REGULARITY OF DENSITIES IN RELAXED AND PENALIZED AVERAGE DISTANCE PROBLEM

Xin Yang Lu

Department of Mathematical Sciences, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, United States.

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. The original formulation of the average distance problem exhibits several undesirable properties. In this paper we propose an alternative variant: we minimize the functional

$$\int_{\mathbb{R}^d\times\Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y) + \lambda L_\gamma + \varepsilon\alpha(\nu) + \varepsilon'\eta(\gamma) + \varepsilon''\|\gamma'\|_{TV},$$

where γ varies among the family of parametrized curves, ν among probability measures on γ, and Π among transport plans between μ and ν. Here $\lambda, \varepsilon, \varepsilon', \varepsilon''$ are given parameters, α is a penalization term on μ, Γ_γ (resp. L_γ) denotes the graph (resp. length) of γ, and $\|\cdot\|_{TV}$ denotes the total variation seminorm. We will use techniques from optimal transport theory and calculus of variations. The main aim is to prove essential boundedness, and Lipschitz continuity for Radon-Nikodym derivative of ν, when (γ, ν, Π) is a minimizer.

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 [7], Smola, Mika, Schölkopf and Williamson [23]). 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 [5, 6], Paolini and Stepanov [20]. A similar formulation, often referred to as "penalized formulation", was introduced by Buttazzo, Mainini and Stepanov introduced in [2]:

Problem 1. *Given $d \geq 2$, a nonnegative, compactly supported measure μ and a parameter $\lambda > 0$, minimize*

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

1991 *Mathematics Subject Classification.* Primary: 49Q20, 49K10, 49Q10, 35B65.

Key words and phrases. nonlocal variational problem, average-distance, optimal transport, Kantorovich potential, regularity.

The author is supported by NSF grant DMS-0635983.

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 Golab'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 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_μ^λ corresponds to finding the "best" one dimensional parameterization, which "balances" approximation error and complexity.

Moreover, in data analysis the unknown if often restricted 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

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

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

$$L_{\gamma^*} := \int_0^1 |(\gamma^*)'| \,\mathrm{d}s \tag{1}$$

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

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

Elements of $\mathcal{C}^*$ will be referred to as *"constant speed parameterized curves"*, while elements of $\mathcal{C}$ will be referred to as *"arc-length parameterized curves"*. Thus if $\gamma \in \mathcal{C}$ then $L_\gamma = a$, and its domain is $[0, L_\gamma]$. The average distance problem becomes:

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

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

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, the 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 2 is related to "principal curves", and the lazy traveling salesman problem (see for instance Polak and Wolanski [21]). 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 [8, 9], Fischer [10], Hastie [12], Hastie and Stuetzle [13], Kégl [14], Kégl and Aetal [15], Ozertem and Erdogmus [19], Tibshirani [24].

However the formulation of Problem 2 still exhibits several undesirable properties when used in data parameterization:

(1) it has been proven (Slepčev [22]) that even assuming $\mu \ll \mathcal{L}^d$ with $\mathrm{d}\mu/\,\mathrm{d}\mathcal{L}^d \in C^\infty$, Problem 1 may admit minimizers which are simple curves failing to be C^1 regular. Moreover, any simple curve minimizing Problem 1 admits a parameterization $\gamma \in \mathcal{C}$ minimizing Problem 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 [18]. In data parameterization, this corresponds to a loss of information, which is undesirable.

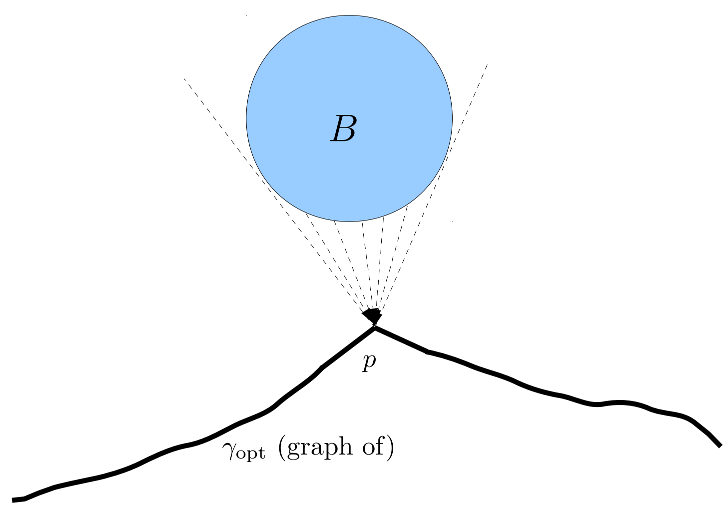

FIGURE 1. In this example from [22], the set $B \subseteq \operatorname{supp}(\mu)$ of positive μ-measure is projected on the single point p (which is a corner), on which C^1 regularity fails.

(2) The aforementioned configuration is a limit case of a more general issue: indeed in the formulation of Problem 2 there is no penalization for very high (even infinite) data concentration on the representation.
(3) In [17] it has been proven that Problem 1 may admit minimizers which are simple curves (thus these admit parameterizations minimizing Problem 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 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 $d_\gamma\big(\gamma(s),\gamma(t)\big) := |s-t|$, $s,t\in[0,L_\gamma]$.

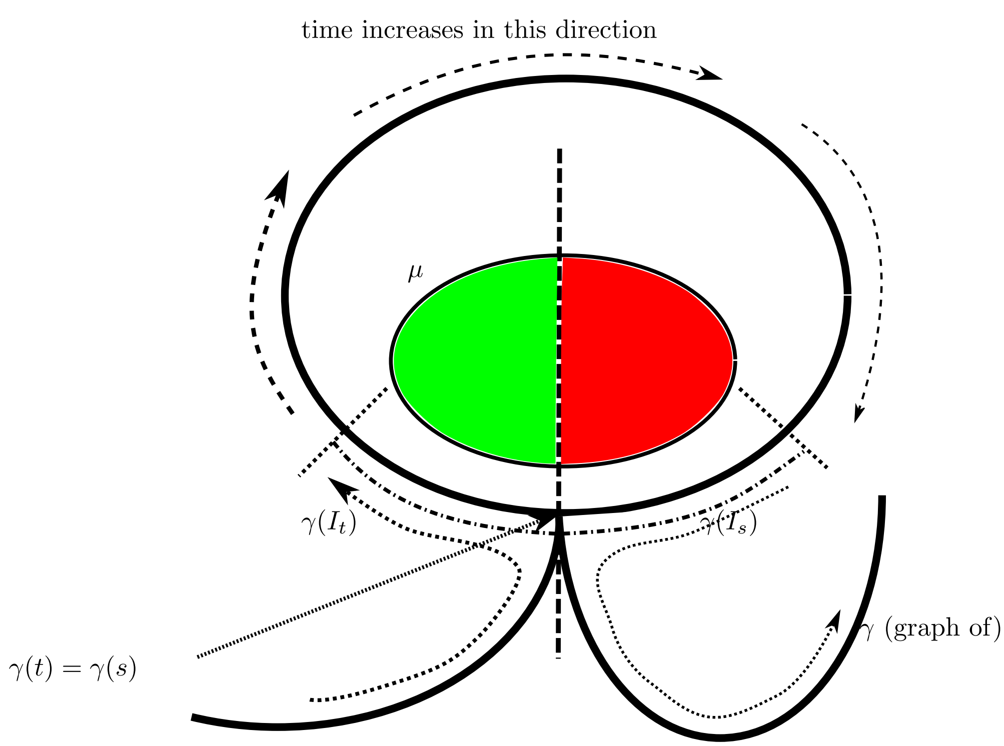

FIGURE 2. In this configuration, assuming $t<s$, points belonging to the red part are projected on $\gamma(I_s)$, while points belonging to the green part are projected on $\gamma(I_t)$. The sets $\gamma(I_s)$ and $\gamma(I_t)$ are distant with respect to d_γ. The colored area is part of $\mathrm{supp}(\mu)$.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

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 points on the curve which is closest. This imposes strong geometric rigidity on minimizers.

Thus we propose an alternative variant:

Problem 3. *Given $d \geq 2$, a measure μ, and parameters $\lambda, \varepsilon, \varepsilon', \varepsilon'' > 0$, $p \geq 1$, $q > 1$, solve*

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

where

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

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

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

$$\int_0^{L_\gamma} \nu^q \,\mathrm{d}\mathcal{L}^1 := \begin{cases} \displaystyle\int_0^{L_\gamma} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}s & \text{if } \nu \ll \mathcal{L}^1,\\ +\infty & \text{otherwise.}\end{cases} \tag{2}$$

Here, and for future reference, $\|\cdot\|_{TV}$ denotes the total variation semi-norm.

The convergence in $\mathcal{T}$ will be detailed in Section 2. Note that the formulation of Problem 3 is quite different from classical average distance problem, and resembles the Monge-Kantorovich problem. Existence of minimizers will be proven in Lemma 2.1. For future reference $\int_0^{L_\gamma} \nu^q \,\mathrm{d}s$ will be referred to 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 to 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, while $\varepsilon''\|\gamma'\|_{TV}$ penalizes large total curvature (the term $\|\gamma'\|_{TV}$ is exactly the generalized total curvature, considered as a measure);
- $\varepsilon\int_0^{L_\gamma} \nu^q \,\mathrm{d}\mathcal{L}^1$ 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 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 points on the curve which are closest). However, projecting on a distant point increases the transport cost, and is advantageous only if it 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 3,
- in Section 3 we prove that for any minimizer (γ,ν,Π) of Problem 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 $1<q\le 2$, then $\mathrm{d}\nu/\,\mathrm{d}\mathcal{L}^1$ is Lipschitz continuous.

2. Preliminaries

The aim of this section is to present preliminary notions and results. The main result is existence of minimizers for Problem 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:

- $\{L_{\gamma_n}\}\to L_\gamma$,
- the sequence $\{\gamma_n^*\}$ converges to γ^* uniformly, where γ^*, γ_n^* denote the constant speed reparameterizations. That is,

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

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 $\{\gamma_n\}\xrightarrow{\mathcal{C}}\gamma$, $\{\nu_n\}\overset{*}{\rightharpoonup}\nu$, and $\{\Pi_n\}\overset{*}{\rightharpoonup}\Pi$.

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 all dimensions greater or equal to 2) for all quantities.

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

The proof will be split over several lemmas. Note that the set

$$\{\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q]<+\infty\}$$

is non empty: indeed choose arbitrary points $x\in\operatorname{supp}(\mu)$, $y\in B(x,1)$, 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

$$\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q](\psi,\mathcal{L}^1_{\llcorner[0,1]},\Pi)\le(\operatorname{diam}\operatorname{supp}(\mu)+1)^p+\lambda+\varepsilon+\varepsilon'<+\infty. \tag{3}$$

In particular, it follows that for any minimizing sequence $\{\gamma_n\}$, it holds $\sup_n\eta(\gamma_n)<+\infty$, $\sup_n\|\gamma_n'\|_{TV}<+\infty$.

Lemma 2.2. *Given $d\ge 2$, a measure μ, parameters $\lambda,\varepsilon,\varepsilon',\varepsilon''>0$, $p\ge 1$, $q>1$, $M\ge\inf\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q]$, and a sequence*

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

then it holds:

(1) length estimate:

$$0 < (M/\varepsilon)^{\frac{1}{1-q}} \le \inf_n L_{\gamma_n} \le \sup_n L_{\gamma_n} \le M/\lambda < +\infty, \tag{4}$$

(2) confinement condition:

$$\bigcup_n \Gamma_{\gamma_n} \subseteq \big(\operatorname{supp}(\mu)\big)_{M^{1/p}+M/\lambda}, \tag{5}$$

where for given $r \ge 0$,

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

Proof. Length estimate. Note that

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

proving the upper bound in (4).

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

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

The last inequality holds since, by Hölder inequality, we have

$$1 = \int_0^{L_{\gamma_n}} \frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\,\mathrm{d}\mathcal{L}^1 \le \left(\int_0^{L_{\gamma_n}} \left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}\mathcal{L}^1\right)^{1/q} L_{\gamma_n}^{\frac{q-1}{q}}$$

$$\Longrightarrow \left(\int_0^{L_{\gamma_n}} \left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q \mathrm{d}\mathcal{L}^1\right)^{1/q} \ge L_{\gamma_n}^{\frac{1-q}{q}}.$$

Since $q > 1$, it follows $L_{\gamma_n}^{1-q} \le M/\varepsilon$, proving the lower bound in (4).

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

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

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

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

Using length estimate $\sup_n L_{\gamma_n} \le M/\lambda$ 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 (5). □

We remark that for any $(\gamma,\nu,\Pi) \in \mathcal{T}$ satisfying $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q](\gamma,\nu,\Pi) < +\infty$ it holds $\nu \ll \mathcal{L}^1$.

Lemma 2.3. *For any $\gamma \in \mathcal{C}$ it holds*

$$\eta(\gamma) < +\infty \Longrightarrow \gamma \text{ injective.} \tag{6}$$

Proof. Assume there exist $t_0, s_0 \in [0, L_\gamma]$ with $t_0 < s_0$, $\gamma(t_0) = \gamma(s_0)$. Choose sufficiently small $r, a > 0$ such that $t_0 + r < s_0 - r - a$, and

$$\begin{aligned}\eta(\gamma) &= \int_0^{L_\gamma}\int_0^{L_\gamma} \frac{|s-t|^2}{|\gamma(s)-\gamma(t)|^2}\,\mathrm{d}s\,\mathrm{d}t \geq \int_{t_0}^{t_0+r}\int_{s_0-r}^{s_0} \frac{|s-t|^2}{|\gamma(s)-\gamma(t)|^2}\,\mathrm{d}s\,\mathrm{d}t \\ &\geq a^2 \int_{t_0}^{t_0+r}\int_{s_0-r}^{s_0} \frac{1}{|\gamma(s)-\gamma(t)|^2}\,\mathrm{d}s\,\mathrm{d}t.\end{aligned}$$

Since γ is arc-length parameterized, it holds

$$\begin{aligned}&|\gamma(s)-\gamma(t)| \leq |\gamma(s)-\gamma(s_0)| + |\gamma(t)-\gamma(t_0)| + |\gamma(s_0)-\gamma(t_0)| \\ &\Longrightarrow |\gamma(s)-\gamma(t)|^2 \leq (|\gamma(s)-\gamma(s_0)| + |\gamma(t)-\gamma(t_0)|)^2 \\ &\qquad \leq 2|\gamma(s)-\gamma(s_0)|^2 + 2|\gamma(t)-\gamma(t_0)|^2 \leq 2|s_0-s|^2 + 2|t_0-t|^2 \\ &\Longrightarrow \frac{1}{|\gamma(s)-\gamma(t)|^2} \geq \frac{1}{2|s-s_0|^2+2|t-t_0|^2},\end{aligned}$$

which gives

$$\int_{t_0}^{t_0+r}\int_{s_0-r}^{s_0} \frac{1}{|\gamma(s)-\gamma(t)|^2}\,\mathrm{d}s\,\mathrm{d}t \geq \int_{t_0}^{t_0+r}\int_{s_0-r}^{s_0} \frac{1}{2|s-s_0|^2+2|t-t_0|^2}\,\mathrm{d}s\,\mathrm{d}t = +\infty,$$

concluding the proof. □

Lemma 2.4. *Given a sequence of piece-wise linear functions* $\{f_n\} : [0,1] \to \mathbb{R}$*, uniformly converging to the identically zero function, and such that* $\int_0^1 |f_n'|\,\mathrm{d}s \to c > 0$*, then* $\|f_n'\|_{TV} \to +\infty$.

We use the proof suggested by a referee.

Proof. Consider an arbitrary $f \in C^\infty([0,1])$. By the mean value theorem, there exists $c \in [0,1]$ such that $f'(c) = f(1) - f(0)$. Therefore,

$$\|f'\|_{L^\infty} \leq 2\|f\|_{L^\infty} + \|f''\|_{L^1}.$$

Direct computation gives

$$\begin{aligned}\left(\int_0^1 |f'(s)|\,\mathrm{d}s\right)^2 &\leq \int_0^1 |f'(s)|^2\,\mathrm{d}s \\ &\leq \int_0^1 |f(s)||f''(s)|\,\mathrm{d}s + f(1)f'(1) - f(0)f'(0) \\ &\leq \|f\|_{L^\infty}\|f''\|_{L^1} + 4\|f\|_{L^\infty}^2 + 2\|f\|_{L^\infty}\|f''\|_{L^1} \\ &\leq 4(\|f\|_{L^\infty}^2 + \|f\|_{L^\infty}\|f'\|_{TV}).\end{aligned}$$

The proof for general f follows by a density argument. □

Lemma 2.5. *Given a sequence of constant (positive) speed curves* $\{\gamma_n\} : [0,1] \to \mathbb{R}^d$*, converging uniformly to* $\gamma : [0,1] \to \mathbb{R}^d$*, such that*

$$\sup_n \|\gamma_n'\|_{TV} < +\infty, \qquad \sup_n \int_0^1\int_0^1 \frac{|s-t|^2}{|\gamma_n(s)-\gamma_n(t)|^2}\,\mathrm{d}s\,\mathrm{d}t < +\infty, \tag{7}$$

then it holds

$$L(\gamma) = \lim_{n\to+\infty} L(\gamma_n).$$

Note that this is a much stronger result then the general lower semicontinuity of length. In particular, due to the curvature penalization, it states that any minimizing sequence $\{\gamma_n\}$ (which surely satisfies (7) in view of (3)), admitting a uniform limit γ, then $L(\gamma_n) \to L(\gamma)$. This will be crucial for the proof of Lemma 2.1. We use the proof suggested by a referee.

Proof. Boundedness of both sequences $\|\gamma_n'\|_{TV}$ and $\|\gamma_n\|_{L^\infty}$ imply boundedness of $\|\gamma_n'\|_{L^1}$. Since the embedding from $BV(0,1)$ into $L^1(0,1)$ is compact, boundedness (and thus, upon subsequences, weak convergence) of γ_n' in $BV(0,1)$ gives strong convergence in $L^1(0,1)$, hence strong convergence of length. $\square$

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 (3))

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

assume without loss of generality $\sup_n \mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', \varepsilon'', p, q](\gamma_n, \nu_n, \Pi_n) \leq 2M$. 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

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

be constant speed reparameterizations. Lemma 2.2 proves that the sequence $\{\gamma_n^*\}$ satisfies the conditions of Ascoli-Arzelà theorem, thus (upon subsequence, which will not be relabeled) there exists $\gamma^* : [0,1] \longrightarrow \mathbb{R}^d$ (not necessarily parameterized by constant speed) such that $\{\gamma_n^*\} \to \gamma^*$ uniformly. Note that (upon subsequence, which we do not relabel) $L_\gamma^* = \lim_{n\to+\infty} L_{\gamma_n^*}$ in view of Lemma 2.5. Define the measures ν_n^* as

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

where $BL_{\gamma_n} := \{t \in [0, L_{\gamma_n}] : t/L_{\gamma_n} \in B\}$. 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

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

Thus

$$\{\nu_n\} \overset{*}{\rightharpoonup} \nu, \qquad \int_0^{L_\gamma} \nu^q \,\mathrm{d}\mathcal{L}^1 \leq \liminf_n \int_0^{L_{\gamma_n}} \nu_n^q \,\mathrm{d}\mathcal{L}^1, \tag{8}$$

where ν is defined as

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

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) $\{\Pi_n\} \overset{*}{\rightharpoonup} \Pi$, and Π is a transport plan between μ and $\gamma_\sharp\nu$ (for further details about stability of transport plans, we refer to [1, 25] and references therein), hence

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d \times \Gamma_{\gamma_n}} |x-y|^p \,\mathrm{d}\Pi_n(x,y) = \int_{\mathbb{R}^d \times \Gamma_\gamma} |x-y|^p \,\mathrm{d}\Pi(x,y). \tag{9}$$

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 $\{g_n\}\to g$ point-wise. Fatou's lemma gives

$$\int_0^1 \int_0^1 g(s,t) \,\mathrm{d}s \,\mathrm{d}t \le \liminf_n \int_0^1 \int_0^1 g_n(s,t) \,\mathrm{d}s \,\mathrm{d}t.$$

Note that $\int_0^1 \int_0^1 g(s,t) \,\mathrm{d}s \,\mathrm{d}t$ and $\int_0^1 \int_0^1 \frac{|s-t|^2}{|\gamma(s)-\gamma(t)^2|} \,\mathrm{d}s \,\mathrm{d}t$ differ by the multiplicative constant L_γ^2, and $\|(\gamma^*)'\|_{TV} = \|\gamma'\|_{TV}$. Similarly for the ratio between

$$\int_0^1 \int_0^1 g_n(s,t) \,\mathrm{d}s \,\mathrm{d}t$$

and

$$\int_0^1 \int_0^1 \frac{|s-t|^2}{|\gamma_n(s) - \gamma_n(t)^2|} \,\mathrm{d}s \,\mathrm{d}t,$$

and $\|(\gamma_n^*)'\|_{TV} = \|\gamma_n'\|_{TV}$. Thus it follows

$$\liminf_{n\to+\infty} \eta(\gamma_n) \ge \eta(\gamma), \qquad \liminf_{n\to+\infty} \|\gamma_n'\|_{TV} \ge \|\gamma'\|_{TV}.$$

Since $\{L_{\gamma_n}\}\to L_\gamma$, combining with (8) and (9) gives

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

and the proof is complete. $\square$

We conclude this section with two simple observations. The first is a Γ-convergence result.

Lemma 2.6. *Given $d \ge 2$, a measure μ, parameters $\lambda, \varepsilon'' > 0$, $p \ge 1$, $q > 1$, sequences $\{\varepsilon_n\}, \{\varepsilon_n'\}\to 0$, and $(\gamma,\nu,\Pi) \in \mathcal{T}$, then:*

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

(10)
$$\liminf_n \mathcal{E}[\mu,\lambda,\varepsilon_n,\varepsilon'_n,\varepsilon'',p,q](\gamma_n,\nu_n,\Pi_n)\geq \int_{\mathbb{R}^d\times\Gamma_\gamma}|x-y|^p\,\mathrm{d}\Pi(x,y)+\lambda L_\gamma+\varepsilon''\|\gamma'\|_{TV};$$

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

(11)
$$\limsup_n \mathcal{E}[\mu,\lambda,\varepsilon_n,\varepsilon'_n,\varepsilon'',p,q](\gamma_n,\nu_n,\Pi_n)\leq \int_{\mathbb{R}^d\times\Gamma_\gamma}|x-y|^p\,\mathrm{d}\Pi(x,y)+\lambda L_\gamma+\varepsilon''\|\gamma'\|_{TV};$$

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

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

proving (10).

To prove (11), note that since by hypothesis there exist $\varepsilon,\varepsilon'>0$ such that $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q](\gamma,\nu,\Pi)<+\infty$, it follows that γ is injective in view of (6), and $\nu\ll\mathcal{L}^1$. Let

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

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

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

proving (11). □

Lemma 2.7. *Given* $d\geq 2$, *a measure* μ, *parameters* $\lambda,\varepsilon',\varepsilon''>0$, $p\geq 1$, $q>1$, *a sequence* $\{\varepsilon_n\}\to 0$, *and* $(\gamma,\nu,\Pi)\in\mathcal{T}$, *then there exists a sequence* $\{(\gamma_n,\nu_n,\Pi_n)\}\xrightarrow{\mathcal{T}}(\gamma,\nu,\Pi)$ *such that*

$$\begin{aligned}\lim_n \mathcal{E}[\mu,\lambda,\varepsilon_n,\varepsilon',\varepsilon'',p,q](\gamma_n,\nu_n,\Pi_n)&=\mathcal{E}[\mu,\lambda,0,\varepsilon',\varepsilon'',p,q](\gamma,\nu,\Pi)\\ &=\int_{\mathbb{R}^d\times\Gamma_\gamma}|x-y|^p\,\mathrm{d}\Pi(x,y)+\lambda L_\gamma\\ &\quad+\varepsilon'\eta(\gamma)+\varepsilon''\|\gamma'\|_{TV}.\end{aligned}\tag{12}$$

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

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

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

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 (12) 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

$$\sum_{j=1}^H a_j\delta_{t_j},\qquad H\in\mathbb{N},\quad \sum_{j=1}^H a_j=1,\quad \{t_j\}\subseteq[0,L_\gamma]$$

can be approximated (in the weak-$*$ topology) by measures of the form $\left(\sum_{j=1}^H a_jf_{n,t}\right)\cdot\mathcal{L}^1_{\llcorner[0,L_\gamma]}$. 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

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

thus

$$(\forall n)\qquad\int_0^{L_\gamma}\left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q\mathrm{d}\mathcal{L}^1\le\varepsilon_n^{-1/2}.\tag{13}$$

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

$$\begin{aligned}&\lim_{n\to\infty}\mathcal{E}[\mu,\lambda,\varepsilon_n,\varepsilon',\varepsilon'',p,q](\gamma_n,\nu_n,\Pi_n)\\&=\lim_{n\to+\infty}\int_{\mathbb{R}^d\times\Gamma_\gamma}|x-y|^p\,\mathrm{d}\Pi_n(x,y)+\lambda L_\gamma\\&\qquad+\varepsilon_n\int_0^{L_\gamma}\left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mathcal{L}^1}\right)^q\mathrm{d}\mathcal{L}^1+\varepsilon'\eta(\gamma)+\varepsilon''\|\gamma'\|_{TV}\\&\overset{(13)}{\le}\int_{\mathbb{R}^d\times\Gamma_\gamma}|x-y|^p\,\mathrm{d}\Pi(x,y)+\lambda L_\gamma+\varepsilon'\eta(\gamma)+\varepsilon''\|\gamma'\|_{TV}.\end{aligned}$$

- 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 $\{\gamma_n\} \xrightarrow{\mathcal{C}} \gamma$. 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

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

Thus (12) is proven. Since for any sequence $\{(\gamma_n, \nu_n, \Pi_n)\} \xrightarrow{\mathcal{T}} (\gamma, \nu, \Pi)$ it holds

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

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

3. Regularity of densities

It follow from the definition that if (γ, ν, Π) is a minimizer of Problem 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 $\lambda, \varepsilon, \varepsilon', \varepsilon'' > 0$, $p \geq 1$, $q > 1$, and a minimizer (γ, ν, Π) of $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', \varepsilon'', p, q]$, 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 $\lambda, \varepsilon, \varepsilon', \varepsilon'' > 0$, $p \geq 1$, $1 < q \leq 2$, and a minimizer (γ', ν', Π') of $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', p, q]$, then ν' has Lipschitz regular density.*

Note that given $K \geq 1$, $a, b \in [0, K]$, $p \geq 1$, then it holds

$$|a^p - b^p| \leq |a - b| p K^{p-1}. \tag{14}$$

The proof is straightforward using mean value theorem, which gives $|a^p - b^p| = |(a-b)p\xi^{p-1}|$ with $a \leq \xi \leq b \leq K$.

To prove Theorems 3.1 and 3.2, we use the technique developed by Buttazzo and Santambrogio in [4]. Since very little modification is required, for most of the proofs, we provide a sketch, and refer to [4] for further details. Similarly to the proof of Theorem 3.1 $\eta(\gamma)$ and $\|\gamma'\|_{TV}$ depend only on γ, not on ν or Π. As the construction in the following lemmas does not alter γ, $\eta(\gamma)$ and $\|\gamma'\|_{TV}$ do not change.

We recall the definition of Kantorovich potential in our specific setting.

Definition 3.3. Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $c(x, y) := |x - y|^p$. Given probability measures μ and ν on $\mathbb{R}^d$, a Kantorovich potential ψ is a function such that

$$\inf_{\pi} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \, \mathrm{d}\pi(x, y) = \int \psi \, \mathrm{d}\mu + \int \psi^c \, \mathrm{d}\nu,$$

where π varies among transport plans between μ and ν, and ψ^c is the c-transform of ψ (that is $\psi^c(y) := \inf_x c(x, y) - \psi(x)$). Such a function ψ will be referred as "Kantorovich potential associated to c, μ, ν".

For further discussion about c-concavity, c-transform, and Kantorovich potentials we refer to [1, 11, 25] and references therein. For future reference we will denote by c_p the cost function $c(x, y) := |x - y|^p$.

Lemma 3.4. *Given probability measures μ, ν, $\tilde{\nu}$, let $\nu_t := \nu + t(\tilde{\nu} - \nu)$ $(t > 0)$, and let ψ_t be the Kantorovich potentials associated to c_p, μ, ν_t $(p > 1)$ such that $\psi_t(x_0) \equiv 0$ for some point x_0. Then (upon subsequence) $\psi_t \to \psi$ uniformly, where ψ is the Kantorovich potential associated to c, μ, ν satisfying $\psi(x_0) = 0$.*

Proof. The proof uses the construction from [4, Lemma 3.4]. Some discussion about uniqueness of Kantorovich potentials is required. It is known that, for transport costs of the form $|x - y|^p$, when one of the measures has compact support and a.e. positive absolutely continuous part, then the Kantorovich potential is (upon addictive constants) unique. Thus we will first assume that μ has convex support, and a.e. positive absolute continuous part. The family $\{\psi_t\}$ is equicontinuous since any c-concave function with respect to c_p is Lipschitz, while equiboundedness follows from $\psi_t(x_0) \equiv 0$. Thus Ascoli-Arzelà theorem gives the existence of a uniform limit $\bar{\psi}$, arising from a certain sequence. The optimality of ψ_t gives

$$\int \psi_t \, \mathrm{d}\mu + \int \psi_t^c \, \mathrm{d}\nu_t \geq \int \varphi \, \mathrm{d}\mu + \int \varphi^c \, \mathrm{d}\nu_t$$

for every c-concave function φ. Passing to the limit $t \downarrow 0$, the uniform convergence of $\{\psi_t\}$ has been proven, while the uniform convergence of $\{\psi_t^c\}$ follows (by definition of c-transform) from $|\psi_t^c(x) - \bar{\psi}^c(x)| \leq \|\psi_t - \bar{\psi}\|_{L^\infty}$. Thus passing to the limit $t \downarrow 0$, along a subsequence we get

$$\int \bar{\psi} \, \mathrm{d}\mu + \int \bar{\psi}^c \, \mathrm{d}\nu \geq \int \varphi \, \mathrm{d}\mu + \int \varphi^c \, \mathrm{d}\nu$$

for every c-concave function φ. Thus $\bar\psi$ is a Kantorovich potential associated to c,μ,ν, and $\bar\psi(x_0)=\psi(x_0)$ ensures $\bar\psi=\psi$, and uniqueness of Kantorovich potential ensures that the whole sequence converges to ψ.

If $\operatorname{supp}(\mu)$ is not convex, or $\mathrm{d}\mu/\mathrm{d}\mathcal{L}^d$ is not a.e. positive, then an approximation argument applies. Clearly, by Lemma 2.2, $\operatorname{supp}(\mu)$ and any minimizer (γ,ν,Π), there exists a convex set K (independent of γ,ν,Π) containing $\operatorname{supp}(\mu)$, Γ_γ and x_0. Consider an arbitrary sequence $\varepsilon_n\searrow 0$, and let $\mu_n:=\dfrac{1}{(\mu+\varepsilon_n\mathcal{L}^d)(\mathbb{R}^d)}(\mu+\varepsilon_n\mathcal{L}^d)$. By construction, μ_n has a.e. positive density. Then, we proceed as in [4, Lemma 3.6] (to which we refer for the detailed arguments): denoting by ψ_n the Kantorovich potentials associated to c,μ_n,ν satisfying $\psi_n(x_0)=0$, by Ascoli-Arzelà theorem, upon subsequence, $\psi_n\to\psi$ uniformly for some ψ. It is then straightforward to verify that ψ is a Kantorovich potential associated to c,μ,ν, satisfying $\psi(x_0)=0$. $\square$

Lemma 3.5. *Let ψ be the Kantorovich potential associated to c_p,μ,ν ($p>1$), with (γ,ν,Π) minimizer of $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q]$. Then there exists a constant l such that*

$$q\nu^{q-1}=l-\psi\qquad \mathcal{L}^1\text{-a.e.}.$$

In particular $q\nu^{q-1}$ is H-Lipschitz regular, $H:=p(\operatorname{diam}K)^{p-1}$ and K is a compact set with minimal diameter in the family of compact sets satisfying confinement condition of Lemma 2.2.

Proof. We will use an an approach based on the Kantorovich potential technique developed by Buttazzo and Santambrogio in [4]. Since any minimizer (γ,ν,Π) satisfies $\nu\ll\mathcal{L}^1_{\llcorner[0,L_\gamma]}$, without an abuse of notation we identify ν with its Radon-Nikodym derivative $\mathrm{d}\nu/\mathrm{d}\mathcal{L}^1_{\llcorner[0,L_\gamma]}$. Thus $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q](\gamma,\nu,\Pi)$ can be written as

$$\begin{aligned}\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q](\gamma,\nu,\Pi)=&\int_{\mathbb{R}^d}\int_0^{L_\gamma}|x-\gamma(t)|^p\,\mathrm{d}\mu(x)\,\mathrm{d}t+\lambda L_\gamma\\&+\varepsilon\int_0^{L_\gamma}\nu^q\,\mathrm{d}s+\varepsilon'\eta(\gamma)+\varepsilon''\|\gamma'\|_{TV}.\end{aligned}$$

Since our construction will modify only ν (and consequently Π, but not γ), let

$$\mathcal{F}(\nu):=T_p(\mu,\nu)+F(\nu),$$

$$T_p(\mu,\nu):=\int_{\mathbb{R}^d}\int_0^{L_\gamma}|x-\gamma(t)|^p\,\mathrm{d}\mu(x)\,\mathrm{d}t,\quad F(\nu):=\varepsilon\int_0^{L_\gamma}\nu^q\,\mathrm{d}s.$$

Although $\mathcal{F}$ depends on several quantities, for the sake of brevity we omit writing them explicitly. Minimality of (γ,π,Π) gives $\nu\in\operatorname{argmin}\ \mathcal{F}$. Note that Lemma 2.2 gives the existence of a compact set K such that $\operatorname{supp}\Pi\subseteq K$ for any optimal plan Π. Thus $|x-y|^p\le H|x-y|$, i.e. c_p is H-Lipschitz. Consider an arbitrary probability measure $\tilde\nu$ with smooth density (with an abuse of notation we identify $\tilde\nu$ with its Radon-Nikodym derivative $\mathrm{d}\tilde\nu/\mathrm{d}\mathcal{L}^1$), and let $\nu_t:=\nu+t(\tilde\nu-\nu)$. Minimality of ν gives

$$T_p(\mu,\nu_t)+F(\nu_t)-T_p(\mu,\nu)-F(\nu)\ge 0. \tag{15}$$

Also

$$T_p(\mu,\nu_t)=\int_{\mathbb{R}^d}\psi_t\,\mathrm{d}\mu+\int_0^{L_\gamma}\psi_t^c\,\mathrm{d}\nu_t,$$
$$T_p(\mu,\nu)=\int_{\mathbb{R}^d}\psi\,\mathrm{d}\mu+\int_0^{L_\gamma}\psi^c\,\mathrm{d}\nu\geq\int_{\mathbb{R}^d}\psi_t\,\mathrm{d}\mu+\int_0^{L_\gamma}\psi_t^c\,\mathrm{d}\nu,$$

where ψ_t (resp. ψ) are Kantorovich potentials associated to c,μ,ν_t (resp. c,μ,ν), hence (15) reads

$$\int_0^{L_\gamma}\psi_t^c\,\mathrm{d}(\nu_t-\nu)+F(\nu_t)-F(\nu)=t\int_0^{L_\gamma}\psi_t^c\,\mathrm{d}(\tilde{\nu}-\nu)+F(\nu_t)-F(\nu)\geq 0.$$

Dividing by t and passing to the limit $t\to 0^+$ gives

$$\int_0^{L_\gamma}\psi^c\,\mathrm{d}(\tilde{\nu}-\nu)+q\int_0^{L_\gamma}\nu^{q-1}(\tilde{\nu}-\nu)\,\mathrm{d}s\geq 0$$

$$\Longrightarrow\int_0^{L_\gamma}(\psi^c+q\nu^{q-1})\tilde{\nu}\,\mathrm{d}s\geq\int_0^{L_\gamma}(\psi^c+q\nu^{q-1})\nu\,\mathrm{d}s.$$

Since c is bounded, it follows $\psi,\psi^c\in L^\infty([0,L_\gamma])$ (for further details we refer to [25, Chapter 2] and references therein). The arbitrariness of $\tilde{\nu}$ allows to make the difference

$$\left|\int_0^{L_\gamma}(\psi^c+q\nu^{q-1})\tilde{\nu}\,\mathrm{d}s-\operatorname{essinf}(\psi^c+q\nu^{q-1})\right|$$

arbitrarily small. Thus

$$\operatorname{essinf}(\psi^c+q\nu^{q-1})\geq\int_0^{L_\gamma}(\psi^c+q\nu^{q-1})\nu\,\mathrm{d}s\geq\operatorname{essinf}(\psi^c+q\nu^{q-1}),$$

that is

$$\psi^c+q\nu^{q-1}=\operatorname{essinf}(\psi^c+q\nu^{q-1})=:l\quad\mathcal{L}^1\text{-a.e.}\tag{16}$$

since ν is a probability measure. Regularity of $q\nu^{q-1}$ follows immediately. □

We need to establish an analogous of Lemma 3.5 when $p=1$. The main issue is the lack of differentiability of the transport cost. An approximation argument (from [4, Lemma 3.7], to which we refer for further details) will be used. Note that (using the same arguments from [4, Lemma 3.5]) for fixed μ, ν, it holds $T_p(\mu,\nu)\xrightarrow{\Gamma}T_1(\mu,\nu)$ in the weak-* topology (here $\xrightarrow{\Gamma}$ denotes Γ-convergence). Indeed it follows from the proof of Lemma 3.5 that c_p is Lipschitz regular with the same Lipschitz constant as p gets close to 1, hence

$$W_1(\mu,\nu)\leq W_p(\mu,\nu)\leq CW_1(\mu,\nu)$$

for some constant depending on H and independent of p.

Lemma 3.6. *Let ψ be the Kantorovich potential associated to c_1,μ,ν, with (γ,ν,Π) minimizer of $\mathcal{E}[\mu,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q]$ (i.e. ν minimizer of $\mathcal{F}$). Then there exists a constant l such that*

$$q\nu^{q-1}=l-\psi\qquad\mathcal{L}^1\text{-a.e.}.$$

Proof. For any $p > 1$, Lemma 3.5 gives the existence of a unique Kantorovich potential associated to c_p, μ, ν_p, where ν_p minimizes $\mathcal{F}$ (for any p, recall that the definition of $\mathcal{F}$ depends on p). Moreover $q\nu_p^{q-1} = -\psi_p$, and all $q\nu_p^{q-1}$ are H-Lipschitz regular (constant H from Lemma 3.5). Thus upon subsequence $\nu_p \to \nu$ and $\psi_p \to \psi$ uniformly. Clearly ψ is Lipschitz regular with Lipschitz constant at most $\liminf_{p\to 1^+} H$ (recall that H depends on p), and consequently c_1-concave. We need to check ψ is a Kantorovich potential associated to c_1, μ, ν. Recall that for any cost function c and real function φ it holds $\varphi^{cc} \geq \varphi$, and φ^{cc} is c-concave function whose c-transform is $\varphi^{ccc} = \varphi^c$. The optimality of ψ_p gives

$$\int \psi_p \, \mathrm{d}\mu + \int \psi_p^{c_p} \, \mathrm{d}\nu_p \geq \int \varphi^{c_p c_p} \, \mathrm{d}\mu + \int \varphi^{c_p c_p} \, \mathrm{d}\nu_p \geq \int \varphi \, \mathrm{d}\mu + \int \varphi^{c_p} d\nu_p. \tag{17}$$

Note that $\{c_p\} \to c_1$ uniformly on compact sets, hence

$$|\varphi_p^{c_p}(x) - \varphi_1^{c_1}(x)| \leq \|c_p - c_1\|_{L^\infty} + \|\varphi_p - \varphi_1\|_{L^\infty},$$

i.e. for any sequence $\{\varphi_p\} \to \varphi_1$, if $\{\varphi_p\} \to \varphi_1$ uniformly then $\{\varphi_p^{c_p}\} \to \varphi_1^{c_1}$ uniformly. Thus passing to the limit in (17) gives

$$\int \psi \, \mathrm{d}\mu + \int \psi^{c_1} \, \mathrm{d}\nu \geq \int \varphi \, \mathrm{d}\mu + \int \varphi^{c_1} \, \mathrm{d}\nu$$

for any φ (thus also for any c_1-concave function φ), concluding the proof. □

Proof. (of Theorem 3.1) Lemmas 3.5 and 3.6 give (for cases $p > 1$ and $p = 1$ respectively) the existence of a constant l such that $q\nu^{q-1} = l - \psi$, where ψ is a Kantorovich potential associated to c_p, μ, ν. In particular, the inverse of the map $t \mapsto t^{q-1}$ (recall that, by hypothesis, we have $q > 1$) is Hölder continuous, thus the density ν is Hölder continuous, hence bounded. □

Proof. (of Theorem 3.2) Lemma 3.5 proved that $q\nu^{q-1}$ is H-Lipschitz regular when $p > 1$. For case $p = 1$, since c_1 is H-Lipschitz regular, ψ and ψ^c are H-Lipschitz (see [25, Chapter 2] for further details), as well as $q\nu^{q-1}$. As by hypothesis $1 < q \leq 2$, and $\nu \in L^\infty$ in view of Theorem 3.1, ν is Lipschitz. □

Scaling properties. The scaling of the energy with respect to homothety is often relevant in data analysis. Given $r > 0$, let

$$T : \mathbb{R}^d \to \mathbb{R}^d, \qquad T(x) := rx.$$

Fix an arbitrary $(\gamma, \nu, \Pi) \in \mathcal{T}$. Let $\gamma_r := T \circ \gamma$, $\nu_r := T_\sharp \nu$, and Π_r optimal plan between $T_\sharp \mu$ and ν_r. By simple change of variable we get

$$\int_{\mathbb{R}^d \times \Gamma_{\gamma_r}} |x-y|^p \, \mathrm{d}\Pi_r(x,y) = r^{p-d} \int_{\mathbb{R}^d \times \Gamma_\gamma} |x-y|^p \, \mathrm{d}\Pi(x,y), \qquad L_{\gamma_r} = rL_\gamma, \tag{18}$$

(19)
$$\int_0^{L_{\gamma_r}} |\nu_r|^q \, \mathrm{d}s = r^{1-q} \int_0^{L_\gamma} |\nu|^q \, \mathrm{d}s, \qquad \eta(\gamma_r) = r^{-2}\eta(\gamma), \qquad \|\gamma_r'\|_{TV} = \|\gamma\|_{TV}.$$

Thus $\mathcal{E}[\mu, \lambda, \varepsilon, \varepsilon', \varepsilon'', p, q]$ does not scale with homothecy, and we infer little information about minimality of (γ_r, ν_r, Π_r) from the minimality of (γ, ν, Π). This because the term transport cost $\int_{\mathbb{R}^d\times\Gamma_\gamma}|x-y|^p\,\mathrm{d}\Pi(x,y)$ is not "weighted" (it has always weight 1). If transport cost is also weighted, then the energy

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

is well-behaved with respect to homothecy: the same arguments from Section 2 give the existence of minimizers for $\bar{\mathcal{E}}[\mu,\xi,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q]$, and (in view of scaling properties (18) and (19))

$$\begin{aligned}
&\bar{\mathcal{E}}[T_\sharp\mu,\xi r^{d-p},\lambda r^{-1},\varepsilon r^{q-1},\varepsilon' r^2,\varepsilon'',p,q](\gamma_r,\nu_r,\Pi_r)\\
&\quad=\frac{\xi}{r^{p-d}}\int_{\mathbb{R}^d\times\Gamma_{\gamma_r}}|x-y|^p\,\mathrm{d}\Pi_r(x,y)+\frac{\lambda}{r}L_{\gamma_r}\\
&\qquad+\frac{\varepsilon}{r^{1-q}}\int_0^{L_{\gamma_r}}\nu_r^q\,\mathrm{d}s+\varepsilon' r^2\eta(\gamma_r)+\varepsilon''\|\gamma_r'\|_{TV}\\
&\quad=\xi\int_{\mathbb{R}^d\times\Gamma_\gamma}|x-y|^p\,\mathrm{d}\Pi(x,y)+\lambda L_\gamma+\varepsilon\int_0^{L_\gamma}\nu_r^q\,\mathrm{d}s+\varepsilon'\eta(\gamma)+\varepsilon''\|\gamma'\|_{TV}\\
&\quad=\bar{\mathcal{E}}[\mu,\xi,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q](\gamma,\nu,\Pi).
\end{aligned}$$

In particular

$$(\gamma_r,\nu_r,\Pi_r)\in\text{argmin }\ \bar{\mathcal{E}}[T_\sharp\mu,\xi r^{d-p},\lambda r^{-1},\varepsilon r^{q-1},\varepsilon' r^2,\varepsilon'',p,q]$$

if and only if

$$(\gamma,\nu,\Pi)\in\text{argmin }\ \bar{\mathcal{E}}[\mu,\xi,\lambda,\varepsilon,\varepsilon',\varepsilon'',p,q].$$

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 Lorenzo Brasco, Filippo Santambrogio, Dejan Slepčev and Eugene Stepanov for valuable comments and suggestions, and to Universidade Nova de Lisboa for its hospitality. The author thanks the anonymous referees for suggesting alternative proofs for Lemmas 2.4 and 2.5, and Theorems 3.1 and 3.2. This research was partly carried out when the author was affiliated with 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** (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 F. Santambrogio, *A mass transportation model for the optimal planning of an urban region*, SIAM J. Math. Anal., **37** (2005), 514–530.

[5] 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), 47–83.

[6] 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.

[7] 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.
[8] T. Duchamp and W. Stuetzle, *Extremal properties of principal curves in the plane*, Ann. Statist., **24** (1996), 1511–1520.
[9] T. Duchamp and W. Stuetzle, *Geometric properties of principal curves in the plane*, in "Robust Statistics, Data Analysis, and Computer Intensive Methods Lecture Notes in Statistics" (ed. H. Rieder), Springer-Verlag (1996), 135–152.
[10] A. Fischer, *Selecting the length of a principal curve within a Gaussian model*, Electron. J. Statist., **7** (2013), 342–363.
[11] W. Gangbo and R.J. McCann, *The geometry of optimal transportation*, Acta Math., **177** (1996), 113–161.
[12] T. Hastie, "Principal curves and surfaces", Ph.D Thesis, Stanford Univ., 1984.
[13] T. Hastie and W. Stuetzle, *Principal curves*, J. Amer. Statist. Assoc., **84** (1989), 502–516.
[14] B. Kégl, "Principal curves: learning, design, and applications", Ph.D thesis, Concordia Univ., 1999.
[15] K. Kégl and K. Aetal, *Learning and design of principal curves*, IEEE Trans. Pattern Anal. Mach. Intell., **22** (2000), 281–297.
[16] A. Lemenant, *A presentation of the average distance minimizing problem*, J. Math. Sci. (N.Y.), **181** (2012), 820–836.
[17] X.Y. Lu, *Example of minimizer of the average distance problem with non closed set of corners*, Rend. Sem. Mat. Univ. Padova, in press.
[18] X.Y. Lu and D. Slepčev, *Properties of minimizers of average distance problem via discrete approximation of measures*, SIAM J. Math. Anal., **45** (2013), 3114–3131.
[19] U. Ozertem and D. Erdogmus, *Locally defined principal curves and surfaces*, J. Mach. Learn. Res., **12** (2011), 1249–1286.
[20] E. Paolini and E. Stepanov, *Qualitative properties of maximum and average distance minimizers in* $\mathbb{R}^n$, J. Math. Sci. (N.Y.), **122** (2004), 3290–3309.
[21] P. Polak and G. Wolanski, *The lazy traveling salesman problem in* $\mathbb{R}^2$, ESAIM: Control Optim. Calc. Var., **13** (2007), 538–552.
[22] D. Slepčev, *Counterexample to regularity in average-distance problem*, Ann. Inst. H. Poincaré (C), **31** (2014), 169–184.
[23] A.J. Smola, S. Mika, B. Schölkopf and R.C. Williamson, *Regularized principal manifolds*, J. Mach. Learn., **1** (2001), 179–209.
[24] R. Tibshirani, *Principal curves revisited*, Stat. Comput., **2** (1992), 183–190.
[25] C. Villani, "Optimal transport, old and new", Grundlehren der mathematischen Wissenschaften, Springer, 2009.

E-mail address: `xinyang@andrew.cmu.edu`