Calculus of Variations

CrossMark

Dual potentials for capacity constrained optimal transport

Jonathan Korman · Robert J. McCann · Christian Seis

Abstract Optimal transportation with capacity constraints, a variant of the well-known optimal transportation problem, is concerned with transporting one probability density $f \in L^1(\mathbf{R}^m)$ onto another one $g \in L^1(\mathbf{R}^n)$ so as to optimize a cost function $c \in L^1(\mathbf{R}^{m+n})$ while respecting the capacity constraints $0 \leq h \leq \bar{h} \in L^\infty(\mathbf{R}^{m+n})$. A linear programming duality for this problem was first proposed by Levin. In this note, we prove under mild assumptions on the given data, the existence of a pair of L^1-functions optimizing the dual problem. Using these functions, which can be viewed as Lagrange multipliers to the marginal constraints f and g, we characterize the solution h of the primal problem. We expect these potentials to play a key role in any further analysis of h. Moreover, starting from Levin's duality, we derive the classical Kantorovich duality for unconstrained optimal transport. In tandem with results obtained in our companion paper [Korman et al. J Convex Anal arXiv:1309.3022 [8] (in press)], this amounts to a new and elementary proof of Kantorovich's duality.

Mathematics Subject Classification Primary 90C46; Secondary 49J45

Communicated by L. Ambrosio.

J. Korman · R. J. McCann (✉) · C. Seis
Department of Mathematics, University of Toronto, Toronto, ON M5S 2E4, Canada
e-mail: mccann@math.toronto.edu

J. Korman
e-mail: jkorman@math.toronto.edu

Present address:
C. Seis
Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60,
53115 Bonn, Germany
e-mail: cseis@math.toronto.edu

1 Introduction

In the optimal transport problem of Monge [11] and Kantorovich [3], one is given distributions $f(x)$ of sources and $g(y)$ of sinks, and is asked which pairing $h(x, y) \geq 0$ of sources with sinks minimizes a given transportation cost $c(x, y)$. The duality theory initiated by Kantorovich provides a key tool for the analysis of this question. It was historically one of the first and has become one of the archetypal examples in the theory of infinite-dimensional linear programming. Note that the optimal transportation problem is concerned only with moving sources to sinks and not with the means by which this might be accomplished.

In the capacity-constrained variant of this problem, a bound $\bar{h}(x, y)$ is imposed on the number of sources at x which can be paired with sinks at y. Such a bound is meant to model the constraints on the means of transportation. For example, in a discrete setting, if bread loaves are transported from bakeries at i to cafés at j by means of trucks (each dedicated to route ij), then the capacity bound models each truck's capacity [6]. From the perspective of the constrained transport problem we sometimes refer to the (unconstrained) optimal transport problem as the case $\bar{h} = \infty$.

The imposition of such a bound has a qualitative effect on the optimal solution and can lead to surprising conclusions about the topology of its support. For generic $c \in C^2$, the unconstrained problem leads to solutions which are singular measures concentrated on sets of zero volume, whereas in the constrained problem all solutions are functions of the form $\bar{h}1_W$ whose support W is a Lebesgue set with positive volume. Examples show that the boundary of the support need neither be smooth nor simply connected [7].

Although a duality theorem for capacity-constrained transport was established by Levin[1] (see Theorem 4.6.14 of [12], for which a new proof is given in [8]), it has not been clear until now whether the dual problem admits solutions. When they exist, such solutions play the role of Lagrange multipliers for the primal problem. However, the compactification techniques used to find them in the unconstrained problem [10,14] fail miserably when $\bar{h} \neq +\infty$. The purpose of this manuscript is to prove that, under suitable hypotheses such as the continuity and strict positivity of f, g and $\bar{h}$ on their supports, presumed compact in $\mathbf{R}^m$, $\mathbf{R}^n$ and $\mathbf{R}^{m+n}$, Levin's dual problem in fact admits solutions $(u, v) \in L^1 \oplus L^1$. This allows us to characterize optimizers h for the primal problem in terms of Lagrange multipliers (u, v). As in the theory for the unconstrained Monge-Kantorovich problem [10,14] which has developed since the work of Brenier [2], we expect our characterization of primal optimizers using dual solutions will be the starting point for any future analysis of their analytic or geometric properties.

As a consequence of our analysis, we are also able to deduce Kantorovich's duality theorem for the unconstrained optimal transportation problem as a singular limit $\bar{h} \to \infty$.

To fix notation, let f and g be probability densities on $X \subset \mathbf{R}^m$ and $Y \subset \mathbf{R}^n$, respectively. We assume X and Y to be compact, unit volume sets without loss of generality. The continuous functions $C(X)$ form a Banach space under the supremum norm $\| \cdot \|_\infty$, whose dual space consists of signed Radon measures $M(X) = C(X)^*$ normed by total variation $\| \cdot \|_{TV}$, according to the Riesz-Markov theorem [13]. Let $C_+(X) = \{c \in C(X) \mid c \geq 0\}$ and $M_+(X) = \{U \in M(X) \mid U \geq 0\}$. We denote the Jordan decomposition of a signed measure into its positive and negative parts by $U = U_+ - U_-$, and set $|U| = U_+ + U_-$.

Fix $0 \leq \bar{h} \in L^\infty(X \times Y)$ and let $\Gamma := \Gamma^{\bar{h}}(f, g)$ denote the set of joint probability densities $h \leq \bar{h}$ with f and g as marginals. Kellerer (1) and Levin (2) give necessary and sufficient

conditions for Γ to be non-empty: $\Gamma \neq \emptyset$ if and only if

$$\int \varphi f dx + \int \psi g dy \leq \iint [\varphi(x) + \psi(y)]_+ \bar{h}(x, y) dx dy \tag{1}$$

for all $\varphi, \psi \in L^\infty$ [4,5], or equivalently if and only if

$$\int_A f dx + \int_B g dy \leq 1 + \iint_{A \times B} \bar{h} dx dy \tag{2}$$

for all Borel $A \times B \subset X \times Y$ [9]. Here $dx = d\mathcal{H}^m(x)$ and $dy = d\mathcal{H}^n(y)$ denote Lebesgue measure. Assuming these conditions are satisfied, the capacity-constrained transportation problem is to compute

$$I^* := \sup_{h \in \Gamma^{\bar{h}}(f,g)} \iint h(x, y) s(x, y) dx dy, \tag{3}$$

where our sign convention is to maximize the surplus $s := -c \in L^1(X \times Y)$ rather than to minimize the cost c. This supremum is known to be attained at an extreme point of the feasible set [6], and the extreme points have been characterized in [7] as those $h \in \Gamma$ given by $h = \bar{h} 1_W$ for some Lebesgue measurable $W \subset X \times Y$.

For $u \in L^1(X)$, $v \in L^1(Y)$ and $s \in L^1(X \times Y)$ define

$$I(u, v) := \iint [s(x, y) + u(x) + v(y)]_+ \bar{h}(x, y) dx dy - \int u f dx - \int v g dy. \tag{4}$$

Although Levin's dual problem is linear rather than convex and was set in $C(X) \oplus C(Y)$ rather than in $L^1(X) \oplus L^1(Y)$, it is easily seen to be equivalent to computing the infimum

$$I_* := \inf_{(u,v) \in L^1(X) \oplus L^1(Y)} I(u, v), \tag{5}$$

by identifying his Lagrange multiplier w conjugate to the capacity constraint $h(x, y) \leq \bar{h}(x, y)$ with $w(x, y) = [s(x, y) + u(x) + v(y)]_+$. Levin asserts

$$I^* = I_*, \tag{6}$$

see also [8]. We would like to know whether the infimum I_* is attained.

Our main theorem is stated below. It gives a condition on continuous strictly positive $\bar{h}, f, g$ which ensures that the infimum in (5) is attained: namely, the set $\Gamma^{\bar{h}/\eta}(f, g)$ of joint probability densities $h \leq \bar{h}/\eta$ with f and g as marginals must be non-empty—not only for $\eta = 1$ as required to have the identity $I_* = I^*$—but for some $\eta > 1$. We note that the condition that $\bar{h}/\eta$ is still admissible is a kind of "interior point" condition which often appears in convex analysis. For example, in transport theory, a similar phenomenon, leading to the existence of a dual potential, was observed in [1].

Main Theorem (Theorem 4.2) *Let f, g and $\bar{h}$ be continuous and strictly positive on the compact, unit-volume sets $X \subset \mathbf{R}^m$, $Y \subset \mathbf{R}^n$ and $X \times Y$ respectively. Fix $\eta > 1$ and $s \in L^1(X \times Y)$. If $\Gamma^{\bar{h}/\eta}(f, g)$ is not empty, then there exist functions $(u, v) \in L^1(X) \oplus L^1(Y)$ such that $I_* = I(u, v)$.*

The functions u and v are commonly referred to as *dual potentials*. Combining this theorem with known results [Proposition 2.1 and Levin's duality (6)], we obtain a characterization of optimality:

Corollary 1.1 (Characterization of optimality) *Under the hypotheses of the main Theorem, any $h \in \Gamma^{\bar{h}}(f, g)$ is optimal if and only if there exist $(u, v) \in L^1(X) \oplus L^1(Y)$ such that*

$$s(x, y) + u(x) + v(y) \begin{cases} \leq 0 & \text{where} \quad h = 0, \\ = 0 & \text{where} \quad 0 < h < \bar{h}, \\ \geq 0 & \text{where} \quad h = \bar{h}. \end{cases} \tag{7}$$

Characterization (7) can be interpreted as a complementary slackness condition.

We now outline our strategy. Proving that the minimum (5) is attained relies on a continuity and compactness argument. At its core is a coercivity estimate (Proposition 3.3) which guarantees L^1-boundedness of (suitably normalized) minimizing sequences. In view of the lack of compactness of the closed unit ball in the non-reflexive Banach space L^1, it is natural to embed L^1 into the larger space of signed Radon measures equipped with the total variation norm, which, as a dual space, is guaranteed better compactness properties by Alaoglu's theorem. Although the number of competitors increases, this embedding does not affect our variational problem: It can be easily shown that the value of the functional $I(u, v)$, suitably extended to singular measures, does not decrease with this extension. Lower semicontinuity properties of $I(u, v)$ allow us to conclude the proof.

The L^1-bounds that we derive for minimizing sequences are in a certain sense uniform in $\bar{h}$. This observation allows us to derive the classical Kantorovich duality from Levin's duality as the singular limit $\bar{h} \to \infty$. In tandem with the results obtained in the companion paper [8], this amounts to a new and *elementary* proof of the Kantorovich duality theorem.

The remainder of this paper is organized as follows: In Sect. 2, we derive weak duality and complementary slackness conditions; Sect. 3 contains the key (coercivity) estimates; Sect. 4 contains the main result (the appendix contains a few elementary facts used in its proof); in Sect. 5, we give a new elementary proof of the Kantorovich duality; we finally conclude this paper with a discussion on future work in Sect. 6.

2 Complementary slackness

In the following proposition, we establish the complementary slackness conditions of linear programming and obtain the inequality $I_* \geq I^*$, that is, the easy part of Levin's duality, as a by-product.

Proposition 2.1 (Complementary slackness) *Fix $(u, v) \in L^1(X) \oplus L^1(Y)$, $(s, \bar{h}) \in (L^1 \oplus L^\infty)(X \times Y)$ and a probability density $0 \leq h \leq \bar{h}$ whose marginals are denoted by f and g. Then*

$$I(u, v) \geq \iint hs \, dx \, dy. \tag{8}$$

Moreover, equality holds if and only if (u, v) and h satisfy the following complementary slackness conditions

$$s + u + v \begin{cases} \leq 0 & \text{in} \quad \{h = 0\}, \\ = 0 & \text{in} \quad \{0 < h < \bar{h}\}, \\ \geq 0 & \text{in} \quad \{h = \bar{h}\}, \end{cases} \tag{9}$$

up to Lebesgue negligible sets.

Clearly, $(u, v) \in L^1(X) \oplus L^1(Y)$ and $h \in L^\infty(X \times Y)$ satisfying (9) are a pair of minimizers and maximizers of (5) and (3), respectively.

Proof Because $h \leq \bar{h}$, we first notice that

$$
\begin{aligned}
I(u, v) &\geq \iint [s + u + v]_+ h \, dx \, dy - \int u f \, dx - \int v g \, dy \\
&\geq \iint (s + u + v) \, h \, dx \, dy - \int u f \, dx - \int v g \, dy \\
&= \iint h s \, dx \, dy
\end{aligned}
$$

to establish (8). Moreover, these inequalities become equalities if and only if $s + u + v \leq 0$ where $\bar{h} > h$ and $s + u + v \geq 0$ where $h > 0$. Combining these two conditions implies (9). □

3 Coercivity in L^1

In this section, we establish the estimates which constitute the core of this paper.

We remark that $I(u + k, v - k) = I(u, v)$ for all $k \in \mathbf{R}$ shows the level sets of I cannot be compact. To resolve this lack of coercivity, we shall eventually assume

$$
\overline{uf} := \int u(x) f(x) dx = \int v(y) g(y) dy =: \overline{vg},
$$

which can always be enforced by an appropriate choice of constant k, and restricts the problem to a codimension 1 subspace of $L^1(X) \oplus L^1(Y)$. We would then like to show $I(u, v) \leq I_* + 1$ implies an L^1-bound on u and v under suitable assumptions on the data f, g, and $\bar{h}$.

Our argument comes in two steps: We first prove a bound on the means $\overline{uf} = \overline{vg}$. We shall use this to establish a bound on the oscillation $\|uf - \overline{uf}\|_{L^1}$ of uf. A similar bound for vg follows by symmetry.

The bound on the means follows immediately from $I(u, v) \leq I_* + 1$ by the following claim.

Lemma 3.1 (Mean bound) *Fix* $(u, v, s) \in L^1(X) \oplus L^1(Y) \oplus L^1(X \times Y)$ *and a probability density* $h \in L^\infty(X \times Y)$ *with marginals* f *and* g. *If* $\bar{h} \geq \eta h$ *for some* $\eta > 1$, *then*

$$
-I(u, v) \leq \overline{uf} + \overline{vg} \leq \frac{I(u, v) + \eta \|hs\|_{L^1}}{\eta - 1}.
$$

Proof One direction is easy: $I(u, v) \geq -\overline{uf} - \overline{vg}$ follows directly from the definition (4). The other direction is a consequence of

$$
\begin{aligned}
I(u, v) &\geq -\overline{uf} - \overline{vg} + \eta \iint h(x, y)[s(x, y) + u(x) + v(y)] dx \, dy \\
&\geq (\eta - 1)(\overline{uf} + \overline{vg}) - \eta \|hs\|_{L^1}
\end{aligned}
$$

since h has f and g as its left and right marginals. □

We next convert upper bounds on f, g, and $1/\bar{h}$ into a bound on the oscillation of uf around its mean. Of course, a symmetrical bound holds for vg.

Lemma 3.2 (Oscillation bound) *Fix* $u, f \in L^1(X)$, $v, g \in L^1(Y)$ *and* $(s, \bar{h}) \in (L^1 \oplus L^\infty)(X \times Y)$. *If* $\bar{h}(x, y) \geq \epsilon f(x) g(y)$ *for some* $0 < \epsilon \leq 1$ *and all* $(x, y) \in X \times Y$, *then*

$$
\frac{\epsilon}{6} \|uf - \overline{uf}\|_{L^1(X)} \leq I(u, v) + \|sfg\|_{L^1} + |\overline{uf}| + |\overline{vg}|. \tag{10}
$$

Proof Let $\sigma = \int |uf - \overline{uf}| dx$ denote oscillation around the mean, i.e. the L^1-distance separating uf from its average value. Set $X^{\pm} = \{x : \pm(u(x)f(x) - \overline{uf}) > \sigma/3\}$ with $|X^+| = \mathcal{H}^m(X^+)$ and $|X^-| = \mathcal{H}^n(X^-)$. From the definition of $\overline{uf}$ we find

$$\int_{\{|uf - \overline{uf}| \le \sigma/3\}} \left(\overline{uf} - u(x)f(x) \right) dx = A^+ - A^-$$

where

$$A^{\pm} = \pm \int_{X^{\pm}} \left(u(x)f(x) - \overline{uf} \right) dx$$
$$\ge 0.$$

On the other hand, from the definition of σ,

$$\sigma = \left(\int_{\{|uf - \overline{uf}| \le \sigma/3\}} + \int_{X^+} + \int_{X^-} \right) |u(x)f(x) - \overline{uf}| dx$$
$$\le (1 - |X^+| - |X^-|)\sigma/3 + A^+ + A^-.$$

Thus

$$A^+ + A^- \ge (2 + |X^+| + |X^-|)\sigma/3 \ge \frac{2}{3}\sigma$$

and

$$A^+ - A^- \ge -(1 - |X^+| - |X^-|)\frac{\sigma}{3} \ge -\frac{\sigma}{3}$$

whence $A^+ \ge \sigma/6$. Thus to control σ it is enough to control A^+.

Now, since g is a probability density

$$A^+ = -|X^+|\overline{uf} + \iint_{X^+ \times Y} ufg \, dxdy$$
$$= \iint_{X^+ \times Y} (s + u + v)fg \, dxdy - \iint_{X^+ \times Y} sfg \, dxdy$$
$$\quad -|X^+|\overline{uf} - \overline{vg} \int_{X^+} f \, dx$$
$$\le \frac{1}{\epsilon} I(u, v) + \|sfg\|_{L^1} + \left(\frac{1}{\epsilon} - |X^+| \right) \overline{uf} + \left(\frac{1}{\epsilon} - \int_{X^+} f \, dx \right) \overline{vg}$$

where in the last estimate we have used the assumption that $\epsilon fg \le \bar{h}$ for some $0 < \epsilon \le 1$ which implies $\iint_{X^+ \times Y}(s + u + v)fg \le \iint_{X^+ \times Y}[s + u + v]_+ fg \le \frac{1}{\epsilon} \iint_{X \times Y}[s + u + v]_+ \bar{h} = \frac{1}{\epsilon}(I(u, v) + \overline{uf} + \overline{vg})$. Using that f is a probability density, $|X^+| \le 1$, and $\epsilon \le 1$ yields

$$\epsilon A^+ \le I(u, v) + \|sfg\|_{L^1} + \left[\overline{uf}\right]_+ + \left[\overline{vg}\right]_+.$$

Since $A^+ \ge \sigma/6$, this yields the desired bound (10), in which we have replaced the positive part $[\cdot]_+$ by absolute value $|\cdot|$ simply for ease of parsing. $\qquad \square$

Proposition 3.3 (Coercivity) *Let* $s \in L^1(X \times Y)$ *and* $\eta h \le \bar{h} \in L^\infty(X \times Y)$, *where* $\eta > 1$ *and h is a probability density having f and g as its marginals. Assume* $\bar{h}(x, y) \ge \epsilon f(x)g(y)$ *and* $\min\{f(x), g(y)\} \ge \epsilon$ *for some* $\epsilon > 0$ *and almost all* $(x, y) \in X \times Y$. *If* $\overline{uf} = \overline{vg}$ *for some* $u \in L^1(X)$ *and* $v \in L^1(Y)$, *then* $I(u, v) \le I_* + 1$ *implies* $\|u\|_{L^1}$ *and* $\|v\|_{L^1}$ *are controlled by a bound which depends only on* I_*, ϵ, η, $\|s\|_{L^1}$ *and* $\|\bar{h}\|_\infty$.

Proof According to Lemma 3.1, the means $\overline{vg} = \overline{uf}$ are bounded in terms of the given I_*, ϵ, η, $\|s\|_{L^1}$ and $\|\bar{h}\|_\infty$. Lemma 3.2 and $\epsilon fg \leq \bar{h}$ then bounds the mean oscillation of uf around $\overline{uf}$ in terms of $\overline{uf} = \overline{vg}$ and the listed parameters. Now $\|u\|_{L^1} \leq \|f^{-1}\|_{L^\infty} \|uf\|_{L^1} \leq \epsilon^{-1} \|uf\|_{L^1}$ combined with $\|uf\|_{L^1} \leq \|\overline{uf}\|_{L^1} + \|uf - \overline{uf}\|_{L^1}$ yields the desired bound for $\|u\|_{L^1}$. A similar bound for $\|v\|_{L^1}$ follows by symmetry. □

Remark 3.4 The bound mentioned in the conclusion of Proposition 3.3 is uniform in $\|\bar{h}\|_{L^\infty} \gg 1$.

4 Existence of optimizers

Proposition 3.3 implies that we can choose a minimizing sequences (u_i, v_i) for (5) which is uniformly bounded in $L^1(X) \oplus L^1(Y)$. Since the L^1 unit ball is non-compact, an L^1 bound is not enough to permit extraction of a limit lying in $L^1(X) \oplus L^1(Y)$. To recover convergence we would like to view (u_i, v_i) in a larger space which is better behaved.

Therefore, let us extend the definition of $I(u, v)$ from L^1 to the space of signed measures with finite total variation which, as a dual Banach space, has better compactness properties. Denoting by by $S \in M(X \times Y)$ the measure with Lebesgue density $s(x, y)$, this extension is given by

$$\tilde{I}(U, V) := \iint \bar{h}d[S + U \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V]_+ - \int f dU - \int g dV$$

for $(U, V) \in M(X) \oplus M(Y)$. (We recall that in our notation, $\mathcal{H}^m$ and $\mathcal{H}^n$ are Lebesgue measures.) The next lemma is routine; it verifies that $\tilde{I}$ is lower semicontinuous with respect to weak-$*$ convergence in $M(X) \oplus M(Y)$.

Lemma 4.1 (Lower semicontinuity) *Let $s \in L^1(X \times Y)$ induce $dS = sd\mathcal{H}^{m+n}$. Given continuous non-negative f, g and $\bar{h}$ on the compact sets $X \subset \mathbf{R}^m, Y \subset \mathbf{R}^n$ and $X \times Y$, the functional $\tilde{I}(U, V)$ behaves lower semicontinuously with respect to weak-$*$ convergence in $M(X) \oplus M(Y)$.*

Proof Choose a bounded sequence in $M(X) \oplus M(Y)$ which converges $(U_i, V_i) \to (U, V)$ when tested against functions in $C(X) \oplus C(Y)$. Then

$$\int f dU = \lim_{i \to \infty} \int f dU_i \quad \text{and} \quad \int g dV = \lim_{i \to \infty} \int g dV_i,$$

so we need only show

$$\liminf_{i \to \infty} \iint \bar{h}d[S + U_i \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_i]_+$$
$$\geq \iint \bar{h}d[S + U \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V]_+. \tag{11}$$

We define the signed Radon measures $d\mu := \bar{h}d(S + U \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V)$ and $d\mu^i := \bar{h}d(S + U_i \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_i)$. Then $\mu^i \to \mu$ weakly-$*$. Indeed, for every $\varphi \in C(X \times Y)$

$$\iint \varphi d\mu^i = \int \bar{h}s\varphi dx dy + \int \left(\int \bar{h}\varphi dy \right) dU^i + \int \left(\int \bar{h}\varphi dx \right) dV^i$$

$$\rightarrow \int \bar{h}s\varphi dx dy + \int \left(\int \bar{h}\varphi dy \right) dU + \int \left(\int \bar{h}\varphi dx \right) dV$$

$$= \iint \varphi d\mu,$$

because the x- and y-marginals of $\bar{h}\varphi$ are both continuous. Moreover, decomposing μ and μ^i into their positive and negative parts $\mu = \mu_+ - \mu_-$ and $\mu^i = \mu^i_+ - \mu^i_-$, (11) becomes

$$\liminf_{i \to \infty} \mu^i_+(X \times Y) \geq \mu_+(X \times Y). \tag{12}$$

This in fact is an immediate consequence of the weak-$*$ convergence $\mu^i \to \mu$: On the one hand, choosing $\varphi \equiv 1$, weak-$*$ convergence implies that $\mu^i_+(X \times Y) - \mu^i_-(X \times Y) = \mu^i(X \times Y) \to \mu(X \times Y) = \mu_+(X \times Y) - \mu_-(X \times Y)$. On the other hand, by the lower semicontinuity of the total variation norm, it is $\mu_+(X \times Y) + \mu_-(X \times Y) = |\mu|(X \times Y) = \|\mu\|_{TV} \leq \liminf_{i \to \infty} \|\mu^i\|_{TV} = \liminf_{i \to \infty} |\mu^i|(X \times Y) = \liminf_{i \to \infty} (\mu^i_+(X \times Y) + \mu^i_-(X \times Y))$. The statement in (12) now follows upon combining both convergence results. □

To prove the next theorem we need to recall some basic facts about signed Borel measures.

A measure μ on $\mathbf{R}^m$ is absolutely continuous with respect to Lebesgue measure $\mathcal{H}^m$, denoted $\mu \ll \mathcal{H}^m$, if for any $\mathcal{H}^m$-measurable set Ω for which $\mathcal{H}^m(\Omega) = 0$, it is also the case that $\mu(\Omega) = 0$. Two signed measures μ_1 and μ_2 are mutually singular, denoted $\mu_1 \perp \mu_2$, if they are concentrated on disjoint sets.

Lebesgue's decomposition theorem states that a σ-finite measure μ can be the uniquely decomposed into its absolutely continuous part and singular part with respect to Lebesgue measure: $\mu = \mu_{ac} + \mu_s$, where $\mu_{ac} \ll \mathcal{H}^m$ and $\mu_s \perp \mathcal{H}^m$; furthermore $\mu_{ac} = h d\mathcal{H}^m$ for a unique $h \in L^1(\mathcal{H}^m)$ called the Radon–Nikodym derivative of μ_{ac} (with respect to $\mathcal{H}^m$). Note that $\mu_s \perp \mu_{ac}$.

The Hahn–Jordan decomposition theorem states that any signed measure μ can be uniquely decomposed into its positive and negative parts: $\mu = \mu_+ - \mu_-$, where μ_+, μ_- are positive measures and $\mu_+ \perp \mu_-$. The variation of μ is denoted by $|\mu| := \mu_+ + \mu_-$.

It is a standard fact that for σ-finite measure μ the Hahn–Jordan decomposition and the Lebesgue decomposition commute: $\mu_+ = [\mu_{ac}]_+ + [\mu_s]_+$ and $\mu_- = [\mu_{ac}]_- + [\mu_s]_-$.

Theorem 4.2 (Existence of optimal potentials) *Let f, g and $\bar{h}$ be continuous and strictly positive on the compact, unit-volume sets $X \subset \mathbf{R}^m$, $Y \subset \mathbf{R}^n$ and $X \times Y$ respectively. Fix $\eta > 1$ and $s \in L^1(X \times Y)$. If $\Gamma^{\bar{h}/\eta}(f, g)$ is not empty, then there exist functions $(u, v) \in L^1(X) \oplus L^1(Y)$ such that $I_* = I(u, v)$ in (5).*

Proof Let (u_i, v_i) be a minimizing sequence for (5): $u_i \in L^1(X)$ and $v_i \in L^1(Y)$ such that $I(u_i, v_i) \to I_*$. Fix $\epsilon \leq \min\{f(x)^{\pm 1}, g(y)^{\pm 1}\}$. Since $\bar{h} > 0$ is bounded away from zero, taking $\epsilon > 0$ smaller if necessary ensures $\bar{h}(x, y) \geq \epsilon f(x)g(y)$ throughout $X \times Y$. Adding a constant to u_i and subtracting the same constant from v_i ensures $\overline{u_i f} = \overline{v_i g}$. Now Proposition 3.3 asserts a bound for $\|u_i\|_{L^1} + \|v_i\|_{L^1}$ independent of i. Letting U_i and V_i denote the measures with Lebesgue densities u_i and v_i and recalling $M(X) = C(X)^*$, Alaoglu's theorem provides a weak-$*$ convergent subsequence also denoted (U_i, V_i) with limit $(U_i, V_i) \to (U, V) \in M(X) \oplus M(Y)$. Since $\tilde{I}(U_i, V_i) = I(u_i, v_i) \to I_*$ by construction, Lemma 4.1 implies $\tilde{I}(U, V) \leq I_*$. We will next argue that $\tilde{I}(U, V) \geq I_*$.

Let $S = s\mathcal{H}^n \otimes \mathcal{H}^m$. Since the Hahn–Jordan and the Lebesgue decompositions of a measure commute, as mentioned at the end of the paragraph above

$$\left[S + U \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V\right]_+ = \left[S + U_{ac} \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_{ac} + U_s \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_s\right]_+$$
$$= \left[S + U_{ac} \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_{ac}\right]_+ + \left[U_s \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_s\right]_+. \tag{13}$$

Therefore

$$\tilde{I}(U, V) = \tilde{I}(U_{ac}, V_{ac}) + \iint \bar{h} \, d\left[U_s \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_s\right]_+ - \int f \, dU_s - \int g \, dV_s$$

$$\geq I(u, v) + \iint h \, d\left[U_s \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_s\right] - \int f \, dU_s - \int g \, dV_s$$

$$= I(u, v),$$

where u and v are the Radon-Nikodym derivatives of U_{ac} and V_{ac} respectively: $u \, dx = dU_{ac}$ and $v \, dy = dV_{ac}$. Since $I(u, v) \geq I_*$ we have $I_* = \tilde{I}(U, V) = I(u, v)$ as desired. $\quad\square$

5 From Levin's to Kantorovich's duality

In this section, we briefly discuss how to derive the classical Kantorovich duality of optimal transport from Levin's duality using some of the insights of previous sections. For simplicity, assume that f and g are strictly positive, continuous probability densities on the compact sets $X \subset \mathbf{R}^m$ and $Y \subset \mathbf{R}^n$, where both X and Y have unit volume. Let $\Gamma(f, g)$ denote the set of all joint probability measures with f and g as marginals. Fix $s \in C(X \times Y)$ continuous and define

$$L_s := \left\{(u, v) \in L^1(X) \oplus L^1(Y) | \, s(x, y) + u(x) + v(y) \leq 0\right\}.$$

We will recover the expression of Kantorovich duality as stated in [14, Theorem 5.10]. We remark that although the elements of L_s need not be continuous, for Lipschitz continuous costs it is well-known that the infimum on the right hand side of (14) below is attained by Lipschitz continuous densities (u, v).

Theorem 5.1 (Duality for the unconstrained problem) *Let $X \subset \mathbf{R}^m$ and $Y \subset \mathbf{R}^n$ be compact unit volume sets equipped with positive continuous probability densities $f \in C(X)$ and $g \in C(Y)$. If $s \in C(X \times Y)$ then*

$$\sup_{H \in \Gamma(f,g)} \iint s(x, y) dH(x, y) = \inf_{(u,v) \in L_s} \left(-\int f(x)u(x)dx - \int g(y)v(y)dy\right), \tag{14}$$

and both infimum and supremum (14) are attained.

Proof Let $I^*(\infty)$ denote the left hand side of (14) and $I_*(\infty)$ denote the right hand side of (14). Note that by a density argument, it is enough to let the sup in $I^*(\infty)$ range over absolutely continuous measures $dH = hdxdy \in \Gamma(f, g)$.

The inequality $I^*(\infty) \leq I_*(\infty)$ follows easily from the following inequality which holds for all $(u, v) \in L_s$ and any density h such that $hdxdy \in \Gamma(f, g)$:

$$\iint shdxdy = \iint h(s + u + v)dxdy - \int fudx - \int gvdy$$

$$\leq -\int fudx - \int gvdy.$$

The last inequality follows from the definition of L_s. Taking sup on the left hand side and inf on the right hand side gives the desired inequality.

Observe that to prove the other inequality, $I^*(\infty) \geq I_*(\infty)$, it is enough to prove existence of $H \in \Gamma(f, g)$ and $(u, v) \in L_s$ satisfying

$$\iint s \, dH \geq - \int f u \, dx - \int g v \, dy.$$

Let $h(x, y) := f(x) g(y)$ and $K := \max_{X \times Y} h(x, y) + 1$, and note that $h \, dx \, dy \in \Gamma(f, g)$. Fix $\eta > 1$ such that $K \geq \eta \max_{X \times Y} h(x, y)$, so that $k \geq \eta h$ for all $k \geq K$. Note also that $\Gamma^{\bar{h}_k}(f, g) \neq \emptyset$ if $\bar{h}_k = k 1_{X \times Y}$.

Defining $I_k(u, v)$, $I_*(k)$, and $I^*(k)$ similarly to $I(u, v)$, I_*, and I^* but with $\bar{h}$ replaced by $\bar{h}_k$, we deduce from Levin's duality $I^*(k) = I_*(k)$ the existence of functions $h_k \in \Gamma^{\bar{h}_k}(f, g)$ and $(u_k, v_k) \in L^1(X) \oplus L^1(Y)$ optimizing $I^*(k)$ and $I_*(k)$ respectively. Levin's duality reads

$$\iint h_k \, s \, dx \, dy = I_k(u_k, v_k) \tag{15}$$

for any $k \geq K$. Since $I_k(u + \alpha, v - \alpha) = I_k(u, v)$ for $\alpha \in \mathbf{R}$, there is no loss of generality to assume that

$$\overline{u_k f} = \overline{v_k g}.$$

By a standard argument the sequence $\{h_k\}_{k \geq K}$ is seen to be weak-$*$ precompact in the space of probability measures and every limit point H satisfies $H \in \Gamma(f, g)$.

In particular, since s is continuous, upon extracting a subsequence, passing to the limit $k \to \infty$ in (15) yields that

$$I^*(\infty) = \iint s \, dH = \lim_{k \to \infty} I_k(u_k, v_k). \tag{16}$$

We next observe that the L^1-bound on the sequences $\{u_k\}_{k \geq K}$ and $\{v_k\}_{k \geq K}$ obtained from Lemmas 3.1 and 3.2 is independent of $k \geq K$. In this context, we state for further references

$$\overline{u_k f} + \overline{v_k g} \leq \frac{I^*(\infty) + \eta ||s||_{L^1} ||h||_\infty}{\eta - 1}, \tag{17}$$

as a consequence Lemma 3.1.

Choosing $\epsilon > 0$ so that $K \geq \epsilon f(x) g(y)$ and $\min\{f(x), g(y)\} \geq \epsilon$, Proposition 3.3 (see Remark 3.4) implies that $||u_k||_{L^1}$, and $||v_k||_{L^1}$ are controlled by a bound which depends only on ϵ, η, $I^*(\infty)$, $||s||_{L^1}$, and $||h||_\infty$, all of which are independent of k.

Arguing as in the proof of Theorem 4.2, we may thus extract a weak-$*$ convergent subsequence (not relabeled) $(u_k, v_k) \to (U, V) \in M(X) \oplus M(Y)$ and obtain

$$\lim_{k \to \infty} I_k(u_k, v_k) \geq \lim_{k \to \infty} \left(- \int u_k f \, dx - \int v_k g \, dy \right) = - \int f \, dU - \int g \, dV,$$

by the continuity of f and g. In particular, we deduce from (15) and (16) that

$$\iint s \, dH \geq - \int f \, dU - \int g \, dV. \tag{18}$$

Moreover, the definition of $\bar{h}_k$ and Lemma 4.1 imply that

$$\iint d[S + U \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V]_+ \leq \liminf_{k \to \infty} \iint [s(x, y) + u_k(x) + v_k(y)]_+ dx dy$$

$$= \liminf_{k \to \infty} \frac{1}{k} \left(I_k(u_k, v_k) + \overline{u_k f} + \overline{v_k g} \right).$$

Since the limit on the right hand side is zero, a consequence of (16) and (17), it follows that $[S + U \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V]_+ = 0$. Now (13) implies that $\left[S + U_{ac} \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_{ac} \right]_+$ and $\left[U_s \otimes \mathcal{H}^n + \mathcal{H}^m \otimes V_s \right]_+$ are both zero. In particular, $[U_s]_+$ and $[V_s]_+$ are both zero, and with $u := dU_{ac}/dx$ and $v := dV_{ac}/dy$ denoting Radon–Nikodym derivatives, we have $[s + u + v]_+ = 0$, or equivalently $s + u + v \in L_s$. It follows that

$$-\int f dU - \int g dV = -\int f u dx - \int g v dy - \int f dU_s - \int g dV_s$$

$$\geq -\int f u dx - \int g v dy.$$

Hence (18) becomes $\iint s \, dH \geq -\int f u dx - \int g v dy$ which implies the desired inequality $I^*(\infty) \geq I_*(\infty)$. We have thus established (14) and existence of optimizers. □

6 Perspectives for future work

Of considerable interest to us is the question of showing some regularity for the minimizing potentials (u, v)—perhaps under stronger restrictions on $(f, g, \bar{h})$ and s. For example, are they continuous or differentiable; might u and v belong to some Hölder or Sobolev space or—as in the unconstrained version $\bar{h} = +\infty$ of the problem [10,14]—inherit Lipschitz and semiconcavity properties from the cost $c = -s$? In view of the characterization (9) for $h \in \Gamma^{\bar{h}}(f, g)$ to maximize the expected value of s, this is closely related to smoothness for the free boundary of the set W such that the optimizer $h = \bar{h} 1_W$. If $m = n$, this set is known to be unique (up to sets of $\mathcal{H}^{2n}$ measure zero) near points where $0 \neq \det[\partial^2 s/\partial x^i \partial y^j]$ [6]. Simple examples show the boundary of W can have isolated singularities [6,7], but is it a smooth hypersurface otherwise? Does W even have finite perimeter? Where the derivative of $s(x, y) + u(x) + v(y)$ is non-vanishing, such questions are related to smoothness of u and v by the implicit function theorem.

References

1. Bertrand, J., Puel, M.: The optimal mass transport problem for relativistic costs. Calc. Var. PDE **46**, 353–374 (2013)
2. Brenier, Y.: Décomposition polaire et réarrangement monotone des champs de vecteurs. C.R. Acad. Sci. Paris Sér. I Math. **305**, 805–808 (1987)
3. Kantorovich, L.: On the translocation of masses. C.R. (Doklady) Acad. Sci. URSS (N.S.) **37**, 199–201 (1942)
4. Kellerer, H.G.: Marginalprobleme für Funktionen. Math. Ann. **154**, 147–156 (1964)
5. Kellerer, H.G.: Maßtheoretische marginalprobleme. Math. Ann. **153**, 168–198 (1964)
6. Korman, J., McCann, R.J.: Optimal transportation with capacity constraints. Trans. Am. Math. Soc. (in press)

7. Korman, J., McCann, R.J.: Insights into capacity constrained optimal transport. Proc. Natl. Acad. Sci **110**, 10064–10067 (2013)

8. Korman, J., McCann, R.J., Seis, C: An elementary approach to linear programming duality with application to capacity constrained transport. J. Convex Anal. arXiv:1309.3022 (in press)

9. Levin, V.L.: The problem of mass transfer in a topological space, and probability measures having given marginal measures on the product of two spaces. Sov. Math. (Doklady) **29**, 638–643 (1984)

10. McCann, R.J., Guillen, N.: Five lectures on optimal transportation: geometry, regularity, and applications. In: Dafni, G. et al. (ed.) Analysis and geometry of metric measure spaces: lecture notes of the séminaire de Mathématiques Supérieure (SMS) Montréal 2011, pp. 145–180. American Mathematical Society, Providence (2013)

11. Monge, G.: Mémoire sur la théorie des déblais et de remblais. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, pp. 666–704 (1781)

12. Rachev, S.T., Rüschendorf, L.: Mass transportation problems, vol. 1. Springer, New York (1998)

13. Reed, M., Simon, B.: Functional analysis.In: Methods of modern mathematical physics, vol. 1, Academic Press, San Diego (1980)

14. Villani, C.: Optimal transport. Old and new. Springer, New York (2009)