INSIGHTS INTO CAPACITY CONSTRAINED OPTIMAL TRANSPORT

JONATHAN KORMAN AND ROBERT J. MCCANN

ABSTRACT. A variant of the classical optimal transportation problem is: among all joint measures with fixed marginals and which are dominated by a given density, find the optimal one. Existence and uniqueness of solutions to this variant were established in [KM11]. In the present manuscript, we expose an unexpected symmetry leading to the first explicit examples in two and more dimensions. These are inspired in part by simulations in one dimension which display singularities and topology and in part by two further developments: the identification of all extreme points in the feasible set, and a new approach to uniqueness based on constructing feasible perturbations.

1. INTRODUCTION

Given fixed distributions of supply and demand, the optimal transportation problem of Monge [Mo81] and Kantorovich [K42] involves pairing supply with demand so as to minimize the average transportation cost $c(x, y)$ between each supplier x and the demander y with whom x is paired. For continuous distributions, this question forms an (the?) archetypal example of an infinite-dimensional linear program. Its relevance to the physics of fluids has been recognized since the work of Brenier [B87] and Cullen and Purser [CP89], while some of its applications to geometry, dynamics, partial differential equations, economics and statistics are described in [MG10] [RR98] [V09] and the references there. It is desirable to introduce congestion effects into this model, as can be attempted in various ways [CJS08]; one of the crudest is simply is to bound the number of suppliers at x who can be paired with demanders at y, for each x and y. Despite its appeal, for continuous distributions of supply and demand, this variant seems not to have been studied until [KM11].

As in all linear programs, if the problem has solutions, at least one of them will be an extreme point of the feasible set. A remaining challenge in the Monge-Kantorovich transportation problem is to arrive at a characterization of the extreme points which yields useful information about the geometry and topology of its solutions [AKM11]. Somewhat surprisingly, such a characterization is much more accessible in our capacity constrained variant; as shown below, it can basically be reduced to a 'bang-bang' (all or nothing) principle. As a corollary, this characterization implies the uniqueness of solutions first established in [KM11]. Moreover, it combines with elementary but obscure symmetries to yield the first explicitly soluble examples in more than one dimension, and with numerical and theoretical considerations to give insights into the geometry and topology aspects of basic examples which — even in one-dimension — still defy explicit solution.

The problem in question is formulated precisely as follows: Given densities $0 \leq f, g \in L^1(\mathbf{R}^d)$ with same total mass $\int f = \int g$, let $\Gamma(f, g)$ denote the set of joint densities $0 \leq h \in L^1(\mathbf{R}^d \times \mathbf{R}^d)$ which have f and g as their marginals, meaning

$$f(x) = \int_{\mathbf{R}^d} h(x, \tilde{y}) d\tilde{y}$$

$$\int_{\mathbf{R}^d} h(\tilde{x}, y) d\tilde{x} = g(y)$$

for Lebesgue almost all $x, y \in \mathbf{R}^d$. A bounded function $c(x, y)$ represents the cost per unit mass for transporting material from $x \in \mathbf{R}^d$ to $y \in \mathbf{R}^d$. The (total) transportation cost of h is denoted $c[h]$, defined by

$$(1) \qquad c[h] := \int_{\mathbf{R}^d \times \mathbf{R}^d} c(x, y) h(x, y) dx dy,$$

is proportional to the expected value of c with respect to h.

Given $0 \leq \bar{h} \in L^1(\mathbf{R}^d \times \mathbf{R}^d)$, we let $\Gamma(f, g)^{\bar{h}}$ denote the set of all $h \in \Gamma(f, g)$ dominated by $\bar{h}$, that is $h \leq \bar{h}$ almost everywhere. The optimization problem we are concerned with — *optimal transport with capacity constraints*—is to minimize the transportation cost (1) among joint densities h in $\Gamma(f, g)^{\bar{h}}$, to obtain the optimal cost

$$(2) \qquad \min_{h \in \Gamma(f,g)^{\bar{h}}} c[h]$$

under the capacity constraint $\bar{h}$.

Notice this problem involves a linear minimization on a convex set $\Gamma = \Gamma(f, g)^{\bar{h}}$, and therefore takes the form of an infinite-dimensional

linear program. When Γ is non-empty, it is not hard to show that the minimum is attained, and that at least one of the minimizers h is an extreme point of Γ, meaning h is not the midpoint of any segment in Γ. Since the possible facet structure of Γ is not obvious, it is harder to determine whether or not this minimizer is unique. A sufficient condition for uniqueness was discovered in [KM11], and is recalled below. It is even harder to envision what the solutions will look like. Perhaps the simplest example involves pairing Lebesgue measure on the unit interval with itself so as to minimize the quadratic transportation cost $c(x, y) = |x - y|^2$ (well-known to be equivalent to $c(x, y) = -x \cdot y$, as in the unconstrained problem [B87]). As the capacity constraint $\bar{h}$ is varied over different constant values, the numerical solutions below display an unexpected variety of strange topological features and analytic singularities begging to be understood. Even though supply equals demand in these examples, and the constraints permit at least some of the demand to be supplied locally at zero cost, the islands of blue in these diagrams show that global optimality may require there to be regions where none of the demand is supplied locally.

In this paper we establish symmetries which explain at least some of the observed structures (Lemma 4.1). These also lead to the first explicit examples of optimizers in higher dimensions (Proposition 4.2). We precede this with a simple description of the extreme points of Γ, based on a perturbation argument. The same idea is also used to substantially simplify the uniqueness argument from [KM11]. The original argument relied on understanding the infinitesimal behaviour of an optimizer near its Lebesgue points in order to argue that any optimizer is geometrically extreme—a property which characterizes extreme points of $\Gamma(f, g)^{\bar{h}}$ (see proposition 3.2). The new proof begins from an 'all or nothing' characterization of extreme points and uses perturbations to argue directly — without asymptotics or blow ups — that every optimizer is extreme. Uniqueness follows easily.

2. Assumptions

We make the following assumptions throughout (see [KM11] for more details).

2.1. **Assumptions on the cost.**
   (C1) $c(x, y)$ is bounded,

(C2) there is a Lebesgue negligible closed set $Z \subset \mathbf{R}^d \times \mathbf{R}^d$ such that $c(x,y) \in C^2(\mathbf{R}^d \times \mathbf{R}^d \setminus Z)$ and,

(C3) $c(x,y)$ is non-degenerate: $\det[D^2_{x^i y^j} c(x,y)] \neq 0$ for all $(x,y) \in \mathbf{R}^d \times \mathbf{R}^d \setminus Z$.

2.2. Assumption on the capacity constraint. $\bar{h}$ is non-negative and Lebesgue integrable: $0 \leq \bar{h} \in L^1(\mathbf{R}^d \times \mathbf{R}^d)$.

Given marginal densities $0 \leq f, g \in L^1(\mathbf{R}^d)$ with same total mass, to avoid talking about the trivial case, we will always assume that a feasible solution exists: $\Gamma = \Gamma(f,g)^{\bar{h}} \neq \emptyset$.

3. Uniqueness: every optimizer is extreme

Our arguments are based crucially on a preparatory lemma from real analysis.

Lemma 3.1 (Marginal-preserving volume exchange). *Fix $\Delta = (1, \ldots, 1) \in \mathbf{R}^d$. If a Lebesgue set $U \subset \mathbf{R}^d \times \mathbf{R}^d$ is non-negligible, then for all $\delta > 0$ sufficiently small there is a subset $V \subset U$ of positive volume such that $(x + \delta\Delta, y), (x, y + \delta\Delta)$ and $(x + \delta\Delta, y + \delta\Delta)$ all belong to U whenever $(x,y) \in V$. Moreover, V may be taken to lie in the interior of a coordinate hypercube of side-length δ. The vertex of this hypercube at which Δ is an outward normal may be chosen to lie at any Lebesgue point z_0 where U has full density. (V may be chosen to have Lebesgue density $1/2^{2d}$ at z_0.)*

Proof. Let $I = [0,1]$ denote the unit interval and I^d the hypercube, so that $H_{ij} = (2i-1)I^d \times (2j-1)I^d$ for $i,j \in \{0,1\}$ define four hypercubes with disjoint interiors in $2d$ dimensions. Let z_0 be a Lebesgue point where U has full density; we may suppose z_0 is the origin without loss of generality. Letting $\delta^{-1}U$ denote the dilation of U around z_0 by factor δ^{-1}, we see the fraction of H_{ij} outside of $\delta^{-1}U$ tends to zero as $\delta \to 0$. For δ sufficiently small, we may assume all four of these fractions to be strictly less than $1/4$. Let $V \subset H_{00}$ be the set of $(x,y) \in H_{00} \cap \delta^{-1}U$ for which $(x + \Delta, y)$, $(x, y + \Delta)$ and $(x + \Delta, y + \Delta)$ also belong to $\delta^{-1}U$. If $(x,y) \in H_{00} \setminus V$, it is because at least one of the four points above does not belong to $\delta^{-1}U$. Thus $H_{00} \setminus V = J_{00} \cup J_{01} \cup J_{10} \cup J_{11}$ where each of the four sets $J_{ij} + (i\Delta, j\Delta) := H_{ij} \setminus \delta^{-1}U$ has volume strictly less than $1/4$. Thus $\mathcal{L}^{2d}[H_{00} \setminus V] < 1$, implying V is a set of positive measure. Discarding from V any points on the boundary of H_{00} and contracting by a factor δ yields the lemma. (The parenthetical remark is obtained by noting $1/4$ is arbitrary in the argument above; taking δ smaller forces the volume of $H_{00} \setminus V$ to be as small as we please. Thus

V fills a larger and larger fraction of the hypercube H_{00} near its vertex, where both H_{00} and hence V have Lebesgue density $1/2^{2d}$.) □

This allows us to give a much nicer characterization of the extreme points of $\Gamma(f,g)^{\bar{h}}$ than any available for the unconstrained problem ($\bar{h} = +\infty$) [AKM11].

Proposition 3.2 (All or nothing characterization of extreme points). *Let $\Gamma = \Gamma(f,g)^{\bar{h}}$ denote the set of joint densities bounded by $\bar{h} \in L^1(\mathbf{R}^n \times \mathbf{R}^n)$ and with marginals $f, g \in L^1(\mathbf{R}^n)$. A density $h \in \Gamma$ is an extreme point of Γ if and only if $h = 1_W \bar{h}$ for some Lebesgue measurable set $W \subset \mathbf{R}^d \times \mathbf{R}^d$.*

Proof. Recall $h \in \Gamma$ implies $0 \le h \le \bar{h}$. If h is extremal, we claim these inequalities cannot both be strict on any subset $U \subset \mathbf{R}^{2d}$ of positive volume. To show the contrapositive, suppose such a U existed. Then for some $\epsilon > 0$ the set $U_\epsilon = \{z \in \mathbf{R}^{2d} \mid \epsilon < h < \bar{h} - \epsilon\}$ would also have positive volume. Lemma 3.1 provides $\delta > 0$ and $V \subset U_\epsilon$ of positive measure such that all four points $(x \pm \frac{\delta}{2}\Delta, y \pm \frac{\delta}{2}\Delta)$ and $(x \pm \frac{\delta}{2}\Delta, y \mp \frac{\delta}{2}\Delta)$ lie in U_ϵ whenever $(x - \frac{\delta}{2}\Delta, y - \frac{\delta}{2}\Delta) \in V$. Setting $I = [0,1]$ and using $i, j \in \{0,1\}$ to define four coordinate hypercubes $H_{ij} = (2i-1)I^d \times (2j-1)I^d$, after translation we may also assume $V \subset H_{00} \setminus \partial H_{00}$. Then

$$(3) \quad \tilde{h}(x,y) := \begin{cases} +1 & \text{if } (x,y) \in V \text{ or } (x - \delta\Delta, y - \delta\Delta) \in V \\ -1 & \text{if } (x - \delta\Delta, y) \in V \text{ or } (x, y - \delta\Delta) \in V \\ 0 & \text{otherwise} \end{cases}$$

is well-defined. Notice that $\tilde{h}$ is constructed using symmetries which ensure its integrals with respect to x and with respect to y both vanish, the other variable being held fixed. In other words, the marginals of $\tilde{h}$ vanish. Also, $\tilde{h}$ is supported in U_ϵ, where we have room to add or subtract ϵ from $h \in (\epsilon, \bar{h} - \epsilon)$. Thus $h_\pm := h \pm \epsilon\tilde{h}$ both belong to Γ; they are distinct since V has positive volume. Expressing $h = \frac{1}{2}(h_+ + h_-)$ as a convex combination of $h_\pm$ shows h is not an extreme point of Γ.

Conversely, we claim any $h = 1_W \bar{h}$ with $W \subset \mathbf{R}^{2d}$ Lebesgue is extreme. To see this, suppose $h = 1_W \bar{h}$ could be decomposed as a convex combination $h = \frac{1}{2}(h_+ + h_-)$ of $h_\pm \in \Gamma$. Since $h_\pm$ are both non-negative, they must both vanish where h does; thus $h_\pm = 0$ outside of W. Since both $h_\pm \le \bar{h}$, they must both coincide with $\bar{h}$ where h does; thus $h_\pm = \bar{h}$ in W. This shows $h_+ = h_-$, establishes extremality of $h = 1_W \bar{h}$, and completes the proof of the proposition. □

More importantly, it allows us to construct a perturbative argument for uniqueness, much simpler than the original proof of [KM11].

Theorem 3.3 (Every optimizer is extreme). *Let the cost $c(x, y)$ satisfy conditions $(C1) - (C3)$, fix $0 \leq \bar{h} \in L^1(\mathbf{R}^d \times \mathbf{R}^d)$ and take $0 \leq f, g \in L^1(\mathbf{R}^d)$ such that $\Gamma := \Gamma(f, g)^{\bar{h}} \neq \emptyset$. If $h \in \Gamma$ is optimal, i.e. $h \in \mathrm{argmin}_{k \in \Gamma} c[k]$, then h is an extreme point of Γ.*

Proof. Suppose $h \in \Gamma$ is not an extreme point of Γ. We shall establish the theorem by constructing a perturbation of h which decreases the cost $c[h]$. Proposition 3.2 asserts $h \neq 1_W \bar{h}$, meaning the set U of Lebesgue points z for h and $\bar{h}$ at which $0 < h < \bar{h}$ has positive volume. Similarly, for sufficiently small $\epsilon > 0$ the set $U_\epsilon = \{z \in U \setminus Z \mid \epsilon < h < \bar{h} - \epsilon\}$ also has positive volume, where Z is the negligible closed set on which hypotheses (C2) $c \in C^2$ and (C3) $\det[D^2_{x^i y^j}] \neq 0$ may fail. Let $z_0 = (x_0, y_0)$ be a point where U_ϵ has full Lebesgue density; we may assume z_0 to be the origin without loss of generality. After a linear transformation of the variable y (as in [MPW10] or §5 of [KM11]), we can also assume $D^2_{x^i y^j} c(z_0) = -\delta_{ij}$ without losing generality. Set $I = [0, 1]$ and $H_{ij} = (2i - j)I^d \times (2j - 1)I^d$ for $i, j \in \{0, 1\}$. Applying Lemma 3.1 in the new coordinates yields $\delta > 0$ and a set $V \subset H_{00}$ of positive measure such that $(x + i\delta\Delta, y + j\delta\Delta) \in U_\epsilon \cap H_{ij}$ for $i, j \in \{0, 1\}$. The perturbation $\tilde{h}_\delta = \tilde{h}$ of (3) is again well-defined, and its marginals vanish. Moreover, $h_\delta = h + \epsilon \tilde{h}_\delta \in \Gamma$ is a feasible competitor since $|\tilde{h}_\delta| \leq 1_{U_\epsilon}$ and $h \in (\epsilon, \bar{h} - \epsilon)$ on U_ϵ. The change in cost produced by this perturbation is

$$
\begin{aligned}
& c[h_\delta] - c[h] \\
= \ & \epsilon \int c(x, y) \tilde{h}_\delta(x, y) d^d x d^d y \\
= \ & \epsilon \int_V [c(x, y) + c(x + \delta\Delta, y + \delta\Delta) - c(x + \delta\Delta, y) - c(x, y + \delta\Delta)] d^d x d^d y \\
= \ & \epsilon \delta^2 \int_V \left[\int_0^1 \int_0^1 \sum_{i,j=1}^n D^2_{x^i y^j} c(x + s\delta\Delta, y + t\delta\Delta) ds dt \right] d^d x d^d y.
\end{aligned}
$$

In this formula, the arguments z of the continuous mixed partials $D^2_{x^i y^j} c$ all lie within distance $\delta\sqrt{2d}$ of a point z_0 at which $\sum_{i,j} D^2_{x^i y^j} c(z_0) = -n$. Thus for δ small enough, the perturbed cost $c[h_\delta - h] < 0$ is negative, precluding optimality of h. The contrapositive implies the only optimizers h of c are extreme points of Γ. $\square$

Corollary 3.4 (Uniqueness of Optimizer). *Under the same hypotheses, the minimum in Theorem 3.3 is uniquely attained.*

Proof. If h_0 and h_1 both minimize $c[h]$ on Γ, then so does $h_{1/2} = \frac{1}{2}(h_0 + h_1)$ since $c[\,\cdot\,]$ is linear and Γ is convex. Theorem 3.3 then asserts extremality of $h_{1/2}$ in Γ, so $h_0 = h_1$. This is the desired uniqueness. $\square$

4. SIMULATIONS AND SYMMETRIES

In case $\bar{h}(x, y) = const$, the problem has symmetries which limit the possible solutions. After introducing these symmetries, we use them to establish a new class of examples for which the optimal transport can be displayed explicitly. These include the two-by-two checkerboard (Example 1.1 of [KM11] or Corollary 4.3 below) as a particular case.

Figure 1 shows a simulation of the optimal solutions with uniform marginals for the distance squared cost on $I \times I$ with $\bar{h} = 3$ and with $\bar{h} = \frac{3}{2}$. Red represents the region W where the $\bar{h}$ constraint is saturated; in the complementary blue region, no transportation occurs. These computer simulations were originally presented to us by Brian Wetton who remarked on the symmetry manifested between the $\bar{h} = 3$ case and the $\bar{h} = \frac{3}{2}$ case. This symmetry is explained by the following lemma, which applies to any pair of Hölder conjugates p and q.
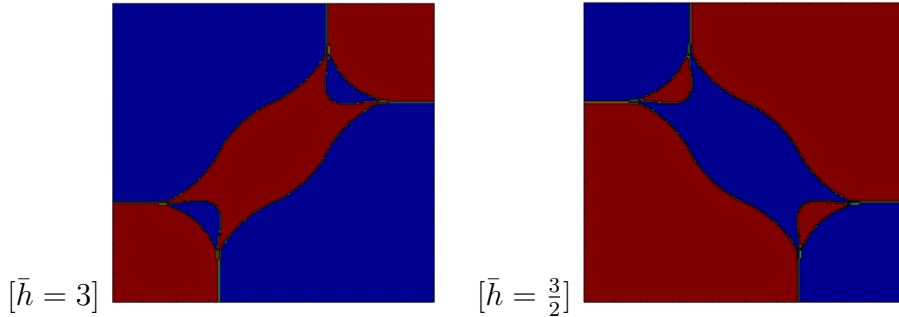

$[\bar{h} = 3]$ $[\bar{h} = \frac{3}{2}]$

FIGURE 1. Red represents the saturation set W given $\bar{h}$. Note the symmetry between (a) $\bar{h} = 3$ and (b) $\bar{h} = 3/2$.

Lemma 4.1 (Symmetries). *Let $\Omega, \Lambda \subset \mathbf{R}^d$ be bounded sets with unit volume. Set $f = 1_\Omega$ and $g = 1_\Lambda$, where 1_Ω denotes the indicator function of the set Ω. Let $\bar{h} = \bar{h}_p := p 1_{\Omega \times \Lambda}$ have constant density $p > 1$ on the product $\Omega \times \Lambda$, and $c(x, y) = -x \cdot y$. Given any set $W \subset \Omega \times \Lambda$, let $R(W)$ denote its image under the reflection $R(x, y) = (x, -y)$, and $\tilde{W} := (\Omega \times \Lambda) \setminus W$ its set theoretic complement. If $p 1_W \in \Gamma^{\bar{h}_p}(1_\Omega, 1_\Lambda)$ then $q 1_{\tilde{W}} \in \Gamma^{\bar{h}_q}(1_\Omega, 1_\Lambda)$, where $p^{-1} + q^{-1} = 1$ are Hölder conjugates. Moreover $c[1_W] + b(\Omega) \cdot b(\Lambda) = -c[1_{\tilde{W}}] = c[1_{R(\tilde{W})}]$, where $b(\Omega) = \int_\Omega x$*

is the center of mass of Ω. Thus $p1_W$ minimizes c on $\Gamma^{\bar{h}_p}(1_\Omega, 1_\Lambda)$ if and only if $q1_{R(\tilde{W})}$ minimizes c on $\Gamma^{q1_{R(\Omega \times \Lambda)}}(1_\Omega, 1_{-\Lambda})$.

Proof. For $W \subset \Omega \times \Lambda$ set $W(x) = \{y \in \mathbf{R}^d \mid (x, y) \in W\}$ and $W^{-1}(y) = \{x \in \mathbf{R}^d \mid (x, y) \in W\}$. Notice $p1_W \in \Gamma^{\bar{h}_p}(1_\Omega, 1_\Lambda)$ if and only if $p|W(x)| = 1$ and $p|W^{-1}(y)| = 1$ for a.e. $(x, y) \in \Omega \times \Lambda$. From $|W(x)| + |\tilde{W}(x)| = 1$ we conclude $|\tilde{W}(x)| = 1 - \frac{1}{p} = \frac{1}{q}$ for a.e. $x \in \Omega$, and similarly $|\tilde{W}^{-1}(y)| = \frac{1}{q}$. Thus $q1_{\tilde{W}} \in \Gamma^{\bar{h}_q}(1_\Omega, 1_\Lambda)$.

On the other hand,

$$
\begin{aligned}
c[1_W] + c[1_{\widetilde{W}}] &= \int_W c(x, y) + \int_{\widetilde{W}} c(x, y) \\
&= \int_{\Omega \times \Lambda} -x \cdot y \\
&= -\int_\Omega x \cdot \int_\Lambda y \\
&= -b(\Omega) \cdot b(\Lambda)
\end{aligned}
$$

and

$$
\begin{aligned}
c[1_{R(\widetilde{W})}] &= \int_{R(\widetilde{W})} c(x, y) \\
&= \int_{\widetilde{W}} c(x, -y) \\
&= -\int_{\widetilde{W}} c(x, y) \\
&= -c[1_{\widetilde{W}}],
\end{aligned}
$$

which imply the remaining assertions. $\qquad\square$

The next proposition shows these elementary symmetries yield a broad class of examples in the self-dual case $p = 2 = q$.

Proposition 4.2 (Universal optimizer for a balanced set with self-dual constraint)**.** *Fix uniform densities $f = 1_\Omega$ and $g = 1_\Lambda$ on two bounded Lebesgue sets $\Omega, \Lambda \subset \mathbf{R}^d$ of unit volume. Let $\bar{h} = \bar{h}_2 := 2 \cdot 1_{\Omega \times \Lambda}$ have constant density 2 on $\Omega \times \Lambda$, and fix $c(x, y) = -x \cdot y$. If Ω and Λ are balanced, meaning $\Omega = -\Omega$, the minimizer $\bar{h}_2 1_W$ of c on $\Gamma^{\bar{h}_2}(1_\Omega, 1_\Lambda)$ satisfies $W = R(\tilde{W}) = -R(\tilde{W})$ (up to sets of measure zero), where $R(x, y) = (x, -y)$. It follows that $W = \{(x, y) \in \Omega \times \Lambda \mid x \cdot y > 0\}$.*

Proof. Note that $\Gamma := \Gamma(1_\Omega, 1_\Lambda)^{\bar{h}} \neq \emptyset$ since it contains $1_{\Omega \times \Lambda}$. Thus there exists $\bar{h}_2 1_W$ minimizing c on Γ as in [KM11]. Lemma 4.1 ensures $\bar{h}_2 1_{R(\tilde{W})}$ also minimizes c on Γ, as do $\bar{h}_2 1_{-W}$ and hence $\bar{h}_2 1_{R(-\tilde{W})} =$

$\bar{h}_2 1_{-R(\tilde{W})}$. The uniqueness established in Corollary 3.4 therefore implies $W = -W = R(\tilde{W}) = -R(\tilde{W})$, up to sets of measure zero. For each Lebesgue point $z_0 = (x_0, y_0) \in W$ of full density, this shows W also contains $(-x_0, -y_0)$ but not $R(z_0) = (x_0, -y_0)$ nor $-R(z_0) = (-x_0, y_0)$. Therefore choose a Lebesgue point $z_0 = (x_0, y_0) \in W$ of full density with $x_0 \neq 0 \neq y_0$ and $x_0 \cdot y_0 \neq 0$, noting almost all points in W have this form. For $r > 0$ sufficiently small, the ball $Z_1 := Z = B_r(z_0)$ will be disjoint from its reflections $Z_3 = -Z$, $Z_2 = R(Z)$, and $Z_4 = -R(Z)$, and moreover $(x, y) \in Z$ will imply $x \cdot y$ has the same sign as $x_0 \cdot y_0$. We claim $x_0 \cdot y_0 > 0$. Otherwise $\tilde{h} = -1_{Z_1 \cup Z_3} + 1_{Z_2 \cup Z_4}$ would be a feasible perturbation, and $c(x, y) + c(-x, -y) > 0 > c(x, -y) + c(-x, y)$ for all $(x, y) \in Z$ shows $\bar{h}_2 1_W + \tilde{h}$ lowers the cost c in $\Gamma(f, g)^{\bar{h}_2}$. This contradicts the minimality of $\bar{h}_2 1_W$. Thus, up to sets of measure zero, W must be contained in $W' := \{(x, y) \in \Omega \times \Lambda \mid x \cdot y > 0\}$. On the other hand, the fact that Ω and Λ are balanced makes it easy to check feasibility of $\bar{h}_2 1_{W'}$. Feasibility of W then shows the containment $W \subset W'$ cannot be strict, apart from a set of measure zero, so $W = W'$ as desired. □

As an immediate corollary, we recover Example 1.1 of [KM11], displayed in Figure 2. Note this analytical example (like the numerical ones preceding) dispels a number of natural conjectures about the optimizing set by demonstrating that its topology need not be simple and its boundary need not be smooth. Symmetry and self-duality gives
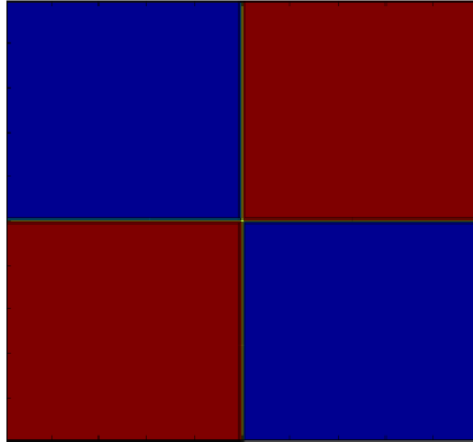

FIGURE 2. The two-by-two checkerboard solves $\bar{h} = 2$.

a much more satisfactory explanation for its singular nature than the

original argument, which was based on guessing a solution to the linear program dual to (2).

Corollary 4.3 (The 2×2 checkerboard revisited)**.** *Taking* $\Omega = \Lambda = [-\frac{1}{2}, \frac{1}{2}]$, *the preceding proposition shows the minimizer* $\bar{h}_2 1_W$ *of* $c[\cdot]$ *on* $\Gamma^{\bar{h}_2}(1_\Omega, 1_\Omega)$ *to be given by* $W = [-\frac{1}{2}, 0]^2 \cup [0, \frac{1}{2}]^2$.

5. Afterword

When transport capacity between x and $y \in \mathbf{R}^d$ is constrained by a density $\bar{h} \in L^1(\mathbf{R}^d \times \mathbf{R}^d)$, the 'all or nothing' (a.k.a. *bang-bang*) characterization of extremal plans $h \in \Gamma^{\bar{h}}(f, g)$ makes optimal transport between f onto g appear easier to analyze than the unconstrained problem $\bar{h} = +\infty$ [AKM11]. Nevertheless, the simple examples with capacity constraints solved above using numerical or theoretical methods display an unexpectedly rich range of phenomena and raise new questions of their own. Although not discussed here, the linear program dual to the capacity constrained problem [KM11] turns out to be more complicated to solve than that of the unconstrained problem $\bar{h} = +\infty$ [RR98] [V09]. We hope to analyze this difficulty in future work.

References

[AKM11] , N. Ahmad, H.K. Kim, and R.J. McCann. Optimal transportation, topology and uniqueness. *Bull. Math. Sci.* **1** (2011) 13–32.

[B87] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs, *C. R. Acad. Sci. Paris Sér. I Math.*, **305** (1987), 805–808.

[CJS08] G. Carlier, C. Jimenez, and F. Santambrogio. Optimal transportation with traffic congestion and Wardrop equilibria. *SIAM J. Control Optim.*, **47** (2008), 1330-1350.

[CP89] M.J.P. Cullen and J. Purser. Properties of the Lagrangian semigeostrophic equations. *J. Atmos. Sci.* **46** (1989) 2684-2697.

[K42] L. Kantorovich. On the translocation of masses, *C.R. (Doklady) Acad. Sci. URSS (N.S.)* **37** (1942), 199–201.

[KM11] J. Korman and R. McCann, Optimal transportation with capacity constraints. To appear in *Trans. Amer. Math. Soc.*

[MPW10] R.J. McCann, B. Pass, and M. Warren. Rectifiability of Optimal Transportation Plans, *Canad. J. Math.* **64** (2012) 924–934.

[MG10] R. J. McCann and Nestor Guillen. Five lectures on optimal transportation: geometry, regularity and applications. In *Analysis and Geometry of Metric-Measure Spaces. Lecture Notes of the Séminaire Mathématiques Supériuere (SMS) Montréal 2011,* G. Dafni et al, eds. (2013) 145-180.

[Mo81] G. Monge. 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*, pages 666–704, 1781.

[RR98] Z. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol I Theory, Vol II Applications*. Springer Verlag, New York, 1998.

[V09] C. Villani. *Optimal Transport, Old and New*, vol. **334** of Grundlehren der Mathematischen Wissenschaften [Fundamental principles of Mathematical Sciences]. Springer, New York, 2009.

E-mail address: `jkorman@math.toronto.edu`

E-mail address: `mccann@math.toronto.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, TORONTO ONTARIO M5S 2E4 CANADA