STUDIA MATHEMATICA 209 (2) (2012)

A general duality theorem for the Monge–Kantorovich transport problem

by

MATHIAS BEIGLBÖCK (Wien), CHRISTIAN LÉONARD (Nanterre)
and WALTER SCHACHERMAYER (Wien)

Abstract. The duality theory for the Monge–Kantorovich transport problem is analyzed in a general setting. The spaces X, Y are assumed to be Polish and equipped with Borel probability measures μ and ν. The transport cost function $c : X \times Y \to [0, \infty]$ is assumed to be Borel. Our main result states that in this setting there is no duality gap provided the optimal transport problem is formulated in a suitably relaxed way. The relaxed transport problem is defined as the limiting cost of the partial transport of masses $1 - \varepsilon$ from (X, μ) to (Y, ν) as $\varepsilon > 0$ tends to zero.

The classical duality theorems of H. Kellerer, where c is lower semicontinuous or uniformly bounded, quickly follow from these general results.

1. Introduction. We consider the *Monge–Kantorovich transport problem* for Borel probability measures μ, ν on Polish spaces X, Y. See [Va, Vb] for an excellent account of the theory of optimal transportation.

The set $\Pi(\mu, \nu)$ consists of all Monge–Kantorovich *transport plans*, that is, Borel probability measures on $X \times Y$ which have X-marginal μ and Y-marginal ν. The *transport costs* associated to a transport plan π are given by

$$\langle c, \pi \rangle = \int_{X \times Y} c(x, y)\, d\pi(x, y). \tag{1.1}$$

In most applications of the theory of optimal transport, the cost function $c : X \times Y \to [0, \infty]$ is lower semicontinuous and only takes values in $\mathbb{R}_+$. But equation (1.1) makes perfect sense if the $[0, \infty]$-valued cost function is only Borel measurable. We therefore assume throughout this paper that $c : X \times Y \to [0, \infty]$ is a Borel measurable function which may very well assume the value $+\infty$ for "many" $(x, y) \in X \times Y$.

An application where the value ∞ occurs in a natural way is transport between measures on Wiener space $X = (C[0, 1], \|\cdot\|_\infty)$, where $c(x, y)$ is

2010 *Mathematics Subject Classification*: Primary 49Q20.
Key words and phrases: Monge–Kantorovich problem, duality.

DOI: 10.4064/sm209-2-4

the squared norm of $x - y$ in the Cameron–Martin space, defined to be ∞ if $x - y$ does not belong to this space. Hence in this situation the set $\{y : c(x,y) < \infty\}$ has ν-measure 0 for every $x \in X$ if the measure ν is absolutely continuous with respect to the Wiener measure on $C[0,1]$ (see [FÜa, FÜb, FÜc, FÜd]).

Going back to the general problem: the (primal) *Monge–Kantorovich problem* is to determine the primal value

$$P := P_c := \inf\{\langle c, \pi\rangle : \pi \in \Pi(\mu,\nu)\} \tag{1.2}$$

and to identify a primal optimizer $\hat{\pi} \in \Pi(\mu,\nu)$. To formulate the dual problem, we define

$$\Psi(\mu,\nu) = \left\{ (\varphi,\psi) : \begin{array}{l} \varphi : X \to [-\infty,\infty), \psi : Y \to [-\infty,\infty) \text{ integrable,} \\ \varphi(x) + \psi(y) \le c(x,y) \text{ for all } (x,y) \in X \times Y \end{array} \right\}. \tag{1.3}$$

The *dual Monge–Kantorovich problem* then consists in determining

$$D := D_c := \sup\Big\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \Big\} \tag{1.4}$$

for $(\varphi,\psi) \in \Psi(\mu,\nu)$. We say that *Monge–Kantorovich duality holds true*, or that *there is no duality gap*, if the primal value P of the problem equals the dual value D, i.e.

$$\inf\{\langle c, \pi\rangle : \pi \in \Pi(\mu,\nu)\} = \sup\Big\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu : (\varphi,\psi) \in \Psi(\mu,\nu)\Big\}. \tag{1.5}$$

There is a long line of research on these questions, initiated already by Kantorovich [Ka] himself and continued by numerous others (we mention [KR, Da, Db, dA, GR, Fe, S, M, MT], see also the bibliographical notes in [Vb, pp. 86, 87]).

The duality (1.5) was established in pleasant generality by H. Kellerer [Kel]. He proved that there is no duality gap provided that c is lower semicontinuous (see [Kel, Theorem 2.2]) or just Borel measurable and bounded by a constant (1) ([Kel, Theorem 2.14]). In [RRa, RRb] the problem is investigated beyond the realm of Polish spaces and a characterization is given of spaces for which duality holds for all bounded measurable cost functions. We also refer to the seminal paper [GM] by W. Gangbo and R. McCann.

We now present a rather trivial example (2) which shows that, in general, there is a duality gap.

(1) Or, more generally, by the sum $f(x) + g(y)$ of two integrable functions f, g.

(2) This is essentially [Kel, Example 2.5].

EXAMPLE 1.1. Consider $X = Y = [0,1]$ and $\mu = \nu$ the Lebesgue measure. Define c on $X \times Y$ to be 0 below the diagonal, 1 on the diagonal and ∞ elsewhere, i.e.

$$c(x,y) = \begin{cases} 0 & \text{for } 0 \le y < x \le 1, \\ 1 & \text{for } 0 \le x = y \le 1, \\ \infty & \text{for } 0 \le x < y \le 1. \end{cases}$$

Then the only finite transport plan is concentrated on the diagonal and leads to costs of one so that $P = 1$. On the other hand, for admissible $(\varphi, \psi) \in \Psi(\mu,\nu)$, it is straightforward to check that $\varphi(x) + \psi(x) > 0$ can hold true for at most countably many $x \in [0,1]$. Hence the dual value equals $D = 0$, so that there is a duality gap.

A common technique in the duality theory of convex optimization is to pass to a *relaxed* version of the problem, i.e., to enlarge the sets over which the primal and/or dual functionals are optimized ([3]). We do so, for the primal problem (1.2), by requiring only the transport of a portion of mass $1-\varepsilon$ from μ to ν, for every $\varepsilon > 0$. Fix $0 \le \varepsilon \le 1$ and define

$$\Pi^\varepsilon(\mu,\nu) = \{\pi \in \mathcal{M}_+(X \times Y) : \|\pi\| \ge 1-\varepsilon,\ p_X(\pi) \le \mu,\ p_Y(\pi) \le \nu\}.$$

Here $\mathcal{M}_+(X \times Y)$ denotes the non-negative Borel measures π on $X \times Y$ with norm $\|\pi\| = \pi(X \times Y)$; by $p_X(\pi) \le \mu$ (resp. $p_Y(\pi) \le \nu$) we mean that the projection of π onto X (resp. onto Y) is dominated by μ (resp. ν). We denote by P^ε the value of the $1-\varepsilon$ partial transportation problem

$$(1.6) \qquad P^\varepsilon := \inf\Big\{\langle c,\pi\rangle = \int\limits_{X\times Y} c(x,y)\,d\pi(x,y) : \pi \in \Pi^\varepsilon(\mu,\nu)\Big\}.$$

This partial transport problem has recently been studied by L. Caffarelli and R. McCann [CM] as well as A. Figalli [Fi]. Their work places the main emphasis on a finer analysis of the Monge problem for the squared Euclidean distance on $\mathbb{R}^n$, and pertains to a fixed $\varepsilon > 0$. In the present paper, we do not deal with these more subtle issues of the Monge problem and always remain in the realm of the Kantorovich problem (1.2). Our emphasis is on the limiting behavior for $\varepsilon \to 0$: we call

$$(1.7) \qquad P_c^{\mathrm{rel}} := P^{\mathrm{rel}} := \lim_{\varepsilon\to 0} P^\varepsilon$$

the *relaxed primal value* of the transport plan. Obviously this limit exists (assuming possibly the value $+\infty$) and $P^{\mathrm{rel}} \le P$.

As a motivation for the subsequent theorem the reader may observe that, in Example 1.1 above, we have $P^{\mathrm{rel}} = 0$ (while $P = 1$). Indeed, it is possible

([3]) We remark that another way of dealing with the situation where there is a duality gap is to consider a "rectification" of the cost function [BP].

to transport the measure $\mu\mathbb{1}_{[\varepsilon,1]}$ to the measure $\nu\mathbb{1}_{[0,1-\varepsilon]}$ with transport cost zero by the partial transport plan $\pi = (\mathrm{id}, \mathrm{id} - \varepsilon)_{\#}\,(\mu\mathbb{1}_{[\varepsilon,1]})$.

We can now formulate our main result.

THEOREM 1.2. *Let X, Y be Polish spaces, equipped with Borel probability measures μ, ν, and let $c : X \times Y \to [0,\infty]$ be Borel measurable. Then there is no duality gap if the primal problem is defined in the relaxed form* (1.7) *while the dual problem is formulated in its usual form* (1.4). *In other words,*

$$P^{\mathrm{rel}} = D. \tag{1.8}$$

We observe that in (1.8) also the value $+\infty$ is possible.

The theorem gives a positive result on the issue of duality in the Monge–Kantorovich problem. Moreover we have $P = P^{\mathrm{rel}}$ and therefore $P = D$ in any of the following cases:

(a) c is lower semicontinuous,
(b) c is uniformly bounded or, more generally,
(c) c is $\mu \otimes \nu$-a.s. finitely valued.

Concerning (a) and (b), it is rather straightforward to check that these assumptions imply $P = P^{\mathrm{rel}}$ (see Corollaries 3.1 and 3.3 below). In particular, the classical duality results of Kellerer quickly follow from Theorem 1.2. To establish that also property (c) is sufficient seems to be more sophisticated and follows from [BS, Theorem 1].

A sufficient condition for attainment in the primal part of the Monge–Kantorovich transport problem is that the cost function c is lower semicontinuous and we have nothing to add here.

To analyze the same question concerning the dual problem we need some preparation. Consider the following alternative definition of P^{rel}. One may relax the transport costs by cutting the maximal transport costs. That is, we could alter the cost function c to $c \wedge M$ for some $M \geq 0$ or to $c \wedge h$ for some $\mu \otimes \nu$-a.s. finite, measurable function $h : X \times Y \to [0,\infty]$. If M resp. h is large this should have a similar effect as ignoring a small mass. Indeed we will establish that

$$\lim_{n\to\infty} P_{c\wedge h_n} = P_c^{\mathrm{rel}} \tag{1.9}$$

for any sequence of measurable functions $h_n : X \times Y \to [0,\infty)$ increasing (uniformly) to ∞.

In Theorem 3.5 below we then prove that we have dual attainment (in the sense of [BS, Section 1.1]) if and only if there exists some finite measurable function $h : X \times Y \to [0,\infty)$ such that

$$P_{c\wedge h} = P_c^{\mathrm{rel}}. \tag{1.10}$$

The paper is organized as follows.

In Section 2 we show Theorem 1.2. The proof is self-contained with the exception of Lemma A.1 which is a consequence of [Kel, Lemma 1.8]. For the convenience of the reader we provide a derivation of Lemma A.1 in the Appendix.

Section 3 deals with consequences of Theorem 1.2. First we rederive the classical duality results of Kellerer. Then we establish (with the help of [BS, Theorem 2]) the alternative characterization of P^{rel} given in (1.9) and the characterization of dual attainment via (1.10).

2. The proof of the duality theorem. The proof of Theorem 1.2 relies on Fenchel's perturbation technique. We refer to the companion paper [BLS] for a didactic presentation of this technique: there we give an elementary version of this argument, where $X = Y = \{1, \ldots, N\}$ with the uniform measure $\mu = \nu$, in which case the optimal transport problem reduces to a finite linear programming problem.

We start with an easy result showing that the relaxed version (1.6) of the optimal transport problem is not "too relaxed", in the sense that the trivial implication of the minmax theorem still holds true.

PROPOSITION 2.1. *Under the assumptions of Theorem* 1.2 *we have*

$$P^{\mathrm{rel}} \geq D.$$

Proof. Let (φ, ψ) be integrable Borel functions such that

$$\varphi(x) + \psi(y) \leq c(x, y) \quad \text{for every } (x, y) \in X \times Y.$$

Let $\pi_n \in \Pi(f_n\mu, g_n\nu)$ be an optimizing sequence for the relaxed problem, where $f_n, g_n \leq \mathbb{1}$, and $\pi_n(X \times Y) = \|f_n\|_{L^1(\mu)} = \|g_n\|_{L^1(\nu)}$ tends to 1. By passing to a subsequence we may assume that $(f_n)_{n=1}^{\infty}$ and $(g_n)_{n=1}^{\infty}$ converge a.s. to $\mathbb{1}$. We may estimate

$$\liminf_{n\to\infty} \int_{X\times Y} c\, d\pi_n \geq \liminf_{n\to\infty} \Big[\int_X \varphi f_n\, d\mu + \int_Y \psi g_n\, d\nu\Big] = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu,$$

where in the last equality we have used Lebesgue's dominated convergence theorem. ∎

The next lemma is a technical result which will be needed in the formalization of the proof of Theorem 1.2.

LEMMA 2.2. *Let V be a normed vector space, $x_0 \in V$, and let $\Phi : V \to (-\infty, \infty]$ be a positively homogeneous* (4) *convex function such that*

$$\liminf_{\|x-x_0\|\to 0} \Phi(x) \geq \Phi(x_0).$$

(4) By positively homogeneous we mean $\Phi(\lambda x) = \lambda\Phi(x)$ for $\lambda \geq 0$, with the convention $0 \cdot \infty = 0$.

If $\Phi(x_0) < \infty$ then, for each $\varepsilon > 0$, there exists a continuous linear functional $v : V \to \mathbb{R}$ such that

$$\Phi(x_0) - \varepsilon \le v(x_0) \quad \textit{and} \quad \Phi(x) \ge v(x) \ \ \textit{for all } x \in V.$$

If $\Phi(x_0) = \infty$ then, for each $M > 0$, there exists a continuous linear functional $v : V \to \mathbb{R}$ such that

$$M \le v(x_0) \quad \textit{and} \quad \Phi(x) \ge v(x) \quad \textit{for all } x \in V.$$

Proof. Assume first that $\Phi(x_0) < \infty$. Let $K = \{(x,t) : x \in V,\, t \ge \Phi(x)\}$ be the epigraph of Φ and $\overline{K}$ its closure in $V \times \mathbb{R}$. Since Φ is assumed to be lower semicontinuous at x_0, we have $\inf\{t : (x_0,t) \in \overline{K}\} = \Phi(x_0)$, hence $(x_0, \Phi(x_0)-\varepsilon) \notin \overline{K}$. By Hahn–Banach, there is a continuous linear functional $w \in V^* \times \mathbb{R}$ given by $w(x,t) = u(x)+st$ (where $u \in V^*$ and $s \in \mathbb{R}$) and $\beta \in \mathbb{R}$ such that $w(x,t) > \beta$ for $(x,t) \in \overline{K}$ and $w(x_0, \Phi(x_0)-\varepsilon) < \beta$. By the positive homogeneity of Φ, we have $\beta < 0$, hence $s > 0$. Also $u(x) + s\Phi(x) \ge \beta$ and by applying positive homogeneity once more we see that β can be replaced by 0. Hence

$$u(x) + s\Phi(x) \ge 0, \quad u(x_0) + s(\Phi(x_0) - \varepsilon) < 0,$$

so we just let $v(x) := -u(x)/s$. In the case $\Phi(x_0) = \infty$ the assertion is proved analogously. ■

We now define the function Φ to which we shall apply the previous lemma.

Let $W = L^1(\mu) \times L^1(\nu)$ and V the subspace of codimension one, formed by the pairs (f,g) such that $\int_X f\,d\mu = \int_Y g\,d\nu$. Let $V_+ = \{(f,g) \in V : f \ge 0,\, g \ge 0\}$, the positive orthant of V. For $(f,g) \in V_+$, we define, by a slight abuse of notation, $\Pi(f,g)$ as the set of non-negative Borel measures π on $X \times Y$ with marginals $f\mu$ and $g\nu$ respectively. With this notation $\Pi(\mathbb{1},\mathbb{1})$ is just the set $\Pi(\mu,\nu)$ introduced above. Define $\Phi : V_+ \to [0,\infty]$ by

$$\Phi(f,g) = \inf\Big\{ \int_{X\times Y} c(x,y)\,d\pi(x,y) : \pi \in \Pi(f,g)\Big\}, \quad (f,g) \in V_+,$$

which is a convex function. By definition we have $\Phi(\mathbb{1},\mathbb{1}) = P$, where P is the primal value of (1.2). Our concern will be the lower semicontinuity of the function Φ at the point $(\mathbb{1},\mathbb{1}) \in V_+$.

PROPOSITION 2.3. *Denote by $\overline{\Phi} : V \to [0,\infty]$ the lower semicontinuous envelope of Φ, i.e., the largest lower semicontinuous function on V dominated by Φ on V_+. Then*

$$\overline{\Phi}(\mathbb{1},\mathbb{1}) = P^{\mathrm{rel}}.$$

Hence the function Φ is lower semicontinuous at $(\mathbb{1},\mathbb{1})$ if and only if $P = P^{\mathrm{rel}}$.

Proof. Let $(\pi_n)_{n=1}^{\infty}$ be an optimizing sequence for the relaxed problem (1.7), i.e., a sequence of non-negative measures on $X \times Y$ such that

$$\lim_{n\to\infty} \int_{X\times Y} c(x,y)\,d\pi_n(x,y) = P^{\mathrm{rel}},$$

$$\lim_{n\to\infty} \|\pi_n\| = \lim_{n\to\infty} \int_{X\times Y} 1\,d\pi_n(x,y) = 1,$$

and such that $p_X(\pi_n) \le \mu$ and $p_Y(\pi_n) \le \nu$. In particular $p_X(\pi_n) = f_n\mu$ and $p_Y(\pi_n) = g_n\mu$ with $(f_n)_{n=1}^{\infty}$ (resp. $(g_n)_{n=1}^{\infty}$) converging to $\mathbb{1}$ in $L^1(\mu)$ (resp. $L^1(\nu)$). It follows that

$$\overline{\Phi}(\mathbb{1},\mathbb{1}) \le \lim_{n\to\infty} \Phi(f_n, g_n) = P^{\mathrm{rel}}.$$

To prove the reverse inequality $\overline{\Phi}(\mathbb{1},\mathbb{1}) \ge P^{\mathrm{rel}}$, fix $\delta > 0$. We have to show that for each $\varepsilon > 0$ there is some $\tilde{\pi} \in \Pi^{\varepsilon}(\mu,\nu)$ such that

$$\overline{\Phi}(\mathbb{1},\mathbb{1}) + \delta \ge \int c\,d\tilde{\pi}. \tag{2.1}$$

Pick $\gamma \in (0,1)$ such that $(1-\gamma)^3 \ge 1-\varepsilon$. Select f, g and $\pi \in \Pi(f,g)$ with $\|f-\mathbb{1}\|_{L^1(\mu)}, \|g-\mathbb{1}\|_{L^1(\nu)} < \gamma$ and $\overline{\Phi}(\mathbb{1},\mathbb{1}) + \delta \ge \int c\,d\pi$. We note for later use that $\|\pi\| = \|f\|_{L^1(\mu)} = \|g\|_{L^1(\nu)} \in (1-\gamma, 1+\gamma)$. Define a Borel measure $\tilde{\pi} \ll \pi$ on $X \times Y$ by

$$\frac{d\tilde{\pi}}{d\pi}(x,y) := \frac{1}{(1+|f(x)-1|)(1+|g(y)-1|)},$$

and set $\tilde{\mu} := p_X(\tilde{\pi}), \tilde{\nu} := p_Y(\tilde{\pi})$. As $d\tilde{\pi}/d\pi \le 1$, we have $\tilde{\pi} \le \pi$ so that (2.1) is satisfied. Also $\tilde{\mu} \le \mu$ and $\tilde{\nu} \le \nu$. Thus it remains to check that $\|\tilde{\pi}\| \ge 1-\varepsilon$.

The function $F(a,b) = 1/((1+a)(1+b))$ is convex on $[0,\infty)^2$ and by Jensen's inequality we have

$$\begin{aligned}
\|\tilde{\pi}\| &= \|\pi\| \int F(|f(x)-1|, |g(y)-1|)\,\frac{d\pi(x,y)}{\|\pi\|} \\
&\ge \|\pi\| F\biggl(\frac{\|f-\mathbb{1}\|_{L^1(\mu)}}{\|\pi\|}, \frac{\|g-\mathbb{1}\|_{L^1(\nu)}}{\|\pi\|}\biggr) \\
&\ge (1-\gamma)\frac{1}{(1+\gamma/(1-\gamma))^2} \ge 1-\varepsilon,
\end{aligned}$$

as required.

The final assertion of the proposition is now obvious. ■

Proof of Theorem 1.2. By the preceding proposition we have to show that

$$\overline{\Phi}(\mathbb{1},\mathbb{1}) = D,$$

where the dual value D of the optimal transport problem is defined in (1.4).

By Lemma 2.2 we know that there are sequences (5) $(\varphi_n, \psi_n)_{n=1}^\infty \subset W^* = L^\infty(\mu) \times L^\infty(\nu)$ such that

$$\lim_{n\to\infty} \langle (\varphi_n, \psi_n), (\mathbb{1}, \mathbb{1}) \rangle = \lim_{n\to\infty} \Big[\int_X \varphi_n \, d\mu + \int_Y \psi_n \, d\nu \Big] = \overline{\Phi}(\mathbb{1}, \mathbb{1}) \in [0, \infty]$$

and

$$\langle (\varphi_n, \psi_n), (f, g) \rangle = \langle \varphi_n, f \rangle + \langle \psi_n, g \rangle \le \Phi(f, g) \quad \text{for all } (f, g) \in V. \tag{2.2}$$

We shall show that (2.2) implies that, for each fixed $n \in \mathbb{N}$, there are representatives (6) $(\tilde{\varphi}_n, \tilde{\psi}_n)$ of (φ_n, ψ_n) such that

$$\tilde{\varphi}_n(x) + \tilde{\psi}_n(y) \le c(x, y)$$

for *all* $(x, y) \in X \times Y$. Indeed, choose any $\mathbb{R}$-valued representatives $(\check{\varphi}_n, \check{\psi}_n)$ of (φ_n, ψ_n) and consider the set

$$C = \{(x, y) \in X \times Y : \check{\varphi}_n(x) + \check{\psi}_n(y) > c(x, y)\}. \tag{2.3}$$

Claim. *For every $\pi \in \Pi(\mu, \nu)$ we have $\pi(C) = 0$.*

Indeed, fix $\pi \in \Pi(\mu, \nu)$ and denote by (f, g) the density functions of the projections $p_X(\pi_{|C})$ and $p_Y(\pi_{|C})$. By (2.2) we have, for $n \ge 1$,

$$\int_{X\times Y} c\mathbb{1}_C \, d\pi \ge \Phi(f, g) \ge \langle \varphi_n, f \rangle + \langle \psi_n, g \rangle = \int_{X\times Y} (\check{\varphi}_n(x) + \check{\psi}_n(y))\mathbb{1}_C \, d\pi(x, y).$$

By the definition of C the first term above can only be greater than or equal to the last term if $\pi(C) = 0$, which readily shows the above claim.

Now we are in a position to apply an innocent looking, but deep result due to H. Kellerer [Kel, Lemma 1.8] (7): a Borel set $C = X \times Y$ satisfies $\pi(C) = 0$, for each $\pi \in \Pi(\mu, \nu)$, if and only if there are Borel sets $M \subseteq X$ and $N \subseteq Y$ with $\mu(M) = \nu(N) = 0$ such that $C \subseteq (M \times Y) \cup (X \times N)$. Choosing such sets M and N for the set C in (2.3), define the representatives $(\tilde{\varphi}_n, \tilde{\psi}_n)$ by $\tilde{\varphi}_n = \check{\varphi}_n \mathbb{1}_{X\setminus M} - \infty \mathbb{1}_M$ and $\tilde{\psi}_n = \check{\psi}_n \mathbb{1}_{Y\setminus N} - \infty \mathbb{1}_N$. We then have $\tilde{\varphi}_n(x) + \tilde{\psi}_n(y) \le c(x, y)$ for *every* $(x, y) \in X \times Y$. As

$$\lim_{n\to\infty} \int_X \tilde{\varphi}_n \, d\mu + \int_Y \tilde{\psi}_n \, d\nu = \overline{\Phi}(\mathbb{1}, \mathbb{1}) = P^{\mathrm{rel}},$$

the proof of Theorem 1.2 is complete. ∎

(5) The dual space V^* of the subspace V of $W = L^1(\mu) \times L^1(\nu)$ equals the quotient of the dual $L^\infty(\mu) \times L^\infty(\nu)$, modulo the annihilator of V, i.e. the one-dimensional subspace formed by the $(\varphi, \psi) \in L^\infty(\mu) \times L^\infty(\nu)$ of the form $(\varphi, \psi) = (a, -a)$, for $a \in \mathbb{R}$.

(6) Strictly speaking, (φ_n, ψ_n) are elements of $L^\infty(\mu) \times L^\infty(\nu)$, i.e. *equivalence classes* of functions. The $[-\infty, \infty[$-valued Borel measurable functions $(\tilde{\varphi}_n, \tilde{\psi}_n)$ will be properly chosen representatives of these equivalence classes.

(7) For the convenience of the reader and in order to keep the present paper self-contained, we provide in the appendix (Lemma A.1) a proof of Kellerer's lemma, which does not rely on duality arguments.

3. Consequences of the duality theorem. Assume first that the Borel measurable cost function $c : X \times Y \to [0,\infty]$ is $\mu \otimes \nu$-almost surely bounded by some constant (8) M. We may then estimate

$$P \le P^\varepsilon + \varepsilon M.$$

Indeed, for $\varepsilon > 0$, every partial transport plan π^ε with marginals $\mu^\varepsilon \le \mu$ and $\nu^\varepsilon \le \nu$ and mass $\|\pi^\varepsilon\| = 1 - \varepsilon$ may be completed to a full transport plan π by letting, e.g.,

$$\pi = \pi^\varepsilon + \varepsilon^{-1}(\mu - \mu^\varepsilon) \otimes (\nu - \nu^\varepsilon).$$

As $c \le M$ we have $\int c\,d\pi \le \int c\,d\pi^\varepsilon + \varepsilon M$. This yields the following corollary due to H. Kellerer [Kel, Theorem 2.2].

Corollary 3.1. *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν, and let $c : X \times Y \to [0,\infty]$ be a Borel measurable cost function which is uniformly bounded. Then there is no duality gap, i.e. $P = D$.*

To establish duality for a lower semicontinuous cost function c, it suffices to note that in this setting also the cost functional Φ is lower semicontinuous:

Lemma 3.2 ([Vb, Lemma 4.3]). *Let $c : X \times Y \to [0,\infty]$ be lower semicontinuous and assume that a sequence of measures π_n on $X \times Y$ converges to a transport plan $\pi \in \Pi(\mu,\nu)$ weakly, i.e. in the topology induced by the bounded continuous functions on $X \times Y$. Then*

$$\int c\,d\pi \le \liminf_{n\to\infty} \int c\,d\pi_n.$$

Corollary 3.3 ([Kel, Theorem 2.6]). *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν, and let $c : X \times Y \to [0,\infty]$ be a lower semicontinuous cost function. Then there is no duality gap, i.e. $P = D$.*

Proof. It follows from Prokhorov's theorem and Lemma 3.2 that the function $\Phi : V_+ \to [0,\infty]$ is lower semicontinuous with respect to the norm topology of V. ∎

We turn now to the question under which assumptions there is dual attainment.

Easy examples show that one cannot expect that the dual problem admits *integrable* maximizers unless the cost function satisfies certain integrability conditions with respect to μ and ν [BS, Examples 4.4, 4.5]. In fact [BS, Example 4.5] takes place in a very "regular" setting, where c is squared Euclidean distance on $\mathbb{R}$. In this case there exist natural candidates (φ, ψ)

(8) In fact, the same argument works provided that $c(x,y) \le f(x) + g(y)$ for integrable functions f, g.

which, however, fail to be dual maximizers in the usual sense as they are not integrable.

The following solution was proposed in [BS, Section 1.1]. If φ and ψ are integrable functions and $\pi \in \Pi(\mu,\nu)$ then

$$\int_X \varphi\, d\mu + \int_Y \psi\, d\nu = \int_{X\times Y} (\varphi(x)+\psi(y))\, d\pi(x,y). \tag{3.1}$$

If we drop the integrability condition on φ and ψ, the left hand side need not make sense. But if we require that $\varphi(x)+\psi(y) \le c(x,y)$ and if π is a finite cost transport plan, i.e. $\int_{X\times Y} c\, d\pi < \infty$, then the right hand side of (3.1) still makes sense, assuming possibly the value $-\infty$, and we set

$$J_c(\varphi,\psi) = \int_{X\times Y} (\varphi(x)+\psi(y))\, d\pi(x,y).$$

It is not difficult to show (see [BS, Lemma 1.1]) that this value does not depend on the choice of the finite cost transport plan π and satisfies $J_c(\varphi,\psi) \le D$. Under the assumption that there exists some finite transport plan $\pi \in \Pi(\mu,\nu)$ we then say that we have *dual attainment* in the optimization problem (1.4) if there exist Borel measurable functions $\hat\varphi : X \to [-\infty,\infty)$ and $\hat\psi : Y \to [-\infty,\infty)$ satisfying $\varphi(x)+\psi(y) \le c(x,y)$, for $(x,y) \in X\times Y$, such that

$$D = J_c(\hat\varphi,\hat\psi).$$

We recall a result established in [BS], generalizing Corollary 3.1. We remark that we do not know how to directly deduce it from Theorem 1.2.

THEOREM 3.4 ([BS, Theorems 1 and 2]). *Let X,Y be Polish spaces equipped with Borel probability measures μ,ν, and let $c : X\times Y \to [0,\infty]$ be a Borel measurable cost function such that $\mu\otimes\nu(\{(x,y) : c(x,y)=\infty\}) = 0$. Then there is no duality gap, i.e. $P = D$. Moreover there exist Borel measurable functions $\varphi : X \to [-\infty,\infty)$ and $\psi : Y \to [-\infty,\infty)$ such that $\varphi(x)+\psi(y) \le c(x,y)$ for all $x \in X$ and $y \in Y$, and $J_c(\varphi,\psi) = D$.*

Using Theorem 3.4 we now obtain an alternative description of P^{rel} and the characterization of dual attainment mentioned in the introduction.

THEOREM 3.5. *Let X,Y be Polish spaces, equipped with Borel probability measures μ,ν, let $c : X\times Y \to [0,\infty]$ be Borel measurable and assume that there exists a finite transport plan. For every sequence of measurable functions $h_n : X\times Y \to [0,\infty]$ such that $h_n \uparrow \infty$ uniformly* [9] *and each h_n*

[9] By saying that h_n increases to ∞ *uniformly*, we mean that for n large enough, $h_n \ge m$ for every given constant $m \in [0,\infty)$. Indeed it is crucial to insist on this strong type of convergence: one can easily construct examples where $h_n(x,y) \uparrow \infty$ for all $(x,y) \in X\times Y$ while $P_{h_n} = 0$ for every $n \in \mathbb{N}$.

is $\mu \otimes \nu$-a.s. finitely valued, we have

$$P_{c\wedge h_n} \uparrow P^{\mathrm{rel}}. \tag{3.2}$$

Moreover, the following are equivalent:

(i) *There is dual attainment, i.e. there exist measurable functions φ, ψ such that $\varphi(x) + \psi(y) \le c(x, y)$ for $x \in X$ and $y \in Y$, and $P^{\mathrm{rel}} = D = J_c(\varphi, \psi)$.*

(ii) *There exists a $\mu \otimes \nu$-a.s. finite function $h : X \times Y \to [0, \infty]$ such that $P^{\mathrm{rel}} = P_{c\wedge h}$.*

Proof. Fix $(h_n)_{n\ge 0}$ as in the statement. To prove (3.2), note that by Theorem 3.4 there exist, for each n, measurable functions $\varphi_n : X \to [-\infty, \infty)$ and $\psi_n : Y \to [-\infty, \infty)$ satisfying $\varphi_n(x) + \psi_n(y) \le c(x, y)$ for all $x \in X$ and $y \in Y$ such that

$$J_c(\varphi_n, \psi_n) = P_{c\wedge h_n}.$$

Thus $P_{c\wedge h_n} \le P^{\mathrm{rel}}$ for each n. To see that $\lim_{n\to\infty} P_{c\wedge h_n} \ge P^{\mathrm{rel}}$, fix $\eta > 0$. As $D = P^{\mathrm{rel}}$ there exists $(\varphi, \psi) \in \Psi(\mu, \nu)$ such that $J(\varphi, \psi) > P^{\mathrm{rel}} - \eta$. Note that, for $M \ge 0$, the pair of functions $(M \wedge (-M \vee \varphi)), M \wedge (-M \vee \psi))$ lies in $\Psi(\mu, \nu)$. Hence we may assume without loss of generality that $|\varphi|$ and $|\psi|$ are uniformly bounded by some constant M. Pick n so that $h_n(x, y) \ge 2M$ for all $x \in X$ and $y \in Y$. It then follows that $c \wedge h_n(x, y) \ge \varphi(x) + \psi(y)$ for all $x \in X$ and $y \in Y$, hence

$$P_{c\wedge h_n} \ge J(\varphi, \psi) > P^{\mathrm{rel}} - \eta,$$

which shows (3.2).

To prove that (ii) implies (i), apply Theorem 3.4 to the cost function $c \wedge h$ to obtain functions φ and ψ satisfying $\varphi(x) + \psi(y) \le (c \wedge h)(x, y)$ and $J_{c\wedge h}(\varphi, \psi) = P_{c\wedge h}$. Then $J_c(\varphi, \psi) = P_{c\wedge h} = P^{\mathrm{rel}} = D$, hence (φ, ψ) is a pair of dual maximizers.

To see that (i) implies (ii), pick dual maximizers φ, ψ and set $h(x, y) := (\varphi(x) + \psi(y))_+$. ■

We close this section with a comment concerning a possible relaxed version of the dual problem.

Remark 3.6. Define

$$D^{\mathrm{rel}} := \sup \left\{ \int \varphi \, d\mu + \int \psi \, d\nu : \begin{array}{l} \varphi, \psi \text{ integrable,} \\ \varphi(x) + \psi(y) \le c(x, y) \ \pi\text{-a.e.} \\ \text{for every finite cost } \pi \in \Pi(\mu, \nu) \end{array} \right\} \ge D$$

where $\pi \in \Pi(\mu, \nu)$ *has finite cost* if $\int_{X\times Y} c \, d\pi < \infty$. It is straightforward to verify that we still have $D^{\mathrm{rel}} \le P$. One might conjecture (and the present authors did so for some time) that, similarly to the situation in Theorem 1.2, duality in the form $D^{\mathrm{rel}} = P$ holds without any additional assumption. For instance this is the case in Example 1.1 and combining the methods of [BGMS] and [BS] one can prove that $D^{\mathrm{rel}} = P$ provided that the Borel measurable cost function $c : X \times Y \to [0, \infty]$ is such that the set $\{c = \infty\}$ is *closed* in the product topology of $X \times Y$. However a rather complicated example constructed in [BLS, Section 4] shows that under the assumptions of Theorem 1.2 it may happen that D^{rel} is strictly smaller than P, i.e. that there is still a duality gap.

Appendix. In our proof of Theorem 1.2 we made use of the following innocent looking result due H. Kellerer:

LEMMA A.1. *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν, let $L \subseteq X \times Y$ be a Borel set and assume that $\pi(L) = 0$ for any $\pi \in \Pi(\mu, \nu)$. Then there exist sets $M \subseteq X$ and $N \subseteq Y$ such that $\mu(M) = \nu(N) = 0$ and $L \subseteq M \times Y \cup X \times N$.*

Lemma A.1 seems quite intuitive and, as we shall presently see, its proof is quite natural provided that the set L is *compact*. However the general case is delicate and relies on relatively involved results from measure theory. H. Kellerer proceeded as follows. First he established various sophisticated duality results. Lemma A.1 is then a consequence of the fact that there is no duality gap in the case when the Borel measurable cost function c is uniformly bounded (Corollary 3.1). To make the present paper more self-contained, we provide a direct proof of Lemma A.1 which does not rely on duality results. Still, most ideas of the subsequent proof are, at least implicitly, contained in the work of H. Kellerer.

Some steps in the proof of Lemma A.1 are (notationally) simpler in the case when $(X, \mu) = (Y, \nu) = ([0, 1], \lambda)$, therefore we give a short argument which shows that it is legitimate to make this additional assumption.

Indeed it is rather obvious that one may reduce to the case that the measure spaces X and Y are free of atoms. A well known result of measure theory (see for instance [Kec, Theorem 17.41]) asserts that for any Polish space Z equipped with a continuous Borel probability measure σ, there exists a measure preserving Borel isomorphism between the spaces (Z, σ) and $([0, 1], \lambda)$. Thus there exist bijections $f : X \to [0, 1]$ and $g : Y \to [0, 1]$, measurable with measurable inverse, such that $f_{\#}\mu = g_{\#}\nu = \lambda$. Hence it is sufficient to consider the case $(X, \mu) = (Y, \nu) = ([0, 1], \lambda)$ and we will do so from now on.

For a measurable set $L \subseteq [0, 1]^2$ we define the functional

$$m(L) := \inf\{\lambda(A) + \lambda(B) : L \subseteq A \times Y \cup X \times B\}.$$

Our strategy is to show that under the assumptions of Lemma A.1, we have $m(L) = 0$. This implies Lemma A.1 in view of the following result.

Lemma A.2. *Let $L \subseteq X \times Y$ be a Borel set with $m(L) = 0$. Then there exist sets $M \subseteq X$ and $N \subseteq Y$ such that $\mu(M) = \nu(N) = 0$ and $L \subseteq M \times Y \cup X \times N$.*

Proof. Fix $\varepsilon > 0$. Since $m(L) = 0$, there exist sets A_n, B_n such that $\mu(A_n) < 1/n$ and $\nu(B_n) < \varepsilon 2^{-n}$ and $L \subseteq A_n \times Y \cup X \times B_n$. Set $A := \bigcap_{n\geq 1} A_n$ and $B := \bigcup_{n\geq 1} B_n$. Then $\mu(A) = 0$, $\nu(B) < \varepsilon$ and

$$L \subseteq \bigcap_{n\geq 1} (A_n \times Y \cup X \times B) = A \times Y \cup X \times B.$$

Iterating this argument with the roles of X and Y exchanged we get the desired conclusion. ∎

The next step proves Lemma A.1 in the case where L is compact.

Lemma A.3. *Assume that $K \subseteq [0,1]^2$ is compact and satisfies $\pi(K) = 0$ for every $\pi \in \Pi(\lambda, \lambda)$. Then $m(K) = 0$.*

Proof. Assume that $\alpha := m(K) > 0$. We have to show that there exists a non-trivial measure π on $X \times Y$, i.e. $\pi(K) > 0$ such that $\operatorname{supp} \pi \subseteq K$ and the marginals of π satisfy $P_X(\pi) \leq \mu$ and $P_Y(\pi) \leq \nu$. We aim to construct increasingly good approximations π_n of such a measure.

Fix n large enough and choose $k \geq 1$ such that $\alpha/3 \leq k/n \leq \alpha/2$. Since K is non-empty, there exist $i_1, j_1 \in \{0, \ldots, n-1\}$ such that

$$((i_1/n, j_1/n) + [0, 1/n]^2) \cap K \neq \emptyset.$$

After $m < k$ steps, assume that we have already chosen $(i_1, j_1), \ldots, (i_m, j_m)$. Since $2m/n < \alpha$, we see that K is not covered by

$$\Big(\bigcup_{l=1}^{m} [i_l/n, (i_l+1)/n]\Big) \times Y \cup X \times \Big(\bigcup_{l=1}^{m} [j_l/n, (j_l+1)/n]\Big).$$

Thus there exist

$$\begin{aligned} i_{m+1} &\in \{0, \ldots, n-1\} \setminus \{i_1, \ldots, i_m\}, \\ j_{m+1} &\in \{0, \ldots, n-1\} \setminus \{j_1, \ldots, j_m\} \end{aligned}$$

such that $((i_{m+1}/n, j_{m+1}/n) + [0, 1/n]^2) \cap K \neq \emptyset$. After k steps we define the measure π_n to be the restriction of $n \cdot \lambda^2$ (i.e. the Lebesgue measure on $[0,1]^2$ multiplied with the constant n) to the set $\bigcup_{l=1}^{k} (i_l/n, j_l/n) + [0, 1/n]^2$. Then the total mass of π_n is bounded from below by $k/n \geq \alpha/3$ and the marginals of π_n satisfy $P_X(\pi_n) \leq \mu$ and $P_Y(\pi_n) \leq \nu$. These properties carry over to every weak-star limit point of the sequence (π_n) and each such limit point π satisfies $\operatorname{supp} \pi \subseteq K$ since K is closed. ∎

The next lemma will enable us to reduce the case of a Borel set L to the case of L compact.

Lemma A.4. *Suppose that a Borel set $L \subseteq [0,1]^2$ satisfies $m(L) > 0$. Then there exists a compact set $K \subseteq L$ such that $m(K) > 0$.*

Lemma A.4 will be deduced from *Choquet's capacitability theorem* ([10]). Before formulating this result we introduce some notation. Given a compact metric space Z, a *capacity* on Z is a map $\gamma : \mathcal{P}(Z) \to \mathbb{R}_+$ such that:

(1) $A \subseteq B \Rightarrow \gamma(A) \le \gamma(B)$.
(2) $A_1 \subseteq A_2 \subseteq \cdots \Rightarrow \sup_{n\ge 1} \gamma(A_n) = \gamma(\bigcup_{n\ge 1} A_n)$.
(3) For every sequence $K_1 \supseteq K_2 \supseteq \cdots$ of compact sets, $\inf_{n\ge 1} \gamma(K_n) = \gamma(\bigcap_{n\ge 1} K_n)$.

A typical example of a capacity is the *outer measure* associated to a finite Borel measure.

Theorem A.5 (Choquet capacitability theorem; see [C] and [Kec, Theorem 30.13]). *Assume that γ is a capacity on a Polish space Z. Then*

$$\gamma(A) = \sup\{\gamma(K) : K \subseteq A,\ K \ \textit{compact}\}$$

for every Borel ([11]) *set $A \subseteq Z$.*

Proof of Lemma A.4. We cannot apply Theorem A.5 directly to the functional m since m fails to be a capacity, even if it is extended in a proper way to all subsets of $[0,1]^2$. A clever trick ([12]) is to replace m by the mapping $\gamma : \mathcal{P}([0,1]^2) \to [0,2]$, defined by

$$\gamma(L) := \inf\Big\{\int f\, d\lambda : f : [0,1] \to [0,1],\\ f(x) + f(y) \ge \mathbb{1}_L(x,y) \text{ for } (x,y) \in [0,1]\Big\}.$$

We then have:

(a) $\gamma(L) \le m(L) \le 4\gamma(L)$ for any Borel set $L \subseteq [0,1]^2$.
(b) γ is a capacity.

To see that (a) holds, notice that $f(x) + f(y) \ge \mathbb{1}_L(x,y)$ implies that $L \subseteq \{f \ge 1/2\} \times Y \cup X \times \{f \ge 1/2\}$ and that $L \subseteq A \times Y \cup X \times B$ yields $\mathbb{1}_{A\cup B}(x) + \mathbb{1}_{A\cup B}(y) \ge \mathbb{1}_L(x,y)$.

To prove (b) it remains to check that γ has properties (2) and (3) of the definition of capacity. To see continuity from below, consider a sequence of sets $A_n \nearrow A$. Pick f_n such that $f_n(x) + f_n(y) \ge \mathbb{1}_{A_n}(x,y)$ pointwise and $\int f\, d\lambda < \gamma(A_n) + 1/n$ for each $n \ge 1$. By Komlos' Lemma there exist

---

([10]) It seems worth noting that Kellerer also employs the Choquet capacitability theorem.

([11]) In fact, the assertion of the Choquet capacitability theorem is true for the strictly larger class of *analytic sets*.

([12]) We thank Richárd Balka and Márton Elekes for showing us this argument (private communication).

$g_n \in \operatorname{conv}\{f_n, f_{n+1}, \ldots\}$ such that (g_n) converges λ-a.s. to a function $g : [0,1] \to [0,1]$. After changing g on a λ-null set if necessary, we have $g(x) + g(y) \geq \mathbb{1}_A(x,y)$ pointwise. By dominated convergence,

$$\int g\, d\lambda = \lim_{n\to\infty} \int g_n\, d\lambda \leq \lim_{n\to\infty} (\gamma(A_n) + 1/n) = \gamma(A).$$

Thus γ has property (2). The proof of (3) follows precisely the same scheme.

An application of Choquet's Theorem A.5 now finishes the proof of Lemma A.4. ■

We have made all the preparations to prove Lemma A.1 and now summarize the necessary steps.

Proof of Lemma A.1. As discussed above, we may assume that $(X,\mu) = (Y,\nu) = ([0,1],\lambda)$. Suppose that the Borel set $L \subseteq [0,1]^2$ satisfies $\pi(L) = 0$ for all $\pi \in \Pi(\mu,\nu)$. Striving for a contradiction, we assume that $m(L) > 0$. By Lemma A.4, we find that there exists a compact set $K \subseteq L$ such that $m(K) > 0$. By Lemma A.3, there is a measure $\pi \in \Pi(\mu,\nu)$ such that $\pi(K) > 0$, hence also $\pi(L) > 0$, contrary to assumption. Thus $m(L) = 0$. By Lemma A.2 we conclude that there exist sets $M \subseteq X$ and $N \subseteq Y$ with $\mu(X) = \nu(N) = 0$ such that $L \subseteq M \times Y \cup X \times N$, hence we are done. ■

Acknowledgements. The first author acknowledges financial support from the Austrian Science Fund (FWF) under grant P21209. The third author acknowledges support from the Austrian Science Fund (FWF) under grant P19456, from the Vienna Science and Technology Fund (WWTF) under grant MA13 and by the Christian Doppler Research Association (CDG). All authors thank A. Pratelli for helpful discussions on the topic of this paper. We also thank R. Balka, F. Delbaen, M. Elekes, M. Goldstern and G. Maresch for their advice.

References

[BGMS] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer, *Optimal and better transport plans*, J. Funct. Anal. 256 (2009), 1907–1927.

[BLS] M. Beiglböck, C. Léonard, and W. Schachermayer, *On the duality of the Monge–Kantorovich transport problem*, in: Proceedings of the Summer School in Grenoble, to appear, 2012.

[BP] M. Beiglböck and A. Pratelli, *Duality for rectified cost functions*, Calc. Var. Partial Differential Equations, to appear.

[BS] M. Beiglböck and W. Schachermayer, *Duality for Borel measurable cost functions*, Trans. Amer. Math. Soc. 363 (2011), 4203–4224.

[CM] L. Caffarelli and R. J. McCann, *Free boundaries in optimal transport and Monge–Ampère obstacle problems*, Ann. of Math. (2) 171 (2010), 673–730.

[C] G. Choquet, *Forme abstraite du théorème de capacitabilité*, Ann. Inst. Fourier (Grenoble) 9 (1959), 83–89.

[dA] A. de Acosta, *Invariance principles in probability for triangular arrays of B-valued random vectors and some applications*, Ann. Probab. 10 (1982), 346–373.

[Da] R. M. Dudley, *Probabilities and Metrics*, Lecture Notes Ser. 45, Matematisk Institut, Aarhus Univ., Aarhus, 1976.

[Db] R. M. Dudley, *Real Analysis and Probability*, Cambridge Stud. Adv. Math. 74, Cambridge Univ. Press, Cambridge, 2002.

[Fe] X. Fernique, *Sur le théorème de Kantorovitch–Rubinstein dans les espaces polonais*, in: Séminaire de Probabilités, XV (Strasbourg, 1979/1980), Lecture Notes in Math. 850, Springer, Berlin, 1981, 6–10.

[FÜa] D. Feyel and A. S. Üstünel, *Measure transport on Wiener space and the Girsanov theorem*, C. R. Math. Acad. Sci. Paris 334 (2002), 1025–1028.

[FÜb] D. Feyel and A. S. Üstünel, *Monge–Kantorovich measure transportation and Monge–Ampère equation on Wiener space*, Probab. Theory Related Fields 128 (2004), 347–385.

[FÜc] D. Feyel and A. S. Üstünel, *Monge–Kantorovich measure transportation, Monge–Ampère equation and the Itô calculus*, in: Stochastic Analysis and Related Topics in Kyoto, Adv. Stud. Pure Math. 41, Math. Soc. Japan, Tokyo, 2004, 49–74.

[FÜd] D. Feyel and A. S. Üstünel, *Solution of the Monge–Ampère equation on Wiener space for general log-concave measures*, J. Funct. Anal. 232 (2006), 29–55.

[Fi] A. Figalli, *The optimal partial transport problem*, Arch. Ration. Mech. Anal. 195 (2010), no. 2, 533–560.

[GR] N. Gaffke and L. Rüschendorf, *On a class of extremal problems in statistics*, Math. Operationsforsch. Statist. Ser. Optim. 12 (1981), 123–135.

[GM] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Math. 177 (1996), 113–161.

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

[KR] L. V. Kantorovich and G. Sh. Rubinshteĭn, *On a space of completely additive functions*, Vestnik Leningrad. Univ. 13 (1958), no. 7, 52–59 (in Russian).

[Kec] A. S. Kechris, *Classical Descriptive Set Theory*, Grad. Texts in Math. 156, Springer, New York, 1995.

[Kel] H. G. Kellerer, *Duality theorems for marginal problems*, Z. Wahrsch. Verw. Gebiete 67 (1984), 399–432.

[M] T. Mikami, *A simple proof of duality theorem for Monge–Kantorovich problem*, Kodai Math. J. 29 (2006), 1–4.

[MT] T. Mikami and M. Thieullen, *Duality theorem for the stochastic optimal control problem*, Stochastic Process. Appl. 116 (2006), 1815–1835.

[RRa] D. Ramachandran and L. Rüschendorf, *A general duality theorem for marginal problems*, Probab. Theory Related Fields 101 (1995), 311–319.

[RRb] D. Ramachandran and L. Rüschendorf, *Duality and perfect probability spaces*, Proc. Amer. Math. Soc. 124 (1996), 2223–2228.

[S] A. Szulga, *On minimal metrics in the space of random variables*, Teor. Veroyatnost. i Primenen. 27 (1982), 401–405 (in Russian).

[Va] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. in Math. 58, Amer. Math. Soc., Providence, RI, 2003.

[Vb] C. Villani, *Optimal Transport. Old and New*, Grundlehren Math. Wiss. 338, Springer, 2009.

Mathias Beiglböck, Walter Schachermayer
Faculty of Mathematics
University of Vienna
Nordbergstrasse 15
1090 Wien, Austria
E-mail: mathias.beiglboeck@univie.ac.at
walter.schachermayer@univie.ac.at

Christian Léonard
Modal-X, Université Paris Ouest
Bât. G, 200 av. de la République
92001 Nanterre, France
E-mail: christian.leonard@u-paris10.fr

Received October 21, 2011
Revised version April 25, 2012 (7334)