A Theory of Transfers: Duality and convolution

Malcolm Bowles* and Nassif Ghoussoub†

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

April 16, 2018

Abstract

We introduce and study the permanence properties of the class of linear transfers between probability measures. This class contains all cost minimizing mass transports, but also martingale mass transports, the Schrödinger bridge associated to a reversible Markov process, and the weak mass transports of Talagrand, Marton, Gozlan and others. The class also includes various stochastic mass transports to which Monge-Kantorovich theory does not apply. We also introduce the cone of convex transfers, which include any $p$-power ($p \geq 1$) of a linear transfer, but also the logarithmic entropy, the Donsker-Varadhan information and certain free energy functionals. This first paper is mostly about exhibiting examples that point to the pervasiveness of the concept in the important work on correlating probability distributions. Duality formulae for general transfer inequalities follow in a very natural way. We also study the infinite self-convolution of a linear transfer in order to establish the existence of generalized weak KAM solutions that could be applied to the stochastic counterpart of Fathi-Mather theory.

1 Introduction

Stochastic control problems and several other analytical and statistical procedures that correlate two probability distributions share many of the useful properties of optimal mass transportation between probability measures. However, these correlations often lack at least two of the useful features of Monge-Kantorovich theory [37]. For one, they are not symmetric, meaning that the problem imposes a specific direction from one of the marginal distributions to the other. Moreover, many of those do not arise as cost minimizing problems associated to functionals $c(x,y)$ that assign “a price for moving one particle $x$ to another $y$.” As such, they are not readily amenable to the duality theory of Monge-Kantorovich. In this paper, we isolate and study a notion of transfers between probability measures that encapsulates both the deterministic and stochastic versions of transport problems studied by Mikami-Thieulin [29] and Barton-Ghoussoub [3], but also includes the weak mass transports of Talagrand [35, 36], Marton [24, 25] and Gozlan et al. [18, 20], the logarithmic entropy, the Donsker-Varadhan information [10], and many other energy correlation functionals.

This first paper introduces the unifying concepts of linear and convex mass transfers and exhibits several examples that illustrate the potential scope of this approach. The underlying idea has been implicit in many related works and should be familiar to the experts. But, as we shall see, the systematic study of these structures add clarity and understanding, allow for non-trivial extensions, and open up a whole new set of interesting problems. The ultimate purpose is to extend many of the remarkable properties enjoyed by standard mass transportations to linear and convex transfers, and hence to the stochastic case, or at least to weak mass transports. This is a vast undertaking. We therefore decided –for this first paper– to give a sample of the results that can be inspired and eventually extended from the standard theory of mass

---

*This is part of the PhD dissertation of this author at the University of British Columbia.
†Partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
transport. We therefore chose to state here the most basic permanence properties of the cones of transfers and to establish general duality formulas for potential comparisons between different transfers that extend the work of Bobkov-Götze [3], Gozlan-Leonard [18], Maurey [27] and others. We also show how the approach of Bernard-Buffoni [1, 2] on Fathi’s weak KAM [11] and Mather theory [26] extend to linear transfers, and therefore could, for example, be applied to the stochastic case. We will pursue this in a forthcoming paper [7]. Furthermore, we shall present in [8] a notion of linear and convex multi-transfers between several probability distributions that will – among other things – extend the theory of multi-marginal mass transportation.

We shall focus here on probability measures on compact spaces, even though the right settings for most applications and examples are complete metric spaces, or at least $\mathbb{R}^n$. This will allow us to avoid the usual functional analytic complications, and concentrate on the algebraic aspects of the theory. The simple compact case will at least point to results that can be expected to hold and be proved – albeit with additional analysis and suitable hypothesis – in more general situations. With this in mind, we shall denote by $C(X)$ (resp. USC(X)) to be the spaces of continuous (resp., upper semi-continuous), (resp., lower semi-continuous) functions on a compact space $X$. The class of signed (resp., probability) measures on $X$ will be denoted by $\mathcal{M}(X)$ (resp., $\mathcal{P}(X)$).

If now $T : \mathcal{M}(X) \times \mathcal{M}(Y) \to \mathbb{R} \cup \{+\infty\}$ is a proper convex functional, we shall denote by $D(T)$ its effective domain, that is the set where it takes finite values. We shall always assume that $D(T) \subset \mathcal{P}(X) \times \mathcal{P}(Y)$, where $\mathcal{P}(X)$ is the set of probability measures on $X$. The “partial domains” of $T$ are then denoted by,

$$D_1(T) = \{\mu \in \mathcal{P}(X) ; \exists \nu \in \mathcal{P}(Y), (\mu, \nu) \in D(T)\} \quad \text{and} \quad D_2(T) = \{\nu \in \mathcal{P}(Y) ; \exists \mu \in \mathcal{P}(X), (\mu, \nu) \in D(T)\}.$$ 

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

**Definition 1** Let $X$ and $Y$ be two compact spaces, and let $T : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a proper convex and weak* lower semi-continuous functional on $\mathcal{M}(X) \times \mathcal{M}(Y)$. We say that

1. $T$ is a **backward linear transfer**, if there exists a map $T^- : C(Y) \to LSC(X)$ such that for each $\mu \in D_1(T)$, the Legendre transform of $T_\mu$ on $\mathcal{M}(Y)$ satisfies:

   $$T_\mu^*(g) = \int_X T^- g(x) \, d\mu(x) \quad \text{for any } g \in C(Y).$$

2. $T$ is a **forward linear transfer**, if there exists a map $T^+ : C(X) \to USC(Y)$ such that for each $\nu \in D_2(T)$, the Legendre transform of $T_\nu$ on $\mathcal{M}(X)$ satisfies:

   $$T_\nu^*(f) = -\int_Y T^+ (-f)(y) \, d\nu(y) \quad \text{for any } f \in C(X).$$

We shall call $T^+$ (resp., $T^-$) the **forward** (resp., **backward**) Kantorovich operator associated to $T$.

By Legendre transform of $T_\nu$, we mean here

$$T_\nu^*(f) = \sup \{ \int_X f \, d\mu - T_\nu(\mu) ; \mu \in \mathcal{P}(X) \} = \sup \{ \int_X f \, d\mu - T(\mu, \nu) ; \mu \in \mathcal{P}(X) \}.$$ 

This is because we are assuming that $T_\nu$ and $T_\mu$ are equal to $+\infty$ whenever $\mu$ and $\nu$ are not probability measures. So, if $T$ is a forward linear transfer on $X \times Y$, then for any $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, we have

$$T(\mu, \nu) = \sup \{ \int_Y T^+ f(y) \, d\nu(y) - \int_X f(x) \, d\mu(x) ; f \in C(X) \},$$

while if $T$ is a backward linear transfer on $X \times Y$, then

$$T(\mu, \nu) = \sup \{ \int_Y g(y) \, d\nu(y) - \int_X T^- g(x) \, d\mu(x) ; g \in C(Y) \}.$$
We shall say that a transfer $T$ is symmetric if
\[
T(\nu, \mu) := T(\mu, \nu) \text{ for all } \mu \in \mathcal{P}(X) \text{ and } \nu \in \mathcal{P}(Y).
\]
Note that if $T$ is a backward linear transfer with Kantorovich operator $T^-$, then $\tilde{T}(\mu, \nu) := T(\nu, \mu)$ is a forward linear transfer with Kantorovich operator $T^+ f = -T^-(f)$. This means that if $T$ is symmetric, then $T^+ f = -T^-(f)$.

The class of linear transfers is quite large and ubiquitous in analysis. To start with, it contains all cost minimizing mass transports, that is functions on $\mathcal{P}(X) \times \mathcal{P}(Y)$ of the form,
\[
\mathcal{T}_c(\mu, \nu) := \inf \left\{ \int_{X \times Y} c(x,y) \, d\pi; \pi \in \mathcal{K}(\mu, \nu) \right\},
\]
where $c(x,y)$ is a 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. The Schrödinger bridge problem associated to a reversible Markov process [12] is also a symmetric backward and forward linear transfer.

Other examples, which are only one-directional linear transfers, are the various Martingale mass transports, the weak mass transports of Marton, Gozlan and collaborators. However, what motivated us to develop the concept of transfers are the stochastic mass transports, which do not minimize a given cost function between point particles, since the cost of transporting a Dirac measure to another is often infinite. This said, we should show however that if the set $\{\delta_x; x \in X\}$ is contained in $D_1(T)$, then we can represent such a linear transfer as a generalized mass transport, a notion recently formalized by Gozlan et al. [20].

Note that we did not specify any property on the maps $T^+$ and $T^-$. However, the fact that they arise from a Legendre transform imposes on them certain properties such as those exhibited in the following.

**Definition 2** If $X$ and $Y$ are two compact spaces, say that a map $T^- : C(Y) \to \text{LSC}(X)$ (resp., $T^+ : C(X) \to \text{USC}(Y)$) is a convex operator (resp., a concave operator), if it satisfies the following conditions:

1. If $f_1 \leq f_2$ in $C(Y)$ (resp., in $C(X)$), then $T^- f_1 \leq T^- f_2$ (resp., $T^+ f_1 \leq T^+ f_2$).

2. For any $\lambda \in [0,1]$, $f_1, f_2$ in $C(Y)$ (resp., in $C(X)$), we have
\[
T^-(\lambda f_1 + (1-\lambda)f_2) \leq \lambda T^- f_1 + (1-\lambda)T^- f_2 \quad (\text{resp., } T^+(\lambda f_1 + (1-\lambda)f_2) \geq \lambda T^+ f_1 + (1-\lambda)T^+ f_2).
\]

3. If $(f_n)_n, f$ in $C(Y)$ (resp., in $C(X)$) is such that $\|f_n - f\| \to 0$, then
\[
T^- f \leq \liminf_n T^- f_n \quad (\text{resp., } T^+ f \geq \limsup_n T^- f_n).
\]

Note that $T^-$ (resp., $T^+$) extend --with the same properties-- to operators $T^- : \text{LSC}(Y) \to \text{LSC}(X)$ (resp., $T^+ : \text{USC}(X) \to \text{LSC}(Y)$).

We leave it to the reader to check that the backward (resp., forward) maps in Definition [1] are necessarily convex (resp., concave) operators. The class of linear transfers has remarkable permanence properties. The two most important ones are stability under inf-convolution and tensorization, which allow to create an even richer class of transfers, such as the ballistic stochastic optimal transport and broken geodesics of transfers. However, a natural and an even richer family of transfers is the class of convex transfers, which are essentially supreme of linear transfers.

**Definition 3** A proper convex and weak* lower semi-continuous functional $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ is said to be a backward convex transfer (resp., forward convex transfer), if there exists a family of backward linear transfers (resp., forward linear transfers) $(T_i)_{i \in I}$ such that for all $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$,
\[
T(\mu, \nu) = \sup_{i \in I} T_i(\mu, \nu).
\]
In other words, a backward convex transfer (resp., forward convex transfer) can be written as:

$$\mathcal{T}(\mu, \nu) = \sup \left\{ \int_Y g(y) \, d\nu(y) - \int_X T_i^- g(x) \, d\mu(x); \, g \in C(Y), \, i \in I \right\},$$  \hfill (7)

respectively,

$$\mathcal{T}(\mu, \nu) = \sup \left\{ \int_Y T_i^+ f(y) \, d\nu(y) - \int_X f(x) \, d\mu(x); \, f \in C(X), \, i \in I \right\},$$  \hfill (8)

where \((T_i^-)_{i \in I}\) (resp., \((T_i^+)_{i \in I}\)) is a family of convex operators from \(C(Y) \to LSC(X)\) (resp., concave operators from \(C(X) \to LSC(Y)\)).

In addition to linear transfers, we shall see that any \(p\)-power \((p \geq 1)\) of a linear transfer is a convex transfer in the same direction. More generally, for any convex increasing real function \(\gamma\) on \(\mathbb{R}^+\) and any linear backward (resp., forward) transfer, the map \(\gamma(\mathcal{T})\) is a convex backward (resp., forward) transfer.

Note that if a \(\mathcal{T}\) is convex backward (resp., forward) transfer, then

$$\mathcal{T}_\mu = (S^-_\mu)^* \quad \text{and} \quad \mathcal{T}_\nu = (S^+_\nu)^*,$$  \hfill (9)

where \(S^-_\mu(g) = \inf_{i \in I} \int_X T_i^- g(x) \, d\mu(x)\) for \(g \in C(Y)\) and \(S^+_\nu(f) = \sup_{i \in I} \int_X T_i^+ (-f) \, d\nu(y)\) for \(f \in C(X)\).

However, we only have

$$\mathcal{T}_\mu^* \leq S^-_\mu \quad \text{and} \quad \mathcal{T}_\nu^* \leq -S^+_\nu,$$  \hfill (10)

since \(S^-_\mu\) (resp., \(S^+_\nu\)) are not necessarily convex (resp., concave). We can therefore introduce the notions of completely convex transfers for when we have equality above, that is when \(S^-_\mu\) is a convex operator (resp., \(S^+_\mu\) is concave) and \(\mathcal{T}_\mu = S^-_\mu\) (resp., \(\mathcal{T}_\nu = -S^+_\nu\)). For instance, this will be the case for the following generalized entropy,

$$\mathcal{T}(\mu, \nu) = \int_X \alpha(\frac{d\nu}{d\mu}) \, d\mu, \quad \text{if} \ \nu << \mu \text{ and } +\infty \text{ otherwise,}$$  \hfill (11)

which is a backward completely convex transfer, whenever \(\alpha\) is a strictly convex lower semi-continuous superlinear real-valued function on \(\mathbb{R}^+\). Another example of a backward completely convex transfer is the Donsker-Varadhan entropy, which is defined as

$$\mathcal{I}(\mu, \nu) := \begin{cases} \mathcal{E}(\sqrt{\mathcal{T}}, \sqrt{\mathcal{T}}), & \text{if } \mu = f \nu, \sqrt{\mathcal{T}} \in \mathbb{D}(\mathcal{E}) \\ +\infty, & \text{otherwise}, \end{cases}$$  \hfill (12)

where \(\mathcal{E}\) is a Dirichlet form with domain \(\mathbb{D}(\mathcal{E})\) on \(L^2(\nu)\).

The important example of the logarithmic entropy

$$\mathcal{H}(\mu, \nu) = \int_X \log(\frac{d\nu}{d\mu}) \, d\nu, \quad \text{if} \ \nu << \mu \text{ and } +\infty \text{ otherwise,}$$  \hfill (13)

is also another backward completely convex transfer. But it is much more as we now focus on a remarkable subset of the cone of completely convex transfers, which is the class of entropic transfers, that we define as follows:

**Definition 4** Let \(\alpha\) (resp., \(\beta\)) be a convex increasing (resp., concave increasing) real function on \(\mathbb{R}\), and let \(\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}\) be a proper (jointly) convex and weak* lower semi-continuous functional. We say that

- \(\mathcal{T}\) is a \(\beta\)-backward transfer, if there exists a convex operator \(T^- : C(Y) \to LSC(X)\) such that for each \(\mu \in D_1(\mathcal{T})\), the Legendre transform of \(\mathcal{T}_\mu\) on \(\mathcal{M}(Y)\) satisfies:

$$\mathcal{T}_\mu^*(g) = \beta \left( \int_X T^- g(x) \, d\mu(x) \right) \quad \text{for any } g \in C(Y).$$
• \( \mathcal{T} \) is a \( \alpha \)-forward transfer, if there exists a concave operator \( T^+: C(X) \rightarrow USC(Y) \) such that for each \( \nu \in D_2(\mathcal{T}) \), the Legendre transform of \( \mathcal{T}_\nu \) on \( \mathcal{M}(X) \) satisfies:

\[
\mathcal{T}_\nu(f) = -\alpha \left( \int_Y T^+(-f)(y) \, d\nu(y) \right) \quad \text{for any } f \in C(X).
\]

So, if \( \mathcal{T} \) is an \( \alpha \)-forward transfer on \( X \times Y \), then for any probability measures \( \mu \in \mathcal{P}(X) \) and \( \nu \in \mathcal{P}(Y) \), we have

\[
\mathcal{T}(\mu, \nu) = \sup \left\{ \alpha \left( \int_Y T^+ f(y) \, d\nu(y) \right) - \int_X f(x) \, d\mu(x); f \in C(X) \right\}, \quad (14)
\]

while if \( \mathcal{T} \) is a \( \beta \)-backward transfer, then

\[
\mathcal{T}(\mu, \nu) = \sup \left\{ \int_Y g(y) \, d\nu(y) - \beta \left( \int_X T^- g(x) \, d\mu(x) \right); g \in C(Y) \right\}. \quad (15)
\]

Entropic transfers are completely convex transfers. A typical example is of course the logarithmic entropy, since it can be written as

\[
\mathcal{H}(\mu, \nu) = \sup \{ \int_X f \, d\nu - \log(\int_X e^f \, d\mu); f \in C(X) \}, \quad (16)
\]

making it a log-backward transfer. The Donsker-Varadhan entropy can also be written as

\[
\mathcal{I}(\mu, \nu) = \sup \{ \int_X f \, d\nu - \log \| P(t) \|_{L^2(\mu)}; f \in C(X) \}, \quad (17)
\]

where \( P(t) \) is an associated (Feynman-Kac) semi-group of operators on \( L^2(\nu) \). More examples of \( \alpha \)-forward transfers and \( \beta \)-backward transfers with readily computable Kantorovich operators can be obtained by convolving entropic transfers with linear transfers of the same direction.

In section 7, we show how the concepts of linear and convex transfers lead naturally to more transparent proofs and vast extensions, of many well known duality formulae for transport-entropy inequalities. In the section 8, we study the limit of the inf-convolution of \( n \)-linear transfers \( \mathcal{T} \) as \( n \rightarrow +\infty \). Our approach reduces the problem to studying the limit of the iterates of the corresponding Kantorovich operator, which lead to fixed points for such a non-linear operator. Note that in the case where \( \mathcal{T} \) is the optimal mass transport minimizing a cost given by the generating function of a Lagrangian \( L \) on a compact manifold \( M \), that is

\[
c^L(y, x) := \inf \left\{ \int_0^1 L(t, \gamma(t), \dot{\gamma}(t)) \, dt; \gamma \in C^1([0, 1], M); \gamma(0) = y, \gamma(1) = x \right\}, \quad (18)
\]

the Kantorovich operator is then given by the Lax-Oleinik semi-group, whose fixed points correspond to weak KAM solutions as described by Fathi [11]. The extension of this result to general transfers allows for a similar approach for the stochastic counterpart of Mather theory. Details will follow in an upcoming paper.

2 First examples of linear mass transfers

*Example 1: The push-forward transfer*

First note that the map \( \mathcal{I} \) on \( \mathcal{P}(X) \times \mathcal{P}(X) \) defined by

\[
\mathcal{I}(\mu, \nu) = \begin{cases} 0 & \text{if } \mu = \nu \\ +\infty & \text{otherwise,} \end{cases} \quad (19)
\]

corresponds to when the Kantorovich operator is the identity. More generally, if \( T \) is a measurable map from \( X \) to \( Y \), then

\[
\mathcal{I}(\mu, \nu) = \begin{cases} 0 & \text{if } T\# \mu = \nu \\ +\infty & \text{otherwise.} \end{cases} \quad (20)
\]
is a backward linear transfer with Kantorovich operator given by $T^- f = f \circ T$.

**Example 2: The trivial Kantorovich transfer**

Any pair of functions $c_1 \in C(X)$, $c_2 \in C(Y)$ defines trivially a linear transfer via

$$T(\mu, \nu) = \int_Y c_2 \, d\nu - \int_X c_1 \, d\mu.$$  

The Kantorovich operators are then $T^+ f = c_2 + \inf(f - c_1)$ and $T^- g = c_1 + \sup(g - c_2)$.

### 2.1 Cost optimizing mass transports are backward and forward linear transfers

**Example 3: Monge-Kantorovich transfers**

Any function $c \in C(X \times Y)$ determines a backward and forward linear transfer. This is Monge-Kantorovich theory of optimal transport. One associates the map $T_c$ on $\mathcal{P}(X) \times \mathcal{P}(Y)$ to be the optimal mass transport between two probability measures $\mu$ on $X$ and $\nu$ on $Y$, that is

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

where $\mathcal{K}(\mu, \nu)$ 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). Monge-Kantorovich theory readily yields that $T_c$ is a linear transfer. Indeed, if we define the operators

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

for any $f \in C(X)$ (resp., $g \in C(Y)$), then Monge-Kantorovich duality yields that for any probability measures $\mu$ on $X$ and $\nu$ on $Y$, we have

$$T_c(\mu, \nu) = \sup \left\{ \int_Y T_c^+ f(y) \, d\nu(y) - \int_X f(x) \, d\mu(x); \, f \in C(X) \right\}$$

$$= \sup \left\{ \int_Y g(y) \, d\nu(y) - \int_X T_c^- g(x) \, d\mu(x); \, g \in C(Y) \right\}.$$  

This means that the Legendre transform $(T_c)_*(g) = \int_X T_c g(x) \, d\mu(x)$ and $T_c^-$ is the corresponding backward Kantorovich operator. Similarly, $(T_c)^-_*(f) = - \int_Y T_c^+(-f)(y) \, d\nu(y)$ on $C(X)$ and $T_c^+$ is the corresponding forward Kantorovich operator. See for example Villani [37].

**Example 4: The Csiszár-Kullback-Pinsker transfer**

This is simply the total variation distance between two probability measures $\nu$ and $\mu$ on $X$, defined by

$$\|\nu - \mu\|_{TV} = \sup \{ |\nu(A) - \mu(A)|; A \text{ measurable subset of } X \},$$  

with forward (resp., backward) Kantorovich operator given by

$$T^+ f(y) = \min \{ \inf_{x \neq y} f(x) + 1, f(y) \}, \text{ while } T^- g(x) = \max \{ \sup_{x \neq y} g(y) - 1, g(x) \}.$$  

It is actually a cost minimizing optimal transport, where the cost is given by the Hamming metric.

**Example 5: The Kantorovich-Rubinstein transfer**

If $d : X \times X \to \mathbb{R}$ is a lower semi-continuous metric on $X$, then

$$T(\mu, \nu) = \|\nu - \mu\|_{\text{Lip}}^* := \sup \left\{ \int_X u \, d(\nu - \mu); u \text{ measurable}, \|u\|_{\text{Lip}} \leq 1 \right\}.$$  

(25)
is a linear transfer, where here \( \|u\|_{\text{Lip}} := \sup_{x \neq y} \frac{|u(y) - u(x)|}{d(x,y)} \). The corresponding forward Kantorovich operator is then the Lipshitz regularization \( T^+ f(x) = \inf \{ f(y) + d(y, x); y \in X \} \), while \( T^- f(x) = \sup \{ f(y) - d(x,y); y \in X \} \). Note that \( T^+ \circ T^- f = T^- f \).

**Example 6: The Brenier-Wasserstein distance** [9]

If \( c(x, y) = \langle x, y \rangle \) on \( \mathbb{R}^d \times \mathbb{R}^d \), and \( \mu, \nu \) are two probability measures of compact support on \( \mathbb{R}^d \), then

\[
W_2(\mu, \nu) = \inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle x, y \rangle \, d\pi; \pi \in \mathcal{K}(\mu, \nu) \right\}.
\]

Here, the Kantorovich operators are

\[
T^+ f(x) = -f^*(-x) \quad \text{and} \quad T^- g(y) = (-g)^*(-y),
\]

where \( f^* \) is the convex Legendre transform of \( f \).

**Example 7: Optimal transport for a cost given by a generating function** (Bernard-Buffoni [11])

This important example links the Kantorovich backward and forward operators with the forward and backward Hopf-Lax operators that solve first order Hamilton-Jacobi equations. Indeed, on a given compact manifold \( M \), consider the cost:

\[
c_{\mu}(y, x) := \inf \left\{ \int_{0}^{1} L(t, \gamma(t), \dot{\gamma}(t)) \, dt; \gamma \in C^1([0, 1], M); \gamma(0) = y, \gamma(1) = x \right\},
\]

where \([0, 1]\) is a fixed time interval, and \( L : TM \to \mathbb{R} \cup \{+\infty\} \) is a given Tonelli Lagrangian that is convex in the second variable of the tangent bundle \( TM \). If now \( \mu \) and \( \nu \) are two probability measures on \( M \), then

\[
T_{c,\mu}(\mu, \nu) := \inf \left\{ \int_{M \times M} c_{\mu}(y, x) \, d\pi; \pi \in \mathcal{K}(\mu, \nu) \right\}
\]

is a linear transfer with forward Kantorovich operator given by \( T^+_1 f(x) = V_f(1, x) \), where \( V_f(t, x) \) being the value functional

\[
V_f(t, x) = \inf \left\{ f(\gamma(0)) + \int_{0}^{t} L(s, \gamma(s), \dot{\gamma}(s)) \, ds; \gamma \in C^1([0, 1], M); \gamma(t) = x \right\}.
\]

Note that \( V_f \) is –at least formally– a solution for the Hamilton-Jacobi equation

\[
\begin{align*}
\partial_t V + H(t, x, \nabla_x V) &= 0 \text{ on } [0, 1] \times M, \\
V(0, x) &= f(x).
\end{align*}
\]

Similarly, the backward Kantorovich potential is given by \( T^-_1 g(y) = W_g(0, y) \), \( W_g(t, y) \) being the value functional

\[
W_g(t, y) = \sup \left\{ g(\gamma(1)) - \int_{t}^{1} L(s, \gamma(s), \dot{\gamma}(s)) \, ds; \gamma \in C^1([0, 1], M); \gamma(t) = y \right\},
\]

which is a solution for the backward Hamilton-Jacobi equation

\[
\begin{align*}
\partial_t W + H(t, x, \nabla_x W) &= 0 \text{ on } [0, 1] \times M, \\
W(1, y) &= g(y).
\end{align*}
\]

**2.2 One-sided linear transfers arising from constrained mass transports**

We now give examples of linear transfers, which do not fit in the framework of Monge-Kantorovich theory. Cost minimizing mass transport with additional constraints give examples of one-directional linear transfers.
We single out the following:

**Example 8: Martingale transports are backward linear transfers** (Henri-Labordère [21], Ghoussoub-Kim-Lim [13])

Consider $MT(\mu, \nu)$ to be the subset of $\mathcal{K}(\mu, \nu)$ consisting of the martingale transport plans, that is the set of probabilities $\pi$ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mu$ and $\nu$, such that for $\mu$-almost $x \in \mathbb{R}^d$, the component $\pi_x$ of its disintegration $(\pi_x)_{x}$ with respect to $\mu$, i.e. $d\pi(x, y) = d\pi_x(y) d\mu(x)$, has its barycenter at $x$. One can also use the probabilistic notation, which amounts to minimize $E_P c(X, Y)$ over all marginals $(X, Y)$ on a probability space $(\Omega, \mathcal{F}, P)$ into $\mathbb{R}^d \times \mathbb{R}^d$ (i.e. $E[Y|X] = X$) with laws $X \sim \mu$ and $Y \sim \nu$ (i.e., $P(X \in A) = \mu(A)$ and $P(Y \in A) = \nu(A)$ for all Borel set $A$ in $\mathbb{R}^d$). Note that in this case, the disintegration of $\pi$ can be written as the conditional probability $\pi_x(A) = \mathbb{P}(Y \in A|X = x)$.

Note that the set $MT(\mu, \nu)$ of martingale transports can be empty, and a classical theorem of Strassen [34] states that it is not if and only if the marginals $\mu$ and $\nu$ are in *convex order*, that is if

$$\int_{\mathbb{R}^d} \varphi \, d\mu \leq \int_{\mathbb{R}^d} \varphi \, d\nu \text{ for every convex function } \varphi \text{ on } \mathbb{R}^d. \tag{32}$$

In that case we will write $\mu \ll_C \nu$, which is sometimes called the *Choquet order for convex functions*. Note that $x$ is the barycenter of a measure $\nu$ if and only if $\delta_x \ll_C \nu$, where $\delta_x$ is Dirac measure at $x$.

If now $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a continuous cost function, then the corresponding martingale transport is a *backward linear transfer* in the following way:

$$T_M(\mu, \nu) = \left\{ \inf_{+\infty} \{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \, d\pi(x, y); \pi \in MT(\mu, \nu) \} \right\} \text{ if } \mu \ll_C \nu \text{ and not.} \tag{33}$$

The backward Kantorovich operator is then given by

$$T^-_M f(x) = f_{c,x}(x), \text{ where } f_{c,x} \text{ is the concave envelope of the function } y \to f(y) + c(x, y),$$

in such a way that $T_M(\mu, \nu) = \sup \{ \int_X f \, d\nu - \int_X T^-_M f \, d\mu; f \in C(X) \}$.

**Example 9: General stochastic transports are backward linear transfers** (Mikami-Thieulin [29])

Given a Lagrangian $L : [0, 1] \times M \times M^* \to \mathbb{R}$, we define the following stochastic counterpart of the optimal transportation problem mentioned above.

$$T_L(\mu, \nu) := \inf \left\{ \mathbb{E} \left[ \int_0^1 L(t, X(t), \beta_X(t, X(t))) \, dt \right] \bigg| X(0) \sim \mu, X(1) \sim \nu, X(\cdot) \in \mathcal{A} \right\} \tag{34}$$

Here $\mathcal{A}$ refers to the set of $\mathbb{R}^d$-valued continuous semimartingales $X(\cdot)$ such that there exists a measurable drift $\beta_X : [0, T] \times C([0, 1]) \to M^*$ where

- $\omega \mapsto \beta_X(t, \omega)$ is $\mathcal{B}(C([0, t]))$-measurable for all $t$.
- $W(t) := X(t) - X(0) - \int_0^t \beta_X(s, X) \, ds$ is a $\sigma[X(s) : s \in [0, t]]$-Brownian motion.

This stochastic transport does not fit in the standard optimal mass transport theory since it does not originate in optimization a cost between two deterministic states. However, under certain conditions on the Lagrangian, Mikami and Thieulin [29] proved that the map $(\mu, \nu) \to T_L(\mu, \nu)$ is jointly convex and weak*-lower semi-continuous on the space of measures and that

$$T_L(\mu, \nu) = \sup \left\{ \int_M f(x) \, d\nu - \int_M V_f(0, x) \, d\mu; f \in C^\infty_b \right\}, \tag{35}$$

where $V_f$ solves the Hamilton-Jacobi-Bellman equation

$$\frac{\partial V}{\partial t} + \frac{1}{2} \Delta V(t, x) + H(t, x, \nabla V) = 0, \quad V(1, x) = f(x). \quad \text{(HJB)}$$
In other words, $T_L$ is a backward linear transfer with a Kantorovich operator being $T_L f = V_f(0, \cdot)$, where $V_f(t, x)$ can be written as

$$V_f(t, x) = \sup_{X \in \mathcal{A}} \left\{ \mathbb{E} \left[ f(X(1)) - \int_0^1 L(s, X(s), \beta X(s, X)) \, ds \, \bigg| X(t) = x \right] \right\}.$$  \hspace{1cm} (36)

**Example 10: Optimally stopped stochastic transport are backward linear transfers** (Ghoussoub-Kim-Palmer [14, 15])

Consider the optimal stopping problem

$$T_L(\mu, \nu) = \inf \left\{ \mathbb{E} \left[ \int_0^T L(t, X(t), \beta X(t, X(t))) \, dt \right] ; X(0) \sim \mu, T \in \mathcal{S}, X(T) \sim \nu, X(\cdot) \in \mathcal{A} \right\},\hspace{1cm} (37)$$

where $\mathcal{S}$ is the set of possibly randomized stopping times. In this case, $T_L$ is a backward linear transfer with Kantorovich potential given by $T_L f = \hat{V}_f(0, \cdot)$, where

$$\hat{V}_f(t, x) = \sup_{X \in \mathcal{A}} \sup_{T \in \mathcal{S}} \left\{ \mathbb{E} \left[ f(X(T)) - \int_t^T L(s, X(s), \beta X(s, X)) \, ds \bigg| X(t) = x \right] \right\},$$  \hspace{1cm} (38)

which is --at least formally-- a solution $\hat{V}_f(t, x)$ of the quasi-variational Hamilton-Jacobi-Bellman inequality,

$$\min \left\{ V_f(t, x) - f(x), -\partial_t V_f(t, x) - H(t, x, \nabla V_f(t, x)) - \frac{1}{2} \Delta V_f(t, x) \right\} = 0.\hspace{1cm} (39)$$

### 2.3 Weak optimal transports

Other examples of linear transfers arise from the work of Marton, who extended the work of Talagrand.

**Example 11: Marton transports are linear transfers** (Marton [24, 25])

These are transports of the following type:

$$\tilde{T}_\gamma, d(\mu, \nu) = \inf \left\{ \int_X \gamma \left( \int_Y d(x, y) d\pi_x(y) \right) d\mu(x) ; \pi \in \mathcal{K}(\mu, \nu) \right\},\hspace{1cm} (40)$$

where $\gamma$ is a convex function on $\mathbb{R}^+$ and $d : X \times Y \rightarrow \mathbb{R}$ is a lower semi-continuous functions. Marton’s weak transfer correspond to $\gamma(t) = t^2$ and $d(x, y) = |x - y|$, which in probabilistic terms reduces to

$$\tilde{T}_2(\mu, \nu) = \inf \left\{ \mathbb{E}[\mathbb{E}[|X - Y|] | Y^2] ; X \sim \mu, Y \sim \nu \right\}.$$  \hspace{1cm} (41)

This is a backward linear transfer with Kantorovich potential

$$T^- f(x) = \sup \left\{ \int_Y f(y) d\sigma(y) - \gamma \left( \int_Y d(x, y) d\sigma(y) \right) ; \sigma \in \mathcal{P}(Y) \right\}.$$  \hspace{1cm} (42)

**Example 12: A barycentric cost function** (Gozlan et al. [20])

Consider the transport

$$T(\mu, \nu) = \inf \left\{ \int_X \| x - \int_Y y d\pi_x(y) \| d\mu(x) ; \pi \in \mathcal{K}(\mu, \nu) \right\}.$$  \hspace{1cm} (42)

Again, this is a backward linear transfer, with Kantorovich potential

$$T^- f(x) = \sup \{ f_\ast(y) - \| y - x \| ; y \in \mathbb{R}^n \}.\hspace{1cm} (42)$$
Moreover, even if the set of Dirac measures \( \{ \delta_x \} \) transfers need not be defined on Dirac measures, a prevalent situation in stochastic transport problems. We now consider whether any transfer \( T \) on \( X \times Y \) arises from a cost minimizing mass transport. Note first that transfers need not be defined on Dirac measures, a prevalent situation in stochastic transport problems. Moreover, even if the set of Dirac measures \( \{ \delta_x \} \) transfers need not be defined on Dirac measures, a prevalent situation in stochastic transport problems.

**Example 13: Schrödinger bridge (Gentil-Leonard-Ripani [12])**

Fix some reference non-negative measure \( R \) on path space \( \Omega = C([0,1], \mathbb{R}^n) \), and let \( (X_t) \) be a random process on \( \mathbb{R}^n \) whose law is \( R \). Denote by \( R_{t1} \) the joint law of the initial position \( X_0 \) and the final position \( X_1 \), that is \( R_{t1} = (X_0, X_1) \# R \). For probability measures \( \mu, \nu \) on \( \mathbb{R}^n \), the maximum entropy formulation of the Schrödinger bridge problem between \( \mu \) and \( \nu \) is defined as

\[
S_R(\mu, \nu) = \inf \int_{\mathbb{R}^n \times \mathbb{R}^n} \log(\frac{d\pi}{dR_{t1}}) \, dx; \pi \in \mathcal{K}(\mu, \nu).
\]  

(43)

For example (See [12]), assume \( R \) is the reversible Kolmogorov continuous Markov process associated with the generator \( \frac{1}{2}(\Delta - \nabla V \cdot \nabla) \) and the initial measure \( m = e^{-V(x)}dx \) for some function \( V \). Then, under appropriate conditions on \( V \) (e.g., if \( V \) is uniformly convex), then

\[
\mathcal{T}(\mu, \nu) = \mathcal{S}_R(\mu, \nu) - \frac{1}{2} \int_{\mathbb{R}^n} \log(\frac{d\mu}{dm}) \, dm - \frac{1}{2} \int_{\mathbb{R}^n} \log(\frac{d\nu}{dm}) \, dm
\]

is a forward linear transfer with Kantorovich operator

\[
T^+ f(x) = \log E_{R_{t1}} e^{f(X_t)} = \log S_1(e^f)(x),
\]

where \((S_t)\) is the semi-group associated to \( R \). It is worth noting that \( \mathcal{T} \) is symmetric, that is \( \mathcal{T}(\mu, \nu) = \mathcal{T}(\nu, \mu) \), which means that it is also a backward linear transfer. Note that when \( V = 0 \), the process is Brownian motion with Lebesgue measure as its initial reversing measure, while when \( V(x) = \frac{|x|^2}{2} \), \( R \) is the path measure associated with the Ornstein-Uhlenbeck process with the Gaussian as its initial reversing measure.

### 3 A representation of linear transfers as generalized optimal mass transports

We now consider whether any transfer \( \mathcal{T} \) on \( X \times Y \) arises from a cost minimizing mass transport. Note first that transfers need not be defined on Dirac measures, a prevalent situation in stochastic transport problems. Moreover, even if the set of Dirac measures \( \{ \delta_x, \delta_y \}; (x, y) \in X \times Y \} \subset D(\mathcal{T}) \), and we can then define a cost function as \( c(x, y) = \mathcal{T}(\delta_x, \delta_y) \), and its associated optimal mass transport \( \mathcal{T}_c(\mu, \nu) \), we then only have

\[
\mathcal{T}_c(\mu, \nu) \geq \mathcal{T}(\mu, \nu).
\]  

(44)

Indeed, for every \( x \in X \), we have

\[
T^{-} g(x) = \mathcal{T}_{\delta_x}^*(g) = \sup \{ \int_Y g(d\nu - \mathcal{T}(\delta_x, \nu); \nu \in \mathcal{P}(Y)) \geq \sup \{ g(y) - c(x, y); y \in Y \} = T_{\delta_x}^{-} g(x),
\]

hence,

\[
\mathcal{T}(\mu, \nu) = \sup \{ \int_Y g(d\nu - \int_X T^{-} g(d\mu); g \in C(Y)) \leq \sup \{ \int_Y g(d\nu - \int_X T_{\delta_x}^{-} g(d\mu); g \in C(Y)] = \mathcal{T}_c(\mu, \nu).
\]

Moreover, the inequality (44) is often strict.

Motivated by the work of Marton and others, Gozlan et al. [20] introduced the notion of weak transport. It consists of considering cost minimizing transport plans, where cost functions between two points are replaced by generalized costs \( c \) on \( X \times \mathcal{P}(Y) \), where \( \sigma \to c(x, \sigma) \) is convex and lower semi-continuous. As the following proposition shows, this notion turns out to be equivalent to the notion of backward linear transfer, at least in the case where Dirac measures belong to the first partial effective domain of the map \( \mathcal{T} \), that is when \( \{ \delta_x; x \in X \} \subset D_1(\mathcal{T}) \).
Proposition 5 Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a functional such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$. Then, $\mathcal{T}$ is a backward linear transfer if and only if there exists a lower semi-continuous function $c : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ with $\sigma \to c(x, \sigma)$ convex on $\mathcal{P}(Y)$ for each $x \in X$ such that for every $\mu \in \mathcal{P}(X)$, and $\nu \in \mathcal{P}(Y)$, we have

$$T(\mu, \nu) = \inf_\pi \left\{ \int_X c(x, \pi_x) \, d\nu(x); \pi \in \mathcal{K}(\mu, \nu) \right\}. \quad (45)$$

The corresponding backward Kantorovich operator is given for every $g \in C(Y)$ by

$$T^- g(x) = \sup_\sigma \left\{ \int_Y g(y) \, d\sigma(y) - T(x, \sigma); \sigma \in \mathcal{P}(Y) \right\}. \quad (46)$$

Note that we have identified here any $\pi \in \mathcal{K}(\mu, \nu)$ with its disintegration that gives a probability kernel $\pi : X \to \mathcal{P}(X)$ such that $\nu(A) = \int_X \pi_x(A) \, d\mu(x)$. 

**Proof:** Consider first a lower semi-continuous function $c : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ with $\sigma \to c(x, \sigma)$ convex on $\mathcal{P}(Y)$ for each $x \in X$, and let

$$T_c(\mu, \nu) := \inf_\pi \left\{ \int_X c(x, \pi_x) \, d\mu(x); \pi \in \mathcal{K}(\mu, \nu) \right\}. \quad (47)$$

We first prove that $T_c$ is a backward linear transfer with a Kantorovich operator given by

$$T^-_c g(x) = \sup_\sigma \left\{ \int_Y g(y) \, d\sigma(y) - c(x, \sigma); \sigma \in \mathcal{P}(Y) \right\}. \quad (48)$$

This will then imply that if $\mathcal{T}$ is any backward linear transfer with Kantorovich operator $T^-$, and if $c(x, \sigma) = \mathcal{T}(\delta_x, \sigma)$, then $T^- g(x) = T^-_c g(x)$ and therefore $T(\mu, \nu) = T_c(\mu, \nu)$.

First, it is easy to show that $T_c$ is a convex lower semi-continuous function on $\mathcal{P}(X) \times \mathcal{P}(Y)$. Consider now the Legendre transform of $(T_c)_\mu$, that is

$$(T_c)_\mu^*(g) = \sup_\nu \left\{ \int_Y g \, d\nu - T_c(\mu, \nu); \nu \in \mathcal{P}(Y) \right\} \quad (49)$$

$$= \sup_\nu \left\{ \int_Y g(y) \, d\nu(y) - \int_X c(x, \pi_x) \, d\mu(x); \nu \in \mathcal{P}(Y), \pi \in \mathcal{K}(\mu, \nu) \right\} \quad (50)$$

$$= \sup_X \left\{ \int_Y g(y) \, d\pi_x(y) \, d\mu(x) - \int_X c(x, \pi_x) \, d\mu(x); \pi \in \mathcal{K}(\mu, \nu) \right\} \quad (51)$$

$$\leq \sup_X \left\{ \int_Y g(y) \, d\sigma(y) \, d\mu(x) - \int_X c(x, \sigma) \, d\mu(x); \sigma \in \mathcal{P}(Y) \right\} \quad (52)$$

$$\leq \int_X \left\{ \sup_\sigma \left\{ \int_Y g(y) \, d\sigma(y) - c(x, \sigma) \right\} \right\} \, d\mu(x) \quad (53)$$

$$= \int_X T^-_c g(x) \, d\mu(x). \quad (54)$$

On the other hand, use your favorite selection theorem to find a measurable selection $x \to \bar{\pi}_x$ from $X$ to $\mathcal{P}(Y)$ such that

$$T^-_c g(x) = \int_Y g(y) \, d\bar{\pi}_x(y) - c(x, \pi_x) \quad \text{for every } x \in X. \quad (55)$$

It follows that

$$(T_c)_\mu^*(g) = \sup_\nu \left\{ \int_Y g \, d\nu - T_c(\mu, \nu); \nu \in \mathcal{P}(Y) \right\} \quad (56)$$

$$\geq \int_X \left\{ \int_Y g(y) \, d\pi_x(y) - c(x, \pi_x) \right\} \, d\mu(x) \quad (57)$$

$$= \int_X T^-_c g(x) \, d\mu(x), \quad (58)$$

which completes the proof.
hence \((T_c)_\ast^\ast(g) = \int_X T_c^{-} g(x) d\mu(x)\) and \(T^{-} = T_c^{-}\).

Conversely, if \(\mathcal{T}\) is a backward linear transfer with \(T^{-}\) as a Kantorovich operator, then by setting \(c(\delta_x, \sigma) = \mathcal{T}(\delta, \sigma)\), we have \(T^{-} = T_c^{-} g\) and we are done.

## 4 Operations on linear mass transfers

Denote by \(\mathcal{LT}_-(X \times Y)\) (resp., \(\mathcal{LT}_+(X \times Y)\)) the class of backward (resp., forward) linear transfers on \(X \times Y\). The following proposition is an immediate consequence of the properties of the Legendre transform.

**Proposition 6** The class \(\mathcal{LT}_-(X \times Y)\) (resp., \(\mathcal{LT}_+(X \times Y)\)) is a convex cone in the space of convex weak* lower continuous functions on \(\mathcal{P}(X) \times \mathcal{P}(Y)\).

1. **(Scalar multiplication)** If \(a \in \mathbb{R}^+\setminus\{0\}\) and \(\mathcal{T}\) is a backward linear transfer with Kantorovich operator \(T^{-}\), then the transfer \((a\mathcal{T})\) defined by \((a\mathcal{T})(\mu, \nu) = a\mathcal{T}(\mu, \nu)\) is also a backward linear transfer with Kantorovich operator on \(C(Y)\) defined by,

\[
T^{-}_a(f) = aT^{-}\left(\frac{f}{a}\right). \tag{48}
\]

2. **(Addition)** If \(\mathcal{T}_1\) and \(\mathcal{T}_2\) are backward linear transfers on \(X \times Y\) with Kantorovich operator \(T_1^{-}\), \(T_2^{-}\) respectively, and such that \(X \subset D(\mathcal{T}_1) \cap D(\mathcal{T}_2)\), then \((\mathcal{T}_1 + \mathcal{T}_2)(\mu, \nu) := \mathcal{T}_1(\mu, \nu) + \mathcal{T}_2(\mu, \nu)\) is a backward linear transfer on \(X \times Y\), with Kantorovich operator given on \(C(Y)\) by

\[
T^{-}_r = \inf\{T_1^{-}(g) + T_2^{-}(f-g)(x) ; g \in C(Y)\}. \tag{49}
\]

Similar statements hold for \(\mathcal{LT}_+ f(X \times Y)\).

**Definition 7** Consider the following operations on transfers.

1. **(Inf-convolution)** Let \(X_1, X_2, X_3\) be 3 spaces, and suppose \(\mathcal{T}_1\) (resp., \(\mathcal{T}_2\)) are functionals on \(\mathcal{P}(X_1) \times \mathcal{P}(X_2)\) (resp., \(\mathcal{P}(X_2) \times \mathcal{P}(X_3)\)). The convolution of \(\mathcal{T}_1\) and \(\mathcal{T}_2\) is the functional on \(\mathcal{P}(X_1) \times \mathcal{P}(X_3)\) given by

\[
\mathcal{T}(\mu, \nu) := \mathcal{T}_1 \ast \mathcal{T}_2 = \inf\{\mathcal{T}_1(\mu, \sigma) + \mathcal{T}_2(\sigma, \nu) ; \sigma \in \mathcal{P}(X_2)\}. \tag{50}
\]

2. **(Tensor product)** If \(\mathcal{T}_1\) (resp., \(\mathcal{T}_2\)) are functionals on \(\mathcal{P}(X_1) \times \mathcal{P}(Y_1)\) (resp., \(\mathcal{P}(X_2) \times \mathcal{P}(Y_2)\)) such that \(X_1 \subset D(\mathcal{T}_1)\) and \(X_2 \subset D(\mathcal{T}_2)\), then the tensor product of \(\mathcal{T}_1\) and \(\mathcal{T}_2\) is the functional on \(\mathcal{P}(X_1 \times X_2) \times \mathcal{P}(Y_1 \times Y_2)\) defined by:

\[
\mathcal{T}_1 \otimes \mathcal{T}_2(\mu, \nu) = \inf \left\{ \int_{X_1 \times X_2} \left( \mathcal{T}_1(x_1, \pi_{x_1, x_2}) + \mathcal{T}_2(x_2, \pi_{x_1, x_2}) \right) d\mu(x_1, x_2) ; \pi \in \mathcal{K}(\mu, \nu) \right\}. \tag{51}
\]

The following easy proposition is important to what follows.

**Proposition 8** The following stability properties hold for the class of backward linear transfers.

1. If \(\mathcal{T}_1\) (resp., \(\mathcal{T}_2\)) is a backward linear transfer on \(X_1 \times X_2\) (resp., on \(X_2 \times X_3\)) with Kantorovich operator \(T_1^{-}\) (resp., \(T_2^{-}\)), then \(\mathcal{T}_1 \ast \mathcal{T}_2\) is also a backward linear transfer on \(X_1 \times X_3\) with Kantorovich operator equal to \(T_1^{-} \circ T_2^{-}\).

2. If \(\mathcal{T}_1\) (resp., \(\mathcal{T}_2\)) is a backward linear transfer on \(X_1 \times Y_1\) (resp., \(X_2 \times Y_2\)) such that \(X_1 \subset D(\mathcal{T}_1)\) and \(X_2 \subset D(\mathcal{T}_2)\), then \(\mathcal{T}_1 \otimes \mathcal{T}_2\) is a backward linear transfer on \((X_1 \times X_2) \times (Y_1 \times Y_2)\), with Kantorovich operator given by

\[
T^{-}_1 g(x_1, x_2) = \sup \left\{ \int_{Y_1 \times Y_2} f(y_1, y_2) d\sigma(y_1, y_2) - \mathcal{T}_1(x_1, \sigma_1) - \mathcal{T}_2(x_2, \sigma_2) ; \sigma \in \mathcal{K}(\sigma_1, \sigma_2) \right\}. \tag{51}
\]
Moreover,
\[ T_1 \otimes T_2 (\mu, \nu_1 \otimes \nu_2) \leq T_1 (\mu_1, \nu_1) + \int_{X_1} T_2 (\mu_2, \nu_2) \, d\mu_1(x_1), \]

where \( d\mu(x_1, x_2) = d\mu_1(x_1) d\mu_2(x_2) \).

Note a similar statement holds for forward linear transfers, modulo order reversals. For example, if \( T_1 \) and \( T_2 \) are forward linear transfer, then \( T_1 \ast T_2 \) is a forward linear transfer on \( X_1 \times X_2 \) with Kantorovich operator equal to \( T_2^+ \circ T_1^+ \).

**Proof:** For 1), we note first that if \( T_1 \) (resp., \( T_2 \)) 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 both \( (T_1 \ast T_2)_{\mu} : \mu \to (T_1 \ast T_2)(\mu, \nu) \) and \( (T_1 \ast T_2)_{\nu} : \nu \to (T_1 \ast T_2)(\mu, \nu) \) are convex and weak*-lower semi-continuous. We now calculate their Legendre transform. For \( g \in C(X_3) \),

\[
(T_1 \ast T_2)^*_\mu (g) = \sup_{\nu \in \mathcal{P}(X_3)} \sup_{\sigma \in \mathcal{P}(X_2)} \left\{ \int_{X_3} g \, d\nu - T_1 (\mu, \sigma) - T_2 (\sigma, \nu) \right\}
= \sup_{\sigma \in \mathcal{P}(X_2)} \{ (T_2)^*_\sigma (g) - T_1 (\mu, \sigma) \}
= \sup_{\sigma \in \mathcal{P}(X_2)} \left\{ \int_{X_2} T^{-}_2 (g) \, d\sigma - T_1 (\mu, \sigma) \right\}
= (T_1)^*_\mu (T^{-}_2 (g))
= \int_{X_1} T^{-}_1 \circ T^{-}_2 g \, d\mu.
\]

In other words, \( T_1 \ast T_2 (\mu, \nu) = \sup \left\{ \int_{X_3} f(x) \, d\mu(x) - \int_{X_1} T^{-}_1 \circ T^{-}_2 g \, d\mu; \, f \in C(X_3) \right\} \).

2) follows immediately from the last section since we are defining the tensor product as a generalized cost minimizing transport, where the cost ion \( X_1 \times X_2 \times \mathcal{P}(Y_1 \times Y_2) \) is simply,

\[ T((x_1, x_2), \pi) = T_1 (x_1, \pi_1) + T_2 (x_1, \pi_2), \]

where \( \pi_1, \pi_2 \) are the marginals of \( \pi \) on \( Y_1 \) and \( Y_2 \) respectively. \( T_1 \otimes T_2 \) is clearly its corresponding backward transfer with \( T^{-} \) being its Kantorovich operator.

More notationally cumbersome but straightforward is how to write the Kantorovich operators of the tensor product \( T^{-} g(x_1, x_2) \) in terms of \( T^{-}_1 \) and \( T^{-}_2 \), in order to establish [52].

**Example 14:** Stochastic ballistic transfer (Barton-Ghoussoub [3])

Consider the stochastic ballistic transportation problem defined as:

\[ B(\mu, \nu) := \inf \left\{ \mathbb{E} \left[ (V, X(0)) + \int_0^T L(t, X, \beta X(t, X)) \, dt \right] \mid V \sim \mu, X(\cdot) \in \mathcal{A}, X(T) \sim \nu \right\}, \]

where we are using the notation of Example 9. Note that this a convolution of the Brenier-Wasserstein transfer of Example 6 with the general stochastic transfer of Example 9. Under suitable conditions on \( L \), one gets that

\[ B(\mu, \nu) = \sup \left\{ \int g \, d\nu - \int \tilde{\psi}_g \, d\mu; \, g \in C_b \right\}, \]

where \( \tilde{h} \) is the concave legendre transform of \( -h \) and \( \tilde{\psi}_g \) is the solution to the Hamilton-Jacobi-Bellman equation

\[ \frac{\partial \psi}{\partial t} + \frac{1}{2} \Delta \psi(t, x) + H(t, x, \nabla \psi) = 0, \quad \psi(1, x) = g(x). \] (HJB)

In other words, \( B \) is a backward linear transform with Kantorovich operator \( T^{-} g = \tilde{\psi}_g \).
Remark 1 (Lifting convolutions to Wasserstein space) Let $X_0, X_1, \ldots, X_n$ be compact spaces, and suppose for each $i = 1, \ldots, n$, we have a cost function $c_i : X_{i-1} \times X_i$, we consider the following cost function on $X_0 \times X_n$, defined by

$$c(y, x) = \inf \left\{ c_1(y, x_1) + c_2(x_1, x_2) + \ldots + c_n(x_{n-1}, x) \mid x_1 \in X_1, x_2 \in X_2, \ldots, x_{n-1} \in X_{n-1} \right\}.$$ 

Let $\mu$ (resp., $\nu$ be probability measures on $X_0$ (resp., $X_n$).

1. The following then holds on Wasserstein space:

$$\tau_c(\mu, \nu) = \inf \{ \tau_{c_1}(\mu, \nu_1) + \tau_{c_2}(\nu_1, \nu_2) + \ldots + \tau_{c_n}(\nu_{n-1}, \nu) \mid \nu_1 \in \mathcal{P}(X_i), i = 1, \ldots, n - 1 \},$$

and the infimum is attained at $\bar{\nu}_1, \bar{\nu}_2, \ldots, \bar{\nu}_{n-1}$.

2. The following duality formula holds:

$$\tau_c(\mu, \nu) = \sup \left\{ \int_{X_n} T^{+}_{c_n} \circ T^{+}_{c_{n-1}} \circ \ldots \circ T^{+}_{c_1} f(x) \, d\nu(x) - \int_{X_0} f(y) \, d\mu(y) \mid f \in C(X_0) \right\}$$

$$= \sup \left\{ \int_{X_n} g(x) \, d\nu(x) - \int_{X_0} T^{-}_{c_n} \circ T^{-}_{c_{n-1}} \ldots \circ T^{-}_{c_1} g(x) \mid g \in C(X_n) \right\}.$$

We note a few more elementary properties of linear transfers.

Proposition 9 Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to [0, \infty]$ be a proper (jointly) convex and weak* lower semi-continuous on $\mathcal{M}(X) \times \mathcal{M}(Y)$. If $\mathcal{T}$ is both a forward and backward linear transfer, and if $\{(x, \delta_{y_1}, \delta_{y_2}, \ldots, \delta_{y_n}) \mid (x, y_1, y_2, \ldots, y_n) \in X \times Y \}$ then for any $g \in C(Y)$ such that $T^+ f \in C(X)$ (resp., $f \in C(X)$ such that $T^- f \in C(Y)$)

$$T^+ \circ T^- g(y) \geq g(y) \quad \text{for } y \in Y,$$

and

$$T^- \circ T^- \circ T^+ g = T^- g \text{ and } T^+ \circ T^- \circ T^+ g = T^+ g.$$ 

Proof: Write for $\nu \in \mathcal{P}(Y)$,

$$\int_Y T^+ \circ T^- g \, d\nu = -T^*_\nu(-T^- g)$$

$$= -\sup \left\{ -\int_X T^*_\delta_x(g) \, d\mu(x) - \mathcal{T}(\mu, \nu) \mid \mu \in \mathcal{P}(X) \right\}$$

$$= \inf \left\{ \int_X T^{*}_{\delta_x}(g) \, d\mu(x) + \mathcal{T}(\mu, \nu) \mid \mu \in \mathcal{P}(X) \right\}$$

$$\geq \inf \left\{ \int_X T^{*}_{\delta_x}(g) \, d\mu(x) + \int_Y g \, d\nu - \int_X T^*_{\delta_x}(g) \, d\mu \mid \mu \in \mathcal{P}(X) \right\}$$

$$= \int_Y g \, d\nu.$$

The last item follows from the above and the positivity property of the Kantorovich operators.

Remark 2 By analogy with the case of cost optimizing mass transports, and assuming that a transfer is both forward and backward, we can say that a function $f \in C(X)$ is $\mathcal{T}$-concave if it is of the form $f = T^- \circ T^+ g$ for some $g \in C(X)$. It follows from the last proposition that

$$\mathcal{T}(\mu, \nu) = \sup \left\{ \int_Y T^+ f(y) \, d\nu(y) - \int_X f(x) \, d\mu(x) \mid f \in C(X), f \text{ is } \mathcal{T}\text{-concave} \right\},$$

Similarly, we can say that a function $g \in C(Y)$ is $\mathcal{T}$-convex if it is of the form $g = T^+ \circ T^- f$ for some $f \in C(Y)$. In this case,

$$\mathcal{T}(\mu, \nu) = \sup \left\{ \int_X g(y) \, d\nu(y) - \int_Y T^- g \, d\mu(x) \mid g \in C(Y), g \text{ is } \mathcal{T}\text{-convex} \right\},$$

14
5 Examples of convex and entropic transfers

We now give a few examples of convex and entropic transfers, which are not necessarily linear transfers. First, recall that the increasing Legendre transform (resp., decreasing Legendre transform) of a function $\alpha : \mathbb{R}^+ \to \mathbb{R}$ (resp., $\beta : \mathbb{R}^+ \setminus \{0\} \to \mathbb{R}$) is defined as

$$\alpha^\#(t) = \sup\{ ts - \alpha(s); s \geq 0 \} \quad \text{resp.,} \quad \beta^\#(t) = \sup\{ -ts + \beta(s); s > 0 \}$$  \hspace{1cm} (60)

By extending $\alpha$ to the whole real line by setting $\alpha(t) = +\infty$ if $t < 0$, and using the standard Legendre transform, one can easily show that $\alpha$ is convex increasing on $\mathbb{R}^+$ if and only if $\alpha^\#$ is convex and increasing on $\mathbb{R}^+$. We then have the following reciprocal formula

$$\alpha(t) = \sup\{ ts - \alpha^\#(s); s \geq 0 \}. \hspace{1cm} (61)$$

Similarly, if $\beta$ is convex decreasing on $\mathbb{R}^+ \setminus \{0\}$, we have

$$\beta(t) = \sup\{ -ts + \beta^\#(s); s > 0 \}. \hspace{1cm} (62)$$

Proposition 10 Let $\alpha : \mathbb{R}^+ \to \mathbb{R}$ (resp., $\beta : \mathbb{R}^+ \setminus \{0\} \to \mathbb{R}$) be a convex (resp., concave) increasing functions.

1. If $T$ is a linear backward (resp., forward) transfer with convex Kantorovich operator $T^-$ (resp., concave Kantorovich operator $T^+$), then $\alpha(T)$ is a backward convex transfer (resp., forward convex transfer) with Kantorovich operators $(T^-)_{s > 0}$ (resp., $(T^+)_{s > 0}$), where

$$T^- f = sT^-(\frac{f}{s}) - \alpha^\#(s) \quad \text{(resp.,} \quad T^+ f = sT^+(\frac{f}{s}) - \alpha^\#(s).$$  \hspace{1cm} (63)

In particular, for any $p \geq 1$, $T^p$ is a convex forward (resp., backward) transfer.

2. If $E$ is a $\beta$-backward transfer with Kantorovich operator $E^-$, then it is a backward convex transfer with Kantorovich operators $(T^-)_{s > 0}$ given by

$$T^- f = sT^- f + (-\beta)^\#(s).$$  \hspace{1cm} (64)

3. Similarly, if $E$ is an $\alpha$-forward transfer with Kantorovich operator $E^+$, then it is a forward convex transfer with Kantorovich operators $(T^+)_{s > 0}$ given by

$$T^+ f = sT^+ f - \alpha^\#(s).$$  \hspace{1cm} (65)

Proof: For 1) it suffices to write

$$\alpha(T(\mu, \nu)) = \sup\{ s \int_Y T^+ f \, d\nu - s \int_X f \, d\mu - \alpha^\#(s); \ s \in \mathbb{R}^+, \ f \in C(X) \}$$

$$= \sup\{ \int_Y sT^+ h \, d\nu - \alpha^\#(s) - \int_X h \, d\mu; \ s \in \mathbb{R}^+, \ h \in C(X) \},$$

which means that $\alpha(T)$ is a forward convex transfer with Kantorovich operators $T^+ f = sT^+(\frac{h}{s}) - \alpha^\#(s)$.

For 2) use the fact that $(-\beta)$ is convex decreasing to write

$$T(\mu, \nu) = \sup\{ \int_Y g \, d\nu - \beta(\int_X T^- g \, d\mu); \ g \in C(Y) \}$$

$$= \sup\{ \int_Y g \, d\nu + \sup\{ \int_X -sT^- g \, d\mu; \ s \in \mathbb{R}^+, \ (\beta)^\#(s); \ g \in C(Y) \}$$

$$= \sup\{ \int_Y g \, d\nu - s \int_X T^- g \, d\mu; \ s \in \mathbb{R}^+, \ (\beta)^\#(s); \ g \in C(Y) \}.$$
Hence $\mathcal{T}$ is the supremum of backward linear transfers.

**Example 14:** General entropic functionals are backward completely convex transfers

Consider the following generalized entropy,

$$T_\alpha(\mu, \nu) = \int_X \alpha\left(\frac{d\nu}{d\mu}\right) d\mu, \quad \text{if} \ \nu << \mu \ \text{and} \ +\infty \ \text{otherwise}, \quad (66)$$

where $\alpha$ is any strictly convex lower semi-continuous superlinear (i.e., $\lim_{t \to +\infty} \frac{\alpha(t)}{t} = +\infty$) real-valued function on $\mathbb{R}^+$. It is then easy to show [19] that

$$T_\alpha^*(f) = \inf \left\{ \int_X [\alpha^\circ(f(x) + t)] d\mu(x) ; t \in \mathbb{R} \right\}.$$ \quad (67)

In other words, $T_\alpha$ is a backward completely convex transfer, with Kantorovich operators

$$T_\alpha^* f(x) = \alpha^\circ(f(x) + t) - t.$$ \quad (68)

**Example 15:** The logarithmic entropy is a backward log-transfer

The relative logarithmic entropy $H(\mu, \nu)$ is defined as

$$H(\mu, \nu) := \int_X \log \left(\frac{d\nu}{d\mu}\right) d\nu \quad \text{if} \ \nu << \mu \ \text{and} \ +\infty \ \text{otherwise}.$$ \quad (69)

It can also be written as

$$H(\mu, \nu) := \int_X h\left(\frac{d\nu}{d\mu}\right) d\mu \quad \text{if} \ \nu << \mu \ \text{and} \ +\infty \ \text{otherwise},$$

where $h(t) = t \log t - t + 1$, which is strictly convex and positive. Since $h^*(t) = e^t - 1$, it follows that

$$H_\mu^*(f) = \inf \left\{ \int_X (e^t e^{f(x)} - 1 - t) d\mu(x) ; t \in \mathbb{R} \right\} = \log \int_X e^f d\mu.$$ \quad (70)

In other words, $H(\mu, \nu) = \sup \{ \int_X f d\nu - \log \int_X e^f d\mu ; f \in C(X) \}$, and $H$ is therefore a backward $\beta$-transfer with $\beta(t) = \log t$, and $E^- f = e^f$ is a convex Kantorovich operator.

$H$ is a convex backward transfer since

$$H(\mu, \nu) = \sup \left\{ \int_X f d\nu - \log \int_X e^f d\mu ; f \in C(X) \right\}$$

$$= \sup \left\{ \int_X f d\nu + \sup \{ \int_X -se^f d\mu - \beta^\circ(s) ; f \in C(X) \} \right\}$$

$$= \sup \left\{ \int_X f d\nu - s \int_X e^f d\mu - \beta^\circ(s) ; s > 0, f \in C(X) \right\}.$$ \quad (71)

In other words, it is backward completely convex with Kantorovich operators $T_s^- f = se^f + \beta^\circ(s)$ where $s > 0$.

**Example 16:** The Fisher-Donsker-Varadhan information is a backward completely convex transfer [10]

Consider an $X$-valued time-continuous Markov process $(\Omega, \mathcal{F}, (X_t)_{t \geq 0}, (\mathbb{P}_x)_{x \in X})$ with an invariant probability measure $\mu$. Assume the transition semigroup, denoted $(P_t)_{t \geq 0}$, to be completely continuous on $L^2(\mu) := L^2(X, \mathcal{B}, \mu)$. Let $\mathcal{L}$ be its generator with domain $\mathbb{D}_2(\mathcal{L})$ on $L^2(\mu)$ and assume the corresponding Dirichlet form $E(g, g) := \langle -\mathcal{L} g, g \rangle_{\mu}$ for $g \in \mathbb{D}_2(\mathcal{L})$ is closable in $L^2(\mu)$, with closure $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$. The Fisher-Donsker-Varadhan information of $\nu$ with respect to $\mu$ is defined by

$$I(\mu | \nu) := \begin{cases} E(\sqrt{\mathcal{J}}, \sqrt{\mathcal{J}}), & \text{if} \ \nu = f \mu, \sqrt{\mathcal{J}} \in \mathbb{D}(\mathcal{E}) \\ +\infty, & \text{otherwise}. \end{cases} \quad (68)$$
In other words, with transfer. The corresponding Feynman-Kac semigroup on $L^2(\mu)$

$$P_t^\mu g(x) := \mathbb{E}_x^\mu g(X_t) \exp \left( \int_0^t u(X_s) \, ds \right).$$

It has been proved in [38] that $\mathcal{I}_\mu^t(f) = \log \|P_t^f\|_{L^2(\mu)}$, which yields that $\mathcal{I}$ is a backward completely convex transfer.

$$\mathcal{I}_\mu(f) = \log \|P_t^f\|_{L^2(\mu)} = \frac{1}{2} \int_X |P_t^f g|^2 \, d\mu; \|g\|_{L^2(\mu)} \leq 1.$$  

In other words, with $\beta(t) = \log t$, we have

$$\mathcal{I}(\mu, \beta) = \sup \left\{ \int_Y f \, d\nu - \frac{1}{2} \log \sup \left\{ \int_X |P_t^f g|^2 \, d\mu; \|g\|_{L^2(\mu)} \leq 1 \right\}; f \in C(X) \right\}.$$  

Hence, it is a backward completely convex transfer, with convex Kantorovich operators $(T_{s,g})_{s,g}$ defined by

$$T_{s,g} f = \frac{1}{2} |P_t^f g|^2 + \frac{1}{2} \beta(s).$$

### 6 Operations on convex and entropic transfers

Denote by $CT^{-}(X \times Y)$ (resp., $CT_{+}(X \times Y)$) the class of backward (resp., forward) completely convex transfers. They are clearly convex cones in the space of convex weak*-lower semi-continuous functions on $\mathcal{P}(X) \times \mathcal{P}(Y)$. They also satisfy the following permanence properties. The most important being that the inf-convolution with linear transfers generate many new examples of convex and entropic transfers.

**Proposition 11** Let $\mathcal{F}$ be a backward completely convex transfer with Kantorovich operators $(F)_i$, Then,

1. If $a \in \mathbb{R}^+ \setminus \{0\}$, then $a\mathcal{F} \in CT_-(X \times Y)$ with Kantorovich operators given by $F_{a,i}^- (f) = aF_i^- (\frac{f}{a})$.

2. If $T$ is a backward linear transport on $Y \times Z$ with Kantorovich operator $T^-$, then $\mathcal{F} \ast T$ is a backward completely convex transfer with Kantorovich operators given by $F_i^- \circ T^-$. 

**Proof:** Immediate. For 2) we calculate the Legendre dual of $(\mathcal{F} \ast T)_\mu$ at $g \in C(Z)$ and obtain,

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

$$= \sup_{\sigma \in \mathcal{P}(Y)} \left\{ T_\mu^*(g) - \mathcal{F}(\mu, \sigma) \right\}$$

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

$$= (\mathcal{F})_\mu^* (T^- (g))$$

$$= \inf_{i \in I} \int_X F_i^- \circ T^- g(x) \, d\mu(x).$$

17
The same properties hold for entropic transfers. That we will denote by $E$ as opposed to $T$ to distinguish them from the linear transfers. We shall use $E^+$ and $E^-$ for their Kantorovich operators.

**Proposition 12** Let $\beta : \mathbb{R} \to \mathbb{R}$ be a concave increasing function and let $E$ be a $\beta$-backward transfer with Kantorovich operator $E^-$. Then,

1. If $\lambda \in \mathbb{R}^+ \setminus \{0\}$, then $\lambda E$ is a $(\lambda \beta)$-backward transfer with Kantorovich operator $E^-_\lambda (f) = E^- \left( \frac{f}{\lambda} \right)$.

2. $E$ is a $((\beta^\circ)^\circ)$-forward convex transfer with Kantorovich operator $E^+ = -E^-(-h)$.

3. If $T$ is a backward linear transfer on $Y \times Z$ with Kantorovich operator $T^-$, then $E \ast T$ is a backward transfer on $X \times Z$ with Kantorovich operator equal to $E^- \circ T^-$. In other words,

$$E \ast T (\mu, \nu) = \sup \left\{ \int_Z g(y) \, d\nu(y) - \beta(\int_X E^- \circ T^- g(x) \, d\mu(x)); \, g \in C(Z) \right\}. \quad (70)$$

**Proof:** 1) is trivial. For 2) note that since $\beta$ is concave and increasing, then

$$\tilde{T}(\nu, \mu) = T(\nu, \mu)$$
$$= \sup \left\{ \int_Y g \, d\nu - \beta(\int_X T^- g \, d\mu); \, g \in C(Y) \right\}$$
$$= \sup \left\{ \int_Y g \, d\nu + \sup_{s > 0} \left\{ \int_X -s T^- g \, d\mu - (-\beta)^\circ(s) \right\}; \, g \in C(X) \right\}$$
$$= \sup \left\{ \int_Y g \, d\nu - s \int_X T^- g \, d\mu - (-\beta)^\circ(s); \, s > 0, \, g \in C(X) \right\}$$
$$= \sup \left\{ s \int_X T^- (-h) \, d\mu - (-\beta)^\circ(s) - \int_Y h \, d\nu; \, s > 0, \, g \in C(X) \right\}$$
$$= \sup \left\{ (-\beta)^\circ \left( \int_X T^- (-h) \, d\mu \right) - \int_Y h \, d\nu; \, s > 0, \, h \in C(X) \right\}.$$ 

In other words, $\tilde{T}$ is a $((\beta^\circ)^\circ)$-forward convex transfer.

For 3) we calculate the Legendre dual of $(E \ast T)_\mu$ at $g \in C(Z)$ and obtain,

$$(E \ast T)_\mu (g) = \sup_{\nu \in P(Z)} \sup_{\sigma \in P(Y)} \left\{ \int_Z g \, d\nu - E(\mu, \sigma) - T(\sigma, \nu) \right\}$$
$$= \sup_{\sigma \in P(Y)} \left\{ T^*_\sigma (g) - E(\mu, \sigma) \right\}$$
$$= \sup_{\sigma \in P(Y)} \left\{ \int_Y T^- g \, d\sigma - E(\mu, \sigma) \right\}$$
$$= (E^{
abla}_\mu(T^-)(g))$$
$$= \beta \left( \int_X E^- \circ T^- g(x) \, d\mu(x) \right).$$

A similar statement holds for $\alpha$-forward transfers where $\alpha$ is now a convex increasing function on $\mathbb{R}^+$. But we then have to reverse the orders. For example, if $T$ (resp., $E$) is a forward linear transfer on $Z \times X$ (resp., a forward $\alpha$-transfer on $X \times Y$) with Kantorovich operator $T^+$ (resp., $E^+$), then $T \ast E$ is a backward $\alpha$-transfer on $Z \times Y$ with Kantorovich operator equal to $E^+ \circ T^-$. In other words,

$$T \ast E (\mu, \nu) = \sup \left\{ \alpha \left( \int_Y E^+ \circ T^+ f(y) \, d\nu(y) \right) - \int_X f(x) \, d\mu(x); \, f \in C(X) \right\}. \quad (71)$$
7 Transfer inequalities

Let \( \mathcal{T} \) be a transfer, and let \( \mathcal{E}_1, \mathcal{E}_2 \) be entropic transfers on \( X \times X \). Standard Transport-Entropy or Transport-Information inequalities are usually of the form

\[
\mathcal{T}(\sigma, \mu) \leq \lambda_1 \mathcal{E}_1(\mu, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(X),
\]

\[
\mathcal{T}(\mu, \sigma) \leq \lambda_2 \mathcal{E}_2(\mu, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(X),
\]

and more generally,

\[
\mathcal{T}(\sigma_1, \sigma_2) \leq \lambda_1 \mathcal{E}_1(\sigma_1, \mu) + \lambda_2 \mathcal{E}_2(\sigma_2, \mu) \quad \text{for all } \sigma_1, \sigma_2 \in \mathcal{P}(X),
\]

where \( \mu \) is a fixed measure, and \( \lambda_1, \lambda_2 \) are two positive reals. In our terminology, Problem 72 (resp., 73), (resp., 74) amount to find \( \lambda_1, \lambda_2 \) such that

\[
(\lambda_1 \mathcal{E}_1) \ast (-\mathcal{T}) (\mu, \mu) \geq 0,
\]

\[
\lambda_2 \mathcal{E}_2) \ast (-\mathcal{T}) (\mu, \mu) \geq 0,
\]

\[
(\lambda_1 \mathcal{E}_1) \ast (-\mathcal{T}) \ast (\lambda_2 \mathcal{E}_2) (\mu, \mu) \geq 0,
\]

where \( \mathcal{T}(\mu, \nu) = \mathcal{T}(\nu, \mu) \). Note for example that

\[
\mathcal{E}_1 \ast (-\mathcal{T}) \ast \mathcal{E}_2 (\mu, \nu) = \inf \{ \mathcal{E}_1(\mu, \sigma_1) - \mathcal{T}(\sigma_1, \sigma_2) + \mathcal{E}_2(\sigma_2, \nu); \sigma_1, \sigma_2 \in \mathcal{P}(Z) \}.
\]

We shall therefore write duality formulas for the transfers \( \mathcal{E}_1 \ast (-\mathcal{T}) \), \( \mathcal{E}_2 \ast (-\mathcal{T}) \) and \( \mathcal{E}_1 \ast (-\mathcal{T}) \ast \mathcal{E}_2 \) between any two measures \( \mu \) and \( \nu \), where \( \mathcal{T} \) is any convex transfer, while \( \mathcal{E}_1, \mathcal{E}_2 \) are entropic transfers.

7.1 Backward convex to backward completely convex inequalities

We would like to prove inequalities such as

\[
\mathcal{F}_2(\sigma, \mu) \leq \mathcal{F}_1(\mu, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(X),
\]

where both \( \mathcal{F}_1 \) and \( \mathcal{F}_2 \) are backward convex transfers. We then apply it to Transport-Entropy inequalities of the form

\[
\mathcal{F}(\sigma, \mu) \leq \lambda \mathcal{E} \ast \mathcal{T}(\mu, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(X),
\]

where \( \mathcal{F} \) is a backward convex transfer, while \( \mathcal{E} \) is a \( \beta \)-entropic transfer and \( \mathcal{T} \) is a backward linear transfer.

Proposition 13 Let \( \mathcal{F}_1 \) be a backward completely convex transfer with Kantorovich operator \( (F^-_{1,i})_{i \in I} \) on \( X_1 \times X_2 \), and \( \mathcal{F}_2 \) is a backward convex transfer on \( X_2 \times X_3 \) with Kantorovich operator \( (F^-_{2,j})_{j \in J} \).

1. The following duality formula hold:

\[
\mathcal{F}_1 \ast -\mathcal{F}_2 (\mu, \nu) = \inf_{f \in C(X_3)} \sup_{j \in J} \left\{ -\int_{X_2} F^-_{1,i} \circ F^-_{2,j} f d\mu - \int_{X_3} f d\nu \right\}.
\]

2. If \( \mathcal{F}_1 \) is a \( \beta \)-backward transfer on \( X_1 \times X_2 \) with Kantorovich operator \( E^-_1 \), then

\[
\mathcal{F}_1 \ast -\mathcal{F}_2 (\mu, \nu) = \inf_{f \in C(X_3)} \left\{ -\beta(\int_{X_2} E^-_1 \circ F^-_{2,j} f d\mu) - \int_{X_3} f d\nu \right\}.
\]
Proof: Write
\[
F_1 \ast -F_2 (\mu, \nu) = \inf_{\sigma \in \mathcal{P}(X_2)} \left\{ \mathcal{F}_1(\mu, \sigma) - \sup_{f \in C(X_3)} \sup_{j \in J} \left\{ \int_{X_3} f \, d\nu - \int_{X_2} F_{2,j}^{-} f \, d\sigma \right\} \right\}
\]
\[
= \inf_{\sigma \in \mathcal{P}(X_2)} \inf_{f \in C(X_3)} \inf_{j \in J} \left\{ \mathcal{F}_1(\mu, \sigma) - \int_{X_3} f \, d\nu + \int_{X_2} F_{2,j}^{-} f \, d\sigma \right\}
\]
\[
= \inf_{f \in C(X_3)} \inf_{j \in J} \left\{ -\mathcal{F}_1(\mu, \sigma) - \int_{X_3} f \, d\nu \right\}
\]
\[
= \inf_{f \in C(X_3)} \inf_{j \in J} \left\{ -\mathcal{F}_1(\mu, \sigma) - \mathcal{F}_1(\mu, \sigma) \right\}
\]
\[
= \inf_{f \in C(X_3)} \inf_{j \in J} \left\{ -\mathcal{F}_1(\mu, \sigma) \right\}
\]
\[
= \inf_{f \in C(X_3)} \inf_{j \in J} \left\{ -\mathcal{F}_1(\mu, \sigma) \right\}
\]

2) If \( F_1 \) is a \( \beta \)-backward transfer on \( X_1 \times X_2 \) with Kantorovich operator \( E_{1}^{-} \), then use in the above calculation that \( (F_1)_{\mu}(F_{2,j}^{-}) = \beta(\int_{X_1} E_{1}^{-} \circ -F_{2,j}^{-} f \, d\mu) \).

Corollary 14 Let \( F \) be a backward convex transfer on \( Y_2 \times X_2 \) with Kantorovich operators \( (F_{i}^{-})_{i \in I} \) and let \( E \) be a backward \( \beta \)-transfer on \( X_1 \times Y_1 \) with Kantorovich operator \( E^{-} \). Let \( T \) be a backward linear transfer on \( Y_1 \times Y_2 \) with Kantorovich operator \( T^{-} \) and \( \lambda > 0 \). Then, for any fixed pair of probability measures \( \mu \in \mathcal{P}(X_1) \) and \( \nu \in \mathcal{P}(X_2) \), the following are equivalent:

1. For all \( \sigma \in \mathcal{P}(Y_2) \), we have \( \mathcal{F}(\sigma, \nu) \leq \lambda E \ast T (\mu, \sigma) \).

2. For all \( g \in C(X_2) \) and \( i \in I \), we have \( \beta\left( \int_{X_1} E_{i}^{-} \circ T^{-} \circ \frac{1}{\lambda} F_{i}^{-} (\lambda g) \, d\mu \right) + \int_{X_2} g \, d\nu \leq 0 \).

In particular, if we apply the above in the case where \( E \) is the logarithmic entropy, that is
\[
\mathcal{H}(\mu, \nu) = \int_{X} \log(\frac{d\nu}{d\mu}) \, d\nu \text{ if } \nu << \mu \text{ and } +\infty \text{ otherwise,}
\]
which is a backward \( \beta \)-transfer with \( \beta(t) = \log t \) and \( E^{-} f = e^f \) as a backward Kantorovich operator.

Corollary 15 Let \( F \) be a backward convex transfer on \( X_2 \times Y_2 \) with Kantorovich operators \( (F_{i}^{-})_{i \in I} \) and let \( E \) be a backward \( \beta \)-transfer on \( X_1 \times Y_1 \), with Kantorovich operator \( E^{-} \). Let \( T \) be a backward linear transfer on \( Y_1 \times Y_2 \) with Kantorovich operator \( T^{-} \) and \( \lambda > 0 \). Then, for any fixed pair of probability measures \( \mu \in \mathcal{P}(X_1) \) and \( \nu \in \mathcal{P}(X_2) \), the following are equivalent:

1. For all \( \sigma \in \mathcal{P}(Y) \), we have \( \mathcal{F}(\sigma, \nu) \leq \lambda \mathcal{H} \ast T (\mu, \sigma) \).

2. For all \( g \in C(X_2) \), we have \( \sup_{i \in I} \int_{X_1} e^{T^{-} \circ \frac{1}{\lambda} F_{i}^{-} (\lambda g)} \, d\mu \leq e^{-\int_{X_2} g \, d\nu} \).

In particular, if \( T \) is the identity transfer and \( F \) is a backward linear transfer, then
\[
\mathcal{F}(\sigma, \nu) \leq \lambda \mathcal{H}(\sigma, \mu) \text{ for all } \sigma \in \mathcal{P}(Y) \iff \int_{X_1} e^{-F_{i}^{-} (\lambda g)} \, d\mu \leq e^{-\frac{\lambda}{\lambda} e^{-\int_{X_2} g \, d\nu}} \text{ for all } g \in C(X_2).
\]
7.2 Forward convex to backward completely convex transfer inequalities

We are now interested in inequalities such as
\[ F_2(\nu, \sigma) \leq F_1(\mu, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \] (84)
where both \( F_1 \) and \( F_2 \) are backward convex transfers, and in particular, Transport-Entropy inequalities of the form
\[ F(\nu, \sigma) \leq \lambda E \star T(\mu, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \] (85)
where \( E \) is a \( \beta \)-entropic transfer and \( T \) is a backward linear transfer. But we can write (86) as
\[ \tilde{F}_2(\sigma, \nu) \leq F_1(\mu, \sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \] (86)
where now \( \tilde{F}_2(\sigma, \nu) = F_2(\nu, \sigma) \) is a forward convex transfer. So, we need to establish the following type of duality.

**Proposition 16** Let \( F_1 \) be a backward completely convex transfer with Kantorovich operator \( (F_{1,i}^-)_{i \in I} \) on \( X_1 \times X_2 \), and let \( F_2 \) be a forward convex transfer on \( X_2 \times X_3 \) with Kantorovich operator \( (F_{2,j}^+)_{j \in J} \).

1. The following duality formula then holds:
\[ F_1 \star - F_2(\mu, \nu) = \inf_{g \in C(X_2)} \inf_{j \in J} \sup_{i \in I} \left\{ - \int_{X_1} F_{1,i}^-(-g) \, d\nu - \int_{X_3} F_{2,j}^+(g) \, d\nu \right\}. \] (87)

2. If \( F_1 \) is a \( \beta \)-backward transfer on \( X_1 \times X_2 \) with Kantorovich operator \( E_1^- \), then
\[ F_1 \star - F_2(\mu, \nu) = \inf_{g \in C(X_2)} \inf_{j \in J} \left\{ -\beta \left( \int_{X_1} E_1^-(-g) \, d\mu \right) - \int_{X_3} F_{2,j}^+(g) \, d\nu \right\}. \] (88)

3. If \( F_1 \) is a backward \( \beta \)-transfer with Kantorovich operator \( E_1^- \), and \( F_2 \) is a forward \( \alpha \)-transfer with Kantorovich operator \( E_2^+ \), then
\[ F_1 \star - F_2(\mu, \nu) = \inf_{g \in C(X_2)} \left\{ -\beta \left( \int_{X_1} E_1^-(-g) \, d\mu \right) - \alpha \left( \int_{X_3} E_2^+ g \, d\nu \right) \right\}. \] (89)

4. In particular, if \( F_1 \) is a backward \( \beta \)-transfer with Kantorovich operator \( E^- \), and \( T \) is a forward linear transfer with Kantorovich operator \( T^+ \), then
\[ E \star - T(\mu, \nu) = \inf_{g \in C(X_2)} \left\{ -\beta \left( \int_{X_1} E^-(-g) \, d\mu \right) - \int_{X_3} T^+ g \, d\nu \right\}. \] (90)
Proof: 1) Assume \( F_1 \) is a backward completely convex transfer with Kantorovich operator \( F_{1,i}^- \), and \( F_2 \) is a forward convex transfer with Kantorovich operator \( F_{2,j}^+ \), then
\[
F_1 \ast -F_2 (\mu, \nu) = \inf_{\sigma \in \mathcal{P}(X_2)} \left\{ \mathcal{F}_1(\mu, \sigma) - \mathcal{F}_2(\sigma, \nu); \sigma \in \mathcal{P}(X_2) \right\}
\]
\[
= \inf_{\sigma \in \mathcal{P}(X_2)} \left\{ \mathcal{F}_1(\mu, \sigma) - \sup_{g \in C(X_2)} \left\{ \sup_{j \in J} \int_{Y_2} F_{2,j}^+ g d\nu - \int_{X_2} g d\sigma \right\} \right\}
\]
\[
= \inf_{\sigma \in \mathcal{P}(X_2)} \inf_{g \in C(X_2)} \inf_{j \in J} \left\{ \mathcal{F}_1(\mu, \sigma) - \int_{X_2} F_{2,j}^+ g d\nu + \int_{X_2} g d\sigma \right\}
\]
\[
= \inf_{g \in C(X_2)} \inf_{j \in J} \left\{ -\mathcal{F}_1^*(g) - \int_{X_1} F_{1,i}^- (g) d\nu - \int_{X_3} F_{2,j}^+ (g) d\nu \right\}
\]
2) If \( \mathcal{E} \) is a \( \beta \)-backward entropic transfer with Kantorovich operator \( E^- \), it suffices to note in the above proof that \( \mathcal{F}_{\beta}^*(g) = \beta(\int_{X_2} E_1(\lambda g) d\mu) \).
3) If now \( \mathcal{E}_2 \) is a forward \( \alpha \)-transfer with Kantorovich operator \( E_2^+ \), then it suffices to note in the above proof that \( (\mathcal{F}_2)^*(g) = \alpha(\int_{X_2} E_2^+ g d\nu) \).

Corollary 17 Let \( \mathcal{F} \) be a backward convex transfer on \( X_2 \times Y_2 \) with Kantorovich operators \( (F_i^-)_{i \in I} \) and let \( \mathcal{E} \) be a backward \( \beta \)-transfer on \( X_1 \times Y_1 \) with Kantorovich operator \( E^- \). Let \( \mathcal{T} \) be a backward linear transfer on \( Y_1 \times Y_2 \) with Kantorovich operator \( T^- \) and \( \lambda > 0 \). Then, for any fixed pair of probability measures \( \mu \in \mathcal{P}(X_1) \) and \( \nu \in \mathcal{P}(X_2) \), the following are equivalent:

1. For all \( \sigma \in \mathcal{P}(Y_2) \), we have \( \mathcal{F}(\nu, \sigma) \leq \lambda \mathcal{E} \ast \mathcal{T} (\mu, \sigma) \).
2. For all \( g \in C(X_2) \), we have \( \beta \left( \int_{X_1} E^- \circ T^- g d\mu \right) \leq \inf_{i \in I} \frac{1}{\lambda} \int_{X_2} F_i^- (\lambda g) d\nu \).

In particular, if \( \mathcal{E}_2 \) is a backward \( \beta_2 \)-transfer on \( X_2 \times Y_2 \) with Kantorovich operator \( E_2^- \), and \( \mathcal{E}_1 \) is a backward \( \beta_1 \)-transfer on \( X_1 \times Y_1 \) with Kantorovich operator \( E_1^- \), then the following are equivalent:

1. For all \( \sigma \in \mathcal{P}(Y_2) \), we have \( \mathcal{E}_2(\nu, \sigma) \leq \lambda \mathcal{E}_1 \ast \mathcal{T} (\mu, \sigma) \).
2. For all \( g \in C(X_2) \) and \( i \in I \), we have \( \beta_1 \left( \int_{X_1} E_i^- \circ T^- g d\mu \right) \leq \frac{1}{\lambda} \beta_2 \left( \int_{X_2} E_2^- (\lambda g) d\nu \right) \).

Proof: Note that here, we need the formula for \( (\mathcal{E} \ast \mathcal{T} ) \ast (-\tilde{\mathcal{F}})(\mu, \nu) \). Since \( \tilde{\mathcal{F}} \) is now a forward convex transfer with Kantorovich operators equal to \( F_i^+(g) = -F_i^-(g) \), we can apply Part 2) of Proposition 16 to \( \mathcal{F}_2 = \frac{1}{\lambda} \tilde{\mathcal{F}} \) and \( \mathcal{F}_1 = \mathcal{E} \ast \mathcal{T} \), which is a \( \beta \)-backward transfer with Kantorovich operator \( E^- \circ T^- \), to obtain
\[
(\mathcal{E} \ast \mathcal{T} ) \ast (-\tilde{\mathcal{F}})(\mu, \nu) = \inf_{g \in C(X_2)} \inf_{j \in J} \left\{ -\beta \left( \int_{X_1} E^- \circ T^- g d\mu \right) + \frac{1}{\lambda} \int_{X_3} F_j^+ (\lambda g) d\nu \right\}
\]
A similar argument applies for 2).
We now apply the above to the case where \( \mathcal{E} \) is the backward logarithmic transfer to obtain,

Corollary 18 Let \( \mathcal{F} \) be a backward convex transfer on \( X_2 \times Y_2 \) with Kantorovich operators \( (F_i^-)_{i \in I} \). Let \( \mathcal{T} \) be a backward linear transfer on \( Y_1 \times Y_2 \) with Kantorovich operator \( T^- \) and \( \lambda > 0 \). Then, for any fixed pair of probability measures \( \mu \in \mathcal{P}(X_1) \) and \( \nu \in \mathcal{P}(X_2) \), the following are equivalent:
1. For all $\sigma \in \mathcal{P}(Y_2)$, we have $\mathcal{F}(\nu, \sigma) \leq \lambda \mathcal{H} * T (\mu, \sigma)$

2. For all $g \in C(X_2)$, we have $\log \left( \int_{X_1} e^{T^* g} \, d\mu \right) \leq \inf_{i \in I} \int_{X_2} \bar{F}_i^-(\lambda g) \, d\nu$.

### 7.3 Maurey-type inequalities

We are now interested in inequalities of the following type: For all $\sigma_1 \in \mathcal{P}(X_1), \sigma_2 \in \mathcal{P}(X_2)$, we have

$$\mathcal{F}(\sigma_1, \sigma_2) \leq \lambda_1 \mathcal{T}_1 * \mathcal{H}_1(\sigma_1, \mu) + \lambda_2 \mathcal{T}_2 * \mathcal{H}_2(\sigma_2, \nu).$$

(91)

This will requires a duality formula for the expression $\tilde{\mathcal{E}}_1 * (-\mathcal{T}) * \mathcal{E}_2$, where $\mathcal{F}$ is a backward convex transfer and $\mathcal{E}_1, \mathcal{E}_2$ are forward entropic transfers.

**Theorem 1** Assume $\tilde{\mathcal{E}}_1$ is a backward convex transfer on $Y_1 \times Y_2$ with Kantorovich operators $(F_{i1}^-)_{i \in I}$, $\mathcal{E}_1$ (resp., $\mathcal{E}_2$) is a forward $\alpha_1$-transfer on $Y_1 \times X_1$ (resp., a forward $\alpha_2$-transfer on $Y_2 \times X_2$) with Kantorovich operator $E_{11}^+$ (resp., $E_{22}^+$), then for any given $(\mu, \nu) \in \mathcal{P}(X_1) \times \mathcal{P}(X_2)$, we have

$$\tilde{\mathcal{E}}_1 * (-\mathcal{F}) * \mathcal{E}_2 (\mu, \nu) = \inf_{i \in I} \inf_{f \in C(X_3)} \left\{ \alpha_1 \left( \int_{X_1} E_{11}^+ \circ F_{i1}^- \, f \, d\mu \right) \right\}$$

(92)

$$+ \inf_{i \in I} \inf_{f \in C(Z)} \left\{ \alpha_2 \left( \int_{X_2} E_{22}^+ \circ F_{i2}^- \, f \, d\nu \right) \right\}.$$

**Proof:** If $\mathcal{E}_1$ a forward $\alpha_1$-transfer on $Y_1 \times X_1$, then $\tilde{\mathcal{E}}_1$ is a backward $-\alpha_1^\oplus$ transfer on $X_1 \times Y_1$ with Kantorovich operator $\tilde{E}_{11}^- g = -E_{11}^+ (-g)$. Apply Proposition 13 with $\mathcal{F}_1 = \tilde{\mathcal{E}}_1$, and $\mathcal{F}_2 = \mathcal{F}$ to get

$$\tilde{\mathcal{E}}_1 * (-\mathcal{F}) (\mu, \nu) = \inf_{f \in C(X_3)} \left\{ \alpha_1 \left( \int_{X_1} E_{11}^+ \circ F_{i1}^- \, f \, d\mu \right) - \int_{X_3} f \, d\nu \right\}$$

Write now,

$$\tilde{\mathcal{E}}_1 * (-\mathcal{F}) * \mathcal{E}_2 (\mu, \nu) = \inf_{\sigma \in \mathcal{P}(Y_2)} \inf_{f \in C(X_3)} \inf_{i \in I} \left\{ \alpha_1 \left( \int_{X_1} E_{11}^+ \circ F_{i1}^- \, f \, d\mu \right) - \int_{X_3} f \, d\nu + \mathcal{E}_2 (\sigma, \nu) \right\}.$$

**Corollary 19** Assume $\tilde{\mathcal{E}}_1$ (resp., $\tilde{\mathcal{E}}_2$) is a forward $\alpha_1$-transfer on $Z_{1} \times X_{1}$ (resp., $\alpha_2$-transfer on $Z_{2} \times X_{2}$) with Kantorovich operator $E_{11}^+$ (resp., $E_{22}^+$). Let $\mathcal{T}_1$ (resp., $\mathcal{T}_2$) be forward linear transfers on $Y_{1} \times Z_{1}$ (resp., $Y_{2} \times Z_{2}$) with Kantorovich operator $T_{11}^+$ (resp., $T_{22}^+$), and let $\mathcal{F}$ be a backward convex transfer on $Y_{1} \times Y_{2}$ with Kantorovich operators $(F_{i1}^-)_{i \in I}$. Then, for any given $\lambda_1, \lambda_2 \in \mathbb{R}^+$ and $\mu, \nu \in \mathcal{P}(X_1) \times \mathcal{P}(X_2)$, the following are equivalent:

1. For all $\sigma_1 \in \mathcal{P}(Y_1), \sigma_2 \in \mathcal{P}(Y_2)$, we have

$$\mathcal{F}(\sigma_1, \sigma_2) \leq \lambda_1 \mathcal{T}_1 * \mathcal{E}_1(\sigma_1, \mu) + \lambda_2 \mathcal{T}_2 * \mathcal{E}_2(\sigma_2, \nu).$$

(93)

2. For all $g \in C(Y_2)$ and all $i \in I$, we have

$$\lambda_1 \alpha_1 \left( \int_{X_1} E_{11}^+ \circ T_{11}^+ \circ \left( \frac{1}{\lambda_1} F_{i1}^- g \right) \, d\mu \right) + \lambda_2 \alpha_2 \left( \int_{X_2} E_{22}^+ \circ T_{22}^+ \left( \frac{1}{\lambda_2} g \right) \, d\nu \right) \geq 0.$$

(94)
Proof: It suffices to apply the above with the forward \( \lambda_i \alpha_i \)-transfers \( F_i := \lambda_i T_i \ast \mathcal{E}_i \), whose Kantorovich operators are \( F_i(g) = E_i^+ \circ T_i^+ (\frac{g}{X_i}) \) for \( i = 1, 2 \).

By applying the above to \( \mathcal{E}_i(\mu, \nu) =: \mathcal{H} \) the forward logarithmic entropy where \( \alpha_i(t) = -\log(-t) \) and Kantorovich operator \( E^+ f = e^{-f} \), we get the following extension of a celebrated result of Maurey [27].

**Corollary 20** Assume \( \mathcal{F} \) is a backward convex transfer on \( Y_1 \times Y_2 \) with Kantorovich operators \( (F_i^-)_{i \in I} \), and let \( T_1 \) (resp., \( T_2 \)) be forward linear transfer on \( Y_1 \times X_1 \) (resp., \( Y_2 \times X_2 \)) with Kantorovich operator \( T_1^+ \) (resp., \( T_2^+ \)), then for any given \( \lambda_1, \lambda_2 \in \mathbb{R}^+ \) and \( (\mu, \nu) \in \mathcal{P}(X_1) \times \mathcal{P}(X_2) \), the following are equivalent:

1. For all \( \sigma_1 \in \mathcal{P}(X_1), \sigma_2 \in \mathcal{P}(X_2) \), we have
   \[
   \mathcal{F}(\sigma_1, \sigma_2) \leq \lambda_1 T_1 \ast \mathcal{H}(\sigma_1, \mu) + \lambda_2 T_2 \ast \mathcal{H}(\sigma_2, \nu).
   \] (95)

2. For all \( g \in C(Y_2) \) and all \( i \in I \), we have
   \[
   \left( \int_{X_1} e^{-T_i^+ \circ F_i^+ (\frac{g}{X_1}) \, d\mu} \right)^{\lambda_1} \left( \int_{X_2} e^{-T_i^+ (\frac{g}{X_2}) \, d\nu} \right)^{\lambda_2} \leq 1.
   \] (96)

If \( T_1 = T_2 \) are the identity transfer, then the above is equivalent to saying that for all \( g \in C(Y_2) \) and all \( i \in I \), we have

\[
\left( \int_{X_1} e^{-\frac{1}{\lambda_1} F_i^+ g \, d\mu} \right)^{\lambda_1} \left( \int_{X_2} e^{\frac{1}{\lambda_2} g \, d\nu} \right)^{\lambda_2} \leq 1.
\] (97)

### 8 Infinite convolution and weak KAM theory on Wasserstein space

Let \( X \) be a compact metric space, and let \( \mathcal{T} \) be a backward linear transfer on \( X \times X \) that is weak*-continuous on \( \mathcal{M}(X) \). Let \( \mathcal{T} \) be the corresponding backward Kantorovich operator. We will be looking for fixed points of \( \mathcal{T} \), which in the case of a transfer induced by a generating function minimizing transport correspond to Fathi’s notion of weak KAM solution ([11], [1]).

**Lemma 21** For each \( n \in \mathbb{N} \), Let \( \mathcal{T}_n = \mathcal{T} \ast \mathcal{T} \ast \ldots \ast \mathcal{T} \) \( n \)-times. Then

1. For all \( \mu, \nu \in \mathcal{P}(X) \), we have
   \[
   \mathcal{T}_n(\mu, \nu) = \sup \left\{ \int_X g(y) \, d\nu(y) - \int_X T^n g(x) ; g \in C(X) \right\}.
   \] (98)

2. The functionals \( \mathcal{T}_n \) are equicontinuous, and there exists a positive constant \( C > 0 \) and a number \( \ell \in \mathbb{R} \) such that
   \[
   |\mathcal{T}_n(\mu, \nu) - \ell n| \leq C \quad \text{for all } \mu, \nu \in \mathcal{P}(X) \text{ and } n \in \mathbb{N}.
   \] (99)

**Proof:** 1) is Immediate from Proposition [8]

For 2), follow an argument of Bernard-Buffoni [1]. Since \( \mathcal{T} \) is weak* continuous and \( X \) is compact, there exists a modulus of continuity \( \delta : [0, \infty) \to [0, \infty) \), with \( \lim_{t \to 0} \delta(t) = \delta(0) = 0 \) such that

\[
|\mathcal{T}(\mu, \nu) - \mathcal{T}(\mu', \nu')| \leq \delta(W_2(\mu, \mu') + W_2(\nu, \nu')) \quad \text{for all } \mu, \mu', \nu, \nu' \in \mathcal{P}(X).
\] (100)

Since for all \( \sigma_1, \ldots, \sigma_{n-1} \in \mathcal{P}(X) \), the map

\[
(\mu, \nu) \mapsto \mathcal{T}(\mu, \sigma_1) + \mathcal{T}(\sigma_1, \sigma_2) + \ldots + \mathcal{T}(\sigma_{n-1}, \nu)
\]
Proof: We note that $T_n$ does too, as the infimum of functions with the same modulus of continuity, also has the same modulus of continuity.

Define the sequences

\[ M_n := \max \{ T_n(\mu, \nu) : \mu, \nu \in \mathcal{P}(X) \} \]
\[ m_n := \min \{ T_n(\mu, \nu) : \mu, \nu \in \mathcal{P}(X) \}. \]

Then $M_n$ is sub-additive, while $m_n$ is super-additive. Indeed, there exists $\mu, \nu \in \mathcal{P}(X)$ such that, for all $\sigma \in \mathcal{P}(X),$

\[ M_{n+k} = T_{n+k}(\mu, \nu) \leq T_n(\mu, \sigma) + T_k(\sigma, \nu) \leq M_n + M_k. \]

The argument for $m_n$ is the same with reverse inequalities.

By the equicontinuity of $T_n$, the difference $M_n - m_n$ is bounded above by a constant $C$ independent of $n.$ Therefore it is well known that $\frac{M_n}{n}$ decreases to $M := \inf_n \frac{M_n}{n}$, while $\frac{m_n}{n}$ increases to $m := \sup_n \frac{m_n}{n}.$ Then the string of inequalities

\[ nM - C \leq M_n - C \leq m_n \leq T_n(\mu, \nu) \leq M_n \leq m_n + C \leq nm + C \]

implies $M \leq m + \frac{2C}{n}$ for all $n$; hence $m = M.$

**Theorem 2** Under the condition that $\ell = 0$, there exists a weak* lower semi-continuous backward linear transfer $T_\infty$ on $X$ with associated Kantorovich operator $T_\infty$ satisfying the following properties.

1. For every $f \in C(X)$, we have $T_\infty \circ T_\infty f = T_\infty f$ and $T_\infty \circ T_\infty f = T_\infty f$.

   Moreover, the fixed points of $T$ and the fixed points of $T_\infty$ are the same.

2. For all $\mu, \nu \in \mathcal{P}(X)$, we have

   \[ T_\infty(\mu, \nu) = T_n \ast T_\infty(\mu, \nu) \quad \text{and} \quad T_\infty(\mu, \nu) = T_\infty \ast T_\infty(\mu, \nu). \quad (101) \]

3. $T_\infty(\mu, \nu) \leq \lim \inf_n T_n(\mu, \nu)$.

4. The set $A := \{ \mu \in \mathcal{P}(X); T_\infty(\mu, \mu) = 0 \}$ is non-empty and for every $\mu, \nu \in \mathcal{P}(X)$, we have

\[ T_\infty(\mu, \nu) = \inf \{ T_\infty(\mu, \sigma) + T_\infty(\sigma, \nu), \sigma \in A \}, \quad (102) \]

and the infimum on $A$ is attained.

**Proof:** We note that $T_\infty$ can be seen as a generalized Peierls barrier and $A$ as a projected Aubry set [11].

1. By Lemma 21, $T_n$ is equicontinuous with modulus of continuity $\delta$. Therefore, for each $f \in C(X), x \mapsto T_n f(x)$ is uniformly continuous with the same modulus of continuity $\delta$. Moreover, $f \mapsto T_n f$ is 1-Lipschitz with respect to the sup-norm. Indeed,

\[ T_n f(x) = \sup_{\nu \in \mathcal{P}(X)} \left\{ \int_X f \, d\nu - T_n(\delta_x, \nu) \right\} \]
\[ = \sup_{\nu \in \mathcal{P}(X)} \left\{ \int_X f \, d\nu - T_n(\delta_y, \nu) + T_n(\delta_y, \nu) - T_n(\delta_x, \nu) \right\} \]
\[ \leq \sup_{\nu \in \mathcal{P}(X)} \left\{ \int_X f \, d\nu - T_n(\delta_y, \nu) + \delta(d(x, y)) \right\} \]
\[ = T_n f(y) + \delta(d(x, y)). \]
Interchanging \( x \) and \( y \) we conclude the continuity. For 1-Lipschitz,

\[
T_n f(x) = \sup_{\nu \in \mathcal{P}(X)} \left\{ \int_X f \, d\nu - T_n(\delta_x, \nu) \right\}
\]

\[
\leq \sup_{\nu \in \mathcal{P}(X)} \left\{ \int_X g \, d\nu - T_n(\delta_x, \nu) \right\} + \|f - g\|_\infty
\]

\[
= T_n g(x) + \|f - g\|_\infty.
\]

Interchanging \( f \) and \( g \) we conclude.

Define \( T_1^\infty f(x) := \limsup_{n \to \infty} T_n f(x) \). The assumption \( \ell = 0 \) ensures that \( T_1^\infty f(x) < \infty \), and \( T_1^\infty \) satisfies the same properties above as \( T \). We have the following monotonicity:

\[
T \circ T_1^\infty f(x) \geq T_1^\infty f(x), \quad \text{for all } f \in C(X) \text{ and all } x \in X.
\] (103)

Indeed, since \( \{\sup_{k \geq n} T_k f\}_n \) is a sequence of continuous functions which pointwise decreases monotonically to the continuous function \( T_1^\infty f \), the sequence \( \sup_{k \geq n} T_k f \) converges uniformly to \( T_1^\infty f \). We conclude by the Lipschitz property of \( T \).

Therefore, for each \( f \in C(X) \) and \( x \in X \), \( \{T_n \circ T_1^\infty f(x)\}_n \) is a monotone sequence. The assumption \( \ell = 0 \) implies it is uniformly bounded in \( n \), hence the pointwise limit exists and is finite, and we may define an operator \( T_\infty \) by

\[
T_\infty f(x) := \lim_{n \to \infty} T_n \circ T_1^\infty f(x).
\] (104)

Note that again \( T_\infty \) satisfies the same properties as \( T \); in particular it is a convex operator. Moreover, \( x \mapsto T_\infty f(x) \) is continuous, and therefore, the convergence is uniform. Then, by the Lipschitz property of \( T \), we get that \( T \circ (T_n \circ T_1^\infty) f \) converges uniformly to \( T \circ T_\infty f \). In other words,

\[
T_\infty f = \lim_{n \to \infty} T_n \circ T_1^\infty f = \lim_{n \to \infty} T_{n+1} \circ T_1^\infty f = T \circ T_\infty f.
\] (105)

Finally, suppose \( f \in C(X) \) is a fixed point of \( T \). Then \( T_n f = f \), so \( T_\infty f = f \), and consequently \( T_n \circ T_1^\infty f = f \). Letting \( n \to \infty \), we get \( T_\infty f = f \).

Conversely, suppose \( f \) is a fixed point of \( T_\infty \). Since \( T_\infty f \) is a fixed point of \( T \) from above, we get that \( Tf = f \).

2. Define

\[
T_\infty(\mu, \nu) := \sup \left\{ \int_X f \, d\nu - \int_X T_\infty f \, d\mu : f \in C(X) \right\}.
\] (106)

Since \( T_\infty \) is a convex operator, \( T_\infty \) is a backward linear transfer. From 1), we get immediately the conclusion of 2), by Proposition \[8\]

3. Note that \( T_\infty f \geq T_1^\infty f = \limsup_n T_n f, \) so

\[
T_\infty(\mu, \nu) = \sup_{f \in C(X)} \left\{ \int_X f \, d\nu - \int_X T_\infty f \, d\mu \right\}
\]

\[
\leq \sup_{f \in C(X)} \left\{ \int_X f \, d\nu - \int_X \limsup_n T_n f \, d\mu \right\}
\]

\[
\leq \sup_{f \in C(X)} \liminf_n \left\{ \int_X f \, d\nu - \int_X T_n f \, d\mu \right\} \quad \text{(Fatou)}
\]

\[
\leq \liminf_n \sup_{f \in C(X)} \left\{ \int_X f \, d\nu - \int_X T_n f \, d\mu \right\}
\]

\[
= \liminf_n T_n(\mu, \nu).
\]
4. This argument is a minor modification of the one given in \[1\], to account for the fact that \( T_\infty \) is apriori only weak* lower semi-continuous. Fix \( \mu, \nu \in \mathcal{P}(X) \). By 2), there exists \( \sigma_1 \in \mathcal{P}(X) \) such that

\[
T_\infty(\mu, \nu) = T_\infty(\mu, \sigma_1) + T_\infty(\sigma_1, \nu).
\]

Similarly, there exists a \( \sigma_2 \) such that

\[
T_\infty(\sigma_1, \nu) = T_\infty(\sigma_1, \sigma_2) + T_\infty(\sigma_2, \nu).
\]

Combining the above two equalities, we obtain

\[
T_\infty(\mu, \nu) = T_\infty(\mu, \sigma_1) + T_\infty(\sigma_1, \sigma_2) + T_\infty(\sigma_2, \nu).
\]

Note also that

\[
T_\infty(\mu, \sigma_1) + T_\infty(\sigma_1, \sigma_2) = T_\infty(\mu, \sigma_2).
\]

This follows from

\[
T_\infty(\mu, \nu) = T_\infty(\mu, \sigma_1) + T_\infty(\sigma_1, \sigma_2) + T_\infty(\sigma_2, \nu) \geq T_\infty(\mu, \sigma_2) + T_\infty(\sigma_2, \nu) = T_\infty(\mu, \nu).
\]

Hence all the inequalities are equalities; in particular (107).

After \( k \) times we have

\[
T_\infty(\mu, \nu) = \sum_{i=0}^{k} T_\infty(\sigma_i, \sigma_{i+1})
\]

where \( \sigma_0 := \mu \) and \( \sigma_{k+1} := \nu \). This inductively generates a sequence \( \{\sigma_k\} \) with the property

\[
\sum_{i=0}^{m} T_\infty(\sigma_i, \sigma_{i+1}) = T_\infty(\sigma_{\ell}, \sigma_{m+1})
\]

whenever \( 0 \leq \ell < m \leq k \). In particular, for any subsequence \( \sigma_{k_j} \), we have

\[
T(\mu, \sigma_{k_1}) + \sum_{j=1}^{m} T_\infty(\sigma_{k_j}, \sigma_{k_{j+1}}) + T_\infty(\sigma_{k_{m+1}}, \nu) = T_\infty(\mu, \nu).
\]

Extract a weak-* convergent subsequence \( \{\sigma_{k_j}\} \) to some \( \bar{\sigma} \in \mathcal{P}(X) \). By weak-* l.s.c. of \( T_\infty \), we have

\[
\liminf_j T_\infty(\sigma_{k_j}, \sigma_{k_{j+1}}) \geq T_\infty(\bar{\sigma}, \bar{\sigma}).
\]

In particular, given \( \epsilon > 0 \), for all but finitely many \( j \),

\[
T_\infty(\sigma_{k_j}, \sigma_{k_{j+1}}) \geq T_\infty(\bar{\sigma}, \bar{\sigma}) - \epsilon.
\]

Therefore, by refining to a further (non-relabeled) subsequence if necessary, we obtain a subsequence \( \{\sigma_{k_j}\} \) satisfying (109) for all \( j \). By further refinement, we may also assume

\[
T_\infty(\mu, \sigma_{k_1}) \geq T_\infty(\mu, \bar{\sigma}) - \epsilon.
\]

Therefore, by refining to a further (non-relabeled) subsequence if necessary, we obtain a subsequence \( \{\sigma_{k_j}\} \) with properties (108), (109), and (110).
Moreover, for all \( m \) large enough (depending on \( \epsilon \)), we have
\[
T_\infty(\sigma_{k,m+1}, \nu) \geq T_\infty(\sigma, \nu) - \epsilon \tag{111}
\]
Applying the inequalities of (109), (110), and (111), to (108), we obtain
\[
T_\infty(\mu, \nu) \geq T_\infty(\mu, \bar{\sigma}) + mT_\infty(\bar{\sigma}, \bar{\sigma}) + T_\infty(\bar{\sigma}, \nu) - (m + 2)\epsilon
\]
for all large enough \( m \). From the fact that \( T_\infty = T_\infty \circ T_\infty \), the above inequality is only possible if
\[
T_\infty(\bar{\sigma}, \bar{\sigma}) \leq \frac{m + 2}{m} - \epsilon \leq 2\epsilon.
\]
As \( \epsilon \) is arbitrary, we obtain \( T_\infty(\bar{\sigma}, \bar{\sigma}) \leq 0 \), and consequently \( T_\infty(\bar{\sigma}, \bar{\sigma}) = 0 \) (the reverse inequality following from \( T_\infty = T_\infty \circ T_\infty \)).

Finally, we note that \( T_\infty(\mu, \nu) = T_\infty(\mu, \sigma_{k,j}) + T_\infty(\sigma_{k,j}, \nu) \) for all \( j \), so at the \( \lim \inf \), we find
\[
T_\infty(\mu, \nu) \geq T_\infty(\mu, \bar{\sigma}) + T_\infty(\bar{\sigma}, \nu).
\]
The reverse inequality is immediate from \( T_\infty = T_\infty \circ T_\infty \).

The following can be seen as an extension of Mather theorem \([1]\).

**Theorem 3** Under the above hypothesis, we have
\[
\inf \{ T(\mu, \mu) ; \mu \in \mathcal{P}(X) \} = 0, \tag{112}
\]
and the infimum is attained by a measure \( \bar{\mu} \in \mathcal{A} \) such that
\[
(\bar{\mu}, \bar{\mu}) \in \mathcal{D} := \{ (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X) : T(\mu, \nu) + T_\infty(\nu, \mu) = 0 \}. \tag{113}
\]

**Proof:** Note that \( \mathcal{D} \) can be seen as a generalized Aubry set \([1]\). By the assumption \( \ell = 0 \), we have \( T(\mu, \mu) \geq 0 \) for all \( \mu \). Therefore it suffices to find \( \bar{\mu} \) such that \( T(\bar{\mu}, \bar{\mu}) = 0 \). We may construct a sequence \( (\mu_k) \subset \mathcal{A} \) such that \( (\mu_k, \mu_{k+1}) \in \mathcal{D} \). The set \( \mathcal{D} \) is convex by convexity of both \( T \) and \( T_\infty \). Therefore, the Cesaro averages belong to \( \mathcal{D} \),
\[
\left( \frac{1}{n} \sum_{k=1}^{n} \mu_k, \frac{1}{n} \sum_{k=1}^{n} \mu_{k+1} \right) \in \mathcal{D}.
\]
Denoting \( \nu_n := \frac{1}{n} \sum_{k=1}^{n} \mu_k \), we have
\[
(\nu_n, \nu_n + \frac{1}{n}(\mu_{n+1} - \mu_1)) \in \mathcal{D}.
\]
Extracting a weak-* convergent subsequence \( \nu_{n_j} \) converging to some \( \bar{\mu} \in \mathcal{A} \) (since, in particular \( \nu_n \in \mathcal{A} \) and \( \mathcal{A} \) is weak-* closed), we use the fact that \( (\nu_n, \nu_n + \frac{1}{n}(\mu_{n+1} - \mu_1)) \in \mathcal{D} \) and weak-* continuity of \( T \) (resp. weak-* l.s.c. of \( T_\infty \)) to find
\[
T(\bar{\mu}, \bar{\mu}) \leq -T_\infty(\bar{\mu}, \bar{\mu}) = 0,
\]
which concludes the proof.
References

[1] P. Bernard & B. Buffoni: *Optimal mass transportation and Mather theory*. J. Eur. Math. Soc., 9 (2007), no. 1, 85-121.

[2] P. Bernard & B. Buffoni: *Weak KAM Pairs and Monge-Kantorovich Duality*, Advanced Studies in Pure Mathematics, 47-2, (2007) 397- 420.

[3] A. Barton, N. Ghoussoub: *Dynamic and Stochastic Propagation of Brenier’s Optimal Mass Transport*, (February 2018) 32 pp.

[4] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control*, Mathematics in Science and Engineering, vol. 139, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1978.

[5] S. G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 80(7):669–696, 2001.

[6] S.G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis.*, 163:1–28, 1999.

[7] M. Bowles, N. Ghoussoub, *A Theory of Transfers II: A stochastic weak KAM theorem*, in preparation (2018).

[8] M. Bowles, N. Ghoussoub, *A Theory of Multi-transfers and Applications*, in preparation (2018).

[9] Y. Brenier: *Polar factorization and monotone rearrangement of vector-valued functions*. Comm. Pure Appl. Math., 44 (1991), 375-417.

[10] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, III. *Comm. Pure Appl. Math.*, 29:389–461, 1976.

[11] A. Fathi, *Weak KAM Theorem in Lagrangian Dynamics*, preliminary version, Lyon, version X, 2018.

[12] I. Gentil, C. Lonard, 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, 3 (2017), pp. 569-600.

[13] N. Ghoussoub, Y.-H. Kim, and T. Lim, *Structure of optimal martingale transport in general dimensions*, Annals of Probability, in press. [http://arxiv.org/abs/1508.01806](http://arxiv.org/abs/1508.01806) (2015).

[14] N. Ghoussoub, Y.-H. Kim and A. Z. Palmer, *Optimal Transport With Controlled Dynamics and Free End Times*, submitted (March 2018) 23 pp.

[15] N. Ghoussoub, Y.-H. Kim and A. Z. Palmer, *PDE Methods for Stochastic Transportation Problems*, In preparation (2018).

[16] D. Gomes, *A stochastic analog of Aubry-Mather theory*, Nonlinearity 10 ( 2002), 271-305.

[17] N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields 16 (2010), no. 4, 635–736.

[18] N. Gozlan and C. Léonard. *A large deviation approach to some transportation cost inequalities*, To appear in *Probability Theory and Related Fields*.

[19] N. Gozlan, C. Roberto, and P.-M. Samson, *From concentration to logarithmic Sobolev and Poincaré inequalities*, J. Funct. Anal. 260 (2011), no. 5, 1491–1522.
[20] N. Gozlan, C. Roberto, P.M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, to appear in J. Funct. Anal., preprint (2014), arXiv:1412.7480v4.

[21] Pierre Henry-Labordère, *Model-free Hedging: A Martingale Optimal Transport Viewpoint*, Chapman and Hall/CRC (2017) 190 pages.

[22] M. Ledoux. *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs 89. American Mathematical Society, Providence RI, 2001.

[23] C. Léonard. *A saddle-point approach to the Monge-Kantorovich optimal transport problem*, ESAIM Control Optim. Calc. Var. 17 (2011), no. 3, 682–704.

[24] K. Marton. *Bounding d̅-distance by informational divergence: a way to prove measure concentration*, Annals of Probability, 24:857–866, 1996.

[25] K. Marton, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. 6 (1996), no. 3, 556–571.

[26] J. N. Mather, *Action minimizing invariant measures for positive definite Lagrangian systems*, Mathematische Zeitschrift, Math. Z. 207 (1991), 169-207.

[27] B. Maurey, *Some deviation inequalities*, Geom. Funct. Anal. 1 (1991), no. 2, 188–197.

[28] T. Mikami, *A simple proof of duality theorem for Monge-Kantorovich problem*, Kodai Math. J. 29 (2006), no. 1, 1–4.

[29] T. Mikami and M. Thieullen, *Duality theorem for the stochastic optimal control problem*. Stoch. Process. Appl. 116 (2006), no. 12, 1815–1835

[30] S. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol I : Theory, Vol. II : Applications*. Probability and its applications. Springer Verlag, New York, 1998.

[31] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205.

[32] P.-M. Samson, *Concentration inequalities for convex functions on product spaces*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 33–52.

[33] P.-M. Samson, *Infimum-convolution description of concentration properties of product probability measures, with applications*, Ann. Inst. H. Poincaré Probab. Statist. 43 (2007), no. 3, 321–338.

[34] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. 36 (1965), 423–439.

[35] M. Talagrand, *New concentration inequalities in product spaces*, Invent. Math. 126 (1996), no. 3, 505–563.

[36] M. Talagrand. *Transportation cost for gaussian and other product measures*, Geometric and Functional Analysis, 6:587–600, 1996.

[37] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.

[38] L. Wu. Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal*, 172:301–376, 2000.