(x))\mu_0(x) + \int (g_0(y) - g_\varepsilon(y))\mu_1(y) \leq 2(\mathsf{QOT}_\varepsilon - \mathsf{OT}).$$

The claim (37) follows as $C^{-1}\varepsilon^{\frac{2}{3}} \leq \mathsf{QOT}_\varepsilon - \mathsf{OT} \leq C\varepsilon^{\frac{2}{3}}$ by [18, Corollary 3.14 and Proposition 4.4]. □

References

[1] J. M. Altschuler, J. Niles-Weed, and A. J. Stromme. Asymptotics for semidiscrete entropic optimal transport. *SIAM J. Math. Anal.*, 54(2):1718–1741, 2022.

[2] E. Bayraktar, S. Eckstein, and X. Zhang. Stability and sample complexity of divergence regularized optimal transport. *Preprint arXiv:2212.00367v1*, 2022.

[3] R. J. Berman. The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampère equations. *Numer. Math.*, 145(4):771–836, 2020.

[4] E. Bernton, P. Ghosal, and M. Nutz. Entropic optimal transport: Geometry and large deviations. *Duke Math. J.*, 171(16):3363–3400, 2022.

[5] M. Blondel, V. Seguy, and A. Rolet. Smooth and sparse optimal transport. volume 84 of *Proceedings of Machine Learning Research*, pages 880–889, 2018.

[6] L. A. Caffarelli. Boundary regularity of maps with convex potentials. II. *Ann. of Math. (2)*, 144(3):453–496, 1996.

[7] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

[8] G. Carlier, P. Pegon, and L. Tamanini. Convergence rate of general entropic optimal transport costs. *Calc. Var. Partial Differential Equations*, 62(4):Paper No. 116, 28, 2023.

[9] S. Chen, J. Liu, and X.-J. Wang. Global regularity for the Monge-Ampère equation with natural boundary condition. *Ann. of Math. (2)*, 194(3):745–793, 2021.

[10] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *J. Optim. Theory Appl.*, 169(2):671–691, 2016.

[11] S. Chewi and A.-A. Pooladian. An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *C. R. Math. Acad. Sci. Paris*, 361:1471–1483, 2023.

[12] R. Cominetti and J. San Martín. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Math. Programming*, 67(2, Ser. A):169–187, 1994.

[13] G. Conforti and L. Tamanini. A formula for the time derivative of the entropic cost and applications. *J. Funct. Anal.*, 280(11):108964, 2021.

[14] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26*, pages 2292–2300. 2013.

[15] E. del Barrio, A. González-Sanz, and J.-M. Loubes. Central limit theorems for general transportation costs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 60(2):847–873, 2024.

[16] A. Dessein, N. Papadakis, and J.-L. Rouas. Regularized optimal transport and the rot mover's distance. *J. Mach. Learn. Res.*, 19(15):1–53, 2018.

[17] S. Eckstein and M. Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Appl. Math. Optim.*, 83(2):639–667, 2021.

[18] S. Eckstein and M. Nutz. Convergence rates for regularized optimal transport via quantization. *Math. Oper. Res.*, 49(2):1223–1240, 2024.

[19] M. Essid and J. Solomon. Quadratically regularized optimal transport on graphs. *SIAM J. Sci. Comput.*, 40(4):A1961–A1986, 2018.

[20] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[21] A. Garriz-Molina, A. González-Sanz, and G. Mordant. Infinitesimal behavior of quadratically regularized optimal transport and its relation with the porous medium equation. *Preprint arXiv:2407.21528v1*, 2024.

[22] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré. Sample complexity of Sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1574–1583. PMLR, 2019.

[23] A. Genevay, M. Cuturi, G. Peyré, and F. Bach. Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems 29*, pages 3440–3448, 2016.

[24] P. Ghosal, M. Nutz, and E. Bernton. Stability of entropic optimal transport and Schrödinger bridges. *J. Funct. Anal.*, 283(9):Paper No. 109622, 2022.

[25] A. González-Sanz and M. Nutz. Quantitative convergence of quadratically regularized linear programs. *Preprint arXiv:2408.04088v1*, 2024.

[26] A. González-Sanz, M. Nutz, and A. Riveros Valdevenito. Monotonicity in quadratically regularized linear programs. *Preprint arXiv:2408.07871v1*, 2024.

[27] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of Wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 5769–5779, 2017.

[28] C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(4):1879–1920, 2012.

[29] L. Li, A. Genevay, M. Yurochkin, and J. Solomon. Continuous regularized Wasserstein barycenters. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17755–17765. Curran Associates, Inc., 2020.

[30] D. Lorenz and H. Mahler. Orlicz space regularization of continuous optimal transport problems. *Appl. Math. Optim.*, 85(2):Paper No. 14, 33, 2022.

[31] D. Lorenz, P. Manns, and C. Meyer. Quadratically regularized optimal transport. *Appl. Math. Optim.*, 83(3):1919–1949, 2021.

[32] O. L. Mangasarian and R. R. Meyer. Nonlinear perturbation of linear programs. *SIAM J. Control Optim.*, 17(6):745–752, 1979.

[33] G. Mena and J. Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems 32*, pages 4541–4551. 2019.

[34] T. Mikami. Optimal control for absolutely continuous stochastic processes and the mass transportation problem. *Electron. Comm. Probab.*, 7:199–213, 2002.

[35] T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields*, 129(2):245–260, 2004.

[36] M. Nutz. *Introduction to Entropic Optimal Transport.* Lecture notes, Columbia University, 2021. https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.

[37] M. Nutz. Quadratically regularized optimal transport: Existence and multiplicity of potentials. *Preprint arXiv:2404.06847v1*, 2024.

[38] M. Nutz and J. Wiesel. Entropic optimal transport: convergence of potentials. *Probab. Theory Related Fields*, 184(1-2):401–424, 2022.

[39] S. Pal. On the difference between entropic cost and the optimal transport cost. *Ann. Appl. Probab.*, 34(1B):1003–1028, 2024.

[40] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

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

[42] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1998.

[43] V. Seguy, B. B. Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. Large scale optimal transport and mapping estimation. In *International Conference on Learning Representations*, 2018.

[44] J. Urbas. On the second boundary value problem for equations of Monge-Ampère type. *J. Reine Angew. Math.*, 487:115–124, 1997.

[45] J. L. Vázquez. *The porous medium equation*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford, 2007.

[46] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.

[47] S. Zhang, G. Mordant, T. Matsumoto, and G. Schiebinger. Manifold learning with sparse regularised optimal transport. *Preprint arXiv:2307.09816v1*, 2023.