Busemann functions on the Wasserstein space

Guomin Zhu¹ · Wen-Long Li² · Xiaojun Cui¹

Received: 18 November 2019 / Accepted: 31 December 2020 / Published online: 27 April 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract
We study rays and co-rays in the Wasserstein space $P_p(\mathcal{X})$ ($p > 1$) whose ambient space $\mathcal{X}$ is a complete, separable, non-compact, locally compact length space. We show that rays in the Wasserstein space can be represented as probability measures concentrated on the set of rays in the ambient space. We show the existence of co-rays for any prescribed initial probability measure. We introduce Busemann functions on the Wasserstein space and show that co-rays are negative gradient lines in some sense.

Mathematics Subject Classification 58E10 · 60B10 · 60H30

1 Introduction

The Wasserstein distance plays an important role in optimal transport theory [2, 24, 28] and shows its advantages in numerous topics such as mean-field games [12], machine learning [3] and Hamilton-Jacobi equations [1, 18]. Given an ambient space $\mathcal{X}$ which is a Polish space, namely complete separable metric space, the Wasserstein space $P_p(\mathcal{X})$ consisting of Borel probability measures on $\mathcal{X}$ with finite $p$-moment is also Polish [28, Theorem 6.18]. Besides, if $\mathcal{X}$ is a length (resp. geodesic) space, then $P_p(\mathcal{X})$ is also a length (resp. geodesic) space [21]. Lisini [21] also characterized geodesics in the Wasserstein space as Borel probability measures on $C([0, T]; \mathcal{X})$ concentrated on the set of geodesics in $\mathcal{X}$. In the setting of $\mathcal{X}$...
to be an Hadamard space, Bertrand and Kloeckner [8] investigated the geometry of $P_2(\mathcal{X})$, especially rays and geodesic boundary.

Busemann functions, introduced by Busemann [11], are powerful tools for studying the topology and geometry of some kinds of non-compact spaces. For example, Cheeger and Gromoll [13] used them to prove the celebrated splitting theorem for manifolds of nonnegative Ricci curvature. Bangert [4] investigated the dynamics on a Riemannian 2-torus by means of Busemann functions on the covering space. For more applications of Busemann functions, we refer to [5,6,19,23,25–27]. Furthermore the Busemann function has been introduced into the study of Lorentzian geometry [7,17,20].

The aim of this paper is to extend Busemann functions to the Wasserstein space and thereby to get similar results as in the conventional case. In a metric space $(Y,d)$, a curve $\gamma \in C(\mathbb{R}_+; Y)$ satisfying $d(\gamma_s, \gamma_t) = |s - t| d(\gamma_0, \gamma_1)$ for any $s, t \geq 0$ where $k_\gamma := d(\gamma_0, \gamma_1)$ is called the speed of $\gamma$. Another ray $\tilde{\gamma}$ is said to be a co-ray from $p$ to $\gamma$ if it is the limit of a sequence $\{\zeta^n\}$ as $n \to \infty$, where $\zeta^n$ is a geodesic connecting $p^n$ and $\gamma_n$ with $p^n \to p$ and $t_n \to +\infty$. Usually a Busemann function is determined by a ray. As a preparation we give a characterization of rays in the Wasserstein space, that is, every ray in $P_p(\mathcal{X})$ can be represented by a probability measure concentrated on the set of rays in $\mathcal{X}$. The main ingredient in the study of Busemann functions is co-ray. A common method to show the existence of co-rays is selecting a sequence of geodesics with some good properties which by Ascoli’s theorem admits a subsequence converging to a co-ray. In the case of $P_p(\mathcal{X})$, the essential difficulty lies in the absence of local compactness: any non-compact Wasserstein space can not be locally compact [2, Remark 7.1.9]. Meanwhile this fact implies that a Wasserstein space is not a G-space on which the Busemann function is initially defined, because G-spaces are finitely compact, i.e. every bounded infinity set has at least one accumulation point.

Without local compactness, our approach consists of two steps. First we construct probability measures $\Pi^n$ on $C(\mathbb{R}_+; \mathcal{X})$ representing the geodesics $\zeta^n$. The sequence $\{\Pi^n\}$ admits some accumulation points in the sense of weak convergence if it is confirmed to be tight. Then we show that time sections of the weak convergent subsequences are actually uniformly integrable, which means that the associated geodesics are convergent pointwisely with respect to the Wasserstein distance $W_p$.

With the assumptions of $p > 1$ and $\mathcal{X}$ to be a complete, separable, non-compact, locally compact length space, the main results of this paper are stated as follows.

**Theorem 1.1** Each point in $P_p(\mathcal{X})$ is the origin of at least one unit-speed ray.

$\mathcal{X}$ is called non-branching if any geodesic $\zeta : [a, b] \to \mathcal{X}$ is uniquely determined by its restriction to a nontrivial time-interval.

**Theorem 1.2** Let $(\mu_t)_{t \geq 0}$ be a unit-speed ray in $P_p(\mathcal{X})$.

(i) For any $\nu_0 \in P_p(\mathcal{X})$, there exists at least one co-ray from $\nu_0$ to $(\mu_t)_{t \geq 0}$;

(ii) Moreover if $\mathcal{X}$ is non-branching, let $(\nu_t)_{t \geq 0}$ be one of the co-rays, then for any $\tau > 0$ the subray $(\nu_{t+\tau})_{t \geq 0}$ is the unique co-ray from $\nu_\tau$ to $(\mu_t)_{t \geq 0}$.

For a unit-speed ray $(\mu_t)_{t \geq 0}$ in $P_p(\mathcal{X})$, the Busemann function

$$b_\mu(v) = \lim_{t \to +\infty} [W_p(v, \mu_t) - t]$$

is well defined on $P_p(\mathcal{X})$. 

\(\textcircled{S} \) Springer
Theorem 1.3 Let \((\mu_t)_{t \geq 0}\) and \((\nu_t)_{t \geq 0}\) be two unit-speed rays in \(P_p(\mathcal{X})\). Assume \(\mathcal{X}\) is non-branching, then \((\nu_t)_{t \geq 0}\) is a co-ray from \(v_0\) to \((\mu_t)_{t \geq 0}\) if and only if
\[
b_\mu(\nu_t) - b_\mu(\nu_s) = s - t
\]
holds for any \(s, t \geq 0\).

We introduce the metric viscosity class (Definition 5.7) in order to avoid the complicated discussions about differential structures of Wasserstein spaces. On a non-compact complete Riemannian manifold, it equals to the set of viscosity solutions to the eikonal equation \([14]\) including all Busemann functions \([15]\). Each Busemann function determined by a unit-speed ray in \(P_p(\mathcal{X})\) still belongs to the metric viscosity class \(V(P_p(\mathcal{X}))\).

Theorem 1.4 If \((\mu_t)_{t \geq 0}\) is a unit-speed ray in \(P_p(\mathcal{X})\), then \(b_\mu \in V(P_p(\mathcal{X}))\).

This paper is organized as follows. Section 2 presents some preliminaries. In Sect. 3 we study the structure of rays in the Wasserstein space. More precisely, we show that each ray can be represented as a probability measure concentrated on the set of rays in the ambient space \(\mathcal{X}\). Section 4 presents the existence of co-rays in \(P_p(\mathcal{X})\) and thus Theorem 1.1 and Theorem 1.2 (i) are proved. To prove a sequence of geodesics is compact in the pointwise convergence topology with respect to \(W_p\), the key point is to show the uniformly integrability of the projections at any fixed time \(\tau \geq 0\). In Sect. 5 we define the Busemann function on the Wasserstein space and obtain some analogue fundamental properties to the conventional case. The proofs of Theorem 1.2 (ii) and Theorems 1.3, 1.4 are completed in this section.

2 Preliminaries

2.1 Convergence in the Wasserstein space

Let \((\mathcal{X}, d)\) be a Polish space. For \(p \geq 1\), the Wasserstein space \(P_p(\mathcal{X})\) of order \(p\) is the set of Borel probability measures with finite \(p\)-moments, i.e.
\[
P_p(\mathcal{X}) = \left\{ \mu \in P(\mathcal{X}) \left| \int_{\mathcal{X}} d(x_0, x)^p d\mu(x) < +\infty \right. \right\} \text{ for fixed } x_0 \in \mathcal{X}.
\]
This space does not depend on the choice of \(x_0\). Given \(\mu, \nu \in P_p(\mathcal{X})\), we denote by \(\Pi(\mu, \nu)\) the set of Borel probability measure on \(\mathcal{X} \times \mathcal{X}\) whose marginals are \(\mu\) and \(\nu\) respectively. Elements in \(\Pi(\mu, \nu)\) are called couplings of \((\mu, \nu)\). The Wasserstein distance of order \(p\) between \(\mu\) and \(\nu\) is defined by
\[
W_p(\mu, \nu) = \left( \min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y) \right)^{1/p}.
\] (2.1)
A coupling \(\pi\) is said to be optimal if it achieves the minimum.

The space \(C([0, T]; \mathcal{X})\) of continuous curves in \(\mathcal{X}\) equipped with the metric \(\rho_T\) is a Polish space, where
\[
\rho_T(\alpha, \beta) = \sup_{0 \leq t \leq T} d(\alpha(t), \beta(t)), \text{ for } \alpha, \beta \in C([0, T]; \mathcal{X}).
\]
For \(\alpha, \beta \in C(\mathbb{R}_+; \mathcal{X})\), we define
\[
\rho(\alpha, \beta) = \sum_{N \in \mathbb{N}} 2^{-N} \frac{\rho_N(\alpha, \beta)}{1 + \rho_N(\alpha, \beta)}.
\]
Then \( C(\mathbb{R}_+; \mathcal{X}) \) is also a Polish space \([29]\).

Let \( P(\mathcal{X}) \) denote the set of Borel probability measures on \( \mathcal{X} \). The support of \( \mu \in P(\mathcal{X}) \) defined by

\[
\text{supp } \mu = \{ x \in \mathcal{X} | \mu(B_r(x)) > 0 \text{ for any } r > 0 \} \tag{2.2}
\]

is the smallest closed set on which \( \mu \) is concentrated. We say that a sequence \( \{\mu^n\} \subset P(\mathcal{X}) \) converges weakly to \( \mu \in P(\mathcal{X}) \), denoted by \( \mu^n \Rightarrow \mu \), if

\[
\lim_{n \to \infty} \int_{\mathcal{X}} f(x) d\mu_n(x) = \int_{\mathcal{X}} f(x) d\mu(x) \tag{2.3}
\]

for every bounded continuous function \( f \) on \( \mathcal{X} \).

Let \( f : \mathcal{X} \to \mathcal{Y} \) be a Borel map between Polish spaces, and \( \lambda \) be a Borel measure on \( \mathcal{X} \). The push-forward of \( \lambda \), denoted by \( f_#\lambda \), is defined by \( (f_#\lambda)[A] = \lambda[f^{-1}(A)] \) for any Borel subset \( A \). Furthermore, if \( f \) is continuous, then \( \mu^n \Rightarrow \mu \) implies \( f_#\mu^n \Rightarrow f_#\mu \) \([9, \text{Theorem 2.7}]\).

**Proposition 2.1** \([2, \text{Proposition 5.1.8}]\) If \( \{\mu^n\} \subset P(\mathcal{X}) \) converges weakly to \( \mu \in P(\mathcal{X}) \), then for any \( x \in \text{supp } \mu \) there exists a sequence \( x^n \in \text{supp } \mu^n \) such that \( \lim_{n \to \infty} x^n = x \).

We call \( S \subset P(\mathcal{X}) \) relatively compact (for the weak topology) if every sequence in \( S \) contains a weakly convergent subsequence. And \( S \) is said to be tight if for any \( \varepsilon > 0 \), there exists a compact set \( K_\varepsilon \subset \mathcal{X} \) such that \( \mu(K_\varepsilon) > 1 - \varepsilon \) for every \( \mu \in S \). Every single probability measure on a Polish space is itself tight (see e.g. \([9, \text{Theorem 1.3}]\)). The following theorem indicates why the tightness makes sense.

**Theorem 2.2** (Prokhorov) Let \( \mathcal{X} \) be a Polish space. \( S \subset P(\mathcal{X}) \) is tight if and only if it is relatively compact in \( P(\mathcal{X}) \).

There is a tightness criterion for a sequence of probability measures on the space \( C(\mathbb{R}_+; \mathcal{X}) \).

**Theorem 2.3** \([29, \text{Theorem 4}]\) Let \( \{P^n\} \) be a sequence of probability measures on \( C(\mathbb{R}_+; \mathcal{X}) \). Then \( \{P^n\} \) is tight if and only if the following two conditions hold:

(i) For each \( t \geq 0 \) and \( \eta > 0 \), there exists a compact set \( K_t \) in \( \mathcal{X} \) such that

\[
P^n \{ x \in C(\mathbb{R}_+; \mathcal{X}) | x(t) \in K_t \} > 1 - \eta, \text{ for any } n \geq 1.
\]

(ii) For each \( j \geq 1 \) and \( \varepsilon, \eta > 0 \), there exists a \( \delta \in (0, 1) \), and an \( n_0 \in \mathbb{N} \) such that

\[
P^n \{ x \in C(\mathbb{R}_+; \mathcal{X}) | w^n(x)(\delta) \leq \varepsilon \} \leq \eta, \text{ for any } n \geq n_0,
\]

where

\[
w^n(x)(\delta) = \sup_{0 \leq s, t \leq j, |s-t| < \delta} d(x(s), x(t)) \text{ is the modulus of continuity of } x|_{[0, j]}.\]

The next theorem provides a more effective description of the convergence with respect to the Wasserstein distance.

**Theorem 2.4** \([2, \text{Theorem 7.1.5}]\) Let \( (\mathcal{X}, d) \) be a Polish space and \( p \geq 1 \). Let \( \{\mu^n\}_{n \in \mathbb{N}} \subset P_p(\mathcal{X}) \) and \( \mu \in P_p(\mathcal{X}) \), then the following statements are equivalent:

(i) \( \mu^n \Rightarrow \mu \) and for some \( x_0 \in \mathcal{X} \),

\[
\lim_{R \to \infty} \limsup_{n \to \infty} \int_{d(x_0, x) \geq R} d(x_0, x)^p d\mu^n(x) = 0;
\]

(ii) \( W_p(\mu^n, \mu) \to 0 \) as \( n \to \infty \).
2.2 Geodesics in the Wasserstein space

Let \((\mathcal{Y}, d)\) be a metric space. The length of a continuous curve \(\zeta : [a, b] \to \mathcal{Y}\) is defined by

\[
L(\zeta) = \sup_{N \in \mathbb{N}} \sup_{a=t_0 < t_1 \ldots < t_N = b} \sum_{i=0}^{N-1} d(\zeta_{t_i}, \zeta_{t_{i+1}}).
\]  

(\mathcal{Y}, d) is said to be a length space if for any \(x, y \in \mathcal{Y}\),

\[
d(x, y) = \inf_{\zeta \in C([0, 1], \mathcal{Y})} \{L(\zeta) : \zeta_0 = x, \zeta_1 = y\}. \tag{2.5}
\]

\(\mathcal{Y}\) is a geodesic space if the infimum in equation (2.5) is attainable for any \(x, y \in \mathcal{Y}\). \(\zeta\) is called a constant-speed minimizing geodesic segment if

\[
d(\zeta_s, \zeta_t) = \frac{|t-s|}{b-a} d(\zeta_a, \zeta_b) \text{ for any } s, t \in [a, b]. \tag{2.6}
\]

For convenience, throughout this paper we use the single word “geodesic” instead. Let \(\mathcal{G}_L(\mathcal{Y})\) denote the set of geodesics defined on the interval \([0, L]\), i.e.

\[
\mathcal{G}_L(\mathcal{Y}) = \{\zeta : [0, L] \to \mathcal{Y} | \zeta \text{ is geodesic}\}. \tag{2.7}
\]

The next statement is a straight corollary to [28, Corollary 7.22] via simple reparameterization. This conclusion is twofold: The Wasserstein space over a complete separable locally compact length space is a geodesic space; Geodesics in such a Wasserstein space can be considered as probability measures concentrated on the set of geodesics in the ambient space. Write \(e_t : \zeta \mapsto \zeta_t\) be the canonical projection at time \(t\).

Proposition 2.5 Let \(p > 1\) and let \((\mathcal{X}, d)\) be a complete separable, locally compact length space. Given \(\mu, \nu \in P_p(\mathcal{X})\), let \(L = W_p(\mu, \nu)\). Then for any continuous curve \((\mu_t)_{0 \leq t \leq L}\) in \(P(\mathcal{X})\) with \(\mu_0 = \mu, \mu_L = \nu\), the following properties are equivalent:

(i) There exists a \(\Pi \in P(C([0, L]; \mathcal{X}))\) concentrated on \(\mathcal{G}_L(\mathcal{X})\) such that

\[
\mu_t = (e_t)_#\Pi \text{ for any } t \in [0, L],
\]

and \((e_0, e_L)_#\Pi\) is an optimal coupling of \((\mu, \nu)\);

(ii) \((\mu_t)_{0 \leq t \leq L}\) is a unit-speed geodesic in the space \(P_p(\mathcal{X})\).

Moreover, for any given \(\mu, \nu \in P_p(\mathcal{X})\), there exists at least one such curve. We denote by \(T(\mu, \nu)\) the set of unit-speed geodesics from \(\mu\) to \(\nu\).

Remark 2.6 Let \(\Pi\) and \((\mu_t)_{0 \leq t \leq L}\) be as mentioned above. Then for \(0 \leq s < t \leq L\), \(\pi_{s,t} := (e_s, e_t)_#\Pi\) is an optimal coupling of \((\mu_s, \mu_t)\). Actually, by the definition of geodesics,

\[
W_p^p(\mu_s, \mu_t) = \left(\frac{t-s}{L}\right)^p W_p^p(\mu_0, \mu_L)
\]

\[
= \int \left[\frac{t-s}{L} d(\gamma_0, \gamma_L)\right]^p d\Pi(\gamma)
\]

\[
= \int d^p(\gamma_s, \gamma_t)d\Pi(\gamma)
\]

\[
= \int d^p(x, y)d\pi_{s,t}(x, y),
\]

where the third equality follows because \(\Pi\) is concentrated on \(\mathcal{G}_L(\mathcal{X})\).
Based on the understanding of geodesics in $P_p(\mathcal{X})$, it is natural to discuss similar properties of rays in the Wasserstein space.

### 3 Characterization of rays in the Wasserstein space

In the rest of this paper, we always assume that $(\mathcal{X}, d)$ is a complete, separable non-compact, locally compact length space and the order $p > 1$. Recall that a ray in $\mathcal{X}$ is a curve $\gamma \in C(\mathbb{R}_+; \mathcal{X})$ satisfying

$$
d(\gamma_s, \gamma_t) = |t - s|d(\gamma_0, \gamma_1) \text{ for any } s, t \geq 0,
$$

where $k_\gamma = d(\gamma_0, \gamma_1)$ is called the speed of $\gamma$.

**Lemma 3.1** Let $\Gamma$ be the set of rays in $\mathcal{X}$, then $\Gamma$ is closed in $(C(\mathbb{R}_+; \mathcal{X}), \rho)$. As a consequence, $(\Gamma, \rho)$ is a Polish space.

**Proof** Let $\{\gamma^n\}$ be a Cauchy sequence of $\Gamma$, then there exists a $\gamma : \mathbb{R}_+ \to \mathcal{X}$ such that $\rho(\gamma^n, \gamma) \to 0$. Denote $k = \lim_{n \to \infty} d(\gamma^n_0, \gamma_1)$. The limit exists since $\lim_{n \to \infty} d(\gamma^n_0, \gamma_1) = 0$ for any $t \geq 0$. For any $t_1, t_2 \geq 0$,

$$
|d(\gamma^n_{t_1}, \gamma^n_{t_2}) - k|t_1 - t_2|| = |t_1 - t_2||d(\gamma^n_0, \gamma^n_1) - k| \to 0.
$$

By the triangle inequality,

$$
|d(\gamma_{t_1}, \gamma_{t_2}) - d(\gamma^n_{t_1}, \gamma^n_{t_2})| \leq d(\gamma^n_{t_1}, \gamma_{t_1}) + d(\gamma^n_{t_2}, \gamma_{t_2}) \to 0.
$$

Combining (3.2) and (3.3), we have $|d(\gamma_{t_1}, \gamma_{t_2}) - k|t_1 - t_2|| = 0$, then the conclusion follows. \hfill \Box

As shown in Proposition 2.5, geodesics in the Wasserstein space are related to different curve spaces unless their lengths are equal. So we introduce the mapping $E_T : C([0, T]; \mathcal{X}) \to C(\mathbb{R}_+; \mathcal{X})$ by

$$(E_T(\xi))_s = \xi_{\min[s, T]}, s \geq 0$$

in order to extend their sample paths onto the common space $C(\mathbb{R}_+; \mathcal{X})$.

**Definition 3.2** Let $(\mu_t)_{0 \leq t \leq L}$ be a geodesic in $P_p(\mathcal{X})$. Let $\Pi \in P(C(\mathbb{R}_+; \mathcal{X}))$ concentrated on $\mathcal{G}_L(\mathcal{X})$ such that $(e_0, e_L)\#\Pi$ is an optimal coupling of $(\mu_0, \mu_L)$ and $\mu_t = (e_t)\#\Pi$ for any $t \in [0, L]$. We call $(E_L)\#\pi$ the lifting of the geodesic.

It can be seen from Proposition 2.5 that each geodesic in $P_p(\mathcal{X})$ admits a lifting.

**Lemma 3.3** For $\mu, \nu \in P_p(\mathcal{X})$, let $\pi$ be an optimal coupling. Then for $R > 0$,

$$
\pi\{x, y \mid d(x, y) > R\} \leq \left[ \frac{W_p(\mu, \nu)}{R} \right]^p.
$$

**Proof** Given $R > 0$, by the definition of $W_p$,

$$
\int_{d(x, y) > R} R^p d\pi(x, y) \leq \int_{d(x, y) > R} d(x, y)^p d\pi(x, y)
\leq \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y)
= W_p^p(\mu, \nu).
$$
So we obtain the inequality (3.4).

**Theorem 3.4** Let \( \{v^n\} \) and \( \{\mu^n\} \) be two sequences in \( P_p(\mathcal{X}) \) such that \( \{v^n\} \) is tight and \( L_n := W_p(\mu^n, v^n) \to +\infty \). For each \( n \), let \( \Pi^n \) be a lifting of an element in \( T(v^n, \mu^n) \), then \( \{\Pi^n\} \) is tight.

**Proof** We will verify the criteria of Theorem 2.3 to obtain the tightness of \( \{\Pi^n\} \). Let \( v^n_t = (e_t)_{#} \Pi^n \) for \( t \leq L_n \).

(i) Fix an arbitrary \( t \geq 0 \). Given any \( \eta > 0 \), there exists an \( N \) such that \( L_n \geq t \) for \( n > N \). Recall that each single probability measure on a Polish space is tight. For \( n \), there exists a compact set \( K^n_t \) such that

\[
\Pi^n \{ \gamma^n \in C(\mathbb{R}_+; \mathcal{X}) \mid \gamma^n_t \in K^n_t \} = v^n_t[K^n_t] > 1 - \eta/2. \tag{3.5}
\]

While for \( n > N \), by the tightness of \( \{v^n_0\} \) there exists a constant \( D_0 \) such that

\[
\Pi^n \{ \gamma^n \mid d(\gamma^n_0, x_0) > D_0 \} = v^n_0\{x \mid d(x, x_0) > D_0\} < \eta/4. \tag{3.6}
\]

By Proposition 2.5, for each \( n \in \mathbb{N} \), \( \pi^n := (e_0, e_{L_n})_{#} \Pi^n \) is an optimal coupling of \( (v^n, \mu^n) \). By Lemma 3.3,

\[
\Pi^n \{ \gamma^n \mid d(\gamma^n_0, \gamma^n_t) > R \} = \pi^n\{(x, y) \mid d(x, y) > R\} \leq \left( \frac{t}{R} \right)^p. \tag{3.7}
\]

For \( D_t \geq t (4/\eta)^{1/p} + D_0 \), from (3.6) and (3.7) we have

\[
\Pi^n \{ \gamma^n \mid d(\gamma^n_0, x_0) > D_t \}
\leq \Pi^n \{ \gamma^n \mid d(\gamma^n_0, x_0) > D_0 \} + \Pi^n \{ \gamma^n \mid d(\gamma^n_0, x_0) \leq D_0, d(\gamma^n_0, \gamma^n_t) > D_t - D_0 \}
\leq \Pi^n \{ \gamma^n \mid d(\gamma^n_0, x_0) > D_0 \} + \Pi^n \{ \gamma^n \mid d(\gamma^n_0, \gamma^n_t) > D_t - D_0 \}
\leq \eta/4 + \left( \frac{t}{D_t - D_0} \right)^p
\leq \eta/2,
\]

which means there is a compact set \( K_t = \{ x \mid d(x, x_0) \leq D_t \} \cup \bigcup_{i=1}^N K^n_t^i \) such that

\[
\Pi^n \{ \gamma^n \mid \gamma^n_t \in K_t \} > 1 - \eta. \tag{3.8}
\]

(ii) For any fixed \( j \geq 1 \), by Definition 3.2, for \( n > j \) the curves \( (v^n_t)_{0 \leq t \leq j} \) are geodesics in \( P_p(\mathcal{X}) \). For any fixed \( \epsilon, \eta > 0 \), let \( \delta < \epsilon \eta^{1/p} \), then

\[
w_{\gamma^n_t}(\delta) = \sup_{0 \leq s, t \leq j \atop |s - t| < \delta} d(\gamma^n_t, \gamma^n_s) = \sup_{0 \leq s, t \leq j \atop |s - t| < \delta} |s - t| d(\gamma^n_0, \gamma^n_t) \leq \delta d(\gamma^n_0, \gamma^n_t). \tag{3.9}
\]

As \( (v^n_t)_{0 \leq t \leq L_n} \) is a unit-speed geodesic for each \( n \), we have \( W_p(v^n_0, v^n_t) = 1 \). By Lemma 3.3,

\[
\Pi^n \{ \gamma^n \mid w_{\gamma^n_t}(\delta) > \epsilon \} \leq \Pi^n \{ \gamma^n \mid d(\gamma^n_0, \gamma^n_t) > \epsilon/\delta \} \leq (\delta/\epsilon)^p < \eta. \tag{3.10}
\]

Hence the tightness of \( \{\Pi^n\} \) follows from Theorem 2.3. \( \square \)

Now we are able to characterize rays in \( P_p(\mathcal{X}) \). A similar representation can be found in [8, Proposition 3.2] for the 2-Wasserstein space over an Hadamard space.

**Corollary 3.5** For a continuous path \((\mu_t)_{t \geq 0}\) in \( P_p(\mathcal{X}) \), the following statements are equivalent:
(i) \((\mu_t)_{t \geq 0}\) is a ray in \(P_p(\mathcal{X})\);
(ii) For any \(t_1, t_2\) with \(0 \leq t_1 < t_2\), \((\mu_t)_{t_1 \leq t \leq t_2}\) is a geodesic;
(iii) There exists a \(\Pi \in P(C(\mathbb{R}_+; \mathcal{X}))\) concentrated on \(\Gamma\) such that for \(0 \leq t_1 < t_2\),
\[\mu_{t_1} = (e_{t_1})_#\Pi \quad \text{and} \quad (e_{t_2})_#\Pi \quad \text{is an optimal coupling of} \quad (\mu_{t_1}, \mu_{t_2}).\]

**Proof** (i) and (ii) are equivalent by the definition of ray and (iii) \(\Rightarrow\) (ii) is obvious.
(ii) \(\Rightarrow\) (iii). For \(j \in \mathbb{N}\), let \(\Pi^j\) be a lifting of \((\mu_t)_{0 \leq t \leq j}\) and
\[\Gamma^j = \{\xi \in C([0, j]; \mathcal{X}) | \xi \text{ is a geodesic}\}.

By Proposition 2.5, \(\Pi^j\) is concentrated on \(E_j(\Gamma^j)\) such that \((e_0, e_j)_#\Pi^j\) is an optimal coupling. By Theorem 3.4 and Theorem 2.2, it admits a subsequence \(\{\Pi^{j'}\}\) which converges weakly to some measure \(\Pi\) on \(C(\mathbb{R}_+; \mathcal{X})\). For any \(\gamma \in \text{supp} \Pi\), by Proposition 2.1, there exists \(\xi^{j'} \in E_{j'}(\Gamma^{j'})\) with
\[\lim_{j' \to \infty} \rho(\xi^{j'}, \gamma) = 0. \tag{3.11}\]

For any \(0 \leq s_1 < s_2\), choose \(T \geq \max\{1, s_2\}\), then (3.11) implies
\[d(\xi^{j'}_{s_1}, \gamma) \to 0 \quad \text{for any} \quad s \in [0, T]. \tag{3.12}\]
Notice that \(d(\xi^{j'}_{s_1}, \xi^{j'}_{s_2}) = (s_2 - s_1)\rho(\xi_0^{j'}, \xi_1^{j'})\) when \(j' > T\). (3.12) yields
\[d(\gamma_{s_1}, \gamma_{s_2}) = (s_2 - s_1)\rho(\gamma_0, \gamma_1). \tag{3.13}\]
It follows that \(\gamma \in \Gamma\). In other words, \(\Pi\) is concentrated on \(\Gamma\).

For \(0 \leq t_1 < t_2\), denote \(\pi^{j'}_{t_1, t_2} = (e_{t_1}, e_{t_2})_#\Pi^{j'}\), \(\pi_{t_1, t_2} = (e_{t_1}, e_{t_2})_#\Pi\), then \(\pi^{j'}_{t_1, t_2} \Rightarrow \pi_{t_1, t_2}\).

By the lower semicontinuity of the map \(\pi \mapsto \int d^p d\pi\) [28, Lemma 4.3],
\[W_p^p(\mu_{t_1}, \mu_{t_2}) \leq \int_{\mathcal{X} \times \mathcal{X}} d^p(x, y) d\pi_{t_1, t_2} \leq \liminf_{j' \to \infty} \int_{\mathcal{X} \times \mathcal{X}} d^p(x, y) d\pi^{j'}_{t_1, t_2}. \tag{3.14}\]
From Remark 2.6 we know that each \(\pi^{j'}_{t_1, t_2}\) is an optimal coupling of \((\mu_{t_1}, \mu_{t_2})\), i.e.
\[\int_{\mathcal{X} \times \mathcal{X}} d^p(x, y) d\pi^{j'}_{t_1, t_2} = W_p^p(\mu_{t_1}, \mu_{t_2}). \tag{3.15}\]
Thus \(\pi_{t_1, t_2}\) is also an optimal coupling of \((\mu_{t_1}, \mu_{t_2})\).

\[\square\]

**4 Existence of co-rays in the Wasserstein space**

In the conventional case, co-rays play a central role in the study of Busemann functions [11]. This notion also makes sense in the present case. The existence of co-rays in the Wasserstein space will be proved in this section.

**Definition 4.1** (Co-ray) Let \((\mu_t)_{t \geq 0}\) be a unit-speed ray in \(P_p(\mathcal{X})\). We say another ray \((v_t)_{t \geq 0}\) is a co-ray from \(\nu_0\) to \((\mu_t)_{t \geq 0}\), if there exist:
- \(\{t_n\} \subset \mathbb{R}_+\) tends to infinity,
- \((v^n_0) \subset P_p(\mathcal{X})\) with \(W_p(\nu^n_0, \nu_0) \to 0\),
- for \(n \in \mathbb{N}\), \((v^n_0)_{0 \leq t \leq t_n} \in T(v^n_0, \mu_{t_n})\) with \(L_n = W_p(v^n_0, \mu_{t_n})\), such that \(\lim_{n \to \infty} W_p(v^n_t, v_t) = 0\) for every \(t \geq 0\).
The gluing lemma (see e.g. [2, Lemma 5.3.2]) is often used in optimal transport to connect two couplings. However this instrument can not meet our demand while probability measures on curve spaces get involved, so we cite a theorem here in order to obtain another version of gluing lemma.

**Theorem 4.2** ([16, Theorem A.1]) Let J be an arbitrary index set. For each $j \in J$, let $S_j, T_j$ be Polish spaces, $S = \prod_{j \in J} S_j, T = \prod_{j \in J} T_j$. Also, let $p_j : S \to S_j, q_j : T \to T_j$ be canonical projections, $\phi_j : S_j \to T_j$ be a measurable map. If $\mu_j$ is a probability measure on $S_j$ and $\lambda$ is a probability measure on $T$ such that $(\phi_j)_# \mu_j = (q_j)_# \lambda, \forall j \in J$, then there exists a probability measure $\sigma$ on $S$ such that:

(i) $(p_j)_# \sigma = \mu_j, \forall j \in J$;
(ii) $((\phi_j \circ p_j))_# \sigma = \lambda$.

The modified gluing lemma allows us to glue two probability measures on curve space and product space respectively together.

**Lemma 4.3** Let $\mathcal{Y}$ be a Polish space, $\alpha \in P(C([0, T]; \mathcal{X})), \beta \in P(\mathcal{X} \times \mathcal{Y})$ and $\pi^i$ be the natural projection onto the $i$-th coordinate, $i = 1, 2$. If $(e_T)_# \alpha = (\pi^1)_# \beta$, then there exists a $\delta \in P(C([0, T]; \mathcal{X}) \times \mathcal{Y})$ such that

$$(\pi^1)_# \delta = \alpha, (e_T \pi^1, \pi^2)_# \delta = \beta.$$

**Proof** Let $S_1 = C([0, T]; \mathcal{X}), T_1 = \mathcal{X}, \phi_1 = e_T; S_2 = \mathcal{Y} = T_2, \phi_2 = Id_\mathcal{Y}; \mu_1 = \alpha, \mu_2 = (\pi^2)_# \beta = (q_2)_# \beta, \lambda = \beta$. Then

$$(\phi_1)_# \mu_1 = (e_T)_# \alpha = (\pi^1)_# \beta = (q_1)_# \beta;
(\phi_1)_# \mu_2 = (Id_\mathcal{Y})_# ((q_2)_# \beta) = (q_2)_# \beta.$$

Applying Theorem 4.2 for $J = \{1, 2\}$, there exists a probability measure $\delta$ on $S_1 \times S_2 = C([0, T]; \mathcal{X}) \times \mathcal{Y}$ such that

$$(\pi^1)_# \delta = (p_1)_# \delta = \mu_1 = \alpha,
(e_T \pi^1, \pi^2)_# \delta = (\phi_1 \circ p_1, \phi_1 \circ p_2)_# \delta = \lambda = \beta.$$

\hfill $\square$

**Theorem 4.4** Let $(\mu_t)_{t \geq 0}$ be a unit-speed ray in $P_\rho(\mathcal{X})$, $\rho > 1$. Given an arbitrary $v_0 \in P_\rho(\mathcal{X})$, for any:

- $\{t_n\}$ increasing to infinity,
- $\{v^p_0\} \subset P_\rho(\mathcal{X})$ with $W_\rho(v^p_0, v_0) \to 0$,
- $(v^p_0)_{0 \leq t \leq L_n} \in T(v^0_0, \mu_{t_n}), n \in \mathbb{N}$ where $L_n = W_\rho(v^0_n, \mu_{t_n})$.

there exists a subsequence of $\{(v^p_t)_{0 \leq t \leq L_n}\}$ which converges to a co-ray from $v_0$ to $(\mu_t)_{t \geq 0}$.

**Proof** Let $\Lambda^n$ be a lifting of $(v^p_t)_{0 \leq t \leq L_n}$. By Theorem 3.4, there exists a subsequence of $\{\Lambda^n\}$, still denoted by the same notation, which converges weakly to a probability measure $\Lambda$ on $C(\mathbb{R}_+; \mathcal{X})$. Then for arbitrary fixed $\tau \geq 0$,

$$v^n_\tau = (e_\tau)_# \Lambda^n \Rightarrow (e_\tau)_# \Lambda = v_\tau.$$

To show that $W_\rho(v^n_\tau, v_\tau) \to 0$ as $n \to \infty$, by Theorem 2.4, it remains to prove that

$$\lim_{R \to \infty} \limsup_{n \to \infty} \int_{d(x_0, z) \geq R} d(x_0, z)^p d v^n_\tau (z) = 0. \quad (4.1)$$
By Proposition 2.5, for each \( n \), there exists \( \alpha^n \in P(C([0, L_n]; \mathcal{X})) \) satisfying
\[
(e_t)_#\alpha^n = \nu^n_t \text{ for } t \in [0, L_n].
\]

By Corollary 3.5, there exists a probability measure \( \beta \) concentrated on \( \Gamma \) such that \( \mu_t = (e_t)_#\beta \) and \( (e_0, e_t)_#\beta \) is an optimal coupling of \( \mu_0, \mu_t \) for any \( t \geq 0 \). For each \( n \), let \( \beta^n = (e_0, e_{L_n})_#\beta \), then it is an optimal coupling of \( \mu_0, \mu_{t_n} \). By Lemma 4.3, we can construct a sequence of probability measures \( \{\Pi^n\} \subset P(C([0, L_n]; \mathcal{X}) \times \mathcal{X}) \) satisfying
\[
(\pi_1)_#\Pi^n = \alpha^n, \ (e_{L_n}\pi_1, \pi_2)_#\Pi^n = \beta^n. \tag{4.2}
\]

By the triangle inequality,
\[
|L_n - t_n| = |W_p(\nu^n_0, \mu_{t_n}) - W_p(\mu_0, \mu_{t_n})| \\
\leq W_p(\nu^n_0, \mu_0) \\
\leq W_p(\nu^n_0, \nu_0) + W_p(\nu_0, \mu_0).
\]

Then
\[
\lim_{n \to \infty} \frac{t_n}{L_n} = 1. \tag{4.3}
\]

There exists an \( N \) such that \( L_n > \tau \) for all \( n > N \). In this case, for \( \Pi^n \)-a.e. \( (\gamma^n, y) \in C([0, L_n]; \mathcal{X}) \times \mathcal{X} \), by the triangle inequality,
\[
d(x_0, \gamma^n_\tau) \leq d(x_0, \gamma^n_0) + d(\gamma^n_0, \gamma^n_\tau) \\
= d(x_0, \gamma^n_0) + \tau/L_n d(\gamma^n_0, \gamma^n_{L_n}) \\
\leq d(x_0, \gamma^n_0) + \tau/L_n [d(x_0, \gamma^n_0) + d(x_0, y) + d(y, \gamma^n_{L_n})] \\
= \left(1 + \frac{\tau}{L_n}\right) d(x_0, \gamma^n_0) + \frac{\tau}{L_n} d(x_0, y) + \frac{\tau}{L_n} d(y, \gamma^n_{L_n}).
\]

Here the first equality follows from that each \( \gamma^n \) is a geodesic.
Applying Jensen’s inequality we have
\[
\int_{d(x_0, z) \geq R} d^p(x_0, z) d\nu^p(z)
\]
\[
= \int_{d(x_0, y^n) \geq R} d^p(x_0, y^n) d\Pi^n(y^n, y)
\]
\[
\leq 3^p \left( \frac{\tau}{L_n} + 1 \right)^p \int_{d(x_0, y^n) \geq R} d^p(x_0, y^n) d\Pi^n(y^n, y)
\]
\[
+ \left( \frac{3\tau}{L_n} \right)^p \int_{d(x_0, y^n) \geq R} d^p(x_0, y) d\Pi^n(y^n, y)
\]
\[
+ \left( \frac{3\tau}{L_n} \right)^p \int_{d(x_0, y^n) \geq R} d^p(y, y^n) d\Pi^n(y^n, y)
\]
\[
\leq 3^p \left( \frac{\tau}{L_n} + 1 \right)^p \int_{d(x_0, y^n) \geq R} d^p(x_0, y^n) d\Pi^n(y^n, y)
\]
\[
+ \left( \frac{3\tau}{L_n} \right)^p W_p^p(\delta_{x_0}, \mu_0) + \left( \frac{3\tau}{L_n} \right)^p \int_{d(x_0, y^n) \geq R} d^p(y, y^n) d\Pi^n(y^n, y)
\]
\[
= 3^p \left( 1 + \frac{\tau}{L_n} \right)^p I_1 + \left( \frac{3\tau}{L_n} \right)^p W_p^p(\delta_{x_0}, \mu_0) + \left( \frac{3\tau}{L_n} \right)^p I_2.
\]

We will consider the three terms above respectively.

(i) Let $M = \sup_n \{ \frac{\tau}{L_n} \}$. For arbitrary $\varepsilon > 0$, since $W_p^p(v^n_0, \nu_0) \to 0$, by Theorem 2.4 there exists an $R_1$ such that
\[
\limsup_{n \to \infty} \int_{d(x_0, z) \geq R_1} d^p(x_0, z) d\nu^p_0(z) < \frac{\varepsilon}{4 \cdot 3^p(1 + M)^p}.
\]

We can further find an $R_2 > R_1$ such that when $R > R_2$,
\[
\left( \frac{R_1 \tau}{R - R_1} \right)^p < \frac{\varepsilon}{4 \cdot 3^p(1 + M)^p}.
\]

In this case we estimate $I_1$ as following:

\[
I_1 = \int_{d(x_0, y^n) \geq R} d^p(x_0, y^n) d\Pi^n(y^n, y)
\]
\[
= \int_{d(x_0, y^n) \geq R_1, d(x_0, y^n) \geq R} d^p(x_0, y^n) d\Pi^n(y^n, y)
\]
\[
+ \int_{d(x_0, y^n) < R_1, d(x_0, y^n) \geq R} d^p(x_0, y^n) d\Pi^n(y^n, y)
\]
\[
\leq \int_{d(x_0, y^n) \geq R_1} d^p(x_0, y^n) d\Pi^n(y^n, y) + R_1^p \cdot \alpha^n \{ y^n | d(y^n, y^n) > R - R_1 \}
\]
\[
\leq \int_{d(x_0, z) \geq R_1} d^p(x_0, z) d\nu^p_0(z) + \left( \frac{R_1 \tau}{R - R_1} \right)^p.
\]

The last inequality is obtained from Lemma 3.3. Consequently,
\[
\limsup_{n \to \infty} 3^p \left( 1 + \frac{\tau}{L_n} \right)^p I_1 \leq \limsup_{n \to \infty} 3^p \left( 1 + \frac{\tau}{L_n} \right)^p \frac{\varepsilon}{2 \cdot 3^p(1 + M)^p} < \frac{\varepsilon}{2}.
\]
(ii) Since $W_p(\mu_0, \mu_1) = 1$, there exists an $R_3$ such that
\[
\int_{d(\eta_0, \eta_1) \geq R_3} d^p(\eta_0, \eta_1) d\beta(\eta) < \frac{\varepsilon}{6 \cdot (3\pi)^p}. \tag{4.7}
\]
From the construction of $\Pi^n$, we have
\[
\int_{d(y, y''_{\lambda^n}) \geq t_n R_3} d^p(y, y''_{\lambda^n}) d\Pi^n(y^n, y) = \int_{d(\eta_0, \eta_n) \geq t_n R_3} d^p(\eta_0, \eta_n) d\beta^n(\eta_0, \eta_n) = t_n^p \int_{d(\eta_0, \eta_1) \geq R_3} d^p(\eta_0, \eta_1) d\beta(\eta).
\]
Due to the tightness of $\{\nu_0^n\}$, there is an $R_4 > R_3$ with
\[
v_0^n[z | d(x_0, z) \geq R_4] < \frac{\varepsilon}{6 \cdot (3\pi R_3)^p} \quad \text{for all } n > N. \tag{4.8}
\]
Applying the inequality (3.4) again, we can obtain another $R_5 > R_4$ such that when $R > R_5$,
\[
\alpha^n[y^n | d(y_0^n, y''_n) > R - R_4] < \frac{\varepsilon}{6 \cdot (3\pi R_3)^p} \quad \text{for all } n > N. \tag{4.9}
\]
In this case,
\[
I_2 = \int_{d(x_0, y''_{\lambda^n}) \geq R} d^p(y, y''_{\lambda^n}) d\Pi^n(y^n, y) = \int_{d(y, y''_{\lambda^n}) \geq t_n R_3} d^p(y, y''_{\lambda^n}) d\Pi^n(y^n, y)
\]
\[
+ \int_{d(x_0, y''_{\lambda^n}) \geq R, d(y, y''_{\lambda^n}) < t_n R_3} d^p(y, y''_{\lambda^n}) d\Pi^n(y^n, y)
\]
\[
+ \int_{d(x_0, y''_{\lambda^n}) \geq R, d(y, y''_{\lambda^n}) < t_n R_3} d^p(y, y''_{\lambda^n}) d\Pi^n(y^n, y)
\]
\[
\leq \int_{d(y, y''_{\lambda^n}) \geq t_n R_3} d^p(y, y''_{\lambda^n}) d\Pi^n(y^n, y)
\]
\[
+ (t_n R_3)^p \Pi^n\{(y^n, y) | d(x_0, y_0^n) \geq R_4\}
\]
\[
+ (t_n R_3)^p \Pi^n\{(y^n, y) | d(y_0^n, y''_n) > R - R_4\}
\]
\[
= t_n^p \int_{d(\eta_0, \eta_1) \geq R_3} d^p(\eta_0, \eta_1) d\beta(\eta)
\]
\[
+ (t_n R_3)^p \nu_0^n[z | d(x_0, z) \geq R_4]
\]
\[
+ (t_n R_3)^p \alpha^n[y^n | d(y_0^n, y''_n) > R - R_4].
\]
Combining this inequality with (4.7), (4.8), (4.9) and (4.3), we have
\[
\limsup_{n \to \infty} \left( \frac{3\pi}{L_n} \right)^p I_2 \leq \frac{\varepsilon}{2}. \tag{4.10}
\]
(iii) Notice that $\lim_{n \to \infty} \left( \frac{3\pi}{L_n} \right)^p W_p^p(\delta_{x_0}, \mu_0) = 0$. We obtain
\[
\limsup_{n \to \infty} \int_{d(x_0, z) \geq R} d^p(x_0, z) d\nu_1^n(z) < \varepsilon \quad \text{when } R > \max\{R_2, R_5\}. \tag{4.11}
\]
We conclude from (4.1) that \( \lim_{n \to \infty} W_p(v^n_t, v) = 0 \) for arbitrary \( t \geq 0 \), thus \( v_t \in P_p(\mathcal{X}) \).

Moreover, for \( t, s \geq 0 \),

\[
W_p(v_t, v_s) = \lim_{n \to \infty} W_p(v^n_t, v^n_s) = |t - s|,
\]

which means \((v_t)_{t \geq 0}\) is also a unit-speed ray in \( P_p(\mathcal{X}) \). Therefore it is a co-ray from \( v_0 \) to \((\mu_t)_{t \geq 0}\).

**Proof of Theorem 1.1** A complete and locally compact length space, by the Hopf-Rinow Theorem (see [10, Theorem 2.5.28]), is boundedly compact, i.e. every closed metric ball is compact. Then each point in \( \mathcal{X} \) is the origin of some unit-speed rays, which is due to [22, Proposition 10.1.1]. It is easily seen that the mapping \( x \mapsto \delta_x \) is an isometric embedding from \( \mathcal{X} \) to \( P_p(\mathcal{X}) \). Let \( \gamma \) be a unit-speed ray in \( \mathcal{X} \) and \( \mu_t = \delta_{\gamma_t} \), then \((\mu_t)_{t \geq 0}\) is a unit-speed ray in \( P_p(\mathcal{X}) \) accordingly. Notice that for each \( v_0 \in P_p(\mathcal{X}) \), co-rays from \( v_0 \) to \((\mu_t)_{t \geq 0}\) are some of the unit-speed rays with the origin \( v_0 \), thus Theorem 1.1 holds.

Theorem 1.2 (i) follows directly from Theorem 4.4. It should be point out that co-ray may not be unique in \( P_p(\mathcal{X}) \) even in the non-branching cases. For some specific \( \mathcal{X} \), there exist more than one co-rays from given point \( p \) to given ray \( \gamma \) (see [11, 22.18] and [14, Example 5.2] for the examples). Since \( x \mapsto \delta_x \) is an isometric embedding, each co-ray \( \gamma' \) to \( \gamma \) determines a co-ray \((\delta_{\gamma'_t})_{t \geq 0}\) to ray \((\delta_{\gamma_t})_{t \geq 0}\), which provides an example on \( P_p(\mathcal{X}) \).

## 5 Busemann functions on the Wasserstein space

**Definition 5.1** Let \((\mu_t)_{t \geq 0}\) be a unit-speed ray in \( P_p(\mathcal{X}) \). The Busemann function associated with \((\mu_t)_{t \geq 0}\) is defined by

\[
b_\mu(v) = \lim_{t \to +\infty} [W_p(v, \mu_t) - t].
\]  

(5.1)

**Remark 5.2** To show that the limit in (5.1) exists and is finite, by triangle inequality,

\[
W_p(v, \mu_0) \geq W_p(\mu_0, \mu_t) - W_p(v, \mu_t) = t - W_p(v, \mu_t).
\]

For \( 0 \leq t_1 \leq t_2 \),

\[
[t_2 - W_p(v, \mu_t)] - [t_1 - W_p(v, \mu_t)] = W_p(\mu_{t_2}, \mu_{t_1}) + W_p(v, \mu_{t_1}) - W_p(v, \mu_{t_2}) \geq 0.
\]

Thereby \( t - W_p(v, \mu_t) \) is bounded and non-decreasing with respect to \( t \).

**Remark 5.3** The Busemann function \( b_\mu \) is Lipschitzian since

\[
|b_\mu(v^2) - b_\mu(v^1)| = \lim_{t \to \infty} |W_p(v^2, \mu_t) - W_p(v^1, \mu_t)| \leq W_p(v^1, v^2).
\]  

(5.2)

Busemann functions have the following fundamental properties. For the conventional case, we refer to [23, Proposition 41].

**Proposition 5.4** Let \((\mu_t)_{t \geq 0}\) be a unit-speed ray in \( P_p(\mathcal{X}) \). If \((v_t)_{t \geq 0}\) is a co-ray from \( v_0 \) to \((\mu_t)_{t \geq 0}\), then

(i) \( b_\mu(v_{t_1}) - b_\mu(v_{t_2}) = t_2 - t_1 \), for any \( t_1, t_2 \geq 0 \);

(ii) \( b_\mu(\lambda) \leq b_\mu(v_0) + b_\mu(v_0) \), for each \( \lambda \in P_p(\mathcal{X}) \).
Proof \( (i) \) Assume the geodesic sequence \((v^n_i)_{0 \leq i \leq L_n} \in T(v^n_0, \mu_{t_n})\) converges to \((v_i)_{\tau \geq 0}\) as \(n \to \infty\). Let \(0 \leq t_1 \leq t_2\), for \(L_n > t_2\) we have \(W_p(v^n_0, \mu_{t_n}) = W_p(v^n_0, v^n_{t_1}) + W_p(v^n_{t_1}, \mu_{t_n})\), \(i = 1, 2\). Then

\[
|W_p(v^n_1, \mu_{t_n}) - W_p(v^n_{t_1}, \mu_{t_n}) - (t_2 - t_1)| = |W_p(v^n_1, \mu_{t_n}) - W_p(v^n_{t_2}, \mu_{t_n}) - [W_p(v^n_{t_1}, \mu_{t_n}) - W_p(v^n_{t_2}, \mu_{t_n})]| \\
\leq W_p(v^n_{t_1}, v^n_{t_2}) + W_p(v^n_{t_2}, v^n_{t_1}) \to 0.
\]

(ii) For \(s, t \geq 0\), by the triangle inequality,

\[
W_p(\lambda, \mu_s) - s \leq W_p(\lambda, v_t) + W_p(v_t, \mu_s) - s = [W_p(\lambda, v_t) - t] + [W_p(v_t, \mu_s) - s] + t.
\]

Letting \(s \to +\infty\) and applying (i) we have

\[
b_{\mu}(v_t) = [W_p(\lambda, v_t) - t] + b_{\mu}(v_t)
\]

Let \(t \to +\infty\), then the inequality follows. \( \square \)

For \( p > 1 \), \( X \) is non-branching if and only if \( P_p(X) \) is non-branching (see [28, Corollary 7.32]). Based on this fact, we have the following lemma.

**Lemma 5.5** Assume \( X \) is non-branching. Let \((\mu_\tau)_{\tau \geq 0}\) be a unit-speed ray in \( P_p(X) \). If there is another unit-speed ray \((v_\tau)_{\tau \geq 0}\) satisfying

\[
b_{\mu}(v_t) - b_{\mu}(v_\tau) = s - t, \text{ for any } s, t \geq 0,
\]

then for any \( \tau > 0 \), the subray \((v_{\tau + \tau})_{\tau \geq 0}\) is the unique co-ray from \(v_\tau\) to \((\mu_\tau)_{\tau \geq 0}\).

**Proof** For \( \tau > 0 \), assume \((\tilde{v}_t)_{t \geq 0}\) is a co-ray from \(v_\tau\) to \((v_t)_{t \geq 0}\). Let

\[
v'_t = \begin{cases} 
    v_t, & 0 \leq t \leq \tau, \\
    \tilde{v}_{t-\tau}, & t \geq \tau.
\end{cases}
\]

We claim that \((v'_t)_{t \geq 0}\) is a ray. Since \((v_t)_{t \geq 0}\) and \((\tilde{v}_t)_{t \geq 0}\) are unit-speed rays, it is obvious for \(s, t \leq \tau\) or \(s, t \geq \tau\) that

\[
W_p(v'_s, v'_t) = |s - t|.
\]

For \(s < \tau < t\), by triangle inequality and the definition of rays,

\[
W_p(v'_s, v'_t) \leq W_p(v'_s, v'_\tau) + W_p(v'_\tau, v'_t) = W_p(v_s, v_\tau) + W_p(v_\tau, \tilde{v}_{t-\tau}) = (\tau - s) + (t - \tau) = t - s.
\]

On the other hand, by (5.3) and Proposition 5.4,

\[
t - s = (\tau - s) + (t - \tau)
\]

\[
= [b_{\mu}(v_s) - b_{\mu}(v_\tau)] + [b_{\mu}(v_\tau) - b_{\mu}(\tilde{v}_{t-\tau})]
\]

\[
= [b_{\mu}(v'_s) - b_{\mu}(v'_\tau)] + [b_{\mu}(v'_\tau) - b_{\mu}(v'_t)]
\]

\[
= b_{\mu}(v'_s) - b_{\mu}(v'_t)
\]

\[
\leq W_p(v'_s, v'_t).
\]
Here the last inequality follows from (5.2). This proves our claim. Since $P_p(\mathcal{X})$ is non-branching, the ray $(v_t')_{t \geq 0}$ coincides with $(v_t)_{t \geq 0}$. \hfill $\Box$

Notice that (5.3) holds for every co-ray. Lemma 5.5 shows the second conclusion of Theorem 1.2.

**Theorem 5.6** Assume $\mathcal{X}$ is non-branching. Let $(\mu_t)_{t \geq 0}$ and $(v_t)_{t \geq 0}$ be two unit-speed rays in $P_p(\mathcal{X})$. If there exists a sequence $\{t_n\}$ tending to 0 such that $(v_{t_n+t_n})_{t \geq 0}$ is the unique co-ray from $v_{t_n}$ to $(\mu_t)_{t \geq 0}$, then $(v_t)_{t \geq 0}$ is a co-ray from $v_0$ to $(\mu_t)_{t \geq 0}$.

**Proof** Extract a decreasing subsequence of $\{t_n\}$, denoted by $\{t_{n'}\}$, such that $t_{n'} \leq 2^{-n'}$. Let $\tau \geq 1$ be a constant. For each $n'$, by Lemma 5.5, $(v_{t_n'+t_n'})_{t \geq 0}$ is the unique co-ray from $v_{t_{n'}}$ to $(\mu_t)_{t \geq 0}$. Then there exists $s_{n'} \geq \max\{\tau, n'\}$ and geodesic $(v_t^{n'})_{0 \leq t \leq L_{n'}} \in T(v_{t_{n'}}, \mu_{s_{n'}})$ such that $W_p(v_t^{n'}, v_{\tau+t_{n'}}) < 2^{-n'}$, where $L_{n'} = W_p(v_{t_{n'}}, \mu_{s_{n'}})$. It follows immediately that

$$W_p(v_t^{n'}, v_{\tau}) < 2^{-n'+1}. \quad (5.5)$$

By Theorem 4.4, there exists a subsequence $\{(v_t^{n''})_{0 \leq t \leq L_{n''}}\}$ converging to a co-ray $(\lambda_t)_{t \geq 0}$ from $v_0$ to $(\mu_t)_{t \geq 0}$. It suffices to show that $(\lambda_t)_{t \geq 0}$ coincides with $(v_t)_{t \geq 0}$. We can see from (5.5) that $\lambda_t = \tau$. Let $\tilde{v}_t^{n''} = v_{t+n''}$, then $(\tilde{v}_t^{n''})_{0 \leq t \leq L_{n''}-\tau}$ converges to the co-ray $(\lambda_t+\tau)_{t \geq 0}$. By Lemma 5.5, $\lambda_t = v_t$ for every $t \geq \tau$. Thus the conclusion follows from that $P_p(\mathcal{X})$ is non-branching. \hfill $\Box$

Combining Proposition 5.4, Lemma 5.5 and Theorem 5.6, we obtain Theorem 1.3.

**Definition 5.7** Let $(\mathcal{Y}, d)$ be a metric space. The set

$$V(\mathcal{Y}) = \left\{ u : \mathcal{Y} \to \mathbb{R} \mid \text{for any } y \in \mathcal{Y}, u(y) = \min_{z \in \mathcal{Y} \setminus \{y\}} \{d(y, z) + u(z)\} \right\}$$

is said to be the metric viscosity class of $\mathcal{Y}$.

For $u \in V$ we mean that:

- for any $x$ and $y$, $u(x) \leq d(x, y) + u(y)$;
- for any $x$, there exists $y \neq x$ such that $u(x) = d(x, y) + u(y)$.

On a non-compact complete Riemannian manifold $(M, g)$, $u$ is a viscosity solution to the eikonal equation $|\nabla u|_g = 1$ if and only if $u \in \mathcal{V}(M)$ (see for instance [14]). Besides, Busemann functions on $(M, g)$ are viscosity solutions to the eikonal equation (see e.g. [15]). Unfortunately, for $P_p(\mathcal{X})$ with $p \neq 2$, viscosity solutions can not be defined as usual because lack of proper differential structure. These facts motivate us to consider such a set for $P_p(\mathcal{X})$.

**Proof of Theorem 1.4** For any $v_0, \lambda \in P_p(\mathcal{X})$, from (5.2) we have $b_{\mu}(v_0) \leq W_p(v_0, \lambda) + b_{\mu}(\lambda)$. By Theorem 4.4, there exists a co-ray $(v_t)_{t \geq 0}$ from $v_0$ to $(\mu_t)_{t \geq 0}$. Thus by Proposition 5.4,

$$b_{\mu}(v_0) = t + b_{\mu}(v_t) = W_p(v_0, v_t) + b_{\mu}(v_t) \leq \min_{\lambda \in P_p(\mathcal{X}) \setminus \{v_0\}} W_p(v_0, \lambda) + b_{\mu}(\lambda).$$

\hfill $\Box$
Acknowledgements  The authors thank the anonymous referee for helpful suggestions.

References

1. Ambrosio, L., Feng, J.: On a class of first order Hamilton–Jacobi equations in metric spaces. J. Differ. Equ. 256(7), 2194–2245 (2014)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows: In Metric Spaces and in the Space of Probability Measures. Springer, New York (2008)
3. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. Int. Conf. Mach. Learn. 214–223, (2017)
4. Bangert, V.: Geodesic rays, Busemann functions and monotone twist maps. Calc. Var. Partial. Differ. Equ. 2(1), 49–63 (1994)
5. Bangert, V., Emmerich, P.: Area growth and rigidity of surfaces without conjugate points. J. Differ. Geom. 94(3), 367–385 (2013)
6. Bangert, V., Gutkin, E.: Insecurity for compact surfaces of positive genus. Geom. Dedicata. 146(1), 165–191 (2010)
7. Beem, J.K.: Global Lorentzian Geometry. Routledge, London (2017)
8. Bertrand, J., Kloeckner, B.: A geometric study of Wasserstein spaces: Hadamard spaces. J. Topol. Anal. 4(04), 515–542 (2012)
9. Billingsley, P.: Convergence of Probability Measures. Wiley, Hoboken (2013)
10. Burago, D., Burago, Y., Ivanov, S.: A course in metric geometry, Vol. 33. American Mathematical Soc. (2001)
11. Busemann, H.: The Geometry of Geodesics. Academic Press, New York (1955)
12. Carmona, R., Delarue, F.: Probabilistic analysis of mean-field games. SIAM J. Control Optim. 51(4), 2705–2734 (2013)
13. Cheeger, J., Gromoll, D.: The splitting theorem for manifolds of nonnegative Ricci curvature. J. Differ. Geom. 6(1), 119–128 (1971)
14. Cui, X.: Viscosity solutions, ends and ideal boundaries. Ill. J. Math. 60(2), 459–480 (2016)
15. Cui, X., Cheng, J.: Busemann functions and barrier functions. Acta Applicandae Mathematicae (2017)
16. de Acosta, A.: Invariance principles in probability for triangular arrays of B-valued random vectors and some applications. Ann. Probab. 346–373, (1982)
17. Galloway, G., Horta, A.: Regularity of Lorentzian Busemann functions. Trans. Am. Math. Soc. 348(5), 2063–2084 (1996)
18. Gangbo, W., Nguyen, T., Tudorascu, A.: Hamilton–Jacobi equations in the Wasserstein space. Methods Appl. Anal. 15(2), 155–184 (2008)
19. Innami, N.: Differentiability of Busemann functions and total excess. Math. Z. 180(3), 235–247 (1982)
20. Jin, L., Cui, X.: Global viscosity solutions for eikonal equations on class A Lorentzian 2-tori. Geom. Dedicata. 193(1), 155–192 (2018)
21. Lisini, S.: Characterization of absolutely continuous curves in Wasserstein spaces. Calc. Var. Partial. Differ. Equ. 28(1), 85–120 (2007)
22. Papadopoulos, A.: Metric spaces, convexity and nonpositive curvature, Vol. 6. European Mathematical Society, (2005)
23. Petersen, P.: Riemannian Geometry, vol. 171. Springer, New York (2006)
24. Santambrogio, F.: Optimal Transport for Applied Mathematicians, pp. 58–63. Birkhäuser, New York (2015)
25. Shiohama, K.: Topology of complete noncompact manifolds. Geometry of geodesics and related topics. Adv. Stud. Pure Math. 3, 423–450 (1984)
26. Sormani, C.: Busemann functions on manifolds with lower Ricci curvature bounds and minimal volume growth. J. Differ. Geom. 48, 557–585 (1998)
27. Sormani, C.: The almost rigidity of manifolds with lower bounds on Ricci curvature and minimal volume growth. Commun. Anal. Geom. 8(1), 159–212 (2000)
28. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, New York (2008)
29. Whitt, W.: Weak convergence of probability measures on the function space C[0, ∞). Ann. Math. Stat. 41(3), 939–944 (1970)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.