Proceedings of the Royal Society of Edinburgh, **152**, 674–700, *2022*
DOI:10.1017/prm.2021.27

Unbalanced optimal total variation transport problems and generalized Wasserstein barycenters

Nhan-Phu Chung and Thanh-Son Trinh
Department of Mathematics, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon, Gyeonggi-do 16419, Korea (phuchung@skku.edu; phuchung82@gmail.com; sontrinh@skku.edu)

(Received 26 January 2021; accepted 7 May 2021)

In this paper, we establish a Kantorovich duality for unbalanced optimal total variation transport problems. As consequences, we recover a version of duality formula for partial optimal transports established by Caffarelli and McCann; and we also get another proof of Kantorovich–Rubinstein theorem for generalized Wasserstein distance $\widetilde{W}_1^{a,b}$ proved before by Piccoli and Rossi. Then we apply our duality formula to study generalized Wasserstein barycenters. We show the existence of these barycenters for measures with compact supports. Finally, we prove the consistency of our barycenters.

Keywords: Unbalanced optimal transport; duality; generalized Wasserstein space; barycenter

2010 *Mathematics Subject Classification:* Primary: 90C46
Secondary: 49J40, 58E30

1. Introduction

In the 2010s, various generalizations of classical optimal transport problems and Wasserstein distances have been introduced and investigated by numerous authors [**2**, **4**, **6**, **7**, **13**, **17**, **18**, **20**, **21**]. Recently, in 2021 we introduced unbalanced optimal entropy problems [**10**] which cover both optimal entropy transport problems in [**18**] and weak optimal transport problems in [**13**]. In [**10**], under certain conditions of entropy functionals we establish a Kantorovich duality for our unbalanced optimal transport problem. Before stating our first main result, let us review our unbalanced optimal entropy problems.

Given a metric space X, we denote by $\mathcal{M}(X)$ and $\mathcal{P}(X)$ the spaces of all Borel non-negative finite measures and probability measures on X, respectively. Let X_1, X_2 be Polish metric spaces and let $C : X_1 \times \mathcal{P}(X_2) \to [0, \infty]$ be a lower semi-continuous function satisfying that $C(x_1, \cdot)$ is convex for every $x_1 \in X_1$. For every $\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)$, we denote $(\gamma_{x_1})_{x_1 \in X_1}$ its disintegration with respect to its first marginal. Let $F_i : [0, \infty) \to [0, \infty]$, $i = 1, 2$ be convex, lower semi-continuous entropy functions with their recession constants $(F_i)'_\infty := \lim_{s\to\infty} F_i(s)/s$. Given $\mu_1 \in \mathcal{M}(X_1), \mu_2 \in \mathcal{M}(X_2)$ and $\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)$, we define

$$\mathcal{F}_i(\gamma_i|\mu_i) := \int_{X_i} F_i(f_i(x_i))\,\mathrm{d}\mu_i(x_i) + (F_i)'_\infty \gamma_i^\perp(X),$$

$$\mathcal{E}(\boldsymbol{\gamma}|\mu_1,\mu_2) := \sum_{i=1}^{2} \mathcal{F}_i(\gamma_i|\mu_i) + \int_{X_1} C(x_1,\gamma_{x_1})\,\mathrm{d}\gamma_1(x_1),$$

where γ_1, γ_2 are the first and second marginals of $\boldsymbol{\gamma}$, and $\gamma_i = f_i\mu + \gamma_i^\perp$ is the Lebesgue decomposition of γ_i with respect to μ_i.

Our unbalanced optimal entropy-transport problem is defined as

$$\mathcal{E}(\mu_1,\mu_2) := \inf_{\boldsymbol{\gamma}\in\mathcal{M}(X_1\times X_2)} \mathcal{E}(\boldsymbol{\gamma}|\mu_1,\mu_2). \tag{1.1}$$

Similarly to optimal entropy-transport problems in [**18**], to handle with problem (1.1) we often assume that F_i is superlinear, i.e. $(F_i)'_\infty = +\infty$ for $i = 1, 2$. This assumption makes the problems easier as we can get rid of the part $(F_i)'_\infty \gamma_i^\perp(X_i)$ in the expression of $\mathcal{F}_i$.

In the first part of the paper, we investigate problem (1.1) for a special case that F_i is not superlinear, $i = 1, 2$. Given $a, b > 0$, we consider the total variation entropy function $F_i(s) := a|s-1|$, $i = 1, 2$ and the cost function $b \cdot C$. In this case, problem (1.1) will become

$$\mathrm{E}^{a,b}(\mu_1,\mu_2) := \inf_{\boldsymbol{\gamma}\in\mathcal{M}(X_1\times X_2)} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1,\mu_2), \tag{1.2}$$

where $\mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1,\mu_2) := a\,|\mu_1-\gamma_1| + a\,|\mu_2-\gamma_2| + b\int_{X_1} C(x_1,\gamma_{x_1})d\gamma_1(x_1)$.

As F_i is not superlinear, to deal with problem (1.2) we need new techniques being different from [**10**, **18**]. We define

$$\begin{aligned}\Phi_I := \Big\{&(\varphi_1,\varphi_2) \in C_b(X_1)\times C_b(X_2) : \varphi_1(x_1), \varphi_2(x_2)\\ &\geqslant -a \text{ for every } x_i \in X_i, i = 1, 2 \quad \text{and } \varphi_1(x_1)\\ &+ q(\varphi_2) \leqslant b\cdot C(x_1,q) \text{ for every } x_1 \in X_1, q \in \mathcal{P}(X_2)\Big\}.\end{aligned} \tag{1.3}$$

Next, we define the functional $J : \mathbb{R} \to (-\infty, +\infty]$ by

$$J(\phi) = \sup_{s>0} \frac{\phi - a|1-s|}{s} = \begin{cases} +\infty & \text{if } \phi > a,\\ \phi & \text{if } -a \leqslant \phi \leqslant a,\\ -a & \text{otherwise.}\end{cases} \tag{1.4}$$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Then we define

$$\begin{aligned}\Phi_J := \Big\{&(\varphi_1, \varphi_2) \in C_b(X_1) \times C_b(X_2) : \varphi_1(x_1), \varphi_2(x_2) \\ &\leqslant a \text{ for every } x_i \in X_i, i = 1, 2 \quad \text{and } J(\varphi_1(x_1)) \\ &+ q(J(\varphi_2)) \leqslant b \cdot C(x_1, q) \text{ for every } x_1 \in X_1, q \in \mathcal{P}(X_2)\Big\}. \qquad (1.5)\end{aligned}$$

Our main result for the first part is a Kantorovich duality of problem (1.2).

THEOREM 1.1. *Let X_1, X_2 be locally compact, Polish metric spaces. Let $C : X_1 \times \mathcal{P}(X_2) \to [0, \infty]$ be a lower semi-continuous function such that $C(x_1, \cdot)$ is convex for every $x_1 \in X_1$. Then for every $\mu_i \in \mathcal{M}(X_i), i = 1, 2$ we have*

$$\begin{aligned}E^{a,b}(\mu_1, \mu_2) &= \sup_{(\varphi_1,\varphi_2)\in\Phi_I} \sum_{i=1}^{2} \int_{X_i} I(\varphi_i(x_i))\, d\mu_i(x_i) \\ &= \sup_{(\varphi_1,\varphi_2)\in\Phi_J} \sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i)\, d\mu_i(x_i),\end{aligned}$$

where

$$I(\varphi) := \inf_{s\geqslant 0} (s\varphi + a|1 - s|) = \begin{cases} a & \text{if } \varphi > a \\ \varphi & \text{if } -a \leqslant \varphi \leqslant a. \\ -\infty & \text{otherwise} \end{cases} \qquad (1.6)$$

We need the local compactness assumption on theorem 1.1 because in our proof we use Riesz representation theorem stating that $\mathcal{M}_s(X)$, the space of all signed Borel measures with finite masses on X, is the dual space of $C_0(X)$, and it is only true for locally compact spaces. However, as the duality results for optimal entropy transport problems in [**18**] were proved for general Polish spaces by a different method, we expect that theorem 1.1 would still hold for these general spaces.

Now we present consequences of theorem 1.1. The first one is that we can get a version of [**6**, corollary 2.6]. Let $X_1 = X_2 = X$ be a Polish space, $\mu_1, \mu_2 \in \mathcal{M}(X)$, $a, b > 0$ and $c_1 : X \times X \to [0, +\infty]$ be a lower semi-continuous function. We define $\hat{X} := X \cup \{\hat{\infty}\}$ by attaching an isolated point $\hat{\infty}$ to X. We endow $\hat{X}$ with the topology induced from the topology of X and the isolated point $\hat{\infty}$. We extend the cost function

$$\hat{c}_1(x, y) := \begin{cases} b \cdot c_1(x, y) & \text{if } x \neq \hat{\infty} \text{ and } y \neq \hat{\infty}, \\ a & \text{if } x \in X, y = \hat{\infty} \text{ or } x = \hat{\infty}, y \in X, \\ 0 & \text{otherwise}, \end{cases} \qquad (1.7)$$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

and measures μ_1, μ_2 to $\hat{X}$ by adding a Dirac measure at infinity: $\hat{\mu}_1 := \mu_1 + |\mu_2|\delta_\infty$, $\hat{\mu}_2 := \mu_2 + |\mu_1|\delta_\infty$. Then the measures $\hat{\mu}_1$ and $\hat{\mu}_2$ have the same masses. We define

$$\Gamma(\hat{\mu}_1, \hat{\mu}_2) := \Big\{ \hat{\gamma} \in \mathcal{M}(\hat{X} \times \hat{X}) : \hat{\gamma}(A \times \hat{X}) = \hat{\mu}_1(A), \hat{\gamma}(\hat{X} \times A) = \hat{\mu}_2(A) \text{ for Borel } A \subset \hat{X} \Big\}.$$

Then we will get a version of [6, corollary 2.6] as follows.

Corollary 1.2. *Given a locally compact, Polish metric space X, $\mu_1, \mu_2 \in \mathcal{M}(X)$, $a, b > 0$, and a lower semi-continuous function $c_1 : X \times X \to [0, +\infty]$. Then*

$$\sup_{\substack{(\hat{\varphi}_1, \hat{\varphi}_2) \in L^1(\hat{\mu}_1) \times L^1(\hat{\mu}_2) \\ \hat{\varphi}_1(x) + \hat{\varphi}_2(y) \leqslant \hat{c}_1(x,y)}} \sum_{i=1}^{2} \int_{\hat{X}} \hat{\varphi}_i(x)\, d\hat{\mu}_i(x) = \inf_{\hat{\gamma} \in \Gamma(\hat{\mu}_1, \hat{\mu}_2)} \int_{\hat{X} \times \hat{X}} \hat{c}_1(x, y)\, d\hat{\gamma}(x, y).$$

Another consequence of theorem 1.1 is that we establish a Kantorovich duality for generalized Wasserstein distance $\widetilde{W}_p^{a,b}$, and a version of Kantorovich–Rubinstein theorem for generalized Wasserstein distance $\widetilde{W}_1^{a,b}$.

Let (X, d) be a metric space. For a function $f : X \to \mathbb{R}$, we denote

$$\|f\|_{Lip} := \sup_{x,y \in X, x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}.$$

Corollary 1.3. *Let (X, d) be a locally compact and Polish metric space. Then for every $a, b > 0, \mu, \nu \in \mathcal{M}(X)$ and $p \geqslant 1$ we have*

1. $\widetilde{W}_p^{a,b}(\mu, \nu)^p = \sup\limits_{(\varphi_1, \varphi_2) \in \Phi_W} \Big\{ \int_X I(\varphi_1(x)) d\mu(x) + \int_X I(\varphi_2(x)) d\nu(x) \Big\}$, *where*

$$\Phi_W := \{(\varphi_1, \varphi_2) \in C_b(X) \times C_b(X) \mid \varphi_1(x) + \varphi_2(y) \leqslant (b \cdot d(x, y))^p \text{ and } \varphi_1(x), \varphi_2(y) \geqslant -a,\ \forall x, y \in X\}.$$

2. $\widetilde{W}_1^{a,b}(\mu, \nu) = \sup \Big\{ \int_X f d(\mu - \nu) : f \in \mathbb{F} \Big\}$, *where*

$$\mathbb{F} := \big\{ f \in C_b(X), \|f\|_\infty \leqslant a, \|f\|_{Lip} \leqslant b \big\}.$$

Note that corollary 1.3 (1) is proved for the case $p = 1$ in [9], and corollary 1.3 (2) is a main result of [21] proved by a different method there. In the second part of the paper, we apply corollary 1.3 to study barycenters of generalized Wasserstein distances. In 2002, Sturm investigated barycenters in non-positive curvature spaces as he showed the existence, uniqueness and contraction of barycenters in such spaces [24]. Because Wassertein spaces are not in the framework of non-positive curvature spaces, to study the existence, uniqueness and properties of Wasserstein barycenters over $\mathbb{R}^n$, Agueh and Carlier introduced dual problems of the primal barycenter problem and used convex analysis to handle them [1]. Recently, barycenters in

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Hellinger–Kantorovich spaces, siblings of Wasserstein spaces, have been investigated in [8, 12].

On the other hand, in 2014, Piccoli and Rossi introduced generalized Wasserstein distances [20] and established a duality Kantorovich–Rubinstein formula and a generalized Benamou–Breiner formula for them [21]. Combining corollary 1.3 with the streamline of Agueh and Carlier's work [1], we study the existence and consistency of generalized Wasserstein barycenters.

More precisely, first we show the existence of generalized Wasserstein barycenters whenever starting measures have compact supports. Second, we introduce and investigate a dual problem of the barycenter problem. Although our barycenters are not unique, we still can establish their consistency as Boissard, Le Gouic and Loubes did in the Wasserstein case [5].

Our paper is organized as follows. In § 2, we review basic notations and generalized Wasserstein distances $\widetilde{W}_p^{a,b}$. In § 3, we prove theorem 1.1, corollaries 1.2 and 1.3. In § 4, we study our primal barycenter problem and its dual problems. We also show the existence and consistency of generalized Wasserstein barycenters in this last section.

2. Preliminaries

Let (X, d) be a metric space. We denote by $\mathcal{M}(X)$ and $\mathcal{P}(X)$ the sets of all non-negative Borel measures with finite mass and all probability Borel measures, respectively.

Given a Borel measure μ, we denote its mass by $|\mu| := \mu(X)$. In the general case, if $\mu = \mu^+ - \mu^-$ is a signed Borel measure then $|\mu| := |\mu^+| + |\mu^-|$. A set $M \subset \mathcal{M}(X)$ is bounded if $\sup_{\mu\in M} |\mu| < \infty$, and it is *tight* if for every $\varepsilon > 0$, there exists a compact subset K_ε of X such that for all $\mu \in M$, we have $\mu(X\backslash K_\varepsilon) \leqslant \varepsilon$.

For every $\mu_1, \mu_2 \in \mathcal{M}(X)$, we say that μ_1 is absolutely continuous with respect to μ_2 and write $\mu_1 \ll \mu_2$ if $\mu_2(A) = 0$ yields $\mu_1(A) = 0$ for every Borel subset A of X. We call that μ_1 and μ_2 are mutually singular and write $\mu_1 \perp \mu_2$ if there exists a Borel subset B of X such that $\mu_1(B) = \mu_2(X\backslash B) = 0$. We write $\mu_1 \leqslant \mu_2$ if for all Borel subset A of X we have $\mu_1(A) \leqslant \mu_2(A)$.

For every $p \geqslant 1$, we denote by $\mathcal{M}_p(X)$ (reps. $\mathcal{P}_p(X)$) the space of all measures $\mu \in \mathcal{M}(X)$ (reps. $\mathcal{P}(X)$) with finite p-moment, i.e. there is some (and therefore any) $x_0 \in X$ such that $\int_X d^p(x, x_0) d\mu(x) < \infty$.

For every measures $\mu_1, \mu_2 \in \mathcal{M}(X)$, a Borel probability measure $\boldsymbol{\pi}$ on $X \times X$ is called a transference plan between μ_1 and μ_2 if

$$|\mu_1|\boldsymbol{\pi}(A \times X) = \mu_1(A) \text{ and } |\mu_2|\boldsymbol{\pi}(X \times B) = \mu_2(B),$$

for every Borel subsets A, B of X. We denote the set of all transference plan between μ_1 and μ_2 by $\Pi(\mu_1, \mu_2)$.

Given measures $\mu_1, \mu_2 \in \mathcal{M}_p(X)$ with the same mass, i.e. $|\mu_1| = |\mu_2|$. The Wasserstein distance between μ_1 and μ_2 is defined by

$$W_p(\mu_1, \mu_2) := \left(|\mu_1| \inf_{\pi\in\Pi(\mu_1,\mu_2)} \int_{X\times X} d^p(x, y)\, \mathrm{d}\boldsymbol{\pi}(x, y)\right)^{1/p}.$$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

For each $\mu_1, \mu_2 \in \mathcal{M}(X)$ with $|\mu_1| = |\mu_2|$, we denote by $\mathrm{Opt}_p(\mu_1, \mu_2)$ the set of all $\boldsymbol{\pi} \in \Pi(\mu_1, \mu_2)$ such that $W_p^p(\mu_1, \mu_2) = |\mu_1| \int_{X\times X} d^p(x, y) d\boldsymbol{\pi}(x, y)$. If (X, d) is a Polish metric space, i.e. (X, d) is complete and separable then $\mathrm{Opt}_p(\mu_1, \mu_2)$ is non-empty [**25**, theorem 1.3].

THEOREM 2.1 (Prokhorov's theorem). *If (X, d) is a Polish metric space then a subset $M \subset \mathcal{M}(X)$ is bounded and tight if and only if M is relatively compact under the weak*-topology.*

We now review the definitions of the generalized Wasserstein distances. They were introduced by Piccoli and Rossi in [**20**, **21**]. For convenience to establish Kantorovich duality formulas for the generalized Wasserstein distances, we adapt slightly the original ones.

DEFINITION 2.2. Let X be a Polish metric space and let $a, b > 0, p \geqslant 1$. For every $\mu_1, \mu_2 \in \mathcal{M}(X)$, the generalized Wasserstein distance $\widetilde{W}_p^{a,b}$ between μ_1 and μ_2 is defined by

$$\widetilde{W}_p^{a,b}(\mu_1, \mu_2) := \left(\inf \{C\left(\widetilde{\mu_1}, \widetilde{\mu_2}\right) \mid \widetilde{\mu_1}, \widetilde{\mu_2} \in \mathcal{M}_p(X), |\widetilde{\mu_1}| = |\widetilde{\mu_2}|\}\right)^{1/p},$$

where $C(\widetilde{\mu_1}, \widetilde{\mu_2}) = a\,|\mu_1 - \widetilde{\mu_1}| + a\,|\mu_2 - \widetilde{\mu_2}| + b^p\, W_p^p(\widetilde{\mu_1}, \widetilde{\mu_2})$.

The following results can be adapted from the proofs of [**20**, proposition 1 and theorem 3].

PROPOSITION 2.3 [**20**, proposition 1]. *If X is a Polish metric space then $(\mathcal{M}(X), \widetilde{W}_p^{a,b})$ is a metric space. Moreover, there exist $\widetilde{\mu_1}, \widetilde{\mu_2} \in \mathcal{M}_p(X)$ such that $|\widetilde{\mu_1}| = |\widetilde{\mu_2}|, \widetilde{\mu_1} \leqslant \mu_1, \widetilde{\mu_2} \leqslant \mu_2$, and $\widetilde{W}_p^{a,b}(\mu_1, \mu_2)^p = C(\widetilde{\mu_1}, \widetilde{\mu_2})$.*

If measures $\widetilde{\mu_1}, \widetilde{\mu_2} \in \mathcal{M}_p(X)$ have the same mass such that $\widetilde{W}_p^{a,b}(\mu_1, \mu_2)^p = C(\widetilde{\mu_1}, \widetilde{\mu_2})$ then we say that $(\widetilde{\mu_1}, \widetilde{\mu_2})$ is an optimal for $\widetilde{W}_p^{a,b}(\mu_1, \mu_2)$.

Let X_1, X_2 be Polish metric spaces. For every $\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)$, we denote its disintegration with respect to its first marginal by $(\gamma_{x_1})_{x_1 \in X_1}$. We also denote by γ_1 and γ_2 the first and second marginals of $\boldsymbol{\gamma}$, i.e.

$$\gamma_1(B_1) = \boldsymbol{\gamma}(B_1 \times X_2) \text{ and } \gamma_2(B_2) = \boldsymbol{\gamma}(X_1 \times B_2) \text{ for Borel sets } B_i \subset X_i.$$

3. Unbalanced optimal total variation transport problems

Let $C : X_1 \times \mathcal{P}(X_2) \to [0, \infty]$ be a lower semi-continuous function such that for every $x_1 \in X_1$ we have

$$C(x_1, tq_1 + (1-t)q_2) \leqslant tC(x_1, q_1) + (1-t)C(x_1, q_2),$$

for every $t \in [0, 1], q_1, q_2 \in \mathcal{P}(X_2)$.

For every $a, b > 0, \mu_i \in \mathcal{M}(X_i), i = 1, 2$ and every $\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)$, we recall

$$\mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2) := a\,|\mu_1 - \gamma_1| + a\,|\mu_2 - \gamma_2| + b \int_{X_1} C(x_1, \gamma_{x_1})\,\mathrm{d}\gamma_1(x_1)$$

Then for every $\mu_i \in \mathcal{M}(X_i), i = 1, 2$ we have

$$\mathrm{E}^{a,b}(\mu_1, \mu_2) := \inf_{\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2)} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \inf_{\boldsymbol{\gamma} \in M} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2),$$

where $M := \{\boldsymbol{\gamma} \in \mathcal{M}(X_1 \times X_2) | \int_X C(x_1, \gamma_{x_1}) d\gamma_1(x_1) < \infty\}$.

Lemma 3.1. *Let X_1, X_2 be Polish metric spaces and $a, b > 0$. For every $\mu_1 \in \mathcal{M}(X_1)$ and $\mu_2 \in \mathcal{M}(X_2)$ we have*

$$E^{a,b}(\mu_1, \mu_2) = \inf_{\boldsymbol{\gamma} \in M} E^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2) = \inf_{\boldsymbol{\gamma} \in M^{\leqslant}(\mu_1, \mu_2)} E^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2),$$

where $M^{\leqslant}(\mu_1, \mu_2) := \{\boldsymbol{\gamma} \in M | \gamma_i \leqslant \mu_i, i = 1, 2\}$.

Proof. It is clear that we only need to prove that

$$\inf_{\boldsymbol{\gamma} \in M} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2) \geqslant \inf_{\boldsymbol{\gamma} \in M^{\leqslant}(\mu_1, \mu_2)} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2).$$

For any $\boldsymbol{\alpha} \in M$, let α_1, α_2 be the first and second marginals of $\boldsymbol{\alpha}$. Suppose that $\alpha_1 = f\mu_1 + \mu_1^{\perp}$ is the Lebesgue decomposition of α_1 with respect to μ_1. We define $\overline{\alpha}_1 := \min\{f, 1\}\mu_1$. Then $\overline{\alpha}_1 \leqslant \mu_1$ and $\overline{\alpha}_1 \leqslant \alpha_1$. By the Radon–Nikodym theorem we get that there exists a measurable function $g : X_1 \to [0, \infty)$ such that $\overline{\alpha}_1 = g\alpha_1$ and $g \leqslant 1$ α_1-a.e.

Next, for every Borel subsets A_i of X_i, $i = 1, 2$, we define

$$\overline{\boldsymbol{\alpha}}(A_1 \times A_2) := \int_{A_1 \times A_2} g(x_1)\,\mathrm{d}\boldsymbol{\alpha}(x_1, x_2).$$

Then $\overline{\boldsymbol{\alpha}}(A_1 \times X_2) = \int_{A_1} g(x_1) d\alpha_1(x_1) = \overline{\alpha}_1(A_1)$ for every Borel subset A_1 of X_1. For any Borel subset A_2 of X_2, we define $\overline{\alpha}_2(A_2) := \int_{X_1 \times A_2} g(x_1) d\boldsymbol{\alpha}(x_1, x_2)$. Then $\overline{\alpha}_2(A_2) = \overline{\boldsymbol{\alpha}}(X_1 \times A_2)$. This means that $\overline{\alpha}_1$ and $\overline{\alpha}_2$ are the first and second marginals of $\overline{\boldsymbol{\alpha}}$. Since $g \leqslant 1$ α_1-a.e one has $\overline{\boldsymbol{\alpha}} \leqslant \boldsymbol{\alpha}$. Moreover, for every Borel function $h : X_1 \times X_2 \to [0, +\infty]$ we have

$$\begin{aligned}
\int_{X_1 \times X_2} h(x_1, x_2)\,\mathrm{d}\overline{\boldsymbol{\alpha}}(x_1, x_2) &= \int_{X_1 \times X_2} h(x_1, x_2) g(x_1)\,\mathrm{d}\boldsymbol{\alpha}(x_1, x_2) \\
&= \int_{X_1} \left(\int_{X_2} h(x_1, x_2) g(x_1)\,\mathrm{d}\alpha_{x_1}(x_2) \right) \mathrm{d}\alpha_1(x_1) \\
&= \int_{X_1} \left(\int_{X_2} h(x_1, x_2)\,\mathrm{d}\alpha_{x_1}(x_2) \right) \mathrm{d}\overline{\alpha}_1(x_1).
\end{aligned}$$

Therefore, by the uniqueness of disintegration we get that $\overline{\alpha}_{x_1} = \alpha_{x_1}$ $\overline{\alpha}_1$-a.e. Then,

$$\int_{X_1} C(x_1, \overline{\alpha}_{x_1})\,\mathrm{d}\overline{\alpha}_1(x_1) \leqslant \int_{X_1} C(x_1, \alpha_{x_1})\,\mathrm{d}\alpha_1(x_1).$$

Notice that as $\overline{\boldsymbol{\alpha}} \leqslant \boldsymbol{\alpha}$, we have $\overline{\alpha}_2 \leqslant \alpha_2$.

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

On the other hand, putting $D := \{x_1 \in X_1 : f(x_1) \leqslant 1\}$ then we get that

$$\begin{aligned}
|\mu_1 - \alpha_1| &= \int_{X_1} |1 - f(x_1)| \,\mathrm{d}\mu_1 + \mu_1^{\perp}(X_1) \\
&= \int_D (1 - f(x_1)) \,\mathrm{d}\mu_1 + \int_{X_1\setminus D} (f(x_1) - 1) \,\mathrm{d}\mu_1 + \mu_1^{\perp}(X_1) \\
&= \int_D \mathrm{d}\mu_1 - \int_D \mathrm{d}\overline{\alpha}_1 + \int_{X_1\setminus D} f(x_1) \,\mathrm{d}\mu_1 - \int_{X_1\setminus D} \mathrm{d}\overline{\alpha}_1 + \mu_1^{\perp}(X_1) \\
&= \int_D \mathrm{d}\mu_1 - \int_D \mathrm{d}\overline{\alpha}_1 + \int_{X_1\setminus D} \mathrm{d}\alpha_1 - \int_{X_1\setminus D} \mathrm{d}\overline{\alpha}_1 - \mu_1^{\perp}(X_1\setminus D) + \mu_1^{\perp}(X_1) \\
&= \int_D \mathrm{d}\mu_1 - \int_D \mathrm{d}\overline{\alpha}_1 + \int_{X_1\setminus D} \mathrm{d}\alpha_1 - \int_{X_1\setminus D} \mathrm{d}\overline{\alpha}_1 + \int_D \mathrm{d}\alpha_1 - \int_D f \mathrm{d}\mu_1 \\
&= |\mu_1 - \overline{\alpha}_1| + \int_{X_1\setminus D} \mathrm{d}\alpha_1 - \int_{X_1\setminus D} \mathrm{d}\overline{\alpha}_1 + \int_D \mathrm{d}\alpha_1 - \int_D \mathrm{d}\overline{\alpha}_1 \\
&= |\mu_1 - \overline{\alpha}_1| + |\alpha_1 - \overline{\alpha}_1| .
\end{aligned}$$

Observe that $|\alpha_1 - \overline{\alpha}_1| = |\alpha_2 - \overline{\alpha}_2|$, one gets

$$|\mu_1 - \overline{\alpha}_1| + |\mu_2 - \overline{\alpha}_2| = |\mu_1 - \alpha_1| - |\alpha_2 - \overline{\alpha}_2| + |\mu_2 - \overline{\alpha}_2| \leqslant |\mu_1 - \alpha_1| + |\mu_2 - \alpha_2| .$$

Hence, we obtain that $\mathrm{E}^{a,b}(\boldsymbol{\alpha}|\mu_1, \mu_2) \geqslant \mathrm{E}^{a,b}(\overline{\boldsymbol{\alpha}}|\mu_1, \mu_2)$.

Applying this process again for $\overline{\boldsymbol{\alpha}}$, we can find a plan $\widehat{\boldsymbol{\alpha}} \in M$ with its marginals are $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ such that $\widehat{\boldsymbol{\alpha}} \leqslant \overline{\boldsymbol{\alpha}}$ and

$$\mathrm{E}^{a,b}\left(\overline{\boldsymbol{\alpha}}|\mu_1, \mu_2\right) \geqslant \mathrm{E}^{a,b}\left(\widehat{\boldsymbol{\alpha}}|\mu_1, \mu_2\right);$$

and $\widehat{\alpha}_2 \leqslant \mu_2$, $\widehat{\alpha}_1 \leqslant \overline{\alpha}_1 \leqslant \mu_1$. Thus, $\widehat{\boldsymbol{\alpha}} \in M^{\leqslant}(\mu_1, \mu_2)$. Therefore, we get that

$$\mathrm{E}^{a,b}\left(\boldsymbol{\alpha}|\mu_1, \mu_2\right) \geqslant \mathrm{E}\left(\overline{\boldsymbol{\alpha}}|\mu_1, \mu_2\right) \geqslant \mathrm{E}^{a,b}\left(\widehat{\boldsymbol{\alpha}}|\mu_1, \mu_2\right) \geqslant \inf_{\boldsymbol{\gamma} \in M^{\leqslant}(\mu_1, \mu_2)} \mathrm{E}^{a,b}\left(\boldsymbol{\gamma}|\mu_1, \mu_2\right).$$

This implies that $\inf_{\boldsymbol{\gamma} \in M} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2) \geqslant \inf_{\boldsymbol{\gamma} \in M^{\leqslant}(\mu_1, \mu_2)} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2)$. □

For every $a, b > 0$ and $(\mu_1, \mu_2) \in \mathcal{M}(X_1) \times \mathcal{M}(X_2)$ we denote by $\mathrm{Opt}^{a,b}(\mu_1, \mu_2)$ the set of all $\boldsymbol{\gamma} \in M^{\leqslant}(\mu_1, \mu_2)$ such that $\mathrm{E}^{a,b}(\mu_1, \mu_2) = \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2)$.

Lemma 3.2. *Let X_1, X_2 be Polish metric spaces. For every $a, b > 0$ and $\mu_i \in \mathcal{M}(X_i), i = 1, 2$ the set $Opt^{a,b}(\mu_1, \mu_2)$ is a non-empty subset of $\mathcal{M}(X_1 \times X_2)$.*

Proof. From lemma 3.1, we choose a sequence of $\boldsymbol{\gamma}^n \in M^{\leqslant}(\mu_1, \mu_2)$ such that

$$\lim_{n\to\infty} \mathrm{E}^{a,b}(\boldsymbol{\gamma}^n|\mu_1, \mu_2) = \mathrm{E}^{a,b}(\mu_1, \mu_2).$$

Then $\gamma_i^n \leqslant \mu_i$ for $i = 1, 2$ and every $n \in \mathbb{N}$. Since $\mu_i \in \mathcal{M}(X_i)$ for $i = 1, 2$, one has $\{\gamma_1^n\}_n$ and $\{\gamma_2^n\}_n$ are tight and bounded. By [**3**, Lemma 5.2.2] one gets that $\{\boldsymbol{\gamma}^n\}_{n\in\mathbb{N}}$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

is also tight and bounded. Thus, by Prokhorov's theorem, passing to a subsequence we can assume that $\lim_{n\to\infty} \boldsymbol{\gamma}^n = \boldsymbol{\gamma}$ under the weak*-topology for some $\boldsymbol{\gamma} \in \mathcal{M}(X \times X)$.

Next, for any Borel subset A_1 of X_1 we have

$$\begin{aligned}\gamma_1(A_1) &= \boldsymbol{\gamma}(A_1 \times X_2)\\ &= \inf\{\boldsymbol{\gamma}(V) : V \subset X_1 \times X_2 \text{ open}, A_1 \times X_2 \subset V\}\\ &\leqslant \inf\{\boldsymbol{\gamma}(U \times X_2) : U \subset X_1 \text{ open}, A_1 \subset U\}.\end{aligned}$$

Applying [**19**, theorem 6.1 page 40] we obtain that $\boldsymbol{\gamma}(U \times X_2) \leqslant \liminf_{n\to\infty} \boldsymbol{\gamma}^n (U \times X_2) \leqslant \mu_1(U)$ for every open subset U of X_1. This yields, $\gamma_1 \leqslant \mu_1$. Similarly, we also have $\gamma_2 \leqslant \mu_2$. Moreover, using [**19**, theorem 6.1 page 40] again we also have that $\limsup_{n\to\infty} |\boldsymbol{\gamma}^n| \leqslant |\boldsymbol{\gamma}| \leqslant \liminf_{n\to\infty} |\boldsymbol{\gamma}^n|$. This implies that $\lim_{n\to\infty} |\boldsymbol{\gamma}^n| = |\boldsymbol{\gamma}|$. Hence, $\lim_{n\to\infty} |\mu_i - \gamma_1^n| = |\mu_i - \gamma_i|$, for $i = 1, 2$.

Applying [**10**, lemma 3.5] we obtain that

$$\liminf_{n\to\infty} \int_{X_1} C(x_1, \gamma^n_{x_1})\, \mathrm{d}\gamma_1^n(x_1) \geqslant \int_{X_1} C(x_1, \gamma_{x_1})\, \mathrm{d}\gamma_1(x_1).$$

So, we get that

$$a\,|\mu_1 - \gamma_1| + a\,|\mu_2 - \gamma_2| + b\int_{X_1} C(x_1, \gamma_{x_1})\, \mathrm{d}\gamma_1(x_1) \leqslant \mathrm{E}^{a,b}\,(\mu_1, \mu_2)\,.$$

This implies that $\mathrm{Opt}^{a,b}(\mu_1, \mu_2)$ is non-empty. □

We recall that the functionals I, J are defined as in (1.6), (1.4) and Φ_I, Φ_J are defined as in (1.3),(1.5), respectively. We also set

$$\Phi_J^0 := \{\varphi = (\varphi_1, \varphi_2) \in C_0(X_1) \times C_0(X_2) : \varphi \in \Phi_J\}.$$

Lemma 3.3. *For every $\mu_1 \in \mathcal{M}(X_1)$ and $\mu_2 \in \mathcal{M}(X_2)$ one has*

$$\sup_{(\varphi_1,\varphi_2)\in\Phi_J} \sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i)\, d\mu_i(x_i) \leqslant \sup_{(\psi_1,\psi_2)\in\Phi_I} \sum_{i=1}^{2} \int_{X_i} I(\psi_i(x_i))\, d\mu_i(x_i).$$

Proof. For every $(\varphi_1, \varphi_2) \in \Phi_J$ and $i \in \{1, 2\}$ we define $\overline{\varphi}_i := J(\varphi_i)$. Then for every $x_i \in X_i$ we have that $\overline{\varphi}_i(x_i) \in [-a, a]$ for $i = 1, 2$. Thus, from (1.6) one has $I(\overline{\varphi}_i) = \overline{\varphi}_i$. Moreover, by the definition of Φ_J, we also get $\overline{\varphi}_1(x_1) + q(\overline{\varphi}_2) \leqslant b \cdot C(x_1, q)$ for every $x_1 \in X_1, q \in \mathcal{P}(X_2)$. Since J is continuous on $(-\infty, a]$ we get that $(\overline{\varphi}_1, \overline{\varphi}_2) \in \Phi_I$. As $\overline{\varphi}_i = J(\varphi_i) \geqslant \varphi_i$ for $i = 1, 2$ we obtain that

$$\sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i)\, \mathrm{d}\mu_i(x_i) \leqslant \sum_{i=1}^{2} \int_{X_i} \overline{\varphi}_i(x_i) \mathrm{d}\mu_i(x_i) = \sum_{i=1}^{2} \int_{X_i} I(\overline{\varphi}_i(x_i))\, \mathrm{d}\mu_i(x_i).$$

Hence, we get the result. □

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

LEMMA 3.4. *Suppose that X_1, X_2 are Polish metric spaces. For every $a, b > 0$ and $\mu_i \in \mathcal{M}(X_i), i = 1, 2$, we have*

$$E^{a,b}(\mu_1, \mu_2) \geqslant \sup_{(\varphi_1,\varphi_2)\in\Phi_I} \sum_{i=1}^{2} \int_{X_i} I\left(\varphi_i(x_i)\right) d\mu_i(x_i).$$

Proof. Let $\mu_i \in \mathcal{M}(X_i), i = 1, 2$. Thanks to lemma 3.2, let $\boldsymbol{\gamma} \in \mathrm{Opt}^{a,b}(\mu_1, \mu_2)$. Then $\gamma_i \leqslant \mu_i$ for $i = 1, 2$. By Radon–Nikodym theorem, there exists a measurable function $f_i : X \to [0, \infty)$ such that $\gamma_i = f_i\mu_i$ and $f_i \leqslant 1$ μ_i-a.e. Therefore, for every $(\varphi_1, \varphi_2) \in \Phi_I$ we get that

$$\begin{aligned}
\mathrm{E}^{a,b}\left(\mu_1, \mu_2\right) &= \sum_{i=1}^{2} \int_{X_i} a\left(1 - f_i(x_i)\right) \mathrm{d}\mu_i(x_i) + b \int_{X_1} C(x_1, \gamma_{x_1})\, \mathrm{d}\gamma_1(x_1) \\
&\geqslant \sum_{i=1}^{2} \int_{X_i} a\left(1 - f_i(x_i)\right) \mathrm{d}\mu_i(x_i) + \int_{X_1} \left(\varphi_1(x_1) + \gamma_{x_1}(\varphi_2)\right) \mathrm{d}\gamma_1(x_1) \\
&= \sum_{i=1}^{2} \int_{X_i} a\left(1 - f_i(x_i)\right) \mathrm{d}\mu_i(x_i) + \int_{X_1} \varphi_1(x_1)\, \mathrm{d}\gamma_1 \\
&\quad + \int_{X_1} \int_{X_2} \varphi_2(x_2)\, \mathrm{d}\gamma_{x_1}(x_2)\, \mathrm{d}\gamma_1(x_1) \\
&= \sum_{i=1}^{2} \int_{X_i} \left(a\left(1 - f_i(x_i)\right) + f_i(x_i)\varphi_i(x_i)\right) \mathrm{d}\mu_i(x_i).
\end{aligned}$$

Furthermore, for all $x_i \in X_i$, since $f_i(x_i) \geqslant 0$, $f_i \leqslant 1$ μ_i-a.e and (1.6) we get

$$\int_{X_i} I(\varphi_i(x_i))\, \mathrm{d}\mu_i(x_i) \leqslant \int_{X_i} (f_i(x_i)\varphi_i(x_i) + a(1 - f_i(x_i))\, \mathrm{d}\mu_i(x_i), \text{ for } i = 1, 2.$$

Hence, we get the result. □

For $i = 1, 2$, we denote by $\mathcal{M}_s(X_i)$ the space of signed Borel measures with a finite mass on X_i. Then for every $a, b > 0$ we define the functional $\mathrm{ET}^{a,b} : \mathcal{M}_s(X_1) \times \mathcal{M}_s(X_2) \to [0, +\infty]$ by

$$\mathrm{ET}^{a,b}(\mu_1, \mu_2) = \begin{cases} \inf_{\boldsymbol{\gamma}\in M} \mathrm{E}^{a,b}(\boldsymbol{\gamma}|\mu_1, \mu_2) & \text{if } (\mu_1, \mu_2) \in \mathcal{M}(X_1) \times \mathcal{M}(X_2), \\ +\infty & \text{otherwise.} \end{cases}$$

LEMMA 3.5. *Let X_1, X_2 be Polish metric spaces and $a, b > 0$. Then*

1. *$ET^{a,b}$ is convex and satisfies that $ET^{a,b}(k\mu_1, k\mu_2) = kET^{a,b}(\mu_1, \mu_2)$, for every $\mu_i \in \mathcal{M}_s(X_i), i = 1, 2$ and $k > 0$.*

2. *If moreover X_1 and X_2 are locally compact then $ET^{a,b}$ is lower semi-continuous under the weak*-topology.*

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Proof. (1) Let $\mu_i \in \mathcal{M}_s(X_i), i = 1, 2$ and $k > 0$. If there exists $i \in \{1, 2\}$ such that $\mu_i \notin \mathcal{M}(X_i)$ then $k\mathrm{ET}^{a,b}(\mu_1, \mu_2) = +\infty = \mathrm{ET}^{a,b}(k\mu_1, k\mu_2)$. So we only need to consider $(\mu_1, \mu_2) \in \mathcal{M}(X_1) \times \mathcal{M}(X_2)$. Let $\boldsymbol{\gamma} \in \mathrm{Opt}^{a,b}(k\mu_1, k\mu_2)$ then one has

$$\begin{aligned}\mathrm{ET}^{a,b}(k\mu_1, k\mu_2) &= a\,|k\mu_1 - \gamma_1| + a\,|k\mu_2 - \gamma_2| + b\int_{X_1} C(x_1, \gamma_{x_1})\,\mathrm{d}\gamma_1(x_1) \\ &= k\Bigg(a\,|\mu_1 - (\gamma_1/k)| + a\,|\mu_2 - (\gamma_2/k)| \\ &\quad + b\int_{X_1} C(x_1, \gamma_{x_1})\,\mathrm{d}(\gamma_1(x_1)/k)\Bigg) \\ &\geqslant k\mathrm{ET}^{a,b}(\mu_1, \mu_2).\end{aligned}$$

Similarly, we also have $k\mathrm{ET}^{a,b}(\mu_1, \mu_2) \geqslant \mathrm{ET}^{a,b}(k\mu_1, k\mu_2)$ and thus $\mathrm{ET}^{a,b}(k\mu_1, k\mu_2) = k\mathrm{ET}^{a,b}(\mu_1, \mu_2)$.

By this homogeneity property of $\mathrm{ET}^{a,b}$, to show that $\mathrm{ET}^{a,b}$ is convex, we only need to prove that

$$\mathrm{ET}^{a,b}(\mu_1, \mu_2) + \mathrm{ET}^{a,b}(\nu_1, \nu_2) \geqslant \mathrm{ET}^{a,b}(\mu_1 + \nu_1, \mu_2 + \nu_2),$$

for every $(\mu_1, \mu_2), (\nu_1, \nu_2) \in \mathcal{M}_s(X_1) \times \mathcal{M}_s(X_2)$. We will consider $(\mu_1, \mu_2), (\nu_1, \nu_2) \in \mathcal{M}(X_1) \times \mathcal{M}(X_2)$ (the other cases are trivial). Let $\boldsymbol{\gamma} \in \mathrm{Opt}^{a,b}(\mu_1, \mu_2)$ and $\overline{\boldsymbol{\gamma}} \in \mathrm{Opt}^{a,b}(\nu_1, \nu_2)$. By the convexity of $C(x_1, \cdot)$ and observe that $((d\gamma_1/d(\gamma_1 + \overline{\gamma}_1))\gamma_{x_1} + (d\overline{\gamma}_1/d(\gamma_1 + \overline{\gamma}_1))\overline{\gamma}_{x_1})_{x_1 \in X_1}$ is the disintegration of $\boldsymbol{\gamma} + \overline{\boldsymbol{\gamma}}$ with respect to $\gamma_1 + \overline{\gamma}_1$, we have that

$$\int_{X_1} C(x_1, (\gamma + \overline{\gamma})_{x_1})\,\mathrm{d}(\gamma_1 + \overline{\gamma}_1) \leqslant \int_{X_1} C(x_1, \gamma_{x_1})\,\mathrm{d}\gamma_1 + \int_{X_1} C(x_1, \overline{\gamma}_{x_1})\,\mathrm{d}\overline{\gamma}_1.$$

This yields,

$$\begin{aligned}\mathrm{ET}^{a,b}(\mu_1, \mu_2) + \mathrm{ET}^{a,b}(\nu_1, \nu_2) &\geqslant a\sum_{i=1}^{2} |(\mu_i + \nu_i) - (\gamma_i + \overline{\gamma}_i)| \\ &\quad + b\int_{X_1} C(x_1, \gamma_{x_1})\,\mathrm{d}\gamma_1(x_1) \\ &\quad + b\int_{X_1} C(x_1, \overline{\gamma}_{x_1})\,\mathrm{d}\overline{\gamma}_1(x_1) \\ &\geqslant \mathrm{ET}^{a,b}(\mu_1 + \nu_1, \mu_2 + \nu_2).\end{aligned}$$

(2) For $i = 1, 2$, let $\{\mu_i^n\} \subset \mathcal{M}(X_i)$ such that $\mu_i^n \to \mu_i \in \mathcal{M}(X_i)$ as $n \to \infty$ under the weak*-topology. Then $\{\mu_i^n\}$ is relatively compact and by Prokhorov's theorem, $\{\mu_i^n\}$ is tight and bounded. For each $n \in \mathbb{N}$ let $\boldsymbol{\gamma}^n \in \mathrm{Opt}^{a,b}(\mu_1^n, \mu_2^n)$ then $\gamma_i^n \leqslant \mu_i^n$ for $i = 1, 2$. This implies that $\{\gamma_i^n\}$ is also tight and bounded. Hence, by Prokhorov's theorem, passing to a subsequence we can assume

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

that $\lim_{n\to\infty}\gamma_i^n = \gamma_i$ for $\gamma_i \in \mathcal{M}(X_i)$. Furthermore, for every $\mu, \nu \in \mathcal{M}(X_i)$ by [**23**, theorem 6.19] we have that

$$|\mu - \nu| = \sup\left\{ \int_{X_i} f d(\mu - \nu) | f \in C_0(X_i), \|f\|_\infty \leqslant 1 \right\}.$$

From this formula, we get that $\liminf_{n\to\infty} |\mu_i^n - \gamma_i^n| \geqslant |\mu_i - \gamma_i|$ for $i = 1, 2$. By [**10**, Lemma 3.5] we get that $\int_{X_1} C(x_1, \gamma_{x_1}) d\gamma_1(x_1)$ is lower semi-continuous under the weak*-topology. Therefore,

$$\begin{aligned}\liminf_{n\to\infty} \mathrm{E}^{a,b}(\mu_1^n, \mu_2^n) &\geqslant a|\mu_1 - \gamma_1| + a|\mu_2 - \gamma_2| + b\int_{X_1} C(x_1, \gamma_{x_1})\, \mathrm{d}\gamma_1(x_1) \\ &\geqslant \mathrm{E}^{a,b}(\mu_1, \mu_2).\end{aligned}$$

This means that $\mathrm{E}^{a,b}$ is lower semi-continuous. Therefore, $\mathrm{ET}^{a,b}$ is also lower semi-continuous since $\mathcal{M}(X_1) \times \mathcal{M}(X_2)$ is closed. □

Proof of theorem 1.1. Denote by $(\mathrm{ET}^{a,b})^*$ the Fenchel conjugate of $\mathrm{ET}^{a,b}$, i.e.

$$(\mathrm{ET}^{a,b})^*(\varphi_1, \varphi_2) := \sup_{(m_1,m_2)} \left\{ \sum_{i=1}^2 \int_{X_i} \varphi_i(x_i)\, \mathrm{d}m_i(x_i) - \mathrm{E}^{a,b}(m_1, m_2) \right\},$$

where (m_1, m_2) runs over $\mathcal{M}_s(X_1) \times \mathcal{M}_s(X_2)$, for every $(\varphi_1, \varphi_2) \in C_0(X_1) \times C_0(X_2)$. Notice that the dual space of $C_0(X_i)$ is $\mathcal{M}_s(X_i)$. By lemma 3.5 we get that

$$(\mathrm{ET}^{a,b})^*(\varphi_1, \varphi_2) = \begin{cases} 0 & \text{if } (\varphi_1, \varphi_2) \in \Phi_E, \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned}\Phi_E := \Bigg\{ (\varphi_1, \varphi_2) \in C_0(X_1) \times C_0(X_2) : \sum_{i=1}^2 \int_{X_i} \varphi_i(x_i)\, \mathrm{d}m_i(x_i) \leqslant \mathrm{ET}^{a,b}(m_1, m_2) \\ \text{for every } (m_1, m_2) \in \mathcal{M}_s(X_1) \times \mathcal{M}_s(X_2) \Bigg\}.\end{aligned}$$

We now check that $\Phi_E = \Phi_J^0$. Let any $(\varphi_1, \varphi_2) \in \Phi_J^0$. Let $m_i \in \mathcal{M}_s(X_i)$, $i = 1, 2$. If $\mathrm{ET}^{a,b}(m_1, m_2) = +\infty$ then it is clear that $\sum_{i=1}^2 \int_{X_i} \varphi_i(x_i) \mathrm{d}m_i(x_i) \leqslant \mathrm{ET}^{a,b}(m_1, m_2)$. Thus, we only consider $(m_1, m_2) \in \mathcal{M}(X_1) \times \mathcal{M}(X_2)$. By lemmas 3.3 and 3.4 we get that

$$\sum_{i=1}^2 \int_{X_i} \varphi_i \mathrm{d}m_i \leqslant \sup_{(\phi_1,\phi_2)\in\Phi_I} \sum_{i=1}^2 \int_{X_i} I(\phi_i)\, \mathrm{d}m_i \leqslant \mathrm{E}^{a,b}(m_1, m_2) = \mathrm{ET}^{a,b}(m_1, m_2).$$

Therefore, $(\varphi_1, \varphi_2) \in \Phi_E$ and thus $\Phi_J^0 \subset \Phi_E$.

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Now, let any $(\varphi_1, \varphi_2) \in \Phi_E$. We will show that $(\varphi_1, \varphi_2) \in \Phi_J^0$. Denote by η the null measure on $X_1 \times X_2$. As $(\varphi_1, \varphi_2) \in \Phi_E$, for every $(m_1, m_2) \in \mathcal{M}(X_1) \times \mathcal{M}(X_2)$ one has

$$\sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i) \mathrm{d}m_i(x_i) \leqslant \mathrm{E}^{a,b}(m_1, m_2) \leqslant \mathrm{E}^{a,b}(\eta | m_1, m_2) = a(|m_1| + |m_2|).$$

For every $z \in X_1$, setting $m_1 := \delta_z$ and m_2 is the null measure on X_2, we obtain that $\varphi_1(z) \leqslant a$. Similarly, we also have $\varphi_2 \leqslant a$ on X_2.

On the other hand, for any $w \in X_1$ and $q \in \mathcal{P}(X_2)$ putting $m_1 := \delta_w, m_2 := q|_B$ and $\overline{\gamma} := \delta_w \otimes q$, where $B := \{x_2 \in X_2 | \varphi_2(x_2) \geqslant -a\}$. Then

$$\begin{aligned} \varphi_1(w) + \int_B \varphi_2 \mathrm{d}q &= \sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i) \, \mathrm{d}m_i(x_i) \leqslant \mathrm{E}^{a,b}(\overline{\gamma} | m_1, m_2) \\ &= a.q(X_2 \backslash B) + b \cdot C(w, q). \end{aligned}$$

From (1.4), if $\varphi_1(w) < -a$ then

$$J(\varphi_1(w)) + q(J(\varphi_2)) \leqslant -a + a = 0 \leqslant b \cdot C(w, q),$$

and if $\varphi_1(w) \geqslant -a$ then

$$\begin{aligned} J(\varphi_1(w)) + q(J(\varphi_2)) &= \varphi_1(w) + \int_B J(\varphi_2) \, \mathrm{d}q + \int_{X_2 \backslash B} J(\varphi_2) \, \mathrm{d}q \\ &= \varphi_1(w) + \int_B \varphi_2 \mathrm{d}q - a.q(X_2 \backslash B) \\ &\leqslant b \cdot C(w, q). \end{aligned}$$

Therefore, $(\varphi_1, \varphi_2) \in \Phi_J^0$ and hence $\Phi_E \subset \Phi_J^0$. Thus, $\Phi_E = \Phi_J^0$.

Moreover, by lemma 3.5 one has $\mathrm{ET}^{a,b}$ is convex and lower semi-continuous. Hence, applying [**11**, proposition 3.1, page 14 and proposition 4.1, page 18] we get that $(\mathrm{ET}^{a,b})^{**} = \mathrm{ET}^{a,b}$. Therefore,

$$\begin{aligned} \mathrm{ET}^{a,b}(\mu_1, \mu_2) &= \sup_{(\varphi_1, \varphi_2) \in C_0(X_1) \times C_0(X_2)} \left\{ \sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i) \, \mathrm{d}\mu_i(x_i) - (\mathrm{ET}^{a,b})^*(\varphi_1, \varphi_2) \right\} \\ &= \sup_{(\varphi_1, \varphi_2) \in \Phi_J^0} \sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i) \, \mathrm{d}\mu_i(x_i) \\ &\leqslant \sup_{(\varphi_1, \varphi_2) \in \Phi_J} \sum_{i=1}^{2} \int_{X_i} \varphi_i(x_i) \, \mathrm{d}\mu_i(x_i). \end{aligned}$$

Now, using lemmas 3.3 and 3.4 we get the result. □

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Proof of corollary 1.2. We define the cost function $C : X \times \mathcal{P}(X) \to [0, \infty]$ by

$$C(x, q) := \int_X c_1(x, y) dq(y),$$

for every $x \in X$ and $q \in \mathcal{P}(X)$. We will check that C is lower semi-continuous on $X \times \mathcal{P}(X)$. Let $(x^n, q^n) \subset X \times \mathcal{P}(X)$ such that $(x^n, q^n) \to (x^0, q^0)$ as $n \to \infty$. Then as c_1 is lower semi-continuous on $X \times X$ and non-negative, by [**10**, lemma 4.2] we get that

$$\liminf_{n\to\infty} C(x^n, q^n) = \liminf_{n\to\infty} \int_X c_1(x^n, y)\, \mathrm{d}q^n(y) \geqslant \int_X c_1(x^0, y)\, \mathrm{d}q^0(y) = C(x^0, q^0).$$

This means that C is lower semi-continuous on $X \times \mathcal{P}(X)$. Next, a one-to-one correspondence between $\boldsymbol{\gamma} \in \mathcal{M}^{\leqslant}(\mu_1, \mu_2)$ and $\hat{\boldsymbol{\gamma}} \in \Gamma(\hat{\mu}_1, \hat{\mu}_2)$ is given by

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + |1 - f_1|\mu_1 \otimes \delta_{\hat{\infty}} + \delta_{\hat{\infty}} \otimes |1 - f_2|\mu_2 + |\boldsymbol{\gamma}|\delta_{(\hat{\infty},\hat{\infty})},$$

where f_i is the Radon–Nikodym derivative of γ_i with respect to μ_i. From this and theorem 1.1 we obtain that

$$\inf_{\hat{\boldsymbol{\gamma}}\in\Gamma(\hat{\mu}_1,\hat{\mu}_2)} \int_{\hat{X}\times\hat{X}} \hat{c}_1(x, y)\, \mathrm{d}\hat{\boldsymbol{\gamma}}(x, y) = \mathrm{E}^{a,b}(\mu_1, \mu_2) = \sup_{(\varphi_1,\varphi_2)\in\Phi_J} \sum_{i=1}^{2} \int_X \varphi_i(x)\, \mathrm{d}\mu_i(x).$$

Now, for any $(\varphi_1, \varphi_2) \in \Phi_J$ we define $\hat{\varphi}_i(x) = J(\varphi_i(x))$ if $x \in X$ and $\hat{\varphi}_i(x) = 0$ if $x = \hat{\infty}$ for $i = 1, 2$. Then $\hat{\varphi}_i \in L^1(\hat{\mu}_i)$ for $i = 1, 2$. As $(\varphi_1, \varphi_2) \in \Phi_J$, for every $x, y \in X$ we have

$$J(\varphi_1(x)) + J(\varphi_2(y)) = J(\varphi_1(x)) + \delta_y(J(\varphi_2)) \leqslant b \cdot C(x, \delta_y) = b \cdot c_1(x, y).$$

Hence $\hat{\varphi}_1(x) + \hat{\varphi}_2(y) \leqslant \hat{c}_1(x, y)$ for every $x, y \in \hat{X}$. Moreover, we also have

$$\int_{\hat{X}} \hat{\varphi}_1 \mathrm{d}\hat{\mu}_1 = \int_X \hat{\varphi}_1 \mathrm{d}\mu_1 + \hat{\varphi}_1(\hat{\infty})|\mu_2| = \int_X J(\varphi_1)\, \mathrm{d}\mu_1 \geqslant \int_X \varphi_1 \mathrm{d}\mu_1.$$

Similarly, $\int_{\hat{X}} \hat{\varphi}_2 d\hat{\mu}_2 \geqslant \int_X \varphi_2 d\mu_2$. Therefore,

$$\sup_{(\varphi_1,\varphi_2)\in\Phi_J} \sum_{i=1}^{2} \int_X \varphi_i(x)\, \mathrm{d}\mu_i(x) \leqslant \sup_{\substack{(\hat{\varphi}_1,\hat{\varphi}_2)\in L^1(\hat{\mu}_1)\times L^1(\hat{\mu}_2)\\ \hat{\varphi}_1(x)+\hat{\varphi}_2(y)\leqslant\hat{c}_1(x,y)}} \sum_{i=1}^{2} \int_{\hat{X}} \hat{\varphi}_i(x)\, \mathrm{d}\hat{\mu}_i(x).$$

This implies that

$$\begin{aligned}\inf_{\hat{\boldsymbol{\gamma}}\in\Gamma(\hat{\mu}_1,\hat{\mu}_2)} \int_{\hat{X}\times\hat{X}} \hat{c}_1(x, y)\, \mathrm{d}\hat{\boldsymbol{\gamma}}(x, y) &\leqslant \sup_{\substack{(\hat{\varphi}_1,\hat{\varphi}_2)\in L^1(\hat{\mu}_1)\times L^1(\hat{\mu}_2)\\ \hat{\varphi}_1(x)+\hat{\varphi}_2(y)\leqslant\hat{c}_1(x,y)}} \sum_{i=1}^{2} \int_{\hat{X}} \hat{\varphi}_i(x)\, \mathrm{d}\hat{\mu}_i(x)\\ &\leqslant \inf_{\hat{\boldsymbol{\gamma}}\in\Gamma(\hat{\mu}_1,\hat{\mu}_2)} \int_{\hat{X}\times\hat{X}} \hat{c}_1(x, y)\, \mathrm{d}\hat{\boldsymbol{\gamma}}(x, y).\end{aligned}$$

Hence, we get the result. □

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Proof of corollary 1.3. (1) Applying theorem 1.1 for $X_1 = X_2 = X$ and $C(x,q) = \int_X c(x,y)dq(y)$, where $c(x,y) = (b \cdot d(x,y))^p$ for every $x, y \in X$ then we get the result.

(2) We use the techniques of the proof of [25, theorem 1.14] to prove (2). For every $(\psi, \varphi) \in \Phi_W$, we define $\varphi^d(x) := \inf_{y \in X}[b \cdot d(x,y) - \varphi(y)]$ for every $x \in X$. Then φ^d is b-Lipschitz function and $\varphi^d(x) \in [-a, a]$ for every $x \in X$. Therefore, $\varphi^d \in \mathbb{F}$. Now we define $\varphi^{dd}(y) := \inf_{x \in X}[b \cdot d(x,y) - \varphi^d(x)]$ for every $y \in X$. Then φ^{dd} is b-Lipschitz and

$$\varphi^d(x) + \varphi^{dd}(y) \leqslant b \cdot d(x,y), \text{ for every } x, y \in X.$$

As $-a \leqslant \varphi^d(x) \leqslant a$ we also get that $-a \leqslant \varphi^{dd}(y) \leqslant a$ for every $y \in X$. Therefore we have $\varphi^{dd} \in \mathbb{F}$ and $(\varphi^d, \varphi^{dd}) \in \Phi_W$.

On the other hand, as $\psi(x) + \varphi(y) \leqslant b \cdot d(x,y)$ for every $x, y \in X$ we get that

$$\psi(x) \leqslant \inf_{y \in X}[b \cdot d(x,y) - \varphi(y)] = \varphi^d(x) \text{ for every } x \in X.$$

Similarly, from the definitions of φ^{dd} we also have $\varphi^{dd}(y) \geqslant \varphi(y)$ for every $y \in Y$. Hence

$$\int_X I(\psi)\, \mathrm{d}\mu + \int_X I(\varphi)\, \mathrm{d}\nu \leqslant \int_X I\left(\varphi^d\right) \mathrm{d}\mu + \int_X I\left(\varphi^{dd}\right) \mathrm{d}\nu.$$

Therefore,

$$\sup_{(\psi,\varphi)\in\Phi_W} \left\{ \int_X I(\psi)\, \mathrm{d}\mu + \int_X I(\varphi)\, \mathrm{d}\nu \right\} \leqslant \sup_{\varphi \in C_b(X)} \left\{ \int_X I\left(\varphi^d\right) \mathrm{d}\mu + \int_X I\left(\varphi^{dd}\right) \mathrm{d}\nu \right\}.$$

As φ^d is b-Lipschitz we get

$$-\varphi^d(x) \leqslant \inf_{y \in X}[b \cdot d(x,y) - \varphi^d(y)].$$

On the other hand, $\inf_{y \in X}[b \cdot d(x,y) - \varphi^d(y)] \leqslant -\varphi^d(x)$. Hence

$$\varphi^{dd}(x) = \inf_{y \in X}[b \cdot d(x,y) - \varphi^d(y)] = -\varphi^d(x).$$

Thus

$$\sup_{(\psi,\varphi)\in\Phi_W} \left\{ \int_X I(\psi)\, \mathrm{d}\mu + \int_X I(\varphi)\, \mathrm{d}\nu \right\} \leqslant \sup_{\varphi \in C_b(X)} \left\{ \int_X I\left(\varphi^d\right) \mathrm{d}\mu + \int_X I\left(\varphi^{dd}\right) \mathrm{d}\nu \right\}$$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

$$= \sup_{\varphi \in C_b(X)} \left\{ \int_X I\left(\varphi^d\right) \mathrm{d}\mu + \int_X I\left(-\varphi^d\right) \mathrm{d}\nu \right\}$$
$$\leqslant \sup_{\varphi \in \mathbb{F}} \left\{ \int_X I\left(\varphi\right) \mathrm{d}\mu + \int_X I\left(-\varphi\right) \mathrm{d}\nu \right\}$$
$$\leqslant \sup_{(\psi,\varphi)\in\Phi_W} \left\{ \int_X I\left(\psi\right) \mathrm{d}\mu + \int_X I\left(\varphi\right) \mathrm{d}\nu \right\}.$$

So we must have equality everywhere and get the result. □

REMARK 3.6. (1) Corollary 1.3 (2) has been proved in [**21**, theorem 2] for the case $a = b = 1$ and $X = \mathbb{R}^n$ by a different method.

(2) [**14**, **15**] Let (X, d) be a Polish metric space. Let $\mathcal{M}^0(X)$ be the set of all $\mu \in \mathcal{M}_s(X)$ such that $\mu(X) = 0$. For every $\mu \in \mathcal{M}^0(X)$, we denote by Ψ_μ the set of all non-negative measures $\boldsymbol{\gamma} \in \mathcal{M}(X \times X)$ such that $\boldsymbol{\gamma}(X \times A) - \boldsymbol{\gamma}(A \times X) = \mu(A)$ for every Borel $A \subset X$. Then we define for every $\mu \in \mathcal{M}^0(X)$,

$$\|\mu\|_d^0 := \inf_{\boldsymbol{\gamma}\in\Psi_\mu} \left\{ \int_{X\times X} d(x,y) d\boldsymbol{\gamma}(x,y) \right\}.$$

Now, on the vector space $\mathcal{M}_s(X)$ we define an extension Kantorovich–Rubinstein norm as following

$$\|\mu\|_d := \inf_{\nu\in\mathcal{M}^0(X)} \left\{ \|\nu\|_d^0 + |\mu - \nu|(X) \right\}, \text{ for every } \mu \in \mathcal{M}_s(X).$$

Then from [**14**, theorem 0] (when X is compact) or [**15**, theorem 1] (when X is a general Polish metric space), applying Hahn–Banach theorem we get that

$$\|\mu\|_d = \sup \left\{ \int_X f d(\mu - \nu) : f \in \mathbb{F} \right\},$$

where $\mathbb{F} := \left\{ f \in C_b(X), \|f\|_\infty \leqslant 1, \|f\|_{Lip} \leqslant 1 \right\}$. We thank Benedetto Piccoli and Francesco Rossi for pointing [**14**] out to us, and we have found [**15**] after that.

Using corollary 1.3 (2) we get another proof of [**22**, lemma 5].

COROLLARY 3.7. *Let X be a locally compact, Polish metric space. For every $\mu, \nu, \eta \in \mathcal{M}(X)$ we have*

$$\widetilde{W}_1^{a,b}(\mu + \eta, \nu + \eta) = \widetilde{W}_1^{a,b}(\mu, \nu).$$

4. Barycenter problem and an its dual problem

Let (X, d) be a locally compact, Polish metric space. For every integer $k \geqslant 2$, we consider k measures $\mu_1, \mu_2, \ldots, \mu_k$ in $\mathcal{M}(X)$ such that $\mathrm{supp}(\mu_i)$ is a compact subset

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

of X for every $i \in \{1, \ldots, k\}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be positive real numbers such that $\sum_{i=1}^k \lambda_i = 1$ and let $K = \bigcup_{i=1}^k \text{supp}(\mu_i)$, we consider the following problem

$$(B) \inf_{\text{supp}(\mu) \subset K} \sum_{i=1}^k \lambda_i \widetilde{W}_2^{a,b} (\mu_i, \mu)^2 .$$

REMARK 4.1. Let $X = \mathbb{R}^d$. For every $m > 0, a, b \geqslant 0$ and $\mu_1, \mu_2 \in \mathcal{M}(X)$ we define

$$\widetilde{W}_{2,m}^{a,b}(\mu_1, \mu_2) := \inf_{\gamma_i \in \mathcal{M}_2(X), \gamma_i \leqslant \mu_i, |\boldsymbol{\gamma}| = m} a \sum_{i=1}^2 |\mu_i - \gamma_i| + b \int_{X \times X} |x - y|^2 \mathrm{d}\boldsymbol{\gamma}(x, y).$$

In [**16**] Kitagawa and Pass introduced and investigated the following partial barycenter problem:

$$\inf_{\mu \in \mathcal{M}(X), |\mu| = m} \sum_{i=1}^k \widetilde{W}_{2,m}^{0,1}(\mu_i, \mu)^2.$$

The methods there are different from us as they study their partial barycenters via multi-marginal optimal transports while we use duality formulations for our barycenter problems in generalized Wasserstein spaces.

THEOREM 4.2. *Problem* (B) *has solutions.*

Proof. For every $\mu \in \mathcal{M}(X)$ such that $\text{supp}(\mu) \subset K$, let $J(\mu) = \sum_{i=1}^k \lambda_i \widetilde{W}_2^{a,b}(\mu_i, \mu)^2$. Let $\{\mu^n\}_{n \in \mathbb{N}}$ be a minimizing sequence of (B). If there exists n_0 such that $\text{supp}(\mu^{n_0}) = X$ then $X = K$, and thus X is compact. Hence, $\{\mu^n\}_{n \in \mathbb{N}}$ is tight. Otherwise, for every $n \in \mathbb{N}$, let $x \notin \text{supp}(\mu^n)$ then there exists an open neighborhood U_x of x such that $\mu^n(U_x) = 0$. Since X is separable and $\{U_x\}_{x \in X \setminus \text{supp}(\mu^n)}$ is an open cover of $X \setminus \text{supp}(\mu^n)$, applying Lindelöf theorem there is a countable subcover $\{U_{x_i}\}_i$. Therefore, $\mu^n(X \setminus \text{supp}(\mu^n)) = 0$. Moreover, $\text{supp}(\mu^n) \subset K$ for every $n \in \mathbb{N}$. Thus, for every $n \in N$, $\mu^n(X \setminus K) = 0$. It implies that $\{\mu^n\}_{n \in \mathbb{N}}$ is tight.

We now prove that $\{\mu^n\}_{n \in \mathbb{N}}$ is bounded. For every $n \in \mathbb{N}$ and every $i \in \{1, 2, \ldots, k\}$, using corollary 1.3 (1) we get that

$$\widetilde{W}_2^{a,b} (\mu^n, \mu_i)^2 = \sup \left\{ \int_X \varphi_1(x) \mathrm{d}\mu^n(x) + \int_X \varphi_2(x) \mathrm{d}\mu_i(x) | \, (\varphi_1, \varphi_2) \in \Phi_W \right\},$$

We set $\varphi_1(x) = a, \varphi_2(x) = -a$ for every $x \in X$ then

$$\lambda_i . \widetilde{W}_2^{a,b} (\mu^n, \mu_i)^2 \geqslant \lambda_i a \mu^n(X) - \lambda_i a \mu_i(X)$$

This yields,

$$|\mu^n| \leqslant \frac{1}{a} J(\mu^n) + \sum_i^k \lambda_i |\mu_i|, \text{ for every } n \in \mathbb{N}.$$

As $\mu_i \in \mathcal{M}(X)$ for every $i \in \{1, 2, \ldots, k\}$ and $J(\mu^n)$ is bounded, we obtain that $\{\mu^n\}_{n \in \mathbb{N}}$ is bounded. Therefore, applying Prokhorov's theorem, passing to a subsequence we can assume that $\mu^n \to \mu$ as $n \to \infty$ in the weak*-topology for some $\mu \in \mathcal{M}(X)$.

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

We now show that $\operatorname{supp}(\mu) \subset K$. As $X\backslash K$ is an open set, applying [**19**, theorem 6.1] we get that

$$0 = \liminf_{n\to\infty} \mu^n(X\backslash K) \geqslant \mu(X\backslash K).$$

Therefore, $X\backslash K \subset X\backslash\operatorname{supp}(\mu)$. Hence, $\operatorname{supp}(\mu) \subset K$.

Next, we will check that $\widetilde{W}_2^{a,b}(\mu^n, \mu) \to 0$ as $n \to \infty$. If $|\mu| = 0$ then we are done. If $|\mu| > 0$ then there exists $N > 0$ such that $|\mu^n| > 0$ for all $n \geqslant N$. For each $n \geqslant N$, we define $\nu^n := |\mu|\mu^n/|\mu^n|$ then $|\nu^n| = |\mu|$. Therefore,

$$\widetilde{W}_2^{a,b}(\mu^n,\mu)^2 \leqslant a|\mu^n - \nu^n| + b^2 W_2^2(\nu^n,\mu) = a\,||\mu^n| - |\mu|| + b^2 W_2^2(\nu^n,\mu).$$

Moreover, since $\mu^n \to \mu$ as $n \to \infty$ one has $\nu^n \to \mu$ as $n \to \infty$. Observe that ν^n and μ are concentrated in compact set K, applying [**26**, definition 6.8 and theorem 6.9] we obtain that $\lim_{n\to\infty} W_2(\nu^n, \mu) = 0$. This yields,

$$\limsup_{n\to\infty} \widetilde{W}_2^{a,b}(\mu^n,\mu)^2 \leqslant a \lim_{n\to\infty} ||\mu^n| - |\mu|| + \lim_{n\to\infty} b^2 W_2^2(\mu^n,\mu) = 0.$$

Notice that $\liminf_{n\to\infty} \widetilde{W}_2^{a,b}(\mu^n,\mu) \geqslant 0$. Therefore, $\lim_{n\to\infty} \widetilde{W}_2^{a,b}(\mu^n,\mu) = 0$. This implies that $\lim_{n\to\infty} J(\mu^n) = J(\mu)$. Hence, we get the result. □

Definition 4.3. Let X be a locally compact, Polish metric space. For every integer $k \geqslant 2$, let $\mu_1, \ldots, \mu_k \in \mathcal{M}(X)$ such that $\operatorname{supp}(\mu_i)$ is compact, for every $i \in \{1, \ldots, k\}$. Let $\lambda_1, \ldots, \lambda_k > 0$ such that $\sum_{i=1}^k \lambda_i = 1$. We say that $\mu \in \mathcal{M}(X)$ is a generalized Wasserstein barycenter of $(\mu_1, \ldots, \mu_k)$ with weights $(\lambda_1, \ldots, \lambda_k)$ if μ is a solution of (B). We denote by $BC((\mu_i, \lambda_i)_{1\leqslant i\leqslant k})$ the set of all generalized Wasserstein barycenters of $(\mu_1, \ldots, \mu_k)$ with weights $(\lambda_1, \ldots, \lambda_k)$.

In general, barycenters in a generalized Wasserstein space are not unique.

Example 4.4. Let $X = \mathbb{R}, a = b = 1$ and $\lambda_1 = \lambda_2 = 1/2$. For every $x \geqslant 0$ let $\mu_1 = \delta_x$ and $\mu_2 = 3\delta_x$. Then we have $\{\mu \in \mathcal{M}(\mathbb{R}) | \operatorname{supp}(\mu) \subset \{x\}\} = \{q\delta_x | q \geqslant 0\}$. For every $q \geqslant 0$, let $(\widetilde{\mu}_1, \widetilde{\mu}_2)$ be an optimal for $\widetilde{W}_2^{1,1}(\delta_x, q\delta_x)$. Since $|\widetilde{\mu}_1| = |\widetilde{\mu}_2|, \widetilde{\mu}_1 \leqslant \delta_x, \widetilde{\mu}_2 \leqslant q\delta_x$, we must have $\widetilde{\mu}_1 = \widetilde{\mu}_2 = r\delta_x$ where $0 \leqslant r \leqslant \min\{q, 1\}$. Hence, we get that

$$\widetilde{W}_2^{1,1}(\delta_x, q\delta_x)^2 = \min\{q + 1 - 2r | 0 \leqslant r \leqslant \min\{q,1\}\}.$$

Similarly, we also get that

$$\widetilde{W}_2^{1,1}(3\delta_x, q\delta_x)^2 = \min\{q + 3 - 2s | 0 \leqslant s \leqslant \min\{q,3\}\}.$$

It is easy to check that

$$\begin{aligned}&\lambda_1.\min\{q + 1 - 2r | 0 \leqslant r \leqslant \min\{q,1\}\}\\ &\quad + \lambda_2.\min\{q + 3 - 2s | 0 \leqslant s \leqslant \min\{q,3\}\} = 1,\end{aligned}$$

and the minimum is attained when $q \in [1, 3]$. Therefore, $BC((\mu_1, \lambda_1), (\mu_2, \lambda_2)) = \{q\delta_x | q \in [1, 3]\}$.

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

We now prove the consistency of barycenters in generalized Wasserstein spaces which has been shown in [**5**, theorem 3.1] for the Wasserstein setting.

THEOREM 4.5. *Let (X, d) be a locally compact, Polish metric space. For every integer $k \geqslant 2$, let $\{\mu_i^n\} \subset \mathcal{M}(X)$ be sequences converging in the generalized Wasserstein distance to compactly supported measure $\mu_i \in \mathcal{M}(X)$ for every $i \in \{1, \ldots, k\}$. Let $K = \bigcup_{i=1}^k supp(\mu_i)$ and let sequences $\lambda_1^n, \ldots, \lambda_k^n > 0$ such that $\sum_{i=1}^k \lambda_i^n = 1$ for every $n \in \mathbb{N}$ and λ_i^n converges to $\lambda_i > 0$ for $i = 1, \ldots, k$. For each $n \in \mathbb{N}$, suppose that $supp(\mu_i^n) \subset K$ for every $i \in \{1, \ldots, k\}$. Then $BC((\mu_i^n, \lambda_i^n)_{1 \leqslant i \leqslant k})$ is a non-empty set for every $n \in \mathbb{N}$. Moreover, for every $n \in \mathbb{N}$, let $\mu_B^n \in BC((\mu_i^n, \lambda_i^n)_{1 \leqslant i \leqslant k})$ then the sequence $\{\mu_B^n\}$ is precompact in $(\mathcal{M}(X), \widetilde{W}_2^{a,b})$ and any its limit point is a generalized Wasserstein barycenter of $(\mu_1, \ldots, \mu_k)$ with weights $(\lambda_1, \ldots, \lambda_k)$.*

Proof. Since $\text{supp}(\mu_i^n) \subset K$ and K is compact, one has $\text{supp}(\mu_i^n)$ is compact for every $n \in \mathbb{N}$ and every $i \in \{1, \ldots, k\}$. Therefore, $BC((\mu_i^n, \lambda_i^n)_{1 \leqslant i \leqslant k})$ is a non-empty set for every $n \in \mathbb{N}$, this follows from theorem 4.2.

We now prove the second part. Since $\mu_B^n \in BC((\mu_i^n, \lambda_i^n)_{1 \leqslant i \leqslant k})$, we get that $\text{supp}(\mu_B^n) \subset \bigcup_{i=1}^k \text{supp}(\mu_i^n) \subset K$, for every $n \in \mathbb{N}$. Then $\mu_B^n(X \backslash K) = 0$ for every $n \in \mathbb{N}$. Therefore, $\{\mu_B^n\}$ is tight. Let $\mu_B \in BC((\mu_i, \lambda_i)_{1 \leqslant i \leqslant k})$. Since $\widetilde{W}_2^{a,b}(\mu_i^n, \mu_i) \to 0$ as $n \to \infty$ for every $i \in \{1, \ldots, k\}$ we get that

$$\lim_{n \to \infty} \widetilde{W}_2^{a,b}(\mu_B, \mu_i^n) = \widetilde{W}_2^{a,b}(\mu_B, \mu_i) < \infty$$

Therefore, $\{\widetilde{W}_2^{a,b}(\mu_B, \mu_i^n)\}_n$ is bounded for every $i \in \{1, \ldots, k\}$. Moreover,

$$\sum_{i=1}^k \lambda_i^n \widetilde{W}_2^{a,b}(\mu_B^n, \mu_i^n)^2 \leqslant \sum_{i=1}^k \lambda_i^n \widetilde{W}_2^{a,b}(\mu_B, \mu_i^n)^2, \text{ for every } n \in \mathbb{N}. \tag{4.1}$$

This yields, $\widetilde{W}_2^{a,b}(\mu_B^n, \mu_i^n)$ is bounded for every $i \in \{1, \ldots, k\}$. As $\mu_i^n \to \mu_i$ as $n \to \infty$ in the weak*-topology, applying [**19**, theorem 6.1] we get that $\lim_{n \to \infty} \mu_i^n(X) = \mu_i(X) < \infty$. Thus, $\{\mu_i^n\}$ is bounded for every $i \in \{1, \ldots, k\}$. Therefore, using corollary 1.3 (1) and by the same arguments as in the proof of theorem 4.2 we obtain that $\{\mu_B^n\}$ is bounded. Hence, applying Prokhorov's theorem, passing to a subsequence we can assume that $\mu_B^n \to \widehat{\mu}_B$ as $n \to \infty$ in the weak*-topology for some $\widehat{\mu}_B \in \mathcal{M}(X)$. Observe that, from $\mu_B^n(X \backslash K) = 0$ for every $n \in \mathbb{N}$ and $X \backslash K$ is an open set, we get that $\widehat{\mu}_B(X \backslash K) = 0$ and thus $\text{supp}(\widehat{\mu}_B) \subset K$. By the same arguments in the proof of theorem 4.2 we also have $\widetilde{W}_2^{a,b}(\mu_B^n, \widehat{\mu}_B) \to 0$ as $n \to \infty$. This implies that the sequence $\{\mu_B^n\}$ is precompact in generalized Wasserstein topology and we also get that

$$\lim_{n \to \infty} \widetilde{W}_2^{a,b}(\mu_B^n, \mu_i^n) = \widetilde{W}_2^{a,b}(\widehat{\mu}_B, \mu_i), \text{ for every } i \in \{1, \ldots, k\}.$$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Hence, since (4.1) we get that

$$\begin{aligned}\sum_{i=1}^{k}\lambda_i\widetilde{W}_2^{a,b}\left(\widehat{\mu}_B,\mu_i\right)^2 &= \lim_{n\to\infty}\sum_{i=1}^{k}\lambda_i^n\widetilde{W}_2^{a,b}\left(\mu_B^n,\mu_i^n\right)^2\\ &\leqslant \lim_{n\to\infty}\sum_{i=1}^{k}\lambda_i^n\widetilde{W}_2^{a,b}\left(\mu_B,\mu_i^n\right)^2\\ &= \sum_{i=1}^{k}\lambda_i\widetilde{W}_2^{a,b}\left(\mu_B,\mu_i\right)^2.\end{aligned}$$

Therefore, $\widehat{\mu}_B \in BC((\mu_i,\lambda_i)_{1\leqslant i\leqslant k})$. □

Next, we will study the a dual problem of problem (B). For every $\lambda>0$ and every function $f\in C_b(K)$ such that $f(x)\leqslant\lambda a$ for every $x\in K$ where $K=\bigcup_{i=1}^{k}\text{supp}(\mu_i)$, we define $S_\lambda f(x) := \inf_{y\in K}\left\{\lambda b^2d^2(x,y)-f(y)\right\}$ and $\overline{S}_\lambda f(x) := \min\left\{S_\lambda f(x),\lambda a\right\}$. For every integer $k\geqslant 2$ and for each $i\in\{1,2,\ldots,k\}$ we define function $H_i : C_b(K)\to\overline{\mathbb{R}}$ by

$$H_i(f) := \begin{cases} -\displaystyle\int_K \overline{S}_{\lambda_i}f(x)d\mu_i(x) & \text{if } f\in F_{\lambda_i}\\ +\infty \quad \text{otherwise,}\end{cases}$$

where $F_{\lambda_i} := \{f\in C_b(K)|f(x)\leqslant\lambda_i a,\forall x\in K\}$. Then H_i is convex on F_{λ_i}.

We denote by $\mathcal{M}_s(K)$ (resp. $\mathcal{M}_c(K)$) the space of signed (resp. non-negative) Radon measures μ with a finite mass on X such that μ is concentrated on K, i.e. $\mu(X\backslash K)=0$. Then $\mathcal{M}_s(K)$ is the dual space of $C_b(K)$, since K is compact. For every $\mu\in\mathcal{M}_s(K)$, the Legendre–Fenchel transform of H_i is

$$\begin{aligned}H_i^*(\mu) &= \sup\left\{\int_K f(x)\,\mathrm{d}\mu(x) - H_i(f)|f\in C_b(K)\right\}\\ &= \sup\left\{\int_K f(x)\,\mathrm{d}\mu(x) - H_i(f)|f\in F_{\lambda_i}\right\}\\ &= \sup\left\{\int_K f(x)d\mu(x) + \int_K \overline{S}_{\lambda_i}f(x)\,\mathrm{d}\mu_i(x)|f\in F_{\lambda_i}\right\}.\end{aligned}$$

We consider the following problem

$$(B^*)\ \sup\left\{\sum_{i=1}^{k}\int_K \overline{S}_{\lambda_i}f_i(x)\,\mathrm{d}\mu_i(x)|f_i\in F_{\lambda_i},\sum_{i=1}^{k}f_i=0\right\}.$$

LEMMA 4.6. *Let X be a locally compact, Polish metric space then* $\inf(B)\geqslant\sup(B^*)$.

Proof. For $i=1,2,\ldots,k$ let any $f_i\in F_{\lambda_i}$ such that $\sum_{i=1}^{k}f_i=0$. Then $\overline{S}_{\lambda_i}f_i(x)+f_i(y)\leqslant\lambda_i b^2d^2(x,y)$ for every $x,y\in K$ and every $i\in\{1,2,\ldots,k\}$. For every $\mu\in$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

$\mathcal{M}_c(K)$, let $\boldsymbol{\gamma}^i \in M^{\leqslant}(\mu,\mu_i)$ be an optimal plan for $\widetilde{W}_2^{a,b}(\mu,\mu_i)$. Since μ_i is concentrated on K for every $i = 1,\dots,k$, we get that

$$\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 = a\left(\mu - \pi^1_\sharp\boldsymbol{\gamma}^i\right)(K) + a\left(\mu_i - \pi^2_\sharp\boldsymbol{\gamma}^i\right)(K) + b^2\int_{K\times K} d^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}^i(x,y).$$

As $\boldsymbol{\gamma}^i \in M^{\leqslant}(\mu,\mu_i)$, by Radon–Nikodym theorem there exist measurable functions $\varphi_1,\varphi_2 : K \to [0,+\infty)$ such that $\pi^1_\sharp\boldsymbol{\gamma}^i = \varphi_1\mu$, $\pi^2_\sharp\boldsymbol{\gamma}^i = \varphi_2\mu_i$ and $\varphi_1 \leqslant 1$ μ-a.e, $\varphi_2 \leqslant 1$ μ_i-a.e. Therefore, we get that

$$\begin{aligned}
\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 &= a\int_K (1-\varphi_1)\,\mathrm{d}\mu + a\int_K(1-\varphi_2)\,\mathrm{d}\mu_i + b^2\int_{K\times K} d^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}^i(x,y)\\
&\geqslant a\int_K (1-\varphi_1)\,\mathrm{d}\mu + a\int_K(1-\varphi_2)\,\mathrm{d}\mu_i\\
&\quad + \frac{1}{\lambda_i}\int_{K\times K}\left[f_i(x) + \overline{S}_{\lambda_i}f_i(y)\right]\mathrm{d}\boldsymbol{\gamma}^i(x,y)\\
&= \int_K\left[a(1-\varphi_1) + \frac{1}{\lambda_i}f_i.\varphi_1\right]\mathrm{d}\mu + \int_K\left[a(1-\varphi_2) + \frac{1}{\lambda_i}\overline{S}_{\lambda_i}f_i.\varphi_2\right]\mathrm{d}\mu_i.
\end{aligned}$$

Moreover, $\varphi_1(x),\varphi_2(x) \geqslant 0$ for every $x \in X$ and $\varphi_1 \leqslant 1$ μ-a.e, $\varphi_2 \leqslant 1$ μ_i-a.e, $f_i(x)/\lambda_i \leqslant a$, $\overline{S}_{\lambda_i}f_i(x)/\lambda_i \leqslant a$ for every $x \in K$. Therefore, we obtain that

$$\begin{aligned}
a(1-\varphi_1(x)) + (f_i(x)/\lambda_i).\varphi_i(x) &\geqslant f_i(x)/\lambda_i,\ \mu-\text{a.e},\\
a(1-\varphi_2(x)) + \left(\overline{S}_{\lambda_i}f_i(x)/\lambda_i\right).\varphi_2(x) &\geqslant \overline{S}_{\lambda_i}f_i(x)/\lambda_i,\ \mu_i-\text{a.e.}
\end{aligned}$$

Hence, for every $i \in \{1,2,\dots,k\}$, we get that

$$\lambda_i\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 \geqslant \int_K f_i(x)\,\mathrm{d}\mu(x) + \int_K \overline{S}_{\lambda_i}f_i(x)\,\mathrm{d}\mu_i(x). \tag{4.2}$$

Thus,

$$\begin{aligned}
\sum_{i=1}^k \lambda_i\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 &\geqslant \sum_{i=1}^k\int_K f_i(x)\,\mathrm{d}\mu(x) + \sum_{i=1}^k\int_K \overline{S}_{\lambda_i}f_i(x)\,\mathrm{d}\mu_i(x)\\
&= \sum_{i=1}^k\int_K \overline{S}_{\lambda_i}f_i(x)\,\mathrm{d}\mu_i(x).
\end{aligned}$$

This yields,

$$\inf\left\{\sum_{i=1}^k \lambda_i\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 \mid \operatorname{supp}(\mu)\subset K\right\} \geqslant \sum_{i=1}^k\int_K \overline{S}_{\lambda_i}f_i(x)\,\mathrm{d}\mu_i(x).$$

Hence, we get the result. □

Lemma 4.7. *Let X be a locally compact, Polish metric space. Then for every $i \in \{1,2,\dots,k\}$ we have $H_i^*(\mu) = \lambda_i\widetilde{W}_2^{a,b}(\mu,\mu_i)^2$ if $\mu \in \mathcal{M}_c(K)$ and $+\infty$ otherwise.*

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

Proof. If $\mu \in \mathcal{M}_s(K)\backslash\mathcal{M}_c(K)$ then there exists $g \in C_b(K), g \leqslant 0$ such that $\int_K g(x)d\mu(x) > 0$. For every $t \in \mathbb{R}, t \geqslant 0$ let $f = t.g$ then $f \in F_{\lambda_i}$ and $\overline{S}_{\lambda_i}(tf(x)) \geqslant 0$ for every $x \in K$. Therefore, $H_i^*(\mu) \geqslant \sup_{t\geqslant 0}\int_K f d\mu = +\infty$.

We now consider $\mu \in \mathcal{M}_c(K)$. Since (4.2), it is clear that $\lambda_i\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 \geqslant H_i^*(\mu)$. So we need to prove that $\lambda_i\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 \leqslant H_i^*(\mu)$. We define

$$\Phi_K := \big\{(\varphi_1,\varphi_2) \in C_b(K)\times C_b(K) : \varphi_1(x)+\varphi_2(y) \leqslant b^2d^2(x,y), \varphi_1(x),\varphi_2(y) \geqslant -a, \text{ for every } x,y \in K\big\}.$$

Let any $(\varphi_1,\varphi_2) \in \Phi_K$ then $\lambda_i\varphi_1(x)+\lambda_i\varphi_2(y) \leqslant \lambda_i b^2 d^2(x,y)$ for every $x,y \in K$ and every $i = 1,\ldots,k$. Therefore, $\lambda_i\varphi_2(y) \leqslant S_{\lambda_i}(\lambda_i\varphi_1(y))$ for every $y \in K$. Observe that $\varphi_2(y) \in [-a,a]$ for every $y \in K$, we get that $\lambda_i\varphi_2(y) \leqslant \overline{S}_{\lambda_i}(\lambda_i\varphi_1(y))$ for every $y \in K$. As $\lambda_i\varphi_1(x) \leqslant \lambda_i a$ for every $x \in K$, one has $\lambda_i\varphi_1 \in F_{\lambda_i}$. Hence, we obtain that

$$\begin{aligned}
&\int_K \lambda_i\varphi_1(x)\,\mathrm{d}\mu(x) + \int_K \lambda_i\varphi_2(y)\,\mathrm{d}\mu_i(y) \\
&\leqslant \int_K \lambda_i\varphi_1(x)\,\mathrm{d}\mu(x) + \int_K \overline{S}_{\lambda_i}\left(\lambda_i\varphi_1(y)\right)\mathrm{d}\mu_i(y) \\
&\leqslant H_i^*(\mu).
\end{aligned}$$

Applying corollary 1.3 (1) we get that

$$\widetilde{W}_2^{a,b}\left(\mu,\mu_i\right)^2 = \sup_{(\varphi_1,\varphi_2)\in\Phi_K}\left\{\int_K \varphi_1(x)\,\mathrm{d}\mu(x) + \int_K \varphi_2(y)\,\mathrm{d}\mu_i(y)\right\} \leqslant \frac{1}{\lambda_i}H_i^*(\mu).$$

Hence, $\lambda_i\widetilde{W}_2^{a,b}(\mu,\mu_i)^2 \leqslant H_i^*(\mu)$ for every $\mu \in \mathcal{M}_c(K)$ and every $i \in \{1,2,\ldots,k\}$. □

Let $F := \{f \in C_b(K)|f(x) \leqslant a$ for every $x \in K\}$. We define $H : C_b(K) \to \overline{\mathbb{R}}$ by $H(f) = \inf\left\{\sum_{i=1}^k H_i(f_i)|f_i \in F_{\lambda_i}, \sum_{i=1}^k f_i = f\right\}$ if $f \in F$ and $+\infty$ otherwise.

LEMMA 4.8. *H is convex on F and $H^*(\mu) = \sum_{i=1}^k H_i^*(\mu)$ for every $\mu \in M_s(K)$.*

Proof. For every $g_1, g_2 \in F$ and every $t \in [0,1]$ we will check that $H(tg_1 + (1-t)g_2) \leqslant tH(g_1) + (1-t)H(g_2)$. Let any $\overline{f}_i, \widehat{f}_i \in F_{\lambda_i}$ such that $\sum_{i=1}^k \overline{f}_i = g_1$ and $\sum_{i=1}^k \widehat{f}_i = g_2$ then $t\overline{f}_i + (1-t)\widehat{f}_i \in F_{\lambda_i}$ and $\sum_{i=1}^k[t\overline{f}_i + (1-t)\widehat{f}_i] = tg_1 + (1-t)g_2$. As H_i is convex on F_{λ_i} for every $i = 1,\ldots,k$, we get that

$$\begin{aligned}
t\sum_{i=1}^k H_i\left(\overline{f}_i\right) + (1-t)\sum_{i=1}^k H_i\left(\widehat{f}_i\right) &= \sum_{i=1}^k\left[tH_i\left(\overline{f}_i\right) + (1-t)H_i\left(\widehat{f}_i\right)\right] \\
&\geqslant \sum_{i=1}^k H_i\left(t\overline{f}_i + (1-t)\widehat{f}_i\right) \\
&\geqslant H\left(tg_1 + (1-t)g_2\right).
\end{aligned}$$

Therefore, $H(tg_1 + (1-t)g_2) \leqslant tH(g_1) + (1-t)H(g_2)$. Hence, H is convex on F.

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

We now show that $H^*(\mu) = \sum_{i=1}^k H_i^*(\mu)$ for every $\mu \in M_s(K)$. For every $\mu \in \mathcal{M}_s(K)$, by definition of the Legendre–Fenchel one has

$$\begin{aligned}H^*(\mu) &= \sup_{f \in C_b(K)} \left\{ \int_K f \mathrm{d}\mu - H(f) \right\} \\ &= \sup_{f \in F} \left\{ \int_K f \mathrm{d}\mu - H(f) \right\} \\ &= \sup_{f \in F} \left\{ \int_K f \mathrm{d}\mu - \inf \left\{ \sum_{i=1}^k H_i(f_i) | f_i \in F_{\lambda_i}, \sum_{i=1}^k f_i = f \right\} \right\} \\ &= \sup_{f \in F} \left\{ \int_K f \mathrm{d}\mu + \sup \left\{ \sum_{i=1}^k \int_K \overline{S}_{\lambda_i} f_i \mathrm{d}\mu_i | f_i \in F_{\lambda_i}, \sum_{i=1}^k f_i = f \right\} \right\}.\end{aligned}$$

For every $f_i \in F_{\lambda_i}$ let $f = \sum_{i=1}^k f_i$ then $f \in F$. Thus, for every $\mu \in \mathcal{M}_s(K)$ we get

$$\begin{aligned}&\sum_{i=1}^k \left(\int_K f_i(x) \, \mathrm{d}\mu(x) + \int_K \overline{S}_{\lambda_i} f_i(x) \, \mathrm{d}\mu_i(x) \right) \\ &\quad = \int_K f(x) \, \mathrm{d}\mu(x) + \sum_{i=1}^k \int_K \overline{S}_{\lambda_i} f_i(x) \, \mathrm{d}\mu_i(x) \\ &\quad \leqslant H^*(\mu).\end{aligned}$$

This yields,

$$\begin{aligned}\sum_{i=1}^k H_i^*(\mu) &= \sum_{i=1}^k \sup \left\{ \int_K f_i(x) \, \mathrm{d}\mu(x) + \int_K \overline{S}_{\lambda_i} f_i(x) \, \mathrm{d}\mu_i(x) | f_i \in F_{\lambda_i} \right\} \\ &= \sup \left\{ \sum_{i=1}^k \left(\int_K f_i(x) \, \mathrm{d}\mu(x) + \int_K \overline{S}_{\lambda_i} f_i(x) \, \mathrm{d}\mu_i(x) \right) | f_i \in F_{\lambda_i} \right\} \\ &\leqslant H^*(\mu).\end{aligned}$$

Conversely, for every $f \in F$ let $G := \left\{ (f_1, \ldots, f_k) | f_i \in F_{\lambda_i}, \sum_{i=1}^k f_i = f \right\}$. Then,

$$\begin{aligned}&\int_K f \mathrm{d}\mu + \sup_{(f_1, \ldots, f_k) \in G} \sum_{i=1}^k \int_k \overline{S}_{\lambda_i} f_i \mathrm{d}\mu_i \\ &\quad = \sup_{(f_1, \ldots, f_k) \in G} \left\{ \int_K f \mathrm{d}\mu + \sum_{i=1}^k \int_k \overline{S}_{\lambda_i} f_i \mathrm{d}\mu_i \right\}\end{aligned}$$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

$$= \sup_{(f_1,\dots,f_k)\in G} \left\{ \int_K \sum_{i=1}^{k} f_i \mathrm{d}\mu + \sum_{i=1}^{k} \int_k \overline{S}_{\lambda_i} f_i \mathrm{d}\mu_i \right\}$$
$$\leqslant \sum_{i=1}^{k} \sup_{f_i\in F_{\lambda_i}} \left\{ \int_K f_i \mathrm{d}\mu + \int_k \overline{S}_{\lambda_i} f_i \mathrm{d}\mu_i \right\}$$
$$= \sum_{i=1}^{k} H_i^*(\mu).$$

Hence, we get the result. □

Inspired by [**1**, proposition 2.2] we get the following theorem.

THEOREM 4.9. *Let* (X, d) *be a locally compact, Polish space then* $\inf(B) = \sup(B^*)$.

Proof. Combining lemmas 4.7 and 4.8 we obtain that

$$\inf(B) = \inf_{\mu\in\mathcal{M}_c(K)} \sum_{i=1}^{k} H_i^*(\mu) = -\left(\sum_{i=1}^{k} H_i^*\right)^*(0) = -H^{**}(0).$$

Furthermore, we also have $\sup(B^*) = -H(0)$. Thus, we only need to prove that $H^{**}(0) = H(0)$. For every $f \in F$, let $f_i \in F_{\lambda_i}$ such that $\sum_{i=1}^{k} f_i = f$. As $f_i(x) \leqslant \lambda_i a$ for every $x \in K$ and every $i = 1, \dots, k$, one has

$$S_{\lambda_i} f_i(x) = \inf_{y\in K} \left\{ \lambda_i b^2 d^2(x, y) - f_i(y) \right\} \geqslant -\lambda_i a \text{ for every } x \in K.$$

Therefore, $H_i(f_i) \leqslant \lambda_i a$. Moreover, since $\overline{S}_{\lambda_i} f(x) \leqslant \lambda_i a$ for every $x \in K$, we also have $H_i(f_i) \geqslant -\lambda_i a$. Hence H is bounded on F. Thanks to lemma 4.8, one has H is convex on F. We denote by $\mathring{F}$ the interior of F, then $\mathring{F}$ is also a convex set. Applying [**11**, lemma 2.1] we get that H is continuous in $\mathring{F}$ endowed with the supremum norm $\|\cdot\|_\infty$. Observe that $0 \in \mathring{F}$, using [**11**, propositions 3.1 and 4.1] we obtain that $H^{**}(0) = H(0)$. Hence, we get the result. □

LEMMA 4.10. *Let* (X, d) *be a Polish metric space. For every* $\lambda > 0$*, let* $f \in F_\lambda$ *then* $S_\lambda f$ *and* $(S_\lambda \circ S_\lambda)f$ *are* $2\lambda b^2 D$*-Lipschitz functions on* K*, where* $D = diam(K)$.

Proof. As K is a compact subset of X then K is bounded and thus $D = \mathrm{diam}(K) < \infty$. Let any $x_1, x_2 \in K$. For every $\varepsilon > 0$, there exists $y_0 \in K$ such that $S_\lambda f(x_2) \geqslant \lambda b^2 d^2(x_2, y_0) - f(y_0) - \varepsilon$. Moreover, it is clear that $S_\lambda f(x_1) \leqslant \lambda b^2 d^2(x_1, y_0) - f(y_0)$. Hence, we get that

$$S_\lambda f(x_1) - S_\lambda f(x_2) \leqslant \lambda b^2 \left[d^2(x_1, y_0) - d^2(x_2, y_0) \right] + \varepsilon \leqslant 2\lambda b^2 D d(x_1, x_2) + \varepsilon.$$

Similarly, $S_\lambda f(x_2) - S_\lambda f(x_1) \leqslant 2\lambda b^2 D d(x_1, x_2) + \varepsilon$. Therefore, $S_\lambda f$ is a $2\lambda b^2 D$-Lipschitz function. By the same arguments above, we also get that $(S_\lambda \circ S_\lambda)f$ is a $2\lambda b^2 D$-Lipschitz function. □

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

THEOREM 4.11. *Let (X, d) be a Polish metric space then problem (B^*) has solutions.*

Proof. Let $f^n = (f_1^n, \dots, f_k^n)$ be a maximizing sequence for (B^*). For each $i \in \{1, \dots, k-1\}$ we define $\widetilde{f_i^n} := (S_{\lambda_i} \circ S_{\lambda_i}) f_i^n$. Then $\widetilde{f_i^n}$ is bounded on K for every $i = 1, \dots, k-1$. Since $S_{\lambda_i} f_i(x) \geqslant -\lambda_i a$ for every $x \in K$ and every $i = 1, \dots, k-1$, we get $\widetilde{f_i^n}(x) = \inf_{y \in K} \left\{\lambda_i b^2 d^2(x,y) - S_{\lambda_i} f_i^n(y)\right\} \leqslant -S_{\lambda_i} f_i^n(x) \leqslant \lambda_i a$ for every $x \in K$.

Moreover, it is easy to see that $f_i^n \leqslant \widetilde{f_i^n}$ on K and $S_{\lambda_i} \widetilde{f_i^n} = S_{\lambda_i} f_i^n$ for every $i = 1, \dots, k-1$. Hence $\overline{S}_{\lambda_i} \widetilde{f_i^n} = \overline{S}_{\lambda_i} f_i^n$ for every $i = 1, \dots, k-1$. For every $n \in \mathbb{N}$, we define $\widetilde{f_k^n} := -\sum_{i=1}^{k-1} \widetilde{f_i^n}$. As $f_i^n \leqslant \widetilde{f_i^n}$ on K, one has $\widetilde{f_k^n} \leqslant -\sum_{i=1}^{k-1} f_i^n = f_k^n$. Thus, $\widetilde{f_k^n}(x) \leqslant \lambda_k a$ for every $x \in K$ and $S_{\lambda_k} \widetilde{f_k^n} \geqslant S_{\lambda_k} f_k^n$ for every $n \in \mathbb{N}$. Thus, $\overline{S}_{\lambda_k} \widetilde{f_k^n} \geqslant \overline{S}_{\lambda_k} f_k^n$ for every $n \in \mathbb{N}$. Therefore, we obtain that

$$\limsup_{n \to \infty} \sum_{i=1}^{k} \int_K \overline{S}_{\lambda_i} \widetilde{f_i^n}(x) \, \mathrm{d}\mu_i(x) \geqslant \lim_{n \to \infty} \sum_{i=1}^{k} \int_K \overline{S}_{\lambda_i} f_i^n(x) \, \mathrm{d}\mu_i(x) = \sup(B^*).$$

Using lemma 4.10 we get that $\widetilde{f_i^n}$ is a $2\lambda_i b^2 D$-Lipschitz function on K for every $i = 1, \dots, k-1$ and every $n \in \mathbb{N}$. As $\widetilde{f_k^n} := -\sum_{i=1}^{k-1} \widetilde{f_i^n}$ and $\sum_{i=1}^{k} \lambda_i = 1$, we obtain that $\widetilde{f_k^n}$ is a $2(1-\lambda_k) b^2 D$-Lipschitz function on K. Then applying Ascoli–Arzela theorem on compact set K and using a standard diagonal argument there exists a subsequence of $\widetilde{f^n} = (\widetilde{f_1^n}, \dots, \widetilde{f_k^n})$ which we still denote by $\left\{\widetilde{f^n}\right\}$ such that $\widetilde{f^n}$ converges uniformly to $\widetilde{f} = (\widetilde{f}_1, \dots, \widetilde{f}_k)$. Then $\widetilde{f}_i \in F_{\lambda_i}$ for every $i \in \{1, \dots, k\}$. As $\sum_{i=1}^{k} \widetilde{f_i^n} = 0$ for every $n \in \mathbb{N}$, we get that $\sum_{i=1}^{k} \widetilde{f}_i = 0$. This yields,

$$\sum_{i=1}^{k} \int_K \overline{S}_{\lambda_i} \widetilde{f}_i(x) \, \mathrm{d}\mu_i(x) \leqslant \sup(B^*) \leqslant \sum_{i=1}^{k} \limsup_{n \to \infty} \int_K \overline{S}_{\lambda_i} \widetilde{f_i^n}(x) \, \mathrm{d}\mu_i(x).$$

Applying Fatou lemma, we obtain that

$$\begin{aligned}
&\sum_{i=1}^{k} \int_K \overline{S}_{\lambda_i} \widetilde{f}_i(x) \, \mathrm{d}\mu_i(x) \\
&\quad \leqslant \sum_{i=1}^{k} \int_K \limsup_{n \to \infty} \overline{S}_{\lambda_i} \widetilde{f_i^n}(x) \, \mathrm{d}\mu_i(x) \\
&\quad = \sum_{i=1}^{k} \int_K \limsup_{n \to \infty} \left[\min\left\{\inf_{y \in K} \left\{\lambda_i b^2 d^2(x,y) - \widetilde{f_i^n}(y)\right\}, \lambda_i a\right\}\right] \mathrm{d}\mu_i(x) \\
&\quad \leqslant \sum_{i=1}^{k} \int_K \min\left\{\inf_{y \in K} \left\{\limsup_{n \to \infty} \left(\lambda_i b^2 d^2(x,y) - \widetilde{f_i^n}(y)\right)\right\}, \lambda_i a\right\} \mathrm{d}\mu_i(x)
\end{aligned}$$

https://doi.org/10.1017/prm.2021.27 Published online by Cambridge University Press

$$= \sum_{i=1}^{k} \int_K \min \left\{ \inf_{y \in K} \left\{ \lambda_i b^2 d^2(x,y) - \widetilde{f}_i(y) \right\}, \lambda_i a \right\} \mathrm{d}\mu_i(x)$$

$$= \sum_{i=1}^{k} \int_K \overline{S}_{\lambda_i} \widetilde{f}_i(x) \, \mathrm{d}\mu_i(x).$$

Therefore, we must have equality everywhere. Hence, we get the result. □

Acknowledgements

Part of this paper was carried out when N.P. Chung visited University of Science, Vietnam National University at Hochiminh city on summer 2019. He is grateful to the math department there for its warm hospitality. The authors were partially supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government nos. NRF-2016R1A5A1008055, NRF-2016R1D1A1B03931922, and NRF-2019R1C1C1007107. We thank the anonymous referee for useful comments, especially for improving the statement of theorem 4.5.

References

1 M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** (2011), 904–924.

2 J.-J. Alibert, G. Bouchitté and T. Champion. A new class of costs for optimal transport planning. *European J. Appl. Math.* **30** (2019), 1229–1263.

3 L. Ambrosio, N. Gigli and G. Savare. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd edn, Lectures in Math (Basel: ETH Zuurich, Birkhauser Verlag, 2008).

4 J. Backhoff-Veraguas, M. Beiglböck and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differential Equations* **58** (2019), 1–28.

5 E. Boissard, T. Le Gouic and J.-M. Loubes. Distribution's template estimate with Wasserstein metrics. *Bernoulli* **21** (2015), 740–759.

6 L. A. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle. *Ann. Math.* **171** (2010), 673–730.

7 L. Chizat, G. Peyré, B. Schmitzer and F.-X. Vialard. Unbalanced optimal transport: dynamic and Kantorovich formulations. *J. Funct. Anal.* **274** (2018), 3090–3123.

8 N.-P. Chung and M.-N. Phung. Barycenters in the Hellinger-Kantorovich space. to appear Applied Mathematics & Optimization.

9 N.-P. Chung and T.-S. Trinh. Duality and quotients spaces of generalized Wasserstein spaces. arXiv:1904.12461.

10 N.-P. Chung and T.-S. Trinh. Weak optimal entropy transport problems. arXiv:2101.04986.

11 I. Ekeland and R. Témam. *Convex analysis and variational problems.* Corrected reprint of the 1976 English edition, Classics in Applied Mathematics, vol. 28 (Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1999). Translated from the French.

12 G. Friesecke, D. Matthes and B. Schmitzer. Barycenters for the Hellinger-Kantorovich distance over $\mathbb{R}^d$. *SIAM J. Math. Anal.* **53** (2021), 62–110.

13 N. Gozlan, C. Roberto, P.-M. Samson and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* **273** (2017), 3327–3405.

14 L. G. Hanin. Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces. *Proc. Am. Math. Soc.* **115** (1992), 345–352.

15 L. G. Hanin. *An extension of the Kantorovich norm.* Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997), Contemp. Math., vol. 226 (Providence, RI: Am. Math. Soc., 1999), pp. 113–130.

16 J. Kitagawa and B. Pass. The multi-marginal optimal partial transport problem. *Forum Math. Sigma.* **3** (2015), E17.

17 S. Kondratyev, L. Monsaingeon and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations* **21** (2016), 1117–1164.

18 M. Liero, A. Mielke and G. Savaré. Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *Invent. Math.* **211** (2018), 969–1117.

19 K. R. Parthasarathy. *Probability measures on metric spaces* (Providence, RI: AMS Chelsea Publishing, 2005). Reprint of the 1967 original.

20 B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.* **211** (2014), 335–358.

21 B. Piccoli and F. Rossi. On properties of the generalized Wasserstein distance. *Arch. Ration. Mech. Anal.* **222** (2016), 1339–1365.

22 B. Piccoli, F. Rossi and M. Tournus. A Wasserstein norm for signed measures, with application to non local transport equation with source term. hal-01665244v3.

23 W. Rudin. *Real and complex analysis*, 3rd edn (New York: McGraw-Hill Book Co., 1987).

24 K.-T. Sturm. *Probability measures on metric spaces of nonpositive curvature*. Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris 2002), Contemp. Math., vol. 338 (Providence, RI, Amer. Math. Soc., 2003), pp. 357–390.

25 C. Villani. *Topics in optimal transportation*. Graduate Studies in Mathematics, vol. 58 (Providence, RI, Amer. Math. Soc., 2003).

26 C. Villani. *Optimal transport. Old and New*. Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 338 (Berlin: Springer-Verlag, 2009).