An elementary approach to linear programming duality with application to capacity constrained transport*

Jonathan Korman[†], Robert J. McCann[†], and Christian Seis[†‡]

September 22, 2014

Abstract

An approach to linear programming duality is proposed which relies on quadratic penalization, so that the relation between solutions to the penalized primal and dual problems becomes affine. This yields a new proof of Levin's duality theorem for capacity-constrained optimal transport as an infinite-dimensional application.

1 Introduction

Given a distribution of sources (manufacturers) $f(x)$ and sinks (consumers) $g(y)$, and a function $c(x,y)$ that measures the cost of transporting a unit of mass from $x \in \mathbf{R}^m$ to $y \in \mathbf{R}^n$, the optimal transport problem of Monge [10] and Kantorovich [2] seeks to minimize the total cost required to transport f to g. We consider a variant of that classical problem by imposing a limitation on the amount of mass that is allowed to be transferred from x to y: The *capacity constrained* optimal transport problem.

For two given probability distributions $f \in L^1(\mathbf{R}^m)$ and $g \in L^1(\mathbf{R}^n)$, and a fixed nonnegative function $\overline{h} \in L^\infty(\mathbf{R}^m \times \mathbf{R}^n)$, we denote by $\Gamma^{\overline{h}}(f,g)$ the set of all nonnegative measurable joint densities that are bounded by $\overline{h}$, i.e., $f(x) = \int h(x,y)\,dy$, $g(y) = \int h(x,y)\,dx$, and $0 \leq h \leq \overline{h}$. Necessary and

 The authors are grateful to Terry Rockafellar for insightful comments on an earlier version of this manuscript. They are also pleased to acknowledge partial support by NSERC grant 217006-08. RJM acknowledges the kind hospitality of the Mathematical Sciences Research Institute at Berkeley CA, where part of this work was performed.

†Department of Mathematics, University of Toronto, Toronto ON Canada M5S 2E4 jkorman@math.toronto.edu, mccann@math.toronto.edu, cseis@math.toronto.edu

‡Current address: Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany

sufficient conditions for $\Gamma^{\overline{h}}(f,g)$ to be nonempty are given by Kellerer [4, 3] and Levin [8], namely $\Gamma^{\overline{h}}(f,g) \neq \emptyset$ if and only if

$$f(A) + g(B) - \bar{h}(A \times B) \leq 1 \quad \text{for any Borel measurable } A \subset \mathbf{R}^m,\ B \subset \mathbf{R}^n.$$

Throughout this article, we will always assume that these conditions are satisfied. Finally, let $c \in L^1_{\text{loc}}(\mathbf{R}^m \times \mathbf{R}^n)$ and

$$I_c(h) := \iint c(x,y)h(x,y)\,dxdy.$$

The *optimal transportation problem with capacity constraints* consists in finding and studying optimal transference plans $h_0 \in \Gamma^{\overline{h}}(f,g)$ for the total cost functional I_c:

$$I_c(h_0) \;=\; \min_{h \in \Gamma^{\overline{h}}(f,g)} I_c(h).$$

Optimal transference plans always exist as can be easily established via the direct method of calculus of variations. Regarding the capacity constrained optimal transport as an infinite-dimensional linear programming problem, it is not surprising that some of the minimizers are extreme points of the convex polytope $\Gamma^{\overline{h}}(f,g)$. Such extreme points can be characterized by $h_0 = \overline{h}\chi_W$ for some Lebesgue measurable set W in $\mathbf{R}^m \times \mathbf{R}^n$ [6]. Under suitable conditions on the cost function c, minimizers are unique [5].

In this short manuscript, we address the linear programming duality for capacity constrained optimal transport. Although such a duality was already established by Levin* (see Theorem 4.6.14 of [11]), we present an alternative proof here. While Rockafellar-Fenchel dualities (including Levin's, and the Kantorovich's duality for classical optimal transport, cf. [12, Ch. 1]) are usually proved using an asbtract minimax argument with the Hahn–Banach theorem at its core, our new proof is rather elementary and is based on a quadratic approximation of the linear program, cf. Section 2. Combining the techniques presented in the following with some of the results derived by the authors in a companion paper [7], we also provide a new *elementary* proof of Kantorovich's duality.

We prove Levin's duality under the additional assumption that the capacity bound $\overline{h}$ is compactly supported, and we write $\overline{W} = \text{spt}(\overline{h})$. Notice that under this hypothesis, h, f, and g are bounded and compactly supported, so that in particular $f,\ g,\ h \in L^p$ for any $1 \leq p \leq \infty$.

Before stating Levin's duality theorem, we introduce some notation. Given a function $\zeta = \zeta(x,y)$ defined on $\mathbf{R}^m \times \mathbf{R}^n$, we write $\langle\zeta\rangle_x$ and $\langle\zeta\rangle_y$ for the x- and y-marginals of ζ, i.e., $\langle\zeta\rangle_x := \int \zeta(x,y)\,dy$ and $\langle\zeta\rangle_y := \int \zeta(x,y)\,dx$. The integral over the product space is denoted by $\langle\langle\zeta\rangle\rangle$, i.e., $\langle\langle\zeta\rangle\rangle := \iint \zeta(x,y)\,dxdy$. Likewise, if $\zeta = \zeta(x)$ or $\zeta = \zeta(y)$, we simply write $\langle\zeta\rangle$ to denote the integral

*In a private communication, Rachev and Rüschendorf attribute Theorem 4.6.14 of [11] to a handwritten manuscript of Levin; we are unsure where or whether it was subsequently published.

over $\mathbf{R}^m$ or $\mathbf{R}^n$, respectively. With the above notation, the total cost functional becomes

$$I_c(h) \ := \ \langle\langle ch \rangle\rangle.$$

We introduce some further notation. Let

$$J(u,v,w) \ := \ -\langle uf \rangle - \langle vg \rangle + \langle\langle w\overline{h} \rangle\rangle,$$

and

$$\begin{aligned} \mathrm{Lip}_c^{\overline{h}} \ &= \ \big\{(u,v,w) \in L^1(f\,dx) \times L^1(g\,dy) \times L^1(\bar{h}\,dxdy) : \\ &\quad u(x) + v(y) - w(x,y) + c(x,y) \geq 0 \text{ and } w(x,y) \leq 0\big\}\,. \end{aligned}$$

Here, we use the notation that $L^1(\mu)$ is the class of all Lebesgue integrable functions with respect to the measure μ. Obviously, $J(u,v,w)$ is well-defined on $\mathrm{Lip}_c^{\overline{h}}$.

Our main result is the following

Theorem 1 (Levin's duality)**.** *Let $0 \leq \overline{h} \in L^\infty(\mathbf{R}^m \times \mathbf{R}^n)$ be compactly supported and $f \in L^1(\mathbf{R}^m)$ and $g \in L^1(\mathbf{R}^n)$ be two probability densities such that $\Gamma^{\overline{h}}(f,g) \neq \emptyset$. Suppose that $c \in L^1_{\mathrm{loc}}(\mathbf{R}^m \times \mathbf{R}^n)$. Then*

$$\min_{h \in \Gamma^{\overline{h}}(f,g)} I_c(h) \ = \ \sup_{(u,v,w) \in \mathrm{Lip}_c^{\overline{h}}} J(u,v,w).$$

In [7], the authors prove (under some additional assumptions) that the supremum on the right is attained by triple of functions.

In the following Section 2, we illustrate the method of this paper by considering an analogous problem in finite-dimensions. The proof of Theorem 1 is presented in Section 3.

2 Finite-dimensional linear programming duality

In this section we illustrate the method of this paper by sketching a non-standard proof of the finite-dimensional linear programming duality. For $A \in \mathbf{R}^{m \times n}$, $c \in \mathbf{R}^n$, and $b \in \mathbf{R}^m$, duality asserts

$$I_* := \inf_{y \geq 0,\, A^T y = c} b \cdot y \ = \ \sup_{Ax \leq b} c \cdot x, \tag{1}$$

where, of course, $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$, cf. [9, Ch. 4]. We understand the inequalities $y \geq 0$ and $Ax \leq b$ componentwise. We shall take for granted that the side of the problem with equality constraints (the infimum above) is (i) feasible and (ii) has sufficient compactness or coercivity properties. For example, $Y_\varepsilon := \{0 \leq y \in \mathbf{R}^m \mid b \cdot y + \frac{1}{2\varepsilon}|A^T y - c|^2 \leq I_*\}$ compact and non-empty for $\epsilon > 0$ sufficiently small is enough. We do not try to formalize (ii), but rather point out that it may also be satisfied in various problems of interest due, for example, to the presence of additional inequality constraints such as upper bounds on the components of y. In particular, this would be the case in the finite-dimensional version of Levin's duality. The advantage of our approach to (1) is that, in the presence of sufficient compactness or coercivity, it generalizes in a straightforward way to infinite-dimensional problems. This we shall see in the subsequent section, where we give a new (and unconditional) proof of Levin's duality, Theorem 1.

The basic idea in our proof of (1) is to relax the equality constraint in the minimization problem by adding a penalizing quadratic term to the linear function. That is, we consider the quadratic function

$$I^\varepsilon(y) := b \cdot y + \frac{1}{2\varepsilon}|A^T y - c|^2,$$

and minimize I^ε over all y such that $y \geq 0$. Relaxing a minimization problem by approximating hard by soft constraints is a fairly standard procedure in theoretical and numerical optimization, as well as in the calculus of variations, whether it be to regularize singular problems or simply to extend the class of admissible competitors (e.g. [1]). In particular, when dealing with constraints of different kinds, as in the capacity constrained optimal transport problem, relaxing some of these constraints eventually simplifies the computation of the Euler–Lagrange equation dramatically, see e.g. Lemma 3 below.

The key observation in our analysis is a duality theorem for the relaxed problem,

$$\min_{y\geq 0} I^\varepsilon(y) \;=\; \max_{Ax\leq b} J^\varepsilon(x), \tag{2}$$

provided that the minimum on the left is attained, and where $J^\varepsilon(x) = c \cdot x - \frac{\varepsilon}{2}|x|^2$. The derivation of the "inf $\geq$ sup"-inequality is standard: Using $y \geq 0$ and $b \geq Ax$, we have

$$\begin{aligned} I^\varepsilon(y) &\geq Ax \cdot y + \frac{1}{2\varepsilon}|A^T y - c|^2 \\ &= c \cdot x + x \cdot \left(A^T y - c\right) + \frac{1}{2\varepsilon}|A^T y - c|^2 \\ &= c \cdot x - \frac{\varepsilon}{2}|x|^2 + \frac{\varepsilon}{2}|x + \frac{1}{\varepsilon}\left(A^T y - c\right)|^2 \\ &\geq J^\varepsilon(x), \end{aligned}$$

and the statement follows upon taking the infimum on the left and the supremum on the right. Moreover, the above inequality turns into an equality for any pair $(y_\varepsilon, x_\varepsilon)$ with $y_\varepsilon \geq 0$, $Ax_\varepsilon \leq b$,

$$x_\varepsilon = \frac{1}{\varepsilon}\left(c - A^T y_\varepsilon\right) \quad \text{and} \quad (b - Ax_\varepsilon) \cdot y_\varepsilon = 0. \tag{3}$$

In particular, if such a pair exists, we must have

$$\min_{y\geq 0} I^\varepsilon(y) \;=\; I^\varepsilon(y_\varepsilon) \;=\; J^\varepsilon(x_\varepsilon) \;=\; \max_{Ax\leq b} J^\varepsilon(x),$$

that is (2) holds. The existence of $(y_\varepsilon, x_\varepsilon)$ with $y_\varepsilon \geq 0$, $Ax_\varepsilon \leq b$, and (3) is a simple but crucial insight: when compactness or coercivity (ii) implies the minimum of I^ε on the non-negative orthant to be attained by y_ε we find

$$0 \leq \frac{\partial I^\varepsilon}{\partial y^i}(y_\varepsilon) = \left(b + \frac{1}{\varepsilon} A \left(A^T y_\varepsilon - c\right)\right)_i$$

for each $i \leq m$, and the derivative vanishes for those i such that $y_\varepsilon^i > 0$. Thus x_ε defined as in the first part of (3) satisfies $b \geq Ax_\varepsilon$, and the second part of (3) holds as well. Hence, $(y_\varepsilon, x_\varepsilon)$ is dual pair with the desired properties. This proves (2) under the assumption that the minimum of I^ε is attained.

Whether the minimum is attained in this finite-dimensional toy problem certainly depends on the particular choice of the matrix A and objective b, and is encoded in our coercivity hypothesis. Notice, however, that existence of minimizers is obvious when including the "capacity constraint" $y \leq \bar{y}$ for some $\bar{y} \in \mathbf{R}^m$ into the problem, which would actually correspond to the real finite-dimensional analog for the problem considered in this paper. To keep the discussion in this section as elementary as possible, we simply drop this capacity constraint and instead invoke (ii) to assume the existence of a minimizer y_ε of I^ε at this point.

There is a remarkable *affine* relation (3) between the maximizer x_ε of J^ε and the minimizer y_ε of I^ε. This relation, however, is not surprising, since (3) can also be derived as the first order necessary condition for the dual maximum problem, which is linear in x_ε since J^ε is quadratic, and in which y_ε plays the role of the Lagrange multiplier associated with the constraint $Ax \leq b$.

Finally, to obtain (1) from (2) requires the limit $\varepsilon \downarrow 0$. Invoking our compactness assumption (ii) once more, we extract a convergent subsequence $y_\varepsilon \to \tilde{y}$ that we do not relabel. Since the feasible set for the constrained problem (1) is also feasible for the penalized problem, we have

$$I^\varepsilon(y_\varepsilon) = b \cdot y_\varepsilon + \frac{1}{2\varepsilon}|A^T y_\varepsilon - c|^2 \leq \inf_{y \geq 0,\, A^T y = c} b \cdot y < +\infty \tag{4}$$

which is finite by hypothesis (i). The limit $\varepsilon \to 0$ along our subsequence shows $A^T \tilde{y} = c$, whence $\tilde{y}$ optimizes the constrained problem:

$$b \cdot \tilde{y} = \inf_{y \geq 0,\, A^T y = c} b \cdot y.$$

Along the same subsequence, from (4) we then deduce

$$\varepsilon |x_\varepsilon|^2 = \frac{1}{\varepsilon}|A^T y_\varepsilon - c|^2 \to 0$$

and

$$J^\varepsilon(x_\varepsilon) = I^\varepsilon(y_\varepsilon) \to \inf_{y \geq 0,\, A^T y = c} b \cdot y$$

as $\varepsilon \to 0$. Thus

$$b \cdot \tilde{y} - \lim_{\varepsilon \to 0} c \cdot x_\varepsilon = \lim_{\varepsilon \to 0} I^\varepsilon(y_\varepsilon) - J^\varepsilon(x_\varepsilon) - \varepsilon |x_\varepsilon|^2 = 0,$$

which establishes (1). Notice the subsequence x_ε need not converge for this argument, nor do we claim the supremum over x is attained.

3 Proof of Levin's duality theorem

Theorem 1 is an immediate consequence of the following two Propositions:

Proposition 1. *The hypotheses of Theorem 1 imply*

$$\inf_{h\in\Gamma^{\overline{h}}(f,g)} I_c(h) \;\geq\; \sup_{(u,v,w)\in \mathrm{Lip}_c^{\overline{h}}} J(u,v,w). \tag{5}$$

Proposition 2. *The hypotheses of Theorem 1 imply existence of a sequence* $\{(u_\varepsilon, v_\varepsilon, w_\varepsilon)\}_{\varepsilon\downarrow 0}$ *in* $\mathrm{Lip}_c^{\overline{h}}$ *such that*

$$I_c(h_0) \;=\; \lim_{\varepsilon\downarrow 0} J(u_\varepsilon, v_\varepsilon, w_\varepsilon), \tag{6}$$

where h_0 *is a minimizer of the form* $h_0 = \overline{h}\chi_W$.

The first Proposition is easily established:

Proof of Proposition 1. For any coupling $h \in \Gamma^{\overline{h}}(f,g)$ with $I_c(h)$ finite, and $(u,v,w) \in \mathrm{Lip}_c^{\overline{h}}$ we have

$$\begin{aligned} I_c(h) &= -\langle uf\rangle - \langle vg\rangle + \langle\langle w\overline{h}\rangle\rangle + \langle\langle (c+u+v-w)\, h\rangle\rangle + \langle\langle w(h-\overline{h})\rangle\rangle \\ &\geq J(u,v,w), \end{aligned}$$

where in the first line we have used the marginal constraint on h and in the second line we applied the definition of $\mathrm{Lip}_c^{\overline{h}}$ together with the fact that $0 \leq h \leq \overline{h}$. Now, the inequality in (5) follows immediately upon taking the supremum on the right and the infimum on the left. □

The remainder of the paper is devoted to the proof of Proposition 2.

We introduce a *relaxed* version of the optimal transportation problem with capacity constraints. Let $\varepsilon > 0$ denote a small number. We define the relaxed transportation cost

$$I_c^\varepsilon(h) \;=\; \langle\langle ch\rangle\rangle + \frac{1}{2\varepsilon}\|\langle h\rangle_x - f\|_2^2 + \frac{1}{2\varepsilon}\;\|\langle h\rangle_y - g\|_2^2$$

using the L^2 norms $\|\cdot\|_2$ on $\mathbf{R}^m$ and $\mathbf{R}^n$. Notice that $I_c^\varepsilon(h_0) = I_c(h_0)$. Furthermore, for $(u,v,w) \in \mathrm{Lip}_c^{\overline{h}}$ such that u and v are both square-integrable, we consider the functional

$$J^\varepsilon(u,v,w) \;:=\; -\langle uf\rangle - \langle vg\rangle + \langle\langle w\overline{h}\rangle\rangle - \frac{\varepsilon}{2}\|u\|_2^2 - \frac{\varepsilon}{2}\|v\|_2^2.$$

We can extend J^ε to a functional all over $\mathrm{Lip}_c^{\overline{h}}$ by setting $J^\varepsilon(u,v,w) := -\infty$ if $u \notin L^2(\mathbf{R}^m)$ or $v \notin L^2(\mathbf{R}^n)$.

In a first step, we derive the analogous statement to Proposition 1 for the relaxed problem.

Lemma 1 (Easy direction of relaxed duality)**.** *For $\epsilon > 0$, the hypotheses of Theorem 1 imply*

$$\inf_{0\le h\le \overline{h}} I_c^\varepsilon(h) \;\ge\; \sup_{(u,v,w)\in \mathrm{Lip}_c^{\overline{h}}} J^\varepsilon(u,v,w). \tag{7}$$

Proof. Without lost of generality we may choose $0 \le h \le \overline{h}$ and $(u,v,w) \in \mathrm{Lip}_c^{\overline{h}}$ such that $I_c^\varepsilon(h)$ and $J^\varepsilon(u,v,w)$ are both finite. A short computation shows that $I_c^\varepsilon(h)$ can be rewritten as

$$\begin{aligned} I_c^\varepsilon(h) \;&=\; -\langle uf\rangle - \langle vg\rangle + \langle\langle w\overline{h}\rangle\rangle - \frac{\varepsilon}{2}\|u\|_2^2 - \frac{\varepsilon}{2}\|v\|_2^2 \\ &\quad + \langle\langle (c+u+v-w)\,h\rangle\rangle + \langle\langle w(h-\overline{h})\rangle\rangle \\ &\quad + \frac{1}{2\varepsilon}\,\|\langle h\rangle_x - f - \varepsilon u\|_2^2 + \frac{1}{2\varepsilon}\,\|\langle h\rangle_y - g - \varepsilon v\|_2^2\,. \end{aligned}$$

By the definition of $\mathrm{Lip}_c^{\overline{h}}$ and $J^\varepsilon(u,v,w)$, recalling that $0 \le h \le \overline{h}$, and observing that the term in the last line is trivially nonnegative, it follows that

$$I_c^\varepsilon(h) \;\ge\; J^\varepsilon(u,v,w).$$

Taking the infimum on the left hand side and the supremum on the right hand side yields (7).

□

We next address existence of minimizers for the relaxed problem.

Lemma 2 (Existence of minimizers and uniqueness of relaxed marginals)**.** *The hypotheses of Theorem 1 imply existence of a minimizer h_ε of I_c^ε, and h_ε can be chosen of the form $h_\varepsilon = \overline{h}\chi_{W_\varepsilon}$ for some Lebesgue measurable set W_ε in $\mathbf{R}^m \times \mathbf{R}^n$. Moreover, if $\tilde{h}_\varepsilon$ is another minimizer of I_c^ε, then $\langle h_\varepsilon\rangle_x = \langle \tilde{h}_\varepsilon\rangle_x$ and $\langle h_\varepsilon\rangle_y = \langle \tilde{h}_\varepsilon\rangle_y$.*

Existence of minimizers and uniqueness of their marginals follow by standard arguments. We provide the proof for the convenience of the reader.

Proof. Since h_0 is admissible for I_c^ε with $I_c^\varepsilon(h_0) = I_c(h_0)$, it follows that $-\|\overline{h}\|_\infty \|c\|_{L^1(\overline{W})} \le \inf I_c^\varepsilon(h) \le I_c(h_0) < \infty$, where the infimum is taken over all admissible h. Let $\{h_\nu\}_{\nu\uparrow\infty}$ denote a minimizing sequence. By the compactly supported bound $0 \le h_\nu \le \overline{h} \in L^\infty$, we see h_ν is bounded in L^p for all $p \ge 1$. The same is true for f and g since $\Gamma^{\overline{h}}(f,g) \ne \emptyset$. Thus we can find an L^∞-function h_ε satisfying $0 \le h_\varepsilon \le \overline{h}$ and we can extract a subsequence converging to h_ε weakly-$\star$ in L^∞, such that the subsequences $\{\langle h_\nu\rangle_x - f\}_{\nu\uparrow\infty}$ and $\{\langle h_\nu\rangle_y - g\}_{\nu\uparrow\infty}$ also converge weakly in L^2 towards $\langle h_\varepsilon\rangle_x - f$ and $\langle h_\varepsilon\rangle_y - g$, respectively. Without relabeling the subsequences, we then have

$$\begin{aligned} \|\langle h_\varepsilon\rangle_x - f\|_2 \;&\le\; \liminf_{\nu\uparrow\infty} \|\langle h_\nu\rangle_x - f\|_2, \\ \|\langle h_\varepsilon\rangle_y - g\|_2 \;&\le\; \liminf_{\nu\uparrow\infty} \|\langle h_\nu\rangle_y - g\|_2, \end{aligned}$$

by the lower semi-continuity of the L^2 norm with respect to weak L^2 convergence. Moreover, since $c \in L^1_{\mathrm{loc}}$ and h_ε, h_ν are supported in $\overline{W}$, weak-$\star$ convergence guarantees that

$$\langle\langle c h_\varepsilon \rangle\rangle = \lim_{\nu\uparrow\infty} \langle\langle c h_\nu \rangle\rangle.$$

Hence, by combining the above (in)equalities, we have

$$I_c^\varepsilon(h_\varepsilon) \;\leq\; \liminf_{\nu\uparrow\infty} I_c^\varepsilon(h_\nu).$$

Since $\{h_\nu\}_{\nu\uparrow\infty}$ was a minimizing sequence, it turns out that h_ε minimizes I_c^ε. By strict convexity of the relaxed optimization problem, h_ε has unique marginals. Moreover, since h_ε minimizes I_c in the class $\Gamma^{\overline{h}}(\langle h_\varepsilon\rangle_x, \langle h_\varepsilon\rangle_y)$, we can choose h_ε geometrically extreme (with respect to h_0): $h_\varepsilon = \overline{h}\chi_{W_\varepsilon}$ for some Lebesgue measurable set $W_\varepsilon \subset \overline{W}$, cf. [6]. □

In the following, we construct an approximate dual triple $(u_\varepsilon, v_\varepsilon, w_\varepsilon)$ by defining

$$u_\varepsilon \;:=\; \frac{1}{\varepsilon}\left(\langle h_\varepsilon\rangle_x - f\right), \tag{8}$$

$$v_\varepsilon \;:=\; \frac{1}{\varepsilon}\left(\langle h_\varepsilon\rangle_y - g\right), \tag{9}$$

$$w_\varepsilon \;:=\; \min\{c + u_\varepsilon + v_\varepsilon, 0\}. \tag{10}$$

The definition of w_ε entails that $c + u_\varepsilon + v_\varepsilon - w_\varepsilon \geq 0$ and $w_\varepsilon \leq 0$. Observe that by Lemma 2 these triples are determined independently of the choice of h_ε. Notice that u_ε and v_ε (but not w_ε) depend linearly on h_ε, echoing our finite dimensional model problem. In Lemma 4 below, we prove that this triple maximizes J^ε in $\mathrm{Lip}_c^{\overline{h}}$, which in turn yields the duality theorem for the relaxed problem. We can pass to the limit $\varepsilon \downarrow 0$ in this duality to prove Proposition 2.

Lemma 3 (Euler–Lagrange equations for relaxed problem)**.** *Taking h_ε and W_ε from Lemma 2, using* (8)–(10) *to define* $(u_\varepsilon, v_\varepsilon, w_\varepsilon)$ *yields*

$$c + u_\varepsilon + v_\varepsilon \begin{cases} \leq 0 & \text{a.e. in } W_\varepsilon, \\ \geq 0 & \text{a.e. in } \overline{W}\setminus W_\varepsilon. \end{cases} \tag{11}$$

Proof of Lemma 3. Let $\zeta \geq 0$ denote an arbitrary smooth test function. We give the argument for the second inequality in (11) by considering the outer perturbation

$$h_\varepsilon^\sigma := h_\varepsilon + \sigma\zeta(\overline{h} - h_\varepsilon) = \begin{cases} h_\varepsilon & \text{a.e. in } W_\varepsilon, \\ \sigma\zeta\overline{h} & \text{a.e. in } \overline{W}\setminus W_\varepsilon. \end{cases}$$

Obviously $h_\varepsilon^0 = h_\varepsilon$ and $0 \leq h_\varepsilon^\sigma \leq \overline{h}$ for $0 \leq \sigma \leq \|\zeta\|_\infty^{-1}$. Hence, by the optimality of h_ε we have $I_c^\varepsilon(h_\varepsilon^0) \leq I_c^\varepsilon(h_\varepsilon^\sigma)$, and a short computation using (8)&(9) yields

$$0 \;\leq\; \left.\frac{d}{d\sigma}\right|_{\sigma=0} I_c^\varepsilon(h_\varepsilon^\sigma) \;=\; \langle\langle (c + u_\varepsilon + v_\varepsilon)\,\zeta(\overline{h} - h_\varepsilon)\rangle\rangle.$$

This estimate holds for all smooth test functions $\zeta \geq 0$. Via the Fundamental Lemma of Calculus of Variations it immediately follows that $(c + u_\varepsilon + v_\varepsilon)(\overline{h} - h_\varepsilon) \geq 0$ almost everywhere. Moreover, since $\overline{h} - h_\varepsilon$ is nonnegative almost everywhere and positive almost everywhere in $\overline{W} \setminus W_\varepsilon$, we deduce the second inequality in (11).

The argument for the first inequality in (11) is proved similarly, we just need to consider the perturbation $h_\varepsilon^\sigma := h_\varepsilon - \sigma\zeta h_\varepsilon$ and argue as above. □

Lemma 4 (A duality theorem for the relaxed problem)**.** *Taking h_ε and W_ε from Lemma 2 and using* (8)–(10) *to define* $(u_\varepsilon, v_\varepsilon, w_\varepsilon)$ *yields*

$$I_c^\varepsilon(h_\varepsilon) \;=\; J^\varepsilon(u_\varepsilon, v_\varepsilon, w_\varepsilon). \tag{12}$$

In particular, $(u_\varepsilon, v_\varepsilon, w_\varepsilon)$ maximizes $J^\varepsilon(u, v, w)$ in $\mathrm{Lip}_c^{\overline{h}}$.

Proof. Using the definition of u_ε and v_ε, we easily compute that

$$J^\varepsilon(u_\varepsilon, v_\varepsilon, w_\varepsilon) \;=\; I_c^\varepsilon(h_\varepsilon) - \langle\langle (c + u_\varepsilon + v_\varepsilon - w_\varepsilon)\, h_\varepsilon\rangle\rangle + \langle\langle w_\varepsilon(\overline{h} - h_\varepsilon)\rangle\rangle.$$

In view of (10) and (11) we see that $(c+u_\varepsilon+v_\varepsilon-w_\varepsilon)h_\varepsilon \equiv 0$ and $w_\varepsilon(\overline{h}-h_\varepsilon) \equiv 0$. Hence, (12) follows.

In view of (7), the triple $(u_\varepsilon, v_\varepsilon, w_\varepsilon)$ is a maximizer of J^ε in $\mathrm{Lip}_c^{\overline{h}}$ because $(u_\varepsilon, v_\varepsilon, w_\varepsilon) \in \mathrm{Lip}_c^{\overline{h}}$ by construction. □

The next result shows that solutions to the relaxed problem approximate the original one "as the soft constraints become harder".

Lemma 5 (Extracting a limit from the penalized problems)**.** *The sequence $\{h_\varepsilon\}_{\varepsilon\downarrow 0}$ defined by Lemma 2 is precompact in the L^∞-weak-$\star$ topology and every limit point h_0 is a minimizer of I_c. Moreover,*

$$\lim_{\varepsilon\downarrow 0} I_c(h_\varepsilon) \;=\; I_c(h_0), \tag{13}$$

$$\lim_{\varepsilon\downarrow 0} \varepsilon\|u_\varepsilon\|_2^2 \;=\; 0, \tag{14}$$

$$\lim_{\varepsilon\downarrow 0} \varepsilon\|v_\varepsilon\|_2^2 \;=\; 0. \tag{15}$$

Proof. Since $0 \leq h_\varepsilon \leq \overline{h}$, we immediately see that a subsequence of $\{h_\varepsilon\}_{\varepsilon\downarrow 0}$ (which we will not relabel) converges weakly-$\star$ in L^∞ to some function $0 \leq \tilde{h} \leq \overline{h}$.

By the optimality of h_ε and since any minimizer $\tilde{h}_0$ of the original $\varepsilon = 0$ problem is admissible in the relaxed problem, we have the trivial bound

$$I_c^\varepsilon(h_\varepsilon) \;\leq\; I_c^\varepsilon(\tilde{h}_0) \;=\; I_c(\tilde{h}_0), \tag{16}$$

and thus weak-$\star$ convergence of $\{h_\varepsilon\}_{\varepsilon>0}$ implies that

$$\langle\langle c\tilde{h}\rangle\rangle \;=\; \lim_{\varepsilon\downarrow 0}\langle\langle c h_\varepsilon\rangle\rangle \;\overset{(16)}{\leq}\; \langle\langle c\tilde{h}_0\rangle\rangle,$$

i.e., $I_c(\tilde{h}) \le I_c(\tilde{h}_0)$. Since $0 \le \tilde{h} \le \overline{h}$, it remains to show that $\tilde{h}$ satisfies the marginal constraints

$$f = \langle \tilde{h} \rangle_x \quad \text{and} \quad g = \langle \tilde{h} \rangle_y,$$

because then $\tilde{h}$ must be a minimizer of I_c, i.e., $I_c(\tilde{h}) = I_c(\tilde{h}_0)$.

Indeed, from (16) we deduce that

$$\|f - \langle h_\varepsilon \rangle_x\|_2^2 + \|g - \langle h_\varepsilon \rangle_y\|_2^2 \;\le\; 2\varepsilon \langle\langle c\tilde{h}_0 \rangle\rangle,$$

which states that $\langle h_\varepsilon \rangle_x \to f$ and $\langle h_\varepsilon \rangle_y \to g$ in L^2. For any smooth and compactly supported test function $\zeta = \zeta(x)$, we write

$$\langle (f - \langle \tilde{h} \rangle_x)\zeta \rangle \;=\; \langle (f - \langle h_\varepsilon \rangle_x)\zeta \rangle + \langle (\langle h_\varepsilon \rangle_x - \langle \tilde{h} \rangle_x)\zeta \rangle.$$

The first integral on the right converges to zero by the L^2-convergence of the marginals stated above. The second integral can be rewritten as $\langle\langle (h_\varepsilon - \tilde{h})\zeta \rangle\rangle$ which converges to zero by L^∞-weak-$\star$ convergence. Invoking the Fundamental Lemma of Calculus of Variations, this proves that $f = \langle \tilde{h} \rangle_x$, and the analogous argument applies for the y-marginals, showing that $g = \langle \tilde{h} \rangle_y$.

Since $I_c(\tilde{h}) \le \liminf_{\varepsilon \downarrow 0} I_c^\varepsilon(h_\varepsilon)$, passing to the limit in (16), the above analysis shows that

$$\min_{h \in \Gamma^{\overline{h}}(f,g)} I_c(h) \;=\; \lim_{\varepsilon \downarrow 0} I_c(h_\varepsilon) \;=\; \lim_{\varepsilon \downarrow 0} I_c^\varepsilon(h_\varepsilon),$$

which implies (13)–(15) by the definition of u_ε and v_ε. □

We are now in the position to prove Proposition 2.

Proof of Proposition 2. We may rewrite identity (12) in terms of $J(u_\varepsilon, v_\varepsilon, w_\varepsilon)$ and $I_c(h_\varepsilon)$, that is

$$J(u_\varepsilon, v_\varepsilon, w_\varepsilon) \;=\; I_c(h_\varepsilon) + \varepsilon \|u_\varepsilon\|_2^2 + \varepsilon \|v_\varepsilon\|_2^2.$$

Invoking (13)–(15), we then have

$$\lim_{\varepsilon \downarrow 0} J(u_\varepsilon, v_\varepsilon, w_\varepsilon) \;=\; I_c(h_0),$$

i.e., equation (6). It remains to recall that $(u_\varepsilon, v_\varepsilon, w_\varepsilon) \in \mathrm{Lip}_c^{\overline{h}}$ by Lemma 4. □

References

[1] Bethuel, F., Brezis, H., and Hélein, F. *Ginzburg-Landau vortices.* Progress in Nonlinear Differential Equations and their Applications, 13. Birkhäuser Boston Inc., Boston, MA, 1994.

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

[3] Kellerer, H. G. Marginalprobleme für Funktionen. *Math. Ann. 154* (1964), 147–156.

[4] Kellerer, H. G. Maßtheoretische Marginalprobleme. *Math. Ann. 153* (1964), 168–198.

[5] Korman, J., and McCann, R. J. Optimal transport with capacity constraints. To appear in *Trans. Amer. Math. Soc.*. Preprint arXiv:1201.6404.

[6] Korman, J., and McCann, R. J. Insights into capacity constrained optimal transport. *Proc. Natl. Acad. Sci. USA 110* (2013), 10064–10067.

[7] Korman, J., McCann, R. J., and Seis, C. Dual potentials for capacity constrained optimal transport. Preprint arXiv:1307.7774, 2014.

[8] Levin, V. L. The problem of mass transfer in a topological space and probability measures with given marginal measures on the product of two spaces. *Dokl. Akad. Nauk SSSR 276*, 5 (1984), 1059–1064.

[9] Luenberger, D. G., and Ye, Y. *Linear and nonlinear programming*, third ed. International Series in Operations Research & Management Science, 116. Springer, New York, 2008.

[10] Monge, G. Mémoire sur la théorie de déblais et de remblais. *Histoire de l'Académie Royale des Science de Paris, avec les Mémoires de Mathématique et de Physique pour la même année* (1781), 666–704.

[11] Rachev, S. T., and Rüschendorf, L. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[12] Villani, C. *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.