Electronic Journal of Differential Equations, Vol. 2023 (2023), No. 22, pp. 1–24.
ISSN: 1072-6691. URL: https://ejde.math.txstate.edu or https://ejde.math.unt.edu

AN OPTIMAL TRANSPORT PROBLEM WITH STORAGE FEES

MOHIT BANSIL, JUN KITAGAWA

Communicated by Nestor Guillen

Abstract. We establish basic properties of a variant of the semi-discrete optimal transport problem in a relatively general setting. In this problem, one is given an absolutely continuous source measure and cost function, along with a finite set which will be the support of the target measure, and a "storage fee" function. The goal is to find a map for which the total transport cost plus the storage fee evaluated on the masses of the pushforward of the source measure is minimized. We prove existence and uniqueness for the problem, derive a dual problem for which strong duality holds, and give a characterization of dual maximizers and primal minimizers. Additionally, we find some stability results for minimizers and a Γ-convergence result as the target set becomes denser and denser in a continuum domain.

1. Introduction

Semi-discrete optimal transport. We begin by recalling the classical optimal transport problem. Suppose X, Y are Polish spaces, $c : X \times Y \to \mathbb{R}$ is a Borel measurable *cost function*, and μ, ν are Borel probability measures on X and Y respectively. Then the *optimal transport problem* or *Monge problem* transporting μ to ν is to find a Borel measurable mapping $T : X \to Y$ such that $T_{\#}\mu = \nu$ (here recall the *pushforward measure* is defined by $T_{\#}\mu(E) = \mu(T^{-1}(E))$ for any measurable $E \subset Y$), and T satisfies

$$\int_X c(x, T(x))\, d\mu(x) = \min_{S_{\#}\mu=\nu} \int_X c(x, S(x))\, d\mu(x). \tag{1.1}$$

If ν is a finite linear combination of delta measures, the above is usually referred to as the *semi-discrete* optimal transport problem.

We will now be interested in the following variant of the semi-discrete optimal transport problem, where we introduce a "storage fee." Fix a finite collection of N points, $Y := \{y_j\}_{j=1}^N$ and a function $F : \mathbb{R}^N \to \mathbb{R}$, and assume μ is a Borel probability measure on a Polish space X. This variant is to find a pair (T, λ) where $\lambda = (\lambda^1, \dots, \lambda^N) \in \mathbb{R}^N$ and $T : X \to Y$ is Borel measurable satisfying

$$T_{\#}\mu = \sum_{j=1}^N \lambda^j \delta_{y_j}$$

2020 *Mathematics Subject Classification.* 49J45, 49K40, 49Q22.
Key words and phrases. Optimal transport; gamma convergence; storage fees; queue penalization.

Submitted June 22, 2022. Published March 2, 2023.

such that

$$\int_X c(x,T(x))\,d\mu(x)+F(\lambda)=\min_{\tilde\lambda\in\mathbb{R}^N,\ \tilde T_{\#}\mu=\sum_{j=1}^N\tilde\lambda^j\delta_{y_j}}\int_X c(x,\tilde T(x))\,d\mu(x)+F(\tilde\lambda). \tag{1.2}$$

Note here that the vector λ is actually uniquely determined by the map T. We will consider a relaxation of this problem which we will refer to as the *primal problem* for the remainder of the paper. To define this relaxation, we write $\Pi(\mu,\nu)$ to denote the space of probability measures on $X\times Y$ whose left and right marginals are μ and ν respectively. Then, we wish to find a pair (γ,λ) where $\lambda\in\mathbb{R}^N$ and $\gamma\in\Pi(\mu,\sum_{j=1}^N\lambda^j\delta_{y_j})$, satisfying

$$\int_{X\times Y}c\,d\gamma+F(\lambda)=\min_{\tilde\lambda\in\mathbb{R}^N,\ \tilde\gamma\in\Pi(\mu,\sum_{j=1}^N\tilde\lambda^j\delta_{y_j})}\int_{X\times Y}c\,d\tilde\gamma+F(\tilde\lambda). \tag{1.3}$$

The above relaxation is the analogue of relaxing the Monge problem (1.1) in classical optimal transport to the *Kantorovich problem*, which we recall is (fixing Borel probability measures μ and ν on any two topological spaces X and Y) the problem of finding a measure $\gamma\in\Pi(\mu,\nu)$ satisfying

$$\int_{X\times Y}c\,d\gamma=\min_{\tilde\gamma\in\Pi(\mu,\nu)}\int_{X\times Y}c\,d\tilde\gamma. \tag{1.4}$$

Once a minimizing pair in the above primal problem (1.3) is found, it is clear the measure γ is a solution in the Kantorovich problem (1.4) with the choice $\nu=\sum_{j=1}^N\lambda^j\delta_{y_j}$. Hence under standard conditions on the cost function and μ, it is easily seen that a solution of (1.3) gives rise to a solution of the Monge version of the problem (1.2). For more details see Remark 4.3.

Previous results. The paper [6] considers the problem presented here in the specific case of cost function given by $c(x,y)=|x-y|^p$ with $p\geq 1$, and storage fee function of the form $F(\lambda)=\sum_{j=1}^N\lambda^jh_j(\lambda^j)$ for some functions h_j (note however, the authors mention their results can be extended to more general cost functions satisfying the condition (4.2)). This previous result gives conditions for optimizers, but does not introduce the dual problem or show stability properties as we do here. We are careful to mention our characterization from Section 4 matches the characterization of optimizers given in [6]: however our methods differ as we use a proof based on the dual formulation, while [6] relies on both the specific form of their function F and an assumption of differentiability. Finally, we mention that [6] also analyzes an associated but different variational problem which we do not discuss, our problem is equivalent to what Crippa, Jimenez, and Pratelli refer to as finding an "optimum," while the above reference deals with the additional problem of finding an "equilibrium."

There are also a number of results in the literature dealing with the so-called bilevel location problem using the framework of optimal transport: this can be viewed as a two level problem in which there is a "lower level problem" equivalent to the problem discussed in this manuscript, followed by a second "upper level problem" consisting of minimizing over the locations $\{y_j\}_{j=1}^N$ in the target domain. The paper [8], analyzes the case when the lower level problem corresponds to our problem with $c(x,y)=|x-y|^2$ in $\mathbb{R}^2\times\mathbb{R}^2$ and $F(\lambda)=\langle a,\lambda\rangle$ for a fixed vector a, and shows existence and uniqueness under certain conditions. The result [5] views the problem in an economic context, their lower level problem is related to a *partial* optimal transport problem with an associated storage fee; note however that their problem is not exactly an optimal transport problem as it arises from the problem of monopolistic pricing, and involves an extra nonlinearity in the definition of Laguerre cells. We emphasize that we do not deal with the "upper level problem", while the above two references also analyze that problem as well.

Contributions of this article. We show the existence of minimizers of the optimal transport problem with storage fees, exhibit a dual problem with strong duality (along with existence of dual maximizers) for a wide class of storage fees F and c, show a sharp characterization of dual and primal extremizers, and produce some stability results. One novelty of our results is that both existence and the characterization of minimizers do not require any assumption of differentiability of F, in contrast with previous results such as [6]. The class of functions is wide enough to allow for storage fee functions which may take the value ∞ at some points in the standard simplex. Additionally, the simple characterization of extremizers can be exploited to form a provably convergent numerical approximation scheme based on a damped Newton method, which the authors have explored in [1].

We also mention the following economics interpretation of the problem with storage fees. A manufacturer has a distribution of factories μ, all producing the same product, and is leasing a finite number of warehouses at the locations y_j. At the end of each production cycle, the manufacturer must ship all of their product to be stored at some combination of the warehouses. The manufacturer can choose how many units of their product is to be stored at each warehouse, but the leasing company will charge a storage fee given by F based on the capacity used. Additionally, there is a cost associated to the transportation itself given by c, and the goal is to minimize the total cost of transport plus storage.

We conclude by mentioning that even in the restricted cases treated in previous work, there is no mention of the associated dual problem, which we have shown has strong duality. The dual problem we exhibit here has a very natural interpretation in terms of the economic analogy mentioned above. Namely, for the classical optimal transport problem (see [12, p. 53]), the dual problem can be interpreted as contracting a third party shipping company: a pair of dual potentials represents a price schedule (the amount the company would charge to pick up and drop off goods at given source and target locations) and the dual functional would be the total price charged for a given distribution of source and target goods. In our dual problem with storage fee (see Theorem 3.3 for precise statement), the dual pair (φ, ψ) again represents a price schedule for pickup and drop off, but the term $F^*(\psi)$ indicates that the third party company will build into their prices the added step of optimizing how to distribute the target mass, accounting for the storage fee that will be charged. While the classical optimal transport problem has a well-known canonical dual problem, as a general optimization our primal problem may have many possible associated dual problems; this interpretation in terms of shipping suggests the particular dual problem presented here represents a canonical choice.

Remark 1.1. The minimization problem with storage fees has potential applications in supervised data clustering. If μ is a distribution of data points known to fall into N clusters, and $y_1, \dots, y_N$ are representative data points from each cluster, solving the optimal transport problem with a storage fee is a method of clustering the data, where the size of each cluster is penalized according to the storage fee function F. The theory developed is flexible enough to allow for combining a convex function to penalize large cluster size, with a hard size constraint ($F \equiv \infty$ when the capacity at a point exceeds a certain threshold) in such a data clustering problem.

As a toy example, one could consider testing whether a handwritten document conforms well to Benford's law of leading digits. It is empirically observed that the probability the digit d occurs as a leading digit in a document tends to follow Benford's law:

$$P(d) \approx \log_{10}(1 + \frac{1}{d}), \quad d = 1, \dots, 9.$$

This has been suggested as a potential method to detect whether numerical figures are unnaturally generated, for example in accounting fraud or fabrication of scientific data. Now consider a handwritten document containing many numbers. Then each leading digit can be considered as a grey scale image, which can be described as a vector in some $\mathbb{R}^n$, suppose the distribution μ describes this collection of digits. Let $Y = \{y_1, \dots, y_9\}$ where each y_i is one representative for the handwritten digit j. Finally suppose some cost function $c(x, y_j)$ is given which is a measure of affinity between a handwritten digit x and the representative digit y_j. Now define the storage fee function $F : \mathbb{R}^9 \to \mathbb{R}$ by

$$F(\lambda) := \sum_{j=1}^{9} f(\lambda^j - \log_{10}(1 + \frac{1}{j})),$$

where $f : \mathbb{R} \to \mathbb{R}$ is a strictly increasing, convex function with $f(0)$; note that F is a convex function hence (modulo conditions on c) all of our results in this article will be applicable. Then a minimizer in (1.3) with this choice of F will give a clustering of the leading digits appearing in the distribution μ, and the value of the functional associated to this minimizer can represent a measure of divergence of the document from Benford's law.

Notation and conventions. We will fix some notation and conventions to be used in the remainder of the paper. We fix positive integers N and n and a collection $Y := \{y_j\}_{j=1}^N$ (however, in Subsection 5.1 we will consider an arbitrary bounded metric space Y, not necessarily finite). We also denote the standard N-simplex by

$$\Lambda := \{\lambda \in \mathbb{R}^N : \sum_{j=1}^{N} \lambda^j = 1, \ \lambda^j \geq 0\},$$

and given a vector $\lambda \in \Lambda$ we write $\nu_\lambda := \sum_{j=1}^N \lambda^j \delta_{y_j}$. We reserve the boldface notation $\mathbf{1}$ for the vector in $\mathbb{R}^N$ whose components are all 1. The space of Borel probability measures on a metric space X will be denoted $\mathcal{P}(X)$. The subset of $\mathcal{P}(X \times Y)$ consisting of measures with left and right marginals equal to $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ respectively will be written as $\Pi(\mu, \nu)$, and $\Pi(\mu)$ will be the subset of $\mathcal{P}(X \times Y)$ consisting of measures with left marginal μ. Projection from $X \times Y$ to X and Y will be written π_X and π_Y. We will write $C_b(S)$ for the set of bounded, continuous functions on a metric space S. As is standard practice, we will refer to weak $*$ convergence of probability measures in duality with $C_b(S)$ as "weak convergence."

We will also identify any real valued function on Y with a vector in $\mathbb{R}^N$ in the obvious way, and always assume that X is separable.

Also given a convex function $f : \mathbb{R}^N \to \mathbb{R}\cup\{\infty\}$, we write $\operatorname{dom}(f) := \{x : f(x) < \infty\}$ to denote its *effective domain*. The function F will be assumed to be lower semicontinuous on $\mathbb{R}^N$ for Section 2, while for Sections 3 and 4 we assume F is a proper, closed, convex function, with $\operatorname{dom}(F) \subset \Lambda$.

Finally we use $|\cdot|$ to denote either the Euclidean norm of a vector, or the Lebesgue measure of a set, which usage should be clear to the reader from context.

Outline. In Section 2 we show existence of minimizers for the relaxed problem (1.3) when F and c are lower semicontinuous. In Subsection 3 we show strong duality when F is convex, and existence of dual maximizers when c is uniformly continuous and bounded (Proposition 3.5). We also provide an estimate on the duality gap when F is not convex which is sharp (Remark 3.6 and Proposition 3.7). In Subsection 3.1 we show some relationships between our dual problem and the classical Kantorovich dual problem. In Section 4, under additional assumptions on c and μ we show the existence of solutions of the primal problem exist in the "Monge" sense of (1.2) along with a sharp characterization of dual and primal extremizers. Finally we show two stability results in Section 5: the first for minimizers under perturbations of the storage fee function, and the second a Γ-convergence result as one takes a sequence of finite target sets that become dense in some domain. Appendix 6 contains the proof of the sharp duality gap Proposition 3.7 while appendix 7 presents an example where a maximizer in the classical Kantorovich problem is not a maximizer in our dual problem.

2. Existence of minimizers

In this section we prove the existence of minimizers for problem (1.3). First we recall some elementary definitions and results.

Definition 2.1. A collection of $\Gamma \subset \mathcal{P}(X)$ is said to be *tight* if for any $\epsilon > 0$, there exists a compact set $K \subset X$ such that $\mu(K) > 1 - \epsilon$ for every $\mu \in \Gamma$.

Lemma 2.2. *Fix $\mu \in \mathcal{P}(X)$, then $\Pi(\mu)$ is tight.*

Proof. Since X is separable, the collection $\{\mu\}$ is tight. Now let $\epsilon > 0$ be given. Choose $K \subset X$, compact so that $\mu(K) > 1 - \epsilon$. Note that since Y is finite, $K \times Y$ is also compact. Then for any $\gamma \in \Pi(\mu)$, we find $\gamma(K \times Y) = \mu(K) > 1 - \epsilon$, hence $\Pi(\mu)$ is tight. $\square$

As a corollary we see that $\Pi(\mu)$ is relatively weakly compact by Prokhorov's Theorem (see [2, Theorem 5.1]). With this compactness in hand, existence of a minimizer follows easily.

Theorem 2.3. *Suppose $c(\cdot, y_j) : X \to (-\infty, \infty]$ is lower semicontinuous for each j, there exists some upper semicontinuous $a \in L^1(\mu)$ with $a : X \to [-\infty, \infty)$ such that $\min_j c(x, y_j) \geq a(x)$ for all $x \in X$, and F is lower semicontinuous. Then there exist minimizers of the primal problem* (1.3)*.*

Proof. Replacing each $c(\cdot, y_j)$ with $c(\cdot, y_j) - a$, we may assume $c(\cdot, y_j) \geq 0$ and is lower semicontinuous for each j.

Let $\{(\gamma_k, \lambda_k)\}_{k=1}^{\infty}$ be a minimizing sequence for (1.3): that is a sequence with $\gamma_k \in \Pi(\mu)$ such that $\int_{X\times Y} c\, d\gamma_k + F(\lambda_k)$ approaches the minimum value in (1.3), where ν_{λ_k} is the right marginal of γ_k. By the above remark $\Pi(\mu)$ is compact and so there is a subsequence of γ_k, which we do not relabel, that converges weakly to some $\gamma_{\min} \in \Pi(\mu)$. We will show that $\gamma_{\min}$ is actually a minimizer. Let $\lambda_{\min}$ be the vector in $\mathbb{R}^N$ so that $\nu_{\lambda_{\min}}$ is the right marginal of $\gamma_{\min}$.

For $\varphi \in C_b(Y) = \mathbb{R}^N$,

$$\lim_{k\to\infty} \int_Y \varphi d\nu_{\lambda_k} = \lim_{k\to\infty} \int_{X\times Y} (\varphi \circ \pi_Y)\, d\gamma_k = \int_{X\times Y} (\varphi \circ \pi_Y)\, d\gamma_{\min} = \int_Y \varphi d\nu_{\lambda_{\min}},$$

meaning ν_{λ_k} converges weakly to $\nu_{\lambda_{\min}}$, hence in particular since $\mathbb{1}_{X\times\{y_j\}} \in C_b(X\times Y)$, we have

$$\lim_{k\to\infty} \lambda_k^j = \lim_{k\to\infty} \gamma_k(X\times\{y_j\}) = \gamma_{\min}(X\times\{y_j\}) = \lambda_{\min}^j. \tag{2.1}$$

By [12, Lemma 4.3], the functional $\gamma \mapsto \int_{X\times Y} c\, d\gamma$ is weakly lower semicontinuous. Since F is lower semicontinuous, we thus obtain

$$\int_{X\times Y} c\, d\gamma_{\min} + F(\lambda_{\min}) \le \liminf_{k\to\infty} \Big(\int_{X\times Y} c\, d\gamma_k + F(\lambda_k) \Big)$$

showing $(\gamma_{\min}, \lambda_{\min})$ is a minimizer. □

3. Dual problem

Our first goal in this section will be to deduce a dual problem associated to our primal problem (1.3). For the following Sections 3 and 4, we will assume that F is a proper, closed, convex function, with $\mathrm{dom}(F) \subset \Lambda$.

Strong duality. To state the dual problem, we first recall a basic concept from convex analysis.

Definition 3.1. Let E be a Banach space. If $G : E \to \mathbb{R}\cup\{\infty\}$ is a proper function (i.e., it is not identically ∞), its *Legendre-Fenchel transform* is the (proper, convex) function $G^* : E^* \to \mathbb{R}\cup\{\infty\}$ defined for any $y \in E^*$ by

$$G^*(y) := \sup_{x\in E} (\langle y, x\rangle - F(x)),$$

where $\langle y, x\rangle$ is the duality pairing between elements of E^* and E.

If $E = E^* = \mathbb{R}^N$, the Legendre-Fenchel transform is called the *Legendre transform*.

Since Λ is compact, we see that F is bounded from below everywhere, as any affine function supporting F from below will be bounded on $\mathrm{dom}(F)$. Thus, since F is proper we see that F^* is actually finitely valued everywhere on $\mathbb{R}^N$ by the definition of Legendre transform.

It is also convenient at this point to introduce the notion of c and c^*-transforms, and c-convexity. Note carefully that, since we are in the semi-discrete case the c-transform of a function defined on X will be a vector in $\mathbb{R}^N$, while the c^*-transform of a vector in $\mathbb{R}^N$ will be a function whose domain is X.

Definition 3.2. If $\varphi : X \to \mathbb{R}\cup\{\infty\}$ (which is not identically ∞) and $\psi \in \mathbb{R}^N$, their c- and c^*-transforms are a vector $\varphi^c \in \mathbb{R}^N$ and a function $\psi^{c^*} : X \to \mathbb{R}\cup\{\infty\}$ respectively, defined by

$$(\varphi^c)^j := \sup_{x\in X}(-c(x, y_j) - \varphi(x)), \quad (\psi^{c^*})(x) := \max_{1\le j\le N}(-c(x, y_j) - \psi^j).$$

If $\varphi : X \to \mathbb{R}\cup\{\infty\}$ is the c^*-transform of some vector in $\mathbb{R}^N$, we say φ *is a c-convex function*. A pair (φ, ψ) with $\varphi : X \to \mathbb{R}\cup\{\infty\}$ and $\psi \in R^N$ is is called a *c-conjugate pair* if $\varphi = \psi^{c^*}$ and $\psi = \psi^{c^*c} = \varphi^c$.

Note that from the definition, if $-\varphi - \psi \le c$, then $-\varphi(x) \le -\psi^{c^*}(x)$ and $-\psi^j \le -(\varphi^c)^j$ for all $x \in X$ and $1 \le j \le N$, while $-\varphi - (\varphi^c) \le c$, $-(\psi^{c^*}) - \psi \le c$ always holds.

Theorem 3.3 (Strong duality)**.** *Suppose c satisfies the same conditions as Theorem 2.3 and F is a proper, closed, convex function, with* $\operatorname{dom}(F) \subset \Lambda$*, then there is strong duality, i.e.,*

$$\begin{aligned}
&\min_{\lambda\in\Lambda,\, \gamma\in\Pi(\mu,\nu_\lambda)} \int_{X\times Y} c\, d\gamma + F(\lambda) \\
&= \sup\big\{ -\int_X \varphi\, d\mu - F^*(\psi) : (\varphi,\psi) \in L^1(\mu)\times\mathbb{R}^N, \\
&\quad -\varphi(x) - \psi^j \le c(x,y_j),\ \forall y_j \in Y,\ \mu\text{-a.e. } x \in X\big\}.
\end{aligned} \tag{3.1}$$

Moreover, it is possible to replace $L^1(\mu)$ by $C_b(X)$ in the right hand side above.

Proof. Since $C_b(X) \subset L^1(\mu)$, the supremum of the dual problem with $C_b(X)$ is a lower bound for the one with $L^1(\mu)$. On the other hand, for any $\gamma \in \Pi(\mu,\nu_\lambda)$ for some $\lambda \in \Lambda$ and $(\varphi,\psi) \in L^1(\mu)\times\mathbb{R}^N$ which is admissible in (3.1), by the definition of F^* (see [9, p.105, Fenchel's inequality]),

$$\begin{aligned}
-\int_X \varphi\, d\mu - F^*(\psi) &\le -\int_X \varphi\, d\mu - \langle\lambda,\psi\rangle + F(\lambda) \\
&= \int_{X\times Y} (-\varphi(x) - \psi(y))\, d\gamma(x,y) + F(\lambda) \\
&\le \int_{X\times Y} c\, d\gamma + F(\lambda).
\end{aligned}$$

Thus it is sufficient to prove that

$$\begin{aligned}
&\min_{\lambda\in\Lambda,\ \gamma\in\Pi(\mu,\nu_\lambda)} \int_{X\times Y} c\, d\gamma + F(\lambda) \\
&\le \sup\big\{ -\int_X \varphi\, d\mu - F^*(\psi) : (\varphi,\psi) \in C_b(X)\times\mathbb{R}^N, \\
&\quad -\varphi(x) - \psi^j \le c(x,y_j), \forall (x,y_j) \in X\times Y\}.
\end{aligned} \tag{3.2}$$

First assume X is compact and each $c(\cdot,y_j)$ is continuous; by subtracting a constant we may assume $c \ge 0$. We first prove the duality statement with $C_b(X)$ in place of $L^1(\mu)$. Let $E = C(X\times Y) = C_b(X\times Y)$ (by compactness of X) and note its dual is given by $E^* = \mathcal{M}(X\times Y)$, the space of Radon measures on $X\times Y$. Then define Θ, $\Xi : E \to \mathbb{R}\cup\{\infty\}$ by

$$\Theta(u) := \begin{cases} 0, & \text{if } u(x,y) \ge -c(x,y),\ \forall (x,y) \in X\times Y \\ \infty, & \text{otherwise} \end{cases}$$

$$\Xi(u) := \begin{cases} -\int_X \varphi\, d\mu + F^*(-\psi), & \\ \quad \exists(\varphi,\psi) \in C_b(X)\times\mathbb{R}^N \text{ s.t. } u(x,y_j) = -\varphi(x) - \psi^j,\ \forall (x,y_j) \in X\times Y, & \\ \infty, \quad \text{otherwise,} & \end{cases}$$

(we will write $u = -\varphi - \psi$ as shorthand for the condition in the first case of Ξ above). It is now necessary to check that Ξ as above is well-defined. Indeed, if $u(x,y_j) = -\varphi_1(x) - \psi_1^j = -\varphi_2(x) - \psi_2^j$ for all $(x,y_j) \in X\times Y$, we can see there exists some $r \in \mathbb{R}$ such that $\varphi_1 = -r + \varphi_2$, and $\psi_1 = \psi_2 + r\mathbf{1}$. Since Λ is contained in a plane orthogonal to $\mathbf{1}$ and $F = \infty$ outside of this plane, a direct verification of the definition implies that for any $\lambda \in \operatorname{dom}(F)$ and $\psi \in \mathbb{R}^N$ such that $\langle \cdot - \lambda, \psi \rangle + F(\lambda) \leq F$ on $\mathbb{R}^N$, we have $\langle \cdot - \lambda, \psi + r\mathbf{1} \rangle + F(\lambda) \leq F$ on $\mathbb{R}^N$ for any $r \in \mathbb{R}$ as well. Since F^* is finite everywhere, by [9, Theorem 23.4] there exists some λ which must be in Λ, such that $\langle \cdot - (-\psi_2), \lambda \rangle + F^*(-\psi_2) \leq F^*$ on $\mathbb{R}^N$. Hence by [9, Theorem 23.5],

$$F^*(-\psi_1) = F^*(-\psi_2 - r\mathbf{1}) = -\langle \psi_2, \lambda \rangle - r\langle \mathbf{1}, \lambda \rangle - F(\lambda) = -r + F^*(-\psi_2). \tag{3.3}$$

Since μ is a probability measure, this shows Ξ is well-defined. It is immediate to see that Θ and Ξ are convex, and that for $u \equiv 1$, $\Theta(u)$, $\Xi(u) < \infty$ and Θ is continuous at u.

We now compute

$$\Theta^*(-\gamma) = \sup_{u \geq -c} \Big(-\int_{X \times Y} u \, d\gamma \Big) = \begin{cases} \int_{X \times Y} c \, d\gamma, & \gamma \geq 0 \\ \infty, & \text{otherwise} \end{cases}$$

and

$$\begin{aligned} \Xi^*(\gamma) &= \sup_{(\varphi, \psi) \in C_b(X) \times \mathbb{R}^N} \Big(\int_{X \times Y} (-\varphi(x) - \psi) \, d\gamma + \int_X \varphi \, d\mu - F^*(-\psi) \Big) \\ &= \begin{cases} \sup_\psi \Big(\langle -\psi, \lambda \rangle - F^*(-\psi) \Big) = F^{**}(\lambda) = F(\lambda), & (\pi_X)_\# \gamma = \mu, (\pi_Y)_\# \gamma = \nu_\lambda \\ \infty, & \text{otherwise} \end{cases} \end{aligned}$$

where we used convexity of F in the last line above.

Next we find (where by an abuse of notation we will write $-\varphi - \psi \leq c$ to denote $-\varphi(x) - \psi^j \leq c(x, y_j)$ for all $x \in X$ and $1 \leq j \leq N$)

$$\begin{aligned} \inf_{z \in E} (\Theta(z) + \Xi(z)) &= \inf_{\tilde{\varphi} + \tilde{\psi} \leq c} \Big(-\int_X \tilde{\varphi} \, d\mu + F^*(-\tilde{\psi}) \Big) \\ &= -\sup_{-\varphi - \psi \leq c} \Big(\int_X (-\varphi) \, d\mu - F^*(\psi) \Big). \end{aligned}$$

Hence by the Fenchel-Rockafellar theorem (see [11, Theorem 1.9]) we have

$$\begin{aligned} \sup_{-\varphi - \psi \leq c} \Big(-\int_X \varphi \, d\mu - F^*(\psi) \Big) &= -\inf_{z \in E} (\Theta(z) + \Xi(z)) \\ &= \min_{\lambda \in \Lambda, \gamma \in \Pi(\mu, \nu_\lambda)} \Big(\int_{X \times Y} c \, d\gamma + F(\lambda) \Big) \end{aligned}$$

proving (3.2) in this case.

Next assume X is a general Polish space, but $c \in C_b(X \times Y)$ is uniformly continuous; again we may assume $c \geq 0$. Fix $\epsilon \in (0, 1)$, then by Ulam's lemma there exists a compact $K \subset X$ such that $\mu(K) > 1 - \epsilon > 0$. Let us define $\tilde{\mu} := \frac{\mathbb{1}_K \mu}{\mu(K)}$, then by Theorem 2.3 there is a minimizer $(\tilde{\gamma}, \lambda)$ of (1.3) with $\tilde{\mu}$ and the restriction of c to $K \times Y$ replacing μ and c. Then we can apply the first case above to find $(\tilde{\varphi}, \tilde{\psi}) \in C_b(K) \times \mathbb{R}^N$, admissible in (3.1), satisfying

$$-\int_X \tilde{\varphi} \, d\tilde{\mu} - F^*(\tilde{\psi}) \geq \int_{K \times Y} c \, d\tilde{\gamma} + F(\lambda) - \epsilon \geq \min_\Lambda F - \epsilon = -F^*(0) - \epsilon,$$

where by an abuse of notation we identify the discrete measure $(\pi_Y)_\# \gamma$ with its vector of weights in Λ, and we have used [9, Theorem 27.1] to obtain the final equality. In particular, this shows that for some $x_0 \in K$, we have $-\tilde{\varphi}(x_0) - F^*(\tilde{\psi}) \geq -F^*(0) - 1$, then by adding a constant to $\tilde{\varphi}$ and subtracting the same multiple of $\mathbf{1}$ from $\tilde{\psi}$, by (3.3) we may assume $\min\big(-\tilde{\varphi}(x_0), -F^*(\tilde{\psi})\big) \geq -\frac{F^*(0)+1}{2}$. Then for any index j, we have $\tilde{\psi}^j \geq -c(x_0, y_j) - \tilde{\varphi}(x_0) \geq -\sup|c| - \frac{F^*(0)+1}{2}$. On the other hand, since F is proper, there exists at least one $\lambda_0 \in \Lambda$ for which $F(\lambda_0) < \infty$, thus we have $F^*(r\lambda_0) \geq \langle r\lambda_0, \lambda_0\rangle - F(\lambda_0) \to \infty$ as $r \to \infty$. Since $F^*(\tilde{\psi})$ is bounded from above, this shows that $\langle \tilde{\psi}, \lambda_0\rangle$ has an upper bound, independent of ϵ. Combined with the previous line, there exists some fixed index j and $M > 0$ independent of ϵ such that $|\tilde{\psi}^j| \leq M$. In turn, this shows $\sup_X |\tilde{\psi}^{c^*}| \leq \sup_{X\times Y}|c| + M$, and we calculate

$$
\begin{aligned}
&-\int_X \tilde{\psi}^{c^*}\, d\mu - F^*(\tilde{\psi}) \\
&= -\mu(K)\int_K \tilde{\psi}^{c^*}\, d\tilde{\mu} - \int_{X\setminus K} \tilde{\psi}^{c^*}\, d\mu - F^*(\tilde{\psi}) \\
&\geq -\mu(K)\int_K \tilde{\varphi}\, d\tilde{\mu} - \mu(X\setminus K)\sup_X |\tilde{\psi}^{c^*}| - F^*(\tilde{\psi}) \\
&\geq (1-\epsilon)\Big(\int_{K\times Y} c\, d\tilde{\gamma} + F(\lambda)\Big) - \epsilon\big(1 + \sup_{X\times Y}|c| + M + \frac{F^*(0)+1}{2}\big).
\end{aligned} \tag{3.4}
$$

Now define $\gamma := \mu(K)\tilde{\gamma} + (\mu\otimes\nu_\lambda)\mathbb{1}_{(X\setminus K)\times Y}$. For $A \subset X$ and $B \subset Y$ Borel we have

$$
\begin{aligned}
\gamma(A\times Y) &= \mu(K)\tilde{\gamma}(A\times Y) + \mu(A\setminus K) = \mu(A\cap K) + \mu(A\setminus K) = \mu(A), \\
\gamma(X\times B) &= \mu(K)\tilde{\gamma}(X\times B) + \mu(X\setminus K)\nu_\lambda(B) \\
&= \mu(K)\nu_\lambda(B) + \mu(X\setminus K)\nu_\lambda(B) = \nu_\lambda(B),
\end{aligned}
$$

thus $\gamma \in \Pi(\mu, \nu_\lambda)$. Then

$$
\begin{aligned}
\int_{X\times Y} c\, d\gamma + F(\lambda) &= \mu(K)\int_{X\times Y} c\, d\tilde{\gamma} + \int_{(X\setminus K)\times Y} c\; d(\mu\otimes\nu_\lambda) + F(\lambda) \\
&\leq \int_{K\times Y} c\, d\tilde{\gamma} + F(\lambda) + \mu(X\setminus K)\sup_{X\times Y}|c| \\
&\leq \int_{K\times Y} c\, d\tilde{\gamma} + F(\lambda) + \epsilon\sup_{X\times Y}|c|,
\end{aligned}
$$

hence combining with (3.4) and taking $\epsilon \to 0$ shows (3.2) in this case.

Finally, suppose c is lower semicontinuous, considering $c - a$ we can assume $c \geq 0$. By [4, Corollary 1.34], c is the limit of an increasing sequence of $c_k \in C_b(X\times Y)$ which are uniformly continuous, by Theorem 2.3 there exist a sequence of minimizers (γ_k, λ_k) to the associated primal problems (1.3). By Lemma 2.2 and arguing as in (2.1), we extract a subsequence where $\lambda_k \to \lambda \in \Lambda$ and γ_k converges weakly to $\gamma \in \Pi(\mu, \nu_\lambda)$. Arguing as in Step 3 of the proof of [11, Theorem 1.3] and using the lower semicontinuity of F, we find that $\int_{X\times Y} c\, d\gamma + F(\lambda) \leq \liminf_{k\to\infty}(\int_{X\times Y} c_k\, d\gamma_k + F(\lambda_k))$. We may then apply (3.2) for each c_k, and use that dual pairs admissible for c_k are admissible for c (since $c_k \leq c$) to obtain (3.2) for c, finishing the proof. $\square$

We will now show the existence of maximizers for the dual problem (3.1). As in the classical optimal transport case (see, for example [10, Proposition 1.11]), we utilize the c- and c^*-transforms of functions to obtain compactness.

Remark 3.4. It turns out a dual maximizing pair in (3.1) is maximizing in the classical optimal transport problem where the target measure is the minimizer in the primal problem (1.3). However this fact is not obvious because of the presence of the term $-F^*(\psi)$ in the dual problem, we will obtain this as a consequence of Lemma 3.10 in Subsection 3.1 below. We also warn the reader, a maximizer in the classical dual problem (3.5) between μ and a minimizing measure ν_λ in the primal problem may *not* be a maximizer in our dual problem 3.1, as illustrated in Example 7.1 in the appendix.

Finally, unlike Theorem 3.3, it is not clear how to extend existence of dual maximizers to more general (for example, lower semicontinuous) c. In the classical Kantorovich duality one can use an argument based on c-cyclical monotonicity as in [12, Theorem 5.10 (iii)], however this approach fails as there is a lack of information of F and F^*. Additionally, for more general c it could be possible that the potential function defined on Y takes the values ∞ in the classical dual problem, which corresponds in our problem to ψ having components equal to ∞, in this case it is not clear what the meaning of $F^*(\psi)$ should be. We hope to explore this question in a future work.

Proposition 3.5. (1) *If $c(\cdot, y_j)$ is lower semicontinuous for each j and $(\varphi, \psi) \in C_b(X) \times \mathbb{R}^N$ is any maximizing pair in* (3.1)*, then $\varphi \equiv \psi^{c^*}$ on $\operatorname{spt}\mu$.*
(2) *If $c(\cdot, y_j)$ is uniformly continuous and bounded for each j, then there exists at least one maximizer in $C_b(X) \times \mathbb{R}^N$ of the dual problem* (3.1) *that is a c-conjugate pair.*

Proof. (1) Let $(\varphi_{\max}, \psi_{\max})$ be a maximizing pair. Recall that $-\varphi_{\max} \le -\psi_{\max}^{c^*}$ on X. Since $-\psi_{\max}^{c^*}$ can be written as a *finite* minimum of lower semicontinous functions, it is also lower semicontinuous, then the set $\{-\varphi_{\max} < -\psi_{\max}^{c^*}\}$ is open; in particular if the inequality $-\varphi_{\max} < -\psi_{\max}^{c^*}$ is strict at any point in $\operatorname{spt}\mu$, it holds on a neighborhood relatively open in X. Then we would have $-\int_X \varphi_{\max}\, d\mu - F^*(\psi_{\max}) < -\int_X \psi_{\max}^{c^*}\, d\mu - F^*(\psi_{\max})$, contradicting that $(\varphi_{\max}, \psi_{\max})$ is a maximizing pair. Thus we must have $\varphi_{\max} \equiv \psi_{\max}^{c^*}$ on $\operatorname{spt}\mu$.

(2) Suppose $c(\cdot, y_j)$ is uniformly continuous and bounded for each j and let $(\varphi_k, \psi_k) \in L^1(\mu) \times \mathbb{R}^N$ be an admissible, maximizing sequence for (3.1). We may assume $\varphi_k = \psi_k^{c^*}$ and $\psi_k = \psi_k^{c^* c}$ for this sequence as $-\int_X \varphi_k\, d\mu \le -\int_X \psi_k^{c^*}\, d\mu$ and

$$\begin{aligned} -F^*(\psi_k) &= \inf_{\lambda \in \operatorname{dom}(F)} (\langle \lambda, -\psi_k \rangle + F(\lambda)) \\ &\le \inf_{\lambda \in \operatorname{dom}(F)} (\langle \lambda, -(\psi_k^{c^* c}) \rangle + F(\lambda)) \\ &= -F^*(\psi_k^{c^* c}), \end{aligned}$$

using that $\lambda^j \ge 0$ for all $\lambda \in \operatorname{dom}(F)$ and $-\psi_k \le -\psi_k^{c^* c}$ componentwise; in particular we may assume each $\varphi_k \in C_b(X)$. Since $(\psi_k + r\mathbf{1})^{c^*}(x) = \psi_k^{c^*}(x) - r$ for any r, the above along with (3.3) implies that replacing ψ_k by $(\psi_k + r\mathbf{1})^{c^* c}$ and taking $\varphi_k = (\psi_k + r\mathbf{1})^{c^*}$ does not reduce the values of $-\int_X \varphi_k\, d\mu - F^*(\psi_k)$ for each k, hence we may assume $\varphi_k(x_0) = 0$ for all k, for some fixed $x_0 \in X$. It can then be seen that boundedness and uniform continuity of the $c(\cdot, y_j)$ are enough to obtain boundedness and equicontinuity of $\{\varphi_k\}_{k=1}^\infty$ as in the proof of [10, Proposition 1.11] hence we can conclude existence of a subsequence, that we do not relabel, of (φ_k, ψ_k) that converges (φ_k uniformly on X and ψ_k in $\mathbb{R}^N$) to some $(\varphi_{\max}, \psi_{\max})$.

Since $-F^*$ is a concave function, finite on all of $\mathbb{R}^N$ by compactness of $\operatorname{dom}(F)$, it is continuous on $\mathbb{R}^N$, hence we obtain that $(\varphi_{\max}, \psi_{\max})$ is a maximizer in (3.1). We can replace the pair by $(\psi_{\max}^{c^*}, \psi_{\max}^{c^*c})$ which only increases the value of the associated functional, hence there exists at least one c-conjugate maximizing pair. □

Remark 3.6. If F is lower semicontinuous but not convex, we may not obtain strong duality, however it is easy to see that the duality gap is bounded by $\|F - F^{**}\|_{L^\infty(\operatorname{dom}(F)\cup\operatorname{dom}(F^{**}))}$, when the minimal value is finite and c is lower semicontinuous. Let $\mathfrak{m}_F$ and $\mathfrak{m}_{F^{**}}$ denote the minimal values in (1.3) with storage fee functions F and F^{**} respectively, assume $\mathfrak{m}_F < \infty$, and $\mathfrak{M}_F$ denote the supremum in the dual problem (3.1) associated to F. Since $F^{**} \le F$, we also have $\mathfrak{m}_{F^{**}} < \infty$. Then by strong duality combined with Proposition 5.1 below, we see (since $F^* = F^{***}$)

$$0 \le \mathfrak{m}_F - \mathfrak{M}_F = \mathfrak{m}_F - \mathfrak{M}_{F^{**}} = \mathfrak{m}_F - \mathfrak{m}_{F^{**}} \le \|F - F^{**}\|_{L^\infty(\operatorname{dom}(F)\cup\operatorname{dom}(F^{**}))}.$$

This bound is essentially sharp, as the following proposition (whose proof we defer to the appendix) illustrates.

Proposition 3.7. *Suppose $F : \Lambda \to \mathbb{R}$ is L-Lipschitz. Then for each $Y = \{y_j\}_{j=1}^N \subset \mathbb{R}^n$ satisfying $\min_{i\neq j} |y_i - y_j| > 2\sqrt{L}$, there exists a $\mu \in \mathcal{P}(\mathbb{R}^n)$ such that $\mathfrak{m}_F - \mathfrak{M}_F = \|F - F^{**}\|_{L^\infty(\Lambda)}$, where we take the cost function $c(x, y) = |x - y|^2$.*

3.1. **Relationship with classical Kantorovich duality.** The dual problem (3.1) bears many similarities to Kantorovich's dual problem for the classical optimal transport problem, and it is natural to explore relationships between maximizers of the two problems; we do so in this subsection. We recall the *classical dual problem* here for clarity.

Definition 3.8 ([11, Theorem 1.3])**.** If μ, ν are Borel probability measures on X and Y respectively and c a Borel measurable cost on $X \times Y$, Kantorovich's dual problem is to find

$$\begin{aligned} \sup \Big\{ -\int_X \varphi\, d\mu - \int_Y \psi d\nu : (\varphi, \psi) \in L^1(\mu) \times L^1(\nu), \\ -\varphi(x) - \psi(y) \le c(x, y),\ \forall (x, y) \in X \times Y \Big\}. \end{aligned} \tag{3.5}$$

First recall the definition.

Definition 3.9 ([9, Section 23])**.** If $F : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ is a convex function, its *subdifferential* at a point $\lambda_0 \in \mathbb{R}^N$ is defined as

$$\partial F(\lambda_0) := \{\psi \in \mathbb{R}^N : F(\lambda) \ge F(\lambda_0) + \langle \lambda - \lambda_0, \psi \rangle, \forall \lambda \in \mathbb{R}^N\}.$$

Next, we prove a key lemma.

Lemma 3.10. *Suppose c is lower semicontinuous, there exists a real valued upper semicontinuous function $a \in L^1(\mu)$ with $c(x, y_j) \ge a(x)$ for all x and j, and $(\gamma_{\min}, \lambda_{\min}) \in \Pi(\mu, \nu_{\lambda_{\min}}) \times \Lambda$ are a minimizing pair in the primal problem* (1.3)*. Then if there exists a $(\varphi_{\max}, \psi_{\max}) \in L^1(\mu) \times \mathbb{R}^N$ which is a maximizing pair in the dual problem* (3.1)*, we must have $F(\lambda_{\max}) < \infty$ and*

$$-F^*(\psi_{\max}) = -\langle \lambda_{\min}, \psi_{\max} \rangle + F(\lambda_{\min})$$

or equivalently by [9, Theorem 23.5]*, $\psi_{\max} \in \partial F(\lambda_{\min})$.*

Proof. Let $(\gamma_{\min}, \lambda_{\min})$ and $(\varphi_{\max}, \psi_{\max})$ respectively be minimizing and maximizing pairs in the primal problem (1.3) and dual problem (3.1). By definition of the Legendre transform, we have

$$-F^*(\psi_{\max}) = \inf_{\lambda\in\Lambda}(-\langle\lambda, \psi_{\max}\rangle + F(\lambda)) \le -\langle\lambda_{\min}, \psi_{\max}\rangle + F(\lambda_{\min}).$$

For the opposite inequality, using Theorem 3.3 we have

$$\begin{aligned}\int_{X\times Y} c\, d\gamma_{\min} + F(\lambda_{\min}) &= \sup_{-\varphi-\psi\le c}\Big(-\int_X \varphi\, d\mu - F^*(\psi)\Big)\\ &= \sup_{-\varphi-\psi\le c}\Big(-\int_X \varphi\, d\mu + \inf_{\lambda\in\Lambda}(-\langle\lambda,\psi\rangle + F(\lambda))\Big)\\ &\le \sup_{-\varphi-\psi\le c}\Big(-\int_X \varphi\, d\mu - \langle\lambda_{\min},\psi\rangle\Big) + F(\lambda_{\min})\\ &\le \int_{X\times Y} c\, d\gamma_{\min} + F(\lambda_{\min}).\end{aligned}$$

Thus we have

$$\begin{aligned}-\int_X \varphi_{\max}\, d\mu - F^*(\psi_{\max}) &= \sup_{-\varphi-\psi\le c}\Big(-\int_X \varphi\, d\mu - F^*(\psi)\Big)\\ &= \sup_{-\varphi-\psi\le c}\Big(-\int_X \varphi\, d\mu - \langle\lambda_{\min},\psi\rangle + F(\lambda_{\min})\Big)\\ &\ge -\int_X \varphi_{\max}\, d\mu - \langle\lambda_{\min},\psi_{\max}\rangle + F(\lambda_{\min}),\end{aligned}$$

since $\varphi_{\max} \in L^1(\mu)$, this yields $-F^*(\psi_{\max}) = -\langle\lambda_{\min}, \psi_{\max}\rangle + F(\lambda_{\min})$. Since F^* is finite everywhere, we have $F(\lambda_{\max}) < \infty$. $\square$

Corollary 3.11. *Assume the same conditions on c as Lemma 3.10. Then:*

1. *If $(\gamma_{\min}, \lambda_{\min}) \in \Pi(\mu, \nu_{\lambda_{\min}}) \times \Lambda$ are a minimizing pair in the primal problem* (1.3)*, then any maximizer of* (3.1) *is a maximizer in the classical Kantorovich dual problem with source measure μ and target measure $\nu_{\lambda_{\min}}$.*
2. *If there is a $\lambda_{\min} \in \Lambda$ such that $(\hat\varphi_{\max}, \hat\psi_{\max})$ is a maximizer in the classical Kantorovich dual problem with source measure μ and target measure $\nu_{\lambda_{\min}}$ and $\hat\psi_{\max} \in \partial F(\lambda_{\min})$, then $(\hat\varphi_{\max}, \hat\psi_{\max})$ is also a maximizer in the dual problem* (3.1)*. Moreover, if $\gamma_{\min}$ is a minimizer of the classical Kantorovich problem* (1.4) *between μ and $\nu_{\lambda_{\min}}$, then $(\gamma_{\min}, \lambda_{\min})$ is a minimizer in* (1.3)*.*

Proof. (1) Let $(\gamma_{\min}, \lambda_{\min}) \in \Pi(\mu, \nu_{\lambda_{\min}}) \times \Lambda$ be a minimizing pair in the primal problem (1.3) and $(\varphi_{\max}, \psi_{\max})$ be a maximizer of (3.1). Then by Theorem 3.3 and Lemma 3.10, $\int_{X\times Y} c\, d\gamma_{\min} + F(\lambda_{\min}) = -\int_X \varphi_{\max}\, d\mu - F^*(\psi_{\max}) = -\int_X \varphi_{\max}\, d\mu - \langle\lambda_{\min}, \psi_{\max}\rangle + F(\lambda_{\min})$, proving the first claim.

(2) Now suppose $\lambda_{\min} \in \mathbb{R}^N$ and $(\hat\varphi_{\max}, \hat\psi_{\max})$ are as in claim (2) above; by [12, Theorem 4.1] there exists $\gamma_{\min} \in \Pi(\mu, \nu_{\lambda_{\min}})$ minimizing (1.4). By [12, Theorem 5.10 (i)] and [9, Theorem 23.5],

$$\begin{aligned}\int_{X\times Y} c\, d\gamma_{\min} + F(\lambda_{\min}) &= -\int_X \hat\varphi_{\max}\, d\mu - \langle\lambda_{\min}, \hat\psi_{\max}\rangle + F(\lambda_{\min})\\ &= -\int_X \hat\varphi_{\max}\, d\mu - F^*(\hat\psi_{\max}),\end{aligned}$$

hence $(\hat{\varphi}_{\max}, \hat{\psi}_{\max})$ maximizes (3.1). Theorem 3.3 then shows $(\gamma_{\min}, \lambda_{\min})$ is a minimizer in (1.3). $\square$

4. Relationship between dual and primal optimizers

In this section we work towards a sharp first order condition characterizing optimizers in (3.1) and (1.3), along with uniqueness of minimizers under some mild conditions.

Proposition 4.1. *Suppose c is lower semicontinuous, there exists a real valued upper semicontinuous function $a \in L^1(\mu)$ with $c(x, y_j) \geq a(x)$ for all x and j. Also let $(\gamma_{\min}, \lambda_{\min}) \in \Pi(\mu, \nu_{\lambda_{\min}}) \times \Lambda$ and $(\varphi_{\max}, \psi_{\max}) \in L^1(\mu) \times \mathbb{R}^N$ be extremizers respectively in (1.3) and (3.1). If $\lambda^j_{\min} > 0$ for some $1 \leq j \leq N$, then we must have $\big(\psi^{c^*c}_{\max}\big)^j = \psi^j_{\max}$ for that index j.*

Proof. Suppose $\lambda^j_{\min} > 0$ for some $1 \leq j \leq N$. Recall that by Proposition 3.5 (1) we must have $\varphi_{\max} \equiv \psi^{c^*}_{\max}$ on $\operatorname{spt}\mu$, and $\psi^k_{\max} \geq (\psi^{c^*c}_{\max})^k$ for all $1 \leq k \leq N$. Suppose by contradiction there is a strict inequality for the index j. Since $(\psi^{c^*}_{\max}, \psi^{c^*c}_{\max})$ is also a maximizer in (3.1) we would then obtain

$$\begin{aligned}
-F^*(\psi^{c^*c}_{\max}) &= -\langle \lambda_{\min}, \psi^{c^*c}_{\max} \rangle + F(\lambda_{\min}) \\
&= -\sum_{k \neq j} (\psi^{c^*c}_{\max})^k \lambda^k_{\min} - (\psi^{c^*c}_{\max})^j \lambda^j_{\min} + F(\lambda_{\min}) \\
&> -\sum_{k \neq j} \psi^k_{\max} \lambda^k_{\min} - \psi^j_{\max} \lambda^j_{\min} + F(\lambda_{\min}) \\
&= -\langle \lambda_{\min}, \psi_{\max} \rangle + F(\lambda_{\min}) = -F^*(\psi_{\max}).
\end{aligned}$$

However, this contradicts that $(\varphi_{\max}, \psi_{\max})$ is a maximizer, thus we must have $\psi_{\max} = (\psi^{c^*c}_{\max})$. $\square$

Next we aim to start with a maximizer in the dual problem and construct a minimizer in the primal problem. To do so, we need an assumption about when minimizers in the classical Kantorovich problem (1.4) can be written as solutions to the Monge problem (1.1).

Definition 4.2. If $\mu \in \mathcal{P}(X)$, we say the cost function c satisfies the condition (μ-Twist) if for each $1 \leq j \leq N$, the function $c(\cdot, y_j) \in L^1(\mu)$ is lower semicontinuous, there exists a real valued upper semicontinuous function $a \in L^1(\mu)$ with $c(x, y_j) \geq a(x)$ for all $x \in X$, and for any $u \in \mathbb{R}$, and each $j \neq k$, we have

$$\mu(\{x \in X : c(x, y_j) = c(x, y_k) + u\}) = 0. \tag{μ-Twist}$$

Remark 4.3. If c satisfies (μ-Twist) then we can apply [12, Theorem 5.30] to find the Kantorovich problem (1.4) has a unique solution for any choice of $\nu = \nu_\lambda$, $\lambda \in \Lambda$, which can be written in the form $(\mathrm{Id} \times T)_\# \mu$ where T is a mapping defined μ-a.e., that is in turn a solution to the Monge problem (1.1). In particular, under these conditions, a solution γ of (1.4) must be supported on the graph of a mapping from X to Y that is single valued μ-a.e. We note that when X is a subset of a smooth Riemannian manifold and μ is absolutely continuous with respect to the Riemannian volume, (μ-Twist) holds under what is usually referred to as the "twist" or "bi-twist" condition (see [12, p. 234]).

Now let $\psi \in \mathbb{R}^N$ and define $\lambda \in \mathbb{R}^N$ by

$$\lambda^j = \mu\Big(\{x \in X : \psi^{c^*}(x) = -c(x, y_j) - \psi^j\}\Big)$$

for each j. When c satisfies (μ-Twist), it is clear that for $j_1 \neq j_2$,

$$\begin{aligned} &\mu(\{x \in X : \psi^{c^*}(x) = -c(x, y_{j_1}) - \psi^{j_1}\} \cap \{x \in X : \psi^{c^*}(x) = -c(x, y_{j_2}) - \psi^{j_2}\}) \\ &\le \mu(\{x \in X : -c(x, y_{j_1}) - \psi^{j_1} = -c(x, y_{j_2}) - \psi^{j_2}\}) = 0, \end{aligned}$$

thus in particular $\lambda \in \Lambda$, and for μ-a.e. x there is a unique index j such that $\psi^{c^*}(x) = -c(x, y_j) - \psi^j$. Defining $T_\psi : X \to Y$ by $T_\psi(x) = y_j$ whenever j is the unique index associated to x, it is clear that $(T_\psi)_\# \mu = \nu_\lambda$, hence by [12, Remark 5.13], we can see that $\gamma_\lambda := (\mathrm{Id} \times T_\psi)_\# \mu$ is a solution to the Kantorovich problem (1.4) with $\nu = \nu_\lambda$.

Proposition 4.4. *Suppose c satisfies* (μ-Twist). *Also suppose $(\hat{\varphi}, \hat{\psi})$ is a maximizing pair in the dual problem* (3.1), *define $\tilde{\lambda} \in \Lambda$ by*

$$\tilde{\lambda}^j := \mu\Big(\{x \in X : \hat{\varphi}(x) = -c(x, y_j) - \hat{\psi}^j\}\Big),$$

and take $\gamma_{\min} \in \Pi(\mu, \nu_{\tilde{\lambda}})$ to be the (unique) solution of the classical Kantorovich problem (1.4) *with $\nu = \nu_{\tilde{\lambda}}$. Then $(\gamma_{\min}, \tilde{\lambda})$ is a minimizing pair in the primal problem* (1.3).

Proof. Let $(\hat{\varphi}, \hat{\psi})$ be a maximizing pair in the dual problem. By Proposition 3.5 we see that $\hat{\varphi} \equiv \hat{\psi}^{c^*}$, and we easily see that replacing $\hat{\psi}$ with $\hat{\psi}^{c^* c}$ does not change the vector $\tilde{\lambda}$, so we make this replacement.

Since $-\hat{\varphi}(x) - \hat{\psi}^j \le c(x, y_j)$ for all x, j, by Kantorovich duality in the classical optimal transport problem [11, Theorem 1.3], we have for any $\lambda \in \Lambda$ that

$$\ell(\lambda) := -\int_X \hat{\varphi}\, d\mu - \langle \lambda, \hat{\psi} \rangle \le \min_{\gamma \in \Pi(\mu, \nu_\lambda)} \int_{X \times Y} c\, d\gamma =: \mathcal{C}(\lambda) < \infty,$$

where finiteness comes from $c(\cdot, y_j) \in L^1(\mu)$ for each j. At the same time by strong duality, Theorem 3.3,

$$\begin{aligned} \inf_{\lambda \in \Lambda} [F(\lambda) + \ell(\lambda)] &= -\int_X \hat{\varphi}\, d\mu + \inf_{\lambda \in \Lambda} [F(\lambda) - \langle \lambda, \hat{\psi} \rangle] \\ &= -\int_X \hat{\varphi}\, d\mu - F^*(\hat{\psi}) \\ &= \min_{\lambda \in \Lambda,\ \gamma \in \Pi(\mu, \nu_\lambda)} \Big(\int_{X \times Y} c\, d\gamma + F(\lambda) \Big) \\ &= \min_{\lambda \in \Lambda} [F(\lambda) + \mathcal{C}(\lambda)]. \end{aligned}$$

Thus we obtain that $F + \ell \le F + \mathcal{C}$ pointwise everywhere on Λ, and the above calculation shows that $F + \ell$ attains its minimum value over Λ, at the same point as $F + \mathcal{C}$; say this point is $\lambda_{\min}$.

Arguing as in the proof of [10, Proposition 7.19] (which can be carried out under the assumption (μ-Twist), note the exact form of the cost function is immaterial), we find that $\mathcal{C}$ is strictly convex on Λ. By Remark 4.3 and the choice of $\tilde{\lambda}$, we have $\mathcal{C}(\tilde{\lambda}) = \ell(\tilde{\lambda})$, hence for any $t \in [0, 1]$ we must have

$$\ell((1-t)\lambda_{\min} + t\tilde{\lambda}) \le \mathcal{C}((1-t)\lambda_{\min} + t\tilde{\lambda})$$

$$
\begin{aligned}
&\le (1-t)\mathcal{C}(\lambda_{\min}) + t\mathcal{C}(\tilde\lambda)\\
&= (1-t)\ell(\lambda_{\min}) + t\ell(\tilde\lambda)\\
&= \ell((1-t)\lambda_{\min} + t\tilde\lambda),
\end{aligned}
$$

i.e., $\ell \equiv \mathcal{C}$ on the segment $[\lambda_{\min}, \tilde\lambda]$.

However the only way for the strictly convex function $\mathcal{C}$ to equal an affine function on $[\lambda_{\min}, \tilde\lambda]$ is if $\lambda_{\min} = \tilde\lambda$. It is then clear that $(\tilde\gamma, \tilde\lambda)$ is a minimizer in the primal problem (1.3). □

The strict convexity of $\mathcal{C}$ demonstrated in the above proof immediately yields the following corollary.

Corollary 4.5. *If c satisfies (μ-Twist), minimizers in the primal problem (1.3) are unique.*

Using the above properties of dual and primal optimizers, we obtain a characterization for optimizers in both problems.

Theorem 4.6. *Assume c satisfies (μ-Twist). If $(\hat\varphi, \hat\psi)$ is a maximizing pair in the dual problem (3.1) and $(\tilde\gamma, \tilde\lambda)$ is a minimizer in the primal problem (1.3), $\hat\psi$ and $\tilde\lambda$ satisfy the conditions:*

(i) $\hat\psi \in \partial F(\tilde\lambda)$

(ii) $\tilde\lambda^j = \mu\big(\{x \in X : -c(x, y_j) - \hat\psi^j = \hat\psi^{c^*}(x)\}\big)$.

Furthermore, if $\tilde\lambda^j > 0$ for some $1 \le j \le N$, we have

(iii) $\hat\psi^j = \big(\hat\psi^{c^* c}\big)^j$.

Conversely, if $\tilde\lambda \in \Lambda$ and $\hat\psi \in \mathbb{R}^N$ are such that conditions (i) and (ii) hold, then defining $T_{\hat\psi}$ as in Remark 4.3, the pairs $(\hat\psi^{c^*}, \hat\psi)$ and $((\mathrm{Id} \times T_{\hat\psi})_{\#}\mu, \tilde\lambda)$ are maximizing and minimizing pairs in the dual and primal problem respectively.

Proof. Under the hypotheses above "conversely", conditions (i) and (ii) follow from Lemma 3.10, Proposition 4.4, and Corollary 4.5. Condition (iii) follows from Proposition 4.1.

Now suppose $\hat\psi \in \mathbb{R}^N$, $\tilde\lambda \in \Lambda$ satisfy (i) and (ii). Then

$$
\begin{aligned}
\sup_{-\varphi-\psi\le c}\Big(-\int_X \varphi\, d\mu - F^*(\psi)\Big) &\ge -\int_X \hat\psi^{c^*}\, d\mu - F^*(\hat\psi)\\
&= -\int_X \hat\psi^{c^*}\, d\mu - \langle\tilde\lambda, \hat\psi\rangle + F(\tilde\lambda)
\end{aligned}
\tag{4.1}
$$

where this last equality follows from condition (i) and [9, Theorem 23.5]. Let $T_{\hat\psi}$ be defined as in Remark 4.3, by condition (ii), we see that $\tilde\gamma := (\mathrm{Id} \times T_{\hat\psi})_{\#}\mu$ is a minimizer in the classical Kantorovich problem (1.4) with $\nu = \nu_{\tilde\lambda}$. Let $x \in X$ be such that $T_{\hat\psi}(x)$ is well-defined. By definition, this means that $\hat\psi^{c^*}(x) + \hat\psi^j = -c(x, y_j)$ where $y_j = T_{\hat\psi}(x)$. Since (see Remark 4.3) the set of such x has full μ measure, the union of $(x, T_{\hat\psi}(x))$ over such x has full $\tilde\gamma$ measure. Thus by [12, Theorem 5.10 and Remark 5.13], we have that $(\hat\psi^{c^*}, \hat\psi)$ is a maximizer in the classical Kantorovich dual problem, and in particular $-\int_X \hat\psi^{c^*}\, d\mu - \langle\tilde\lambda, \hat\psi\rangle =$

$\inf_{\gamma\in\Pi(\mu,\nu_{\tilde\lambda})}\Big(\int_{X\times Y} c\,d\gamma\Big)$. Thus we can calculate,

$$\begin{aligned}
-\int_X \hat\psi^{c^*}\,d\mu-\langle\tilde\lambda,\hat\psi\rangle+F(\tilde\lambda)&=\inf_{\gamma\in\Pi(\mu,\nu_{\tilde\lambda})}\Big(\int_{X\times Y} c\,d\gamma\Big)+F(\tilde\lambda)\\
&\geq \inf_{\lambda\in\Lambda,\ \gamma\in\Pi(\mu,\nu_\lambda)}\Big(\int_{X\times Y} c\,d\gamma+F(\lambda)\Big)\\
&\geq \sup_{-\varphi-\psi\leq c}\Big(-\int_X\varphi\,d\mu-F^*(\psi)\Big).
\end{aligned}$$

Combining this with (4.1), we have

$$\begin{aligned}
\int_{X\times Y} cd\tilde\gamma+F(\tilde\lambda)&=\inf_{\gamma\in\Pi(\mu,\nu_{\tilde\lambda})}\Big(\int_{X\times Y} c\,d\gamma\Big)+F(\tilde\lambda)\\
&=\inf_{\lambda\in\Lambda,\gamma\in\Pi(\mu,\nu_\lambda)}\Big(\int_{X\times Y} c\,d\gamma+F(\lambda)\Big)
\end{aligned}$$

hence $(\tilde\gamma,\tilde\lambda)$ is a minimizing pair in the primal problem. The above calculations also yield

$$\sup_{-\varphi-\psi\leq c}\Big(-\int_X\varphi\,d\mu-F^*(\psi)\Big)=-\int_X\hat\psi^{c^*}\,d\mu-F^*(\hat\psi),$$

thus $(\hat\psi^{c^*},\hat\psi)$ is a maximizing pair in the dual problem. □

5. Stability of F

In this section we show two kinds of stability for our primal problem (1.3). The first is stability of minimizers under perturbations of the storage fee function F when the finite target set Y is fixed. The second is Γ-convergence of a sequence of objective functionals obtained when one takes a sequence of target sets Y_k that become suitably dense in some continuous domain Y.

5.1. **Stability for a fixed finite target.**

First we estimate the change in the minimum value of the problem.

Proposition 5.1. *Let F_1 and $F_2:\mathbb{R}^N\to\mathbb{R}\cup\{\infty\}$ be lower semicontinuous and proper, c be lower semicontinuous and bounded, and write $\mathfrak{m}_{F_i}$ for the minimum value attained in* (1.3) *with some fixed measure μ and the choice $F=F_i$, $i=1$ or $i=2$. Then if both of $\mathfrak{m}_{F_i}$ are finite,*

$$\|\mathfrak{m}_{F_1}-\mathfrak{m}_{F_2}|\leq\|F_1-F_2\|_{L^\infty(\mathrm{dom}(F_1)\cup\mathrm{dom}(F_2))}$$

Proof. Let the pair $(\tilde\gamma_2,\tilde\lambda_2)$ achieve the minimum value in $\mathfrak{m}_{F_2}$, which exists by Theorem 2.3. Then both $\int_{X\times Y} cd\tilde\gamma_2$ and $F_2(\tilde\lambda_2)$ are finite. If $F_1(\tilde\lambda_2)=\infty$ we have $\|F_1-F_2\|_{L^\infty(\mathrm{dom}(F_1)\cup\mathrm{dom}(F_2))}=\infty$ and the desired inequality is trivial, thus we may assume $F_1(\tilde\lambda_2)$ is finite. Then

$$\begin{aligned}
\mathfrak{m}_{F_1}-\mathfrak{m}_{F_2}&\leq\Big(\int_{X\times Y} cd\tilde\gamma_2+F_1(\tilde\lambda_2)\Big)-\Big(\int_{X\times Y} cd\tilde\gamma_2+F_2(\tilde\lambda_2)\Big)\\
&=F_1(\tilde\lambda_2)-F_2(\tilde\lambda_2)\\
&\leq\|F_1-F_2\|_{L^\infty(\mathrm{dom}(F_1)\cup\mathrm{dom}(F_2))}.
\end{aligned}$$

The same argument reversing the roles of F_1 and F_2 completes the proof. □

The above statement shows that if F_k converges to F uniformly, then $\mathfrak{m}_{F_k}$ converges to $\mathfrak{m}_F$. Next we prove that the minimizing plans weakly converge to a minimizer of the original problem.

Theorem 5.2. *Suppose c satisfies* (μ-Twist). *Let $(\tilde{\gamma}, \tilde{\lambda})$ and $(\tilde{\gamma}_k, \tilde{\lambda}_k)$ minimize* (1.3) *with storage fee functions F and F_k for each k respectively, where F, F_k are all proper, convex functions with compact essential domains contained in Λ. If*

$$\lim_{k\to\infty} \|F_k - F\|_{L^\infty(\mathrm{dom}(F_k)\cup \mathrm{dom}(F))} = 0,$$

then $\tilde{\lambda}_k$ converges to $\tilde{\lambda}$, and $\tilde{\gamma}_k$ converges weakly to $\tilde{\gamma}$.

Proof. Let $\mathcal{C}_F(\lambda) = \inf_{\gamma\in\Pi(\mu,\nu_\lambda)} \int_{X\times Y} c\, d\gamma + F(\lambda)$ and define $\mathcal{C}_{F_k}(\lambda)$ analogously. By the proof of Proposition 4.4 and Corollary 4.5, we see that $\mathcal{C}_F$ and $\mathcal{C}_{F_k}$ are strictly convex functions on Λ each of which have unique minimizers, given by $\tilde{\lambda}$ and $\tilde{\lambda}_k$ respectively. By (μ-Twist) and the properness of F and F_k, all $\mathcal{C}_{F_k}(\tilde{\lambda}_k)$ and $\mathcal{C}_F(\tilde{\lambda})$ are finite, then by Proposition 5.1,

$$\begin{aligned}
\big|\mathcal{C}_F(\tilde{\lambda}_k) - \mathcal{C}_F(\tilde{\lambda})\big| &\le |\mathcal{C}_{F_k}(\tilde{\lambda}_k) - \mathcal{C}_F(\tilde{\lambda})| + |\mathcal{C}_{F_k}(\tilde{\lambda}_k) - \mathcal{C}_F(\tilde{\lambda}_k)| \\
&= |\mathfrak{m}_{F_k} - \mathfrak{m}_F| + |F_k(\tilde{\lambda}_k) - F(\tilde{\lambda}_k)| \\
&\le 2\|F_k - F\|_{L^\infty(\mathrm{dom}(F_k)\cup\mathrm{dom}(F))} \to 0, \quad k\to\infty.
\end{aligned}$$

By the compactness of Λ, any subsequence of $\{\tilde{\lambda}_k\}_{k=1}^\infty$ has a convergent subsequence, by the above calculation and strict convexity of $\mathcal{C}_F$ on Λ all of these subsequential limits must be $\tilde{\lambda}$, hence we must have $\lim_{k\to\infty} \tilde{\lambda}_k = \tilde{\lambda}$.

Now suppose by contradiction that $\tilde{\gamma}_k$ does not converge weakly to $\tilde{\gamma}$. Since $\Pi(\mu)$ is weakly compact by Lemma 2.2, we can extract a subsequence (which we do not relabel) which converges weakly to some limiting measure that is not $\tilde{\gamma}$, say $\hat{\gamma}$. By the above paragraph combined with (2.1) we have $\tilde{\lambda} = \lim_{k\to\infty} \tilde{\lambda}_k = \hat{\lambda}$ where $\hat{\lambda}$ is such that the right marginal of $\hat{\gamma}$ is $\nu_{\hat{\lambda}}$. We then have

$$\begin{aligned}
&\int c d\hat{\gamma} + F(\hat{\lambda}) \\
&= \int c d\tilde{\gamma}_k + F_k(\tilde{\lambda}_k) + \Big(F(\hat{\lambda}) - F(\tilde{\lambda}_k)\Big) + \Big(F(\tilde{\lambda}_k) - F_k(\tilde{\lambda}_k)\Big) + \Big(\int c d\hat{\gamma} - \int c d\tilde{\gamma}_k\Big) \\
&\le \mathfrak{m}_{F_k} + \Big(F(\hat{\lambda}) - F(\tilde{\lambda}_k)\Big) + \|F_k - F\|_{L^\infty(\mathrm{dom}(F_k)\cup\mathrm{dom}(F))} + \Big(\int c d\hat{\gamma} - \int c d\tilde{\gamma}_k\Big).
\end{aligned}$$

Letting k go to infinity we see that $\int c d\hat{\gamma} + F(\hat{\lambda}) \le \mathfrak{m}_F$ by Proposition 5.1, the lower semicontinuity of F, and the fact that $\tilde{\gamma}_k$ converges weakly to $\hat{\gamma}$. Hence $\hat{\gamma}$ is a minimizer and by Corollary 4.5 we see that $\hat{\gamma} = \tilde{\gamma}$ as desired. $\square$

Γ-convergence. We now turn to Γ-convergence. For this subsection, we assume (X, d_X) is a Polish space which may not be compact, and we will suppose (Y, d_Y) is a bounded Polish space, not necessarily finite.

Definition 5.3. Recall that if Ω is a first countable topological space and $G_k : \Omega \to \mathbb{R}\cup\{\pm\infty\}$ is a sequence of functions, we say the G_k Γ-*converge to* $G : \Omega \to \mathbb{R}\cup\{\pm\infty\}$ if

(1) $G(\omega) \le \liminf_{k\to\infty} G_k(\omega_k)$ whenever $\lim_{k\to\infty} \omega_k = \omega$.

(2) For any $\omega \in \Omega$, there exists a *recovery sequence* $\{\omega_k\}_{k=1}^{\infty}$ s.t. $\lim_{k\to\infty} \omega_k = \omega$ and
$G(\omega) \geq \limsup_{k\to\infty} G_k(\omega_k)$.

Before stating our second stability result, we recall some classical definitions in optimal transport theory.

Definition 5.4. If (X, d_X) is a Polish space, we write for the set of probability measures with finite moment,

$$\mathcal{P}_1(X) := \{\mu \in \mathcal{P}(X) : \exists x_0 \in X \text{ with } \int_X d_X(x, x_0)\, d\mu(x) < \infty\}.$$

Also for $\mu_1, \mu_2 \in \mathcal{P}_1(X)$, we write

$$W_1^{d_X}(\mu_1, \mu_2) := \inf_{\gamma \in \Pi(\mu_1,\mu_2)} \int_{X\times X} d_X(x_1, x_2)\, d\gamma(x_1, x_2).$$

It is well known that $W_1^{d_X}$ is a metric on $\mathcal{P}_1$ (see [12, Chapter 6]).

Our second stability result is as follows.

Theorem 5.5. *Suppose there is a sequence of cost functions $c_k \in C_b(X \times Y)$ converging uniformly to some $c \in C_b(X \times Y)$, and $\mathcal{F} : \mathcal{P}(Y) \to \mathbb{R} \cup \{\pm\infty\}$ is sequentially weakly continuous. Also suppose for each $k \in \mathbb{N}$, the finite set $Y_k := \{y_{k,j}\}_{j=1}^{N_k} \subset Y$ is an ϵ_k-net of Y (i.e., for any $y \in Y$ there is a $y_{k,j}$ such that $d_Y(y, y_{k,j}) < \epsilon_k$) with $\epsilon_k \searrow 0$, and define the sets*

$$\Pi_k(\mu) := \{\gamma \in \Pi(\mu) : \operatorname{spt}((\pi_Y)_\#\gamma) \subset Y_k\}.$$

Also define the functionals $\mathcal{C}_k$, $\mathcal{C} : \mathcal{P}(X \times Y) \to \mathbb{R} \cup \{\pm\infty\}$ by

$$\mathcal{C}_k(\gamma) := \begin{cases} \int_{X\times Y} c_k\, d\gamma + \mathcal{F}((\pi_Y)_\#\gamma), & \gamma \in \Pi_k(\mu), \\ \infty, & \textit{otherwise} \end{cases}$$

$$\mathcal{C}(\gamma) := \begin{cases} \int_{X\times Y} c\, d\gamma + \mathcal{F}((\pi_Y)_\#\gamma), & \gamma \in \Pi(\mu), \\ \infty, & \textit{otherwise} \end{cases}$$

Then

(1) *$\mathcal{C}_k$ Γ-converges to $\mathcal{C}$, where the underlying topology on $\mathcal{P}(X \times Y)$ is that of weak convergence of measures.*
(2) *Suppose $\{\mu_k\}_{k=1}^{\infty} \subset \mathcal{P}_1(X)$ converges in $W_1^{d_X}$ to $\mu \in \mathcal{P}_1(X)$. Then if each $\Pi_k(\mu)$ is replaced by $\Pi_k(\mu_k)$ in the definition of $\mathcal{C}_k$, again $\mathcal{C}_k$ Γ-converges to $\mathcal{C}$.*

Remark 5.6. At first glance, the condition of weak continuity of $\mathcal{F}$ may seem restrictive. However, by [4, Remark 1.25 and Proposition 1.28], if $\mathcal{C}_k$ Γ-converges to $\mathcal{C}$, we must have that $\mathcal{F}$ is weakly lower semicontinuous. At the same time, considering the example where $\mathcal{F}$ is 0 for some fixed measure which is not discrete, and 1 otherwise shows that weak lower semicontinuity is not enough to produce a recovery sequence to obtain Γ-convergence. Essentially, one would need that for any measure ν, there is a sequence of discrete measures weakly converging to ν along which $\mathcal{F}$ is continuous, which is not so far off from simply requiring weak continuity.

Before giving the proof, we recall a form of disintegration of measures which follows from [7, Chapter III-70 and 72].

Theorem 5.7 (Disintegration of measures)**.** *If X_1, X_2 are Polish spaces, for any $\rho \in \mathcal{P}(X_1 \times X_2)$ there is a map $X_1 \ni x \mapsto \rho^x \in \mathcal{P}(X_2)$, defined uniquely $(\pi_{X_1})_\# \rho$-a.e., such that for any bounded or nonnegative Borel function φ on $X_1 \times X_2$, the map $X_1 \ni x_1 \mapsto \int_{X_2} \varphi(x_1, x_2) d\rho^{x_1}(x_2)$ is Borel, and*

$$\int_{X_1 \times X_2} \varphi d\rho = \int_{X_1} \int_{X_2} \varphi(x_1, x_2) d\rho^{x_1}(x_2) d((\pi_{X_1})_\# \rho)(x_1).$$

Proof of Theorem 5.5. First suppose a sequence γ_k weakly converges to some $\gamma \in \mathcal{P}(X \times Y)$. If $(\pi_X)_\# \gamma \neq \mu$, since the sequence of left marginals $(\pi_X)_\# \gamma_k$ converge weakly to $(\pi_X)_\# \gamma$, we must have that $(\pi_X)_\# \gamma_k \neq \mu$ (or $\neq \mu_k$ in case (2)) for all k sufficiently large. Thus in this case we have

$$\liminf_{k \to \infty} \mathcal{C}_k(\gamma_k) = \infty \geq \mathcal{C}(\gamma).$$

If $(\pi_X)_\# \gamma = \mu$, since the sequence of right marginals $(\pi_Y)_\# \gamma_k$ converge weakly to $(\pi_Y)_\# \gamma$,

$$\begin{aligned} \liminf_{k \to \infty} \mathcal{C}_k(\gamma_k) &= \liminf_{k \to \infty} \Big(\int_{X \times Y} c_k \, d\gamma_k + \mathcal{F}((\pi_Y)_\# \gamma_k) \Big) \\ &\geq \int_{X \times Y} c \, d\gamma + \mathcal{F}((\pi_Y)_\# \gamma) = \mathcal{C}(\gamma). \end{aligned}$$

Next we fix a $\gamma \in \mathcal{P}(X \times Y)$, we aim to produce a recovery sequence. If the left marginal of γ is not μ, we have $\mathcal{C}(\gamma) = \infty$ and we can take the constant sequence $\{\gamma\}$ as the recovery sequence, thus assume $(\pi_X)_\# \gamma = \mu$. For each $k \in \mathbb{N}$ and $1 \leq j \leq N_k$, we define $\tilde{K}_{k,j}$ as the Voronoi cell associated with $y_{k,j}$, that is, $\tilde{K}_{k,j} := \{y \in Y : d_Y(y, y_{k,j}) \leq \min_{1 \leq i \leq N_k} d_Y(y, y_{k,i})\}$. Then, we define

$$K_{k,1} := \tilde{K}_{k,1}, \quad K_{k,j} := \tilde{K}_{k,j} \setminus \cup_{i=1}^{j-1} K_{k,i}, \quad 2 \leq j \leq N_k,$$

and the maps $\varphi_k : Y \to Y_k$ by $\varphi_k(x) := y_{k,j}$ whenever $x \in K_{k,j}$, this map is well-defined by the disjointness of the collection $\{K_{k,j}\}_{j=1}^{N_k}$ for each fixed k.

We will first consider case (1). Define $\gamma_k := (\pi_X \times (\varphi_k \circ \pi_Y))_\# \gamma \in \mathcal{P}(X \times Y)$. Then $(\pi_X)_\# \gamma_k = (\pi_X)_\# \gamma = \mu$ for each k, and since the range of φ_k is contained in Y_k, it is clear that $\operatorname{spt}((\pi_Y)_\# \gamma_k) \subset Y_k$ hence $\gamma_k \in \Pi_k(\mu)$. Now suppose $\varphi \in C_b(X \times Y)$ is L-Lipschitz for some $L > 0$, and fix $y \in Y$ with $y \in K_{i,j_1(k)}$ for some index $j_1(k)$. Since Y_k is an ϵ_k net of Y, for some $y_{i,j_2(k)} \in Y_k$ we have $d_Y(y, y_{i,j_2(k)}) \leq \epsilon_k$. Hence for any $x \in X$,

$$\begin{aligned} |\varphi(x, \varphi_k(y)) - \varphi(x, y)| &\leq L d_Y(y, \varphi_k(y)) = L d_Y(y, y_{i,j_1(k)}) \\ &\leq L d_Y(y, y_{i,j_2(k)}) < L\epsilon_k, \end{aligned} \tag{5.1}$$

which proves $\lim_{k \to \infty} \varphi(x, \varphi_k(y)) = \varphi(x, y)$ for all $(x, y) \in X \times Y$ where the convergence is uniform over $X \times Y$. Since γ is a probability measure, we have

$$\lim_{k \to \infty} \int_{X \times Y} \varphi(x, y) \, d\gamma_k(x, y) = \lim_{k \to \infty} \int_{X \times Y} \varphi(x, \varphi_k(y)) \, d\gamma(x, y) = \int_{X \times Y} \varphi \, d\gamma.$$

As it is sufficient to test for convergence against bounded, Lipschitz functions (see [3, Remark 8.3.1]) this proves weak convergence of γ_k to γ; in particular it also proves weak convergence of the $(\pi_Y)_\# \gamma_k$ to $(\pi_Y)_\# \gamma$. Thus the uniform convergence of the c_k and weak continuity of $\mathcal{F}$ gives that

$$\limsup_{k \to \infty} \mathcal{C}_k(\gamma_k) = \limsup_{k \to \infty} \Big(\int_{X \times Y} c_k \, d\gamma_k + \mathcal{F}((\pi_Y)_\# \gamma_k) \Big) \leq \mathcal{C}(\gamma).$$

Now we consider case (2). Since μ and all μ_k have finite first moments, there exist $\sigma_k \in \Pi(\mu, \mu_k)$ which achieve the minimum in the definition of $W_1^{d_X}(\mu, \mu_k)$ (say, by [12, Theorem 4.1]). It is clear that the unique minimizer achieving $W_1^{d_X}(\mu, \mu) = 0$ is $(\mathrm{Id} \times \mathrm{Id})_\# \mu$, while for some $x_0 \in X$,

$$\sup_k \int_{X\times X} d_X(x_1, x_2)\, d\sigma_k(x_1, x_2) \le \sup_k \int_X d_X(x, x_0)\, d\mu_k(x) + \int_X d_X(x, x_0)\, d\mu(x) < \infty,$$

hence by [12, Theorem 5.20] we see σ_k converges weakly to $(\mathrm{Id} \times \mathrm{Id})_\# \mu$. Now if σ_k^x and γ^x are the disintegrations of σ_k and γ respectively, with respect to their left marginals (both μ) as given in Theorem 5.7, we define a linear functional γ_k acting on bounded Borel functions φ on $X \times Y$ by

$$\gamma_k(\varphi) := \int_X \int_Y \int_X \varphi(z, y)\, d\sigma_k^x(z) d((\varphi_k)_\# \gamma^x)(y)\, d\mu(x).$$

We claim that $\gamma_k \in \mathcal{P}(X \times Y)$. Fix $\epsilon > 0$. If $\psi \in C_b(Y)$ is L-Lipschitz, by (5.1) we have $\big| \int_Y \psi d((\varphi_k \circ \pi_Y)_\# \gamma) - \int_Y \psi d((\pi_Y)_\# \gamma) \big| \le L\epsilon_k$, and again by [3, Remark 8.3.1] we see $(\varphi_k \circ \pi_Y)_\# \gamma$ weakly converges to $(\pi_Y)_\# \gamma$. Since $\{\mu_k\}_{i=1}^\infty$ is also weakly convergent, for any $\epsilon > 0$, by Prokhorov's theorem there are compact sets $K_1 \subset X$ and $K_2 \subset Y$ such that $\mu_k(X \setminus K_1) < \epsilon$ and $(\varphi_k \circ \pi_Y)_\# \gamma(Y \setminus K_2) < \epsilon$ for all k. Then, if $\varphi \in C_b(X \times Y)$ is identically zero on the compact set $K := K_1 \times K_2$, we have

$$\begin{aligned}
\gamma_k(\varphi) &\le \sup |\varphi| \int_X \int_{Y\setminus K_2} \int_{X\setminus K_1} d\sigma_k^x(z) d((\varphi_k)_\# \gamma^x)(y)\, d\mu(x) \\
&= \sup |\varphi| \int_X \sigma_k^x(X \setminus K_1) \gamma^x(\varphi_k^{-1}(Y \setminus K_2))\, d\mu(x) \\
&\le \sup |\varphi| \Big(\int_X \sigma_k^x(X \setminus K_1)^2\, d\mu(x) \Big)^{1/2} \Big(\int_X \gamma^x(\varphi_k^{-1}(Y \setminus K_2))^2\, d\mu(x) \Big)^{1/2} \\
&\le \sup |\varphi| \Big(\int_X \sigma_k^x(X \setminus K_1)\, d\mu(x) \Big)^{1/2} \Big(\int_X \gamma^x(\varphi_k^{-1}(Y \setminus K_2))\, d\mu(x) \Big)^{1/2} \\
&= \sup |\varphi| \sqrt{\sigma_k(X \times (X \setminus K_1))} \sqrt{\gamma(X \times \varphi_k^{-1}(Y \setminus K_2))} \\
&= \sup |\varphi| \sqrt{\mu_k(X \setminus K_1)} \sqrt{((\varphi_k \circ \pi_Y)_\# \gamma)(Y \setminus K_2)} \\
&\le \epsilon \sup |\varphi|,
\end{aligned}$$

where we have used Hölder's inequality for the third line, and that σ_k^x and γ^x are probability measures to obtain the inequality in the fourth line. Hence by [3, Theorem 7.10.6], γ_k is a Radon measure on $X \times Y$, and it is clear that $\gamma_k \ge 0$ with $\gamma_k(X \times Y) = 1$, hence the claim is proved. Next, for any Borel $\psi : X \to [0, \infty]$ we have

$$\begin{aligned}
\int_{X\times Y} \psi(x)\, d\gamma_k(x, y) &= \int_X \int_Y \int_X \psi(z)\, d\sigma_k^x(z) d((\varphi_k)_\# \gamma^x)(y)\, d\mu(x) \\
&= \int_X \int_X \psi(z)\, d\sigma_k^x(z)\, d\mu(x) \\
&= \int_{X\times X} \psi(x_2)\, d\sigma_k(x_1, x_2) \\
&= \int_X \psi(x)\, d\mu_k(x),
\end{aligned}$$

hence the left marginal of γ_k is μ_k and we have $\gamma_k \in \Pi_k(\mu_k)$, in particular $\mathcal{C}_k(\gamma_k)$ is finite. Finally suppose $\varphi \in C_b(X \times Y)$ is L-Lipschitz. Then by (5.1),

$$
\begin{aligned}
&\Big| \int_{X\times Y} \varphi \, d\gamma_k - \int_{X\times Y} \varphi \, d\gamma \Big| \\
&\le \Big| \int_X \int_Y \int_X \varphi(z, \varphi_k(y)) \, d\sigma_k^x(z) \, d\gamma^x(y) \, d\mu(x) \\
&\quad - \int_X \int_Y \int_X \varphi(z, y) \, d\sigma_k^x(z) \, d\gamma^x(y) \, d\mu(x) \Big| \\
&\quad + \Big| \int_X \int_Y \int_X \varphi(z, y) \, d\sigma_k^x(z) \, d\gamma^x(y) \, d\mu(x) - \int_X \int_Y \varphi(x, y) \, d\gamma^x(y) \, d\mu(x) \Big| \\
&\le \int_X \int_Y \int_X |\varphi(z, \varphi_k(y)) - \varphi(z, y)| \, d\sigma_k^x(z) \, d\gamma^x(y) \, d\mu(x) \\
&\quad + \int_X \int_Y \int_X |\varphi(z, y) - \varphi(x, y)| \, d\sigma_k^x(z) \, d\gamma^x(y) \, d\mu(x) \\
&\le L\epsilon_k + L \int_X \int_Y \int_X d_X(z, x) \, d\sigma_k^x(z) \, d\gamma^x(y) \, d\mu(x) \\
&= L(\epsilon_k + \int_{X\times X} d_X(z, x) \, d\sigma_k(z, x)) \\
&= L(\epsilon_k + W_1^{d_X}(\mu_k, \mu)) \to 0
\end{aligned}
$$

as $k \to \infty$. Thus as in case (1), we see γ_k weakly converges to γ. A similar argument yields $\limsup_{k\to\infty} \mathcal{C}_k(\gamma_k) \le \mathcal{C}(\gamma)$. This completes the proof of Γ-convergence in both cases. $\square$

6. Appendix

Proof of Proposition 3.7. Let us write $R := \min_{i\neq j} |y_i - y_j|$. Let $\lambda_\infty \in \Lambda$ be such that $F(\lambda_\infty) - F^{**}(\lambda_\infty) = \|F - F^{**}\|_{L^\infty(\Lambda)}$, and define

$$
\mu = \sum_{i=1}^N \frac{\lambda_\infty^i \mathbb{1}_{B_{\frac{R}{4}}(y_i)}(x)}{|B_{\frac{R}{4}}(y_i)|} dx.
$$

Now suppose the pair $(\tilde{T}, \tilde{\lambda})$ achieves the minimum in (1.2) with $c(x, y) = |x - y|^2$, storage fee function F, and this μ; since c satisfies (μ-Twist) such a pair exists by Theorem 2.3 combined with Remark 4.3. We claim that $\tilde{\lambda} = \lambda_\infty$. Let us write $B_i := B_{R/4}(y_i)$ for brevity, clearly the B_i are disjoint hence the map T defined by $T(x) = y_i$ whenever $x \in B_i$ is well-defined μ-a.e. We also see that $T_\# \mu = \sum_{i=1}^N \lambda_\infty^i \delta_{y_i}$. Let us also write $\tilde{\mathcal{L}}_i := \tilde{T}^{-1}(\{y_i\})$ which form a partition up to sets of μ-measure zero. Then if $x \in B_i$ with $\tilde{T}(x) = y_j$ and $i \neq j$, we see $|x - \tilde{T}(x)| \ge |T(x) - \tilde{T}(x)| - |x - T(x)| \ge 3R/4$, while for any j, $\tilde{\lambda}^j = \mu(\tilde{\mathcal{L}}_j) =$

$\sum_{i=1}^N \frac{\lambda^i_\infty |B_i\cap\tilde{\mathcal{L}}_j|}{|B_i|}$. Hence

$$\begin{aligned}
&\int (|x-\tilde T(x)|^2 - |x-T(x)|^2)\, d\mu(x)\\
&= \sum_{i=1}^N \frac{\lambda^i_\infty}{|B_i|}\sum_{j\neq i}\int_{B_i\cap\tilde{\mathcal{L}}_j} (|x-\tilde T(x)|^2 - |x-T(x)|^2) dx\\
&\geq \frac{R^2}{2}\sum_{i=1}^N\sum_{j\neq i}\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_j|}{|B_i|}\\
&= \frac{R^2}{2}\Big(\sum_{i=1}^N\sum_{j=1}^N\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_j|}{|B_i|} - \sum_{i=1}^N\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_i|}{|B_i|}\Big)\\
&= \frac{R^2}{2}\Big(\sum_{j=1}^N\tilde\lambda^j - \sum_{i=1}^N\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_i|}{|B_i|}\Big)\\
&= \frac{R^2}{2}\Big(1 - \sum_{i=1}^N\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_i|}{|B_i|}\Big).
\end{aligned} \tag{6.1}$$

On the other hand,

$$\begin{aligned}
|\tilde\lambda-\lambda_\infty|_1 &:= \sum_{j=1}^N|\tilde\lambda^j-\lambda^j_\infty| = \sum_{j=1}^N\Big|\sum_{i=1}^N\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_j|}{|B_i|}-\lambda^j_\infty\Big|\\
&= \sum_{j=1}^N\Big|\sum_{i\neq j}\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_j|}{|B_i|}+\lambda^j_\infty\Big(\frac{|B_j\cap\tilde{\mathcal{L}}_j|}{|B_j|}-1\Big)\Big|\\
&\leq \sum_{j=1}^N\sum_{i\neq j}\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_j|}{|B_i|}+\sum_{j=1}^N\lambda^j_\infty\Big(1-\frac{|B_j\cap\tilde{\mathcal{L}}_j|}{|B_j|}\Big)\\
&= \sum_{j=1}^N\sum_{i=1}^N\frac{\lambda^i_\infty|B_i\cap\tilde{\mathcal{L}}_j|}{|B_i|}+1-2\sum_{j=1}^N\frac{\lambda^j_\infty|B_j\cap\tilde{\mathcal{L}}_j|}{|B_j|}\\
&= 2\Big(1-\sum_{j=1}^N\frac{\lambda^j_\infty|B_j\cap\tilde{\mathcal{L}}_j|}{|B_j|}\Big).
\end{aligned}$$

Thus combining this with (6.1) we have

$$\begin{aligned}
|\tilde\lambda-\lambda_\infty|_1 &\leq \frac{4}{R^2}\int(|x-\tilde T(x)|^2-|x-T(x)|^2)\,d\mu(x)\\
&\leq \frac{4}{R^2}(F(\lambda_\infty)-F(\tilde\lambda))\\
&\leq \frac{4L}{R^2}|\tilde\lambda-\lambda_\infty|\leq|\tilde\lambda-\lambda_\infty|.
\end{aligned}\tag{6.2}$$

In turn this implies that for at most one index i, we can have $|\tilde\lambda^i-\lambda^i_\infty|\neq 0$. However, since $1=\sum_i\tilde\lambda^i=\sum_i\lambda^i_\infty$, this actually implies $\tilde\lambda=\lambda_\infty$ as claimed. By the uniqueness in Remark 4.3, this also shows $\tilde T\equiv T$ μ-a.e.

By [9, Corollary 13.3.3] and since $F^{***}=F^*$, we see that F^{**} also has Lipschitz constant L, then a calculation similar to (6.2) with F^{**} replacing F shows that the pair (T, λ_∞) minimizes (1.2) with storage fee function F^{**}. Thus, by the strong duality Theorem 3.3 we have

$$\mathfrak{m}_F - \mathfrak{M}_F = \mathfrak{m}_F - \mathfrak{m}_{F^{**}} = F(\lambda_\infty) - F^{**}(\lambda_\infty) = \|F - F^{**}\|_{L^\infty(\Lambda)}.$$

□

7. Classical dual maximizers may not be dual maximizers

Example 7.1. Let $X = \mathbb{R}$, $N = 2$, $y_1 = -1$, $y_2 = 1$, $\mu = \mathbb{1}_{[-1,-1/2]\cup[1/2,1]}(x)dx$, and $c(x, y) = |x-y|^2/2$. Also suppose $F(\lambda_1, \lambda_2) = f(\lambda_1)$ on Λ, where f is a smooth, convex function with a strict minimum at $1/2$, satisfying $f(1-\lambda) = f(\lambda)$ for all $\lambda \in [0, 1]$. By Theorem 2.3 there is a minimizer (γ, λ) of (1.3), and by [11, Remark 2.19 (iv)], γ must be given by $(\mathrm{Id} \times T_\lambda)_\# \mu$ for some increasing map $T_\lambda : \mathbb{R} \to \mathbb{R}$, defined by

$$T_\lambda(x) = \begin{cases} -1, & x \le -1 + \lambda_1, \\ 1, & x > -1 + \lambda_1, \end{cases} \quad \text{if } \lambda_1 \le \frac{1}{2},$$

and

$$T_\lambda(x) = \begin{cases} -1, & x \le \lambda_1, \\ 1, & x > \lambda_1, \end{cases} \quad \text{if } \lambda_1 > \frac{1}{2}.$$

Then, if $\lambda_1 \le 1/2$,

$$\begin{aligned} &\int_{\mathbb{R}} c(x, T_\lambda(x))\, d\mu(x) + F(\lambda) \\ &= \frac{1}{2}\Big(\int_{-1}^{-1+\lambda_1} (x+1)^2 dx + \int_{-1+\lambda_1}^{-1/2} (x-1)^2 dx + \int_{\frac{1}{2}}^{1} (x-1)^2 dx \Big) + f(\lambda_1) \\ &= f(\lambda_1) + \frac{\lambda_1^2}{3} - \frac{2\lambda_1}{3} + \frac{19}{24}, \end{aligned}$$

which is minimized at $\lambda_1 = 1/2$. By symmetry the case $\lambda_1 \ge 1/2$ has a minimum when $\lambda_1 = 1/2$, hence $\lambda = (1/2, 1/2)$, and the optimal map $T := T_{(1/2,1/2)}$ satisfies $T([-1, -1/2]) = -1$, $T([1/2, 1]) = 1$. For $\psi := (\psi^1, \psi^2) \in \mathbb{R}^2$, let

$$\varphi_\psi := \max\big(-\frac{|x+1|^2}{2} - \psi^1, -\frac{|x-1|^2}{2} - \psi^2 \big).$$

If $\psi^2 - \psi^1 \in [-1, 1]$, then $-\frac{|x+1|^2}{2} - \psi^1 \ge -\frac{|x-1|^2}{2} - \psi^2$ on $[-1, -1/2]$ and vice versa on $[1/2, 1]$. Thus

$$\begin{gathered} -\varphi_\psi(x) - \psi^1 = c(x, T(x)), \quad \forall x \in [-1, -\frac{1}{2}], \\ -\varphi_\psi(x) - \psi^2 = c(x, T(x)), \quad \forall x \in [\frac{1}{2}, 1], \end{gathered}$$

and by [12, Remark 5.13], for all such ψ the pair (φ_ψ, ψ) is a maximizer in the classical dual problem from μ to $\frac{1}{2}(\delta_{-1} + \delta_1)$ (identifying $\psi \in \mathbb{R}^2$ with a function in $L^1(\frac{1}{2}(\delta_{-1} + \delta_1))$).

By the choice of F, we see that $\partial F(1/2, 1/2) = \{(r, r) : r \in \mathbb{R}\}$, which does not contain all ψ as above. Hence by Lemma 3.10, not all classical dual maximizers are maximizers in (3.1).

Acknowledgments. J. Kitagawa was supported by the National Science Foundation grants DMS-1700094 and DMS-2000128. Part of this research was performed while J. Kitagawa was visiting the Institute for Mathematical and Statistical Innovation (IMSI), while supported by the National Science Foundation (Grant No. DMS-1929348). M. Bansil was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650604.

References

[1] Mohit Bansil, Jun Kitagawa; *A Newton Algorithm for Semidiscrete Optimal Transport with Storage Fees*, SIAM J. Optim. **31** (2021), no. 4, 2586–2613. MR 4329988

[2] Patrick Billingsley; *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1999, A Wiley-Interscience Publication. MR 1700749

[3] V. I. Bogachev; *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007. MR 2267655

[4] Andrea Braides; *Γ-convergence for beginners*, Oxford Lecture Series in Mathematics and its Applications, vol. 22, Oxford University Press, Oxford, 2002. MR 1968440

[5] Guillaume Carlier, Lina Mallozzi; *Optimal monopoly pricing with congestion and random utility via partial mass transport*, J. Math. Anal. Appl. **457** (2018), no. 2, 1218–1231. MR 3705351

[6] Gianluca Crippa, Chloé Jimenez, Aldo Pratelli; *Optimum and equilibrium in a transport problem with queue penalization effect*, Adv. Calc. Var. **2** (2009), no. 3, 207–246. MR 2537021

[7] Claude Dellacherie, Paul-André Meyer; *Probabilities and potential*, North-Holland Mathematics Studies, vol. 29, North-Holland Publishing Co., Amsterdam-New York, 1978. MR 521810

[8] Lina Mallozzi, Antonia Passarelli di Napoli; *Optimal transport and a bilevel location-allocation problem*, J. Global Optim. **67** (2017), no. 1-2, 207–221. MR 3596835

[9] R. Tyrrell Rockafellar; *Convex analysis*, Princeton University Press, 1970.

[10] Filippo Santambrogio; *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015, Calculus of variations, PDEs, and modeling. MR 3409718

[11] Cédric Villani; *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR 1964483 (2004e:90003)

[12] Cédric Villani; *Optimal transport: Old and new*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009. MR 2459454 (2010f:49001)

Mohit Bansil
Department of Mathematics, University of California, Los Angeles, CA 90095, USA
Email address: `mbansil@math.ucla.edu`

Jun Kitagawa
Department of Mathematics, Michigan State University, 619 Red Cedar Road, C212 Wells Hall, East Lansing, MI 48824, USA
Email address: `kitagawa@math.msu.edu`