REGULARITY OF DENSITIES IN RELAXED AND PENALIZED AVERAGE DISTANCE PROBLEM

XIN YANG LU

ABSTRACT. The average distance problem finds application in data parameterization, which involves "representing" the data using lower dimensional objects. From a computational point of view it is often convenient to restrict the unknown to the family of parameterized curves. However this formulation exhibits several undesirable properties. In this paper we propose an alternative variant: the average distance functional is replaced by a transport cost, and the unknown is composed both by a curve and by a "projected measure", on which an L^q penalization term is added. Moreover we will add a term penalizing non injectivity. We will use techniques from optimal transport theory and calculus of variations. The main aim is to prove essential boundedness, and a variant of Lipschitz continuity for Radon-Nikodym derivative of projected measures for minimizers.

Keywords. nonlocal variational problem, average-distance problem, regularity
Classification. 49Q20, 49K10, 49Q10, 35B65

1. INTRODUCTION

The average distance problem was first proposed for mathematical modeling of optimization problems, such as urban planning and image processing, and for application in statistics. It also finds application in data parameterization, where given a data distribution, the aim is to find a lower dimensional object "representing" such data (see for instance Drineas, Frieze, Kannan, Vempala and Vinay [6], Smola, Mika, Schölkopf and Williamson [21]). The average distance problem was first analyzed by Buttazzo, Oudet and Stepanov in [3], where several qualitative properties of minimizers were proven. Further results were proven in Buttazzo and Stepanov [4, 5], Paolini and Stepanov [18]. A similar formulation, often referred as "penalized formulation", was introduced by Buttazzo, Mainini and Stepanov introduced in [2]:

Problem 1.1. *Given $d \geq 2$, a nonnegative, compactly supported measure μ and a parameter $\lambda > 0$, minimize*

$$E^\lambda_\mu : \mathcal{A} \longrightarrow [0, +\infty), \qquad E^\lambda_\mu(\cdot) := F_\mu(\cdot) + \lambda \mathcal{H}^1(\cdot),$$

where

$$F_\mu : \mathcal{A} \longrightarrow [0, +\infty), \qquad F_\mu(\Sigma) := \int_{\mathbb{R}^d} \inf_{y \in \Sigma} |x - y| \, \mathrm{d}\mu(x),$$

$$\mathcal{A} := \{\Sigma \subseteq \mathbb{R}^d : \Sigma \text{ compact, path-wise connected, } \mathcal{H}^1(\Sigma) < +\infty\}.$$

Existence of minimizers follows from Blaschke's selection theorem and and Gołąb's theorem. For future reference any considered measure will be assumed nonnegative, compactly supported, probability measure. The choice to work with probability measures is done for the

sake of simplicity, and results proven in this paper can be easily extended to finite measures. Problem 1.1 could be used to parameterize data clouds, i.e. representing a distribution of data point using lower dimensional objects, in this case elements of $\mathcal{A}$. Let

- μ be the distribution of data points,
- Σ (the unknown) be the set parameterizing the data points.

Thus $F_\mu(\Sigma)$ represents the "error" of such representation, while $\lambda\mathcal{H}^1(\Sigma)$ is the cost associated to its complexity. Although it is possible to consider penalizations terms of the form $G(\Sigma)$ (instead of $\lambda\mathcal{H}^1(\Sigma)$), with G satisfying some natural conditions (e.g. G non decreasing with respect to set inclusion, etc.), this is outside the scope of this paper. Thus minimizing E^λ_μ corresponds to finding the "best" one dimensional parameterization, which "balances" approximation error and complexity.

Moreover, from a computational point of view it is often advantageous to restrict the unknown to the family of parameterized curves. We need first to define the "length" of a parameterized curve, as defining it as $\mathcal{H}^1$-measure of the graph is not natural, since injectivity is not imposed and points (of the graph) can be visited multiple times. Let

$$\mathcal{C}^* := \{\gamma^* : [0,1] \longrightarrow \mathbb{R}^d : \gamma^* \text{ differentiable } \mathcal{L}^1\text{-a.e. with } |(\gamma^*)'| \text{ constant } \mathcal{L}^1\text{-a.e.}\},$$

and define the "length" of a curve $\gamma^* \in \mathcal{C}^*$ as its total variation

$$L_{\gamma^*} := \|\gamma^*\|_{TV} = \sup_n \left(\sup_{0=t_0<\cdots<t_{n-1}<t_n=1} \sum_{j=1}^{n} |\gamma^*(t_j) - \gamma^*(t_{j-1})| \right). \tag{1.1}$$

For the sake of simplicity, we will work with elements of

$$\mathcal{C} := \{\gamma : [0,a] \longrightarrow \mathbb{R}^d : a \geq 0,\ \gamma \text{ differentiable } \mathcal{L}^1\text{-a.e. with } |\gamma'| = 1\ \mathcal{L}^1\text{-a.e.}\}.$$

Elements of $\mathcal{C}^*$ will be referred as *"constant speed parameterized curves"*, while elements of $\mathcal{C}$ will be referred as *"arc-length parameterized curves"*. A natural way to define the "length" of an arc-length parameterized curve $\gamma : [0,a] \longrightarrow \mathbb{R}^d$ is the following:

(1) if $a = 0$, then $L_\gamma := 0$,
(2) if $a > 0$, then $L_\gamma := L_{\gamma^*}$ where $\gamma^* \in \mathcal{C}^*$, $\gamma^* : [0,1] \longrightarrow \mathbb{R}^d$, $\gamma^*(t) := \gamma(ta)$.

Thus by construction $L_\gamma = a$, and the domain of a curve $\gamma \in \mathcal{C}$ is $[0, L_\gamma]$. The average distance problem becomes:

Problem 1.2. *Given $d \geq 2$, a nonnegative, compactly supported measure μ and a parameter $\lambda > 0$, minimize*

$$\tilde{E}^\lambda_\mu : \mathcal{C} \longrightarrow [0,+\infty), \qquad \tilde{E}^\lambda_\mu(\gamma) := \tilde{F}_\mu(\gamma) + \lambda L_\gamma,$$

where

$$\tilde{F}_\mu : \mathcal{C} \longrightarrow [0,+\infty), \qquad \tilde{F}_\mu(\gamma) := \int_{\mathbb{R}^d} \inf_{y\in\Gamma_\gamma} |x-y| \, \mathrm{d}\mu, \qquad \Gamma_\gamma := \gamma([0, L_\gamma]).$$

For future reference notation L_γ will denote the "length" of γ, while Γ_γ will denote its graph. More details on the space $\mathcal{C}$ (including its topology) will be discussed in Section 2. In many applications the integrand $\inf_{y\in\Gamma_\gamma} |x-y|$ can be replaced by $\inf_{y\in\Gamma_\gamma} |x-y|^p$ for some $p \geq 1$.

Choice $p = 2$ is the most common. Note that in this case, if the reference measure μ is discrete, i.e.

$$\mu := \sum_j a_j \delta_{x_j}, \qquad \sum_j a_j = 1, \quad a_j \geq 0 \quad \forall j,$$

then

$$\tilde{F}_\mu(\gamma) = \sum_j a_j |x_j - y_j|^2, \qquad y_j \in \operatorname{argmin}_{y \in \Gamma_\gamma} |x_j - y| \quad \forall j,$$

i.e. $\tilde{F}_\mu(\gamma)$ is the (weighted) mean square distance of points x_j from the graph of γ. Problem 1.2 is related to "principal curves", and the lazy traveling salesman problem (see for instance Polak and Wolanski [19]). Principal curves are widely used in statistics and machine learning. For a (highly non exhaustive) list of references about the literature (both theoretical and applied) on principal curves, we cite Duchamp and Stuetzle [7, 8], Fischer [9], Hastie [10], Hastie and Stuetzle [11], Kégl [12], Kégl and Aetal [13], Ozertem and Erdogmus [17], Tibshirani [22].

However the formulation of Problem 1.2 still exhibits several undesirable properties when used in data parameterization:

(1) it has been proven (Slepčev [20]) that even assuming $\mu \ll \mathcal{L}^d$ with $\mathrm{d}\mu/\,\mathrm{d}\mathcal{L}^d \in C^\infty$, Problem 1.1 may admit minimizers which are simple curves failing to be C^1 regular. Moreover, any simple curve minimizing Problem 1.1 admits a parameterization $\gamma \in \mathcal{C}$ minimizing Problem 1.2, and a positive amount of mass is projected on any point on which C^1 regularity fails. For further details about "projections", we refer to Section 2 of [16]. In data parameterization, this corresponds to a loss of information, which is undesirable.

(2) The aforementioned configuration is a limit case of a more general issue: indeed in the formulation of Problem 1.2 there is no penalization for very high (even infinite) data concentration on the representation.

(3) In [15] it has been proven that Problem 1.1 may admit minimizers which are simple curves (thus these admit parameterizations minimizing Problem 1.2) whose set of non differentiability is not closed. This makes difficult to "control" the set on which C^1 regularity fails.

(4) Injectivity is not guaranteed, but highly desired: indeed given a minimizer γ of Problem 1.2, there are two "natural" choices of distances:
   - for data points, Euclidean distance is the natural choice,
   - on the representation γ however, the natural distance is the path distance d_γ, defined as $d_\gamma\big(\gamma(s), \gamma(t)\big) := |s - t|$, $s, t \in [0, L_\gamma]$.

   Clearly, if γ is not injective, then there exist s, t satisfying $s < t$ and $\gamma(s) = \gamma(t)$. Thus these two distances are not equivalent, and data points which are "close" (with respect to Euclidean distance) can be projected on points which are "distant" (with respect to d_γ). This is undesirable. Figure 2 is a schematic representation of this situation.

(5) The functional $\tilde{F}_\mu$ forces any point to project on one of the closest points on the curve. This imposes strong geometric rigidity on minimizers.

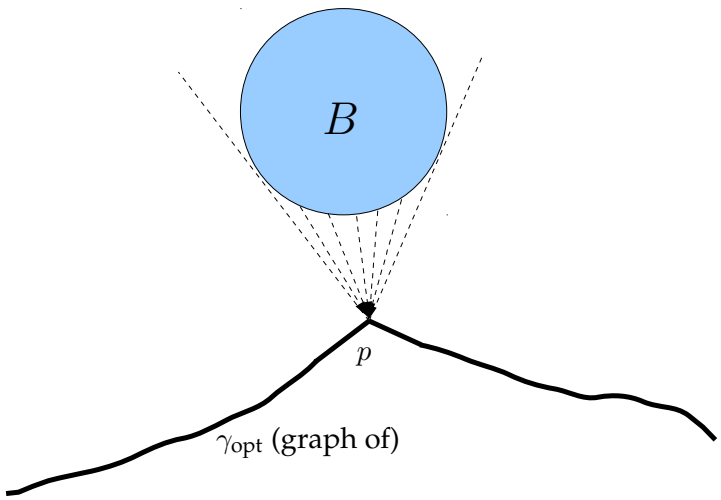

FIGURE 1. In this example from [20], the set $B \subseteq \operatorname{supp}(\mu)$ of positive μ-measure is projected on the single point p, on which C^1 regularity fails.

Thus we propose an alternative variant:

Problem 1.3. *Given $d \geq 2$, a measure μ, and parameters $\lambda, \varepsilon, \varepsilon' > 0$, $p \geq 1$, $q > 1$, solve*

$$\min_{(\gamma,\nu,\Pi)\in\mathcal{T}} \mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q](\gamma,\nu,\Pi),$$

where

$$\mathcal{T} := \{(\gamma,\nu,\Pi) : \gamma \in \mathcal{C},\ \nu \textit{ probability measure on } [0,L_\gamma],\\ \Pi \textit{ transport plan between } \mu \textit{ and } \gamma_\sharp\nu\},$$

$$\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q](\gamma,\nu,\Pi) := \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) + \lambda L_\gamma + \varepsilon \int_0^{L_\gamma} \nu^q \,\mathrm{d}\mathcal{L}^1 + \varepsilon'\eta(\gamma),$$

$$\eta(\gamma) := \int_0^{L_\gamma}\int_0^{L_\gamma} \left(\frac{|t-s|}{|\gamma(t)-\gamma(s)|}\right)^2 \mathrm{d}t\,\mathrm{d}s,$$

$$\int_0^{L_\gamma} \nu^q \,\mathrm{d}\mathcal{L}^1 := \begin{cases} \displaystyle\int_0^{L_\gamma} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}\mathcal{L}^1 & \textit{if } \nu \ll \mathcal{L}^1, \\ +\infty & \textit{otherwise.} \end{cases} \tag{1.2}$$

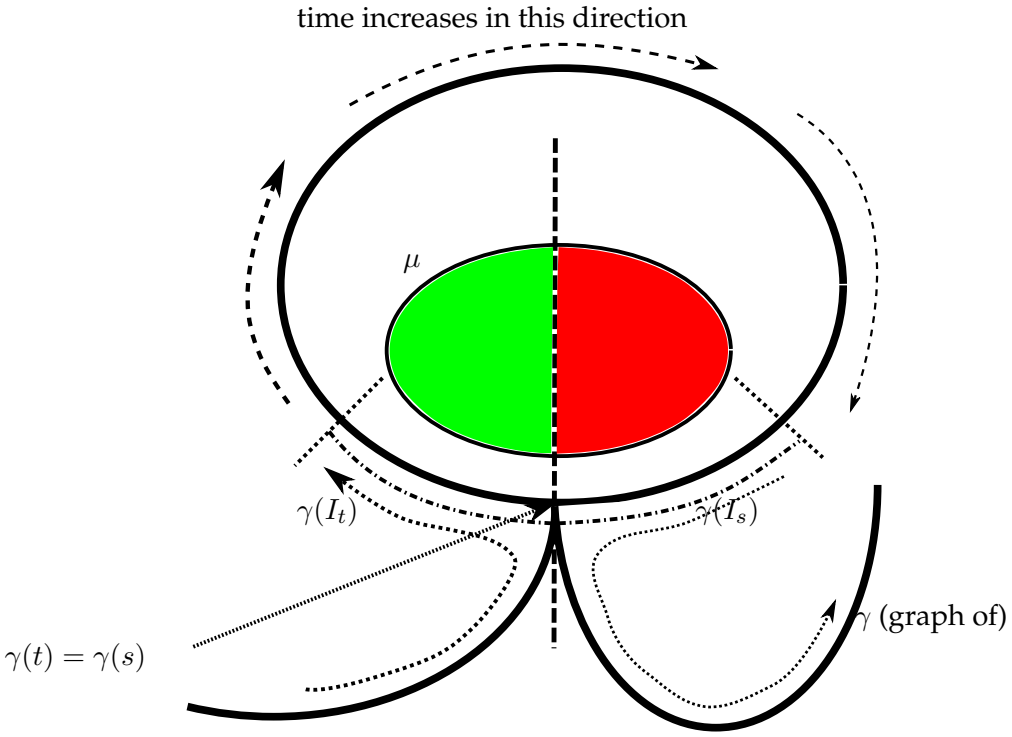

FIGURE 2. In this configuration, assuming $t < s$, points belonging to the red part are projected on $\gamma(I_s)$, while points belonging to the green part are projected on $\gamma(I_t)$. The sets $\gamma(I_s)$ and $\gamma(I_t)$ are distant with respect to d_γ. The colored area is part of $\mathrm{supp}(\mu)$. Time increases along the direction of dotted arrows.

The convergence in $\mathcal{T}$ will be detailed in Section 2. Note that the formulation of Problem 1.3 is quite different from classical average distance problem, and resembles the Monge-Kantorovich problem. Definition (1.2) is justified in view of Lemma 2.3. Existence of minimizers will be proven in Lemma 2.1. For future reference $\int_0^{L_\gamma} \nu^q \, \mathrm{d}\mathcal{L}^1$ will be referred as *"density penalization term"*, while with an abuse of notation, the transport cost $\int_{\mathbb{R}^d \times \Gamma_\gamma} |x-y|^p \, \mathrm{d}\Pi(x,y)$ will be referred as *"average distance term"*. The transport plan Π is more a technical expedient, and will play a marginal role in the following. Given $x \in \mathrm{supp}(\mu)$, $y \in \Gamma_\gamma$, we will say that "x projects on y" if $(x,y) \in \mathrm{supp}(\Pi)$. Note that:

- $\varepsilon' \eta(\gamma)$ penalizes non injectivity;
- $\varepsilon \int_0^{L_\gamma} \nu^q \, \mathrm{d}\mathcal{L}^1$ penalizes high concentrations of data on Γ_γ. In particular it diverges if a positive amount of data is projected on a singleton;

- the functional $\tilde{F}_\mu(\gamma)$ (from Problem 1.2) is replaced by $\int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y)$, allowing data points to be projected on any point (not just the closest ones). However, projecting on a distant point increases transport cost, but can be advantageous if such choice decreases the density penalization term.

The aim of this paper is to prove essential boundedness (Theorem 3.1) and "Lipschitz regularity" (Theorem 3.2) for $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$, when (γ,ν,Π) is a minimizer. Note that $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ is well defined upon $\mathcal{L}^1$-negligible sets. This paper will be structured as follows:

- in Section 2 we introduce preliminary notations and results, and prove existence of minimizers for Problem 1.3,
- in Section 3 we prove that for any minimizer (γ,ν,Π) of Problem 1.3, the Radon-Nikodym derivative $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ is essentially bounded. Moreover, if the exponent q appearing in the density penalization term is assumed $q=2$, then $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ satisfies a variant of Lipschitz continuity.

2. Preliminaries

The aim of this section is to present preliminary notions and results. The main result is existence of minimizers for Problem 1.3. We endow the space $\mathcal{C}$ with the following convergence: given a sequence $\{\gamma_n\}\subseteq\mathcal{C}$, we say $\{\gamma_n\}$ converges to $\gamma\in\mathcal{C}$ (and write $\{\gamma_n\}\xrightarrow{\mathcal{C}}\gamma$) if:

- $\{L_{\gamma_n}\}\to L_\gamma$,
- upon time inversion, i.e. replacing (for suitable indices) γ_n with $\tilde{\gamma}_n$ defined as $\tilde{\gamma}_n(t):=\gamma_n(L_{\gamma_n}-t)$, the sequence $\{\gamma_n^*\}$ converges to γ^* uniformly, where
$$\gamma^*:[0,1]\longrightarrow\mathbb{R}^d,\quad \gamma^*(t):=\gamma(tL_\gamma),$$
$$\gamma_n^*:[0,1]\longrightarrow\mathbb{R}^d,\quad \gamma_n^*(t):=\gamma_n(tL_{\gamma_n}),\qquad n=1,2,\cdots$$
denote the constant speed reparameterizations.

The convergence in $\mathcal{C}$ induces a "natural" convergence in $\mathcal{T}$: we say that a sequence $\{(\gamma_n,\nu_n,\Pi_n)\}\subseteq\mathcal{T}$ converges to $(\gamma,\nu,\Pi)\in\mathcal{T}$ (and write $\{(\gamma_n,\nu_n,\Pi_n)\}\xrightarrow{\mathcal{T}}(\gamma,\nu,\Pi)$) if $\{\gamma_n\}\xrightarrow{\mathcal{C}}\gamma$, $\{\nu_n\}\overset{*}{\rightharpoonup}\nu$, $\{\Pi_n\}\overset{*}{\rightharpoonup}\Pi$.

The first issue is existence of minimizers. For the sake of brevity we will omit writing the dependency on dimension (since all results will be valid for any dimension) for all quantities.

Lemma 2.1. *Given $d\geq 2$, a measure μ, parameters $\lambda,\varepsilon,\varepsilon'>0$, $p\geq 1$, $q>1$, the functional $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q]$ admits minimizers in $\mathcal{T}$.*

The proof will be split over several lemmas. Note that the set $\{\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q]<+\infty\}$ is non empty: indeed choose arbitrary points $x\in\mathrm{supp}(\mu)$, $y\in B(x,1)$, and let
$$\psi:[0,1]\longrightarrow\mathbb{R}^d,\ \psi(t):=(1-t)x+ty.$$
Let Π be an arbitrary optimal plan between μ and $\psi_\sharp\mathcal{L}^1_{\llcorner[0,1]}$. Then direct computation gives

$$\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q](\psi,\mathcal{L}^1_{\llcorner[0,1]},\Pi)\leq(\mathrm{diam}\,\mathrm{supp}(\mu)+1)^p+\lambda+\varepsilon+\varepsilon'<+\infty. \tag{2.1}$$

Lemma 2.2. *Given $d \geq 2$, a measure μ, parameters $\lambda, \varepsilon, \varepsilon' > 0$, $p \geq 1$, $q > 1$, $M \geq \inf_{\mathcal{C}} \mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q]$, and a sequence $\{(\gamma_n, \nu_n, \Pi_n)\} \subseteq \mathcal{T} \cap \{\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q] \leq M\}$, then it holds:*

(1) length estimate:

$$0 < (M/\varepsilon)^{\frac{1}{1-q}} \leq \inf_n L_{\gamma_n} \leq \sup_n L_{\gamma_n} \leq M/\lambda < +\infty, \tag{2.2}$$

(2) confinement condition:

$$\bigcup_n \Gamma_{\gamma_n} \subseteq \big(\operatorname{supp}(\mu)\big)_{M^{1/p}+M/\lambda}, \tag{2.3}$$

where for given $r \geq 0$,

$$\big(\operatorname{supp}(\mu)\big)_r := \left\{ x \in \mathbb{R}^d : \inf_{z \in \operatorname{supp}(\mu)} |x - z| \leq r \right\}.$$

Proof. Length estimate. Note that

$$(\forall n) \quad \lambda L_{\gamma_n} \leq \mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma_n, \nu_n, \Pi_n) \leq M \Longrightarrow L_{\gamma_n} \leq M/\lambda,$$

proving the upper bound in (2.2).

Fix an arbitrary n. Condition $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma_n, \nu_n, \Pi_n) \leq M$ gives

$$\begin{aligned} M \geq \mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma_n, \nu_n, \Pi_n) &\geq \varepsilon \int_0^{L_{\gamma_n}} \left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1} \right)^q \mathrm{d}\mathcal{L}^1 \geq \varepsilon \int_0^{L_{\gamma_n}} \left(\frac{1}{L_{\gamma_n}} \int_0^{L_{\gamma_n}} \frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1} \, \mathrm{d}\mathcal{L}^1 \right)^q \mathrm{d}\mathcal{L}^1 \\ &= \varepsilon \int_0^{L_{\gamma_n}} \left(\frac{1}{L_{\gamma_n}} \right)^q \mathrm{d}\mathcal{L}^1 = \varepsilon L_{\gamma_n}^{1-q}. \end{aligned}$$

Since $q > 1$, it follows $L_{\gamma_n}^{1-q} \leq M/\varepsilon$, proving the lower bound in (2.2).

Confinement condition. Note that for any n and $\xi \geq 0$, if $\Gamma_{\gamma_n} \cap \big(\operatorname{supp}(\mu)\big)_{(M+\xi)^{1/p}} = \emptyset$ then

$$\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma_n, \nu_n, \Pi_n) \geq \int_{\mathbb{R}^d \times \Gamma_{\gamma_n}} |x - y|^p \, \mathrm{d}\Pi(x, y) \geq M + \xi.$$

Since $\{(\gamma_n, \nu_n, \Pi_n)\} \subseteq \mathcal{T} \cap \{\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q] \leq M\}$, it follows

$$(\forall n)(\forall \xi > 0) \quad \Gamma_{\gamma_n} \cap \big(\operatorname{supp}(\mu)\big)_{(M+\xi)^{1/p}} \neq \emptyset.$$

Using length estimate $\sup_n L_{\gamma_n} \leq M/\lambda$ gives

$$(\forall n)(\forall \xi > 0) \quad \Gamma_{\gamma_n} \subseteq \big(\operatorname{supp}(\mu)\big)_{(M+\xi)^{1/p}+M/\lambda},$$

and the arbitrariness of ξ proves (2.3). □

Lemma 2.3. *Given $d \geq 2$, a measure μ, parameters $\lambda, \varepsilon, \varepsilon' > 0$, $p \geq 1$, $q > 1$, and $(\gamma, \nu, \Pi) \in \mathcal{T}$ satisfying $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma, \nu, \Pi) < +\infty$, then $\nu \ll \mathcal{L}^1$.*

Proof. Lebesgue decomposition theorem gives $\nu = \nu_a + \nu_s$, where $\nu_a \ll \mathcal{L}^1$, $\nu_s \perp \mathcal{L}^1$. Assume by contradiction $\nu_s \neq 0$, i.e. there exists a $\mathcal{L}^1$- negligible set $A \subseteq [0, L_\gamma]$ such that $\nu_s(A) = a > 0$. Let

$\{A_n\}$ be a sequence of open sets satisfying $A_n \supseteq A$ and $\mathcal{L}^1(A_n) = 1/n$ (for any $n \in \mathbb{N}$). Then it holds

$$\liminf_n \int_{A_n} \nu_s^q \, \mathrm{d}\mathcal{L}^1 \geq \liminf_n \int_{A_n} \int_{A_n} \left(\frac{a}{1/n} \right)^q \mathrm{d}\mathcal{L}^1 = \liminf_n \int_{A_n} a^q n^{q-1} = +\infty,$$

which contradicts $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma, \nu, \Pi) < +\infty$. □

Note that

$$(\forall \gamma \in \mathcal{C}) \quad \eta(\gamma) < +\infty \Longrightarrow \gamma \text{ injective.} \tag{2.4}$$

This because if there exist $t, s \in [0,1]$ with $t < s$, $\gamma(t) = \gamma(s)$, the integrand $\left(\frac{|s-t|}{|\gamma(s) - \gamma(t)|} \right)^2$ would have an asymptote of order two in t, thus $\eta(\gamma) = +\infty$.

Now it is possible to prove Lemma 2.1.

Proof. (of Lemma 2.1) Consider a minimizing sequence $\{(\gamma_n, \nu_n, \Pi_n)\}$. Since (in view of (2.1))

$$\inf_{\mathcal{T}} \mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q] \leq (\operatorname{diam} \operatorname{supp}(\mu) + 1)^p + \lambda + \varepsilon + \varepsilon' =: M,$$

assume without loss of generality $\sup_n \mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma_n, \nu_n, \Pi_n) \leq 2M$. Lemma 2.2 gives c_1, c_2 such that $c_2 \geq \sup_n L_{\gamma_n} \geq \inf_n L_{\gamma_n} \geq c_1 > 0$. Let

$$\gamma_n^* : [0,1] \longrightarrow \mathbb{R}^d, \quad \gamma_n^*(t) := \gamma_n(t L_{\gamma_n}), \qquad n = 1, 2, \cdots$$

be constant speed reparameterizations. Lemma 2.2 proves that the sequence $\{\gamma_n^*\}$ satisfies conditions of Ascoli-Arzelà theorem, thus (upon subsequence, which will not be relabeled) there exists $\gamma^* : [0,1] \longrightarrow \mathbb{R}^d$ such that $\{\gamma_n^*\} \to \gamma^*$ uniformly, and $L_{\gamma^*} := \lim_n L_{\gamma_n^*} > 0$. This implies $\{\gamma_n\} \xrightarrow{\mathcal{C}} \gamma$, where

$$\gamma : [0, L_{\gamma^*}] \longrightarrow \mathbb{R}^d, \quad \gamma(t) := \gamma^*(t / L_{\gamma^*}).$$

Define the measures ν_n^* as

$$\nu_n^*(B) := \nu_n(B L_{\gamma_n}) \quad \text{for any } \mathcal{L}^1\text{-measurable set } B \subseteq [0,1], \qquad n = 1, 2, \cdots,$$

where $B L_{\gamma_n} := \{t \in [0, L_{\gamma_n}] : t / L_{\gamma_n} \in B\}$. Since $\{(\gamma_n, \nu_n, \Pi_n)\}$ is a minimizing sequence, it follows

$$\sup_n \int_0^{L_{\gamma_n}} \nu_n^q \, \mathrm{d}\mathcal{L}^1 < +\infty \Longrightarrow \nu_n^* \ll \mathcal{L}^1, \qquad n = 1, 2, \cdots.$$

Let $f_n := \mathrm{d}\nu_n^* / \mathrm{d}\mathcal{L}^1$, $n = 1, 2, \cdots$. Since ν_n are nonnegative, it follows $f_n \geq 0$ for any n, and $\int_0^{L_{\gamma_n}} \nu_n^q \, \mathrm{d}\mathcal{L}^1$ differs from $\int_0^1 f_n^q \, \mathrm{d}\mathcal{L}^1$ by the multiplicative constant L_{γ_n}. This yields

$$\sup_n \int_0^{L_{\gamma_n}} \nu_n^q \, \mathrm{d}\mathcal{L}^1 < +\infty \Longrightarrow \sup_n \int_0^1 f_n^q \, \mathrm{d}\mathcal{L}^1 < +\infty,$$

i.e. the sequence $\{f_n\}$ is bounded in $L^q([0,1])$. Thus there exists $f \in L^q([0,1])$ such that (upon subsequence, which will not be relabeled) $\{f_n\} \rightharpoonup f$, which implies

$$\{\nu_n^*\} = \{f_n \cdot \mathcal{L}^1\} \overset{*}{\rightharpoonup} f \cdot \mathcal{L}^1 =: \nu^*,$$

and

$$\int_0^1 f^q \,\mathrm{d}\mathcal{L}^1 = \|f\|_{L^q}^q \le \liminf_n \|f_n\|_{L^q}^q = \liminf_n \int_0^1 f_n^q \,\mathrm{d}\mathcal{L}^1.$$

Since $\{L_{\gamma_n}\} \to L_\gamma$, it follows

$$\{\nu_n\} \xrightarrow{*} \nu, \qquad \int_0^{L_\gamma} \nu^q \,\mathrm{d}\mathcal{L}^1 \le \liminf_n \int_0^{L_{\gamma_n}} \nu_n^q \,\mathrm{d}\mathcal{L}^1, \tag{2.5}$$

where ν is defined as

$$\nu(B) := \nu^*(B/L_{\gamma_n}) \quad \text{for any } \mathcal{L}^1\text{-measurable set } B \subseteq [0, L_\gamma], \qquad B/L_\gamma := \{t \in [0,1] : tL_\gamma \in B\}.$$

Note that $\Gamma_{\gamma_n} \subseteq \mathbb{R}^d$, thus $\gamma_{n\sharp}\nu_n$ (resp. Π_n) is also a measure on $\mathbb{R}^d$ (resp. $\mathbb{R}^d \times \mathbb{R}^d$). Thus

$$\int_{\mathbb{R}^d \times \Gamma_{\gamma_n}} |x-y|^p \,\mathrm{d}\Pi_n(x,y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \,\mathrm{d}\Pi_n(x,y),$$

eliminating any problem that a moving domain of integration may generate. Prokhorov's theorem gives the existence of Π such that (upon subsequence, which will not be relabeled) $\{\Pi_n\} \xrightarrow{*} \Pi$, and Π is a transport plan between μ and $\gamma_\sharp\nu$ (for further details about stability of transport plans, we refer to [1], [23] and references therein), yielding

$$\lim_n \int_{\mathbb{R}^d \times \Gamma_{\gamma_n}} |x-y|^p \,\mathrm{d}\Pi_n(x,y) = \int_{\mathbb{R}^d \times \Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y). \tag{2.6}$$

It remains to prove lower semicontinuity for $\varepsilon'\eta(\cdot)$. Let

$$g_n : [0,1] \times [0,1] \longrightarrow \mathbb{R}, \qquad g_n(t,s) := \left(\frac{|s-t|}{|\gamma_n^*(s) - \gamma_n^*(t)|}\right)^2.$$

$$g : [0,1] \times [0,1] \longrightarrow \mathbb{R}, \qquad g(t,s) := \left(\frac{|s-t|}{|\gamma^*(s) - \gamma^*(t)|}\right)^2.$$

Since $\{\gamma_n^*\} \to \gamma^*$ uniformly, it follows $\{g_n\} \to g$ pointwisely. Fatou's lemma gives

$$(\forall s \in [0,1]) \quad \int_0^1 g(s,t) \,\mathrm{d}t \le \liminf_n \int_0^1 g_n(s,t) \,\mathrm{d}t,$$

i.e.

$$\begin{aligned}\eta(\gamma^*) = \int_0^1 \int_0^1 g(s,t) \,\mathrm{d}t \,\mathrm{d}s &\le \int_0^1 \left(\liminf_n \int_0^1 g_n(s,t) \,\mathrm{d}t\right) \mathrm{d}s \\ &\le \liminf_n \int_0^1 \int_0^1 g_n(s,t) \,\mathrm{d}t \,\mathrm{d}s = \liminf_n \eta(\gamma_n^*).\end{aligned}$$

Since $\eta(\gamma_n)$ (resp. $\eta(\gamma)$) differs from $\eta(\gamma_n^*)$ (resp. $\eta(\gamma^*)$) by the multiplicative constant $L_{\gamma_n}^2$ (resp. L_γ^2), it follows

$$\liminf_n \eta(\gamma_n) \ge \eta(\gamma).$$

Since $\{L_{\gamma_n}\} \to L_\gamma$, combining with (2.5) and (2.6) gives

$$\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q](\gamma,\nu,\Pi) \le \liminf_n \mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q](\gamma_n,\nu_n,\Pi_n),$$

and the proof is complete. □

We conclude this section with two simple observations.

Lemma 2.4. *Given $d \geq 2$, a measure μ, parameters $\lambda > 0$, $p \geq 1$, $q > 1$, sequences $\{\varepsilon_n\}, \{\varepsilon'_n\} \to 0$, and $(\gamma, \nu, \Pi) \in \mathcal{T}$, then it holds:*

- *any sequence $\{(\gamma_n, \nu_n, \Pi_n)\} \overset{\mathcal{T}}{\to} (\gamma, \nu, \Pi)$, satisfies*

$$\liminf_n \mathcal{E}[\mu, \lambda, \varepsilon_n, \varepsilon'_n, p, q](\gamma_n, \nu_n, \Pi_n) \geq \int_{\mathbb{R}^d \times \Gamma_\gamma} |x - y|^p \, \mathrm{d}\Pi(x, y) + \lambda L_\gamma; \tag{2.7}$$

- *assume there exist $\varepsilon, \varepsilon' > 0$ such that $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma, \nu, \Pi) < +\infty$. Then there exists a sequence $\{(\gamma_n, \nu_n, \Pi_n)\} \overset{\mathcal{T}}{\to} (\gamma, \nu, \Pi)$, such that*

$$\limsup_n \mathcal{E}[\mu, \lambda, \varepsilon_n, \varepsilon'_n, p, q](\gamma_n, \nu_n, \Pi_n) \leq \int_{\mathbb{R}^d \times \Gamma_\gamma} |x - y|^p \, \mathrm{d}\Pi(x, y) + \lambda L_\gamma; \tag{2.8}$$

Proof. Fix an arbitrary $(\gamma, \nu, \Pi) \in \mathcal{T}$. Consider an arbitrary sequence $\{(\gamma_n, \nu_n, \Pi_n)\} \overset{\mathcal{T}}{\to} (\gamma, \nu, \Pi)$. It holds

$$\begin{aligned}
\liminf_n \int_{\mathbb{R}^d \times \Gamma_{\gamma_n}} |x - y|^p \, \mathrm{d}\Pi_n(x, y) + \lambda L_{\gamma_n} + \varepsilon_n \int_0^{L_{\gamma_n}} \nu_n^q \, \mathrm{d}\mathcal{L}^1 + \varepsilon'_n \eta(\gamma_n) & \\
\geq \liminf_n \int_{\mathbb{R}^d \times \Gamma_{\gamma_n}} |x - y|^p \, \mathrm{d}\Pi_n(x, y) + \lambda L_{\gamma_n} & \\
\geq \int_{\mathbb{R}^d \times \Gamma_\gamma} |x - y|^p \, \mathrm{d}\Pi(x, y) + \lambda L_\gamma, &
\end{aligned}$$

proving (2.7).

To prove (2.8), note that since by hypothesis there exist $\varepsilon, \varepsilon' > 0$ such that $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma, \nu, \Pi) < +\infty$, it follows that γ is injective in view of (2.4), and $\nu \ll \mathcal{L}^1$ in view of Lemma 2.3. Let

$$\gamma_n := \gamma, \quad \nu_n := \nu, \quad \Pi_n := \Pi, \qquad n = 1, 2, \cdots.$$

By construction $\{(\gamma_n, \nu_n, \Pi_n)\} \overset{\mathcal{T}}{\to} (\gamma, \nu, \Pi)$, and

$$\int_0^{L_{\gamma_n}} \nu_n^q \, \mathrm{d}\mathcal{L}^1 = \int_0^{L_{\gamma_n}} \nu^q \, \mathrm{d}\mathcal{L}^1 < +\infty, \quad \eta(\gamma_n) = \eta(\gamma) < +\infty, \qquad n = 1, 2, \cdots,$$

thus

$$\begin{aligned}
\lim_n \int_{\mathbb{R}^d \times \Gamma_{\gamma_n}} |x - y|^p \, \mathrm{d}\Pi_n(x, y) + \lambda L_{\gamma_n} + \varepsilon_n \int_0^{L_{\gamma_n}} \nu_n^q \, \mathrm{d}\mathcal{L}^1 + \varepsilon'_n \eta(\gamma_n) & \\
= \int_{\mathbb{R}^d \times \Gamma_\gamma} |x - y|^p \, \mathrm{d}\Pi(x, y) + \lambda L_\gamma, &
\end{aligned}$$

proving (2.8). □

Lemma 2.5. *Given $d \geq 2$, a measure μ, parameters $\lambda, \varepsilon' > 0$, $p \geq 1$, $q > 1$, a sequence $\{\varepsilon_n\} \to 0$, and $(\gamma, \nu, \Pi) \in \mathcal{T}$, then there exists a sequence $\{(\gamma_n, \nu_n, \Pi_n)\} \overset{\mathcal{T}}{\to} (\gamma, \nu, \Pi)$ such that*

$$\lim_n \mathcal{E}[\mu, \lambda, \varepsilon_n, \varepsilon', p, q](\gamma_n, \nu_n, \Pi_n) = \mathcal{E}[\mu, \lambda, 0, \varepsilon', p, q](\gamma, \nu, Pi) = \int_{\mathbb{R}^d \times \Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) + \lambda L_\gamma + \varepsilon' \eta(\gamma). \tag{2.9}$$

In particular $\{\mathcal{E}[\mu, \lambda, \varepsilon_n, \varepsilon', p, q]\} \overset{\Gamma}{\to} \mathcal{E}[\mu, \lambda, 0, \varepsilon', p, q]$ as $n \to +\infty$.

Before the proof, note that for fixed γ, the quantity

$$\mathcal{E}[\mu, \lambda, 0, \varepsilon', p, q](\gamma, \nu, \Pi) = \int_{\mathbb{R}^d \times \Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) + \lambda L_\gamma + \varepsilon' \eta(\gamma)$$

is minimum when

$$\int_{\mathbb{R}^d \times \Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) = \int_{\mathbb{R}^d} \inf_{z \in \Gamma_\gamma} |x-z|^p \,\mathrm{d}\mu(x),$$

since only the average distance term depends on ν and Π.

Proof. If $\eta(\gamma) = +\infty$ then (2.9) follows. Thus assume $\eta(\gamma) < +\infty$, i.e. γ is injective.

- Case $L_\gamma > 0$.

Let $\gamma_n := \gamma$, $n = 1, 2, \cdots$. Note that for any $t \in [0, L_\gamma]$ the measure δ_t (Dirac measure in t) can be approximated (in the weak-$*$ topology) by measures of the form $f_{n,t} \cdot \mathcal{L}^1_{\llcorner [0,L_\gamma]}$, where $f_{n,t} := k_n \chi_{I_t(k_n)}$, $\{k_n\} \to +\infty$, χ denotes the characteristic function of the subscripted set, and $I_t(k_n)$ is an arbitrary interval containing t such that $\mathcal{L}^1(I_t(k_n)) = 1/k_n$. Thus any measure of the form

$$\Big(\sum_{j=1}^H a_j \delta_{t_j}\Big) \cdot \mathcal{L}^1_{\llcorner [0,L_\gamma]}, \qquad H \in \mathbb{N}, \quad \sum_{j=1}^H a_j = 1, \quad \{t_j\} \subseteq [0, L_\gamma]$$

can be approximated (in the weak-$*$ topology) by measures of the form $\Big(\sum_{j=1}^H a_j f_{n,t}\Big) \cdot \mathcal{L}^1_{\llcorner [0,L_\gamma]}$. Thus ν can be approximated (in the weak-$*$ topology) by a sequence of measures $\{\nu_n\}$ the form

$$\nu_n := \left(\sum_{j=1}^{H_n} a_{j,n} f_{n,t_{j,n}}\right) \cdot \mathcal{L}^1_{\llcorner [0,L_\gamma]},$$

for suitable choices of $\{H_n\} \subseteq \mathbb{N}$, $\{a_{j,n}\} \subseteq [0,1]$, $\sum_{j,n} a_{j,n} = 1$, $\{t_{j,n}\} \subseteq [0, L_\gamma]$. Choosing $k_n := \varepsilon_n^{1/(2-2q)}$ gives

$$(\forall n, t) \quad \int_0^{L_\gamma} f_{n,t}^q \,\mathrm{d}\mathcal{L}^1 \leq k_n^{q-1} = \varepsilon_n^{-1/2},$$

thus

$$(\forall n) \quad \int_0^{L_\gamma} \left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}\mathcal{L}^1 \leq \varepsilon_n^{-1/2}. \tag{2.10}$$

For any n, choose an optimal plan Π_n between μ and $\gamma_\sharp\nu_n$. Since $\{\nu_n\} \overset{*}{\rightharpoonup} \nu$, it follows (upon subsequence, which will not be relabeled) $\{\Pi_n\} \overset{*}{\rightharpoonup} \Pi$, and

$$\begin{aligned}\lim_n \mathcal{E}[\mu,\lambda,\varepsilon_n,\varepsilon',p,q](\gamma_n,\nu_n,\Pi_n) &= \lim_n \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi_n(x,y) + \lambda L_\gamma \\ &\quad + \varepsilon_n \int_0^{L_\gamma} \left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}\mathcal{L}^1 + \varepsilon'\eta(\gamma) \\ &\overset{(2.10)}{\leq} \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) + \lambda L_\gamma + \varepsilon'\eta(\gamma).\end{aligned}$$

- Case $L_\gamma = 0$.

This implies $\nu = \delta_0$. Choose an arbitrary unit vector $w \in \mathbb{R}^d$, let $\{P\} := \Gamma_\gamma$ and

$$\gamma_n : [0,\xi_n] \longrightarrow \mathbb{R}^d, \quad \gamma_n(t) := P + tw, \quad \xi_n := \varepsilon_n^{1/(2q-2)} \qquad n = 1,2,\cdots.$$

By construction $\{\gamma_n\} \overset{\mathcal{C}}{\to} \gamma$. Let

$$\nu_n := \xi_n^{-1} \cdot \mathcal{L}^1_{\llcorner[0,\xi_n]}, \qquad n = 1,2,\cdots,$$

and direct computation gives

$$(\forall n) \quad \int_0^{\xi_n} \left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}\mathcal{L}^1 \leq \varepsilon_n^{-1/2}.$$

By construction $\{\nu_n\} \overset{*}{\rightharpoonup} \nu$. For any n choose an optimal plan Π_n between μ and $\gamma_{n\sharp}\nu_n$, and (note that Π_n can be considered as measure on $\mathbb{R}^d$, thus eliminating any problem potentially related to a moving domain of integration) upon subsequence (which will not be relabeled) $\{\Pi_n\} \overset{*}{\rightharpoonup} \Pi$. Since by construction $\{\eta(\gamma_n)\} \to 0$, it follows

$$\begin{aligned}\lim_n \mathcal{E}[\mu,\lambda,\varepsilon_n,\varepsilon',p,q](\gamma_n,\nu_n,\Pi_n) &= \lim_n \int_{\mathbb{R}^d\times\Gamma_{\gamma_n}} |x-y|^p \,\mathrm{d}\Pi_n(x,y) + \lambda\xi_n \\ &\quad + \varepsilon_n \int_0^{\xi_n} \left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}\mathcal{L}^1 + \varepsilon'\eta(\gamma_n) \\ &= \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) = \int_{\mathbb{R}^d} |x-P|^p \,\mathrm{d}\mu(x).\end{aligned}$$

Thus (2.9) is proven. Since for any sequence $\{(\gamma_n,\nu_n,\Pi_n)\} \overset{\mathcal{T}}{\to} (\gamma,\nu,\Pi)$ it holds

$$\begin{aligned}\liminf_n \int_{\mathbb{R}^d\times\Gamma_{\gamma_n}} |x-y|^p \,\mathrm{d}\Pi_n(x,y) + \lambda L_{\gamma_n} + \varepsilon_n \int_0^{L_{\gamma_n}} \nu_n^q \,\mathrm{d}\mathcal{L}^1 + \varepsilon'\eta(\gamma_n) \\ \geq \liminf_n \int_{\mathbb{R}^d\times\Gamma_{\gamma_n}} |x-y|^p \,\mathrm{d}\Pi_n(x,y) + \lambda L_{\gamma_n} + \varepsilon'\eta(\gamma_n) \\ \geq \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) + \lambda L_\gamma + \varepsilon'\eta(\gamma),\end{aligned}$$

it follows $\{\mathcal{E}[\mu,\lambda,\varepsilon_n,\varepsilon',p,q]\} \xrightarrow{\Gamma} \mathcal{E}[\mu,\lambda,0,\varepsilon',p,q]$ as $n\to+\infty$. □

3. Regularity of densities

In Lemma 2.3 it has been proven that if (γ,ν,Π) is a minimizer of Problem 1.3 then $\nu \ll \mathcal{L}^1$. In this section further regularity properties will be analyzed. The main results are:

Theorem 3.1. *(Essential boundedness) Given $d\geq 2$, a measure μ, parameters $\lambda,\varepsilon,\varepsilon'>0$, $p\geq 1$, $q>1$, and a minimizer (γ,ν,Π) of $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,q]$, then $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1 \in L^\infty$.*

Theorem 3.2. *("Lipschitz continuity") Given $d\geq 2$, a measure μ, parameters $\lambda,\varepsilon,\varepsilon'>0$, $p\geq 1$, and a minimizer (γ',ν',Π') of $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,2]$, then for any time $t\in[0,L_\gamma]$, it holds:*

(1) upon suitably modifying $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ on a $\mathcal{L}^1$-negligible set, $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ is continuous in t,

(2) denoting by Λ the set of Lebesgue points of $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$, for arbitrary sequences $\{t_n\}\to t$, $\{s_n\}\to t$, $\{t_n\},\{s_n\}\subseteq\Lambda$, there exists n_0 such that

$$(\forall n\geq n_0)\quad \left|\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}(t_n)-\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}(s_n)\right|\leq H|t_n-s_n|,$$

where

$$H:=p(M^{1/p}+M/\lambda)^{p-1}/\varepsilon,\qquad M:=(\operatorname{diam}\operatorname{supp}(\mu)+1)^p+\lambda+\varepsilon+\varepsilon'.$$

In particular, conclusions (1) and (2) imply that for any $t\in[0,L_\gamma]$ there exists a function g such that:

- *g is continuous in t,*
- *$\mathcal{L}^1(\{\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1\neq g\})=0$,*
- *for any sequence $\{t_n\}\to t$ it holds $|g(t_n)-g(t)|\leq H|t_n-t|$.*

Choice $q=2$ in Theorem 3.2 is due to technical reasons, as noted in Remark I in the following. The next lemma will be useful.

Lemma 3.3. *Given $K\geq 1$, $a,b\in[0,K]$, $p\geq 1$, then it holds $|a^p-b^p|\leq|a-b|pK^{p-1}$.*

Proof. Assume by symmetry $a\geq b$. If $p\in\mathbb{N}$, then

$$a^p-b^p=(a-b)\sum_{j=0}^{p-1}a^{p-1-j}b^j \overset{a,b\leq K}{\leq} (a-b)pK^{p-1}.$$

If $p\notin\mathbb{N}$, let $g(p):=\dfrac{a^p-b^p}{a-b}$, and note that the only issue is the diagonal $\{a=b\}$. Moreover direct computation gives $\{\nabla g=0\}\subseteq\{a=b\}$. For any sufficiently small $\delta\ll 1$, letting $a=(1+\delta)b$ gives $g(a,b)\approx pb^{p-1}\leq pK^{p-1}$. Note also that for $a=K$, $b\mapsto g(K,b)$ is maximum when $\dfrac{K^p-b^p}{K-b}=pb^{p-1}$, concluding the proof. concluding the proof. □

Proof. (of Theorem 3.1) Note that the term $\eta(\gamma)$ depends only on γ, not on ν or Π. As the following construction does not alter γ, the term $\eta(\gamma)$ does not change.

Choose an arbitrary $S \gg 1$, and let $A_S := \{S \leq \mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1 \leq S^{4/3}\}$. Clearly $\mathcal{L}^1(A_S)S \leq 1$. Assume $\mathcal{L}^1(A_S) > 0$. The goal is to find an upper bound for S. Since (γ, ν, Π) is a minimizer, it follows

$$\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q](\gamma, \nu, \Pi) \leq (\operatorname{diam}\operatorname{supp}(\mu) + 1)^p + \lambda + \varepsilon + \varepsilon' =: M.$$

Lemma 2.2 gives the existence of constants $c := (M/\varepsilon)^{\frac{1}{1-q}} < 1$, $C := M/\lambda$ such that $C \geq L_\gamma \geq c > 0$. The set $B := \{\,\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1 \leq 2/c\} \subseteq [0, L_\gamma]$ satisfies $\mathcal{L}^1(B) \geq 1 - c/2$. Let

$$\nu'_S := \nu - \nu_{\llcorner A_S} + \frac{\nu(A_S)}{\mathcal{L}^1(B)}\mathcal{L}^1_{\llcorner B}.$$

In particular, for any $S > 2/c$ it holds $A_S \cap B = \emptyset$. Let Π'_S be an optimal plan between μ and $\gamma_\sharp\nu'_S$. Consider arbitrary $x \in \operatorname{supp}(\mu)$, $y, y' \in \Gamma_\gamma$. Lemma 2.2 gives $\big||x-y| - |x-y'|\big| \leq M^{1/p} + M/\lambda$, where clearly $M \geq 1$. Thus applying Lemma 3.3 (with $a := |x-y|$, $b := |x-y'|$, $K := M^{1/p} + M/\lambda$) gives

$$\begin{aligned}\big||x-y|^p - |x-y'|^p\big| &\leq \big||x-y| - |x-y'|\big|p(M^{1/p} + M/\lambda)^{p-1} \\ &\leq L_\gamma p(M^{1/p} + M/\lambda)^{p-1},\end{aligned}$$

i.e.

$$\begin{aligned}\int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|\,\mathrm{d}\Pi'_S(x,y) - \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|\,\mathrm{d}\Pi(x,y) &\leq L_\gamma p(M^{1/p} + M/\lambda)^{p-1}\nu(A_S) \\ &\leq Cp(M^{1/p} + M/\lambda)^{p-1}S^{4/3}\mathcal{L}^1(A_S). \qquad (3.1)\end{aligned}$$

Recalling that $\left.\dfrac{\mathrm{d}\nu'_S}{\mathrm{d}\mathcal{L}^1}\right|_{A_S} = 0$ for any $S > 2/c$, it follows

$$\begin{aligned}\int_B \left(\frac{\mathrm{d}\nu'_S}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 - \int_B \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 &= \int_B \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} + \frac{\nu(A_S)}{\mathcal{L}^1(B)}\right)^2 \mathrm{d}\mathcal{L}^1 - \int_B \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 \\ &= \frac{2\nu(A_S)}{\mathcal{L}^1(B)}\nu(B) + \left(\frac{\nu(A_S)}{\mathcal{L}^1(B)}\right)^2 \mathcal{L}^1(B) \\ &\leq \frac{2\nu(B)}{\mathcal{L}^1(B)}S^{4/3}\mathcal{L}^1(A_S) + \frac{S^{8/3}\mathcal{L}^1(A_S)^2}{\mathcal{L}^1(B)},\end{aligned}$$

and

$$\int_{A_S} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 - \int_{A_S} \left(\frac{\mathrm{d}\nu'_S}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 \geq S^2\mathcal{L}^1(A_S). \qquad (3.2)$$

Combining estimates (3.1), (3.2) and the minimality of (γ, ν, Π) (compared against (γ, ν'_S, Π'_S)) gives

$$\begin{aligned}Cp(M^{1/p} + M/\lambda)^{p-1}S^{4/3}\mathcal{L}^1(A_S) + \varepsilon\frac{2\nu(B)}{\mathcal{L}^1(B)}S^{4/3}\mathcal{L}^1(A_S) + \varepsilon\frac{S^{8/3}\mathcal{L}^1(A_S)^2}{\mathcal{L}^1(B)} \\ \geq \varepsilon S^2\mathcal{L}^1(A_S). \qquad (3.3)\end{aligned}$$

Note that

$$\lim_{S\to+\infty} \frac{Cp(M^{1/p}+M/\lambda)^{p-1}S^{4/3}\mathcal{L}^1(A_S)}{S^2\mathcal{L}^1(A_S)} = 0.$$

Since $\mathcal{L}^1(B) \geq 1 - c/2$, it follows

$$\lim_{S\to+\infty} \frac{2\nu(B)S^{4/3}\mathcal{L}^1(A_S)/\mathcal{L}^1(B)}{S^2\mathcal{L}^1(A_S)} = 0,$$

and

$$\begin{aligned}\lim_{S\to+\infty} \frac{\varepsilon S^{8/3}\mathcal{L}^1(A_S)^2/\mathcal{L}^1(B)}{S^2\mathcal{L}^1(A_S)} &\leq \lim_{S\to+\infty} \frac{\varepsilon S^{8/3}\mathcal{L}^1(A_S)^2}{(1-c/2)S^2\mathcal{L}^1(A_S)} \\ &= \lim_{S\to+\infty} \frac{\varepsilon S^{2/3}\mathcal{L}^1(A_S)}{1-c/2} \leq \lim_{S\to+\infty} \frac{\varepsilon S^{-1/3}}{1-c/2} = 0.\end{aligned}$$

Let S^* be the maximum value for which inequality (3.3) holds, thus for any $S > S^*$ the minimality of (γ, ν, Π) cannot hold, and similarly for any $S > S^*$ assumption $\mathcal{L}^1(A_S) > 0$ cannot hold. However, if $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1 \notin L^\infty$, then there exists a sequence $\{S_j\} \to +\infty$ such that $\mathcal{L}^1(A_{S_j}) > 0$ for any j, and $A_{S_j} \cap A_{S_{j'}} = \emptyset$ whenever $j \neq j'$. The aforementioned arguments would construct a competitor contradicting the minimality of (γ, ν, Π). Thus the only possibility is $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1 \in L^\infty$. □

For future reference the exponent q appearing in the density penalization term will be assumed $q = 2$. This mainly due to technical reasons, as noted in Remark I in the following.

Proposition 3.4. *Given $d \geq 2$, a measure μ, parameters $\lambda, \varepsilon, \varepsilon' > 0$, $p \geq 1$, and a minimizer (γ, ν, Π) of $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, 2]$, then it holds:*

- *for any $t \in [0, L_\gamma]$ there exist no sequences of Borel sets $\{A_n\}$, $\{B_n\}$ and $\{c_{1,n}\}, \{c_{2,n}\}$ such that*

$$(\forall n) \quad \inf_n c_{1,n} - c_{2,n} > 0, \qquad (\forall \xi)(\exists n_0) : \ (\forall n \geq n_0) \quad A_n \cup B_n \subseteq B(t, \xi),$$

$$\mathcal{L}^1(A_n) > 0, \quad \mathcal{L}^1(B_n) > 0, \quad \left.\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right|_{A_n} \geq c_{1,n} > c_{2,n} \geq \left.\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right|_{B_n}, \qquad n = 1, 2, \cdots.$$

Proof. Assume (for the sake of contradiction) there exist $t \in [0, L_\gamma]$, $\{A_n\}$, $\{B_n\}$, $\{c_{1,n}\}$, $\{c_{2,n}\}$, c such that

$$\inf_n c_{1,n} - c_{2,n} \geq c > 0, \qquad (\forall \xi)(\exists n_0) : \ (\forall n \geq n_0) \quad A_n \cup B_n \subseteq B(t, \xi),$$

$$\mathcal{L}^1(A_n) > 0, \quad \mathcal{L}^1(B_n) > 0, \quad \left.\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right|_{A_n} \geq c_{1,n} > c_{2,n} \geq \left.\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right|_{B_n}, \qquad n = 1, 2, \cdots.$$

Clearly such $\{A_n\}$, $\{B_n\}$ are disjoint for any n, and it can be assumed $\mathcal{L}^1(A_n) = \mathcal{L}^1(B_n)$ (since if $\mathcal{L}^1(A_n) > \mathcal{L}^1(B_n)$, there exists $A'_n \subseteq A_n$ satisfying $\mathcal{L}^1(A'_n) = \mathcal{L}^1(B_n)$, and similarly if $\mathcal{L}^1(B_n) > \mathcal{L}^1(A_n)$). Let

$$l_n := \mathcal{L}^1(A_n) = \mathcal{L}^1(B_n), \quad d_n := \mathrm{diam}(A_n \cup B_n), \quad n = 1, 2, \cdots.$$

The goal is to construct $\tilde{\nu}_n$ such that $(\gamma, \tilde{\nu}_n, \tilde{\Pi}_n)$ (with $\tilde{\Pi}_n$ arbitrary optimal plan between μ and $\gamma_\sharp\tilde{\nu}_n$) contradicts the minimality of (γ, ν, Π). Consider an index n. Choose $C_n \subseteq A_n$ such that $\nu(C_n) = \big(\nu(A) - \nu(B)\big)/2$. Let

$$\tilde{\nu}_n := \nu - \nu_{\llcorner C_n} + \frac{\nu(A_n) - \nu(B_n)}{2\mathcal{L}^1(B_n)}\mathcal{L}^1_{\llcorner B_n}.$$

Choose an optimal transport plan $\tilde{\Pi}_n$ between μ and $\gamma_\sharp\tilde{\nu}_n$. Consider arbitrary $x \in \operatorname{supp}(\mu)$, $y \in \gamma(C_n)$, $y' \in \gamma(B_n)$. Lemma 2.2 gives

$$\big||x-y| - |x-y'|\big| \le M^{1/p} + M/\lambda, \qquad M := (\operatorname{diam}\operatorname{supp}(\mu) + 1)^p + \lambda + \varepsilon + \varepsilon',$$

thus Lemma 3.3 (applied with $a := |x-y|$, $b := |x-y'|$, $K := M^{1/p} + M/\lambda$) gives

$$\begin{aligned}\big||x-y|^p - |x-y'|^p\big| &\le \big||x-y| - |x-y'|\big|p(M^{1/p} + M/\lambda)^{p-1}\\ &\le d_n p(M^{1/p} + M/\lambda)^{p-1},\end{aligned}$$

i.e.

$$\int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\tilde{\Pi}_n(x,y) - \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) \le \frac{\nu(A_n) - \nu(B_n)}{2} p(M^{1/p} + M/\lambda)^{p-1} d_n.$$

Direct computation gives

$$\begin{aligned}&\int_{A_n\cup B_n} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 - \int_{A_n\cup B_n} \left(\frac{\mathrm{d}\tilde{\nu}_n}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1\\ &\qquad\ge \frac{\nu(A_n)^2 + \nu(B_n)^2}{l_n} - \frac{(\nu(A_n) + \nu(B_n))^2}{2l_n}\\ &\qquad\ge \frac{(\nu(A_n) - \nu(B_n))^2}{2l_n}. \qquad (3.4)\end{aligned}$$

Combining (3.4) with the minimality of (γ, ν, Π) gives

$$\varepsilon\frac{\big(\nu(A_n) - \nu(B_n)\big)^2}{2l_n} \le \frac{\nu(A_n) - \nu(B_n)}{2} p(M^{1/p} + M/\lambda)^{p-1} d_n,$$

i.e.

$$\frac{\nu(A_n) - \nu(B_n)}{l_n} \le \frac{d_n}{\varepsilon} p(M^{1/p} + M/\lambda)^{p-1}.$$

Since $\nu(A_n) - \nu(B_n) \ge cl_n$, it follows

$$c \le \frac{p(M^{1/p} + M/\lambda)^{p-1}}{\varepsilon} d_n. \qquad (3.5)$$

This argument can be repeated for any n, and inequality (3.5) is false for any sufficiently large n since $\{d_n\} \to 0$. Thus the proof is complete. □

We present some comments on the conclusion of Proposition 3.4. In Corollaries 3.5 and 3.6 we will use the same notation from Proposition 3.4.

Corollary 3.5. *The conclusion of Proposition 3.4 implies*

$$(\forall t \in [0, L_\gamma],\ \xi > 0)(\exists \delta > 0): \quad \operatorname{esssup}_{[t-\delta,t+\delta]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \operatorname{essinf}_{[t-\delta,t+\delta]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \leq \xi, \tag{3.6}$$

where esssup *(resp.* essinf*) denotes the essential supremum (resp. essential infimum).*

Proof. Assume the opposite holds, i.e. there exist $t \in [0, L_\gamma]$, $\xi > 0$ and a sequence $\{\delta_n\} \to 0$ such that

$$\operatorname{esssup}_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \operatorname{essinf}_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \geq \xi \qquad n = 1, 2, \cdots .$$

Then (for any n) there exist A_n, $B_n \subseteq [t - \delta_n, t + \delta_n]$ such that

$$\mathcal{L}^1(A_n) > 0,\ \mathcal{L}^1(B_n) > 0, \qquad \inf_{A_n} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \sup_{B_n} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \geq \frac{\xi}{2}, \qquad n = 1, 2, \cdots .$$

Letting

$$c_{1,n} := \inf_{A_n} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}, \qquad c_{2,n} := \sup_{B_n} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}, \qquad n = 1, 2, \cdots$$

concludes the proof. □

Corollary 3.6. *The conclusion of Proposition 3.4 implies that for any $t \in [0, L_\gamma]$ there exists a function g continuous in t such that $\mathrm{d}\nu/\mathrm{d}\mathcal{L}^1 = g$ $\mathcal{L}^1$-a.e., i.e. conclusion (1) of Theorem 3.2.*

Proof. Note that if at a given time t_0, a function f satisfies

$$(\forall \xi > 0)(\exists \delta > 0): \qquad \sup_{[t_0-\delta,t_0+\delta]} f - \inf_{[t_0-\delta,t_0+\delta]} f \leq \xi, \tag{3.7}$$

then there exists $k \in \mathbb{R}$ such that for any sequence $\{t_n\} \to t_0$ it holds $\{f(t_n)\} \to k$, i.e. f can be made continuous in t_0 by imposing $f(t_0) := k$. Consider an arbitrary time $t \in [0, L_\gamma]$. Since the thesis states a local property, we need only to consider times close to t. Choose a sequence $\{\xi_n\} \to 0$, and Corollary 3.5 gives

$$(\forall n)(\exists \delta_n > 0): \quad \operatorname{esssup}_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \operatorname{essinf}_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \leq \xi_n.$$

Thus for any n there exists a $\mathcal{L}^1$-negligible set $E_n \subseteq [t - \delta_n, t + \delta_n]$ such that upon suitably modifying $\mathrm{d}\nu/\mathrm{d}\mathcal{L}^1$ on E_n gives

$$\sup_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \inf_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \leq \xi_n,$$

i.e. upon suitably modifying $\mathrm{d}\nu/\mathrm{d}\mathcal{L}^1$ on $\bigcup_n E_n$ (clearly $\mathcal{L}^1$-negligible) gives

$$(\forall n)(\exists \delta_n > 0): \qquad \sup_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \inf_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \leq \xi_n.$$

Then for arbitrary $\xi > 0$, choosing $\xi_n \leq \xi$ gives

$$\sup_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \inf_{[t-\delta_n,t+\delta_n]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \leq \xi_n \leq \xi,$$

i.e.

$$(\forall \xi > 0)(\exists \delta > 0): \quad \sup_{[t-\delta,t+\delta]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} - \inf_{[t-\delta,t+\delta]} \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1} \leq \xi.$$

Thus (upon suitably modifying $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ on $\bigcup_n E_n$) there exists a value $h \in \mathbb{R}$ such that imposing $\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}(t) := h$ makes $\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}$ continuous in t. Letting

$$g : [0, L_\gamma] \longrightarrow \mathbb{R}, \qquad g := \begin{cases} \mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1 & \text{on } [0, L_\gamma] \setminus \left(\bigcup_n E_n \cup \{t\}\right), \\ h & \text{on } \bigcup_n E_n \cup \{t\} \end{cases}$$

concludes the proof. □

Remark I. The choice $q = 2$ is due to technical reasons, as estimate (3.4) involves computing the difference

$$\nu(A_n)^2 + \nu(B_n)^2 - \big(\nu(A_n) + \nu(B_n)\big)^2/2.$$

However, we are unable to prove that for any $q > 1$ there exists a constant $M_q > 0$ (depending only on q) such that

$$\nu(A_n)^q + \nu(B_n)^q - \big(\nu(A_n) + \nu(B_n)\big)^q/2 \geq M_q \left(\nu(A_n) - \nu(B_n)\right)^q.$$

This would allow to extend the result for any $q > 1$ (or for any $q > 1$ for which a similar estimate holds), by using the same argument found in the proof of Proposition 3.4.

Now we can prove conclusion *(2)* of Theorem 3.2.

Proof. (of conclusion *(2)* of Theorem 3.2) Consider an arbitrary $t \in [0, L_\gamma]$: note that conclusion *(2)* states a local property, thus we need only to consider times close to t. If $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ is constant in a neighborhood of t, then conclusion *(2)* follows with $g := \mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$. Otherwise, choose two (mutually disjoint) sequences $\{t_n\} \to t$, $\{s_n\} \to t$, $\{t_n\}, \{s_n\} \subseteq \Lambda$ (with Λ denoting the set of Lebesgue points of $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$) such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}(t_n) > \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}(s_n)$ for any $n \in \mathbb{N}$. For any n choose a sequence $\{\delta_{n,j}\}$, such that $\{\delta_{n,j}\} \to 0$ if $j \to +\infty$ or $n \to +\infty$, and let

$$I_{n,j} := (t_n - \delta_{n,j}, t_n + \delta_{n,j}), \qquad J_{n,j} := (s_n - \delta_{n,j}, s_n + \delta_{n,j}), \qquad j = 1, 2, \cdots.$$

Clearly choosing sufficiently small $\delta_{n,j}$ ensures $I_{n,j} \cap J_{n,j} = \emptyset$ for any n, j. Since $\{t_n\}, \{s_n\} \subseteq \Lambda$, it follows (upon choosing sufficiently small $\delta_{n,j}$) $\nu(I_{n,j}) > \nu(J_{n,j})$ for any n, j. Fix arbitrary n, j. Choose $C_{n,j} \subseteq I_{n,j}$ such that $\nu(C_{n,j}) = \big(\nu(I_{n,j}) - \nu(J_{n,j})\big)/2$, let

$$\tilde{\nu}_{n,j} := \nu - \nu_{\llcorner C_{n,j}} + \frac{\nu(I_{n,j}) - \nu(J_{n,j})}{2\mathcal{L}^1(J_{n,j})} \mathcal{L}^1_{\llcorner J_{n,j}}, \tag{3.8}$$

and choose an arbitrary optimal plan $\tilde{\Pi}_{n,j}$ between μ and $\gamma_\sharp\tilde{\nu}_{n,j}$. Note that

$$\begin{aligned}\int_{I_{n,j}\cup J_{n,j}} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 - \int_{I_{n,j}\cup J_{n,j}} \left(\frac{\mathrm{d}\tilde{\nu}_{n,j}}{\mathrm{d}\mathcal{L}^1}\right)^2 \mathrm{d}\mathcal{L}^1 \\ \geq \frac{\nu(I_{n,j})^2+\nu(J_{n,j})^2}{\delta_{n,j}} - \frac{2}{\delta_{n,j}}\left(\frac{\nu(I_{n,j})+\nu(J_{n,j})}{2}\right)^2 \\ = \frac{\big(\nu(I_{n,j})-\nu(J_{n,j})\big)^2}{2\delta_{n,j}}. \qquad (3.9)\end{aligned}$$

Note that by construction, any point $x \in \mathrm{supp}(\mu)$ transported by Π on $\gamma(C_{n,j})$ is transported by $\tilde{\Pi}_{n,j}$ on $\gamma(J_n)$, i.e. for μ-a.e. x such that there exists $y \in \gamma(C_{n,j}) \subseteq \gamma(I_{n,j})$ satisfying $(x,y) \in \mathrm{supp}(\Pi)$, there exists also $y' \in \gamma(J_{n,j})$ such that $(x,y') \in \mathrm{supp}(\tilde{\Pi})$. Since

$$(\forall n,j) \qquad \sup_{z\in I_{n,j},\, w\in J_{n,j}} |z-w| \leq |t_n - s_n| + 2\delta_{n,j},$$

and γ is arc-length parameterized, it follows

$$(\forall n,j) \qquad \sup_{z\in I_{n,j},\, w\in J_{n,j}} |\gamma(z)-\gamma(w)| \leq |t_n - s_n| + 2\delta_{n,j}. \qquad (3.10)$$

Since (γ,ν,Π) is a minimizer, it follows

$$\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',p,2](\gamma,\nu,\Pi) \leq (\mathrm{diam}\,\mathrm{supp}(\mu)+1)^p + \lambda + \varepsilon + \varepsilon' =: M,$$

and applying Lemma 2.2 (with $\{(\gamma_n,\nu_n,\Pi_n)\} \equiv (\gamma,\nu,\Pi)$) gives

$$\sup_{x\in\mathrm{supp}(\mu),\, y\in\Gamma_\gamma} |x-y| \leq M^{1/p} + M/\lambda. \qquad (3.11)$$

Note that $M^{1/p} + M/\lambda \geq 1$. Applying Lemma 3.3 (with $a := |x-y|$, $b := |x-y'|$, $K := M^{1/p}+M/\lambda$) yields

$$\big||x-y|^p - |x-y'|^p\big| \leq \big||x-y| - |x-y'|\big| \cdot p(M^{1/p}+M/\lambda)^{p-1},$$

which gives

$$\begin{aligned}\int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\tilde{\Pi}_{n,j}(x,y) - \int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) \\ \overset{(3.10)}{\leq} \nu(C_{n,j})\big(|t_n-s_n|+2\delta_{n,j}\big)p(M^{1/p}+M/\lambda)^{p-1} \\ = \frac{\nu(I_{n,j})-\nu(J_{n,j})}{2}\big(|t_n-s_n|+2\delta_{n,j}\big)p(M^{1/p}+M/\lambda)^{p-1}. \qquad (3.12)\end{aligned}$$

Combining estimates (3.9) and (3.12) with the minimality of (γ,ν,Π) gives

$$\frac{\nu(I_{n,j})-\nu(J_{n,j})}{2}\big(|t_n-s_n|+2\delta_{n,j}\big)p(M^{1/p}+M/\lambda)^{p-1} \geq \frac{\varepsilon\big(\nu(I_{n,j})-\nu(J_{n,j})\big)^2}{2\delta_{n,j}}. \qquad (3.13)$$

Note that the above construction can be repeated for any n, j, and since $\{t_n\}$, $\{s_n\} \subseteq \Lambda$, passing to the limit $j \to +\infty$ inequality (3.13) becomes

$$\left| \frac{d\nu}{d\mathcal{L}^1}(t_n) - \frac{d\nu}{d\mathcal{L}^1}(s_n) \right| \leq \frac{p(M^{1/p} + M/\lambda)^{p-1}}{\varepsilon} |t_n - s_n|,$$

concluding the proof. □

Acknowledgements. The author warmly thanks the Center of Nonlinear Analysis (NSF grant DMS-0635983), where part of this research was carried out, and acknowledges the support by ICTI and FCT (grant UTA_CMU/MAT/0007/2009). The author is grateful to Filippo Santambrogio, Dejan Slepčev and Eugene Stepanov for valuable comments and suggestions, and to Universidade Nova de Lisboa for its hospitality. This research started when the author was postdoc at Center of Nonlinear Analysis of Carnegie Mellon University, and completed at Instituto Superior Técnico.

REFERENCES

[1] L. Ambrosio, N. Gigli and G. Savaré, "Gradient flow in metric spaces and in the space of probability measures", Second Editon, Lectures in Mathematics, ETH Zürich, Birkenhäuser Verlag, Basel, 2005

[2] G. Buttazzo, E. Mainini and E. Stepanov, *Stationary configurations for the average distance functional and related problems* , Control Cybernet., **38-4A** (2009), 1107–1130

[3] G. Buttazzo, E. Oudet and E. Stepanov, *Optimal transportation problems with free Dirichlet regions* , Progr. Nonlinear Differential Equations Appl., **51** (2002), 41–65

[4] G. Buttazzo and E. Stepanov, *Minimization problems for average distance functionals*, in "Calculus of Variations: Topics from the Mathematical Heritage of Ennio De Giorgi" (ed. D. Pallara), Quaderni di Matematica, Seconda Università di Napoli (2004)

[5] G. Buttazzo and E. Stepanov, *Optimal transportation networks as free Dirichlet regions for the Monge-Kantorovich problem*, Ann. Sc. Norm. Sup. Pisa Cl. Sci., **2** (2003), 631–678

[6] P. Drineas, A. Frieze, R. Kannan, S. Vempala and V. Vinay, *Clustering large graphs via the singular value decomposition*, Mach. Learn., **56** (2004), 9–33

[7] T. Duchamp and W. Stuetzle, *Extremal properties of principal curves in the plane*, Ann. Statist., **24-4** (1996), 1511–1520

[8] T. Duchamp and W. Stuetzle, *Geometric properties of principal curves in the plane*, "Robust Statistics, Data Analysis, and Computer Intensive Methods Lecture Notes in Statistics" (ed. H. Rieder), Springer-Verlag (1996)

[9] A. Fischer, *Selecting the length of a principal curve within a Gaussian model*, Electron. J. Statist., **7** (2013), 342–363

[10] T. Hastie, *Principal curves and surfaces*, PhD Thesis, Stanford Univ., 1984

[11] T. Hastie and W. Stuetzle, *Principal curves*, J. Amer. Statist. Assoc., **84** (1989), 502–516

[12] B. Kégl, *Principal curves: learning, design, and applications*, PhD thesis, Concordia Univ., 1999

[13] K. Kégl and K. Aetal, *Learning and design of principal curves*, IEEE Trans. Pattern Anal. Mach. Intell., **22-3** (2000), 281–297

[14] A. Lemenant, *A presentation of the average distance minimizing problem*, POMI, **390** (2011), 117–146

[15] X.Y. Lu, *Example of minimizer of the average distance problem with non closed set of corners*, Preprint (available at `http://cvgmt.sns.it/paper/2305/`)

[16] X.Y. Lu and D. Slepčev, *Properties of minimizers of average distance problem via discrete approximation of measures*, SIAM J. Math. Anal., **45-5** (2013), 3114–3131

[17] U. Ozertem and D. Erdogmus, *Locally defined principal curves and surfaces*, J. Mach. Learn. Res., **12-4** (2011), 1249–1286

[18] E. Paolini and E. Stepanov, *Qualitative properties of maximum and average distance minimizers in* $\mathbb{R}^n$, J. of Math. Sci. (N. Y.), **122-3** (2004), 3290–3309

[19] P. Polak and G. Wolanski, *The lazy traveling salesman problem in* $\mathbb{R}^2$, ESAIM: Control Optim. Calc. Var., **13** (2007), 538–552
[20] D. Slepčev, *Counterexample to regularity in average-distance problem*, Ann. Inst. H. Poincaré (C), **31-1** (2014), pp. 169–184
[21] A.J. Smola, S. Mika, B. Schölkopf and R.C. Williamson, *Regularized principal manifolds*, J. Mach. Learn., **1** (2001), 179–209
[22] R. Tibshirani, *Principal curves revisited*, Stat. Comput., **2** (1992), 183–190
[23] C. Villani, "Optimal transport, old and new", Grundlehren der mathematischen Wissenschaften, Springer, 2009

(1) DEPARTMENT OF MATHEMATICS, INSTITUTO SUPERIOR TÉCNICO, 1 AVENIDA ROVISCO PAIS, 1049-001 LISBON, PORTUGAL

(2) DEPARTMENT OF MATHEMATICAL SCIENCES, CARNEGIE MELLON UNIVERSITY, 5000 FORBES AVENUE, 15213 PITTSBURGH, PA, UNITED STATES, EMAIL: XINYANG@ANDREW.CMU.EDU