



## LINEAR TRANSFERS AS MINIMAL COSTS OF DILATIONS OF MEASURES IN BALAYAGE ORDER

NASSIF GHOUSSOUB

*Dedicated to Professor David Preiss on the occasion of his 75th birthday*

ABSTRACT. Linear transfers between probability distributions were introduced in [5, 6] in order to extend the theory of optimal mass transportation while preserving the important duality established by Kantorovich. It is shown here that  $\{0, +\infty\}$ -valued linear transfers can be characterized by balayage of measures with respect to suitable cones of functions à la Choquet, while general linear transfers extend balayage theory by requiring the “sweeping out” of measures to optimize certain cost functionals. We study the dual class of Kantorovich operators, which are natural and manageable extensions of Markov operators. It is also an important subclass of capacities, and could be called “convex functional Choquet capacities,” since they play for non-linear maps the same role that convex envelopes do for arbitrary numerical functions. A forthcoming paper [7] will study their ergodic properties and their applications.

### 1. INTRODUCTION

Given a closed convex cone  $\mathcal{A}$  of lower semi-continuous functions on a compact metric space  $\Omega$  that is stable under finite maxima and containing the non-negative constant functions, one defines a partial order on the set  $\mathcal{P}(\Omega)$  of probability measures via the relation

$$(1.1) \quad \mu \preceq_{\mathcal{A}} \nu \text{ if } \int_{\Omega} \phi d\mu \leq \int_{\Omega} \phi d\nu \text{ for all } \phi \in \mathcal{A}.$$

This is the so-called *balayage order with respect to  $\mathcal{A}$*  and cones with the above properties will be called *balayage cones*. They often appear in convex analysis, probability theory, analysis in several complex variables and potential theory. Here are a few examples.

- The cone of lower semi-continuous convex functions on a convex compact set  $\Omega$  in a locally convex topological vector space.
- The cone of subharmonic (resp., plurisubharmonic) functions on a domain  $\Omega$  in  $\mathbb{R}^n$  (resp.,  $\mathbb{C}^n$ ).
- The cone of excessive (resp., supermedian) functions associated to a homogenous Markov process.

2020 *Mathematics Subject Classification*. 31C45, 47N10, 49L20, 58E30.

*Key words and phrases*. Linear transfer, optimal mass transport, balayage of measures, functional capacities, Kantorovich operators.

- The cone of 1-Lipschitz functions on a compact metric space  $\Omega$ , as well as its tangent cone at a given  $f$ , i.e.,  $\mathcal{A}_f = \{\lambda(f-g); \lambda \geq 0, g \text{ is 1-Lipschitz on } \Omega\}$ .

We are interested in balayage as a transport procedure between probability distributions that are sometimes supported on different compact metric spaces  $X$  and  $Y$ . We shall then consider closed convex cones  $\mathcal{A}$  of lower semi-continuous functions on the disjoint union  $\Omega = X \sqcup Y$  and define a *restricted balayage* only on pairs of probabilities  $(\mu, \nu)$  in  $\mathcal{P}(X) \times \mathcal{P}(Y)$  -as opposed to  $\mathcal{P}(X \sqcup Y)$ - via:

$$(1.2) \quad \mu \preceq_{\mathcal{A}} \nu \text{ if and only if } \int_X \phi d\mu \leq \int_Y \phi d\nu \text{ for all } \phi \in \mathcal{A}.$$

Note the difference between (1.1) and (1.2). For example, the first pre-order is a transitive relation while the second is not, even if  $X = Y$ . Still, a celebrated theorem of Strassen [25] can be applied whenever (1.2) holds to deduce the following (See for example [10]).

**Proposition 1.1.** *For every pair  $(\mu, \nu)$  in  $\mathcal{P}(X) \times \mathcal{P}(Y)$  with  $\mu \preceq_{\mathcal{A}} \nu$ , where  $\mathcal{A}$  is a balayage cone on  $X \sqcup Y$ , there exists a Markov kernel (or a transport plan)  $\pi \in \mathcal{P}(X \times Y)$  whose marginals are  $(\mu, \nu)$  and such that the disintegration  $(\pi_x)_{x \in X}$  of  $\pi$  with respect to  $\mu$ , i.e.,*

$$(1.3) \quad \pi(X \times B) = \nu(B) \text{ and } \pi(A \times B) = \int_A \pi_x(B) d\mu(x) \text{ for all Borel } A \subset X, B \subset Y,$$

also satisfies

$$(1.4) \quad \delta_x \preceq_{\mathcal{A}} \pi_x \quad \mu - \text{a.e.}$$

These are sometimes called  $\mathcal{A}$ -dilations of  $\mu$  into  $\nu$ . The set of such  $\mathcal{A}$ -dilations will be denoted  $\mathcal{K}_{\mathcal{A}}(\mu, \nu)$ . The extreme points of the set  $\{(\mu, \nu); \mu \preceq_{\mathcal{A}} \nu\}$  are then pairs of the form  $(\delta_x, \eta)$  where  $\delta_x \preceq_{\mathcal{A}} \eta$ .

In this paper, we consider the problem of finding  $\mathcal{A}$ -dilations of  $\mu$  into  $\nu$  that minimize appropriate cost functionals of the form  $c : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ , that is

$$(1.5) \quad \mathcal{T}(\mu, \nu) = \inf \left\{ \int_X c(x, \pi_x) d\mu(x); \pi \in \mathcal{K}_{\mathcal{A}}(\mu, \nu) \right\}.$$

This can be seen as a problem of finding an optimal Strassen disintegration of transport plans between two probability measures in balayage order. This is known as the *optimal martingale transport* [16,21] whenever  $\mathcal{A}$  is the cone of convex functions, and  $c(x, \sigma) = \int_Y d(x, y) d\sigma(y)$  with  $d(x, y)$  being the cost of moving mass from  $x$  to  $y$ .

One of our objectives is to characterize those convex functions  $\mathcal{T}$  on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  that can be described as a value function of an optimal transport problem as in (1.5) for a suitable cost function  $c$  and an appropriate balayage cone  $\mathcal{A}$ . These turn out to be the *backward linear transfers* between probability distributions, a notion we introduced in [5,6] in order to extend the theory of optimal mass transportation while preserving the important duality established by Kantorovich.

*Backward linear transfers* are essentially lower semi-continuous convex functionals  $\mathcal{T}$  on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  whose partial maps  $\mathcal{T}_{\mu} : \nu \mapsto \mathcal{T}(\mu, \nu)$  are such that their

Legendre-Fenchel transforms lead to a map  $\mu \rightarrow \mathcal{T}_\mu^*$  that is *linear* with respect to  $\mu$ . This class of functionals is rich enough to naturally extend the cone of convex lower semi-continuous energies on Wasserstein space, Markov operators, and cost minimizing mass transports, but also many other couplings between probability measures to which Monge-Kantorovich theory does not readily apply. Examples include *optimal martingale transports* [16, 21], *optimal Skorokhod embeddings* [17, 18], optimal stochastic transports [13, 19, 23] and the *weak mass transports* of Talagrand [26], Marton [22], Gozlan et al. [20]. The class also includes various couplings such as the *Schrödinger bridge* associated to a reversible Markov process [15]. Here is the precise definition:

Let  $\mathcal{M}(X)$  is the class of signed Radon measures on  $X$ . The domain of a functional  $\mathcal{T} : \mathcal{M}(X) \times \mathcal{M}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  will be denoted

$$(1.6) \quad D(\mathcal{T}) := \{(\mu, \nu) \in \mathcal{M}(X) \times \mathcal{M}(Y); \mathcal{T}(\mu, \nu) < +\infty\}.$$

We consider for each  $\mu \in \mathcal{P}(X)$  (resp.,  $\nu \in \mathcal{P}(Y)$ ), the partial maps  $\mathcal{T}_\mu$  on  $\mathcal{P}(Y)$  (resp.,  $\mathcal{T}_\nu$  on  $\mathcal{P}(X)$ ) given by  $\nu \rightarrow \mathcal{T}(\mu, \nu)$  (resp.,  $\mu \rightarrow \mathcal{T}(\mu, \nu)$ ).

We let  $C(X)$  (resp.,  $B(X)$ ) (resp.,  $USC(X)$ ) (resp.,  $LSC(X)$ ) be the space of continuous (resp., Borel measurable) (resp., bounded above, proper and upper semi-continuous), (resp., bounded below, proper and lower semi-continuous) functions on a compact space  $X$ .

**Definition 1.2.** A functional  $\mathcal{T} : \mathcal{M}(X) \times \mathcal{M}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  is said to be a **backward linear transfer** (resp., **forward linear transfer**) if

- (1)  $\mathcal{T} : \mathcal{M}(X) \times \mathcal{M}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, convex, bounded below, and weak\* lower semi-continuous.
- (2)  $D(\mathcal{T}) \subset \mathcal{P}(X) \times \mathcal{P}(Y)$ .
- (3) There exists a map  $T^-$  (resp.,  $T^+$ ) from  $C(Y)$  (resp.,  $C(X)$ ) into the space of bounded above (resp., below) Borel-measurable functions on  $X$  (resp., on  $Y$ ) such that for all  $\mu \in \mathcal{P}(X)$  and  $g \in C(Y)$  (resp.,  $\nu \in \mathcal{P}(Y)$  and  $f \in C(X)$ ),

$$(1.7) \quad \mathcal{T}_\mu^*(g) = \int_X T^- g d\mu, \quad (\text{resp.}, \mathcal{T}_\nu^*(f) = - \int_Y T^+(-f) d\nu),$$

where  $\mathcal{T}_\mu^*$  (resp.,  $\mathcal{T}_\nu^*$ ) is the Fenchel-Legendre transform of  $\mathcal{T}_\mu$  (resp.,  $\mathcal{T}_\nu$ ) with respect to the duality of  $C(X)$  and  $\mathcal{M}(X)$  (resp.,  $C(Y)$  and  $\mathcal{M}(Y)$ ).

Note that since  $D(\mathcal{T}) \subset \mathcal{P}(X) \times \mathcal{P}(Y)$ , the Legendre transforms are simply

$$(1.8) \quad \mathcal{T}_\mu^*(g) := \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \mathcal{T}(\mu, \sigma) \right\} \quad \text{resp.}, \quad \mathcal{T}_\nu^*(f) := \sup_{\sigma \in \mathcal{P}(X)} \left\{ \int_X f d\sigma - \mathcal{T}(\sigma, \nu) \right\}.$$

The map  $T^-$  (resp.,  $T^+$ ) is then called the **backward (resp., forward) Kantorovich operator** associated to the transfer  $\mathcal{T}$ . We shall restrict our analysis to backward transfers since if  $\mathcal{T}$  is a forward linear transfer with  $T^+$  as a forward Kantorovich operator, then  $\tilde{\mathcal{T}}(\mu, \nu) = \mathcal{T}(\nu, \mu)$  is a backward linear transfer with  $T^- f = -T^+(-f)$  being the corresponding backward Kantorovich operator.

Note that since  $T^-$  and  $T^+$  arise from a Legendre transform, they must satisfy certain additional properties that we exhibit in Section 2. These -mostly non-linear- Kantorovich operators are important extensions of *Markov operators* and are ubiquitous in mathematical analysis and differential equations. They appear as the *réduite operators* in potential theory, as *filling scheme operators* in ergodic theory, as well as in Monge-Kantorovich duality of optimal mass transport (hence the name). They also include maps that associate to an initial (resp., final) state the solution at a given time  $t$  (resp., initial time) of a first or second-order *Hamilton-Jacobi equation*. They are actually general value functions in *dynamic programming principles* ([14] Section II.3) and also appear in the mathematical theory of *image processing* [2].

Linear transfers that can only take 0 or  $+\infty$  values are particularly interesting since we shall be able to characterize them completely in terms of the classical notion of balayage.

**Definition 1.3.** Say that a subset  $\mathcal{S}$  of  $\mathcal{P}(X) \times \mathcal{P}(Y)$  is a *backward transfer set* if its characteristic function

$$(1.9) \quad \mathcal{T}(\mu, \nu) := \begin{cases} 0 & \text{if } (\mu, \nu) \in \mathcal{S}, \\ +\infty & \text{otherwise,} \end{cases}$$

is a backward linear transfer.

**Definition 1.4.** 1) Say that a functional  $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  is **standard** if

$$(1.10) \quad \text{for each } x \in X, \text{ there is } \sigma \in \mathcal{P}(Y) \text{ with } \mathcal{T}(\delta_x, \sigma) < +\infty.$$

2) A set  $\mathcal{S} \subset \mathcal{P}(X) \times \mathcal{P}(Y)$  is then said to be **standard**<sup>1</sup> if its characteristic function is *standard*, that is if for each  $x \in X$ , there is  $\sigma \in \mathcal{P}(Y)$  with  $(\delta_x, \sigma) \in \mathcal{S}$ .

A typical example of a transfer set is given by the classical notion of convex order.

*Balayage of measures in convex order:* If  $X$  is a convex compact subset of a locally convex topological vector space, consider the functional

$$(1.11) \quad \mathcal{B}(\mu, \nu) = \begin{cases} 0 & \text{if } \mu \prec \nu \\ +\infty & \text{otherwise,} \end{cases}$$

and where  $\mu \prec \nu$  if and only if  $\int_X \phi d\mu \leq \int_X \phi d\nu$  for all convex continuous functions  $\phi$ . It follows from standard Choquet theory [12], that  $\mathcal{B}$  is a backward linear transfer whose Kantorovich operator is  $T^-f = \hat{f}$ , where  $\hat{f}$  is the upper semi-continuous concave upper envelope of  $f$ . It is also a forward linear transfer with  $T^+f$  being the lower semi-continuous convex lower envelope of  $f$ . In other words,

$$(1.12) \quad \mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec \nu\}$$

is a standard transfer set.

Balayage orders (though “restricted”) with respect to appropriate cones turned out to characterize transfer sets, as we shall prove in Section 3 the following. For

<sup>1</sup>It turned out that, motivated by some work of Dubins and Savage, Dellacherie-Meyer [8] had called such sets **gambling houses**. We shall adopt this terminology throughout this paper.

that, recall that a *positively 1-homogenous operator*  $T$  is one that verifies  $T(\lambda f) = \lambda Tf$  for any  $\lambda \geq 0$ .

**Theorem 1.5.** *Let  $\mathcal{S}$  be a weak\*-compact convex subset of  $\mathcal{P}(X) \times \mathcal{P}(Y)$ . The following are then equivalent:*

- (1)  $\mathcal{S}$  is a gambling house.
- (2)  $\mathcal{S}$  is a standard backward transfer set.
- (3)  $\mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \mu \prec_{\mathcal{A}} \nu\}$ , where  $\mathcal{A}$  is a balayage cone on  $X \sqcup Y$ .
- (4)  $(\mu, \nu) \in \mathcal{S}$  if and only if there exists  $\pi \in \mathcal{K}(\mu, \nu)$  such that  $(\delta_x, \pi_x) \in \mathcal{S}$  for  $\mu$ -almost  $x \in X$ .
- (5)  $\mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \nu \leq T_{\#}\mu\}$ , where  $T : C(Y) \rightarrow USC(X)$  is a positively 1-homogenous Kantorovich operator and  $T_{\#}\mu(g) := \int_X Tg d\mu$  for every  $g \in C(Y)$ .

**Definition 1.6.** Sets of the form (1.12) will be called **restricted (resp., true) balayage sets** if the cone  $\mathcal{A}$  is in  $LSC(X \sqcup Y)$  (resp., if  $X = Y$  and  $\mathcal{A} \subset LSC(X)$ ).

Note the two basic differences between an *extended balayage* and a *true balayage*, even when  $X = Y$ . For one, a probability measure is comparable to itself in a true balayage order, that is the set of ordered pairs must containing the diagonal  $\{(\mu, \mu); \mu \in \mathcal{P}(X)\}$ . Another difference is that a true balayage is clearly transitive, that is if  $\mu \prec_{\mathcal{A}} \sigma$  and  $\sigma \prec_{\mathcal{A}} \nu$  then  $\mu \prec_{\mathcal{A}} \nu$ . Neither property makes sense for an extended balayage, even if  $X = Y$ . We prove in Section 4 that these two properties actually suffice to have a backward transfer set represented by a true balayage.

**Definition 1.7.** Say that a subset  $\mathcal{S}$  of  $\mathcal{P}(X) \times \mathcal{P}(X)$  is **transitive** if

$$(1.13) \quad (\mu, \sigma) \in \mathcal{S} \text{ and } (\sigma, \nu) \in \mathcal{S} \Rightarrow (\mu, \nu) \in \mathcal{S}.$$

**Theorem 1.8.** *Let  $\mathcal{S}$  be a subset of  $\mathcal{P}(X) \times \mathcal{P}(X)$ . Then the following are equivalent:*

- (1)  $\mathcal{S}$  is a transitive backward transfer set containing the diagonal  $\{(\mu, \mu); \mu \in \mathcal{P}(X)\}$ .
- (2)  $\mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{A}} \nu\}$ , where  $\mathcal{A}$  is a balayage cone in  $C(X)$ .
- (3)  $\mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \nu \leq T_{\#}\mu\}$ , where  $T : C(X) \rightarrow USC(X)$  is an idempotent positively 1-homogenous Kantorovich operator such that  $Tf \geq f$  for all  $f \in C(X)$ .

Actually, we shall show that any standard backward transfer set is contained in a minimal transfer set represented by a true balayage.

The following example illustrates the difference between these two notions of balayage.

*The prescribed Markov transfer:* Let  $T : C(X) \rightarrow C(X)$  be a bounded linear positive operator such that  $T1 = 1$ , i.e., a Markov operator, then one can associate a backward linear transfer in the following way:

$$(1.14) \quad \mathcal{T}_T(\mu, \nu) := \begin{cases} 0 & \text{if } \nu = T^*(\mu) \\ +\infty & \text{otherwise,} \end{cases}$$

where  $T^* : \mathcal{M}(X) \rightarrow \mathcal{M}(X)$  is the adjoint operator. This is an example of a transfer set which cannot arise from a true balayage with respect to a cone  $\mathcal{A}$  in  $C(X)$ , unless

$T$  is the identity. On the other hand, we shall see in Section 4 that it is associated to a restricted balayage, and that the smallest true balayage set containing it is the one corresponding to balayage with respect to the cone of  $T$ -superharmonic functions.

So far, we have only encountered positively 1-homogenous Kantorovich operators. The next examples are not.

*Optimal mass transport:* Cost minimizing mass transports à la Monge-Kantorovich [27] are functionals on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  of the form,

$$(1.15) \quad \mathcal{T}_c(\mu, \nu) := \inf \left\{ \int_{X \times Y} c(x, y) d\pi; \pi \in \mathcal{K}(\mu, \nu) \right\},$$

where  $c(x, y)$  is a lower semi-continuous cost function on the product measure space  $X \times Y$ , and  $\mathcal{K}(\mu, \nu)$  is the set of probability measures  $\pi$  on  $X \times Y$  whose marginal on  $X$  (resp. on  $Y$ ) is  $\mu$  (resp.,  $\nu$ ) (*i.e.*, the transport plans). A consequence of the Monge-Kantorovich theory is that cost minimizing transports  $\mathcal{T}_c$  are both forward and backward linear transfers with Kantorovich operators given for any  $f \in C(X)$  (resp.,  $g \in C(Y)$ ), by

$$(1.16) \quad T_c^+ f(y) = \inf_{x \in X} \{c(x, y) + f(x)\} \quad \text{and} \quad T_c^- g(x) = \sup_{y \in Y} \{g(y) - c(x, y)\}.$$

*Optimal weak mass transport:* A typical example of a backward linear transfer is the following natural generalization of optimal transport: The **optimal weak transport**, formally introduced by Gozlan et. al. [20], in order to include previous work by Talagrand, Merton and others.

**Definition 1.9.** Let  $c : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  be a bounded below, lower semi-continuous function such that for each  $x \in X$ , the function  $\sigma \mapsto c(x, \sigma)$  is proper and convex. The *optimal weak transport* problem with cost  $c$  from  $\mu \in \mathcal{P}(X)$  to  $\nu \in \mathcal{P}(Y)$  is

$$(1.17) \quad \mathcal{V}_c(\mu, \nu) := \inf_{\pi} \left\{ \int_X c(x, \pi_x) d\mu(x); \pi \in \mathcal{K}(\mu, \nu) \right\}$$

where  $(\pi_x)_x$  is the disintegration of  $\pi$  with respect to  $\mu$ .

As shown in [1] and [20],  $\mathcal{V}_c(\mu, \nu) = \sup \{ \int_Y g d\nu - \int_X T^- g d\mu \}$ , where

$$(1.18) \quad T^- g(x) := \sup \left\{ \int_Y g d\sigma - c(x, \sigma); \sigma \in \mathcal{P}(Y) \right\},$$

a duality that will be key to our study. The effective domain of  $\mathcal{V}_c$  is not necessarily a backward transfer set, but we shall show in Section 3 that the domain of every backward linear transfer is contained in a minimal backward transfer set.

*Optimal balayage transport with cost:* This example which combines balayage and optimal transport can be described as follows:

**Definition 1.10.** Assume  $\mathcal{A}$  is a balayage cone in  $LSC(X \sqcup Y)$ , and let  $c : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  be as in definition (1.9). The following is also a backward linear

transfer:

$$(1.19) \quad \mathcal{B}_{c,\mathcal{A}}(\mu, \nu) = \begin{cases} \inf\{\int_X c(x, \pi_x) d\mu(x); \pi \in \mathcal{K}_{\mathcal{A}}(\mu, \nu)\} & \text{if } \mu \preceq_{\mathcal{A}} \nu, \\ +\infty & \text{otherwise.} \end{cases}$$

The corresponding Kantorovich operator is then

$$(1.20) \quad T^- f(x) = \sup\{\int_Y f d\sigma - c(x, \sigma); (x, \sigma) \in X \times \mathcal{P}(Y) \text{ and } \delta_x \prec_{\mathcal{A}} \sigma\}.$$

We shall refer to it as an **optimal restricted balayage transport** with cost  $c$ .

If  $X = Y$  and  $\mathcal{A}$  is a balayage cone in  $LSC(X)$ , then  $\mathcal{B}_{c,\mathcal{A}}$  is said to be an **optimal balayage transport** with cost  $c$ .

The following summarizes our main results. It roughly shows that mass transport theory is essentially balayage theory (or potential theory) with a “cost” assigned for moving distributions around.

**Theorem 1.11.** *Let  $\mathcal{T}$  be a standard functional on  $\mathcal{P}(X) \times \mathcal{P}(Y)$ . Then the following are equivalent:*

- (1)  $\mathcal{T}$  is a backward linear transfer.
- (2)  $\mathcal{T}$  is an optimal restricted balayage transport, i.e.,  $\mathcal{T} = \mathcal{B}_{c,\mathcal{A}}$  with cost function  $c(x, \sigma) := \mathcal{T}(\delta_x, \sigma)$  and where  $\mathcal{A}$  is a balayage cone on  $X \sqcup Y$ .

Moreover, for any  $(\mu, \nu) \in D(\mathcal{T})$ , there exists  $\pi \in \mathcal{K}_{\mathcal{A}}(\mu, \nu)$  such that  $(\delta_x, \pi_x) \in D(\mathcal{T})$ ,  $\delta_x \prec_{\mathcal{A}} \pi_x$  for  $\mu$ -a.e.  $x \in X$  and  $\mathcal{T}(\mu, \nu) = \int_X \mathcal{T}(\delta_x, \pi_x) d\mu(x)$ .

Furthermore,

- $D(\mathcal{T}) = D(\mathcal{B}_{c,\mathcal{A}}) \subset \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \mu \prec_{\mathcal{A}} \nu\}$ , and we have equality if and only if  $D(\mathcal{T})$  is a backward transfer set, in which case,

$$D(\mathcal{T}) = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \exists \pi \in \mathcal{K}_{\mathcal{A}}(\mu, \nu), \delta_x \prec_{\mathcal{A}} \pi_x \text{ for } \mu\text{-a.e. } x \in X\}.$$

- If  $X = Y$ , then  $D(\mathcal{T}) \subset \widehat{D(\mathcal{B}_{c,\hat{\mathcal{A}}})} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\hat{\mathcal{A}}} \nu\}$ , where  $\hat{\mathcal{A}}$  is a cone of lower semi-continuous functions on  $X$ .
- $D(\mathcal{T}) = \widehat{D(\mathcal{B}_{c,\hat{\mathcal{A}}})}$  if and only if  $D(\mathcal{T})$  is a transitive backward transfer set containing the diagonal  $\{(\delta_x, \delta_x); x \in X\}$ .

Note that (2) can be seen as an extension of Strassen’s theorem since the latter correspond to when  $\mathcal{T}$  is the zero-cost balayage transfer.

There are two important connections that emerge from our analysis:

- (1) The class  $\mathcal{K}(Y, X)$  of backward Kantorovich operators in the smallest “manageable” convex set in the space  $\mathcal{F}(Y, X)$  of maps from  $C(Y)$  to  $USC(X)$  that is stable under pointwise suprema, while containing the Markov operators.
- (2) Essentially any map  $T : C(Y) \rightarrow USC(X)$  has a lower “envelope” that is a backward Kantorovich operator. Moreover,  $\mathcal{K}(Y, X)$  is an important subclass of functional capacities that could be called “convex functional Choquet capacities,” since essentially any functional capacity has a backward Kantorovich operator lower envelope for the standard order on capacities.

These connections make the cone of Kantorovich operators particularly rich and flexible as it echoes the role of convex functions vis-a-vis general numerical functions.

Here is an outline of the paper. In Section 2, we exhibit the duality between linear transfers and Kantorovich operators as well as some of their properties that will be needed later. We also show that backward Kantorovich operators are Choquet functional capacities. In Section 3, we prove Theorem 1.5 and show that if  $\mathcal{T}$  is a standard backward linear transfer, then its effective domain  $D(\mathcal{T})$  is contained in a minimal restricted balayage set. In Section 4, we prove Theorem 1.8 and show that if  $X = Y$ , then  $D(\mathcal{T})$  is contained in a minimal true balayage set. In Section 5, we show that if  $\mathcal{T}$  is a standard backward linear transfer, then it can be represented as an *optimal weak transport*, and characterize those whose domains  $D(\mathcal{T})$  are backward transfer sets. This is then used in Section 6 to construct backward Kantorovich operators (resp., Choquet-Kantorovich operators) as envelopes of general maps from  $C(Y)$  to  $USC(X)$  (resp., functional capacities).

## 2. DUALITY AND KANTOROVICH OPERATORS AS CHOQUET FUNCTIONAL CAPACITIES

We shall restrict our study throughout to the case where  $X$  and  $Y$  are two compact metric spaces. Extensions to the non-compact case are of course possible and sometimes desirable, but could add complications and technical hypotheses that would take a way from this natural setting. Recall that  $USC(X)$  is the cone of proper, bounded above, upper semi-continuous functions on  $X$ .

**Definition 2.1.** A **backward Kantorovich operator** is a map  $T^- : C(Y) \rightarrow USC(X)$  that satisfies the following properties:

- (1)  $T^-$  is *monotone*, i.e., if  $g_1 \leq g_2$  in  $C(Y)$ , then  $T^-g_1 \leq T^-g_2$ .
- (2)  $T^-$  is *affine on the constants*, i.e., for any  $c \in \mathbb{R}$  and  $g \in C(Y)$ ,  $T^-(g+c) = T^-g + c$ .
- (3)  $T^-$  is a *convex operator*, that is for any  $\lambda \in [0, 1]$ ,  $g_1, g_2$  in  $C(Y)$ , we have

$$T^-(\lambda g_1 + (1 - \lambda)g_2) \leq \lambda T^-g_1 + (1 - \lambda)T^-g_2.$$

- (4)  $T^-$  is *lower semi-continuous* in the sense that if  $g_n \rightarrow g$  in  $C(Y)$  for the sup norm, then  $\liminf_{n \rightarrow \infty} T^-g_n \geq T^-g$ .

Similarly, a **forward Kantorovich operator** is a map  $T^+ : C(X) \rightarrow LSC(Y)$  that satisfies (1) and (2) above, but instead of (3) and (4), it satisfies:

- (3')  $T^+$  is a *concave operator*.
- (4')  $T^+$  is *upper semi-continuous*.

This class of operators is in a way “dual” to the class of backward linear transfers.

**Theorem 2.2.** *Let  $X$  and  $Y$  be two compact metric spaces. Then, the following assertions are equivalent:*

- (1)  $T^-$  is a backward Kantorovich operator from  $C(Y)$  to  $USC(X)$ .
- (2) There exists a backward linear transfer  $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  such that for all  $\mu \in \mathcal{P}(X)$  and  $g \in C(Y)$ ,

$$(2.1) \quad \mathcal{T}_\mu^*(g) = \int_X T^-g d\mu.$$



- (3) *There exists a proper lower semi-continuous function  $c : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  such that for all  $x \in X$ , the functional  $\sigma \mapsto c(x, \sigma)$  is convex, and for any  $g \in C(Y)$ ,*

$$(2.2) \quad T^-g(x) := \sup\left\{\int_Y g d\sigma - c(x, \sigma); \sigma \in \mathcal{P}(Y)\right\}.$$

We shall need the following lemma.

**Lemma 2.3.** *Let  $T : C(Y) \rightarrow USC(X)$  be a backward Kantorovich operator, and suppose  $x \in X$  is such that  $T0(x) > -\infty$ . Then, there exists  $\sigma \in \mathcal{P}(Y)$  such that*

$$(2.3) \quad \sup_{g \in C(Y)} \left\{ \int_Y g d\sigma - Tg(x) \right\} < +\infty.$$

**Proof:** Note first that the hypothesis yields that  $Tg(x) > -\infty$  for every  $g \in C(Y)$ , since by the first 2 properties of a Kantorovich operator, we have for every  $g \in C(Y)$ ,  $Tg - \inf_Y g = T(g - \inf_Y g) \geq T0$ . It follows that the functional  $\Phi_x : C(Y) \rightarrow \mathbb{R}$  defined by  $\Phi_x(g) = Tg(x)$  is clearly convex, lower semi-continuous and finite. Hence, by the Hahn-Banach theorem, it is the supremum of all affine functionals below it, that is, for every  $g \in C(Y)$ ,

$$\Phi_x(g) = \sup\left\{\int_Y g d\sigma + k; \sigma \in \mathcal{M}(Y), k \in \mathbb{R}, \sigma + k \leq \Phi_x \text{ on } C(Y)\right\}.$$

By compactness, there is  $k_0 \in \mathbb{R}$  and  $\sigma_0 \in \mathcal{M}(Y)$  such that  $\sigma_0 + k_0 \leq \Phi_x$  and

$$\Phi_x(0) = k_0 = \sup\{k; \sigma \in \mathcal{M}(Y), k \in \mathbb{R}, \sigma + k \leq \Phi_x \text{ on } C(Y)\}.$$

Since  $\Phi_x$  is affine on constants, we have  $\sigma_0(Y) + k_0 \leq \Phi_x(1) = 1 + \Phi_x(0) = 1 + k_0$ , hence  $\sigma_0(Y) \leq 1$ . On the other hand,

$$\Phi_x(-1) + 1 = \Phi_x(0) = \int_Y (-1) d\sigma_0 + k_0 + \sigma_0(Y) \leq \Phi_x(-1) + \sigma_0(Y),$$

hence  $\sigma_0(Y) \geq 1$  and therefore it is equal to 1.

To show that  $\sigma_0$  is a non-negative measure, let  $f \in C(Y)$  with  $0 \leq f \leq 1$ , then

$$1 - \sigma_0(f) = \sigma_0(1) - \sigma_0(f) = \sigma_0(1 - f) \leq \Phi_x(1 - f) - k_0.$$

Using the monotonicity of  $\Phi_x$ , we obtain

$$\sigma_0(f) \geq 1 - \Phi_x(1 - f) + k_0 = -\Phi_x(-f) + k_0 \geq -\Phi_x(0) + k_0 = 0.$$

It follows that  $\sigma_0 \in \mathcal{P}(Y)$ ,  $\sigma_0 + k_0 \leq \Phi_x$  on  $C(Y)$  and therefore for every  $g \in C(Y)$ , we have

$$Tg(x) = \sup\left\{\int_Y g d\sigma + k; \sigma \in \mathcal{M}(Y), k \in \mathbb{R}, \sigma + k \leq \Phi_x \text{ on } C(Y)\right\} \geq \int_Y g d\sigma_0 + k_0.$$

**Proof of Theorem 2.2:** (1)  $\Rightarrow$  (2) since if  $T^-$  is a backward Kantorovich operator then  $T^-0 \in USC(X)$ , that is (2.3) holds for some  $x \in X$ , and the following functional

$$(2.4) \quad \mathcal{T}(\mu, \nu) = \begin{cases} \sup_{g \in C(Y)} \left\{ \int_Y g d\nu - \int_X T^-g d\mu \right\} & \text{for all } (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y) \\ +\infty & \text{otherwise,} \end{cases}$$

is therefore proper.  $\mathcal{T}$  is also bounded below since

$$\begin{aligned} \inf\{\mathcal{T}(\mu, \nu); (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)\} &\geq \inf\left\{-\int_X T^-0 d\mu; \mu \in \mathcal{P}(X)\right\} \\ &> -\sup_X T^-(0) \\ &> -\infty, \end{aligned}$$

because  $T^-0$  is upper semi-continuous. It is clear that  $\mathcal{T}$  is jointly convex and lower semi-continuous on  $\mathcal{M}(X) \times \mathcal{M}(Y)$  since  $T^-$  is valued in  $USC(X)$ . Now define for  $\mu \in \mathcal{P}(X)$ , the functional  $F_\mu(g) := \int_X T^-g d\mu$ .

Assuming  $F_\mu(g) \neq -\infty$  for some  $g \in C(Y)$  and since  $g \mapsto T^-g$  is convex and lower semi-continuous in the sense defined above,  $F_\mu$  is then a convex and lower semi-continuous on  $C(Y)$ . By Fenchel-Moreau duality, we will be done if we show that  $F_\mu^*(\nu) = \mathcal{T}(\mu, \nu)$  for all  $\nu \in \mathcal{M}(Y)$ , where

$$F_\mu^*(\nu) = \sup\left\{\int_Y g d\nu - \int_X T^-g d\mu; g \in C(Y)\right\} \text{ on } \mathcal{M}(Y).$$

The equality holds when  $\nu \in \mathcal{P}(Y)$  by the definition of  $\mathcal{T}$ , so suppose  $\nu \in \mathcal{M}(Y)$  with  $\nu(Y) = \lambda \neq 1$ . Taking  $g(x) \equiv n \in \mathbb{Z}$ , we have  $T^-(g) = T^-(0+n) = n+T^-(0)$ , and therefore

$$F_\mu^*(\nu) \geq n\lambda - \int_X T^-(n) d\mu = n(\lambda - 1) - \int_X T^-(0) d\mu.$$

With  $n \rightarrow \pm\infty$ , depending on if  $\lambda < 1$  or  $\lambda > 1$ , we deduce  $F_\mu^*(\nu) = +\infty$ . Hence  $F_\mu^*(\nu) = \mathcal{T}(\mu, \nu)$  for all  $\nu \in \mathcal{M}(Y)$ , and it follows that

$$F_\mu(g) = \sup_{\nu \in \mathcal{M}(Y)} \left\{ \int_Y g d\nu - F_\mu^*(\nu) \right\} = \sup_{\nu \in \mathcal{M}(Y)} \left\{ \int_Y g d\nu - \mathcal{T}(\mu, \nu) \right\} = \mathcal{T}_\mu^*(g).$$

If now  $F_\mu \equiv -\infty$ , that is if  $F_\mu(g) = \int_X T^-g d\mu = -\infty$  for every  $g \in C(Y)$ , then formula (2.4) gives that  $\mathcal{T}_\mu \equiv +\infty$ , and again  $\mathcal{T}_\mu^*(g) = -\infty = \int_X T^-g d\mu$ .

(2)  $\Rightarrow$  (1): Suppose now  $\mathcal{T}$  is a backward linear transfer, and let  $T^-$  be the operator associated to it in Definition 1.2. Since  $\mathcal{T}$  is proper, we have  $\mathcal{T}(\mu, \nu) < +\infty$  for some  $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ , hence

$$\int_X T^-0 d\mu = \mathcal{T}_\mu^*(0) = \sup\{-\mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(Y)\} \geq -\mathcal{T}(\mu, \nu) > -\infty,$$

from which follows that there is  $x \in X$  such that  $T^-0(x) > -\infty$ .

To show that  $T^-$  is a Kantorovich operator, note that for every  $g \in C(Y)$  and any  $x \in X$ ,

$$(2.5) \quad T^-g(x) = \mathcal{T}_{\delta_x}^*(f) = \sup_{\nu \in \mathcal{P}(Y)} \left\{ \int_Y g d\nu - \mathcal{T}(\delta_x, \nu) \right\}.$$

It is clear that  $T^-$  is monotone, convex and affine on the constants. To show that  $T^-g \in USC(X)$  for every  $g \in C(Y)$ , note first that  $T^-g$  is proper since  $T^-0(x) > -\infty$  and the monotonicity yields that it is also the case for any  $g \in C(Y)$ . Moreover, if  $x_n \rightarrow x$  in  $X$  so that -up to a subsequence-  $\limsup_{n \rightarrow \infty} T^-g(x_n) = \lim_{j \rightarrow \infty} T^-g(x_{n_j})$ , then there is nothing to prove if the latter is equal to  $-\infty$ . However, if  $\lim_{j \rightarrow \infty} T^-g(x_{n_j}) > -\infty$ , then we let  $\nu_j$  achieve the supremum above when

$x = x_{n_j}$  (the supremum is achieved by upper semi-continuity of  $\nu \mapsto \int_Y g d\nu - \mathcal{T}(\delta_x, \nu)$  on the compact space  $\mathcal{P}(Y)$ ). By the weak\* compactness of  $\mathcal{P}(Y)$ , we may extract a further subsequence if necessary and assume  $\nu_j \rightarrow \bar{\nu}$  for some  $\bar{\nu} \in \mathcal{P}(Y)$ . It follows from the weak\*- lower semi-continuity of  $\mathcal{T}$

$$\begin{aligned} \limsup_{n \rightarrow \infty} T^-g(x_n) &= \lim_{j \rightarrow \infty} T^-g(x_{n_j}) \leq \int_Y g d\bar{\nu} - \mathcal{T}(\delta_x, \bar{\nu}) \\ &\leq \sup_{\nu \in \mathcal{P}(Y)} \left\{ \int_Y g d\nu - \mathcal{T}(\delta_x, \nu) \right\} \\ &= T^-g(x), \end{aligned}$$

hence  $T^-g \in USC(X)$ .

This also yields the following Lipschitz property: for any  $g, h$  in  $C(Y)$ , and any  $x \in X$ , where  $T^-h(x) > -\infty$ ,

$$(2.6) \quad T^-g(x) - T^-h(x) \leq \|g - h\|_\infty,$$

since if  $T^-g(x) > -\infty$  (otherwise there is nothing to prove), then

$$\begin{aligned} T^-g(x) &= \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \mathcal{T}(\delta_x, \sigma) \right\} \\ &\leq \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y h d\sigma - \mathcal{T}(\delta_x, \sigma) \right\} + \|g - h\|_\infty \\ &= T^-h(x) + \|g - h\|_\infty. \end{aligned}$$

This immediately yields the lower semi-continuity property since if  $g_n \rightarrow g \in C(Y)$ , then,  $T^-g(x) \leq T^-g_n(x) + \|g - g_n\|_\infty$ , and  $T^-g \leq \liminf_{n \rightarrow \infty} T^-g_n(x)$ .

(2)  $\Rightarrow$  (3) is immediate by taking  $c(x, \sigma) = \mathcal{T}(\delta_x, \sigma)$  and using the definition of a Legendre transform.

That (3)  $\Rightarrow$  (1) is immediate from expression (2.2), which was used in (2)  $\Rightarrow$  (1) to establish the 4 properties of a Kantorovich operator.

Finally, note that for each  $\mu \in \mathcal{P}(X)$ ,  $\mathcal{T}_\mu$  is convex, lower semi-continuous and bounded below such that  $\mathcal{T}_\mu^*(g) = \int_X T^-g d\mu$ . It follows that for each  $\nu \in \mathcal{P}(Y)$ ,

$$\mathcal{T}(\mu, \nu) = \mathcal{T}_\mu(\nu) = \mathcal{T}_\mu^{**}(\nu) = \sup_{g \in C(Y)} \left\{ \int_Y g d\nu - \mathcal{T}_\mu^*(g) \right\} = \sup_{g \in C(Y)} \left\{ \int_Y g d\nu - \int_X T^-g d\mu \right\}.$$

**Proposition 2.4.** *Consider the class  $\mathcal{K}(Y, X)$  of backward Kantorovich operators from  $C(Y)$  to  $USC(X)$ . Then,*

- (1) *If  $T_1$  and  $T_2$  are in  $\mathcal{K}(Y, X)$ , and  $\lambda \in [0, 1]$ , then  $\lambda T_1 + (1 - \lambda)T_2 \in \mathcal{K}(Y, X)$ .*
- (2) *If  $\lambda \in \mathbb{R}^+$  and  $T \in \mathcal{K}(Y, X)$ , then the map  $(\lambda \cdot T)f := \frac{1}{\lambda}T(\lambda f)$  for any  $f \in C(Y)$ , belongs to  $\mathcal{K}(Y, X)$ .*
- (3) *If  $T_1$  and  $T_2$  are in  $\mathcal{K}(Y, X)$ , then the map  $T_1 \star T_2$  defined for  $f \in C(Y)$  and  $x \in X$  by*

$$(T_1 \star T_2)f(x) := \sup_{\sigma \in \mathcal{P}(Y)} \inf_{g, h \in C(Y)} \int_Y (f - g - h) d\sigma + T_1g(x) + T_2h(x)$$

*is in  $\mathcal{K}(Y, X)$*

- (4) If  $T_1$  and  $T_2$  are in  $\mathcal{K}(Y, X)$ , then the map  $(T_1 \vee T_2)f(x) := T_1f(x) \vee T_2f(x)$  for any  $f \in C(Y)$ , and  $x \in X$  belongs to  $\mathcal{K}(Y, X)$ .
- (5) If  $(T_i)_{i \in I}$  is a family in  $\mathcal{K}(Y, X)$ , such that for every  $f \in C(Y)$ , the function  $\sup_{i \in I} T_i f$  is bounded above on  $X$ , then the operator  $f \rightarrow \overline{\sup_{i \in I} T_i f}$  is a Kantorovich operator, where here  $\bar{g}(x) = \inf\{h(x); h \in C(X), h \geq g \text{ on } X\}$  is the smallest upper semi-continuous function above  $g$ .

*Proof.* (1) and (2) are immediate, while (3) follows from the fact that

$$T_1 \star T_2 f(x) = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y f d\sigma - \mathcal{T}_1(x, \sigma) - \mathcal{T}_2(x, \sigma) \right\},$$

where  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are the linear transfers corresponding to  $T_1$  and  $T_2$  respectively. In other words, in view of Theorem 2.2, the convex cost associated with  $T_1 \star T_2$  is

$$c(x, \sigma) = \mathcal{T}_1(\delta_x, \sigma) + \mathcal{T}_2(\delta_x, \sigma),$$

hence  $T_1 \star T_2$  is a Kantorovich operator.

4) is immediate from the definition of a Kantorovich operator, while for 5) first notice that the operator  $T_\infty f := \sup_{i \in I} T_i f$  clearly satisfies properties (1), ..., (4) of a backward Kantorovich operator. Indeed, the monotonicity, convexity, and affine with respect to constants, properties of  $S_\infty$  are readily inherited from each  $S_i$  being Kantorovich. The map  $f \rightarrow T_\infty f(x)$  is also lower semi-continuous. The first three properties extend to the operator  $f \rightarrow \overline{T_\infty f}$ . It remains to check the lower semi-continuity. For this, suppose  $g_k \rightarrow g$  in  $C(X)$  for the sup norm. If  $\liminf_k T_\infty g_k(x) = +\infty$ , there is nothing to prove. Otherwise, we have from (2.6) that  $T_i g(x) \leq T_i g_k(x) + \|g - g_k\|_\infty$ , so that  $\sup_i T_i g(x) \leq \sup_i T_i g_k(x) + \|g - g_k\|_\infty$  and therefore

$$\overline{\sup_i T_i g(x)} \leq \overline{\sup_i T_i g_k(x)} + \|g - g_k\|_\infty,$$

since  $\overline{f + c} = \overline{f} + c$  for any constant  $c$ , hence  $\overline{T_\infty g}(x) \leq \liminf_{k \rightarrow \infty} \overline{T_\infty g_k}(x)$ . □

**Remark 2.5.** The duality allows to construct and identify new Kantorovich operators. Indeed, let  $I : \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  be a bounded below, convex, weak\*-lower semi-continuous function on  $\mathcal{P}(Y)$  and consider

$$(2.7) \quad \mathcal{I}(\mu, \nu) := I(\nu) \quad \text{for all } (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y),$$

so that  $\mathcal{I}$  is a backward linear transfer with corresponding backward Kantorovich operator  $T^-g \equiv I^*(g)$  (hence a constant function of  $x$ ). If now  $\mathcal{T}$  is a general linear transfer, then the following functional on  $\mathcal{P}(X) \times \mathcal{P}(Y)$ ,

$$(2.8) \quad \mathcal{T}_I(\mu, \nu) := \inf \left\{ \int_X [\mathcal{T}(x, \pi_x) + I(\pi_x)] d\mu(x); \pi \in \mathcal{K}(\mu, \nu) \right\}$$

is also a linear transfer, and if  $\mathcal{T}$  is given by an optimal mass transport  $\mathcal{T}_c$  associated to a proper lower semi-continuous cost function  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ , whose backward Kantorovich operator is  $T_c g(x) := \sup\{g(y) - c(x, y); y \in Y\}$ , then

$$(2.9) \quad \mathcal{T}_I(\mu, \nu) = \inf \left\{ \int_X [c(x, y) d\pi_x(y) + I(\pi_x)] d\mu(x); \pi \in \mathcal{K}(\mu, \nu) \right\},$$

and the associated Kantorovich operator is given by

$$\begin{aligned} T_I^- g(x) &:= T^- \oplus I^*(g)(x) = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \mathcal{T}(x, \sigma) - I(\sigma) \right\} \\ &= \sup \left\{ \int_Y (g(y) - c(x, y)) d\sigma(y) - I(\sigma); \sigma \in \mathcal{P}(Y) \right\} \\ &= I^*(g_{c,x}), \end{aligned}$$

where  $g_{c,x}$  is the function  $y \mapsto g(y) - c(x, y)$ .

- (1) If  $I$  is the potential energy functional  $I(\nu) = \epsilon \int_Y V(y) d\nu(y)$ , where  $V$  is a bounded below lower semi-continuous potential on  $Y$ , then  $T_I^- g = T_c^-(g - \epsilon V)$ .
- (2) Less obvious is the case when  $I$  is the relative entropy  $H_{\nu_0}$  with respect to  $\nu_0$ , i.e.,

$$(2.10) \quad H_{\nu_0}(\nu) := \begin{cases} \int_Y \frac{d\nu}{d\nu_0} \log\left(\frac{d\nu}{d\nu_0}\right) d\nu_0 & \text{if } \nu \ll \nu_0 \\ +\infty & \text{otherwise.} \end{cases}$$

In this case,  $T_{\epsilon H} f(x) = \epsilon \log \int_Y e^{\frac{g(y)-c(x,y)}{\epsilon}} d\nu_0(y)$ , which corresponds to an entropic regularization of  $T_c$ .

Denote by  $USC_f(X)$  (resp.,  $USC_b(X)$ ) the cone of functions in  $USC(X)$  that are finite (resp., bounded below), and by  $USC_\sigma(X)$  the closure of  $USC(X)$  with respect to monotone increasing limits.

**Definition 2.6.** Say that a Kantorovich operator  $T$  is **standard** (resp., **regular**) if  $T$  maps  $C(Y)$  to  $USC_f(X)$  (resp.,  $USC_b(X)$ ).

The following proposition follows readily from the above theorem.

**Proposition 2.7.** *Let  $T$  be a Kantorovich operator and  $\mathcal{T}$  its associated linear transfer. Then, the following are equivalent:*

- (1)  $T$  is **standard** (resp., **regular**).
- (2) The function  $T(0)$  is finite (resp., bounded below).
- (3) The corresponding linear transfer  $\mathcal{T}$  satisfies

$$(2.11) \quad \inf_{\sigma \in \mathcal{P}(Y)} \mathcal{T}(x, \sigma) < +\infty \text{ for all } x \in X,$$

$$(2.12) \quad (\text{resp., } k := \sup_{x \in X} \inf_{\sigma \in \mathcal{P}(Y)} \mathcal{T}(x, \sigma) < +\infty.)$$

If  $T$  is a standard Kantorovich operator, then  $T$  is a contraction from  $C(Y)$  to  $USC_f(X)$ , that is

$$(2.13) \quad \sup_{x \in X} |T_-g(x) - T_-h(x)| \leq \|g - h\|_\infty.$$

A regular operator is clearly standard and  $T + k$  then maps  $C_+(Y)$  to  $USC_+(X)$ . Let  $F(X)$  (resp  $F_+(X)$ ) (resp.,  $F_b(X)$ ) denote the class of all extended numerical (resp., non-negative) (resp., bounded below) functions on  $X$ .

**Definition 2.8.** A Choquet functional capacity is a map  $T : F_+(Y) \rightarrow F_+(X)$  such that

- (1)  $T$  is monotone, i.e.,  $f \leq g \Rightarrow Tf \leq Tg$ .
- (2)  $T$  maps  $USC(Y)$  to  $USC(X)$  and if  $g_n, g \in USC(Y)$  and  $g_n \downarrow g$ , then  $T^-g_n \downarrow T^-g$ .
- (3) If  $g_n, g \in F_+(Y)$  with  $g_n \uparrow g$ , then  $T^-g_n \uparrow T^-g$ .

The class  $USC_\sigma(X)$  will denote the closure of  $USC(X)$  with respect to monotone increasing limits.

**Theorem 2.9.** Let  $\mathcal{T}$  be a backward linear transfer and let  $T^- : C(Y) \rightarrow USC(X)$  be the associated Kantorovich operator. Then

- (1)  $T^-$  can be extended to be a functional from  $F(Y)$  to  $F(X)$  via the formula

$$(2.14) \quad T^-g(x) = \sup \left\{ \int_Y^* g d\nu - \mathcal{T}(\delta_x, \nu); \nu \in \mathcal{P}(Y), (\delta_x, \nu) \in D(\mathcal{T}) \right\},$$

where  $\int_Y^* g d\nu$  is the outer integral of  $g$  with respect to  $\nu$ .

Moreover,  $T^-$  maps bounded above functions on  $Y$  to bounded above functions on  $X$ . It is monotone and satisfies  $T^-g + c = T^-(g + c)$  for every  $g \in F(Y)$  and  $c \in \mathbb{R}$ .

- (2) For any upper semi-continuous functions  $g$  on  $Y$ , we have

$$(2.15) \quad T^-g(x) := \inf \{ T^-h(x); h \in C(Y), h \geq g \},$$

and  $T^-$  maps  $USC_b(Y)$  to  $USC(X)$ .

- (3) If  $T^-$  is standard (resp., regular), then it maps  $USC_b(Y)$  to  $USC_f(X)$  (resp.,  $USC_b(Y)$ ).
- (4) If  $T^-$  is regular, then the map  $T^- + k$  is a functional capacity that maps  $F_+(Y)$  to  $F_+(X)$ , and consequently, if  $g$  is a  $K$ -analytic function that is bounded below on  $Y$ , then

$$(2.16) \quad T^-g(x) := \sup \{ T^-h(x); h \in USC(Y), h \leq g \}.$$

*Proof.* 1) Note first that formula (2.14) makes sense for any function  $g : Y \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  provided one uses the outer integral  $\int^* g d\nu$  in formula (2.14). Moreover,  $T^-g(x) \leq \sup_{y \in Y} g(y) - m_{\mathcal{T}}$ , where  $m_{\mathcal{T}}$  is a lower bound for  $\mathcal{T}$ , hence if  $g$  is bounded above then,  $T^-g$  is also bounded above. The monotonicity and the affine properties on constants are clear. Finally note that since there exists  $(\delta_x, \sigma) \in D(\mathcal{T})$ , we have for any  $g \in F_b(Y)$ ,  $T^-g(x) \geq \inf_Y g - \mathcal{T}(\delta_x, \sigma) > -\infty$ , hence  $T^-g$  is proper.

2. We now show that the extension as defined in (2.14) and still denoted  $T^-$ , maps  $USC(Y)$  to  $USC(X)$ . Indeed, if  $g \in USC(Y)$  and  $h_n \searrow g$  is a decreasing sequence of continuous functions converging to  $g$ , then,  $T^-g \leq \inf_n T^-h_n$  so that if  $\inf_n T^-h_n(x) = -\infty$ , we have equality at that point. On the other hand, if  $\inf_n T^-h_n(x) > -\infty$ , then

$$T^-g(x) \leq \inf_n T^-h_n(x) = \inf_n \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y h_n d\sigma - \mathcal{T}(\delta_x, \sigma) \right\} = \inf_n \int_Y h_n d\sigma_n - \mathcal{T}(\delta_x, \sigma_n),$$

where the supremum is achieved for some probability measure  $\sigma_n$  because  $\sigma \mapsto \int_Y h_n d\sigma - \mathcal{T}(\delta_x, \sigma)$  is weak\* upper semi-continuous and bounded above on the compact set  $\mathcal{P}(Y)$ .

Let  $(n_k)_k$  be a subsequence such that  $\inf_n \int_Y h_n d\sigma_n - \mathcal{T}(\delta_x, \sigma_n) = \lim_{k \rightarrow \infty} \int_Y h_{n_k} d\sigma_{n_k} - \mathcal{T}(\delta_x, \sigma_{n_k})$  and by weak\* compactness of  $\mathcal{P}(Y)$ , extract a further increasing subsequence (that we call again  $n_k$ ) so that  $\sigma_{n_k} \rightarrow \bar{\sigma}$ . For any  $j \leq k$ , we have  $h_{n_k} \leq h_{n_j}$ , hence for this fixed  $j$ , we have that  $h_{n_j} \in C(Y)$  and so  $\int h_{n_j} d\sigma_{n_k} \rightarrow \int h_{n_j} d\bar{\sigma}$  as  $k \rightarrow \infty$ , and since  $\mathcal{T}$  is lower semi-continuous, we obtain

$$\inf_n \int_Y h_n d\sigma_n - \mathcal{T}(\delta_x, \sigma_n) \leq \lim_{k \rightarrow \infty} \int_Y h_{n_j} d\sigma_{n_k} - \mathcal{T}(\delta_x, \sigma_{n_k}) \leq \int_Y h_{n_j} d\bar{\sigma} - \mathcal{T}(\delta_x, \bar{\sigma}).$$

Finally  $\int_Y h_{n_j} d\bar{\sigma} \rightarrow \int_Y g d\bar{\sigma}$  by monotone convergence, and we obtain

$$\begin{aligned} T^-g(x) &\leq \inf_n T^-h_n(x) \leq \inf_n \int_Y h_n d\sigma_n - \mathcal{T}(\delta_x, \sigma_n) \\ &\leq \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \mathcal{T}(\delta_x, \sigma) \right\} \\ &= T^-g(x). \end{aligned}$$

It follows that  $T^-g(x) := \inf\{T^-h(x); h \in C(Y), h \geq g\}$ , and  $T^-g$  is upper semi-continuous. If now  $g \in USC_f(Y)$ , then  $T^-g$  is proper and therefore belongs to  $USC(X)$ .

3) Suppose now that  $g_n, g \in USC(Y)$  are such that  $g_n \searrow g$ . We claim that  $T^-g_n \searrow T^-g$ . Indeed, by the monotonicity property of  $T^-$ , we have  $T^-g(x) \leq \liminf_n T^-g_n(x)$ . If for some  $n$ ,  $T^-g_n(x) = -\infty$ , then  $T^-g(x) = -\infty$  and there is nothing to prove. Otherwise,  $T^-g_n(x) > -\infty$  for all  $n$ , in which case the expression (2.14) is finite. The map  $\sigma \mapsto \int_Y g_n d\sigma - \mathcal{T}(\delta_x, \sigma)$  is weak\* upper semi-continuous (since  $\sigma \mapsto \int_Y h d\sigma$  is weak\* upper semi-continuous for any  $h \in USC(Y)$ ), so it achieves its supremum at some  $\sigma_n$ , i.e.,  $T^-g_n(x) = \int g_n d\sigma_n - \mathcal{T}(\delta_x, \sigma_n)$ . Extract now an increasing subsequence  $n_k$  so that  $\limsup_n T^-g_n(x) = \lim_k T^-g_{n_k}(x)$  and  $\sigma_{n_k} \rightarrow \bar{\sigma}$ . Similarly to the proof above, we get by the monotonicity of  $g_n$ , that

$$(2.17) \quad T^-g_{n_k}(x) \leq \int g_{n_j} d\sigma_{n_k} - \mathcal{T}(\delta_x, \sigma_{n_k}) \quad \text{for fixed } j \leq k.$$

As  $g_{n_j} \in USC(Y)$  and  $\sigma_{n_k} \rightarrow \bar{\sigma}$ , it follows that  $\limsup_{k \rightarrow \infty} \int g_{n_j} d\sigma_{n_k} \leq \int g_{n_j} d\bar{\sigma}$ . Hence upon taking  $\limsup_{k \rightarrow \infty}$  in (2.17), we get that  $\limsup_n T^-g_n(x) \leq \int_Y g_{n_j} d\bar{\sigma} - \mathcal{T}(\delta_x, \bar{\sigma})$ . Now let  $j \rightarrow +\infty$  and use monotone convergence to conclude that

$$\limsup_n T^-g_n(x) \leq \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int g d\sigma - \mathcal{T}(\delta_x, \sigma) \right\} = T^-g(x).$$

Suppose now  $g_n, g \in F(Y)$  with  $g_n \nearrow g$ , we shall show that  $T^-g_n \nearrow T^-g$ . Indeed, again, by the monotonicity property for  $T^-$ , we have  $T^-g(x) \geq \limsup_n T^-g_n(x)$ . On the other hand,

$$T^-g_n(x) = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y^* g_n d\sigma - \mathcal{T}(\delta_x, \sigma) \right\} \geq \int_Y^* g_n d\sigma - \mathcal{T}(\delta_x, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(Y),$$

hence by monotone convergence,  $\liminf_n T^-g_n(x) \geq \int^* g d\sigma - \mathcal{T}(\delta_x, \sigma)$  for all  $\sigma$ . Taking the supremum over  $\sigma$  yields  $\liminf_n T^-g_n(x) \geq \sup_{\sigma \in \mathcal{P}(Y)} \{ \int_Y^* g d\sigma - \mathcal{T}(\delta_x, \sigma) \} = T^-g(x)$ .

4) If now  $T$  is a regular Kantorovich operator, then  $T$  maps  $USC_b(Y)$  to  $USC_b(X)$  and  $T + k$  maps  $F_+(Y)$  to  $F_+(X)$ . It follows from the above established properties that  $T + k$  is a functional capacity. Note that for each  $x \in X$ , the set function  $T^x(A) = T(\chi_A)(x) + k$  is then a non-negative regular Choquet capacity. Moreover, for any Radon measure  $\mu$ , the functional  $(T + k)_\# \mu$  defined as  $(T + k)_\# \mu(f) = \int_X (Tf(x) + k) d\mu(x)$  is also a regular capacity.

A celebrated theorem of Choquet now yields that any  $K$ -analytic function  $g \in F_+(Y)$ , is ‘‘capacitable’’, that is

$$T^-g(x) = \sup\{T^-h(x); h \in USC(Y), h \leq g\}.$$

This holds in particular for any bounded above function  $g \in USC_\sigma(Y)$ , i.e., the closure of  $USC(Y)$  with respect to monotone increasing limits. Actually, it can be readily be seen that  $T^-g(x) \geq \sup\{T^-h(x); h \in USC(Y), h \leq g\}$ , and for an increasing  $h_n \nearrow g$ , with  $h_n \in USC(Y)$ , we have  $T^-g = \sup_n T^-h_n$  by the 3rd property of the capacity.  $\square$

**Corollary 2.10.** *Let  $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  be a backward linear transfer. Then,*

(1) *For any  $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ , we have*

$$\mathcal{T}(\mu, \nu) = \sup \left\{ \int_Y g d\nu - \int_X T^-g d\mu; g \in USC_b(Y) \right\}.$$

(2) *The Legendre transform formula (1.7) for  $\mathcal{T}_\mu$  extends from  $C(Y)$  to  $USC(Y)$ ; that is, for  $\mu \in \mathcal{P}(X)$ , we have for any  $g \in USC(Y)$ ,*

$$(2.18) \quad \mathcal{T}_\mu^*(g) := \sup\left\{ \int_Y g d\sigma - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(Y) \right\} = \int_X T^-g d\mu.$$

*Proof.* 1. For  $g \in USC_b(Y)$ , take a monotone decreasing sequence  $g_n \in C(Y)$  with  $g_n \downarrow g$ . Since  $T^-$  is a capacity, we have

$$\lim_{n \rightarrow \infty} \left( \int_Y g_n d\nu - \int_X T^-g_n d\mu \right) = \int_Y g d\nu - \int_X T^-g d\mu,$$

from which we conclude  $\mathcal{T}(\mu, \nu) \geq \sup \left\{ \int_Y g d\nu - \int_X T^-g d\mu; g \in USC_b(Y) \right\}$ . The reverse inequality is immediate since  $C(Y) \subset USC_b(Y)$ .

2. Let  $g \in USC(Y)$  and take  $g_n \searrow g$  with  $g_n \in C(Y)$ . We have

$$\int_Y g_n d\sigma - \mathcal{T}(\mu, \sigma) \leq \sup\left\{ \int_Y g_n d\sigma - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(Y) \right\} = \int_X T^-g_n d\mu$$

so that by the capacity property of  $T^-$ , we have  $\int_Y g d\sigma - \mathcal{T}(\mu, \sigma) \leq \int_X T^-g d\mu$ , and consequently

$$(2.19) \quad \sup\left\{ \int_Y g d\sigma - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(Y) \right\} \leq \int_X T^-g d\mu.$$



On the other hand, by monotonicity,  $T^-g \leq T^-g_n$ , so

$$(2.20) \quad \int_X T^-g \, d\mu \leq \int_X T^-g_n \, d\mu \leq \sup\left\{\int_Y g_n \, d\sigma - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(Y)\right\}.$$

If now  $\sup\{\int_Y g_n \, d\sigma - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(Y)\} = -\infty$  for some  $n$ , there is nothing left to prove. Otherwise, The supremum on the right-hand side of (2.20) is achieved by some  $\sigma_n$ . Extract an increasing subsequence  $n_j$  so that  $\sigma_{n_j} \rightarrow \sigma$  for some  $\sigma \in \mathcal{P}(Y)$ . Then if  $i \leq j$ , we have  $g_{n_j} \leq g_{n_i}$ , so that

$$\int_X T^-g \, d\mu \leq \int_Y g_{n_i} \, d\sigma_{n_j} - \mathcal{T}(\mu, \sigma_{n_j}) \quad \text{for } i \leq j$$

where upon sending  $j \rightarrow \infty$  yields  $\int_X T^-g \, d\mu \leq \int_Y g_{n_i} \, d\sigma - \mathcal{T}(\mu, \sigma)$ , and finally as  $i \rightarrow +\infty$ , monotone convergence yields the reverse inequality of (2.19).  $\square$

Now that standard Kantorovich operators can be extended so as to map  $USC_b(Y)$  into  $USC_b(X)$ , we can compose them in the following way.

**Proposition 2.11.** *Let  $X_1, \dots, X_n$  be  $n$  compact spaces, and suppose for each  $i = 1, \dots, n$ ,  $\mathcal{T}_i$  is a regular backward linear transfer on  $\mathcal{P}(X_{i-1}) \times \mathcal{P}(X_i)$  with Kantorovich operator  $T_i^- : C(X_i) \rightarrow USC_f(X_{i-1})$ . For any probability measures  $\mu$  on  $X_1$  (resp.,  $\nu$  on  $X_n$ ), define*

$$\begin{aligned} &\mathcal{T}_1 \star \mathcal{T}_2 \dots \star \mathcal{T}_n(\mu, \nu) \\ &:= \inf\{\mathcal{T}_1(\mu, \sigma_1) + \mathcal{T}_2(\sigma_1, \sigma_2) \dots + \mathcal{T}_n(\sigma_{n-1}, \nu); \sigma_i \in \mathcal{P}(X_i), i = 1, \dots, n - 1\}. \end{aligned}$$

Then,  $\mathcal{T} := \mathcal{T}_1 \star \mathcal{T}_2 \dots \star \mathcal{T}_n$  is a linear backward transfer with a Kantorovich operator

$$(2.21) \quad T^- = T_1^- \circ T_2^- \circ \dots \circ T_n^-.$$

In other words, the following duality formula holds:

$$(2.22) \quad \mathcal{T}(\mu, \nu) = \sup\left\{\int_{X_n} g(y) \, d\nu(y) - \int_{X_1} T_1^- \circ T_2^- \circ \dots \circ T_n^- g(x); g \in C(X_n)\right\}.$$

*Proof.* By an obvious induction, it suffices to show the proposition for two transfers. For that, note that since  $\mathcal{T}_1$  (resp.,  $\mathcal{T}_2$ ) is jointly convex and weak\*-lower semi-continuous on  $\mathcal{P}(X_1) \times \mathcal{P}(X_2)$  (resp.,  $\mathcal{P}(X_2) \times \mathcal{P}(X_3)$ ), then  $(\mathcal{T}_1 \star \mathcal{T}_2)_\mu : \nu \rightarrow (\mathcal{T}_1 \star \mathcal{T}_2)(\mu, \nu)$  is convex and weak\*-lower semi-continuous. Consider their corresponding Kantorovich operator  $T_1^-$  (resp.,  $T_2^-$ ) from  $USC_b(X_2)$  to  $USC_b(X_1)$  (resp.,  $USC_b(X_3)$  to  $USC_b(X_2)$ ) and calculate the following Legendre transform:

For  $g \in C(X_3)$ ,

$$\begin{aligned}
 (\mathcal{T}_1 \star \mathcal{T}_2)_\mu^*(g) &= \sup_{\nu \in \mathcal{P}(X_3)} \sup_{\sigma \in \mathcal{P}(X_2)} \left\{ \int_{X_3} g \, d\nu - \mathcal{T}_1(\mu, \sigma) - \mathcal{T}_2(\sigma, \nu) \right\} \\
 &= \sup_{\sigma \in \mathcal{P}(X_2)} \{ (\mathcal{T}_2)_\sigma^*(g) - \mathcal{T}_1(\mu, \sigma) \} \\
 &= \sup_{\sigma \in \mathcal{P}(X_2)} \left\{ \int_{X_2} T_2^-(g) \, d\sigma - \mathcal{T}_1(\mu, \sigma) \right\} \\
 &= (\mathcal{T}_1)_\mu^*(T_2^-(g)) \\
 &= \int_{X_1} T_1^- \circ T_2^- g \, d\mu.
 \end{aligned}$$

In other words,  $\mathcal{T}_1 \star \mathcal{T}_2(\mu, \nu) = \sup \{ \int_{X_3} g(x) \, d\nu(x) - \int_{X_1} T_1^- \circ T_2^- g \, d\mu; f \in C(X_3) \}$ , which means that  $\mathcal{T}_1 \star \mathcal{T}_2$  is a backward linear transfer on  $X_1 \times X_3$  with Kantorovich operator equal to  $T_1^- \circ T_2^-$ . □

### 3. GAMBLING HOUSES AND POSITIVELY HOMOGENOUS KANTOROVICH OPERATORS

In this section we study  $\{0, +\infty\}$ -valued linear transfers, i.e., those corresponding to a 0-cost  $c$  in expression (2.2).

**Definition 3.1.** Let  $\mathcal{S}$  be a non-empty subset of  $\mathcal{P}(X) \times \mathcal{P}(Y)$ .

- (1) Recall that  $\mathcal{S}$  is a **gambling house** if for every  $x \in X$ , there exists  $\sigma \in \mathcal{P}(Y)$  such that  $(\delta_x, \sigma) \in \mathcal{S}$ , that is, if its characteristic function

$$\mathcal{T}_{\mathcal{S}}(\mu, \nu) := \begin{cases} 0 & \text{if } (\mu, \nu) \in \mathcal{S} \\ +\infty & \text{otherwise,} \end{cases}$$

is a standard functional.

- (2) Say that a subset  $\mathcal{S}$  of  $\mathcal{P}(X) \times \mathcal{P}(Y)$  is a **backward transfer set** if its characteristic function  $\mathcal{T}_{\mathcal{S}}$  is a backward linear transfer.

It is clear that transfer sets are convex weak\*-compact subsets of  $\mathcal{P}(X) \times \mathcal{P}(Y)$ .

**Proposition 3.2.** *A standard backward Kantorovich operator  $T$  is 1-positively homogenous if and only there exists a closed gambling house  $\mathcal{S}$  such that*

$$(3.1) \quad T^-g(x) = \sup \left\{ \int_X g \, d\sigma; (\delta_x, \sigma) \in \mathcal{S} \right\},$$

in which case

$$(3.2) \quad \overline{\text{conv}(\mathcal{S})} = \{ (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \nu \leq T_{\#}\mu \text{ on } C(Y) \},$$

where  $T_{\#}\mu(g) := \int_X Tg \, d\mu$  for every  $g \in C(Y)$ .

*Proof.* Note that if  $T^-$  is a 1-positively homogenous Kantorovich operator, then the corresponding backward linear transfer

$$(3.3) \quad \mathcal{T}(\mu, \nu) = \sup_{g \in C(Y)} \left\{ \int_Y g \, d\nu - \int_X T^-g \, d\mu \right\} \quad \text{for all } (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$$

can only take values in  $\{0, +\infty\}$ , since then

$$\mathcal{T}(\mu, \nu) \geq n \sup_{g \in C(Y)} \left\{ \int_Y g d\nu - \int_X T^- g d\mu \right\},$$

for every  $n$ , which means it is  $+\infty$  unless  $\int_Y g d\nu - \int_X T^- g d\mu \leq 0$  for every  $g \in C(Y)$ . □

In other words,

$$\mathcal{T}(\mu, \nu) = \begin{cases} 0 & \text{if } (\mu, \nu) \in \mathcal{D} \\ +\infty & \text{otherwise,} \end{cases}$$

where  $\mathcal{D} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \nu \leq T_{\#}\mu \text{ on } C(Y)\}$ . It is clear that if  $T^-$  is standard, then  $\mathcal{S} := \mathcal{D}$  is a closed and convex gambling house.

Conversely, if  $\mathcal{S}$  is a closed gambling house, then (3.1) defines a standard map that satisfies properties (1)–(4) of a backward Kantorovich operator. It remains to show that for a fixed  $g \in C(Y)$ ,  $T^-g$  is upper semi-continuous. For that assume  $x_n \rightarrow x$  in  $X$ . Assuming  $\limsup_{n \rightarrow \infty} T^-g(x_n) > -\infty$  (since otherwise there is nothing to prove), consider a subsequence such that  $\limsup_{n \rightarrow \infty} T^-g(x_n) = \lim_{j \rightarrow \infty} T^-g(x_{n_j})$ . Since  $\mathcal{S}$  is closed, the set  $\mathcal{S}_j := \{\sigma \in \mathcal{P}(Y); (x_{n_j}, \sigma) \in \mathcal{S}\}$  is compact in  $\mathcal{P}(Y)$ , hence there is  $\nu_j$  in  $\mathcal{S}_j$  such that

$$T^-g(x_{n_j}) = \sup \left\{ \int_X g d\sigma; (\delta_{x_{n_j}}, \sigma) \in \mathcal{S} \right\} = \int_X g d\nu_j.$$

By the weak\* compactness of  $\mathcal{P}(Y)$ , we may extract a further subsequence if necessary and assume  $\nu_j \rightarrow \bar{\nu}$  for some  $\bar{\nu} \in \mathcal{P}(Y)$ . Since  $\mathcal{S}$  is closed, we have  $(\delta_x, \bar{\nu}) \in \mathcal{S}$ . It follows that

$$\limsup_{n \rightarrow \infty} T^-g(x_n) = \lim_{j \rightarrow \infty} T^-g(x_{n_j}) = \int_Y g d\bar{\nu} \leq \sup_{\nu \in \mathcal{P}(Y)} \left\{ \int_Y g d\nu; (\delta_x, \nu) \in \mathcal{S} \right\} = T^-g(x),$$

hence  $T^-g \in USC(X)$ .

It is clear that  $\mathcal{S}$ , hence  $\overline{\text{conv}(\mathcal{S})}$  is contained in  $\mathcal{D} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \nu \leq T_{\#}\mu \text{ on } C(Y)\}$ . Conversely, by the Hahn-Banach theorem, every  $(\delta_x, \nu)$  in  $\mathcal{D}$  belongs to  $\mathcal{S}$ , hence  $\overline{\text{conv}(\mathcal{S})} = \mathcal{D}$  since the extreme points of the latter are of the form  $(\delta_x, \nu)$  for some  $x \in X$  and  $\nu \in \mathcal{P}(Y)$ .

**Theorem 3.3.** *Let  $\mathcal{S}$  be a non-empty subset of  $\mathcal{P}(X) \times \mathcal{P}(Y)$ . The following are then equivalent:*

- (1)  $\mathcal{S}$  is a convex weak\*-closed gambling house.
- (2)  $\mathcal{S}$  is a standard backward transfer set.
- (3)  $\mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \nu \leq T_{\#}\mu\}$ , where  $T : C(Y) \rightarrow USC_f(X)$  is a positively 1-homogenous Kantorovich operator and  $T_{\#}\mu(g) := \int_X Tg d\mu$  for every  $g \in C(Y)$ .
- (4)  $\mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \mu \prec_{\mathcal{A}} \nu\}$ , where  $\mathcal{A}$  is a balayage cone on  $X \sqcup Y$ .
- (5)  $\mathcal{S}$  is convex weak\*-compact in  $\mathcal{P}(X) \times \mathcal{P}(Y)$  and  $(\mu, \nu) \in \mathcal{S}$  if and only if there exists  $\pi \in \mathcal{K}(\mu, \nu)$  such that  $(\delta_x, \pi_x) \in \mathcal{S}$  for  $\mu$ -almost  $x \in X$ , where  $(\pi_x)_x$  is a disintegration of  $\pi$  with respect to  $\mu$ .

*Proof.* Note that 1), 2) and 3) follow from the above proposition, while the restricted Strassen theorem mentioned in the introduction yields that 4) implies 5). In order to show that 5) implies 2), we compute the Legendre transform of  $\mathcal{B}_\mu$  where  $\mathcal{B}$  is the characteristic function of  $\mathcal{S}$  and show that it is a backward linear transfer with Kantorovich operator

$$(3.4) \quad T^-g(x) := \sup_{(\delta_x, \sigma) \in \mathcal{S}} \int_Y g(y) d\sigma(y).$$

Indeed,  $\mathcal{B}_\mu^*(g) = \sup \{ \int_Y g d\nu ; \nu \in \mathcal{P}(X), (\mu, \nu) \in \mathcal{S} \}$  while Strassen's theorem gives that for any  $(\mu, \nu) \in \mathcal{S}$ , there is  $(x, \pi_x) \in \mathcal{S}$  such that

$$\int_Y g(y) d\nu(y) = \int_Y \left[ \int_X g(y) d\pi_x(y) \right] d\mu(x) \leq \int_X \left( \sup_{(\delta_x, \sigma) \in \mathcal{S}} \int_Y g(y) d\sigma(y) \right) d\mu(x).$$

hence  $\mathcal{B}_\mu^*(g) \leq \int_X T^-g(x) d\mu$ .

On the other hand, for each  $x \in X$ , the supremum of  $\sigma \rightarrow \int_X g(y) d\sigma(y)$  is achieved on the set  $\{ \sigma \in \mathcal{P}(X); (\delta_x, \sigma) \in \mathcal{S} \}$  since the latter is weak\* closed in  $\mathcal{P}(X)$  and therefore is weak\* compact. By a standard selection theorem, there is a measurable selection  $x \rightarrow \sigma_x$ , where for each  $x$ ,  $\sigma_x$  is where the maximum is achieved. Note that since  $\delta_x \prec_A \sigma_x$ , we have that  $\mu \prec_A \tilde{\nu}$ , where  $\tilde{\nu}(A) := \int_X \sigma_x(A) d\mu(x)$ . It follows that

$$\int_X T^-g(x) d\mu \leq \int_X \left( \sup_{(\delta_x, \sigma) \in \mathcal{S}} \int_X g(y) d\sigma(y) \right) d\mu(x) = \int_X g(y) d\tilde{\nu}(y) \leq \mathcal{B}_\mu^*(g).$$

Note that the above also show that  $D_1(\mathcal{B}) = \mathcal{P}(X)$ , that is  $\mathcal{S}$  is a gambling house, where

$$D_1(\mathcal{T}) = \{ \mu \in \mathcal{P}(X); \exists \nu \in \mathcal{P}(Y), (\mu, \nu) \in D(\mathcal{T}) \}.$$

It remains to define a suitable balayage cone. For that, we establish the following general result. □

**Theorem 3.4.** *Let  $\mathcal{T}$  be a standard backward linear transfer on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  with backward Kantorovich operator  $T^-$ , then  $D(\mathcal{T}) \subset D(\mathcal{T}_r)$ , where  $\mathcal{T}_r$  is a backward linear transfer on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  whose Kantorovich operator is given by the positively 1-homogenous recession operator associated with  $T^-$ , i.e.,*

$$(3.5) \quad T_r^-g(x) := \lim_{\lambda \rightarrow +\infty} \frac{T^-(\lambda g)(x)}{\lambda}.$$

Moreover, there is a balayage cone  $\mathcal{A}$  on  $X \sqcup Y$  such that the following are equivalent:

- (1)  $(\mu, \nu) \in D(\mathcal{T}_r)$ .
- (2)  $\mu \prec_{\mathcal{A}} \nu$ .
- (3)  $\int_X T_r^-g d\mu \geq \int_Y g d\nu$  for every  $g \in C(Y)$ .

Furthermore,  $D(\mathcal{T}_r)$  is the smallest transfer set containing  $D(\mathcal{T})$ , hence  $D(\mathcal{T})$  is a transfer set if and only if  $D(\mathcal{T}_r) = D(\mathcal{T})$ .

*Proof.* Without loss of generality, we may assume  $\mathcal{T} \geq 0$  (otherwise consider  $\mathcal{T} - C$  where  $C \in \mathbb{R}$  is a lower bound for  $\mathcal{T}$ ). Let

$$(3.6) \quad \mathcal{T}_{\text{finite}}(\mu, \nu) := \begin{cases} 0 & \text{if } (\mu, \nu) \in D(\mathcal{T}), \\ +\infty & \text{otherwise.} \end{cases}$$

It is immediate that  $\mathcal{T}_{\text{finite}}$  is a proper, bounded below, convex, and weak\* lower semi-continuous function with  $D_1(\mathcal{T}_{\text{finite}}) = D_1(\mathcal{T})$ . We have for  $\mu \in D_1(\mathcal{T}_{\text{finite}})$ ,

$$(3.7) \quad (\mathcal{T}_{\text{finite}})_\mu^*(g) = \sup \left\{ \int_Y g d\nu ; \nu \in \mathcal{P}(Y), \mathcal{T}(\mu, \nu) < +\infty \right\}.$$

Since for  $\lambda > 0$ , we have

$$\int_X \frac{T^-(\lambda g)}{\lambda} d\mu = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \frac{1}{\lambda} \mathcal{T}(\mu, \sigma) \right\},$$

it follows that  $\liminf_{\lambda \rightarrow +\infty} \int_X \frac{T^-(\lambda g)}{\lambda} d\mu \geq \sup \{ \int_Y g d\sigma ; \mathcal{T}(\mu, \sigma) < +\infty \} = (\mathcal{T}_{\text{finite}})_\mu^*(g)$ . On the other hand, since  $\mathcal{T}$  is non-negative, we have

$$\limsup_{\lambda \rightarrow +\infty} \int_X \frac{T^-(\lambda g)}{\lambda} d\mu \leq \sup \left\{ \int_Y g d\sigma ; \sigma \in \mathcal{P}(Y), \mathcal{T}(\mu, \sigma) < +\infty \right\} = (\mathcal{T}_{\text{finite}})_\mu^*(g),$$

so we conclude that for  $\mu \in D_1(\mathcal{T}_{\text{finite}})$ ,

$$(3.8) \quad (\mathcal{T}_{\text{finite}})_\mu^*(g) = \lim_{\lambda \rightarrow +\infty} \int_X \frac{T^-(\lambda g)}{\lambda} d\mu.$$

To show that  $(\mathcal{T}_{\text{finite}})_\mu^*(g) = \int_X T_r^- g d\mu$ , it remains to justify that

$$\lim_{\lambda \rightarrow +\infty} \int_X \frac{T^-(\lambda g)}{\lambda} d\mu = \int_X \lim_{\lambda \rightarrow +\infty} \frac{T^-(\lambda g)}{\lambda} d\mu.$$

This is simply by monotone convergence, since as  $\mathcal{T}$  is non-negative, the function  $\lambda \rightarrow \frac{T^-(\lambda g)(x)}{\lambda}$  is monotone increasing, i.e., if  $\lambda_2 \geq \lambda_1$ ,

$$\begin{aligned} \frac{T^-(\lambda_2 g)(x)}{\lambda_2} &= \sup_{\nu \in \mathcal{P}(Y)} \left\{ \int_Y g d\nu - \frac{1}{\lambda_2} \mathcal{T}(\delta_x, \nu) \right\} \\ &\geq \sup_{\nu \in \mathcal{P}(Y)} \left\{ \int_Y g d\nu - \frac{1}{\lambda_1} \mathcal{T}(\delta_x, \nu) \right\} \\ &= \frac{T^-(\lambda_1 g)(x)}{\lambda_1}. \end{aligned}$$

Note now that  $T_r^-$  is a positively 1-homogenous Kantorovich operator, hence by Proposition 2.2, the functional defined on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  by

$$\mathcal{T}_r(\mu, \nu) = \sup \left\{ \int_Y g d\mu - \int_X T_r^- g d\nu ; g \in C(Y) \right\}$$

and  $+\infty$  outside  $\mathcal{P}(X) \times \mathcal{P}(Y)$ , is a backward linear transfer that can only take values 0 or  $+\infty$ . Moreover,  $D(\mathcal{T}) \subset D(\mathcal{T}_r)$  since if  $\mathcal{T}(\mu, \nu) < +\infty$ , then  $\mu \in D_1(\mathcal{T}_{\text{finite}})$  and

$$0 = \mathcal{T}_{\text{finite}}(\mu, \nu) = (\mathcal{T}_{\text{finite}})_\mu^{**}(\nu) = \sup_{g \in C(Y)} \left\{ \int_Y g d\nu - \int_X T_r^- g d\mu \right\},$$

hence for all  $g \in C(Y)$ , we have  $\int_X T_r^- g, d\mu \geq \int_Y g d\nu$  and therefore  $\mathcal{T}_r(\mu, \nu) \leq 0$ .

Consider now the cone

$$(3.9) \quad \mathcal{A} := \{\phi \in LSC(X \sqcup Y); T_r^-(-\phi_Y) \leq -\phi_X\} = \{\phi \in LSC(X \sqcup Y); T_r^+(\phi_Y) \geq \phi_X\},$$

where  $\phi_X$  (resp.,  $\phi_Y$ ) is the restriction of  $\phi$  to the component  $X$  (resp.,  $Y$ ) and  $T_r^+ g = -T_r^-(-g)$ . Then  $\mathcal{A}$  is a balayage cone and the following holds:

$$(\mu, \nu) \in D(\mathcal{T}_r) \text{ if and only if } \mu \preceq_{\mathcal{A}} \nu.$$

Indeed, the convexity, lower semi-continuity and positive homogeneity of  $T_r^-$  clearly yield that  $\mathcal{A}$  is a closed convex cone of  $LSC(X \sqcup Y)$  containing the constants. The fact that  $T_r^-$  is monotone increasing yields that if  $\phi^1, \phi^2$  are in  $\mathcal{A}$ , then

$$\begin{aligned} T_r^-(-(\phi_Y^1 \vee \phi_Y^2)) &= T_r^-((-\phi_Y^1) \wedge (-\phi_Y^2)) \leq T_r^-(-\phi_Y^1) \wedge T_r^-(-\phi_Y^2) \\ &\leq (-\phi_X^1) \wedge (-\phi_X^2) = -(\phi_X^1 \vee \phi_X^2). \end{aligned}$$

It follows that  $\mathcal{A}$  is a balayage cone.

Suppose now that  $(\mu, \nu) \in D(\mathcal{T}_r)$ , then

$$(3.10) \quad 0 = \mathcal{T}_r(\mu, \nu) = \sup_g \left\{ \int_Y g d\nu - \int_X T_r^-(g) d\mu \right\},$$

hence  $\int_X T_r^-(g) d\mu \geq \int_Y g d\nu$  for all  $g \in C(Y)$ . Take now any  $\phi \in \mathcal{A}$ , then by assumption,

$$(3.11) \quad \int_X -\phi_X d\mu \geq \int_X T_r^-(-\phi_Y) d\mu \geq \int_Y (-\phi_Y) d\nu,$$

which translates into  $\int_X \phi_X d\mu \leq \int_Y \phi_Y d\nu$  for every  $\phi \in \mathcal{A}$ , hence  $\mu \prec_{\mathcal{A}} \nu$ .

Conversely, suppose  $\mu \preceq_{\mathcal{A}} \nu$ . For  $g \in C(Y)$ , consider the function  $\phi \in LSC(X \sqcup Y)$  defined by  $\phi_X := -T_r^-g$  and  $\phi_Y := -g$ . Then  $\phi \in \mathcal{A}$ , so that

$$(3.12) \quad -\int_X T_r^-g d\mu = \int_X \phi_X d\mu \leq \int_Y \phi_Y d\nu = -\int_Y g d\nu,$$

hence  $\int_X T_r^-g d\mu \geq \int_Y g d\nu$ , which implies that  $\mathcal{T}_r(\mu, \nu) = 0$  and so  $(\mu, \nu) \in D(\mathcal{T}_r)$ . It follows that  $D(\mathcal{T}_r)$  is a balayage, hence a transfer set.

Suppose now that  $\mathcal{S}$  is a transfer set containing  $D(\mathcal{T})$ , then its Kantorovich operator is given by

$$T_{\mathcal{S}}f(x) = \sup \left\{ \int_Y g d\sigma; (\delta_x, \sigma) \in \mathcal{S} \right\} \geq \sup \left\{ \int_Y g d\sigma; (\delta_x, \sigma) \in D(\mathcal{T}) \right\} = T_r^-f(x).$$

If  $(\mu, \nu) \in D(\mathcal{T}_r)$ , then

$$0 = \sup \left\{ \int_Y g d\mu - \int_X T_r^-g d\nu; g \in C(Y) \right\} \geq \sup \left\{ \int_Y g d\mu - \int_X T_{\mathcal{S}}g d\nu; g \in C(Y) \right\},$$

which means that  $(\mu, \nu) \in \mathcal{S}$  and therefore  $D(\mathcal{T}) \subset D(\mathcal{T}_r) \subset \mathcal{S}$ . □

Now we explore when the domain  $D(\mathcal{T})$  of a linear transfer is a transfer set.

**Proposition 3.5.** *Let  $\mathcal{T}$  be a standard backward linear transfer,  $T^-$  the corresponding Kantorovich operator, and  $T_r^-$  the associated recession operator. Then the following properties are equivalent:*

- (1)  $D(\mathcal{T})$  is a transfer set.
- (2)  $D(\mathcal{T}) = D(\mathcal{T}_r) := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \int_Y g d\nu \leq \int_X T_r^- g d\mu, \forall g \in C(Y)\}$ .
- (3)  $D(\mathcal{T})$  is weak\*-closed in  $\mathcal{P}(X) \times \mathcal{P}(Y)$ .
- (4)  $D_1(\mathcal{T})$  is weak\*-closed in  $\mathcal{P}(X)$ .
- (5)  $D_1(\mathcal{T}) = \mathcal{P}(X)$ .

*Proof.* That (1) and (2) are equivalent follows from the preceding proposition. 2) yields 3) since  $T_r^- g$  is upper semi-continuous for every  $g \in C(Y)$ . 3) implies 4) is immediate. For 5) note that  $\mathcal{E} := \{\delta_x; x \in X\}$  is the set of extreme points of  $\mathcal{P}(X)$  and  $\mathcal{E} \subset D_1(\mathcal{T})$  since  $\mathcal{T}$  is standard. The rest follows from Krein-Milman's theorem and the fact that  $D_1(\mathcal{T})$  is closed and convex.

If now (5) holds, then by the above proposition, we have  $(\mathcal{T}_{\text{finite}})_\mu^*(g) = \int_X T_r^- g d\mu$ , for any  $\mu \in D_1(\mathcal{T}_{\text{finite}}) = \mathcal{P}(X)$ . It follows that  $D(\mathcal{T}) = D(\mathcal{T}_r)$ , hence it is a transfer set. □

**Remark 3.6.** Note that in general  $D(\mathcal{T}) \neq D(\mathcal{T}_r)$  and  $D(\mathcal{T})$  is not necessarily a transfer set, since we cannot say much about  $(\mathcal{T}_{\text{inf}})_\mu^*$  when  $\mu \notin D_1(\mathcal{T})$ . However, suppose for every  $\mu \in D_1(\mathcal{T})$ , there exists  $\alpha \in L^1(\mu)$  such that

$$(3.13) \quad \mathcal{T}(x, \sigma) \leq \alpha(x) \text{ for all } \sigma \in \mathcal{P}(Y) \text{ such that } \mathcal{T}(x, \sigma) < +\infty,$$

then  $D(\mathcal{T}) = D(\mathcal{T}_r)$  and  $D(\mathcal{T})$  is a transfer set. Indeed, in this case

$$\begin{aligned} T^- g(x) &= \sup\left\{ \int_Y g d\sigma - \mathcal{T}(x, \sigma); \sigma \in \mathcal{P}(Y) \right\} \\ &\geq \sup\left\{ \int_Y g d\sigma; \mathcal{T}(x, \sigma) < +\infty \right\} - \alpha(x) \\ &\geq T_r^- g(x) - \alpha(x), \end{aligned}$$

from which follows that if  $\mathcal{T}_r(\mu, \nu) < +\infty$ , then  $\mathcal{T}(\mu, \nu) \leq \mathcal{T}_r(\mu, \nu) + \int_X \alpha(x) d\mu < +\infty$ .

Conversely, if  $D(\mathcal{T})$  is a transfer set then we shall see that in this case

$$(3.14) \quad D(\mathcal{T}) = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \exists \pi \in \mathcal{K}(\mu, \nu), \mathcal{T}(\delta_x, \pi_x) < +\infty \mu\text{-a.s}\}$$

$$(3.15) \quad = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \exists \pi \in \mathcal{K}(\mu, \nu), \int_X \mathcal{T}(\delta_x, \pi_x) d\mu(x) < +\infty\},$$

which indicates a type of uniform boundedness on  $\mathcal{T}$  that is slightly weaker than (3.13).

#### 4. TRANSFER SETS AND TRUE BALAYAGE OF PROBABILITY MEASURES

We now consider the case where we can have a true balayage.

**Theorem 4.1.** *Let  $\mathcal{S}$  be a standard backward transfer subset of  $\mathcal{P}(X) \times \mathcal{P}(X)$  and let  $T^-$  be its associated Kantorovich operator. Then*

- (1)  $\mathcal{S}$  is contained in a true balayage set  $\hat{\mathcal{S}} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\bar{c}} \nu\}$ , where  $\mathcal{A}$  is the balayage cone of lower semi-continuous  $T^+$ -superharmonic functions on  $X$ .

The Kantorovich operator associated to  $\hat{\mathcal{S}}$ , is given for every  $f \in C(X)$  and every  $x \in X$  by the expression,

$$(4.1) \quad \hat{T}f(x) := \inf\{\varphi(x); -\varphi \in \mathcal{A} \text{ and } \varphi \geq f \text{ on } X\}.$$

- (2)  $\hat{\mathcal{S}}$  is minimal in the sense that if  $\mathcal{R}$  is any other true balayage set containing  $\mathcal{S}$ , then  $\mathcal{R} \supset \hat{\mathcal{S}}$ .

*Proof.* Since  $\mathcal{S} \subset \mathcal{P}(X) \times \mathcal{P}(X)$  is a backward transfer set, we let  $\mathcal{T}$  be its characteristic function in such a way that  $\mathcal{S} = D(\mathcal{T})$ , with backward Kantorovich operator

$$T^-g(x) := \sup_{(\delta_x, \sigma) \in \mathcal{S}} \int_X g(y) d\sigma(y),$$

which is positively 1-homogenous. According to Theorem 3.4, there is also a cone  $\mathcal{C} \subset LSC(X \sqcup X)$  such that

$$\mathcal{S} = D(\mathcal{T}) = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{C}} \nu\},$$

and that  $\mu \prec_{\mathcal{C}} \nu$  if and only if  $\int_X T^-g d\mu \geq \int_X g d\nu$  for every  $g \in C(X)$ . Consider the cone

$$\bar{\mathcal{C}} = \{f \in LSC(X); T(-f) \leq -f\} = \{f \in LSC(X); T^+(f) \geq f\}$$

and the set  $\bar{\mathcal{S}} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\bar{\mathcal{C}}} \nu\}$ .

Note that  $\bar{\mathcal{S}} \supset \mathcal{S}$ , since if  $\int_X T^-g d\mu \geq \int_X g d\nu$  for every  $g \in C(X)$ , then for any  $f \in \bar{\mathcal{C}}$ , we have

$$-\int_X f d\mu \geq \int_X T(-f) d\mu \geq \int_X -f d\nu.$$

Moreover,  $\bar{\mathcal{C}}$  is a balayage cone on  $X$  and therefore  $\bar{\mathcal{S}}$  is a transfer set containing the diagonal  $\{(\mu, \mu); \mu \in \mathcal{P}(X)\}$ . Let now

$$\bar{T}^-f(x) = \sup_{(\delta_x, \sigma) \in \bar{\mathcal{S}}} \int_X f(y) d\sigma(y)$$

be the Kantorovich operator corresponding to  $\bar{\mathcal{S}}$ , and consider the cone

$$(4.2) \quad \mathcal{A} = \{g \in LSC(X); g(x) \leq \int_X g d\nu \text{ for all } (\delta_x, \nu) \in \bar{\mathcal{S}}\}.$$

It is clear that  $\mathcal{A}$  is a closed convex balayage cone. We claim that for any  $f \in C(X)$ ,

$$(4.3) \quad f(x) \leq \bar{T}^-f(x) \leq \hat{f}(x) := \inf\{\varphi(x); -\varphi \in \mathcal{A} \text{ and } \varphi \geq f \text{ on } X\}.$$

Indeed, since  $(\delta_x, \delta_x) \in \bar{\mathcal{S}}$ , we have that  $\bar{T}^-f(x) \geq f(x)$ . Moreover, for any  $\phi \in -\mathcal{A}$ ,  $\phi \geq f$  on  $X$ , we have

$$\bar{T}^-f(x) = \sup_{(\delta_x, \sigma) \in \bar{\mathcal{S}}} \int_X f(y) d\sigma(y) \leq \sup_{(\delta_x, \sigma) \in \bar{\mathcal{S}}} \int_X \phi(y) d\sigma(y) \leq \phi(x),$$

the last inequality holding since  $\phi \in -\mathcal{A}$ . Hence (4.3) is verified.

Now the mapping  $f \rightarrow \hat{f}$  is a backward Kantorovich operator, hence the functional

$$\hat{\mathcal{T}}(\mu, \nu) := \sup_{f \in C(X)} \left\{ \int_X f d\nu - \int_X \hat{f} d\mu \right\}$$



is a backward linear transfer, and since  $f \rightarrow \hat{f}$  is positively 1-homogenous,  $\hat{\mathcal{T}}$  is necessarily  $\{0, +\infty\}$ -valued. We let  $\hat{\mathcal{S}} := D(\hat{\mathcal{T}})$  and claim that

$$(4.4) \quad \hat{\mathcal{S}} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{A}} \nu\}.$$

Indeed, if  $(\mu, \nu) \in \hat{\mathcal{S}}$ , then  $\int_X \hat{f} d\mu \geq \int_X f d\nu$  for every  $f \in C(X)$ , hence for every  $\psi \in \mathcal{A}$ , we have  $\int_X -\psi d\mu = \int_X \widehat{(-\psi)} d\mu \geq \int_X -\psi d\nu$ , that is  $\int_X \psi d\mu \leq \int_X \psi d\nu$ , hence  $\mu \prec_{\mathcal{A}} \nu$ .

For the reverse implication assume  $(\mu, \nu) \notin \hat{\mathcal{S}}$ , then

$$\sup_{f \in C(X)} \left\{ \int_X f d\nu - \int_X \hat{f} d\mu \right\} = +\infty,$$

hence there is  $f \in C(X)$  such that

$$(4.5) \quad \int_X \hat{f} d\nu \geq \int_X f d\nu \geq \int_X \hat{f} d\mu + 1.$$

Note now that for every  $\epsilon > 0$ , and for each  $x \in X$ , we may choose  $\phi_{\epsilon, x} \in -\mathcal{A}$ ,  $\phi_{\epsilon, x} \geq f$ , such that  $\phi_{\epsilon, x}(x) \leq \hat{f}(x) + \frac{\epsilon}{2}$ . By continuity, there exists an open neighbourhood  $B_{r_x}(x) \subset X$  such that  $\phi_{\epsilon, x}(x') \leq f(x') + \epsilon$  for all  $x' \in B_{r_x}(x)$ . The collection  $\{B_{r_x}(x)\}_{x \in X}$  is an open cover of  $X$ , hence by compactness, there exists a finite subcover  $\{B_{r_{x_i}}(x_i)\}_{i=1}^n$  of  $X$ . Define

$$\phi_{\epsilon}(x) := \min\{\phi_{\epsilon, x_1}(x), \dots, \phi_{\epsilon, x_n}(x)\}$$

It follows that  $\phi_{\epsilon} \in -\mathcal{A}$  since  $\mathcal{A}$  is closed under maxima. Moreover, each  $x \in X$  belongs to  $B_{r_{x_i}}(x_i)$  for some  $i \in \{1, \dots, n\}$ , hence

$$\hat{f}(x) \leq \phi_{\epsilon}(x) \leq \hat{f}(x) + \epsilon \quad \text{for all } x \in X.$$

Combining this with (4.5), we get

$$\int_X \phi_{\epsilon} d\nu \geq \int_X \hat{f} d\nu \geq \int_X \hat{f} d\mu + 1 \geq \int_X \phi_{\epsilon} d\mu - \epsilon + 1.$$

In other words, there is a function  $\psi \in \mathcal{A}$ , namely  $\psi = -\phi_{\epsilon}$  such that  $\int_X \psi d\mu > \int_X \psi d\nu$ , which contradicts the fact that  $\mu \prec_{\mathcal{A}} \nu$ .

We now show that

$$(4.6) \quad \bar{\mathcal{S}} = \hat{\mathcal{S}}, \mathcal{A} = \bar{\mathcal{C}}, \text{ and } \bar{T}^- f = \hat{f} \text{ for all } f \in C(X).$$

Indeed,  $\bar{\mathcal{S}} \subset \hat{\mathcal{S}}$ , since  $\bar{T}^- f \leq \hat{f}$ , hence  $\hat{\mathcal{T}} \leq \bar{\mathcal{T}}$ ,  $\bar{\mathcal{S}} = D(\bar{\mathcal{T}}) \subset D(\hat{\mathcal{T}}) = \hat{\mathcal{S}}$  and  $\bar{T} f \leq \hat{f}$ . On the other hand, if  $\phi \in \bar{\mathcal{C}}$ , then  $\phi(x) \leq \int_X \phi d\mu$  for every  $(\delta_x, \mu) \in \bar{\mathcal{S}}$ , hence  $\bar{\mathcal{C}} \subset \mathcal{A}$ , from which follows that  $\bar{T}^- f \geq \hat{f}$  and  $\hat{\mathcal{S}} \subset \bar{\mathcal{S}}$ . Therefore,  $\mathcal{A} = \bar{\mathcal{C}}$ , the latter being the cone of  $T^+$ -superharmonic functions, and (4.6) is established.

(2) To show that  $\bar{\mathcal{S}}$  is minimal, assume  $\mathcal{B}$  is a balayage cone in  $C(X)$  such that

$$\mathcal{R} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{B}} \nu\},$$

and  $\mathcal{S} \subset \mathcal{R}$ . For every  $\phi \in \mathcal{B}$ , we then have  $\phi(x) \leq \int_X \phi d\mu$  for every  $(\delta_x, \mu) \in \mathcal{S} \subset \bar{\mathcal{S}}$ , hence  $\mathcal{B} \subset \mathcal{A}$ , from which follows that  $\hat{\mathcal{S}} = \bar{\mathcal{S}} \subset \mathcal{R}$ . □

**Remark 4.2.** There are actually several other transfer sets between  $\mathcal{S}$  and  $\hat{\mathcal{S}}$ . Indeed, let

$$T_1 f := T^- f \vee f, \quad T_2 f := \overline{\lim_n \uparrow T_1^n f}, \quad \text{and } T_3 f := \lim_n \downarrow T_1^n \hat{f}.$$

They are all positively 1-homogenous Kantorovich operators such that

$$(4.7) \quad T^- f \leq T_1 f \leq T_2 f \leq T_3 f \leq \bar{T} f = \hat{f}.$$

Indeed, the relation between  $T, T_1$  and  $T_2$  is immediate. Also, for any  $f \in C(X)$ , we have

$$T^- f(x) = \sup_{(\delta_x, \sigma) \in \mathcal{S}} \int_X f(y) d\sigma(y) \leq \sup_{(\delta_x, \sigma) \in \bar{\mathcal{S}}} \int_X f(y) d\sigma(y) = \bar{T}^- f(x) = \hat{f}(x),$$

hence  $T_1 f \leq \hat{f}$ , and therefore  $T_1 \hat{f} \leq \hat{f}$ ,  $T_1^n f \leq \hat{f}$ , and  $T_1^{n+1} \hat{f} \leq T_1^n \hat{f} \leq \hat{f}$ , since  $\bar{T} f = \hat{f}$  is idempotent. It follows that  $T_2$  and  $T_3$  are well defined, are Kantorovich operators, and are both below  $\bar{T} f = \hat{f}$ . Finally, note that  $T_1^n f \leq T_1^n \hat{f}$ , hence the inequalities in (4.7) hold.

For  $i = 1, 2, 3$ , let

$$\mathcal{S}_i = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \nu \leq T_i \# \mu\},$$

where  $T_i \# \mu(f) = \int_X T_i f d\mu$ . It is clear that they are all transfer sets with

$$\mathcal{S} \subset \mathcal{S}_1 \subset \mathcal{S}_2 \subset \mathcal{S}_3 \subset \hat{\mathcal{S}},$$

where  $\mathcal{S}_1$  amounts to adding the diagonal to  $\mathcal{S}$ , while  $\hat{\mathcal{S}} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \nu \leq \hat{\mu}\}$ , with the familiar sublinear functional  $\hat{\mu}(f) = \mu(\hat{f})$  from Choquet theory.

Finally, we note why the process does not necessarily stop at  $T_2$ . Indeed, by setting  $U_\infty f := \lim_n \uparrow T_1^n f$ , we can use that  $T_1$  is a functional capacity to deduce that  $T_1(U_\infty f) = U_\infty f$  and  $U_\infty \circ U_\infty f = U_\infty f$ . It is clear that  $U_\infty f \leq \hat{f}$ . On the other hand, since  $U_\infty f = U(U_\infty f) = T^-(U_\infty f) \vee U_\infty f$ , it follows that  $T^-(U_\infty f) \leq U_\infty f$ .

The process would stop if  $U_\infty f$  were upper semi-continuous, since then  $U_\infty f \in -\bar{\mathcal{C}}$ , and since  $U_\infty f \geq f$ , it would follow that  $U_\infty f \geq \hat{f}$  and therefore  $U_\infty f = \hat{f}$ . This happens for instance if  $T^-$  satisfies for any  $f \in C(X)$ ,  $x, y \in X$ ,

$$(4.8) \quad T^- f(x) - T^- f(y) \leq \omega(d(x, y)),$$

where  $d$  is the metric on  $X$  and  $\omega$  is the modulus of uniform continuity of  $f$ . This occurs for example in the case where  $X$  is the closure of an open bounded domain  $O$  in  $\mathbb{C}^n$  (resp.,  $\mathbb{R}^n$ ), and the Kantorovich operators are of the form

$$T^- f(x) := \sup_{v \in \mathbb{R}^n} \left\{ \int_0^{2\pi} f(x + e^{i\theta} v) \frac{d\theta}{2\pi}; x + \bar{\Delta} v \subset O \right\},$$

(resp.,

$$T^- f(x) = \sup_{r \geq 0} \left\{ \int_B f(x + ry) dm(y); x + r\bar{B} \subset O \right\},$$

where  $\Delta = \{z \in \mathbb{C}, |z| < 1\}$  is the open unit disc in  $\mathbb{C}$ ,  $B$  is the open unit ball in  $\mathbb{R}^n$  centered at 0, and  $m$  is normalized Lebesgue measure on  $\mathbb{R}^n$ . Condition (4.8) is verified in either case, and we have  $\lim_n \uparrow T_1^n f = \hat{f}$ , where in the first case the

envelope  $\hat{f}$  corresponds to the cone of plurisubharmonic functions on  $O$ , while in the second it corresponds to the cone of subharmonic functions on  $O$ . In other words,  $T = T_1$  and  $T_2 = T_3 = \bar{T}$ . However, it is not always the case and we had to add the upper semi-continuous regularization to be able to define  $T_2 f = \overline{U_\infty f}$ .

**Corollary 4.3.** *Let  $\mathcal{S}$  be a subset of  $\mathcal{P}(X) \times \mathcal{P}(X)$ . Then the following are equivalent:*

- (1)  $\mathcal{S}$  is a transitive backward transfer set containing the diagonal  $\{(\mu, \mu); \mu \in \mathcal{P}(X)\}$ .
- (2)  $\mathcal{S}$  is a backward transfer set with a Kantorovich operator  $T$  satisfying  $T^2 = T \geq I$ .
- (3)  $\mathcal{S} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{A}} \nu\}$ , where  $\mathcal{A}$  is a balayage cone in  $LSC(X)$ .

*Proof.* That 3) implies 1) follows from Proposition 1.5 with the two facts that the diagonal is now contained in  $\mathcal{S}$  and that the latter is transitive, following readily from the true balayage relation with respect to a cone in  $LSC(X)$ .

To show that 1) implies 2), we consider for all  $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X)$ , the functional

$$\mathcal{T}_2(\mu, \nu) = \inf\{\mathcal{T}(\mu, \sigma) + \mathcal{T}(\sigma, \nu); \sigma \in \mathcal{P}(X)\}.$$

It is a backward linear transfer with Kantorovich operator  $T^2 = T \circ T$  by Proposition 2.11. Since  $\mathcal{T}$  is transitive, it is clear that  $D(\mathcal{T}_2) \subset D(\mathcal{T})$ . On the other hand, since  $\mathcal{T}_2(\mu, \nu) \leq \mathcal{T}(\mu, \mu) + \mathcal{T}(\mu, \nu)$  and  $D(\mathcal{T})$  contains the diagonal, we have  $D(\mathcal{T}) \subset D(\mathcal{T}_2)$ . Since both  $\mathcal{T}$  and  $\mathcal{T}_2$  are valued in  $\{0, +\infty\}$ , it follows that  $\mathcal{T} = \mathcal{T}_2$  and consequently,  $T = T^2$ .

To show that 2) implies 3), we use Theorem 4.1 to write  $\mathcal{S} \subset \hat{\mathcal{S}}$ , where  $\hat{\mathcal{S}} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{A}} \nu\}$ , and  $\mathcal{A} = \{g \in LSC(X); g(x) \leq \int_X g d\nu \text{ for all } (\delta_x, \nu) \in \mathcal{S}\}$ . Moreover,  $Tf(x) \leq \hat{f}$ . Since  $Tf(x) = T^2 f(x) = \sup_{(\delta_x, \sigma) \in \mathcal{S}} \int_X Tf(y) d\sigma(y)$ , it follows that  $Tf \in -\mathcal{A}$ , and since  $Tf \geq f$ , we conclude that  $Tf \geq \hat{f}$ . In other words,  $Tf = \hat{f}$  and  $\mathcal{S} = \hat{\mathcal{S}}$ . □

The following is now immediate.

**Corollary 4.4.** *Let  $T : C(Y) \rightarrow C(X)$  be a Markov operator (i.e.,  $T$  positive linear continuous and  $T1 = 1$ ), and consider the set*

$$\mathcal{S} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \nu = T^*(\mu)\},$$

where  $T^* : \mathcal{M}(X) \rightarrow \mathcal{M}(Y)$  is the adjoint operator. Then,

- (1)  $\mathcal{S}$  is a backward linear transfer, whose Kantorovich operator is  $T$  itself.
- (2)  $\mathcal{S}$  cannot be a true balayage set with respect to a cone  $\mathcal{A}$  in  $C(X)$ , unless  $X = Y$  and  $T$  is the identity.
- (3)  $\mathcal{S}$  is a restricted balayage set for the cone  $\mathcal{A} = \{(f, g) \in LSC(X \sqcup X); f \leq Tg\}$ .
- (4) The smallest true balayage set containing  $\mathcal{T}$  is the one corresponding to the cone of superharmonic functions, i.e.,  $\bar{\mathcal{C}} = \{f \in LSC(X); f \leq Tf\}$ .

**Corollary 4.5.** *Let  $\mathcal{T}$  be a standard backward linear transfer on  $\mathcal{P}(X) \times \mathcal{P}(X)$  and let  $T$  be the corresponding Kantorovich operator. Then,*

- (1) *There exists a true balayage cone  $\mathcal{A}$  in  $C(X)$  such that*

$$D(\mathcal{T}) \subset D(\mathcal{T}_r) \subset \widehat{D(\mathcal{T}_r)} = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{A}} \nu\},$$

where  $D(\mathcal{T}_r)$  is the domain of the recession operator  $T_r$  of  $T$ .

- (2)  *$\widehat{D(\mathcal{T}_r)}$  is the smallest true balayage set containing  $D(\mathcal{T})$ .*  
 (3) *If  $D(\mathcal{T})$  is a transitive transfer set containing the diagonal, then  $D(\mathcal{T}) = D(\mathcal{T}_r) = \widehat{D(\mathcal{T}_r)}$  and the true balayage with respect to  $\mathcal{A}$  is equivalent to the restricted balayage associated with  $T_r^-$ .*

*Proof.* By Theorem 3.4,  $D(\mathcal{T}_r)$  is the smallest transfer set containing  $D(\mathcal{T})$ , with the recession operator  $T_r^-$  being the corresponding Kantorovich operator for  $\mathcal{T}_r$ . The transfer set  $\widehat{D(\mathcal{T}_r)}$  is the balayage set associated to  $D(\mathcal{T}_r)$  by Theorem 4.1.

If  $D(\mathcal{T})$  is a transitive transfer set, then by the minimality properties of both  $D(\mathcal{T}_r)$  and  $\widehat{D(\mathcal{T}_r)}$ , we obtain that  $D(\mathcal{T}) = D(\mathcal{T}_r) = \widehat{D(\mathcal{T}_r)}$ . By the uniqueness of the Kantorovich operator associated to a transfer set, we have  $T_r^-g = \hat{g}$ .

Finally, consider the cone  $\mathcal{C} \subset LSC(X \sqcup X)$  such that  $D(\mathcal{T}_r) = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X); \mu \prec_{\mathcal{C}} \nu\}$ , and that  $\mu \prec_{\mathcal{C}} \nu$  if and only if  $\int_X T_r^-g d\mu \geq \int_X g d\nu$  for every  $g \in C(X)$ . We claim that  $\mu \prec_{\mathcal{A}} \nu$  if and only if  $\mu \prec_{\mathcal{C}} \nu$ , which means that the restricted balayage is equivalent to the true balayage. Indeed, if  $\mu \prec_{\mathcal{A}} \nu$ , i.e.  $\int_X \varphi d\nu \leq \int_X \varphi d\mu$  for all functions  $\varphi$  in  $-\mathcal{A}$ , then for all  $g \in C(X)$ ,

$$\int_X g d\nu \leq \int_X \hat{g} d\nu \leq \int_X \hat{g} d\mu = \int_X T_r^-g d\mu,$$

hence  $\mu \prec_{\mathcal{C}} \nu$ . On the other hand, if  $\mu \prec_{\mathcal{C}} \nu$  then  $\int_X \hat{g} d\mu = \int_X T_r^-g d\mu \geq \int_X g d\nu$  for every  $g \in C(X)$ , the inequality holds for all  $g \in -\mathcal{A}$  in which case  $g = \hat{g}$ . This implies  $\mu \prec_{\mathcal{A}} \nu$ . □

### 5. REPRESENTATION OF LINEAR TRANSFERS AS OPTIMAL BALAYAGE TRANSPORT WITH COST

In this section we show that every backward linear transfer can be represented as the minimal cost of an optimal balayage transport of measures, by combining our previous results with the following properties of optimal weak transports established in [20] and [1]. We shall sketch a proof of the latter fact using results in [4] since we can then exhibit another interesting representation of a linear transfer as a somewhat classical mass transport of measures but on the enlarged state space  $X \times \mathcal{P}(Y)$ .

**Theorem 5.1.** *Let  $c: X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  be a bounded below, lower semi-continuous functional such that for each  $x \in X$ ,  $\sigma \rightarrow c(x, \sigma)$  is proper and convex. Consider the value functional of an optimal weak transport, i.e, for a pair  $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ ,*

$$(5.1) \quad \mathcal{V}_c(\mu, \nu) := \inf_{\pi} \left\{ \int_X c(x, \pi_x) d\mu(x); \pi \in \mathcal{K}(\mu, \nu) \right\},$$

where  $(\pi_x)_x$  is the disintegration of  $\pi$  with respect to  $\mu$ , Then,

- (1)  $\mathcal{V}_c$  is a standard backward linear transfer with corresponding Kantorovich operator given by  $T^-g(x) := \sup \{ \int_Y g d\sigma - c(x, \sigma); \sigma \in \mathcal{P}(Y) \}$ .

- (2) Conversely, if  $\mathcal{T}$  is a standard backward linear transfer, then  $\mathcal{T} = \mathcal{V}_c$ , where  $\mathcal{V}_c$  is the optimal weak transport associated with the cost function  $c(x, \sigma) := \mathcal{T}(\delta_x, \sigma)$ .

Moreover, for any  $(\mu, \nu)$  in  $D(\mathcal{T})$ , there exists a transport  $\pi \in \mathcal{K}(\mu, \nu)$  such that  $(\delta_x, \pi_x) \in D(\mathcal{T})$  for  $\mu$ -almost  $x \in X$  and

$$(5.2) \quad \mathcal{T}(\mu, \nu) = \int_X \mathcal{T}(\delta_x, \pi_x) d\mu(x).$$

Furthermore,  $D(\mathcal{T})$  is a backward transfer set if and only if

$$(5.3) \quad D(\mathcal{T}) = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \exists \pi \in \mathcal{K}(\mu, \nu), \mathcal{T}(\delta_x, \pi_x) < +\infty \mu\text{-a.s.}\}.$$

**Lemma 5.2.** Under the above hypothesis on the cost  $c$ , we have  $\mathcal{V}_c(\mu, \nu) = \hat{\mathcal{V}}_c(\mu, \nu)$ , where

$$(5.4) \quad \hat{\mathcal{V}}(\mu, \nu) := \inf_{P \in \Lambda(\mu, \nu)} \int_{X \times \mathcal{P}(Y)} c(x, p) dP(x, p),$$

$$(5.5) \quad \Lambda(\mu, \nu) := \left\{ P \in \mathcal{P}(X \times \mathcal{P}(Y)) \mid \text{proj}_X P = \mu, b(\text{proj}_{\mathcal{P}(Y)}(P)) = \nu \right\},$$

and  $b$  is the barycentric map on probability measures.

*Proof.* Recall that for a probability measure  $P \in \mathcal{P}(\mathcal{P}(Y))$ , its barycenter  $b(P)$  is the unique probability in  $\mathcal{P}(Y)$  determined by

$$(5.6) \quad b(P)(g) = \int_{\mathcal{P}(Y)} p(g) dP(p) \quad \forall g \in C(Y).$$

For any  $\pi \in \mathcal{K}(\mu, \nu)$ , we let  $(\pi_x)_{x \in X}$  be its disintegration with respect to  $\mu$ , and consider the measurable map  $\kappa_\pi: X \rightarrow X \times \mathcal{P}(Y)$  defined by  $x \mapsto (x, \pi_x)$ . Let now  $J(\pi) := (\kappa_\pi)_\#(\mu)$ , which yields a map from  $\mathcal{P}(X \times Y)$  to  $\mathcal{P}(X \times \mathcal{P}(Y))$  such that for  $\pi \in \mathcal{K}(\mu, \nu)$ , we have  $J(\pi) \in \Lambda(\mu, \nu)$  and  $\int_X c(x, \pi_x) \mu(dx) = \int_{X \times \mathcal{P}(Y)} c(x, p) J(\pi)(dx, dp)$ , hence

$$\mathcal{V}_c(\mu, \nu) = \inf_{\pi \in \mathcal{K}(\mu, \nu)} \int_X c(x, \pi_x) \mu(dx) \geq \inf_{P \in \Lambda(\mu, \nu)} \int_{X \times \mathcal{P}(Y)} c(x, p) P(dx, dp) = \hat{\mathcal{V}}_c(\mu, \nu).$$

For the converse, associate to any probability measure  $P \in \mathcal{P}(X \times \mathcal{P}(Y))$ , its intensity, which is the unique measure  $I(P) \in \mathcal{P}(X \times Y)$  such that

$$(5.7) \quad \int_{X \times Y} f(x, y) I(P)(dx, dy) = \int_{X \times \mathcal{P}(Y)} \int_Y f(x, y) p(dy) P(dx, dp) \quad \forall f \in C(X \times Y).$$

Now letting  $P \in \Lambda(\mu, \nu)$ , we easily see that  $I(P) \in \mathcal{K}(\mu, \nu)$  and  $I(P)_x = \int_{\mathcal{P}(Y)} p P_x(dp)$  for  $\mu$ -a.e.  $x$ . Using convexity we get that

$$\begin{aligned} \int_{X \times \mathcal{P}(Y)} c(x, p) P(dx, dp) &= \int_X \int_{\mathcal{P}(Y)} c(x, p) P_x(dp) \mu(dx) \geq \int_X c(x, I(P)_x) \mu(dx) \\ &\geq \inf_{\pi \in \mathcal{K}(\mu, \nu)} \int_X c(x, \pi_x) \mu(dx). \end{aligned}$$

It follows that  $\mathcal{V}(\mu, \nu) = \hat{\mathcal{V}}(\mu, \nu)$ .  $\square$

**Lemma 5.3.** *Let  $c: X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  be as above, then*

- (1) *The map  $\pi \rightarrow \int_X c(x, \pi_x) d\mu(x)$  is weak\*-lower semi-continuous and convex on  $\mathcal{K}(\mu, \nu)$ .*
- (2) *For  $\mu \in \mathcal{P}(X)$ , the map  $\nu \rightarrow \mathcal{V}_c(\mu, \nu)$  is convex and weak\*-lower semi-continuous on  $\mathcal{P}(Y)$ .*

*Proof.* (1) To show lower semicontinuity of the map  $\pi \rightarrow \int_X c(x, \pi_x) d\mu(x)$ , let  $\pi^k \rightarrow \pi$  in  $\mathcal{K}(\mu, \nu)$  and set  $P^k = J(\pi^k)$ . We may assume that  $\liminf_k \int_X c(x, \pi_x^k) d\mu = \lim_k \int_X c(x, \pi_x^k) d\mu$  by selecting a subsequence. Since  $\mathcal{P}(X \times \mathcal{P}(Y))$  is compact, let  $P$  be an accumulation point of  $\{P^k\}_k$  and consider a subsequence converging to  $P$ . Observe that

$$\int_X c(x, \pi_x^k) d\mu = \int_{X \times \mathcal{P}(Y)} c(x, p) P^k(dx, dp).$$

Since  $P \rightarrow \int_{X \times \mathcal{P}(Y)} c(x, p) P(dx, dp)$  is lower semi-continuous, we have

$$\liminf_k \int_{X \times \mathcal{P}(Y)} c(x, p) P^k(dx, dp) \geq \int_{X \times \mathcal{P}(Y)} c(x, p) P(dx, dp).$$

The  $X$ -marginal of  $P$  equals the  $X$ -marginal of  $\pi$ , which is  $\mu$ . Letting  $P_x(dp)$  be a disintegration of  $P$  with respect to  $\mu$ , we get from the convexity of  $c(x, \cdot)$  that

$$\liminf_k \int_X c(x, \pi_x^k) d\mu \geq \int_X \int_{\mathcal{P}(Y)} c(x, p) P_x(dp) d\mu \geq \int_X c\left(x, \int_{\mathcal{P}(Y)} p(dy) P_x(dp)\right) d\mu.$$

Now we note that  $\pi_x(dy) = \int_{\mathcal{P}(Y)} p(dy) P_x(dp)$  for  $\mu$ -a.e.  $x$ . Indeed, if  $f \in C(X \times Y)$ , then

$$\int_{X \times Y} f(x, y) \pi^k(dx, dy) \rightarrow \int_{X \times Y} f(x, y) \pi(dx, dy).$$

Moreover, the function  $F(x, p) := \int_Y f(x, y) p(dy)$  is also continuous on  $X \times \mathcal{P}(Y)$ , hence  $\int F dP^k \rightarrow \int F dP$ . We then deduce

$$\int_{X \times Y} f(x, y) \pi(dx, dy) = \int_{X \times \mathcal{P}(Y)} F dP = \int_{X \times \mathcal{P}(Y)} \int_Y f(x, y) p(dy) P(dx, dp),$$

hence  $\pi_x(dy) = \int_{\mathcal{P}(Y)} p(dy) P_x(dp)$  for  $\mu$ -almost every  $x$ . We finally obtain

$$\liminf_k \int_X c(x, \pi_x^k) d\mu \geq \int_X c(x, \pi_x) d\mu.$$

The convexity follows from the convexity of  $c$  in the second variable.

(2) To prove the convexity of  $\nu \rightarrow \mathcal{V}_c(\mu, \nu)$ , fix  $\nu_1, \nu_2 \in \mathcal{P}(Y)$  and find for a fixed  $\epsilon > 0$ ,  $\pi^1 \in \mathcal{K}(\mu, \nu_1)$  and  $\pi^2 \in \mathcal{K}(\mu, \nu_2)$  such that  $\int_X c(x, \pi_x^i) d\mu(x) \leq \mathcal{V}_c(\mu, \nu_i) + \epsilon$  for  $i = 1, 2$ . Define  $\pi \in \mathcal{P}(X \times Y)$  via  $d\pi(x, y) := (\lambda d\pi_x^1(y) + (1 - \lambda) d\pi_x^2(y)) d\mu(x)$ . With  $\nu_\lambda := \lambda \nu_1 + (1 - \lambda) \nu_2$ , we have  $\pi \in \mathcal{K}(\mu, \nu_\lambda)$ , and therefore, by convexity of  $c$  in the second variable, we have

$$\begin{aligned} \mathcal{V}_c(\mu, \nu_\lambda) &\leq \int_X c(x, \pi_x) d\mu(x) \leq \int_X \lambda c(x, \pi_x^1) d\mu(x) + \int_X (1 - \lambda) c(x, \pi_x^2) d\mu(x) \\ &\leq \lambda \mathcal{V}_c(\mu, \nu_1) + (1 - \lambda) \mathcal{V}_c(\mu, \nu_2) + \epsilon. \end{aligned}$$

We now show that  $\nu \mapsto \mathcal{V}_c(\mu, \nu)$  is weak\* lower semi-continuous. Using Lemma 5.2, it suffices to show that  $\nu \mapsto \hat{\mathcal{V}}_c(\mu, \nu)$  is weak\* lower semi-continuous. For that, let  $\nu_n \rightarrow \nu$  and for  $\epsilon > 0$ , select for each  $n$ ,  $P^n \in \Lambda(\mu, \nu_n)$  such that  $\int_{X \times \mathcal{P}(Y)} c(x, p) P^n(dx, dp) \leq \hat{\mathcal{V}}_c(\mu, \nu_n) + \epsilon$ . Now use the compactness of  $\mathcal{P}(X \times \mathcal{P}(Y))$  to find a further subsequence (which we again relabel to  $n$ ) so that  $P^n \rightarrow P$  for some  $P$ , which is necessarily in  $\Lambda(\mu, \nu)$  since the intensity function is continuous. It follows that

$$\begin{aligned} \hat{\mathcal{V}}_c(\mu, \nu) &\leq \int_{X \times \mathcal{P}(Y)} c(x, p) P(dx, dp) \\ &\leq \liminf_{n \rightarrow \infty} \int_{X \times \mathcal{P}(Y)} c(x, p) P^n(dx, dp) \\ &\leq \liminf_{n \rightarrow \infty} \hat{\mathcal{V}}_c(\mu, \nu_n) + \epsilon, \end{aligned}$$

which concludes the proof of lower semi-continuity. □

**Proof of Theorem 5.1:** To show that  $\mathcal{V} := \mathcal{V}_c$  is a backward linear transfer, note first that the above lemma yields that  $\nu \mapsto \mathcal{V}_c(\mu, \nu)$  is convex and lower semi-continuous. We now show that for every  $\mu \in \mathcal{P}(X)$ , its Legendre transform is  $g \mapsto \int_X T^-g d\mu$ , where

$$T^-g(x) = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - c(x, \sigma) \right\}.$$

Indeed, if  $\mu \in D_1(\mathcal{V}_c)$ , then for  $g \in C(Y)$ ,

$$\begin{aligned} -\infty < \mathcal{V}_\mu^*(g) &= \sup_{\nu \in \mathcal{P}(Y)} \left\{ \int_Y g d\nu - \mathcal{V}(\mu, \nu) \right\} \\ &= \sup_{\nu \in \mathcal{P}(Y)} \sup_{\pi \in \mathcal{K}(\mu, \nu)} \left\{ \int_Y g(y) d\nu(y) - \int_X c(x, \pi_x) d\mu(x) \right\} \\ (5.8) \qquad &= \sup_{\pi \in \mathcal{K}(\mu, \cdot)} \left\{ \int_X \left[ \int_Y g(y) d\pi_x(y) - c(x, \pi_x) \right] d\mu(x) \right\} \\ (5.9) \qquad &\leq \int_X \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - c(x, \sigma) \right\} d\mu(x) \\ &= \int_X T^-g d\mu. \end{aligned}$$

On the other hand, the lower semi-continuity of  $c$  yields that for each  $x$  the supremum in (5.9) is achieved by some  $\pi_x$  in a measurable way so that

$$T^-g(x) = \int_Y g(y) d\pi_x(y) - c(x, \pi_x) \text{ for every } x \in X.$$

Define  $\pi \in \mathcal{P}(X \times Y)$  via  $d\pi(x, y) := d\pi_x(y) d\mu(x)$ . Denoting  $\nu := \text{Proj}_Y \# \pi$ , we have  $\pi \in \mathcal{K}(\mu, \nu)$ . Hence, by using (5.8), we get that

$$\mathcal{V}_\mu^*(g) \geq \int_X \left[ \int_Y g(y) d\pi_x(y) d\mu(x) - \int_X c(x, \pi_x) d\mu(x) \right] = \int_X T^-g(x) d\mu(x),$$

and  $\mathcal{V}_\mu^*(g) = \int_X T^-g(x)d\mu(x)$ , hence  $\mathcal{V}_c$  is a backward linear transfer and

$$\mathcal{V}_c(\mu, \nu) = \sup_{g \in C(X)} \left\{ \int_X g d\nu - \int_X T^-g d\mu \right\} \quad \text{for all } (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y).$$

If now  $\mu \notin D_1(\mathcal{V}_c)$ , that is if  $\mathcal{V}_c(\mu, \nu) = +\infty$  for every  $\nu \in \mathcal{P}(Y)$ , then we claim that  $\int_X T^-g d\mu = -\infty$  for all  $g \in C(Y)$ . Indeed if not, we have for some  $g \in C(Y)$ ,

$$\begin{aligned} -\infty < \int_X T^-g(x) d\mu(x) &= \int_X \sup_{\sigma} \left\{ \int_Y g d\sigma - \mathcal{V}_c(\delta_x, \sigma) \right\} d\mu \\ &= \int_X \int_Y \{g(y)d\sigma_x(y) - \mathcal{V}_c(\delta_x, \sigma_x)\} d\mu(x), \end{aligned}$$

where  $x \rightarrow \sigma_x$  is a measurable selection of where  $\sigma \rightarrow \sup_{\sigma} \{ \int_Y g d\sigma - \mathcal{V}_c(\delta_x, \sigma) \}$  is attained on  $\mathcal{P}(Y)$ . Let now  $\nu := \int_X \sigma_x d\mu(x)$  and note that by the definition of  $\mathcal{V}_c$ , we have  $\mathcal{V}_c(\mu, \nu) \leq \int_X \mathcal{V}_c(\delta_x, \sigma_x) d\mu(x)$ , hence

$$\int_X T^-g(x) d\mu(x) = \int_Y g(y) d\nu(y) - \int_X \mathcal{V}_c(\delta_x, \sigma_x) d\mu(x) \leq \int_Y g(y) d\nu(y) - \mathcal{V}_c(\mu, \nu).$$

It follows that  $\mathcal{V}_c(\mu, \nu) < +\infty$ , hence  $\mu \in D_1(\mathcal{V}_c)$ , which is a contradiction. This means that  $\mathcal{V}_\mu^*(g) = \int_X T^-g d\mu$  for every  $\mu \in \mathcal{P}(X)$  and  $\mathcal{V}_c$  is a linear backward transfer.

(2) For the reverse implication, assume  $\mathcal{T}$  is a standard backward linear transfer with backward Kantorovich operator  $T^-$ , and define  $c(x, \sigma) := \mathcal{T}(\delta_x, \sigma)$ . Then  $c$  is bounded below, lower semi-continuous such that  $\sigma \mapsto c(x, \sigma)$  is proper and convex. By the first part,  $\mathcal{V}_c$  is a standard backward transfer with Kantorovich operator  $T^-g(x) = \sup_{\sigma \in \mathcal{P}(Y)} \{ \int_Y g d\sigma - c(x, \sigma) \}$  for every  $x \in X$ , which is the same for  $\mathcal{T}$ . Hence, for all  $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$ , we have

$$\mathcal{T}(\mu, \nu) = \sup_{g \in C(Y)} \left\{ \int_Y g d\nu - \int_X T^-g d\mu \right\} = \mathcal{V}_c(\mu, \nu) \quad \text{and } \mathcal{T} = \mathcal{V}_c.$$

The attainment in the disintegration (5.2) follows from the lower semi-continuity of  $\pi \rightarrow \int_X \mathcal{T}(\delta_x, \pi_x) d\mu$ , which implies that the infimum in (5.1) is attained. Note now that

$D(\mathcal{T}) = D(\mathcal{V}_c) = \{(\mu, \nu); \exists \pi \in \mathcal{K}(\mu, \nu) \text{ with } (x, \pi_x) \in D(\mathcal{T}) \text{ and } \mathcal{T}(\delta_x, \pi_x) \in L^1(\mu)\}$ , and by Theorem 1.5,

$$D(\mathcal{T}) \subset D(\mathcal{T}_r) = \{(\mu, \nu); \exists \pi \in \mathcal{K}(\mu, \nu) \text{ with } (\delta_x, \pi_x) \in D(\mathcal{T}_r) \quad \mu - \text{a.s.}\}.$$

If now  $D(\mathcal{T})$  is a backward transfer, then  $D(\mathcal{V}_c) = D(\mathcal{T}) = D(\mathcal{T}_r)$ , which establishes (5.3).

*Proof of Theorem 1.11.* To prove that  $\mathcal{B}_{c, \mathcal{A}}$  is a backward linear transfer, we consider the zero-cost balayage transfer associated to the balayage cone  $\mathcal{A}$ , that is

$$c_{\mathcal{A}}(x, \sigma) = \begin{cases} 0 & \text{if } \delta_x \prec_{\mathcal{A}} \sigma, \\ +\infty & \text{otherwise.} \end{cases}$$

We can then write

$$(5.10) \quad B_{c, \mathcal{A}}(\mu, \nu) := \inf \left\{ \int_X \{c(x, \pi_x) + c_{\mathcal{A}}(x, \pi_x)\} d\mu(x); \pi \in \mathcal{K}(\mu, \nu) \right\}.$$



Note that the cost  $\tilde{c}(x, \sigma) = c(x, \sigma) + c_{\mathcal{A}}(x, \sigma)$  is a proper, bounded below, lower semi-continuous functional that is convex in the second variable. and therefore Theorem 5.1 applies to yield that  $\mathcal{B}_{c, \mathcal{A}}$  is a backward linear transfer on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  with backward Kantorovich operator

$$T^-g(x) = \sup\left\{\int_Y g d\sigma - c(x, \sigma); \sigma \in \mathcal{P}(Y), \delta_x \prec_{\mathcal{A}} \sigma\right\}.$$

(ii) Suppose now  $\mathcal{T}$  is a backward linear transfer, then by Theorem 5.1, we can write

$$\mathcal{T}(\mu, \nu) = \inf\left\{\int_X \mathcal{T}(\delta_x, \pi_x) d\mu(x); \pi \in \mathcal{K}(\mu, \nu)\right\}.$$

By Theorem 3.4,  $D(\mathcal{T})$  is contained in a balayge set  $\mathcal{S} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \mu \prec_{\mathcal{A}} \nu\}$ , where  $\mathcal{A}$  is a cone on  $X \sqcup Y$ . This means that if  $(\mu, \nu) \in D(\mathcal{T})$ , then there exists  $\pi \in K(\mu, \nu)$  whose disintegration  $(\delta_x, \pi_x) \in \mathcal{S}$ . Now let  $c(x, \sigma) = \mathcal{T}(\delta_x, \sigma)$  and consider the optimal restricted balayage transfer

$$\mathcal{B}_{c, \mathcal{A}}(\mu, \nu) := \begin{cases} \inf\left\{\int_X c(x, \pi_x) d\mu(x); \pi \in \mathcal{K}_{\mathcal{A}}(\mu, \nu)\right\} & \text{if } \mu \prec_{\mathcal{A}} \nu \\ +\infty & \text{otherwise,} \end{cases}$$

Note that  $\mathcal{B}_{c, \mathcal{A}}$  can be written as in (5.10) and therefore  $\mathcal{T} \leq \mathcal{B}_{c, \mathcal{A}}$ .

On the other hand, we claim that  $\mathcal{B}_{c, \mathcal{A}} \leq \mathcal{T}$ . Indeed, it is clearly the case when  $(\mu, \nu) \notin D(\mathcal{T})$ . If however  $(\mu, \nu) \in D(\mathcal{T})$ , then  $\mu \prec_{\mathcal{A}} \nu$  and since

$$\mathcal{K}_{\mathcal{T}}(\mu, \nu) := \{\pi \in \mathcal{K}(\mu, \nu); (x, \pi_x) \in D(\mathcal{T}), \int_X \mathcal{T}(x, \pi_x) d\mu(x) < +\infty\} \subset \mathcal{K}_{\mathcal{A}}(\mu, \nu),$$

we have

$$\mathcal{T}(\mu, \nu) = \inf\left\{\int_X c(x, \pi_x) d\mu(x); \pi \in \mathcal{K}_{\mathcal{T}}(\mu, \nu) \geq \mathcal{B}_{c, \mathcal{A}}(\mu, \nu)\right\},$$

and we are done.

If now  $D(\mathcal{T})$  is a transfer set, then  $D(\mathcal{T}) = D(\mathcal{T}_r) = \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y); \mu \prec_{\mathcal{A}} \nu\}$ , and

$$\mathcal{T}(\mu, \nu) = \inf\left\{\int_X c(x, \pi_x) d\mu(x); \pi \in \mathcal{K}_{\mathcal{A}}(\mu, \nu)\right\}.$$

The remaining case is similar. □

## 6. THE KANTOROVICH ENVELOPE AND THE CHOQUET-KANTOROVICH CAPACITY

In this section, we show how the duality between Kantorovich operators and linear transfers give this subclass of Choquet functional capacities a great flexibility for constructing new ones.

The above observations motivate the search for Kantorovich operators as lower envelopes of arbitrary (standard) maps from  $C(Y)$  to  $USC(X)$ . This will be done in the sequel. We shall make frequent use of the following lemma.

**Lemma 6.1.** *Let  $M$  be a weak\*-compact set of  $\mathcal{P}(Y)$  and consider the following functional on  $C(Y)$*

$$(6.1) \quad Pf := \sup\left\{\int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \sigma \in M\right\}.$$

*Then,  $P$  is a positively 1-homogenous subadditive functional from  $C(Y)$  to  $\mathbb{R}$  and*

$$(6.2) \quad Pf = \sup\left\{\int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq P \text{ on } C_+(Y)\right\}.$$

*Proof.* The claimed Kantorovich properties of  $P$  are obvious. Since  $P$  is also sub-additive, then by the Hahn-Banach theorem, we have for any  $f \in C(K)$ ,

$$Pf = \sup\left\{\int_Y f d\sigma; \sigma \in \mathcal{M}(Y), \sigma \leq P \text{ on } C_+(Y)\right\}.$$

If now  $f \in C(Y)$ , there exists  $\sigma \in \mathcal{M}(Y)$ ,  $\sigma \leq P$  such that

$$P(f+1) = \sup\left\{\int_Y (f+1)d\tau; \tau \in \mathcal{M}(Y), \tau \leq P\right\} = \int_Y f d\sigma + \sigma(Y).$$

It follows that

$$1 + Pf = P(f+1) = \int_Y f d\sigma + \sigma(Y) \leq Pf + \sigma(Y) \leq Pf + P1 \leq Pf + 1,$$

hence  $\sigma(Y) = 1$  and  $\int_Y f d\sigma = Pf$ . Moreover, since  $\sigma \leq P$ , we have for any open set  $O$ ,  $1 - \sigma(O) = \sigma(Y \setminus O) \leq P(Y \setminus O)$ , hence  $0 \leq 1 - P(Y \setminus O) \leq \sigma(O)$  and  $\sigma$  is therefore a probability measure. Claim (6.2) follows.  $\square$

**Definition 6.2.** Say that a map  $T : C(Y) \rightarrow USC(X)$  is **standard** if for every  $x \in X$ , there is  $\nu \in \mathcal{P}(Y)$  such that

$$(6.3) \quad \sup_{g \in C(Y)} \left\{ \int_Y g d\nu - Tg(x) \right\} < +\infty.$$

**Theorem 6.3.** Let  $T : C(Y) \rightarrow USC(X)$  be a standard map. Then,

(1) The map  $\underline{T}$  defined for every  $g \in C(Y)$  by the expression

$$(6.4) \quad \underline{T}g(x) := \sup_{\sigma \in \mathcal{P}(Y)} \inf_{h \in C(Y)} \left\{ \int_Y (g-h) d\sigma + Th(x) \right\}$$

is a Kantorovich operator such that  $\underline{T} \leq T$ .

(2) For every  $g \in C(Y)$ , we have

$$(6.5) \quad \underline{T}g(x) = \sup\{Sg(x); S \text{ Kantorovich operator } S \leq T\},$$

and  $T = \underline{T}$  if and only if  $T$  is a Kantorovich operator.

(3) If  $T$  is positively 1-homogenous, then  $\underline{T}$  is also positively 1-homogenous and

$$(6.6) \quad \underline{T}f(x) = \sup\left\{\int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq T^x \text{ on } C_+(Y)\right\}.$$

(4) If  $X = Y$  and  $T \geq I$ , where  $I$  is the identity operator, then  $\underline{T} \geq I$ . Similarly, if  $T \geq T^2$ , then  $\underline{T} \geq \underline{T}^2$ .

We shall say that  $\underline{T}$  is the **Kantorovich envelope** of  $T$ .

*Proof.* 1) Consider the following cost functional  $c : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ ,

$$c(x, \sigma) := \sup_{g \in C(Y)} \left\{ \int_Y g d\sigma - Tg(x) \right\},$$

and note that  $c$  is weak\*-lower continuous on  $X \times \mathcal{P}(Y)$  and is bounded below since  $T(0) \in USC(X)$ , hence bounded above. Moreover, for every  $x \in X$ ,  $c(x, \cdot)$  is convex

and by condition (6.3) it is proper. Theorem 5.1 then yields that the weak optimal transport

$$\underline{T}(\mu, \nu) := \inf \left\{ \int_X c(x, \pi_x) d\mu; \pi \in \mathcal{K}(\mu, \nu) \right\}$$

is a standard backward linear transfer, with a corresponding backward Kantorovich operator

$$\underline{T}g(x) = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - c(x, \sigma) \right\} = \sup_{\sigma \in \mathcal{P}(Y)} \inf_{h \in C(Y)} \left\{ \int_Y (g - h) d\sigma + Th(x) \right\}.$$

Note that for any  $g \in C(Y)$ ,

$$\begin{aligned} \underline{T}g(x) &= \sup_{\sigma \in \mathcal{P}(Y)} \inf_{h \in C(Y)} \left\{ \int_Y g d\sigma - \int_Y h d\sigma + Th(x) \right\} \\ &\leq \inf_{h \in C(Y)} \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \int_Y h d\sigma + Th(x) \right\} \\ &= \inf_{h \in C(Y)} \left\{ \sup(g - h) + Th(x) \right\} \\ &\leq Tg(x). \end{aligned}$$

2) If now  $S$  is another backward Kantorovich operator such that  $S \leq T$ , then use Sion’s min-max principle [24], that  $h \rightarrow Sh(x)$  is convex and lower semi-continuous on  $C(Y)$  and that  $S(h + c) = Sh + c$  whenever  $c$  is a constant, to write

$$\begin{aligned} \underline{T}g(x) &= \sup_{\sigma \in \mathcal{P}(Y)} \inf_{h \in C(Y)} \left\{ \int_Y g d\sigma - \int_Y h d\sigma + Th(x) \right\} \\ &\geq \sup_{\sigma \in \mathcal{P}(Y)} \inf_{h \in C(Y)} \left\{ \int_Y g d\sigma - \int_Y h d\sigma + Sh(x) \right\} \\ &= \inf_{h \in C(Y)} \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \int_Y h d\sigma + Sh(x) \right\} \\ &= \inf_h \left\{ \sup(g - h) + Sh(x) \right\} \\ &= \inf_h \left\{ S[\sup(g - h) + h](x) \right\} \\ &\geq Sg(x). \end{aligned}$$

Consider now the operators

$$S_\infty f(x) = \sup \{ Sf(x); S \text{ Kantorovich } S \leq T \} \quad \text{and} \quad \hat{S}_\infty f(x) = \widehat{S_\infty f}(x),$$

where here  $\hat{g}$  is the upper semi-continuous envelope of  $g$ . It is clear that  $\underline{T} \leq S_\infty \leq \hat{S}_\infty \leq T$ , and that  $\underline{T} = S_\infty$  since  $\hat{S}_\infty$  is a Kantorovich operator.

3) Finally, if  $T$  is positively 1-homogenous, then formula (6.4) yields immediately that  $\underline{T}$  is positively 1-homogenous. In this case,

$$\underline{T}^x g := \underline{T}g(x) = \underline{T}_r g(x) = \sup \left\{ \int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \underline{T}(x, \sigma) < +\infty \right\}.$$

Use now Lemma 6.1 to write for every  $x \in X$ ,

$$\underline{T}g(x) = \sup \left\{ \int_Y g d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq \underline{T}^x \right\} \leq \sup \left\{ \int_Y g d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq T^x \right\}.$$

On the other hand,

$$\begin{aligned} \underline{T}g(x) &= \sup_{\sigma \in \mathcal{P}(Y)} \inf_{h \in C(Y)} \left\{ \int_Y g d\sigma - \int_Y h d\sigma + Th(x) \right\} \\ &\geq \sup_{\sigma \in \mathcal{P}(Y), \sigma \leq T^x} \inf_{h \in C(Y)} \left\{ \int_Y g d\sigma - \int_Y h d\sigma + Th(x) \right\} \\ &\geq \sup \left\{ \int_Y g d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq T^x \right\}, \end{aligned}$$

which establishes (6.6).

4) If  $T \geq I$ , then  $\underline{T} \geq I$  since the identity operator  $I$  is clearly a Kantorovich operator. Now suppose  $T^2 \leq T$ . Since  $\underline{T}^2$  is a Kantorovich operator, it suffices to note that  $\underline{T}^2 \leq T^2 \leq T$ , hence  $\underline{T}^2 \leq \underline{T}$ .  $\square$

Dually, we have implicitly shown the existence of an upper linear transfer envelope associated to any standard bounded below, weak\* lower semi-continuous functional on  $\mathcal{P}(X) \times \mathcal{P}(Y)$ . Indeed, we have the following.

**Proposition 6.4.** *Let  $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$  be a standard, convex, bounded below and weak\* lower semi-continuous functional, and let  $\underline{\mathcal{T}}$  be the optimal weak transport associated to the cost  $c(x, \sigma) := \mathcal{T}(\delta_x, \sigma)$ , that is  $\underline{\mathcal{T}}(\mu, \nu) := \inf \{ \int_X c(x, \pi_x) d\mu; \pi \in \mathcal{K}(\mu, \nu) \}$ .  $\underline{\mathcal{T}}$  is then a standard backward linear transfer satisfying  $\mathcal{T} \leq \underline{\mathcal{T}}$ . Moreover, if  $\mathcal{S}$  is any standard backward linear transfer such that  $\mathcal{T} \leq \mathcal{S}$ , then  $\underline{\mathcal{T}} \leq \mathcal{S}$*

*Proof.* Note that by the first part of Theorem 5.1,  $\underline{\mathcal{T}}$  is a backward linear transfer with backward Kantorovich operator  $T^-g(x) = \sup_{\sigma \in \mathcal{P}(Y)} \{ \int_Y g d\sigma - c(x, \sigma) \}$ . To show that  $\mathcal{T} \leq \underline{\mathcal{T}}$ , note that since  $\mathcal{T}$  is jointly convex and lower semi-continuous, then for each  $g \in C(Y)$ , the functional

$$\mu \rightarrow (\mathcal{T}_\mu)^*(g) = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \mathcal{T}(\mu, \sigma) \right\}$$

is upper semi-continuous and concave. It follows from Jensen's inequality that

$$(\mathcal{T}_\mu)^*(g) \geq \int_X (\mathcal{T}_{\delta_x})^*(g) d\mu(x) = \int_X T^-g(x) d\mu(x),$$

hence  $\mathcal{T}(\mu, \nu) = (\mathcal{T}_\mu)^{**}(\nu) \leq \sup_{g \in C(Y)} \{ \int_Y g d\nu - \int_X T^-g d\mu \} = \underline{\mathcal{T}}(\mu, \nu)$ .

$\underline{\mathcal{T}}$  is the smallest backward linear transfer greater than  $\mathcal{T}$ , since if  $\mathcal{S}$  is a backward linear transfer and  $\mathcal{T} \leq \mathcal{S}$ , then  $\underline{\mathcal{T}} \leq \underline{\mathcal{S}}$ , and  $\underline{\mathcal{S}} = \mathcal{S}$  by Theorem 6.3.  $\square$

**Remark 6.5.** Suppose  $\mathcal{T}$  is any convex, bounded below, and weak\* lower semi-continuous functional on  $\mathcal{P}(X) \times \mathcal{P}(Y)$  that is finite on the set of Dirac measures  $\{(\delta_x, \delta_y); x \in X, y \in Y\}$ . One can then define a cost function  $c(x, y) = \mathcal{T}(\delta_x, \delta_y)$ , and the associated optimal transport  $\mathcal{T}_c(\mu, \nu)$ . To compare  $\mathcal{T}$  with  $\mathcal{T}_c$ , note that

$$\mathcal{T}_{\delta_x}^*(g) = \sup \left\{ \int_Y g d\nu - \mathcal{T}(\delta_x, \nu); \nu \in \mathcal{P}(Y) \right\} \geq \sup \{ g(y) - c(x, y); y \in Y \} = T_c^-g(x)$$

and so  $\mathcal{T}(\mu, \nu) \leq \underline{\mathcal{T}}(\mu, \nu) \leq \mathcal{T}_c(\mu, \nu)$ . In many cases, it is not possible to define a proper cost  $c(x, y) = \mathcal{T}(\delta_x, \delta_y)$ , i.e.  $c$  is identically  $+\infty$ . This is the case for many

stochastic transport problems where, for example, transport via Brownian motion makes it impossible for a Dirac measure to be transported to another Dirac measure; see for instance [17].  $\square$

**The Choquet-Kantorovich envelope of a functional capacity**

Given two real-valued set functions  $S, T$  on a compact space  $Y$ , we shall say that

$$S \leq_{\mathcal{K}} T \text{ if and only if } S(K) \leq T(K) \text{ for any compact set } K \subset Y.$$

For any set function  $P$  on  $Y$ , we follow Choquet and associate the functional

$$(6.7) \quad \hat{P}(f) = \int_0^{+\infty} P(f \geq \alpha) d\alpha,$$

on the set of non-negative functions  $f$  on  $Y$ . We call  $\hat{P}$  **the Choquet extension of  $P$** . Note that

$$S \leq_{\mathcal{K}} T \text{ if and only if } \hat{S}(f) \leq \hat{T}(f) \text{ for all } f \in USC_+(Y).$$

**Definition 6.6.** Say that a functional  $P : USC(Y) \rightarrow \mathbb{R}$  is **saturated** if for any  $\sigma \in \mathcal{P}(Y)$ ,

$$(6.8) \quad \sigma \leq_{\mathcal{K}} P \text{ if and only if } \sigma \leq P \text{ on } USC_+(Y).$$

If  $P$  is a capacity, then  $\hat{P}$  is also a capacity that coincides with  $P$  on the characteristic functions of sets.  $\hat{P}$  can be seen as an extension of  $P$  to non-negative functions in such a way that it is still monotone increasing, satisfies  $\hat{P}f = \lim_n \uparrow \hat{P}f_n$  if  $f = \lim_n \uparrow f_n$ , and  $\hat{P}f = \lim_n \downarrow \hat{P}f_n$  if  $f, f_n$  are upper semi-continuous and  $f = \lim_n \downarrow f_n$ . We also note that  $\hat{P}$  is positively 1-homogenous, and that it is saturated.

Similarly to the Choquet functional extension  $\hat{P}$  of  $P$ , we follow Dellacherie [9] and introduce the following functional extension  $\tilde{P}$  of the set function  $P$ .

**Definition 6.7.** Say that a set function  $P$  is **common** if there is  $\nu \in \mathcal{P}(Y)$  such that  $\nu \leq_{\mathcal{K}} P$ .

For such set functions, we define for  $f \in USC(Y)$ ,

$$(6.9) \quad \tilde{P}f = \sup\left\{ \int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} P \right\}.$$

**Lemma 6.8.** *Let  $P$  be a non-negative common set function. Then,*

- (1)  $\tilde{P}$  is a saturated functional on  $USC(Y)$ .
- (2) If  $P$  originates from a saturated functional on  $USC(Y)$ , then  $\tilde{P} \leq P$  on  $USC_+(Y)$ .
- (3) If  $P$  is given on  $USC(Y)$  by  $Pf = \sup\{\int_Y f d\sigma; \sigma \in M\}$ , where  $M$  is weak\*-compact subset of  $\mathcal{P}(Y)$ , then  $P \leq \tilde{P}$  on  $USC_+(Y)$ , hence if  $P$  is also saturated then  $P = \tilde{P}$  on  $USC_+(Y)$ .
- (4) We always have  $\tilde{P} = \tilde{\tilde{P}}$  on  $USC_+(Y)$ .

*Proof.* 1) That  $\tilde{P}$  is saturated is clear. 2) If now  $P$  is saturated, then  $\tilde{P} \leq P$ , since

$$\tilde{P}f = \sup\left\{ \int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} P \right\} = \sup\left\{ \int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq P \right\}.$$

3) If  $Pf = \sup\{\int_Y fd\sigma; \sigma \in M\}$ , where  $M$  is weak\*-compact subset of  $\mathcal{P}(Y)$ , then Lemma 6.1 gives that

$$Pf = \sup\{\int_Y fd\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq P\} \leq \sup\{\int_Y fd\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} P\} = \tilde{P}f,$$

and therefore if  $P$  is also saturated, then  $P = \tilde{P}$ .

4) Since  $\tilde{P}$  is saturated, we always have  $\tilde{\tilde{P}} \leq \tilde{P}$  by the first item. On the other hand, the set  $M = \{\sigma \in \mathcal{P}(Y); \sigma \leq_K P\}$  is equal to  $\{\sigma \in \mathcal{P}(Y); \sigma \leq \hat{P}\}$ , where  $\hat{P}$  is the Choquet extension of  $P$ , hence it is weak\*-compact. It follows from item 3) that  $\tilde{P} \leq \tilde{\tilde{P}}$ , hence we have equality.  $\square$

**Definition 6.9.** Say that a map  $T : F_+(Y) \rightarrow F_+(X)$  is **common (resp., saturated)** if for every  $x \in X$ ,  $T^x$  is common (resp., saturated).

**Theorem 6.10.** Let  $T$  be a common functional capacity from  $F_+(Y) \rightarrow F_+(X)$ . Then,

(1) The map  $\tilde{T}$  defined for every  $g \in C(Y)$  by the expression

$$(6.10) \quad \tilde{T}f(x) = \tilde{T}^x(f) = \sup\{\int_Y fd\sigma; \sigma \in \mathcal{P}(Y), \sigma \leq_K T^x\}$$

is a saturated positively 1-homogenous Kantorovich operator such that  $\tilde{T} \leq_{\mathcal{K}} T$ .

(2) If  $T$  is a positively 1-homogenous Kantorovich operator, then  $T \leq \tilde{T}$  on  $C_+(Y)$ , and for any  $g \in C(Y)$ ,

$$(6.11) \quad \tilde{T}g(x) = \sup\{Sg(x); S \text{ is a positively 1-homogenous Kantorovich operator } \hat{S} \leq \hat{T}\}.$$

(3) If  $T$  is saturated, then  $\tilde{T} \leq T$ .

(4) If  $T$  is a saturated positively 1-homogenous Kantorovich operator, then  $T = \tilde{T}$  on  $C_+(Y)$ .

(5)  $\tilde{T}$  is minimal among saturated positively 1-homogenous Kantorovich operators greater than  $T$ .

We shall say that  $\tilde{T}$  is the **Dellacherie envelope of  $T$** .

*Proof.* 1) Note first that  $\tilde{T}$  as defined in (6.10) is clearly a saturated positively 1-homogenous operator such that  $\tilde{T} \leq_{\mathcal{K}} T$ . To show that it is a Kantorovich operator, it suffices to prove that the corresponding cost functional

$$(6.12) \quad \tilde{c}(x, \sigma) = \begin{cases} 0 & \text{if } \sigma \leq_K T^x \\ +\infty & \text{otherwise,} \end{cases}$$

is lower semi-continuous. Note that  $\tilde{c}$  is proper since  $T$  is common. Let now  $x_n \rightarrow x$  in  $X$  and  $\nu_n \rightarrow \nu$  weak\* in  $\mathcal{P}(Y)$ . For any compact set  $K \subset Y$  and any  $\epsilon > 0$ , and since  $T^x$  is a capacity, there is an open set  $O_{2\epsilon} \supset \bar{O}_\epsilon \supset O_\epsilon \supset K$  such that  $T^x(O_{2\epsilon}) - T^x(K) < \epsilon$ . Since  $\sigma \rightarrow \sigma(O_\epsilon)$  is lower semi-continuous, and  $x \rightarrow T^x(\bar{O}_\epsilon)$  is upper semi-continuous, we have

$$\sigma(K) \leq \sigma(O_\epsilon) \leq \liminf_n \sigma_n(O_\epsilon) \leq \liminf_n T^{x_n}(O_\epsilon) \leq \limsup_n T^{x_n}(\bar{O}_\epsilon) \leq T^x(\bar{O}_\epsilon) \leq T^x(K) + \epsilon.$$

It follows that  $\sigma \leq_K T^x$  and  $\tilde{c}$  is therefore lower semi-continuous.

2) If now  $S$  is positively 1-homogenous Kantorovich operator such that  $S \leq_{\mathcal{K}} T$ , then since

$$Sg(x) = S_r g(x) = \sup \left\{ \int_Y g \, d\sigma; \mathcal{T}_S(x, \sigma) < +\infty \right\},$$

we can apply Lemma 6.1 and write

$$\begin{aligned} Sg(x) &= \sup \left\{ \int_Y g \, d\sigma; \sigma \leq S^x \right\} \\ &\leq \sup \left\{ \int_Y g \, d\sigma; \sigma \leq_{\mathcal{K}} S^x \right\} \\ &\leq \sup \left\{ \int_Y g \, d\sigma; \sigma \leq_{\mathcal{K}} T^x \right\} \\ &= \tilde{T}g(x). \end{aligned}$$

Formula (6.11) follows along the same line as the proof of (6.5)

3) follows from Lemma 6.8.

4) follows from 2) and 3).

5) Finally, if  $R$  is a saturated positively 1-homogenous Kantorovich operator such that  $R \geq T$ , then clearly  $R = \tilde{R} \geq \tilde{T}$ .  $\square$

If now  $T : F_+(Y) \rightarrow F_+(X)$  is a functional capacity, then

$$\hat{T}f(x) := \hat{T}^x(f) = \int_0^{+\infty} T^x(f \geq \alpha) \, d\alpha,$$

is a functional capacity that is positively 1-homogenous and saturated (i.e., for each  $x \in X$ ,  $T^x$  is a saturated capacity). The following notion of saturation is more appropriate for functional capacities.

**Definition 6.11.** Say that a functional capacity  $T : F_+(Y) \rightarrow F_+(X)$  is **K-saturated** if for any Kantorovich operator  $S : USC(Y) \rightarrow USC(X)$  we have that

$$(6.13) \quad S \leq_{\mathcal{K}} T \quad \text{if and only if} \quad S \leq T \quad \text{on } USC_+(Y).$$

Let us also say that a functional capacity  $T : F_+(Y) \rightarrow F_+(X)$  is **compactly standard** if for every  $x \in X$ , there is  $\nu \in \mathcal{P}(Y)$  such that

$$(6.14) \quad \sup_{K \text{ compact}} \{ \sigma(K) - T\chi_K(x) \} < +\infty.$$

**Theorem 6.12.** Let  $T : F_+(Y) \rightarrow F_+(X)$  be a compactly standard functional capacity. Then,

(1) The map  $\bar{T}$  defined for every  $g \in C(Y)$  by the expression

$$(6.15) \quad \bar{T}g(x) := \sup_{\sigma \in \mathcal{P}(Y)} \inf_{O \text{ open}} \left\{ \int_Y (g - \chi_O) \, d\sigma + T\chi_O(x) \right\},$$

is a  $K$ -saturated Kantorovich operator such that  $\bar{T} \leq_{\mathcal{K}} T$ .

(2) If  $T$  is a Kantorovich operator then  $T \leq \bar{T}$ , and for any  $g \in C(Y)$ , we have

$$(6.16) \quad \bar{T}g(x) = \sup \{ Sg(x); S \text{ Kantorovich operator with } \hat{S} \leq \hat{T} \}.$$

(3) If  $T$  is  $K$ -saturated, then  $\bar{T} \leq T$ .

- (4) If  $T$  is a  $K$ -saturated Kantorovich operator, then  $T = \bar{T}$ .
- (5)  $\bar{T}$  is minimal among  $K$ -saturated Kantorovich operators greater than  $T$ .

We shall then say that  $\bar{T}$  is the **Choquet-Kantorovich envelope** of  $T$ .

**Lemma 6.13.** Let  $T$  be a functional capacity from  $F_+(Y)$  to  $F_+(X)$ .

- (1) If for some  $x \in X$  and  $\sigma \in \mathcal{P}(Y)$ ,  $\bar{c}(x, \sigma) := \sup_{K \text{ compact } \subset Y} \{\sigma(K) - T^x(K)\} < +\infty$ , then

$$(6.17) \quad \bar{c}(x, \sigma) = \sup_{O \text{ open in } Y} \{\sigma(O) - T^x(\bar{O})\}.$$

- (2) If  $T$  is a standard functional capacity, then  $(x, \sigma) \rightarrow \bar{c}(x, \sigma)$  is lower semi-continuous such that for each  $x \in X$ ,  $\sigma \rightarrow \bar{c}(x, \sigma)$  is convex and proper. Moreover,

$$(6.18) \quad \bar{c}(x, \sigma) \leq \sup_{f \in C(Y)} \left\{ \int_Y f d\sigma - T f(x) \right\}.$$

*Proof.* 1) Note that

$$\sup_{O \text{ open}} \{\sigma(O) - T^x(\bar{O})\} \leq \sup_{O \text{ open}} \{\sigma(\bar{O}) - T^x(\bar{O})\} \leq \sup_{K \text{ compact in } Y} \{\sigma(K) - T^x(K)\}.$$

On the other hand, if  $\bar{c}(x, \sigma) < +\infty$ , then since  $T^x$  is a capacity, we have for every  $K_\epsilon$  compact such that  $\bar{c}(x, \sigma) - \epsilon \leq \sigma(K_\epsilon) - T^x(K_\epsilon) \leq \bar{c}(x, \sigma)$ , an open set  $O_{2\epsilon} \supset \bar{O}_\epsilon \supset O_\epsilon \supset K_\epsilon$  such that  $T^x(O_{2\epsilon}) - T^x(K_\epsilon) < \epsilon$ , hence

$$\bar{c}(x, \sigma) - 2\epsilon \leq \sigma(O_\epsilon) - T^x(O_{2\epsilon}) \leq \sigma(O_\epsilon) - T^x(\bar{O}_\epsilon) \leq \sup_{O \text{ open}} \{\sigma(O) - T^x(\bar{O})\},$$

which implies (6.17).

2) It follows that  $\bar{c}$  is lower semi-continuous since  $\sigma \rightarrow \sigma(O)$  is lower semi-continuous for every open set, and  $x \rightarrow T^x(\bar{O})$  is upper semi-continuous since  $T$  maps  $USC(Y)$  to  $USC(X)$  and  $\bar{O}$  is compact. Also,  $\sigma \rightarrow \bar{c}(x, \sigma)$  is clearly convex.

Now again, since  $T$  is a functional capacity, we have for every open  $O$  such that

$$\bar{c}(x, \sigma) - \epsilon \leq \sigma(O_\epsilon) - T^x(\bar{O}_\epsilon) \leq \bar{c}(x, \sigma),$$

a continuous function  $f_\epsilon \geq \chi_{\bar{O}_\epsilon}$  such that  $T^x(f_\epsilon) - T^x(\bar{O}_\epsilon) \leq \epsilon$ , hence

$$\bar{c}(x, \sigma) - 2\epsilon \leq \int_Y f_\epsilon d\sigma - T^x(f_\epsilon) \leq \sup_{f \in C(Y)} \left\{ \int_Y f d\sigma - T f(x) \right\},$$

which implies (6.18). □

*Proof.* (of Theorem 6.12) Consider the cost functional  $\bar{c} : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ ,

$$\bar{c}(x, \sigma) := \sup_{K \text{ compact}} \{\sigma(K) - T^x(K)\},$$

and note that condition (6.3) ensures that for every  $x \in X$ ,  $\bar{c}(x, \cdot)$  is proper. Proposition 6.4 then yields that the weak transport  $\bar{T}(\mu, \nu) := \inf\{\int_X \bar{c}(x, \pi_x) d\mu; \pi \in \mathcal{K}(\mu, \nu)\}$  is a standard backward linear transfer, with a corresponding backward Kantorovich operator

$$\bar{T}g(x) = \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \bar{c}(x, \sigma) \right\} = \sup_{\sigma \in \mathcal{P}(Y)} \inf_{O \text{ open}} \left\{ \int_Y (g - \chi_O) d\sigma + T\chi_{\bar{O}}(x) \right\}.$$



We now claim that  $\bar{T}$  is  $K$ -saturated. Indeed, if  $S$  is a Kantorovich operator such that  $\hat{S} \leq \hat{T}$ , then again use Sion's min-max theorem and the properties of  $S$  to write for any  $g \in USC(Y)$ ,

$$\begin{aligned} \bar{T}g(x) &= \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \bar{c}(x, \sigma) \right\} \\ &= \sup_{\sigma \in \mathcal{P}(Y)} \inf_{K \text{ compact}} \left\{ \int_Y g d\sigma - \sigma(K) + T^x(K) \right\} \\ &\geq \sup_{\sigma \in \mathcal{P}(Y)} \inf_{K \text{ compact}} \left\{ \int_Y g d\sigma - \sigma(K) + S^x(K) \right\} \\ &\geq \sup_{\sigma \in \mathcal{P}(Y)} \inf_{f \in C(Y)} \left\{ \int_Y g d\sigma - \int_Y f d\sigma + Sf(x) \right\} \\ &= \inf_{f \in C(Y)} \sup_{\sigma \in \mathcal{P}(Y)} \left\{ \int_Y g d\sigma - \int_Y f d\sigma + Sf(x) \right\} \\ &= \inf_{f \in C(Y)} \left\{ \sup_{y \in Y} (g - f) + Sf(x) \right\} \\ &= \inf_{f \in C(Y)} \left\{ S[\sup_{y \in Y} (g - f) + f](x) \right\} \\ &\geq Sg(x). \end{aligned}$$

Write now for any  $g \in USC(Y)$ ,

$$\begin{aligned} \bar{T}g(x) &= \sup_{\sigma \in \mathcal{P}(Y)} \inf_{K \text{ compact}} \left\{ \int_Y g d\sigma - \sigma(K) + T^x(K) \right\} \\ &\leq \inf_{K \text{ compact}} \left\{ \sup_Y (g - \chi_K) + T^x(K) \right\}, \end{aligned}$$

which yields that  $\bar{T}^x(A) \leq T^x(A)$  for any compact set  $A$  in  $Y$ .

- 2) The proof of (6.16) follows the same lines as the proof of (6.5).
- 3) If  $T$  is  $K$ -saturated, then  $\bar{T} \leq T$  since  $\bar{T} \leq_{\mathcal{K}} T$ .
- 4) follows by combining 2) and 3). 5) If  $R$  is a  $K$ -saturated Kantorovich operator such that  $R \geq T$ , then clearly  $R = \bar{R} \geq \bar{T}$ . □

**Definition 6.14.** A capacity  $P$  on a compact space  $Y$  is said to be:

- (1) *Subadditive of order infinity* if for every compact set  $K \subset Y$ , any  $n \in \mathbb{N}$ , and any finite family of compact sets  $(K_i)_{i=1}^m$  such that  $n\chi_K \leq \sum_{i=1}^m \chi_{K_i}$ , then
 
$$(6.19) \quad nP(K) \leq \sum_{i=1}^m P(K_i).$$
- (2) *Strictly subadditive of order infinity* if for every compact set  $K \subset Y$ , any  $n \in \mathbb{N}$ , any  $k \in \mathbb{N} \cup \{0\}$ , and any finite family of compact sets  $(K_i)_{i=1}^m$  such that  $k + n\chi_K \leq \sum_{i=1}^m \chi_{K_i}$ , then
 
$$(6.20) \quad kP(Y) + nP(K) \leq \sum_{i=1}^m P(K_i).$$
- (3) *Strongly subadditive* if for any two compact subsets  $A, B$  of  $Y$ ,
 
$$(6.21) \quad P(A \cup B) + P(A \cap B) \leq P(A) + P(B).$$

It is known that (6.21)  $\Rightarrow$  (6.20)  $\Rightarrow$  (6.19) and that

- $P$  is subadditive of order infinity if  $P(A) = \sup\{\sigma(A); \sigma \in \mathcal{M}(Y), \sigma \leq_{\mathcal{K}} P\}$  for any compact set  $A \subset Y$  (Anger and Lembcke [3]).
- $P$  is strictly subadditive of order infinity if  $P(A) = \sup\{\sigma(A); \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} P\}$  for any compact set  $A \subset Y$  (Anger and Lembcke [3]).
- $P$  is strongly subadditive capacity if and only if  $\hat{P}$  is subadditive on  $C(Y)$  (Choquet [11]).

If now  $T : F_+(Y) \rightarrow F_+(X)$  is a functional capacity, then say that  $T$  is *subadditive of order infinity* (resp., *strictly subadditive of order infinity*) (resp., *strongly subadditive*), if for any  $x \in X$ ,  $T^x$  is a regular capacity that satisfies (6.19) (resp., (6.20)), (resp., (6.21)).

**Theorem 6.15.** *Let  $T : F_+(Y) \rightarrow F_+(X)$  be a functional capacity, then*

- (1)  *$T$  is strictly subadditive of order infinity if and only if  $\hat{T} = \hat{S}$  for some saturated positively 1-homogenous Kantorovich operator  $S$ , in which case  $\bar{T}$  is also strictly subadditive of order infinity.*
- (2)  *$T$  is strongly subadditive if and only if  $\hat{T}$  is a Kantorovich operator, in which case  $\bar{T}$  is also strongly subadditive.*

*Proof.* 1) If  $T$  is strictly subadditive of order infinity then for any compact set  $A$  in  $Y$ ,

$$T^x(A) = \sup\{\sigma(A); \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} T^x\},$$

which means that  $\tilde{T}$  and  $\bar{T}$  exist,  $\tilde{T} \leq \bar{T}$  by the maximality of the latter, and  $\tilde{T}(\chi_A) = \bar{T}(\chi_A) = T(\chi_A)$  for any compact  $A$ , that is,  $\hat{T} = \hat{\bar{T}}$ . Note that this implies that  $\bar{T}$  is strictly subadditive of order infinity since then,

$$\begin{aligned} \bar{T}^x(A) &= T^x(A) = \sup\{\sigma(A); \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} T^x\} \\ &\geq \sup\{\sigma(A); \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} \bar{T}^x\} \\ &\geq \sup\{\sigma(A); \sigma \in \mathcal{P}(Y), \sigma \leq \tilde{T}^x\} \\ &= \tilde{T}^x(A) \\ &= \bar{T}^x(A). \end{aligned}$$

Conversely, if  $\hat{T} = \hat{S}$  for some positively 1-homogenous Kantorovich operator  $S$ , then by the maximality property of  $\tilde{T}$ , we have  $S \leq \tilde{T}$ . It follows that  $\hat{T} = \hat{S} \leq \hat{\tilde{T}} \leq \hat{\bar{T}} \leq \hat{T}$ , and therefore we have equality, from which follows that for every compact set  $A$ ,

$$T^x(A) = \tilde{T}^x(A) = \sup\{\sigma(A); \sigma \in \mathcal{P}(Y), \sigma \leq_{\mathcal{K}} T^x\},$$

which means that  $T$  is strictly subadditive of order infinity,

2) If now  $T$  is strongly subadditive, it is then strictly subadditive of order infinity and  $\hat{T} = \hat{\tilde{T}} = \hat{\bar{T}}$ . Moreover,  $\hat{T}$  is subadditive and therefore  $\hat{\tilde{T}}$  is subadditive, and  $\tilde{T}$  is then strongly subadditive. By a remark of Dellacherie [9], since  $\tilde{T}$  is a supremum of measures over a weakly compact subset in  $\mathcal{P}(Y)$ ,  $\tilde{T}$  is strongly subadditive if and only if  $\hat{\tilde{T}} = \tilde{T}$ . It follows that  $\hat{T} = \tilde{T}$ . Note that this implies that  $\bar{T}$  is strongly subadditive since  $\hat{\bar{T}} = \tilde{T}$  and the latter is subadditive.

Conversely, assume  $\hat{T}$  is a Kantorovich operator. Since it is also positively 1-homogenous, it is subadditive and therefore  $T$  is then strongly subadditive by Choquet's criterium.  $\square$

## REFERENCES

- [1] J. J. Alibert, G. Bouchitté and T. Champion, *A new class of costs for optimal transport planning*, European Journal of Applied Mathematics **30** (2019), 1229–1263.
- [2] L. Alvarez, F. Guichard and P. L. Lions, *Axioms and fundamental equations of image processing*, Arch. Rational Mech. Anal. **123** (1993), 199–257.
- [3] B. Anger and J. Lembcke, *Infinitely Subadditive Capacities as Upper Envelopes of Measures*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **68** (1985), 403–414.
- [4] J. Backhoff-Veraguas, M. Beiglböck and G. Pammer, *Existence, duality, and cyclical monotonicity for weak transport costs*, Calculus of Variations and PDE **58** (2019), 1–28.
- [5] M. Bowles and N. Ghoussoub, *A Theory of Transfers: Duality and convolution*, (April 16, 2018, Revised October 24, 2018) 41 pp. <https://arxiv.org/abs/1804.08563>
- [6] M. Bowles and N. Ghoussoub, *Mather Measures and Ergodic Properties of Kantorovich Operators*, (May 7, 2019, Revised June 20, 2019, 2nd revision on October 24, 2019) 109 pp. <https://arxiv.org/abs/1905.05793>
- [7] M. Bowles and N. Ghoussoub, *Ergodic properties of Kantorovich operators*, In preparation, 2023.
- [8] C. Dellacherie and P. A. Meyer, *Probabilités et Potentiel*, Chapter X, Hermann, 1997.
- [9] C. Dellacherie, *Quelques commentaires sur les prolongements de capacités*, Séminaire de probabilités (Strasbourg), tome **5** (1971), 77–81.
- [10] K. J. Ciosmak, *Applications of Strassen's theorem and Choquet theory to optimal transport problems, to uniformly convex functions and to uniformly smooth functions*, Nonlinear Analysis, Volume 232, July 2023, 113267.
- [11] G. Choquet, *Theory of capacities*, Ann. Inst. Fourier **5** (1953/54), 131–295.
- [12] G. Choquet, *Lectures on Analysis*, W. A. Benjamin, Inc. 1968.
- [13] S. Dweik, N. Ghoussoub, Y. Kim and A. Palmer, *Stochastic Optimal Transport With Free End Time*, Ann. de l'Institut Henri Poincaré, Prob. & Stats, **57** (2021), 700–725.
- [14] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag, vol. 25, New York, 1993.
- [15] I. Gentil, C. Leonard and L. Ripani, *About the analogy between optimal transport and minimal entropy*, Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6, Université Paul Sabatier **26** (2017), 569–600.
- [16] N. Ghoussoub, Y. Kim, T. Lim *Structure of optimal martingale transport in general dimensions*, Ann. of Probability **47** (2019), 109–164.
- [17] N. Ghoussoub, Y. Kim and A. Palmer, *A solution to the Monge transport problem for Brownian martingales*, Ann. of Probability **49** (2021), 877–907.
- [18] N. Ghoussoub, Y. Kim and A. Palmer, *PDE methods for optimal Skorokhod embeddings*, Calc. of Variations and PDEs, Calc. Var. **58** (2019): 113.
- [19] N. Ghoussoub, Y. Kim and A. Palmer, *Optimal stopping of stochastic transport minimizing submartingale costs*, Trans. AMS **374** (2021), 6963–6989.
- [20] C. Gozlan, P. Roberto, M. Samson and P. Tetali, *Kantorovich duality for general transport costs and applications*, J. Func. Analysis **273** (2017), 3327–3405.
- [21] P. Henry-Labordère, *Model-free Hedging: A Martingale Optimal Transport Viewpoint*, Chapman and Hall/CRC, 2017.
- [22] K. Marton, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), 556–571.
- [23] T. Mikami and M. Thieullen, *Duality theorem for the stochastic optimal control problem*. Stoch. Process. Appl. **116** (2006), 1815–1835
- [24] M. Sion, *On general minimax theorems*, Pacific J. Math. **8** (1958), 171–176.

- [25] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439.
- [26] M. Talagrand, *Transportation cost for gaussian and other product measures*, Geometric and Functional Analysis **6** (1996), 587–600.
- [27] C. Villani, *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.

*Manuscript received December 11 2022*  
*revised March 3 2023*

NASSIF GHOUSSOUB

Department of Mathematics, The University of British Columbia Vancouver BC Canada V6T 1Z2

*E-mail address:* `nassif@math.ubc.ca`