Mather Measures and Ergodic Properties of Kantorovich Operators

Malcolm Bowles* and Nassif Ghoussoub†

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

May 07, 2019, revised on June 20, 2019, 2^{d} revision on October 24, 2019

Abstract

We introduce and study the class of *linear transfers* between probability distributions and *the dual class of Kantorovich operators* between function spaces. Linear transfers can be seen as an extension of convex lower semi-continuous energies on Wasserstein space, of cost minimizing mass transports, as well as many other couplings between probability measures to which Monge-Kantorovich theory does not readily apply. Basic examples include *balayage of measures*, *martingale transports*, *optimal Skorokhod embeddings*, and the *weak mass transports* of Talagrand, Marton, Gozlan and others. The class also includes various stochastic mass transports such as the *Schrödinger bridge* associated to a reversible Markov process, and the Arnold-Brenier variational principle for the incompressible Euler equations.

We associate to most linear transfers, a critical constant, a corresponding *effective linear transfer* and *additive eigenfunctions* to their dual Kantorovich operators, that extend Mané's critical value, Aubry-Mather invariant tori, and Fathi's weak KAM solutions for Hamiltonian systems. This amounts to studying the asymptotic properties of Kantorovich operators, which appear as non-linear counterparts of the Markov operators in classical ergodic theory. This allows for the extension of Mather theory to other settings such as its stochastic counterpart and the framework of ergodic optimization in the holonomic case.

We also introduce *the class of convex transfers*, which includes p-powers ($p \geqslant 1$) of linear transfers, the logarithmic entropy, the Donsker-Varadhan information, optimal mean field plans, and certain free energies as functions of two probability measures, i.e., where the reference measure is also a variable. Duality formulae for general transfer inequalities follow in a very natural way. This paper is an expanded version of a previously posted but not published work by the authors [13].

*This is part of the PhD dissertation of this author at the University of British Columbia.

†Both authors have been partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Contents

1 Introduction

Our main objective is to study the ergodic properties of *Kantorovich operators,* which are at the heart of the theory of mass transport summarized in the books of Villani [64] and Santambrogio [55], as well as the so-called weak KAM theory developed by Mather [49], Fathi [25], Aubry [1], Mané [45] and many others. Consider two compact spaces X and Y, and the corresponding spaces $C(Y)$ (resp., $USC(X)$) of continuous functions on Y (resp., bounded above upper semi-continuous functions on X). A *backward Kantorovich operator* is a map (mostly non-linear) $T^-: C(Y) \to USC(X)$ verifying the following 3 properties:

a) T^- is *monotone*, i.e., $f_1 \leqslant f_2$ in $C(Y)$, then $T^- f_1 \leqslant T^- f_2$.

b) T^- is *a convex operator*, that is for any $\lambda \in [0,1]$, f_1, f_2 in $C(Y)$, we have

$$T^-(\lambda f_1 + (1-\lambda) f_2) \leqslant \lambda T^- f_1 + (1-\lambda) T^- f_2.$$

c) T^- is *affine on the constants*, i.e., for any $c \in \mathbb{R}$ and $f \in C(Y)$, there holds

$$T^-(f+c) = T^- f + c.$$

Forward Kantorovich operators $T^+: C(X) \to LSC(Y)$ are those that verify (a), (c), and the *concave counterpart* of (b), that is

$$T^+(\lambda f_1 + (1-\lambda) f_2) \geqslant \lambda T^+ f_1 + (1-\lambda) T^+ f_2,$$

where $LSC(Y)$ is the space of bounded below lower semi-continuous functions on Y.
We shall say that T^- (resp., T^+) is non-trivial if there is at least one function $f \in C(Y)$ (resp., $C(X)$) such that $T^- f \not\equiv -\infty$ (resp., $T^+ f \not\equiv +\infty$).

Kantorovich operators are important extensions of Markov operators and are ubiquitous in mathematical analysis and differential equations. Even affine operators of the form

$T^- f(x) = Tf(x) - A(x)$, where T is a Markov operator and A is a given function (observable) allows the asymptotic theory of Kantorovich operators to incorporate ergodic optimization for expanding dynamical systems. Non-linear Kantorovich operators also appear for example as the maps that associate to an initial state of a Hamilton-Jacobi equation the solution at a given time t, as general value functions in dynamic programming principles ([26] Section II.3), and also in the mathematical theory of image processing [3].

The rich structure of Kantorovich operators stems from their duality -via Legendre transform- with certain lower semi-continuous and convex functionals $\mathcal{T}$ on $\mathcal{M}(X)\times\mathcal{M}(Y)$, where $\mathcal{M}(K)$ is the space of signed measures on a compact space K equipped with the weak*-topology in duality with $C(K)$. Indeed, to any map $T^- : C(Y) \to USC(X)$ (resp., $T^+ : C(X) \to LSC(Y)$), one can associate a corresponding convex and lower semi-continuous functional $\mathcal{T}_{T^-}$ (resp., $\mathcal{T}_{T^+}$) on $\mathcal{M}(X)\times\mathcal{M}(Y)$ via the following –possibly infinite– expressions: If $(\mu,\nu)\in\mathcal{P}(X)\times\mathcal{P}(Y)$, where $\mathcal{P}(K)$ denotes the space of probability measures on K, then set

$$\mathcal{T}_{T^-}(\mu,\nu) = \sup\Big\{\int_Y g\,d\nu - \int_X T^- g\,d\mu;\ g\in C(Y)\Big\}, \tag{1}$$

(resp.,

$$\mathcal{T}_{T^+}(\mu,\nu) = \sup\Big\{\int_Y T^+ f\,d\nu - \int_X f\,d\mu;\ f\in C(X)\Big\}), \tag{2}$$

If $(\mu,\nu)\notin\mathcal{P}(X)\times\mathcal{P}(Y)$, then set $\mathcal{T}_{T^-}(\mu,\nu)=+\infty$ (resp., $\mathcal{T}_{T^+}(\mu,\nu)=+\infty$).

Dually, we introduce the following notions.

Definition 1.1. Let $\mathcal{T} : \mathcal{M}(X)\times\mathcal{M}(Y)\to\mathbb{R}\cup\{+\infty\}$ be a bounded below functional with a non-empty effective domain $D(\mathcal{T})$.

1. We say that $\mathcal{T}$ is *a backward (resp., forward) linear coupling*, if
$$D(\mathcal{T})\subset\mathcal{P}(X)\times\mathcal{P}(Y), \tag{3}$$
and
$$\mathcal{T}=\mathcal{T}_{T^-}\quad(\text{resp., } \mathcal{T}=\mathcal{T}_{T^+}), \tag{4}$$
for some $T^- : C(Y)\to USC(X)$ (resp., $T^+ : C(X)\to LSC(Y)$).

2. We say that $\mathcal{T}$ is *a backward (resp., forward) linear transfer*, if it is a linear coupling with T^- (resp., T^+) being backward (resp., forward) Kantorovich operators.

It is easy to see that in either case, $\mathcal{T}$ is then a proper, bounded below, lower semi-continuous and convex functional on $\mathcal{M}(X)\times\mathcal{M}(Y)$. Moreover, if we consider for each $\mu\in\mathcal{M}(X)$ (resp., $\nu\in\mathcal{M}(Y)$) the partial maps $\mathcal{T}_\mu$ on $\mathcal{P}(Y)$ (resp., $\mathcal{T}_\nu$ on $\mathcal{P}(X)$) given by $\nu\to\mathcal{T}(\mu,\nu)$ (resp., $\mu\to\mathcal{T}(\mu,\nu)$), their Legendre transforms are then the following functionals on $C(Y)$ (resp., $C(X)$) defined by,

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

and

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

respectively, since $\mathcal{T}_\nu$ and $\mathcal{T}_\mu$ are equal to $+\infty$ whenever μ and ν are not probability measures. Note also that

$$\mathcal{T}_\mu^*(g) \leqslant \int_X T^- g(x)\, d\mu(x) \quad \text{for any } g \in C(Y), \tag{5}$$

(resp.,

$$\mathcal{T}_\nu^*(g) \leqslant -\int_Y T^+(-f)(x)\, d\nu(x) \quad \text{for any } f \in C(X). \tag{6}$$

We shall later prove that we have equality if and only if $\mathcal{T}$ is a linear transfer.

Note that if $\mathcal{T}$ is a backward linear coupling with an operator T^-, then $\tilde{\mathcal{T}}(\mu,\nu) := \mathcal{T}(\nu,\mu)$ is a forward linear coupling with the operator $\tilde{T}^+ f = -T^-(-f)$. We shall therefore focus on the properties of backward linear couplings and transfers since their "forward counterparts" could be derived from that relation. There are however special characteristics to those that are simultaneously forward and backward linear transfers (see Sections 3 and 6). We shall say that *a coupling $\mathcal{T}$ is symmetric* if $\tilde{\mathcal{T}} = \mathcal{T}$, in which case $T^+ f = -T^-(-f)$.

The "partial domain" of $\mathcal{T}$ will be denoted by

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

The following characterization of linear transfers is the starting point of our analysis.

Theorem 1.2. *Let $\mathcal{T} : \mathcal{M}(X) \times \mathcal{M}(Y)$ be a functional such that $D(\mathcal{T}) \subset \mathcal{P}(X) \times \mathcal{P}(Y)$ and $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$. Then, the following are equivalent:*

1. *$\mathcal{T}$ is a backward linear transfer.*
2. *There is a map $T : C(Y) \to USC(X)$ such that for each $\mu \in D_1(\mathcal{T})$, $\mathcal{T}_\mu$ is convex lower semi-continuous on $\mathcal{P}(Y)$ and*

$$\mathcal{T}_\mu^*(g) = \int_X Tg(x)\, d\mu(x) \quad \textit{for any } g \in C(Y). \tag{7}$$

3. *There exists a proper bounded below lower semi-continuous function $c : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ with $\sigma \to c(x,\sigma)$ convex such that for any $(\mu,\nu) \in \mathcal{M}(X) \times \mathcal{M}(Y)$,*

$$\mathcal{T}(\mu,\nu) = \begin{cases} \inf_\pi \{\int_X c(x,\pi_x)\, d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\} & \textit{if } \mu,\nu \in \mathcal{P}(X) \times \mathcal{P}(Y), \\ +\infty & \text{otherwise,} \end{cases} \tag{8}$$

where $\mathcal{K}(\mu,\nu)$ is the set of probability measures π on $X \times Y$ whose marginal on X (resp. on Y) is μ (resp., ν) (i.e., the transport plans), and $(\pi_x)_x$ is the disintegration of π with respect to μ.

This characterization makes a link between linear transfers and mass transport theory, and also explains the terminology we chose. Indeed, the class of linear transfers contains all cost minimizing mass transports, that is functionals on $\mathcal{P}(X) \times \mathcal{P}(Y)$ of the form,

$$\mathcal{T}_c(\mu,\nu) := \inf\Big\{\int_{X\times Y} c(x,y))\, d\pi; \pi \in \mathcal{K}(\mu,\nu)\Big\}, \tag{9}$$

where $c(x,y)$ is a continuous cost function on the product measure space $X \times Y$. A consequence of the Monge-Kantorovich theory is that cost minimizing transports $\mathcal{T}_c$ are both forward and backward linear transfers with Kantorovich operators given for any $f \in C(X)$ (resp., $g \in C(Y)$) by

$$T_c^+ f(y) = \inf_{x\in X}\{c(x,y) + f(x)\} \quad \text{and} \quad T_c^- g(x) = \sup_{y\in Y}\{g(y) - c(x,y)\}. \tag{10}$$

However, many couplings between probability measures cannot be formulated as optimal mass transportation problems, since they 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. Moreover, they are often not symmetric, meaning that the problem imposes a specific direction from one of the marginal distributions to the other. The notion of *transfers* between probability measures is therefore much more encompassing than mass transportation, yet is still amenable to –at least a one-sided version– of the duality theory of Monge-Kantorovich [64].

The notion of linear transfer is general enough to encapsulate all bounded below convex lower semi-continuous functions on Wasserstein space and Markov operators, but also the Choquet-Mokobodzki balayage theory [19, 53], the deterministic version of optimal mass transport (e.g., Villani [64], Ambrosio-Gigli-Savare [4]), their stochastic counterparts (Mikami-Thieulin [52]), Barton-Ghoussoub [8] and others), optimal Skorokhod embeddings (Ghoussoub-Kim-Pallmer [34, 35]), the Schrödinger bridge, and the Arnold-Brenier variational descriptions of the incompressible Euler equation. Linear transfers turned out to be essentially equivalent to the notion of *weak mass transports* recently developed by Gozlan et al. [38, 40]), and motivated by earlier work of Talagrand [62, 63], Marton [47, 48] and others.

This paper has two objectives. First, it 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. In other words, there are by now enough examples that share common structural features that the situation warrants the formalization of their unifying concept. The ultimate purpose is to extend many of the remarkable properties enjoyed by energy functionals on Wasserstein space and standard optimal mass transportations to a larger class of couplings that is stable under addition, convex combinations, convolutions, and tensorizations. We exhibit the basic permanence properties of the convex cones of transfers, and extend several results known for mass transports including general duality formulas for inequalities between various transfers that extend the work of Bobkov-Götze [8], Gozlan-Leonard [38], Maurey [50] and others.

The second objective is to show that the approach of Bernard-Buffoni [6, 7] to the Fathi-Mather weak KAM theory ([25] [49]), which is based on optimal mass transport associated to a cost given by a generating function of a Lagrangian, extend to transfers and therefore applies to other couplings, including stochastic transportation. We do that by associating to any linear transfer a corresponding *effective linear transfer* in the same way that weak KAM theory associates an *effective Lagrangian* (and Hamiltonian) to many problems of the calculus of variations [25, 21]. With such a perspective, Mather theory seems to rely on the ergodic properties of the nonlinear Kantorovich operators as opposed to classical ergodic theory, which deals with linear Markov operators.

We shall focus here on probability measures on compact spaces, even though the right settings for most applications and examples are complete metric spaces, Riemannian manifolds, 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. In the case of $\mathbb{R}^n$, which is the setting for many examples stated below, the right duality is between the space $Lip(\mathbb{R}^n)$ of all bounded and Lipschitz functions and the space of Radon measures with finite first moment.

In Section 3, we study in detail the duality between Kantorovich operators and linear transfers. We actually associate to essentially any map $T : C(Y) \to USC(X)$ (resp., any convex functional $\mathcal{T}$ on $\mathcal{P}(X) \times \mathcal{P}(Y)$) an "optimal" Kantorovich map $\overline{T}$ (resp., linear transfer $\overline{\mathcal{T}}$) that can be seen as "envelopes".

Proposition 1.3. (The transfer envelope of a correlation functional) *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a bounded below lower semi-continuous functional that is convex in each of the variables such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$. Then, there exists a functional $\overline{\mathcal{T}} \geqslant \mathcal{T}$ on $\mathcal{P}(X) \times \mathcal{P}(Y)$ that is the smallest backward linear transfer above $\mathcal{T}$.*

Dually, we say that T^- *is proper at* $x \in X$, if

$$\inf_{\nu\in\mathcal{P}(Y)} \sup_{g\in C(Y)} \left\{ \int_Y g\, d\nu - T^- g(x)\right\} < +\infty. \tag{11}$$

This then implies that $Tf(x) > -\infty$ for every $f \in C(Y)$, and translates into the condition that the associated coupling $\mathcal{T}$ is *proper* as a convex function in the following way:

$$\delta_x \in D_1(\mathcal{T}) := \{\mu \in \mathcal{P}(X); \exists \nu \in \mathcal{P}(Y), \mathcal{T}^-(\mu, \nu) < +\infty\}. \tag{12}$$

Proposition 1.4. (The Kantorovich envelope of a non-linear map) *Let $T : C(Y) \to USC(X)$ be a proper map. Then, there exists $\overline{T} : C(Y) \to USC(X)$ that is the largest Kantorovich operator below T on $C(Y)$.*

In anticipation to the study of the ergodic properties of a Kantorovich operators, where we will need to consider iterates of T, we proceed to extend in Section 4 any Kantorovich operator $T : C(Y) \to USC(X)$ to a map from $USC(Y)$ into $USC(X)$ while retaining properties (a), (b) and (c) that characterize Kantorovich operators.

In section 5, we exhibit a large number of (basic) examples of linear transfers which do not fit in standard mass transport theory. The various optimal *martingale mass transports* and *weak mass transports* of Marton, Gozlan and collaborators are examples of one-directional linear transfers. 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.

In Section 6, we show that the class of linear transfers has remarkable permanence properties under various operations. The most important one for our study is the stability under inf-convolution: If $\mathcal{T}_1$ (resp., $\mathcal{T}_2$) are backward linear transfers on $\mathcal{P}(X_1) \times \mathcal{P}(X_2)$ (resp., $\mathcal{P}(X_2) \times \mathcal{P}(X_3)$), then their *inf-convolution*

$$\mathcal{T}(\mu, \nu) := \mathcal{T}_1 \star \mathcal{T}_2(\mu, \nu) = \inf\{\mathcal{T}_1(\mu, \sigma) + \mathcal{T}_2(\sigma, \nu); \sigma \in \mathcal{P}(X_2)\} \tag{13}$$

is a backward linear transfer on $\mathcal{P}(X_1) \times \mathcal{P}(X_3)$. This leads to an even richer class of transfers, such as *the ballistic stochastic optimal transport*, *broken geodesics of transfers*, and projections onto certain subsets of Wasserstein space.

In anticipation to the extension of Mather theory, and motivated by the work of Bernard-Buffoni [6], we study in Section 7 those linear transfers that are *distance-like*, that is satisfy the *triangular inequality*,

$$\mathcal{T}(\mu,\nu) \leqslant \mathcal{T}(\mu,\sigma) + \mathcal{T}(\sigma,\nu) \quad \text{for all } \mu,\nu,\sigma \in \mathcal{P}(X), \tag{14}$$

as well as the $\mathcal{T}$-Lipschitz functionals on the set $\mathcal{A} = \{\mu \in \mathcal{P}(X); \mathcal{T}(\mu,\mu) = 0\}$.

Theorem 1.5. *Let $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ with T^- as a Kantorovich operator. Assume that $\mathcal{T}$ satisfies (14) and that for all $\mu,\nu \in \mathcal{P}(X)$,*

$$\mathcal{T}(\mu,\nu) = \inf\{\mathcal{T}(\mu,\sigma) + \mathcal{T}(\sigma,\nu);\ \sigma \in \mathcal{A}\}. \tag{15}$$

The following then hold:

1. *A functional Φ on $\mathcal{A}$ is $\mathcal{T}$-Lipschitz if and only if there exists a function $f \in C(X)$ such that*
$$\Phi(\mu) = \int_X f\,d\mu = \int_X T^- f\,d\mu \quad \textit{for all } \mu \in \mathcal{A}. \tag{16}$$
2. *If $\mathcal{T}$ is also a forward transfer with T^+ as a Kantorovich operator, then*
$$\Phi(\mu) = \int_X f\,d\mu = \int_X T^- f\,d\mu = \int_X T^+ \circ T^- f\,d\mu \quad \textit{for all } \mu \in \mathcal{A}. \tag{17}$$

We note that the functions $\psi_0 = T^- f$ and $\psi_1 = T^+ \circ T^- f$ are *conjugate* in the sense that $\psi_0 = T^- \psi_1$ and $\psi_1 = T^+ \psi_0$.

In Sections 8-10 we associate to any given linear transfer $\mathcal{T}$, a distance-like transfer $\mathcal{T}_\infty$, by exploiting the ergodic properties of the corresponding Kantorovich operators. For each $n \in \mathbb{N}$, we let $\mathcal{T}_n = \mathcal{T} \star \mathcal{T} \star \dots . \star \mathcal{T}$ be the transfer obtained from a backward linear transfer $\mathcal{T}$ by iterating its convolution n-times. The Kantorovich operator associated to $\mathcal{T}_n$ is given by the n-th iterate $(T^-)^n$ of the Kantorovich operator T^- associated to $\mathcal{T}$. We will be interested in the limiting behavior of $\mathcal{T}_n$ and $(T^-)^n$ as n goes to infinity. The following identifies a critical constant associated to a given linear transfer.

Theorem 1.6. *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ and let $T := T^-$ be its backward Kantorovich operator. Assume*

$$\mathcal{T}(\mu_0,\mu_0) < +\infty \textit{ for some probability measure } \mu_0. \tag{18}$$

1. *Then, there exists a finite constant $c(\mathcal{T})$ such that*
$$c(\mathcal{T}) := \sup_n \frac{1}{n} \inf_{\mu,\nu \in \mathcal{P}(X)} \mathcal{T}_n(\mu,\nu) = \inf_n \frac{1}{n} \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}_n(\mu,\mu). \tag{19}$$

   It will be called the "Mané constant" associated to $\mathcal{T}$.

2. *It is also characterized by*
$$c(\mathcal{T}) = \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu), \tag{20}$$
and the probabilitiy distributions where the infimum is attained will be called "Mather measures" for $\mathcal{T}$.

3. *Moreover, $c(\mathcal{T})$ is the unique constant for which there may be $u \in C(X)$ such that*
$$Tu + c = u. \tag{21}$$
Such a function u will be called a "backward weak KAM solution" for T.

Similar definitions can be made for forward linear transfers. Actually, when $\mathcal{T}$ is continuous on $\mathcal{P}(X) \times \mathcal{P}(X)$ for the Wasserstein metric, much more can be said since we should be able to associate to $\mathcal{T}$ an *idempotent transfer* $\mathcal{T}_\infty$, i.e., one that verify $\mathcal{T} \star \mathcal{T} = \mathcal{T}$, in which case its corresponding Kantorovich map T_∞ is idempotent for the composition operation (i.e., $T_\infty^2 = T_\infty$), while its range correspond to all weak KAM solutions for $\mathcal{T}$. The most known ones are the Monge optimal mass transport or more generally, the Rubinstein-Kantorovich mass transports, where the cost $c(x, y)$ is a distance on a metric space. In reality, many more examples satisfy this property, such as transfers induced by convex energies with 0 as an infimum, the balayage transfer, and certain optimal Skorokhod embeddings in Brownian motion. The following shows that one can associate such an idempotent transfer under equi-continuity conditions on $\mathcal{T}$.

Theorem 1.7. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ that is continuous for the Wasserstein metric, and let $T := T^- : C(X) \to C(X)$ be the corresponding backward Kantorovich operator. Then, there exist a Mané critical value $c = c(\mathcal{T}) \in \mathbb{R}$ and an idempotent backward linear transfer $\mathcal{T}_\infty$ such that if T_∞ is its corresponding idempotent Kantorovich operator, then the following hold:*

1. *For every $f \in C(X)$ and $x \in X$, $\lim\limits_{n \to +\infty} \frac{T^n f(x)}{n} = -c(\mathcal{T})$:*

2. *$\mathcal{T}_\infty$ is the largest linear transfer below $\liminf_n (\mathcal{T}_n - nc)$ and $\mathcal{T}_\infty = (\mathcal{T} - c) \star \mathcal{T}_\infty$;*

3. *$T \circ T_\infty f + c = T_\infty f$ for all $f \in C(X)$, that is $u := T_\infty f$ is a backward weak KAM solution.*

4. *The set $\mathcal{A} := \{\mu \in \mathcal{P}(X); \mathcal{T}_\infty(\mu, \mu) = 0\}$ is non-empty and for every $\mu, \nu \in \mathcal{P}(X)$, we have*
$$\mathcal{T}_\infty(\mu, \nu) = \inf\{\mathcal{T}_\infty(\mu, \sigma) + \mathcal{T}_\infty(\sigma, \nu), \sigma \in \mathcal{A}\}, \tag{22}$$
and the infimum on $\mathcal{A}$ is attained.

5. *The Mané constant $c(\mathcal{T}) = \inf\{\mathcal{T}(\mu, \mu); \mu \in \mathcal{P}(X)\}$ is attained by a probability $\bar{\mu}$ in $\mathcal{A}$.*

6. *If $\mathcal{T}$ is also a forward transfer, then similar results hold for the forward operator T^+. Moreover, the associated effective transfer $\mathcal{T}_\infty$ can then be expressed as*
$$\mathcal{T}_\infty(\mu, \nu) = \sup\Big\{ \int_X f^+ \, d\nu - \int_X f^- \, d\mu;\ (f^-, f^+) \in \mathcal{I}\Big\}, \tag{23}$$

where

$$\mathcal{I} = \big\{(f^-, f^+);\ f^- \text{ (resp., } f^+\text{) is a backward (resp., forward) weak KAM solution} \\ \text{and } \int_X f^- d\mu = \int_X f^+ d\mu \text{ for all } \mu \in \mathcal{A}\big\}.$$

By analogy with the weak KAM theory of Mather-Aubry-Fathi –briefly described in the next paragraph– we shall say that $\mathcal{T}_\infty$ (resp., T^∞) is *the effective transfer* or the *generalized Peierls barrier* (resp., *effective Kantorovich operator*) associated to $\mathcal{T}$. The set $\mathcal{A}$ is the analogue of the *projected Aubry set*, and

$$\mathcal{D} := \{(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X)\ :\ \mathcal{T}(\mu, \nu) + \mathcal{T}_\infty(\nu, \mu) = c(\mathcal{T})\}$$

can be seen as a *generalized Aubry set* [25].

As mentioned above, the effective transfer $\mathcal{T}_\infty$ is obtained by an infinite inf-convolution process, while T_∞ is obtained by an infinite iteration procedure, which lead to fixed points (additive eigenfunctions) for such a non-linear operator. The same procedure actually applies for any semi-group of backward linear transfers (for the convolution operation) and the corresponding semi-group of Kantorovich maps (for the composition operation). This will be established in Section 7 for an equicontinuous semi-group of backward linear transfers.

In Section 9, we deal with the case of a general linear transfer, where we do not assume continuity of $\mathcal{T}$, but that the corresponding Kantorovich operator T maps $C(X)$ to $USC(X)$. We then consider the following measure of the oscillation of the iterates of $\mathcal{T}$:

$$K(n) := \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}_n(\mu, \mu) - \inf_{\mu, \nu \in \mathcal{P}(X)} \mathcal{T}_n(\mu, \nu). \tag{24}$$

Note that Theorem 1.6 already asserts that $\frac{K(n)}{n}$ decreases to zero, but we shall need a slightly stronger condition to prove in section 8 the existence of weak KAM solutions.

Theorem 1.8. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ such that its corresponding Kantorovich operator maps $C(X)$ to $USC(X)$. Assume (18) and the following two conditions:*

$$\sup_{x \in X} \inf_{\sigma \in \mathcal{P}(X)} \mathcal{T}(x, \sigma) < +\infty, \tag{25}$$

and

$$\liminf_n K(n) < +\infty. \tag{26}$$

1. *Then, there exists a backward weak KAM solution for $\mathcal{T}$ at the level $c := c(\mathcal{T})$.*

2. *The Mané constant c is unique in the following sense*

$$\begin{aligned} c(\mathcal{T}) &= \sup\{d \in \mathbb{R}; \text{there exists } u \in USC(X) \text{ with } Tu + d \leqslant u\} \\ &= \inf\{d \in \mathbb{R}; \text{there exists } v \in USC(X) \text{ with } Tv + d \geqslant v\}. \end{aligned} \tag{27}$$

Note that (25) merely states that the function $T1$ is bounded below, while (18) yields that $T1$ is not identically $-\infty$. This will allow us to prove the following.

Theorem 1.9. *Let $\mathcal{T}$ be a backward linear transfer that is also bounded above on $\mathcal{P}(X) \times \mathcal{P}(X)$, then*

$$\frac{\mathcal{T}_n(\mu,\nu)}{n} \to c \quad \textit{uniformly on } \mathcal{P}(X) \times \mathcal{P}(X). \tag{28}$$

Moreover, there exists an idempotent operator $T_\infty : C(X) \to USC(X)$ such that for each $f \in C(X)$, $T_\infty f$ is a backward weak KAM solution for $\mathcal{T}$.

In Section 10, we use a regularization procedure to show that many of the conclusions in Theorem 1.7 can hold for transfers that are neither necessarily continuous nor bounded. This holds for example when the following condition is satisfied.

$$\inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu,\mu) = \inf_{\mu,\nu \in \mathcal{P}(X)} \mathcal{T}(\mu,\nu), \tag{29}$$

which holds in many situations. This will allow us to prove the following general result.

Theorem 1.10. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(X)$, where X is a bounded domain in $\mathbb{R}^n$. Then, for every $\lambda \in (0,1)$, there exists a convex function φ, a constant $c \in \mathbb{R}$ and a function $g \in USC(X)$ such that*

$$T^- g + c = \lambda\, g(\nabla\varphi) + (1-\lambda)\, g. \tag{30}$$

Note that if φ is the quadratic function, then g is a weak KAM solution for T.

To make the connection with Mather-Aubry-Fathi theory, consider $\mathcal{T}_t$ to be the cost minimizing transport

$$\mathcal{T}_t(\mu,\nu) = \inf\{\int_{M\times M} c_t(x,y)\, \mathrm{d}\pi(x,y)\,;\, \pi \in \mathcal{K}(\mu,\nu)\}, \tag{31}$$

where

$$c_t(x,y) := \inf\{\int_0^t L(\gamma(s),\dot\gamma(s))\, \mathrm{d}s\,;\, \gamma \in C^1([0,t];M), \gamma(0)=x, \gamma(t)=y\}, \tag{32}$$

for some given (time-independent) *Tonelli Lagrangian* L possessing suitable regularity properties on a compact state space M. The backward Kantorovich operators associated to $\mathcal{T}_t$ are nothing but the Lax-Oleinik semi-group S_t^-, $t > 0$, defined as

$$S_t^- u(x) := \inf\{u(\gamma(0)) + \int_0^t L(\gamma(s),\dot\gamma(s))\, \mathrm{d}s\,;\, \gamma \in C^1([0,t];M), \gamma(t)=x\}. \tag{33}$$

Recall from [25] that a function $u \in C(M)$ is said to be a *negative weak KAM solution* if for some $c \in \mathbb{R}$, we have

$$S_t^- u + ct = u \quad \text{for all } t \geqslant 0, \tag{34}$$

these solutions are then given by any function in the range of the effective Kantorovich map associated to $(S_t^-)_t$. Actually, these solutions were obtained this way by Bernard and Buffoni [6, 7], who capitalized on the fact that in this case, the transfers $(\mathcal{T}_t)_t$ are actually given by optimal mass transports associated to the cost c_t, and that the Lax-Oleinik semi-groups are obtained via Monge-Kantorovich theory. This general asymptotic theory applies to both the linear setting such as the heat semi-group and to non-linear contexts including the Schrödinger bridge. It also applies to settings where transfers are neither given by optimal transport problems nor are they continuous on Wasserstein space.

In section 11, we apply the general theory to the following semi-group of stochastic optimal mass transports: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with normal filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$, and define $\mathcal{A}_{[0,t]}$ to be the set of continuous semi-martingales $X : \Omega \times [0,t] \to M$ such that there exists a Borel measurable drift $\beta : [0,t] \times C([0,t]) \to \mathbb{R}^d$ for which

1. $\omega \mapsto \beta(s,\omega)$ is $\mathcal{B}(C([0,s]))_+$-measurable for all $s \in [0,t]$, where $\mathcal{B}(C([0,s]))$ is the Borel σ-algbera of $C[0,s]$.
2. $W(s) := X(s) - X(0) - \int_0^s \beta(s')\,\mathrm{d}s'$ is a $\sigma(X(s)\,;\, 0 \leqslant s \leqslant t)$ is an M-valued Brownian motion.

For each β, we shall denote the corresponding X by X^β in such a way that

$$dX^\beta(t) = \beta(t)dt + dW(t). \tag{35}$$

The stochastic transport from $\mu \in \mathcal{P}(M)$ to $\nu \in \mathcal{P}(M)$ on the interval $[0,t]$, $t > 0$, is then defined as

$$\mathcal{T}_t(\mu,\nu) := \inf\left\{\mathbb{E}\int_0^t L(X^\beta(s),\beta(s))\,\mathrm{d}s\,;\, X^\beta(0) \sim \mu, X^\beta(t) \sim \nu, X^\beta \in \mathcal{A}_{[0,t]}\right\}. \tag{36}$$

Note that these couplings do not fit in the Monge-Kantorovich framework as they are not optimal mass transportations that correspond to a cost function between two states, but they are backward linear transfers according to our definition thanks to the work of Mikami-Tieullin [52]. In this case, they only have backward Kantorovich operators given by the stochastic Lax-Oleinik operator,

$$S_t f(x) := \sup_{X\in\mathcal{A}_{[0,t]}} \left\{\mathbb{E}\left[\left(f(X(t)) - \int_0^t L(X(s),\beta_X(s,X))\,\mathrm{d}s\right)|X(0) = x\right]\right\}, \tag{37}$$

in such a way that

$$\mathcal{T}_t(\mu,\nu) = \sup\big\{\int_M u(y)\,d\nu(y) - \int_M S_t u(x)\,d\mu(x); u \in C(M)\big\}. \tag{38}$$

In addition, for each end-time $T > 0$, $u(t,x) = S_{T-t}u(x)$ is a viscosity solution to the following backward Hamilton-Jacobi-Bellman equation

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\Delta u(t,x) + H(x,\nabla u(t,x)) &= 0, \quad \text{on } [0,T) \times M \\ \qquad\qquad\qquad\qquad u(T,x) &= u(x) \quad \text{on } M. \end{cases} \tag{39}$$

The existing of corresponding *stochastic weak KAM solutions* (i.e., fixed points for $u \to S_t u + ct$) will then be viscosity solutions of second order stationary Hamilton-Jacobi-Bellman equation

$$\frac{1}{2}\Delta u(x) + H(x,\nabla u(x)) = c, \quad x \in M. \tag{40}$$

We shall consider the case of a torus, already studied by Gomez [36], and capitalize on his work to show that just like in the deterministic case, the Mané constant c, for which there exists a backward weak KAM solution is unique and is connected to a stochastic analogue of Mather's problem via

$$c = \inf\{\mathcal{T}_1(\mu,\mu); \mu \in \mathcal{P}(M)\} = \inf\{\int_{TM} L(x,v)\, dm(x,v); m \in \mathcal{N}_0(TM)\}, \qquad (41)$$

where $\mathcal{N}_0(TM)$ is the set of probability measures m on phase space that verify for every $\varphi \in C^{1,2}([0,1]\times M)$,

$$\int_{[0,1]}\int_{TM}\left[\partial_t\varphi(x,t) + v\cdot\nabla\varphi(x,t) + \frac{1}{2}\Delta\varphi(x,t)\right] dm(x,v)\, dt = \int_{TM}[\varphi(x,1)-\varphi(x,0)]\, dm(x,v). \qquad (42)$$

The *stochastic Mather measures* are those that are minimizing Problem (41).

In section 12, we introduce a natural and richer family of transfers: the class of *convex transfers.*

Definition 1.11. A proper convex and weak* lower semi-continuous functional $\mathcal{T}: \mathcal{M}(X)\times\mathcal{M}(Y) \to \mathbb{R}\cup\{+\infty\}$ is said to be a *backward convex coupling* (resp., *forward convex coupling*), if there exists a family of maps $T_i^-: C(Y) \to USC(X)$ (resp., $T_i^+: C(X) \to LSC(Y)$) such that:
If $(\mu,\nu) \in \mathcal{P}(X)\times\mathcal{P}(Y)$, then

$$\mathcal{T}(\mu,\nu) = \sup\big\{\int_Y g(y)\, d\nu(y) - \int_X T_i^- g(x)\, d\mu(x);\; g \in C(Y), i \in I\big\}, \qquad (43)$$

(resp.,

$$\mathcal{T}(\mu,\nu) = \sup\big\{\int_Y T_i^+ f(y)\, d\nu(y) - \int_X f(x)\, d\mu(x);\; f \in C(X), i \in I\big\}, \qquad (44)$$

If $(\mu,\nu) \notin \mathcal{P}(X)\times\mathcal{P}(Y)$, then $\mathcal{T}(\mu,\nu) + \infty$.

In other words,

$$\mathcal{T}(\mu,\nu) = \sup_{i\in I}\mathcal{T}_i(\mu,\nu), \qquad (45)$$

where each $\mathcal{T}_i$ is a linear transfer on $\mathcal{P}(X)\times\mathcal{P}(Y)$ induced by each T_i^- (resp., T_i^+). Note that we do not assume in general that each T_i^- (resp., T_i^+) is a Kantorovich operator. Typical examples are p-powers (for $p \geqslant 1$) of a linear transfer, which will then be a convex couplings in the same direction. More generally, for any convex increasing real function γ on $\mathbb{R}^+$ and any linear backward (resp., forward) transfer, the map $\gamma(\mathcal{T})$ is a backward (resp., forward) convex coupling. Actually, in this case, each of the associated $\mathcal{T}_i$ can be taken to be a linear transfer.

Note that a convex coupling $\mathcal{T}$ of the form (45) only implies that for $g \in C(Y)$ (resp., $f \in C(X)$),

$$\mathcal{T}_\mu^*(g) \leqslant \inf_{i\in I}\textstyle\int_X T_i^- g(x)\, d\mu(x) \quad \text{and} \quad \mathcal{T}_\nu^*(f) \leqslant \inf_{i\in I}\textstyle\int_Y -T_i^+(-f)(y)\, d\nu(y). \qquad (46)$$

We therefore introduce the following stronger notion.

Definition 1.12. Say that $\mathcal{T}$ is a backward convex transfer (resp., forward convex transfer) if for $g \in C(Y)$ (resp., $f \in C(X)$),

$$\mathcal{T}_\mu^*(g) = \inf_{i\in I}\textstyle\int_X T_i^- g(x)\, d\mu(x) \quad (\text{resp., } \mathcal{T}_\nu^*(f) = \inf_{i\in I}\textstyle\int_Y -T_i^+(-f)(y)\, d\nu(y)). \qquad (47)$$

Again, the $T_i's$ are not necessarily Kantorovich maps, i.e., they don't correspond to Legendre transforms of linear transfers $\mathcal{T}_i's$, however, the map $g \to \inf\limits_{i\in I} \int_X T_i^- g(x)\, d\mu(x)$ does in this case possess the properties of a Legendre transform. We give an example in Section 12 of a convex coupling that is not a convex transfer.

Typical examples of convex backward transfers include *generalized entropies* of the following form, but as a function of both measures, i.e., including the reference measure,

$$\mathcal{T}(\mu,\nu) = \int_X \alpha(\frac{d\nu}{d\mu})\, d\mu, \quad \text{if } \nu << \mu \text{ and } +\infty \text{ otherwise}, \tag{48}$$

whenever α is a strictly convex lower semi-continuous superlinear real-valued function on $\mathbb{R}^+$.

The *Donsker-Varadhan information* is defined as

$$\mathcal{I}(\mu,\nu) := \begin{cases} \mathcal{E}(\sqrt{f},\sqrt{f}), & \text{if } \mu = f\nu, \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty, & \text{otherwise}, \end{cases} \tag{49}$$

where $\mathcal{E}$ is a Dirichlet form with domain $\mathbb{D}(\mathcal{E})$ on $L^2(\nu)$. It is another example of a backward completely convex transfer, since it can also be written as

$$\mathcal{I}(\mu,\nu) = \sup\{\int_X f\, d\nu - \log \|P_1^f\|_{L^2(\mu)};\, f \in C(X)\}, \tag{50}$$

where P_t^f is an associated (Feynman-Kac) semi-group of operators on $L^2(\mu)$.

The important example of the *logarithmic entropy*

$$\mathcal{H}(\mu,\nu) = \int_X \log(\frac{d\nu}{d\mu})\, d\nu, \quad \text{if } \nu << \mu \text{ and } +\infty \text{ otherwise}, \tag{51}$$

is of course one of them, but it is much more as we now focus on a remarkable subset of the class of convex transfers: the class of *entropic transfers*, defined as follows:

Definition 1.13. Let α (resp., β) 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 *β-entropic backward transfer*, if there exists a map $T^- : C(Y) \to USC(X)$ such that for each $\mu \in D_1(\mathcal{T})$, the Legendre transform of $\mathcal{T}_\mu$ on $\mathcal{M}(Y)$ satisfies:

$$\mathcal{T}_\mu^*(g) = \beta\left(\int_X T^- g(x)\, d\mu(x)\right) \quad \text{for any } g \in C(Y). \tag{52}$$

- $\mathcal{T}$ is an *α-entropic forward transfer*, if there exists a map $T^+ : C(X) \to LSC(Y)$ such that for each $\nu \in D_2(\mathcal{T})$, the Legendre transform of $\mathcal{T}_\nu$ on $\mathcal{M}(X)$ satisfies:

$$\mathcal{T}_\nu^*(f) = -\alpha\left(\int_Y T^+(-f)(y)\, d\nu(y)\right) \quad \text{for any } f \in C(X). \tag{53}$$

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

$$\mathcal{T}(\mu,\nu) = \sup\left\{\alpha\left(\int_Y T^+ f(y)\, d\nu(y)\right) - \int_X f(x)\, d\mu(x);\, f \in C(X)\right\}, \tag{54}$$

while if $\mathcal{T}$ is a β-entropic backward transfer, then

$$\mathcal{T}(\mu,\nu)=\sup\Big\{\int_Y g(y)\,d\nu(y)-\beta\left(\int_X T^- g(x)\,d\mu(x)\right);\ g\in C(Y)\Big\}. \tag{55}$$

Again, the associated maps T^- and T^+ are not necessarily Kantorovich maps, however, the map $g\to\beta\left(\int_X T^-g(x)\,d\mu(x)\right)$ and $f\to\alpha\left(\int_X T^+f(x)\,d\nu(x)\right)$ inherit special (convexity and lower semi-continuity) properties from the fact that they are Legendre transforms.

We observe in Section 12 that entropic transfers are completely convex transfers. A typical example is of course the logarithmic entropy, since it can be written as

$$\mathcal{H}(\mu,\nu)=\sup\{\int_X f\,d\nu-\log(\int_X e^f\,d\mu);\ f\in C(X)\}, \tag{56}$$

making it a log-entropic backward transfer. More examples of α-entropic forward transfers and β-entropic backward transfers can be obtained by convolving entropic transfers with linear transfers of the same direction.

In section 13, 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, such as Maurey-type inequalities of the following type [50]: Given linear transfers $\mathcal{T}_1,\mathcal{T}_2$, entropic transfers $\mathcal{H}_1,\mathcal{H}_2$ and a convex transfer $\mathcal{F}$, find a reference pair $(\mu,\nu)\in\mathcal{P}(X_1)\times\mathcal{P}(X_2)$ such that

$$\mathcal{F}(\sigma_1,\sigma_2)\leqslant\lambda_1\mathcal{T}_1\star\mathcal{H}_1(\sigma_1,\mu)+\lambda_2\mathcal{T}_2\star\mathcal{H}_2(\sigma_2,\nu)\quad\text{for all }(\sigma_1,\sigma_2)\in\mathcal{P}(X_1)\times\mathcal{P}(X_2). \tag{57}$$

This is then equivalent to the non-negativity of an expression of the form $\tilde{\mathcal{E}}_1\star(-\mathcal{T})\star\mathcal{E}_2$, which could be obtained from the following dual formula:

$$\tilde{\mathcal{E}}_1\star(-\mathcal{F})\star\mathcal{E}_2\,(\mu,\nu)=\inf_{i\in I}\inf_{f\in C(X_3)}\left\{\alpha_1\Big(\int_{X_1}E_1^+\circ F_i^- f\,d\mu\Big)+\alpha_2\Big(\int_{X_2}E_2^+(f)\,d\nu\Big)\right\}, \tag{58}$$

where $\mathcal{F}$ is a convex backward transfer on $Y_1\times Y_2$ with Kantorovich family $(F_i^-)_{i\in I}$, $\mathcal{E}_1$ (resp., $\mathcal{E}_2$) is a forward α_1-transfer on $Y_1\times X_1$ (resp., a forward α_2-transfer on $Y_2\times X_2$) with Kantorovich operator E_1^+ (resp., E_2^+).

2 First examples of linear mass transfers

The class of linear transfers is quite large and ubiquitous in analysis.

2.1 Convex energies on Wasserstein space are linear transfers

The class of linear transfers is a natural extension of the convex energies on Wasserstein space.

Example 2.1: Convex energies

If $I:\mathcal{P}(Y)\to\mathbb{R}$ is a bounded below convex weak*-lower semi-continuous functions on $\mathcal{P}(Y)$. One can then associate a backward linear transfer

$$\mathcal{T}(\mu,\nu)=I(\nu)\quad\text{for all }(\mu,\nu)\in\mathcal{P}(X)\times\mathcal{P}(Y), \tag{59}$$

in such a way that the corresponding Kantorovich map is $T^- : C(Y) \to \mathbb{R} \subset C(X)$ is $T^- f(x) = I^*(f)$ for every $x \in X$.

For example, if I is the linear functional $I(\nu) = \int_Y V(y)\, d\nu(y)$, where V is a lower semi-continuous potential on Y, then for every $x \in X$,

$$T^- f(x) = \sup_{y \in Y} (f(y) - V(y)).$$

If I is the relative entropy with respect to Lebesgue measure, that is $I(\nu) = \int_Y \log \frac{d\nu}{dy} dy$ when ν is absolutely continuous with respect to Lebesgue measure and $+\infty$ otherwise, then it induces a linear transfer with backward Kantorovich map being for all x,

$$T^- f(x) = \log \int_Y e^f\, dy.$$

The same holds for the variance functional $I(\nu) := -\mathrm{var}(\nu) := |\int_Y y\, d\nu|^2 - \int_Y |y|^2\, d\nu(y)$, where the associated Kantorovich map is given by

$$T^- f(x) = \sup\{\widehat{f+q}(z) - |z|^2; z \in Y\},$$

where q is the quadratic function $q(x) = \frac{1}{2}|x|^2$ and $\hat{g}$ is the concave envelope of the function g. See (5.1) below.

2.2 Mass transfers with positively homogenous Kantorovich operators

To any Markov operator, i.e., bounded linear positive operator $T : C(Y) \to C(X)$ such that $T1 = 1$, one can associate a backward linear transfer in the following way:

$$\mathcal{T}_T(\mu, \nu) = \begin{cases} 0 & \text{if } T^*(\mu) = \nu \\ +\infty & \text{otherwise,} \end{cases} \tag{60}$$

where $T^* : \mathcal{M}(X) \to \mathcal{M}(Y)$ is the adjoint operator. It is then easy to see that $T^- = T$ is the corresponding backward Kantorovich map. If now $\pi_x = T^*(\delta_x)$, then one can easily see that $T^- f(x) = \int_Y f(y) d\pi_x(y)$ and that

$$\mathcal{T}_T(\mu, \nu) = 0 \quad \text{if and only if} \quad \nu(B) = \int_X \pi_x(B)\, d\mu(x) \text{ for any Borel } B \subset Y.$$

Conversely, any probability measure π on $X \times Y$ induces a forward and backward linear transfer in the following way:

$$\mathcal{I}_\pi(\mu, \nu) = \begin{cases} 0 & \text{if } \mu = \pi_1 \text{ and } \nu = \pi_2. \\ +\infty & \text{otherwise,} \end{cases} \tag{61}$$

where π_1 (resp., π_2) is the first (resp., second) marginal of π. In this case,

$$T^- f(x) = \int_Y f(y) d\pi_x(y) \quad \text{and} \quad T^+ f(y) = \int_X f(x) d\pi_y(x), \tag{62}$$

where $(\pi_x)_x$ (resp., $(\pi_y)_y$) is the disintegration of π with respect to π_1 (resp., π_2). Note however, that we don't necessarily have here that $x \to \pi_x$ is weak*-continuous, that is T maps $L^1(Y, \pi_2) \to L^1(X, \pi_1)$ and not necessarily $C(Y)$ to $C(X)$.

Example 2.2: The prescribed push-forward transfer
If σ is a continuous map from X to Y, then

$$\mathcal{I}_\sigma(\mu, \nu) = \begin{cases} 0 & \text{if } \sigma_\# \mu = \nu \\ +\infty & \text{otherwise,} \end{cases} \tag{63}$$

is a backward linear transfer with Kantorovich operator given by $T^- f = f \circ \sigma$.

The identity transfer corresponds to when $X = Y$ and $\sigma(x) = x$, in which case the corresponding Kantorovich operators are the identity map, that is $T^+ f = T^- f = f$.

Example 2.3: The prescribed Balayage transfer
Given a convex cone of continuous functions $\mathcal{A} \subset C(X)$, where X is a compact space, one can define an order relation between probability measures μ, ν on X, called the $\mathcal{A}$-balayage, in the following way.

$$\mu \prec_{\mathcal{A}} \nu \quad \text{if and only if} \quad \int_X \varphi \, d\mu \leqslant \int_X \varphi \, d\nu \text{ for all } \varphi \text{ in } \mathcal{A}.$$

Suppose now that $T : C(X) \to C(X)$ is a Markov operator such that $\delta_x \prec_{\mathcal{A}} \pi_x := T^*(\delta_x)$ for all $x \in X$, we will then call it – as well as its associated transfer $\mathcal{T}_T$ – an $\mathcal{A}$-dilation. Similarly, a probability measure π on $X \times X$ is *an $\mathcal{A}$-dilation* if $\delta_x \prec_{\mathcal{A}} \pi_x$, where $(\pi_x)_x$ is the disintegration of π with respect to its first marginal π_1. To each $\mathcal{A}$-dilation π, one can define a backward linear transfer as above.

Example 2.4: The prescribed Skorokhod transfer
Writing $Z \sim \rho$ if Z is a random variable with distribution ρ, and letting $(B_t)_t$ denote Brownian motion, and $\mathcal{S}$ the corresponding class of –possibly randomized– stopping times. For a fixed $\tau \in \mathcal{S}$, one can associate a backward linear transfer in the following way:

$$\mathcal{T}_\tau(\mu, \nu) = \begin{cases} 0 & \text{if } B_0 \sim \mu \text{ and } B_\tau \sim \nu. \\ +\infty & \text{otherwise.} \end{cases} \tag{64}$$

Its backward Kantorovich operator is then $T^- f(x) = \mathbb{E}^x[f(B_\tau)]$, where the expectation is with respect to Brownian motion satisfying $B_0 = x$.

2.3 Optimal linear transfers with zero cost

Let $\mathcal{C}$ be a class of positive bounded linear operators T from $C(Y) \to C(X)$ such that $T1 = 1$. We can then consider the following correlation,

$$\mathcal{T}_{\mathcal{C}}(\mu, \nu) = \begin{cases} 0 & \text{if there exists } T \in \mathcal{C} \text{ with } T^*(\mu) = \nu \\ +\infty & \text{otherwise.} \end{cases} \tag{65}$$

In other words,

$$\mathcal{T}_{\mathcal{C}}(\mu, \nu) = \inf\{\mathcal{T}_T(\mu, \nu); T \in \mathcal{C}\}. \tag{66}$$

We now give a few interesting examples, where $\mathcal{T}_{\mathcal{C}}$ is again a linear mass transfer.

Example 2.5: The null transfer

This is simply the map $\mathcal{N}(\mu,\nu)=0$ for all probability measures μ on X and ν on Y. It is easy to see that it is both a backward and forward linear transfer with Kantorovich operators,

$$T^- f \equiv \sup_{y\in Y} f(y) \quad \text{and} \quad T^+ f \equiv \inf_{x\in X} f(x). \tag{67}$$

Note that

$$\begin{aligned}\mathcal{N}(\mu,\nu) &= \inf\{\mathcal{T}_\sigma(\mu,\nu); \sigma\in C(X;Y)\}\\ &= \inf\{\mathcal{T}_\pi(\mu,\nu); \pi \text{ is a transfer plan on } X\times Y\}\\ &= \inf\{\mathcal{T}_T(\mu,\nu); T\in C(Y.X), T \text{ positive and } T1=1\},\end{aligned}$$

where $\mathcal{I}_\sigma$ and $\mathcal{I}_\pi$ are the push-forward transfers defined in Example 2.2. This is a particular case, i.e., when the cost is trivial, of a relaxation result of Kantorovich (e.g., see Villani [64]).

Example 3.6: The Balayage transfer

Let $\mathcal{A}$ be a proper closed convex cone in $C(X)$, and define now *the balayage transfer* $\mathcal{B}$ on $\mathcal{P}(X)\times\mathcal{P}(X)$ via

$$\mathcal{B}(\mu,\nu)=\begin{cases} 0 & \text{if } \mu\prec_{\mathcal{A}}\nu\\ +\infty & \text{otherwise.}\end{cases} \tag{68}$$

A generalized version of a Theorem of Strassen [60] yields the following relations:

Proposition 2.1. *Assume the cone $\mathcal{A}$ separates the points of X and that it is stable under finite suprema. Then, for any two probability measures μ,ν on X, the following are equivalent:*

1. $\mu\prec_{\mathcal{A}}\nu$.
2. *There exists an $\mathcal{A}$-dilation π on $X\times X$ such that $\mu=\pi_1$ and $\nu=\pi_2$.*

From this follows that

$$\mathcal{B}(\mu,\nu)=\inf\{\mathcal{B}_\pi(\mu,\nu); \pi \text{ is an } \mathcal{A}\text{-dilation}\}. \tag{69}$$

Moreover, a generalization of Choquet theory developed by Mokobodoski and others [53] yields that for every $\mu\in\mathcal{P}(X)$, we have

$$\sup\{\int_X f\, d\sigma;\, \mu\prec_{\mathcal{A}}\sigma\}=\int_X \hat{f}\, d\mu,$$

where

$$\hat{f}(x)=\inf\{g(x); g\in -\mathcal{A},\, g\geqslant f \text{ on } X\}=\sup\{\int_X f d\sigma;\, \epsilon_x\prec_{\mathcal{A}}\sigma\}.$$

It follows that $\mathcal{B}^*_\mu(f)=\int_X \hat{f}\, d\mu$, which means that $\mathcal{B}$ is a backward linear transfer whose Kantorovich operator is $T^- f=\hat{f}$.

$\mathcal{B}$ is also a forward linear transfer with a forward Kantorovich operator is $T^+ f=\check{f}$, where

$$\check{f}(x)=\sup\{h(x); h\in\mathcal{A},\, h\leqslant f \text{ on } X\}=\inf\{\int_X f d\sigma;\, \epsilon_x\prec_{\mathcal{A}}\sigma\}.$$

- A typical example is when X is a convex compact space in a locally convex topological vector space and $\mathcal{A}$ is the cone of continuous convex functions. In this case, $T^- f = \hat{f}$ (resp., $T^+ f = \check{f}$) is the concave (resp., convex) envelope of f, and which was the context of the original Choquet theory.

- If X is a bounded subset of a normed space $(E, \|\cdot\|)$, then $\mathcal{A}$ can be taken to be the cone of all norm-Lipschitz convex functions.

- If X is an interval of the real line, then one can consider $\mathcal{A}$ to be the cone of increasing functions.

- If X is a pseudo-convex domain of $\mathbb{C}^n$, then one can take $\mathcal{A}$ to be the cone of Lipschitz plurisubharmonic functions (see [30]). In this case, if φ is a Lipschitz function, then the Lipschitz plurisubharmonic envelope of φ, i.e., the largest Lipschitz PSH function below φ is given by the formula

$$\check{\varphi}(x) = \inf\{\int_0^{2\pi} \varphi(P(e^{i\theta}) \frac{d\theta}{2\pi}; P : \mathbb{C} \to X \text{ polynonial with } P(0) = x\}.$$

  Note that $\hat{\varphi} = -\check{\psi}$, where $\psi = -\varphi$.

Example 2.7: The Skorokhod transfer

Again, letting $\mathcal{S}$ be the class of –possibly randomized– Brownian stopping times, and define

$$\mathcal{S}K(\mu, \nu) = \begin{cases} 0 & \text{if } B_0 \sim \mu \text{ and } B_\tau \sim \nu \text{ for some } \tau \in \mathcal{S}, \\ +\infty & \text{otherwise.} \end{cases} \tag{70}$$

The following is a classical result of Skorokhod. See, for example [32] for a proof in higher dimension.

Proposition 2.2. *Let $\mathcal{A}$ be the cone of Lipschitz subharmonic functions on a domain Ω in $\mathbb{R}^n$. Then, the following are equivalent for two probability measures μ and ν on Ω.*

1. *$\mu \prec_{\mathcal{A}} \nu$ (i.e, μ and ν are in subharmonic order).*

2. *There exists a stopping time $\tau \in \mathcal{S}$ such that $B_0 \sim \mu$ and $B_\tau \sim \nu$.*

This means that $\mathcal{S}K$ is a backward linear transfer with Kantorovich operator given by $T^- f = f_{**}$, which is the smallest Lipschitz superharmonic function above f. This can also be written as $T^- f = J_f$, where $J_f(x)$ is a viscosity solution for the heat variational inequality,

$$\max\{f(x) - J(x), \Delta J(x)\} = 0. \tag{71}$$

Another representation for J_f is given by the following dynamic programming principle,

$$J_f(x) := \sup_{\tau \in \mathcal{S}} \mathbb{E}^x \Big[f(B_\tau)\Big]. \tag{72}$$

2.4 Mass transfers minimizing a transport cost between two points

The examples in this subsection correspond to cost minimizing transfers, where a cost $c(x,y)$ of moving state x to y is given.

Example 2.8: Monge-Kantorovich transfers

Any proper, bounded below, function c on $X \times Y$ determines a backward and forward linear transfer. This is Monge-Kantorovich theory of optimal transport. One associates the map $\mathcal{T}_c$ on $\mathcal{P}(X) \times \mathcal{P}(Y)$ to be the optimal mass transport between two probability measures μ on X and ν on Y, that is

$$\mathcal{T}_c(\mu,\nu) := \inf\big\{\int_{X\times Y} c(x,y))\,d\pi; \pi \in \mathcal{K}(\mu,\nu)\big\}, \tag{73}$$

where $\mathcal{K}(\mu,\nu)$ is the set of probability measures π on $X \times Y$ whose marginal on X (resp. on Y) is μ (resp., ν) *(i.e., the transport plans)*. Monge-Kantorovich theory readily yields that $\mathcal{T}_c$ is a linear transfer. Indeed, if we define the operators

$$T_c^+ f(y) = \inf_{x\in X}\{c(x,y) + f(x)\} \quad \text{and} \quad T_c^- g(x) = \sup_{y\in Y}\{g(y) - c(x,y)\}, \tag{74}$$

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

$$\begin{aligned} \mathcal{T}_c(\mu,\nu) &= \sup\big\{\int_Y T_c^+ f(y)\,d\nu(y) - \int_X f(x)\,d\mu(x);\ f \in C(X)\big\} \\ &= \sup\big\{\int_Y g(y)\,d\nu(y) - \int_X T_c^- g(x)\,d\mu(x);\ g \in C(Y)\big\}. \end{aligned}$$

This means that the Legendre transform $(\mathcal{T}_c)^*_\mu(g) = \int_X T_c^- g(x)\,d\mu(x)$ and T_c^- is the corresponding backward Kantorovich operator. Similarly, $(\mathcal{T}_c)^*_\nu(f) = -\int_Y T_c^+(-f)(y)\,d\nu(y)$ on $C(X)$ and T_c^+ is the corresponding forward Kantorovich operator. See for example Villani [64].

Example 2.9: The trivial Kantorovich transfer

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

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

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

Example 2.10: The Kantorovich-Rubinstein transport

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

$$\mathcal{T}(\mu,\nu) = \|\nu - \mu\|^*_{\rm Lip} := \sup\left\{\int_X u\,d(\nu-\mu); u \text{ measurable}, \|u\|_{\rm Lip} \leqslant 1\right\} \tag{75}$$

is a linear transfer, where here $\|u\|_{\rm Lip} := \sup_{x\neq y} \frac{|u(y)-u(x)|}{d(x,y)}$. The corresponding forward Kantorovich operator is then the Lipschitz 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 2.11: The Brenier-Wasserstein distance [15]

We mention this important example even though it is not in a compact setting. If $c(x,y) = \langle x, y\rangle$ on $\mathbb{R}^d \times \mathbb{R}^d$, and μ, ν are two probability measures of compact support on $\mathbb{R}^d$, then

$$\mathcal{W}_2(\mu,\nu) = \inf\Big\{\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x, y\rangle \, d\pi; \pi \in \mathcal{K}(\mu,\nu)\Big\}.$$

Here, the Kantorovich operators are

$$T^+ f(x) = -f^*(-x) \quad \text{and} \quad T^- g(y) = (-g)^*(-y), \tag{76}$$

where f^* is the convex Legendre transform of f.

Example 2.12: Optimal transport for a cost given by a generating function (Bernard-Buffoni [6])

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:

$$c^L(y,x) := \inf\{\int_0^1 L(t, \gamma(t), \dot\gamma(t))\, dt; \gamma \in C^1([0,1),M); \gamma(0) = y, \gamma(1) = x\}, \tag{77}$$

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 μ and ν are two probability measures on M, then

$$\mathcal{T}_L(\mu,\nu) := \inf\Big\{\int_{M\times M} c^L(y,x)\, d\pi; \pi \in \mathcal{K}(\mu,\nu)\Big\}$$

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

$$V_f(t,x) = \inf\Big\{f(\gamma(0)) + \int_0^t L(s,\gamma(s),\dot\gamma(s))\, ds; \gamma \in C^1([0,1),M); \gamma(t) = x\Big\}. \tag{78}$$

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

$$\begin{cases} \partial_t V + H(t,x,\nabla_x V) &= 0 \text{ on } [0,1]\times M, \\ V(0,x) &= f(x). \end{cases} \tag{79}$$

Similarly, the backward Kantorovich potential is given by $T_1^- g(y) = W_g(0,y)$, $W_g(t,y)$ being the value functional

$$W_g(t,y) = \sup\Big\{g(\gamma(1)) - \int_t^1 L(s,\gamma(s),\dot\gamma(s))\, ds; \gamma \in C^1([0,1),M); \gamma(t) = y\Big\}, \tag{80}$$

which is a solution for the backward Hamilton-Jacobi equation

$$\begin{cases} \partial_t W + H(t,x,\nabla_x W) &= 0 \text{ on } [0,1]\times M, \\ W(1,y) &= g(y). \end{cases} \tag{81}$$

3 Envelopes and representation of Linear transfers

The following relates mass transfers with the optimal weak transports of Gozlan et al. [40].

3.1 Representation of linear transfers as weak transports

Theorem 3.1. *Let $\mathcal{T} : \mathcal{M}(X) \times \mathcal{M}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a functional such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$. Then, the following are equivalent:*

1. *$\mathcal{T}$ is a backward linear transfer.*
2. *There is a map $T : C(Y) \to USC(X)$, such that for each $\mu \in D_1(\mathcal{T})$, $\mathcal{T}_\mu$ is convex lower semi-continuous on $\mathcal{P}(Y)$ and*
$$\mathcal{T}_\mu^*(g) = \textstyle\int_X Tg(x)\, d\mu(x) \quad \textit{for any } g \in C(Y). \tag{82}$$
3. *There exists a bounded below lower semi-continuous function $c : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ with $\sigma \to c(x, \sigma)$ convex such that for any pair $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$,*
$$\mathcal{T}(\mu,\nu) = \begin{cases} \inf_\pi\{\int_X c(x,\pi_x)\, d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\}, & \textit{if } \mu,\nu \in \mathcal{P}(X)\times\mathcal{P}(Y),\\ +\infty & \text{otherwise.}\end{cases} \tag{83}$$
where $(\pi_x)_x$ is the disintegration of π with respect to μ.

The proof of this theorem will be split in Propositions 3.2, 3.3 and 3.4, where we can provide more details about the needed conditions. The first establishes the easy equivalence between (1) and (2).

Proposition 3.2. *1) If $\mathcal{T}$ is a backward linear transfer with Kantorovich operator T, then*
$$\mathcal{T}_\mu^*(g) = \textstyle\int_X Tg(x)\, d\mu(x) \quad \textit{for any } g \in C(Y). \tag{84}$$
2) Conversely, if $\mathcal{T}$ satisfies (2) in the above Theorem, then $\mathcal{T}$ is a backward linear transfer and T is a Kantorovich operator.

Proof: 1) Since
$$\mathcal{T}(\mu,\nu) = \begin{cases} \sup\{\int_Y g\, d\nu - \int_X T^- g\, d\mu;\ g \in C(Y)\} & \text{if } \mu,\nu \in \mathcal{P}(X)\times\mathcal{P}(Y),\\ +\infty & \text{otherwise.}\end{cases} \tag{85}$$
we have that $\mathcal{T}_\mu \geqslant \Gamma^*_{T,\mu}$, where $\Gamma_{T,\mu}$ is the convex lower semi-continuous function on $C(Y)$ defined by $\Gamma_{T,\mu}(g) = \int_X Tg(x)\, d\mu(x)$ since T is a Kantorovich operator. Moreover, $\mathcal{T}_\mu = \Gamma^*_{T,\mu}$ on the probability measures on Y. If now ν is a positive measure with $\lambda := \nu(Y) > 1$, then
$$\begin{aligned}\Gamma^*_{T,\mu}(\nu) &= \sup\Big\{\int_Y g(y)\, d\nu(y) - \int_X Tg(x)\, d\mu(x);\ g \in C(Y)\Big\}\\ &\geqslant n\lambda - \int_X T(n)\, \mathrm{d}\mu\\ &= n(\lambda-1) - \int_X T(0)\, \mathrm{d}\mu,\end{aligned}$$

where we have used property (3) to say that $T(n) = n + T(0)$. Hence $\Gamma^*_{T,\mu}(\nu) = +\infty$. A similar reasoning applies to when $\lambda < 1$, and it follows that $\mathcal{T}_\mu = \Gamma^*_{T,\mu}$ and therefore $(\mathcal{T}_\mu)^* = \Gamma^{**}_{T,\mu} = \Gamma_{T,\mu}$ since the latter is convex and lower semi-continuous on $C(Y)$.

2) Conversely, it is clear that since $\mathcal{T}_\mu$ is convex lower semi-continuous, we have for any $\mu \in D_1(\mathcal{T})$,

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

Moreover, $Tg(x) = (\mathcal{T}_{\delta_x})^*(g)$, which easily implies that T is a Kantorovich operator.

3.2 Linear transfers and Kantorovich operators as envelopes

We now associate to any convex lower semi-continuous functional on $\mathcal{P}(X) \times \mathcal{P}(Y)$ a backward and a forward linear transfer. This is closely related to the work of Gozlan et al. [40], who introduced the notion of *weak transport*. These are 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. We now show that this notion is essentially 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 $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$. We shall prove the following.

Proposition 3.3. *Let $c : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a bounded below, lower semi-continuous function such that $\sigma \to c(x,\sigma)$ is convex, and define for any pair $(\mu,\nu) \in \mathcal{M}(X) \times \mathcal{M}(Y)$, the functional*

$$\mathcal{T}_c(\mu,\nu) = \begin{cases} \inf_\pi\{\int_X c(x,\pi_x)\,d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\}, & \text{if } \mu,\nu \in \mathcal{P}(X) \times \mathcal{P}(Y), \\ +\infty & \text{otherwise.} \end{cases} \tag{86}$$

where $(\pi_x)_x$ is the disintegration of π with respect to μ.

Then, $\mathcal{T}_c$ is a backward linear transfer with Kantorovich operator

$$T_c^- g(x) = \sup\{\int_Y g(y)\,d\sigma(y) - c(x,\sigma); \sigma \in \mathcal{P}(Y)\}. \tag{87}$$

Proof: We first compute the Legendre transform of the functional $(\mathcal{T}_c)_\mu$. Since $\mathcal{T}_c$ is $+\infty$ outside of the probability measures, we can write

$$\begin{aligned}
(\mathcal{T}_c)^*_\mu(g) &= \sup\{\int_Y g\,d\nu - \mathcal{T}_c(\mu,\nu); \nu \in \mathcal{P}(Y)\} \\
&= \sup\{\int_Y g(y)\,d\nu(y) - \int_X c(x,\pi_x)\,d\mu(x); \nu \in \mathcal{P}(Y), \pi \in \mathcal{K}(\mu,\nu)\} \\
&= \sup\{\int_X \int_Y g(y) d\pi_x(y)\,d\mu(x) - \int_X c(x,\pi_x)\,d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\} \\
&\leqslant \sup\{\int_X \int_Y g(y) d\sigma(y)\,d\mu(x) - \int_X c(x,\sigma)\,d\mu(x); \sigma \in \mathcal{P}(Y)\} \\
&\leqslant \int_X \{\sup_{\sigma \in \mathcal{P}(Y)} \{\int_Y g(y) d\sigma(y) - c(x,\sigma)\,d\mu\}\} \\
&= \int_X T_c^- g(x) d\mu(x).
\end{aligned}$$

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 $T_c^- g(x) = \int_Y g(y) d\bar{\pi}_x(y) - c(x, \bar{\pi}_x)$ for every $x \in X$. It follows that

$$\begin{aligned}(\mathcal{T}_c)_\mu^*(g) &= \sup\{\int_Y g\, d\nu - \mathcal{T}_c(\mu,\nu); \nu \in \mathcal{P}(Y)\} \\ &= \sup\{\int_Y g\, d\nu - \int_X c(x,\pi_x) d\mu(x);\ \nu \in \mathcal{P}(Y), \pi \in \mathcal{K}(\mu,\nu)\}.\end{aligned}$$

Let $\bar{\nu}(A) = \int_X \bar{\pi}_x(A)\, d\mu(x)$. Then, $\bar{\pi}(A \times B) = \int_A \bar{\pi}_x(B)\, d\mu$ belongs to $\mathcal{K}(\mu, \bar{\nu})$, hence

$$\begin{aligned}(\mathcal{T}_c)_\mu^*(g) &\geqslant \int_Y g\, d\bar{\nu} - \int_X c(x,\bar{\pi}_x) d\mu(x) \\ &= \int_Y \int_X g(y) d\bar{\pi}_x(y)\ d\mu(x) - \int_X c(x,\bar{\pi}_x) d\mu(x) \\ &= \int_X \{\int_Y g(y) d\bar{\pi}_x(y) - c(x,\bar{\pi}_x)\}\, d\mu(x) \\ &= \int_X T_c^- g(x) d\mu(x),\end{aligned}$$

hence $(\mathcal{T}_c)_\mu^*(g) = \int_X T_c^- g(x) d\mu(x)$.

We now show that $\mathcal{T}_\mu$ is convex. For that let $\nu = \lambda\nu_1 + (1-\lambda)\nu_2$ and find $(\pi_x^1)_x$ and $(\pi_x^2)_x$ in $\mathcal{P}(Y)$ such that

$$\int_X \pi_x^i d\mu(x) = \nu_i \quad \text{and} \quad \int_X c(x,\pi_x^i)\, d\mu(x) \leqslant \mathcal{T}_c(\mu,\nu_i) + \epsilon \text{ for } i = 1, 2.$$

It is clear that the plan defined by $\pi(A \times B) := \int_A (\lambda\pi_x^1(B) + (1-\lambda)\pi_x^2(B)) d\mu(x)$ belongs to $\mathcal{K}(\mu,\nu)$ and therefore, using the convexity of c in the second variable, we have

$$\begin{aligned}\mathcal{T}_c(\mu,\nu) &\leqslant \int_X c(x,\pi_x)\, d\mu(x) \leqslant \int_X \lambda c(x,\pi_x^i)\, d\mu(x) + \int_X (1-\lambda) c(x,\pi_x^i)\, d\mu(x) \\ &\leqslant \lambda \mathcal{T}_c(\mu,\nu_1) + (1-\lambda)\mathcal{T}_c(\mu,\nu_2) + \epsilon.\end{aligned}$$

It follows that for every $\nu \in \mathcal{P}(Y)$,

$$\mathcal{T}_\mu(\nu) = (\mathcal{T}_c)_\mu^{**} = \sup\{\int_Y f(y) d\nu - \int_X T_c^- f d\mu; f \in C(Y)\}.$$

Moreover, it is easy to see that T_c satisfies properties a), b) and c) of a Kantorovich operator. We can therefore conclude that $\mathcal{T}$ is a backward linear transfer. This establishes that 3) implies 1) in Theorem 3.1. □

That (1) implies 3) in Theorem 3.1 will follow from the following general result.

Proposition 3.4. *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a bounded below lower semi-continuous functional that is convex in each of the variables such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$, and consider $\overline{\mathcal{T}}$ to be the backward linear transfer associated to $c(x,\sigma) = \mathcal{T}(\delta_x, \sigma)$ by the previous proposition, and let*

$$\tilde{\mathcal{T}}(\mu,\nu) := \int_X \mathcal{T}(x,\nu) d\mu(x). \tag{88}$$

Then, $\mathcal{T} \leqslant \overline{\mathcal{T}} \leqslant \tilde{\mathcal{T}}$, and

1. $\overline{\mathcal{T}}$ *is the smallest backward linear transfer greater than* $\mathcal{T}$*.*

2. $\overline{\mathcal{T}}$ *is the largest backward linear transfer smaller than* $\tilde{\mathcal{T}}$*.*

Proof: Note that

$$T^- g(x) = \sup\{\int_Y g(y)\, d\sigma(y) - \mathcal{T}(\delta_x, \sigma); \sigma \in \mathcal{P}(Y)\}, \tag{89}$$

and therefore for each $x \in X$, we have for each $g \in C(Y)$

$$T^- g(x) = (\overline{\mathcal{T}}_{\delta_x})^*(g) = (\mathcal{T}_{\delta_x})^*(g). \tag{90}$$

To show that $\overline{\mathcal{T}} \leqslant \tilde{\mathcal{T}}$, write for an arbitrary $\mu \in \mathcal{P}(X)$,

$$\begin{aligned}
(\overline{\mathcal{T}}_\mu)^*(g) &= \int_X T_c^- g(x) d\mu(x) \\
&= \int_X \sup_\sigma\{\int_Y g\, d\sigma - \mathcal{T}(x, \sigma); \sigma \in \mathcal{P}(Y)\} d\mu(x) \\
&\geqslant \sup_\sigma\{\int_Y g\, d\sigma - \int_X \mathcal{T}(x, \sigma) d\mu(x); \sigma \in \mathcal{P}(Y)\} \\
&= \sup\{\int_Y g\, d\sigma - \tilde{\mathcal{T}}(\mu, \sigma); \sigma \in \mathcal{P}(Y)\} \\
&= (\tilde{\mathcal{T}}_\mu)^*(g),
\end{aligned}$$

hence $\overline{\mathcal{T}} \leqslant \tilde{\mathcal{T}}$ since both of them are convex in the second variable.
Note that $\mathcal{T}(\delta_x, \nu) = \tilde{\mathcal{T}}(\delta_x, \nu)$, hence if $\mathcal{S} \leqslant \tilde{\mathcal{T}}$ and $\mathcal{S}$ is a backward linear transfer with S^- as a Kantorovich operator, then

$$S^- g(x) = (S_{\delta_x})^*(g) \geqslant (\tilde{\mathcal{T}}_{\delta_x})^*(g) = (\overline{\mathcal{T}}^*_{\delta_x}(g) = T^- g(x),$$

and therefore $\mathcal{S} \leqslant \overline{\mathcal{T}}$. It follows that $\overline{\mathcal{T}}$ is the greatest backward linear transfer smaller than $\tilde{\mathcal{T}}$.

To show that $\mathcal{T} \leqslant \overline{\mathcal{T}}$, note that since $\mathcal{T}$ is jointly convex and lower semi-continuous, then for each $f \in C(Y)$, the functional

$$\mu \to (\mathcal{T}_\mu)^*(f) := \sup\{\int_Y f d\sigma - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(Y)\}$$

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

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

hence

$$\mathcal{T}(\mu, \nu) = (\mathcal{T}_\mu)^{**}(\nu) \leqslant \sup\{\int_Y f d\nu - \int_X T^- f d\mu; f \in C(Y)\} = \overline{\mathcal{T}}(\mu, \nu).$$

If now $\mathcal{S} \geqslant \mathcal{T}$, then $\overline{\mathcal{S}} \geqslant \overline{\mathcal{T}}$, and if $\mathcal{S}$ is a linear transfer, then $\mathcal{S} = \overline{\mathcal{S}} \geqslant \overline{\mathcal{T}}$, and therefore $\overline{\mathcal{T}}$ is the smallest backward linear transfer greater than $\mathcal{T}$. □

Remark 3.5. A similar construction can be done to associate a forward linear transfer $\overline{\mathcal{T}}_+$ to a given functional $\mathcal{T}$ on $\mathcal{P}(X)\times\mathcal{P}(Y)$ provided $\{\delta_y; y\in Y\}\subset D_2(\mathcal{T})$. Note that one can then define $\mathcal{T}$ as a backward (resp., forward) linear transfer if $\mathcal{T}=\overline{\mathcal{T}}_-$ (resp., if $\mathcal{T}=\overline{\mathcal{T}}_+$).

Remark 3.6. Any lower continuous convex functional $\mathcal{T}$ on $\mathcal{P}(X)\times\mathcal{P}(Y)$ that is finite on the set of Dirac measures gives rise to a backward and forward optimal mass transport $\mathcal{T}_c(\mu,\nu)$ associated to the cost function $c(x,y)=\mathcal{T}(\delta_x,\delta_y)$. It is then easy to see that

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

hence

$$\mathcal{T}_c(\mu,\nu)\geqslant\overline{\mathcal{T}}(\mu,\nu)\geqslant\mathcal{T}(\mu,\nu). \tag{91}$$

However, the inequality (91) is often strict. Moreover, transfers need not be defined on Dirac measures, a prevalent situation in stochastic transport problems.

Dually, we give the following characterization of Kantorovich operators, which in particular, yields a uniqueness statement for the duality between them and linear transfers.

Theorem 3.7. *Let $T: C(Y)\to USC(X)$ be a map such that for every $x\in X$,*

$$\inf_{\nu\in\mathcal{P}(Y)}\sup_{g\in C(Y)}\big\{\textstyle\int_Y g\,d\nu-Tg(x)\big\}<+\infty. \tag{92}$$

Then, there exists a Kantorovich operator $\overline{T}$ such that

1. *$\overline{T}$ is the largest Kantorovich operator smaller than T on $C(Y)$.*
2. *$\overline{T}$ can be written as*

$$\overline{T}f(x)=\sup_{\sigma\in\mathcal{P}(Y)}\inf_{g\in C(Y)}\{\int_Y (f-g)\,d\sigma+Tg(x)\}. \tag{93}$$

PROOF. Consider the functional on $\mathcal{M}(X)\times\mathcal{M}(Y)$ given by

$$\mathcal{T}(\mu,\nu)=\begin{cases}\sup\big\{\int_Y g\,d\nu-\int_X Tg\,d\mu;\ g\in C(Y)\big\} & \text{if } \mu,\nu\in\mathcal{P}(X)\times\mathcal{P}(Y),\\ +\infty & \text{otherwise.}\end{cases} \tag{94}$$

Note that $\mathcal{T}$ is bounded below, convex, lower semi-continuous functional and condition (92) means that $\{\delta_x; x\in X\}\subset D_1(\mathcal{T})$. Hence Proposition 3.3 applies to yield a backward linear transfer $\overline{\mathcal{T}}$ with a corresponding backward Kantorovich operator defined as $\overline{T}f(x)=\sup\limits_\sigma\{\int f d\sigma-\mathcal{T}(\delta_x,\sigma)\}$. Note now that

$$\begin{aligned}(\mathcal{T}_{\delta_x})^* f&=\overline{T}f(x)\\ &=\sup_\sigma\{\int f d\sigma-\mathcal{T}(\delta_x,\sigma)\}\\ &=\sup_\sigma\inf_g\{\int f d\sigma-\int g d\sigma+Tg(x)\}\\ &\leqslant\inf_g\sup_\sigma\{\int f d\sigma-\int g d\sigma+Tg(x)\}\\ &=\inf_g\{\sup(f-g)+Tg(x)\}\\ &=Tf(x),\end{aligned}$$

If now S is a Kantorovich map such that $S \leqslant T$, then

$$
\begin{aligned}
\overline{T}f(x) &= \sup_\sigma \{ \int f d\sigma - \mathcal{T}(\delta_x, \sigma)\} \\
&= \sup_\sigma \inf_g \{ \int f d\sigma - \int g d\sigma + Tg(x)\} \\
&\geqslant \sup_\sigma \inf_g \{ \int f d\sigma - \int g d\sigma + Sg(x)\} \\
&= \inf_g \sup_\sigma \{ \int f d\sigma - \int g d\sigma + Sg(x)\} \\
&= \inf_g \{ \sup(f-g) + Sg(x)\} \\
&= \inf_g \{ S[\sup(f-g) + g](x)\} \\
&\geqslant Sf(x).
\end{aligned}
$$

where the last three steps used the fact that S satisfies properties (a), (b) and (c) of a Kantorovich operator.

3.3 Powers and recessions of linear transfers

Proposition 3.8. *Let $\mathcal{T}$ be a convex coupling on $\mathcal{P}(X) \times \mathcal{P}(Y)$ of the form*

$$
\mathcal{T}(\mu, \nu) := \sup_{i \in I} \mathcal{T}_i(\mu, \nu) \tag{95}
$$

where for each $i \in I$, $\mathcal{T}_i(\mu, \nu) = \sup\limits_{f \in C(Y)} \{\int_Y f d\nu - \int_X T_i f d\mu\}$ for some map $T_i : C(Y) \to USC(X)$. Assume $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$ and consider the envelope $\overline{\mathcal{T}}$ of $\mathcal{T}$ and the corresponding Kantorvich operator $\overline{T}$. Then,

1. *$\overline{T}$ is given by the formula*

$$
\overline{T}f(x) = \sup_{\sigma \in \mathcal{P}(Y)} \inf_{g \in C(Y)} \{ \int_Y (f-g)\, d\sigma + \inf_i T_i g(x)\}. \tag{96}
$$

and therefore satsifies $\overline{T}f \leqslant \inf_i T_i f$ on $C(Y)$.

2. *If each T_i is a Kantorovich operator, then $\overline{T}f = \inf_i T_i f$ if and only if $f \to \inf\limits_i T_i f(x)$ is convex .*

Proof. Note first that

$$
\begin{aligned}
\mathcal{T}(\mu, \nu) = \sup_i \mathcal{T}_i(\mu, \nu) &= \sup_i \sup_{f \in C(Y)} \{ \int_Y f d\nu - \int_X T_i f d\mu\} \\
&= \sup\{ \int_Y f d\nu - \inf_i \int_X T_i f d\mu; f \in C(Y)\}
\end{aligned}
$$

and $(\mathcal{T}_\mu)^*(f) \leqslant \inf_i \int_X T_i f d\mu$. Moreover, $f \to (\mathcal{T}_\mu)^*(f)$ is the convex envelope of $f \to \inf_i \int_X T_i f d\mu$. If now $\overline{T}$ is the Kantorovich operator for the envelope $\overline{\mathcal{T}}$, then

$$\overline{T}f(x) = \sup_{\sigma \in \mathcal{P}(Y)} \inf_{g \in C(Y)} \{\int_Y (f-g)\, d\sigma + \inf_i T_i f\},$$

and consequently, $\overline{T}f(x) = (\mathcal{T}_{\delta_x})^*(f) \leqslant \inf_i T_i f(x)$.

If $f \to \inf_i T_i f(x)$ is convex and lower semi-continuous for any $x \in X$, then the envelope property of $f \to \overline{T}f(x)$ yield that $\overline{T}f(x) = \inf_i T_i f(x)$. Note that if each T_i is a Kantorovich operator, then $f \to \inf_i T_i f$ satisfies properties (a) and (c) of a Kantorovich operator but not necessarily the convexity assumption (b).

Corollary 3.9. *Let $\mathcal{T}$ be a backward linear transfer with Kantorovich operator T^-.*

1. *If $\alpha : \mathbb{R}^+ \to \mathbb{R}$ is a convex increasing function on $\mathbb{R}$, then $\alpha(\mathcal{T})$ is a backward convex transfer, whose envelope $\overline{\alpha(\mathcal{T})}$ has a Kantorovich operator equal to*

$$T^-_\alpha f = \inf_{s>0}\{sT^-(\frac{f}{s}) + \alpha^{\oplus}(s)\}, \tag{97}$$

where $\alpha^{\oplus}(t) = \sup\{ts - \alpha(s); s \geqslant 0\}$.

2. *In particular, if $\mathcal{W}_c(\mu,\nu) := \inf\big\{\int_{X\times Y} c(x,y))\, d\pi; \pi \in \mathcal{K}(\mu,\nu)\big\}$ is the Monge-Kantorovich transport associated to a cost c, and $p > 1$, then*

$$\mathcal{W}_c^p(\mu,\nu) \leqslant \overline{\mathcal{W}}_c^p(\mu,\nu) = \inf_\pi\{\int_X \mathcal{W}_c^p(\delta_x, \pi_x)\, d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\}.$$

Proof: It suffices to note that $\alpha(t) = \sup\{ts - \alpha^{\oplus}(s); s \geqslant 0\}$, hence

$$\begin{aligned}
\alpha(\mathcal{T}(\mu,\nu)) &= \sup\big\{s\int_Y f\, d\nu - s\int_X T^- f\, d\mu - \alpha^{\oplus}(s);\ s \in \mathbb{R}^+, f \in C(Y)\big\} \\
&= \sup\big\{\int_Y h\, d\nu - s\int_X T^-(\frac{h}{s})\, d\mu - \alpha^{\oplus}(s);\ s \in \mathbb{R}^+, h \in C(Y)\big\}.
\end{aligned}$$

Therefore $\alpha(\mathcal{T})$ is a convex coupling and its envelope $\overline{\alpha(\mathcal{T})}$ has a Kantorovich operator

$$T^-_\alpha f \leqslant \inf_{s>0}\{sT^-(\frac{f}{s}) + \alpha^{\oplus}(s)\}. \tag{98}$$

Note however that for each $s > 0$, $T^-_s f := sT^-(\frac{h}{s}) + \alpha^{\oplus}(s)$ is a backward Kantorovich operator. Moreover, the function $(s,f) \to sT^-(\frac{f}{s}) + \alpha^{\oplus}(s)$ is jointly convex on $\mathbb{R}^+ \times C(Y)$, hence the infimum in s is convex in f and therefore we have equality in (97).

Corollary 3.10. *Let $\mathcal{T}$ be a backward linear transfer with Kantorovich operator T^-. Then, the functional*

$$\mathcal{T}_f(\mu,\nu) = \begin{cases} 0 & \text{if } \mathcal{T}(\mu,\nu) < +\infty \\ +\infty & \text{otherwise,} \end{cases} \tag{99}$$

is a backward linear transfer with Kantorovich operator equal to

$$T^-_r f(x) = \lim_{\lambda\to+\infty} \frac{T^-(\lambda f)(x)}{\lambda}. \tag{100}$$

Proof: Given any bounded below, lower semi-continuous and convex functional $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$, we can consider the Kantorovich operator that generates its backward linear transfer envelope $\bar{\mathcal{T}}$, that is

$$T^- g(x) = \sup\{\int_Y g(y)\, d\sigma(y) - \mathcal{T}(\delta_x, \sigma); \sigma \in \mathcal{P}(Y)\}, \tag{101}$$

and its corresponding recession function

$$T_r^- f(x) = \lim_{\lambda \to +\infty} \frac{T^-(\lambda f)(x)}{\lambda}.$$

It is then clear that T_r^- is a Kantorovich operator and

$$T_r^- f(x) = \sup\{\int_Y f d\sigma; \sigma \in \mathcal{P}(Y), \mathcal{T}(\delta_x, \sigma) < +\infty\}. \tag{102}$$

The corresponding linear transfer

$$\mathcal{T}_r(\mu, \nu) = \sup\big\{\int_Y g(y)\, d\nu(y) - \int_X T_r^- g(x)\, d\mu(x);\ g \in C(Y)\big\}. \tag{103}$$

Since T_r^- is positively homogenous, its associated transfer $\mathcal{T}_r$ can only take the values 0 and $+\infty$. It is also clear that $\mathcal{T}_r$ is the envelope of $\mathcal{T}_f$, and therefore $\mathcal{T}_f \leqslant \bar{\mathcal{T}}_f = \mathcal{T}_r$. We now show that if $\mathcal{T}$ is a linear transfer, then $\mathcal{T}_r \leqslant \mathcal{T}_f$. Indeed, assume that $\mathcal{T}_f(\mu, \nu) = 0$, then $\mathcal{T}(\mu, \nu) < +\infty$, hence for every $f \in C(X)$ we have

$$\int_X T^-(tf) d\mu \geqslant t \int_X f\ d\nu - \mathcal{T}(\mu, \nu),$$

hence by dividing by t and letting $t \to \infty$, we get from the monotone convergence theorem that $\int_X T_r f d\mu \geqslant \int_X f d\nu$ and hence $\mathcal{T}_r(\mu, \nu) \leqslant 0 = \mathcal{T}_f(\mu, \nu)$.

Remark 3.11. Note that the above shows that for a general backward linear transfer $\mathcal{T}$ with Kantorovich operator T^- and Recession operator T_r^-, we have

$$\mathcal{T}(\mu, \nu) < +\infty \quad \text{if and only if} \quad \int_X T_r f d\mu \geqslant \int_X f d\nu \text{ for every } f \in C(Y). \tag{104}$$

The latter condition can be seen as a *generalized order condition* between μ and ν that extends the notion of convex order. Indeed, if $\mathcal{T}$ is the balayage transfer, then $T^- f = T_r^- f = \hat{f}$, which is the concave envelope of f, and the condition does coincide with the convex order between measures.

4 Extension of Kantorovich operators

In order to study the ergodic properties of a Kantorovich operator $T : C(X) \to USC(X)$, one needs to iterate it and therefore it is necessary to extend it to an operator $T : USC(X) \to USC(X)$ and eventually to $T : USC_\sigma(X) \to USC_\sigma(X)$ with the same properties (a), (b), (c) of a Kantorovich operator.

In order to define such an extension, we assume that T is proper so that we can associate a linear transfer $\mathcal{T}$ on $\mathcal{P}(X) \times \mathcal{P}(Y)$ in such a way that

$$Tf(x) = \sup\{\int_Y f d\nu - \mathcal{T}(\delta_x, \nu);\, \nu \in \mathcal{P}(Y)\} \quad \text{for every } f \in C(Y). \tag{105}$$

We shall then extend T in such a way that (105) holds for every $f \in USC_\sigma(Y)$. Properties (a), (b) and (c) will then follow.

4.1 Extension of Kantorovich operators from $C(Y)$ to $USC_\sigma(Y)$

Theorem 4.1. *Let $\mathcal{T}$ be a backward linear transfer such that $\{\delta_x; x \in X\} \subset \mathcal{D}_1(\mathcal{T})$, and let $T : C(Y) \to USC_\sigma(X)$ be the associated Kantorovich operator.*

1. *For $f \in USC(Y)$, define $\hat{T} f(x) := \inf\{Tg(x)\,;\, g \in C(Y),\, g \geqslant f\}$, then*

$$\hat{T} f(x) = \sup\{\textstyle\int_Y f d\nu - \mathcal{T}(\delta_x, \nu);\, \nu \in \mathcal{P}(Y)\}, \tag{106}$$

and $\hat{T}$ maps $USC(Y)$ to $USC_\sigma(X)$.

Moreover, if $T : C(Y) \to USC(X)$, then $\hat{T}$ maps $USC(Y)$ to $USC(X)$.

2. *For $f \in USC_\sigma(Y)$, define $\hat{\underset{\smile}{T}} f := \sup\{\hat{T} g\,;\, g \in USC(Y)\,, g \leqslant f\}$, then*

$$\hat{\underset{\smile}{T}} f(x) = \sup\{\int_Y f d\nu - \mathcal{T}(\delta_x, \nu);\, \nu \in \mathcal{P}(Y)\}, \tag{107}$$

and $\hat{\underset{\smile}{T}}$ maps $USC_\sigma(Y)$ to $USC_\sigma(X)$.

Proof. 1) It is clear that for any $g \in C(Y)$, $g \geqslant f$,

$$\sup\{\int_Y f \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)\,;\, \sigma \in \mathcal{P}(Y)\} \leqslant \sup\{\int_Y g \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)\,;\, \sigma \in \mathcal{P}(Y)\} = Tg(x).$$

Therefore $\sup\{\int_Y f \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)\,;\, \sigma \in \mathcal{P}(Y)\} \leqslant \inf\{Tg(x)\,;\, g \in C(Y), g \geqslant f\} = \hat{T} f(x)$.

On the other hand, let $g_n \searrow f$ be a decreasing sequence of continuous functions. Then,

$$\hat{T} f(x) \leqslant Tg_n(x) = \sup\{\int_Y g_n \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)\,;\, \sigma \in \mathcal{P}(Y)\} = \int_Y g_n \,\mathrm{d}\sigma_n - \mathcal{T}(\delta_x, \sigma_n),$$

for some probability measure σ_n. Consider an increasing subsequence n_k so that $\sigma_{n_k} \to \bar{\sigma}$. Then for any $j \leqslant k$, $\hat{T} f(x) \leqslant \int_Y g_{n_j} \,\mathrm{d}\sigma_{n_k} - \mathcal{T}(\delta_x, \sigma_{n_k})$ where we have used the fact that $g_{n_k} \leqslant g_{n_j}$ whenever $j \leqslant k$. For this fixed j, we have that $g_{n_j} \in C(Y)$ and so $\int g_{n_j} \,\mathrm{d}\sigma_{n_k} \to \int g_{n_j} \,\mathrm{d}\bar{\sigma}$ as $k \to \infty$. Hence we obtain

$$\hat{T} f(x) \leqslant \lim_{k\to\infty} \int_Y g_{n_j} \,\mathrm{d}\sigma_{n_k} - \liminf_{k\to\infty} \mathcal{T}(\delta_x, \sigma_{n_k}) \leqslant \int_Y g_{n_j} \,\mathrm{d}\bar{\sigma} - \mathcal{T}(\delta_x, \bar{\sigma}).$$

Finally we take a limit as $j \to \infty$ to obtain $\hat{T} f(x) \leqslant \sup\{\int_Y f \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)\,;\, \sigma \in \mathcal{P}(Y)\}$. It follows that $\hat{T} f$ satisfies (106) and therefore $\hat{T} f \in USC_\sigma$. Note that $\hat{T} f$ is bounded above since $\hat{T} f(x) \leqslant \sup_{y\in Y} f(y) - m_{\mathcal{T}}$, where $m_{\mathcal{T}}$ is a lower bounded for $\mathcal{T}$.

If now $T : C(Y) \to USC(X)$, then $\hat{T}$ is in $USC(X)$ by its definition.
2) For $f \in USC_\sigma(Y)$, we use the first part to write for any $g \in USC(Y)$, $g \leqslant f$,

$$\sup\{\int f \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)\,;\, \sigma \in \mathcal{P}(Y)\} \geqslant \tilde{T} g(x)$$

and so it is greater than $\hat{\underset{\smile}{T}} f(x)$. On the other hand, for an increasing $g_n \nearrow f$, $g_n \in USC(Y)$,

$$\hat{\underset{\smile}{T}} f(x) \geqslant \hat{T} g_n(x) \geqslant \int g_n \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma), \quad \text{for any } \sigma \in \mathcal{P}(Y).$$

By the monotone convergence of g_n to f, we may take the limit as $n \to \infty$ in the above inequality, and conclude

$$\hat{\underset{\smile}{T}} f(x) \geqslant \int f \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma) \quad \text{for any } \sigma \in \mathcal{P}(Y),$$

whereby taking the supremum in σ yields $\hat{\underset{\smile}{T}} f(x) \geqslant \sup\{\int f \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)\,;\, \sigma \in \mathcal{P}(Y)\}$, and we are done showing that $\hat{\underset{\smile}{T}}$ maps $USC_\sigma(Y)$ to $USC_\sigma(X)$, while satisfying (107).

The following continuity properties of T along monotone sequences of $USC(Y)$ and $USC_\sigma(Y)$ will be crucial for Sections 8 and 9.

Lemma 4.2. *Let $\mathcal{T}$ be a backward linear transfer as above, and let T denote its corresponding Kantorovich operator, extended to $USC_\sigma(Y)$.*

1. *If $f_n \in USC(Y)$, $f \in USC_\sigma(Y)$ with $f_n \searrow f$, then $\lim_{n\to\infty} T f_n = T f$.*
2. *If $f_n \in USC_\sigma(Y)$, $f \in USC_\sigma(Y)$ with $f_n \nearrow f$, then $\lim_{n\to\infty} T f_n = T f$.*

Proof. 1) By monotonicity, $Tf \leqslant \liminf_n T f_n$. On the other hand let σ_n achieve the supremum in the definition of $T f_n(x)$, i.e.,

$$T f_n(x) = \int f_n \,\mathrm{d}\sigma_n - \mathcal{T}(\delta_x, \sigma_n).$$

Extract an increasing subsequence n_k so that $\limsup_n T f_n(x) = \lim_k T f_{n_k}(x)$ and $\sigma_{n_k} \to \bar{\sigma}$. Then as before, we have $T f_{n_k}(x) \leqslant \int f_{n_j} \,\mathrm{d}\sigma_{n_k} - \mathcal{T}(\delta_x, \sigma_{n_k})$ for fixed $j \leqslant k$. As $f_{n_j} \in USC(Y)$, it follows that $\limsup_n T f_n(x) \leqslant \int f_{n_j} \,\mathrm{d}\bar{\sigma} - \mathcal{T}(\delta_x, \bar{\sigma})$. Then we let $j \to \infty$ and use monotone convergence.

2) Again, by monotoncity, $Tf \geqslant \limsup_n T f_n(x)$. On the other hand, $T f_n(x) \geqslant \int f_n \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)$ for all σ. Hence by monotone convergence, $\liminf_n T f_n(x) \geqslant \int f \,\mathrm{d}\sigma - \mathcal{T}(\delta_x, \sigma)$ for all σ. Taking the supremum over σ yields $\liminf_n T f_n(x) \geqslant T f(x)$.

Remark 4.3. *We note that In general, the operator T cannot be extended to the class $C_{\delta\sigma\delta}(Y) = USC_{\sigma\delta}(Y)$, and the continuity property of T in item (1) of the above lemma cannot be extended to sequence $f_n \in USC_\sigma(Y)$.*

Corollary 4.4. *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a backward linear transfer such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$. Then,*

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

$$\begin{aligned}\mathcal{T}(\mu,\nu) &= \sup\Big\{\int_Y g(y)\,d\nu(y) - \int_X T^- g(x)\,d\mu(x);\ g\in LSC(Y)\Big\}\\ &= \sup\Big\{\int_Y g(y)\,d\nu(y) - \int_X T^- g(x)\,d\mu(x);\ g\in USC(Y)\Big\}\\ &= \sup\Big\{\int_Y g(y)\,d\nu(y) - \int_X T^- g(x)\,d\mu(x);\ g\in USC_\sigma(Y)\Big\}.\end{aligned}$$

2. *The Legendre transform formula (82) for* $\mathcal{T}_\mu$*, which holds for continuous functions on* Y*, also holds for* $g\in USC_\sigma(Y)$*, that is*

$$\mathcal{T}_\mu^*(g) := \sup\{\textstyle\int_Y g\,\mathrm{d}\sigma - \mathcal{T}(\mu.\sigma);\sigma\in\mathcal{P}(Y)\} = \int_X T^- g\,d\mu \quad \textit{for all } g\in USC_\sigma(Y). \tag{108}$$

Proof. It is clear that $\mathcal{T}(\mu,\nu)$ is smaller than all the expressions on its right. It is also clear that it suffices to show that

$$\mathcal{T}(\mu,\nu) \geqslant \sup\Big\{\int_Y g(y)\,d\nu(y) - \int_X T^- g(x)\,d\mu(x);\ g\in USC_\sigma(Y)\Big\}.$$

For that recall from Section 2 that for any $g\in USC_\sigma(Y)$ we have for every $x\in X$,

$$T^- g(x) = \sup\{\int_Y g(y)\,d\sigma(y) - \mathcal{T}(x,\sigma);\sigma\in\mathcal{P}(Y)\}. \tag{109}$$

Take now any $\pi\in\mathcal{K}(\mu,\nu)$ and its disintegration $(\pi_x)_x$ in such a way that $\nu(A)=\int_X \pi_x(A)\,\mathrm{d}\mu(x)$, then

$$T^- g(x) \geqslant \int_Y g(y)\,d\pi_x(y) - \mathcal{T}(x,\pi_x),$$

hence,

$$\begin{aligned}\int_Y g(y)\,d\nu(y) - \int_X T^- g(x)\,d\mu(x) &\leqslant \int_Y g(y)\,d\nu(y) - \int_X\int_Y g(y)\,d\pi_x(y)\,d\mu(x) + \int_X \mathcal{T}(x,\pi_x)\,d\mu(x)\\ &\leqslant \int_X \mathcal{T}(x,\pi_x)\,d\mu.\end{aligned}$$

It follows from Theorem 3.1 that

$$\int_Y g(y)\,d\nu(y) - \int_X T^- g(x)\,d\mu(x) \leqslant \mathcal{T}(\mu,\nu), \tag{110}$$

and (1) is done.

For (2) note first that (110) yields that $\int_X T^- g(x)\,d\mu(x) \geqslant \mathcal{T}_\mu^*(g)$. On the other hand, assume $g\in USC(Y)$ and use (109) to find a measurable selection $x\to\bar\pi_x$ from X to $\mathcal{P}(Y)$ such that

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

and let $\sigma(A) = \int_X \bar{\pi}_x(A)\, d\mu(x)$, then

$$\begin{aligned}\int_X T^- g(x)\, d\mu(x) &= \int_X \int_Y g(y) d\bar{\pi}_x(y)\, d\mu(x) - \int_Z c(x, \pi_x)\, d\mu(x) \\ &\leqslant \int_Y g(y)\, d\sigma(y) - \mathcal{T}(\mu, \sigma),\end{aligned}$$

hence $\int_X T^- g(x)\, d\mu(x) \leqslant \mathcal{T}_\mu^*(g)$. Note now that (108) carries through increasing limits, hence it also holds for $g \in USC_\sigma(Y)$.

4.2 Conjugate functions for bi-directional transfers

Suppose now that $\mathcal{T}$ is both a backward and forward transfer with Kantorovich operators T^- and T^+. We have the following notion motivated by the theory of mass transport.

Definition 4.5. *Say that a pair $(f_1, f_2) \in USC(Y) \times LSC(X)$ are conjugate if:*

$$T^- f_1 = f_2 \quad \textit{and} \quad T^+ f_2 = f_1. \tag{111}$$

The following proposition shows in particular that for any function $g \in C(Y)$, the couple $(T^- g, T^+ \circ T^- g)$ form a conjugate pair.

Proposition 4.6. *Suppose $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ is both a forward and backward linear transfer, and that $\{(\delta_x, \delta_y); (x, y) \in X \times Y\} \subset D(\mathcal{T})$. Assume that $T^- : C(Y) \to USC(X)$ and that $T^+ : C(X) \to LSC(Y)$, then for any $g \in C(Y)$ (resp.,$f \in C(X)$)*

$$T^+ \circ T^- g(y) \geqslant g(y) \textit{ for } y \in Y, \qquad T^- \circ T^+ f(x) \leqslant f(x) \textit{ for } x \in X, \tag{112}$$

and

$$T^- \circ T^+ \circ T^- g = T^- g \quad \textit{and} \quad T^+ \circ T^- \circ T^+ f = T^+ f. \tag{113}$$

In particular,

$$\mathcal{T}(\mu, \nu) = \sup\Big\{ \int_Y T^+ \circ T^- g(y)\, d\nu(y) - \int_X T^- g\, d\mu(x);\ g \in C(Y) \Big\} \tag{114}$$

$$= \sup\Big\{ \int_Y T^+ f(y)\, d\nu(y) - \int_X T^- \circ T^+ f\, d\mu(x);\ f \in C(X) \Big\}. \tag{115}$$

Proof: Note that $USC(X) \subset LSC_\delta(X)$, hence for $\nu \in \mathcal{P}(Y)$,

$$\begin{aligned}\int_Y T^+ \circ T^- g\, d\nu &= -\mathcal{T}_\nu^*(-T^- g) \\ &= -\sup\{-\int_X T_{\delta_x}^*(g)\, d\mu(x) - \mathcal{T}(\mu, \nu); \mu \in \mathcal{P}(X)\} \\ &= \inf\{\int_X T_{\delta_x}^*(g)\, d\mu(x) + \mathcal{T}(\mu, \nu); \mu \in \mathcal{P}(X)\} \\ &\geqslant \inf\{\int_X T_{\delta_x}^*(g)\, d\mu(x) + \int_Y g\, d\nu - \int_X T_{\delta_x}^*(g)\, d\mu; \mu \in \mathcal{P}(X)\} \\ &= \int_Y g\, d\nu.\end{aligned}$$

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

5 Linear transfers which are not mass transports

We now give examples of linear transfers, which do not fit in the framework of Monge-Kantorovich theory.

5.1 Linear transfers associated to weak mass transports

Weak mass transportations also arise from the work of Marton, who extended the work of Talagrand. The paper of Gozlan et al. [40] exhibit many examples of which we single out the following.

Example 4.2: Marton transports are backward linear transfers (Marton [47, 48])

These are transports of the following type:

$$\mathcal{T}_{\gamma,d}(\mu,\nu) = \inf\left\{\int_X \gamma\left(\int_Y d(x,y)d\pi_x(y)\right)\, d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\right\}, \tag{116}$$

where γ is a convex function on $\mathbb{R}^+$ and $d : X \times Y \to \mathbb{R}$ 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

$$\mathcal{T}_2(\mu,\nu) = \inf\left\{\mathbb{E}[\mathbb{E}[|X-Y|\,|Y]^2]; X \sim \mu, Y \sim \nu\right\}. \tag{117}$$

This is a backward linear transfer with Kantorovich potential

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

We now give applications to transfers that are mostly dependent on the barycenter of the measures involved.

Proposition 5.1. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(Y)$, where Y is convex compact such that for some lower semi-continuous functional $c : X \times Y \to \mathbb{R}$, we have*

$$\mathcal{T}(x,\sigma) = c(x, \int_Y y\, d\sigma(y)) \quad \text{for all } x \in X \text{ and } \sigma \in \mathcal{P}(Y),$$

where $\int_Y y\, d\sigma(y)$ denotes the barycenter of σ. Then, for every $f \in C(Y)$,

$$T^- f(x) = \sup\{\hat{f}(y) - c(x,y); y \in Y\},$$

where $\hat{f}$ is the concave envelope of f, i.e., the smallest concave usc function above f.

Proof: Note that z is the barycenter of a probability measure σ if and only if $\delta_z \prec_{\mathcal{C}} \sigma$ where $\mathcal{C}$ is the cone of convex functions. Write now

$$\begin{aligned} T^- f(x) &= \sup\{\int_Y f\, d\sigma - c(x, \int_Y y\, d\sigma(y)); \sigma \in \mathcal{P}(Y)\} \\ &= \sup_{z\in Y} \sup\{\int f\, d\sigma - c(x,y); \sigma \in \mathcal{P}(Y), \delta_y \prec \sigma\} \\ &= \sup_{z\in Y}\{\hat{f}(z) - c(x,z)\}. \end{aligned}$$

Example 4.3: A barycentric cost function (Gozlan et al. [40])

Consider the (weak) transport

$$\mathcal{T}(\mu,\nu) = \inf\left\{\int_X \|x - \int_Y y d\pi_x(y)\| \, d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\right\}. \tag{118}$$

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

$$T^- f(x) = \sup\{\hat{f}(y) - \|y - x\|; y \in \mathbb{R}^n\},$$

where $\hat{f}$ is the concave envelope of f.

Note that the same holds if one uses other cones for balayage, such as the cone of subharmonic or plurisubharmonic functions.

Example 4.4: The variance functional

If the transfer is given by the variance functional

$$\mathcal{T}(\mu,\nu) := I(\nu) = -\mathrm{var}(\nu) := |\int_Y y \, d\nu|^2 - \int_Y |y|^2 \, d\nu(y),$$

then, by letting q be the quadratic function $q(y) = |y|^2$, we have

$$\begin{aligned} T^- f(x) =& \sup\{\int_Y f \, d\sigma - |\int_Y y \, d\sigma|^2 + \int_Y |y|^2 \, d\sigma(y); \sigma \in \mathcal{P}(Y)\} \\ =& \sup\{\int_Y (f+q) \, d\sigma - |\int_Y y \, d\sigma|^2; \sigma \in \mathcal{P}(Y)\} \\ =& S^-(f+q)(x), \end{aligned}$$

where S^- is the Kantorovich operator associated to the transfer $\mathcal{S}(\mu,\nu) := |\int_Y y \, d\sigma|^2$, which only depends on the barycenter and therefore $S^- g = \sup\{\hat{g}(z) - |z|^2; z \in Y\}$. It follows that

$$T^- f = \sup\{\widehat{f+q}(z) - |z|^2; z \in Y\}.$$

Cost minimizing mass transport with additional constraints give examples of one-directional linear transfers. We single out the following:

Example 4.5: Martingale transports are backward linear transfers

Martingale transports are $\mathcal{C}$-dilations where $\mathcal{C}$ is the cone of convex continuous functions on $\mathbb{R}^n$. If $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a continuous cost function, then define the weak cost as

$$\tilde{c}(x,\sigma) = \begin{cases} \int_{\mathbb{R}^d} c(x,y) \, d\sigma(y) & \text{if } \delta_x \prec_C \sigma, \\ +\infty & \text{if not.} \end{cases} \tag{119}$$

The corresponding martingale transport is then

$$\mathcal{T}_M(\mu,\nu) = \inf\{\int_{\mathbb{R}^d \times \mathbb{R}^d} \tilde{c}(x,\pi_x) \, d\,\mu; \pi \in \mathcal{K}(\mu,\nu)\}.$$

Equivalently, if μ,ν are two probability measures we then 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

π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, such that for μ-almost $x \in \mathbb{R}^d$, the component π_x of its disintegration $(\pi_x)_x$ with respect to μ, i.e. $d\pi(x, y) = d\pi_x(y)d\mu(x)$, has its barycenter at x. As mentioned above,

$$MT(\mu, \nu) \neq \emptyset \text{ if and only if } \mu \prec_C \nu. \tag{120}$$

One can also use the probabilistic notation, which amounts to minimize $\mathbb{E}_P\, c(X, Y)$ over all martingales (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 π can be written as the conditional probability $\pi_x(A) = \mathbb{P}Y \in A|X = x$.

The martingale transport can be written as

$$\mathcal{T}_M(\mu, \nu) = \begin{cases} \inf\{\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi(x, y); \pi \in MT(\mu, \nu)\} & \text{if } \mu \prec_C \nu \\ +\infty & \text{if not.} \end{cases} \tag{121}$$

This s a backward linear transfer with a backward Kantorovich operator given by

$T_M^- f(x) = \hat{f}_{c,x}(x)$, where $\hat{f}_{c,x}$ is the concave envelope of the function $f_{c,x} : y \to f(y) - c(x, y)$.

See Henri-Labordère [42] and Ghoussoub-Kim-Lim [31] for higher dimensions.

Example 4.6: Schrödinger bridge (Gentil-Leonard-Ripani [29])

Let M be a compact Riemannian manifold and fix some reference non-negative measure R on path space $\Omega = C([0, 1], M)$. Let $(X_t)_t$ be a random process on M whose law is R, and denote by R_{01} the joint law of the initial position X_0 and the final position X_1, that is $R_{01} = (X_0, X_1)_\# R$. For example (see [29]), 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.

For probability measures μ and ν on M, define

$$\mathcal{T}_{R_{01}}(\mu, \nu) := \inf\{\int_M \mathcal{H}(r_1^x, \pi_x) d\mu(x)\,;\, \pi \in \mathcal{K}(\mu, \nu),\, d\pi(x, y) = d\mu(x)d\pi_x(y)\} \tag{122}$$

where $dR_{01}(x, y) = dm(x)dr_1^x(y)$ is the disintegration of R_{01} with respect to its initial measure m. By Theorem **??**, $\mathcal{T}_{R_{01}}$ is a backward linear transfer (corresponding to the weak cost $c(x, p) = \mathcal{H}(r_1^x, p)$). Its Kantorovich operator is given by

$$T^- f(x) = \log E_{R^x} e^{f(X_1)} = \log S_1(e^f)(x),$$

where (S_t) is the semi-group associated to R.

The transfer (122) is associated to the maximum entropy formulation of the Schrödinger bridge problem in the following way: Define the entropic transportation cost between μ and ν via the formula

$$\mathcal{S}_R(\mu, \nu) = \inf\{\int_{M \times M} \log(\frac{d\pi}{dR_{01}})\, d\pi; \pi \in \mathcal{K}(\mu, \nu)\}. \tag{123}$$

Then, under appropriate conditions on V (e.g., if V is uniformly convex), then

$$\mathcal{T}_{R_{01}}(\mu, \nu) = \mathcal{S}_R(\mu, \nu) - \int_M \log(\frac{d\mu}{dm})\, d\mu.$$

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.

5.2 One-sided transfers associated to stochastic mass transport

Let M be a manifold (compact manifold or $\mathbb{R}^n$) and consider a Lagrangian on phase space $L : TM \to [0, \infty)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with normal filtration $\{\mathcal{F}_t\}_{t \geqslant 0}$, and define $\mathcal{A}_{[0,t]}$ to be the set of continuous semi-martingales $X : \Omega \times [0, t] \to M$ such that there exists a Borel measurable drift $\beta : [0, t] \times C([0, t]) \to \mathbb{R}^d$ for which

1. $\omega \mapsto \beta(s, \omega)$ is $\mathcal{B}(C([0, s]))_+$-measurable for all $s \in [0, t]$, where $\mathcal{B}(C([0, s]))$ is the Borel σ-algbera of $C[0, s]$.
2. $W(s) := X(s) - X(0) - \int_0^s \beta(s')\,\mathrm{d}s'$ is a $\sigma(X(s)\,;\, 0 \leqslant s \leqslant t)$ M-valued Brownian motion.

For each β, we shall denote the corresponding X by X^β in such a way that

$$dX^\beta(t) = \beta(t)dt + dW(t). \tag{124}$$

Example 4.7: Stochastic mass transport between two probability measures
Consider the following functional $\mathcal{T} : \mathcal{P}(M) \times \mathcal{P}(M) \to \mathbb{R} \cup \{+\infty\}$ defined for any pair of probability measures μ_0 and μ_1 on M via the formula:

$$\mathcal{T}(\mu_0, \mu_1) := \inf \left\{ \mathbb{E} \int_0^1 L(X^\beta(s), \beta(s))\,\mathrm{d}s\,;\, X(0) \sim \mu_0, X(1) \sim \mu_1, X \in \mathcal{A}_{[0,1]} \right\}, \tag{125}$$

This stochastic transport does not fit in the standard optimal mass transport theory since it does not originate in the optimization according to a cost between two deterministic states. However, it still enjoy a dual formulation (first proven in Mikami-Thieullin [52] for the space $\mathbb{R}^d$) that permits it to be realised as a backward linear transfer. In fact, by introducing the operator $T_t : C(M) \to USC(M)$ via the formula

$$T_t f(x) := \sup_{X \in \mathcal{A}_{[0,t]}} \left\{ \mathbb{E}\left[f(X(t))|X(0) = x\right] - \mathbb{E}\left[\int_0^t L(X(s), \beta_X(s, X))\,\mathrm{d}s | X(0) = x\right]\right\}, \tag{126}$$

then the duality relation between $\mathcal{T}$ and T_t can be readily detailed. Indeed, an adaptation of the proofs of Mikami-Thieullin [52] yields the following.

Proposition 5.2. *Under suitable conditions on L (for example if $L(x, \beta) = \frac{1}{2}|\beta|^2$), the following assertions hold:*

1. *$\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(M) \times \mathcal{P}(M)$ with Kantorovich operator T_1. is the unique viscosity solution of*

$$\frac{\partial u}{\partial t}(t, x) + \frac{1}{2}\Delta_x u(t, x) + H(x, \nabla_x u(t, x)) = 0, \quad (t, x) \in [0, 1) \times M, \tag{127}$$

with $u(1, x) = f(x)$.

2. *In particular, for any pair of probability measures μ_0 and μ_1 on M, we have*

$$\mathcal{T}(\mu_0, \mu_1) = \sup\{\int_M u(1,x)d\mu_1(x) - \int_M u(0,x)d\mu_0(x); u(t,x) \text{ solution of (127)}\}. \tag{128}$$

Example 4.8: Stochastic mass transport with fixed distribution at all time

Suppose now $\mu_0 \in \mathcal{P}(M)$ and $\mu \in \mathcal{P}([0,1] \times M)$. If the latter has Lebesgue measure as a first marginal, then we can disintegrate it and write it as $d\mu = d\mu_t\, dt$, where μ_t is a probability measure on M. Consider the following functional on $\mathcal{P}(M) \times \mathcal{P}([0,1] \times M)$,

$$\begin{aligned}\mathcal{T}(\mu_0, \nu) &:= \mathcal{T}(\mu_0, (\mu_t)_{t>0}) \\ &:= \inf\left\{\mathbb{E}\int_0^1 L(X^\beta(s), \beta(s))\,\mathrm{d}s\,;\, X \in \mathcal{A}_{[0,1]}, X(t) \sim \mu_t\ \forall t \in [0,1]\right\}, \end{aligned} \tag{129}$$

if the first marginal of μ is Lebesgue measure and $+\infty$ otherwise.

Proposition 5.3. *Under suitable conditions on L (for example if $L(x,\beta) = \frac{1}{2}|\beta|^2$), the following assertions hold:*

1. *$\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(M) \times \mathcal{P}([0,1] \times M)$ with corresponding Kantorovich operator $T : C([0,1] \times M) \to C(M)$ defined for any $f \in C([0,1] \times M)$ as $u_f(0,x)$, where u_f is a bounded continuous viscosity solution of the following Hamilton-Jacobi equation,*

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\Delta_x u(t,x) + H(x, \nabla_x u(t,x)) + f(t,x) = 0, \quad (t,x) \in [0,1) \times M, \tag{130}$$

with $u_f(1,x) = 0$.

2. *In particular, for any probability measures $\mu_0 \in \mathcal{P}(M)$ and $\nu \in \mathcal{P}([0,1] \times M)$, we have*

$$\mathcal{T}(\mu_0, \mu) = \sup\{\int_0^1 \int_M f(t,x)d\mu_t(x)dt - \int_M u_f(0,x)d\mu_0; u_f \text{ solves (130)}\}. \tag{131}$$

Example 4.9: The Arnold-Brenier variational principle for the incompressible Euler equation

In [16, 17, 18], Brenier proposed several relaxed versions of the Arnold geodesic formulation of the incompressible Euler equation. We describe the following model which, strictly speaking is not stochastic, yet we include it in this section for comparison purposes.

For a smooth domain D in $\mathbb{R}^d$ consider the space

$$H_t^1(\mathbb{R}^d) = \{\xi \in L^2([0,1], \mathbb{R}^d) \text{ such that } \dot{\xi} \in L^2([0,1], \mathbb{R}^d)\}$$

and denote by $H_t^1(D)$ the subset of $H_t^1(\mathbb{R}^d)$ consisting of those paths valued in D.
For any $(s,t) \in [0,1]^2$, we consider the projections $\pi_{s,t} : C([0,1]; D) \to D \times D$ (resp., $\pi_t : C([0,1]; D) \to D$) defined by $\pi_{s,t} f = (f(s), f(t))$ (resp., $\pi_t f = f(t)$).

For $\mu \in \mathcal{P}(C([0,1]; D))$, we denote by $\mu_{s,t} := (\pi_{s,t})_{\#}\mu$ $\mu_t := (\pi_t)_{\#}\mu$. Similarly, for $\nu \in \mathcal{P}(D \times D)$, we denote by ν_0 and ν_1 its first (resp., second) marginal on D.

If λ is the normalized Lebesgue measure on D, we consider the functional

$$\mathcal{T}(\mu, \nu) = \begin{cases} \inf\{\mathbb{E}_\mu \int_0^1 \frac{1}{2}|\dot{\xi}|^2 dt; & \text{if } \mu_t = \lambda,\ \forall t \in [0,1] \text{ and } \mu_{0,1} = \nu \\ +\infty & \text{otherwise.} \end{cases}$$

Proposition 5.4. *The following assertions hold:*

1. $\mathcal{T}$ *is a backward linear transfer on* $\mathcal{P}(C([0,1];D)) \times \mathcal{P}(D \times D)$ *with corresponding Kantorovich operator* $T : C([0,1] \times D) \to C(D \times D)$ *defined for any* $f \in C([0,1] \times D)$ *as*

$$Tf(x,y) = \inf\{\int_0^1 [\frac{1}{2}|\dot{\xi}|^2 - f(t,\xi_t)]dt; \xi \in H_t^1(D),\, \xi(0) = x, \xi(1) = y\}. \tag{132}$$

2. $T_f(x,y) = u_f(0,x,y)$*, where* u_f *is a bounded continuous viscosity solution of the Hamilton-Jacobi equation,*

$$\frac{\partial u}{\partial t}(t,x,y) + \frac{1}{2}|\nabla u(t,x,y)|^2 + f(t,x) = 0, \quad (t,x,y) \in [0,1) \times D \times D, \tag{133}$$

with $u_f(1,x,y) = 0$.

5.3 Transfers associated to optimally stopped stochastic transports

In dimension $d \geqslant 2$, there are many different types of martingales. If one chooses those that essentially follow a Brownian path, then we have the following linear transfers.

Example 4.10: Optimal subharmonic Martingale transfers (Ghoussoub-Kim-Palmer [24])

Confining the problem to a convex bounded domain O in $\mathbb{R}^d$, then if (μ,ν) are in *subharmonic order,* i.e. $\mu \prec_{SH} \nu$, where SH is the cone of subharmonic functions on O, we set,

$$\mathcal{P}_c(\mu,\nu) = \inf_{\pi \in \mathcal{BM}(\mu,\nu)} \int_{O\times O} c(x,y)\pi(dx,dy), \tag{134}$$

where each $\pi \in \mathcal{BM}(\mu,\nu)$ is a probability measure on $O \times O$ with marginals μ and ν, satisfying $\delta_x \prec_{SH} \pi_x$ for $\mu-$a.e. x, where π_x is the disintegration of $\pi(dx,dy) = \pi_x(dy)\mu(dx)$. Otherwise, set $\mathcal{P}_c(\mu,\nu) = +\infty$.

By a remarkable theorem of Skorokhod [59], such transport plans π can be seen as joint distributions of $(B_0, B_\tau) \sim \pi$, where $B_0 \sim \mu$, $B_\tau \sim \nu$ and τ is a possibly randomized stopping time for the Brownian filtration. See for example [32]. The above problem associated to a cost c can then be formulated as

$$\mathcal{P}_c(\mu,\nu) = \inf_\tau \Big\{\mathbb{E}\big[c(B_0,B_\tau)\big];\ B_0 \sim \mu \quad \& \quad B_\tau \sim \nu\Big\}, \tag{135}$$

where $(B_t)_t$ is Brownian motion starting with distribution μ and ending at a stopping time τ such that B_τ realizes the distribution ν.
In [24] it is shown that $\mathcal{P}_c$ is a backward linear transfer with a backward Kantorovich map given by $T^- f(x) = J_f(x,x)$, where

$$J_f(x,y) = \sup_{\tau \leqslant \tau_O} \mathbb{E}\big[\psi(B_\tau^y) - c(x, B_\tau^y)\big], \tag{136}$$

and τ_O is the first exit time of the set O. Under some regularity assumptions on f and c, and for each fixed $x \in \overline{O}$, the function $y \mapsto J_f(x,y)$ is the unique viscosity solution to the obstacle problem for $u \in C(\overline{O})$:

$$\begin{cases} u(y) \geqslant f(y) - c(x,y), \text{ for } y \in O, \\ u(y) = f(y) - c(x,y) \text{ for } y \in \partial O, \\ \Delta u(y) \leqslant 0 \text{ for } y \in O, \\ \Delta u(y) = 0 \text{ whenever } u(y) > f(y) - c(x,y), \end{cases}$$

as well as the unique minimizer of the variational problem

$$\inf\Big\{ \int_O |\nabla u|^2 dy;\ u \geqslant f - c(x,\cdot),\, u \in H^1(O)\Big\}.$$

Example 4.11: Optimally stopped stochastic transport [33, 24]

Given a Lagrangian $L : [0,1] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, consider the optimal stopping problem

$$\mathcal{T}_L(\mu,\nu) = \inf\left\{ \mathbb{E}\left[\int_0^\tau L(t, X(t), \beta_X(t, X(t)))\, dt\right]; X(0) \sim \mu, \tau \in \mathcal{S}, X_\tau \sim \nu, X(\cdot) \in \mathcal{A}\right\}, \tag{137}$$

where $\mathcal{S}$ is the set of possibly randomized stopping times, and $\mathcal{A}$ is the class of processes defined in Section 4.3. In this case, $\mathcal{T}_L$ is a backward linear transfer with Kantorovich potential given by $T_L^- f = \hat{V}_f(0,\cdot)$, where

$$\hat{V}_f(t,x) = \sup_{X \in \mathcal{A}} \sup_{T \in \mathcal{S}} \left\{ \mathbb{E}\left[f(X(T)) - \int_t^T L(s, X(s), \beta_X(s,X))\, ds \Big| X(t) = x\right]\right\}, \tag{138}$$

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

$$\min\left\{\ V_f(t,x) - f(x), -\partial_t V_f(t,x) - H\big(t, x, \nabla V_f(t,x)\big) - \tfrac{1}{2}\Delta V_f(t,x)\ \right\} = 0. \tag{139}$$

In Section 9, we shall deal in detail with optimal stochastic transports as a semi-group of backward linear transfers in conjunction with a stochastic Mather theory.

6 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 $X \times Y$.

Proposition 6.1. *The class $\mathcal{LT}_-(X \times Y)$ (resp., $\mathcal{LT}_+(X \times Y)$) is a convex subcone in the cone 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,*

$$T_a^-(f) = aT^-(\frac{f}{a}). \tag{140}$$

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 the sum defined as*

$$\mathcal{T}_1 \oplus \mathcal{T}_2)(\mu,\nu) := \inf\{\int_X \left\{\mathcal{T}_1(x,\pi_x) + \mathcal{T}_2(x,\pi_x)\right\} d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\} \tag{141}$$

is a backward linear transfer on $X \times Y$, with Kantorovich operator given on $C(Y)$ by

$$\begin{aligned} T^- f(x) &= \sup\{\int_Y f\, d\sigma - \mathcal{T}_1(x,\sigma) - \mathcal{T}_2(x,\sigma); \sigma \in \mathcal{P}(Y)\} \\ &= \inf\{T_1^- g(x) + T_2^-(f-g)(x); g \in C(Y)\}. \end{aligned}$$

6.1 Convolution and tensor products of transfers

Definition 6.2. *Consider the following operations on transfers.*

1. **(Dual Sum)** *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 $\mathcal{T}_1 \square \mathcal{T}_2$ is defined as the transfer whose Kantorovich operator is $T_1 + T_2$, that is*

$$\mathcal{T}_1 \square \mathcal{T}_2(\mu,\nu) = \sup\{\int_Y f d\nu - \int_X (T_1 f + T_2 f) d\mu; f \in C(Y)\} \tag{142}$$

2. *(**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*

$$\mathcal{T}(\mu,\nu) := \mathcal{T}_1 \star \mathcal{T}_2 = \inf\{\mathcal{T}_1(\mu,\sigma) + \mathcal{T}_2(\sigma,\nu);\ \sigma \in \mathcal{P}(X_2)\}. \tag{143}$$

3. *(**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:*

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

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

The following easy proposition is important to what follows.

Proposition 6.3. *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 $\mathcal{T}_1 \star \mathcal{T}_2$ 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*

$$T^- g(x_1, x_2) = \sup\{\int_{Y_1 \times Y_2} f(y_1, y_2) d\sigma(y_1, y_2) - \mathcal{T}_1(x_1, \sigma_1) - \mathcal{T}_2(x_2, \sigma_2);\ \sigma \in \mathcal{K}(\sigma_1, \sigma_2)\}. \tag{144}$$

Moreover,

$$\mathcal{T}_1 \otimes \mathcal{T}_2(\mu, \nu_1 \otimes \nu_2) \leqslant \mathcal{T}_1(\mu_1, \nu_1) + \int_{X_1} \mathcal{T}_2(\mu_2^{x_1}, \nu_2)\, d\mu_1(x_1), \tag{145}$$

where $d\mu(x_1, x_2) = d\mu_1(x_1) d\mu_2^{x_1}(x_2)$.

Note that a similar statement holds for forward linear transfers, modulo order reversals. For example, if $\mathcal{T}_1$ and $\mathcal{T}_2$) are forward linear transfer, then $\mathcal{T}_1 \star \mathcal{T}_2$ is a forward linear transferon $X_1 \times X_3$ with Kantorovich operator equal to $T_2^+ \circ T_1^+$.

Proof: For 1), we note first that if $\mathcal{T}_1$ (resp., $\mathcal{T}_2$) is jointly convex and weak*-lower semi-continuous on $\mathcal{P}(X_1) \times \mathcal{P}(X_2)$ (resp., $\mathcal{P}(X_2) \times \mathcal{P}(X_3)$), then both $(\mathcal{T}_1 \star \mathcal{T}_2)_\nu : \mu \to (\mathcal{T}_1 \star \mathcal{T}_2)(\mu, \nu)$ and $(\mathcal{T}_1 \star \mathcal{T}_2)_\mu : \nu \to (\mathcal{T}_1 \star \mathcal{T}_2)(\mu.\nu)$ are convex and weak*-lower semi-continuous. We now calculate their Legendre transform. For $g \in C(X_3)$,

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

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

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,

$$\mathcal{T}((x_1, x_2), \pi) = \mathcal{T}_1(x_1, \pi_1) + \mathcal{T}_2(x_1, \pi_2),$$

where π_1, π_2 are the marginals of π on Y_1 and Y_2 respectively. $\mathcal{T}_1 \otimes \mathcal{T}_2$ 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 (145).

Remark 6.4. Note that if $\mathcal{T}$ is any backward linear transfer on $X \times Y$, and $\mathcal{T}_\sigma$ is the one induced by a point transformation $\sigma : Z \to X$, then one can easily check that for $\mu \in \mathcal{P}(Z)$

and $\nu \in \mathcal{P}(Y)$, we have $\mathcal{T}_\sigma \star \mathcal{T}(\mu,\nu) = \mathcal{T}(\sigma_\# \mu, \nu)$, and its backward Kantorovich operator is given by $\tilde{T} f(z) = (T^- f)(\sigma(z))$. Similarly, if $\tau : Z \to Y$, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Z)$, then $\mathcal{T} \star \mathcal{T}_\tau(\mu,\nu) = \mathcal{T}(\mu, \tau_\# \nu)$, hence

$$\mathcal{T}_\sigma \star \mathcal{T} \star \mathcal{T}_\tau(\mu,\nu) = \mathcal{T}(\sigma_\# \mu, \tau_\# \nu).$$

6.2 Hopf-Lax formulae and projections in Wasserstein space

By an obvious induction on the convolution property enjoyed by linear transfers, one can immediately show the following.

Proposition 6.5. *Let $X_0, X_1, ..., X_n$ be $(n+1)$ compact spaces, and suppose for each $i = 1, ..., n$, we have functionals $\mathcal{T}_i$ on $\mathcal{P}(X_{i-1}) \times \mathcal{P}(X_i)$. For any probability measures μ on X_0 (resp., ν on X_n), define*

$$\mathcal{T}(\mu,\nu) = \inf\{\mathcal{T}_1(\mu,\nu_1) + \mathcal{T}_2(\nu_1,\nu_2) ... + \mathcal{T}_n(\nu_{n-1},\nu);\, \nu_i \in \mathcal{P}(X_i), i = 1, ..., n-1\}. \quad (146)$$

If each $\mathcal{T}_i$ is a linear forward (resp., backward) transfer with a corresponding Kantorovich operator $T_i^+ : C(X_i) \to C(X_{i+1})$ (resp., $T_i^- : C(X_i) \to C(X_{i-1})$), then $\mathcal{T}$ is a linear forward (resp., backward) transfer with a Kantorovich operator given by

$$T^+ = T_n^+ \circ T_{n-1}^+ \circ ... \circ T_1^+ \ \textit{(resp., } T^- = T_1^- \circ T_2^- \circ ... \circ T_n^- \textit{)}$$

In other words, the following duality formula holds:

$$\mathcal{T}_c(\mu,\nu) = \sup\big\{ \int_{X_n} T_n^+ \circ T_{n-1}^+ \circ ... T_1^+ f(y)\, d\nu(y) - \int_{X_0} f(x)\, d\mu(x);\, f \in C(X_0)\big\} \quad (147)$$

respectively,

$$\mathcal{T}_c(\mu,\nu) = \sup\big\{ \int_{X_n} g(y)\, d\nu(y) - \int_{X_0} T_1^- \circ T_2^- \circ ... \circ T_n^- g(x);\, g \in C(X_n)\big\}. \quad (148)$$

The convolution of two linear transfers associated to optimal mass transports with costs c_1 and c_2 respectively, is also a mass transport corresponding to a cost functional given by the convolution $c_1 \star c_2$. However, the above calculus allows us to convolute a mass transport with a general linear transfer, and to define a broken geodesic problems for stochastic processes.

Proposition 6.6. (Lifting convolutions to Wasserstein space) *Let $X_0, X_1, ..., X_n$ be compact spaces, and suppose for each $i = 1, ..., n$, we have a cost function $c_i : X_{i-1} \times X_i$, its corresponding optimal mass transport*

$$\mathcal{T}_{c_i}(\mu,\nu) = \inf\{\int_{X_{i-1} \times X_i} c_i(x,y)\, d\pi; \pi \in \mathcal{K}(\mu.\nu)\},$$

and its forward and backward transfers $T_{c_i}^+$ and $T_{c_i}^-$ defined in Example 3.7. Consider the following cost function on $X_0 \times X_n$, defined by

$$c(x,x') = \inf\big\{c_1(x,x_1) + c_2(x_1,x_2) .... + c_n(x_{n-1},x');\, x_1 \in X_1, x_2 \in X_2, ..., x_{n-1} \in X_{n-1}\big\}.$$

Let μ (resp., ν be probability measures on X_0 (resp., X_n), then the following holds

$$\mathcal{T}_c(\mu,\nu) = \inf\{\mathcal{T}_{c_1}(\mu,\nu_1) + \mathcal{T}_{c_2}(\nu_1,\nu_2) ... + \mathcal{T}_{c_n}(\nu_{n-1},\nu);\ \nu_i \in \mathcal{P}(X_i), i = 1, ..., n-1\}, \quad (149)$$

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

$$\begin{aligned} \mathcal{T}_c(\mu,\nu) &= \sup\{\int_{X_n} T^+_{c_n} \circ T^+_{c_{n-1}} \circ ... T^+_{c_1} f(x)\, d\nu(x) - \int_{X_0} f(y)\, d\mu(y);\ f \in C(X_0)\} \quad (150) \\ &= \sup\{\int_{X_n} g(x)\, d\nu(x) - \int_{X_0} T^-_{c_n} \circ T^-_{c_{n-1}} ... \circ T^-_{c_1} g(x);\ g \in C(X_n)\}. \quad (151) \end{aligned}$$

Proof: It suffices to verify these formulas in the case of two cost functions. We do so using duality by noting that both $\mathcal{T}_{c_1 \star c_2}$ and $\mathcal{T}_{c_1} \star \mathcal{T}_{c_2}$ have the same backward Kantorovich map equal to

$$\begin{aligned} T^-_{c_1} \circ T^-_{c_2} f(x) &= \sup_{x_1 \in X_1} \{T^-_{c_2} f(x_1) - c_1(x, x_1)\} \\ &= \sup_{x_1 \in X_1, x_2 \in X_2} \{f(x_2) - c_2(x_1, x_2) - c_1(x, x_1)\} \\ &= \sup_{x_2 \in X_2} \{f(x_2) - c(x, x_2)\} = T^-_c f(x). \end{aligned}$$

This is illustrated by the following example.

Example 5.1: The ballistic transfer (Barton-Ghoussoub [8])
Let L be a Tonelli Lagrangian, then the deterministic ballistic mass transport is defined as

$$\underline{\mathcal{B}}_d(\mu,\nu) := \inf\left\{\mathbb{E}\left[\langle V, X(0)\rangle + \int_0^T L(t, X, \dot{X}(t))\, dt\right];\ V \sim \mu,\ X \in \mathcal{A},\ X(T) \sim \nu\right\}, \quad (152)$$

where $\mathcal{A}$ is the space of random processes X_t such that $\dot{X} \in L^2[0,T], M)$. This corresponds to the following cost functional defined on phase space $M^* \times M$ by

$$b(v,x) := \inf\{\langle v, \gamma(0)\rangle + \int_0^1 L(t, \gamma(t), \dot{\gamma}(t))\, dt; \gamma \in C^1([0,T), M); \gamma(1) = x\}. \quad (153)$$

It is then clear that

$$b(t,v,x) = \inf\{\langle v, y\rangle + c(t,y,x);\ y \in M\}, \quad (154)$$

where the cost c is given by the generating function associated to L in Example 3.11, which means that $\underline{\mathcal{B}}_d$ is the convolution of the Brenier cost with the cost induced by the Lagrangian L. The corresponding forward Kantorovich operator is then

$$T^+_b f(x) = T^+_c \circ T^+_2 f(x) = V_{\tilde{f}}(1,x), \quad (155)$$

where $V_{\tilde{f}}(T,x)$ is the final state of the solution of the Hamilton-Jacobi equation (81) starting at $T^+_2 f(x) := \tilde{f}(x) = -f^*(-x)$. So, if μ (resp., ν) is a given probability measure on M^* (resp., M), then we have

$$\begin{aligned} \mathcal{T}_b(\mu,\nu) &= \inf\{\int_{M^* \times M} b(v,x)\, d\pi;\ \pi \in \mathcal{K}(\mu,\nu)\} \quad (156) \\ &= \sup\left\{\int_M V_{\tilde{f}}(T,x)\, d\nu(x) - \int_{M^*} f(v)\, d\mu(v);\ f \text{ convex on } M^*\right\}. \quad (157) \end{aligned}$$

A similar formula holds for the backward Kantorovich operator.

However, we can now convolute a mass transport with a general linear transfer as in the following example.

Example 5.2: Stochastic ballistic transfer (Barton-Ghoussoub [8])

Consider the stochastic ballistic transportation problem defined as:

$$\underline{\mathcal{B}}(\mu,\nu) := \inf\left\{\mathbb{E}\left[\langle V, X(0)\rangle + \int_0^T L(t, X^\beta, \beta(t))\,dt\right]\middle| V\sim\mu, X(\cdot)\in\mathcal{A}, X(T)\sim\nu\right\}, \quad (158)$$

where we are using the notation of Example 4.4. Note that this a convolution of the Brenier-Wasserstein transfer of Example 3.12 with the general stochastic transfer of Example 4.4. Under suitable conditions on L, one gets that

$$\underline{\mathcal{B}}(\mu,\nu) = \sup\left\{\int g\,d\nu - \int \widetilde{\psi_g}\,d\mu; g\in C_b\right\}, \quad (159)$$

where $\widetilde{h}$ is the concave legendre transform of $-h$ and ψ_g is the solution to the Hamilton-Jacobi-Bellman equation

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

In other words, $\underline{\mathcal{B}}$ is a backward linear transform with Kantorovich operator $T^-g = \widetilde{\psi_g}$.

Example 5.3: Broken geodesics on Wasserstein space

Let L be a Lagrangian as above, then for any finite sequence of times $t_1 < t_1 < .... < t_n$, we consider the cost functions $c_i, i = 1, ..., n$,

$$c_i(x,y) = c_{t_i,t_{i+1}} = \inf\left\{\int_{t_i}^{t_{i+1}} L(t,\gamma(t),\dot\gamma(t))\,dt; \gamma(t_i) = x, \gamma(t_{i+1}) = y\right\}.$$

The theory of broken geodesics consist of finding for any fixed x, y, the critical points of the function $(t_1, t_2, ..., t_n) \to c_{t_1,...,t_n}(x,y)$ given by

$$c_{t_1,...,t_n}(x,y) = \inf\{c_1(x,x_1) + c_2(x_1,x_2).... + c_n(x_{n-1},y);\ x_1, x_2, ..., x_{n-1}\in M\}. \quad (160)$$

Thanks to Proposition 6.5, one can consider a broken geodesic problem for stochastic processes by considering for any finite sequence of times $t_1 < t_1 < .... < t_n$ the backward transfer

$$\mathcal{T}_{t_i,t_{i+1}}(\mu,\nu) = \inf\left\{\mathbb{E}\left[\int_{t_i}^{t_{i+1}} L(t, X(t), \beta_X(t, X(t)))\,dt\right]; X(t_i)\sim\mu, X_{t_{i+1}}\sim\nu, X(\cdot)\in\mathcal{A}\right\}, \quad (161)$$

where again $\mathcal{A}$ is the class of processes defined in Section 4.3.

This stochastic transport does not fit in the standard optimal mass transport theory since it does not originate in optimizing a cost between two deterministic states. However, by a result of Mikami-Thieulin [52], $\mathcal{T}_{t_i,t_{i+1}}$ is a backward linear transfer with Kantorovich potential given by $T_{i+1,i}f = V_f(t_i,\cdot)$, where

$$V_f(t,x) = \sup_{X\in\mathcal{A}} \mathbb{E}\left[f(X(T)) - \int_t^{t_{i+1}} L(s, X(s), \beta_X(s,X))\,ds\middle| X(t) = x\right], \quad (162)$$

which is –at least formally– a solution of the Hamilton-Jacobi equation

$$\begin{cases} \partial_t V + H(t,x,\nabla_x V) + \frac{1}{2}\Delta V &= 0 \text{ on } (t_i, t_{i+1}) \times M, \\ V(t_{i+1}, y) &= f(y). \end{cases} \tag{163}$$

One can then define the bacward linear transfer

$$\mathcal{T}_{t_1,\ldots,t_n}(\mu,\nu) = \inf\left\{\mathcal{T}_{t_1,t_2}(\mu,\sigma_1) + \mathcal{T}_{t_2,t_3}(\sigma_1,\sigma_2)\ldots + \mathcal{T}_{t_{n-1},t_n}(\sigma_{n-1},\nu);\ \sigma_1,\ldots\sigma_{n-1} \in \mathcal{P}(M)\right\}, \tag{164}$$

in such a way that

$$\mathcal{T}_{t_1,\ldots,t_n}(\mu,\nu) = \sup\big\{\int_M f(x)\,d\nu(x) - \int_M T_{t_2,t_1} \circ \ldots \circ T_{t_n,t_{n-1}} f(y)\,d\mu(y);\ f \in C(M)\big\}. \tag{165}$$

The broken stochastic geodesics consist of finding for any pair (μ,ν), the critical points of the function $(t_1, t_2, \ldots, t_n) \to \mathcal{T}_{t_1,\ldots,t_n}(\mu,\nu)$ on Wasserstein space. □

Example 5.4: Projection on the set of balayées of a given measure

Let $\mathcal{T}$ be a linear transfer on $X \times Y$ and K a closed convex set of probability measures on Y. We consider the following minimization problem

$$\inf\{\mathcal{T}(\mu,\sigma); \sigma \in K\}, \tag{166}$$

which amounts to finding "the projection" of μ on K, when the "distance" is given by the transfer $\mathcal{T}$. In some cases, the set $K := \mathcal{C}(\nu)$ is a convex compact subset of $\mathcal{P}(Y)$ that depends on a probability measure ν in such a way that the following map

$$\mathcal{S}(\sigma,\nu) = \begin{cases} 0 & \text{if } \sigma \in \mathcal{C}(\nu) \\ +\infty & \text{otherwise.} \end{cases}$$

is a backward transfer on $Y \times Y$. It then follows that

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

If now T^- (resp., S^-) are the backward Kantorovich operators for $\mathcal{T}$ (resp., $\mathcal{S}$), then by Proposition 13.4, the Kantorovich operator for $\mathcal{T} \star \mathcal{S}$ is $T^- \circ S^-$, that is

$$\inf\{\mathcal{T}(\mu,\sigma); \sigma \in \mathcal{C}(\nu)\} = \sup\{\int_Y g\,d\nu - \int_X T^- \circ S^- g\,d\mu; g \in C(Y)\}. \tag{167}$$

Here is an example motivated by a recent result in [41].

Consider now the problem

$$\mathcal{P}(\mu,\nu) = \inf\{\mathcal{T}_c(\mu,\sigma); \sigma \prec_C \nu\}, \tag{168}$$

where $\mathcal{T}_c$ is the optimal mass transport associated to a cost $c(x,y)$ on $X \times Y$, and $\prec_C$ is the convex order on a convex compact set Y. Then,

$$\mathcal{P}(\mu,\nu) = \mathcal{T}_c \star \mathcal{B}(\mu,\nu)$$

where $\mathcal{B}$ is the Balayage transfer. It follows that $\mathcal{P}$ is a linear transfer with backward Kantorovich operator given by the composition of those for $\mathcal{T}_c$ and $\mathcal{B}$, that is

$$T^- f(x) = \sup\{\hat{f}(y) - c(x,y); y \in Y\},$$

where $\hat{f}$ is the concave envelope of f on Y. We note that this is the same Kantorovich operator as for the (weak) barycentric transport (See Proposition 5.1). In other words, we can then deduce the following result of Gozlan-Juillet [41]. Write

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

Corollary 6.7. *Let c be a lower semi-continuous cost functional on $X \times Y$, where Y is convex compact. Then the following holds:*

1. $\mathcal{T}_c \star \mathcal{B} = \mathcal{T}_B^c$.
2. $\mathcal{T}_c \oplus \mathcal{B} = \mathcal{T}_M^c$, *where the latter is the martingale transport of Example 4.4.*

Similar manipulations can be done when the balayage is given by the cones of subharmonic or plurisubharmonic functions.

7 Distance-like transfers

Suppose now that $\mathcal{T}$ is a functional on $\mathcal{P}(X) \times \mathcal{P}(Y)$ satisfying the triangular inequality, that is

$$\mathcal{T}(\mu,\nu) \leqslant \mathcal{T}(\mu,\sigma) + \mathcal{T}(\sigma,\nu) \quad \text{for all } \mu,\, \nu \text{ and } \sigma \text{ in } \mathcal{P}(X), \tag{169}$$

which translates into $\mathcal{T} \leqslant \mathcal{T} \star \mathcal{T}$ and if $\mathcal{T}$ is a backward transfer to $T^- \circ T^- \leqslant T^-$.

Note that if in addition $\mathcal{T}(\mu,\mu) = 0$ for every $\mu \in \mathcal{P}(X)$, then $\mathcal{T} = \mathcal{T} \star \mathcal{T}$. We shall call such a transfer *idempotent*. It is easy to see that $\mathcal{T}$ is idempotent if and only if $(T^-)^2 = T^-$ on $USC(X)$.

7.1 Characterization of $\mathcal{T}$-Lipschitz functions on Wasserstein space

Proposition 7.1. *Let $\mathcal{T}$ be a backward linear transfer on a compact space X and T^- be its associated backward Kantorovich map. If $\mathcal{T}$ satisfies (169), then*

$$\mathcal{T}(\mu,\nu) \geqslant \sup\{\int_X T^- f\, d(\nu-\mu); f \in C(X)\} \quad \textit{for any } \mu \in \mathcal{P}(X) \textit{ and } \nu \in \mathcal{A}. \tag{170}$$

Moreover, if $\mathcal{T}$ is also a forward transfer, then for any $\mu,\nu \in \mathcal{A}$.

$$\mathcal{T}(\mu,\nu) = \sup\{\int_X T^- f\, d(\nu-\mu); f \in C(X)\} = \sup\{\int_X T^- \circ T^+ f\, d(\nu-\mu); f \in C(X)\}. \tag{171}$$

Proof: The proof is straightforward since for every $\nu \in \mathcal{A}$, we have

$$\int_X T^- g\, d\nu \geqslant \int_X g\, d\nu \quad \text{for every } g \in C(X). \tag{172}$$

while if $\mathcal{T}$ satisfies (169), then $\int_X (T^-)^2 g\, d\mu \leqslant \int_X T^- g\, d\mu$ for every $\mu \in \mathcal{P}(X)$.

If now $\mathcal{T}$ is also a forward transfer, then $\int_X T^+ f\, d\nu = \inf\{\int_X f d\sigma + \mathcal{T}(\sigma, \nu); \sigma \in \mathcal{P}(X)\}$, hence for every $\mu \in \mathcal{A}$,

$$\int_X T^+ g\, d\mu \leqslant \int_X g\, d\mu \leqslant \int_X T^- g\, d\mu \text{ for every } g \in C(X).$$

Since by (112) of Proposition 4.6, we have

$$T^+ \circ T^- g(y) \geqslant g(y) \text{ for } y \in Y, \qquad T^- \circ T^+ f(x) \leqslant f(x) \text{ for } x \in X, \tag{173}$$

it follows that for every $\mu \in \mathcal{A}$

$$\int_X T^+ \circ T^- g\, d\mu = \int_X T^- g\, d\mu \text{ for every } g \in C(X).$$

and

$$\int_X T^- \circ T^+ f\, d\mu = \int_X T^+ f\, d\mu \text{ for every } g \in C(X).$$

Assertion (**??**) follows by recalling from Proposition 4.6 that

$$\begin{aligned} \mathcal{T}(\mu, \nu) &= \sup\Big\{ \int_Y T^+ \circ T^- g(y)\, d\nu(y) - \int_X T^- g\, d\mu(x);\ g \in C(Y) \Big\} \\ &= \sup\Big\{ \int_Y T^+ f(y)\, d\nu(y) - \int_X T^- \circ T^+ f\, d\mu(x);\ f \in C(X) \Big\}. \end{aligned}$$

The above proposition states that the maps $\mu \to \int_X T^- f\, d\mu$ and $\mu \to \int_X T^+ \circ T^- g\, d\mu$ are 1-Lipschitz for the metric-like $\mathcal{T}$ on the subset $\mathcal{A}$ of Wasserstein space. We now show the converse, that is all Lipschitz maps on $\mathcal{A}$ are of this form.

Theorem 7.2. *Suppose $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ is bounded below, weak*-lower semi-continuous and convex metric-like functional such that $\mathcal{A} := \{\mu \in \mathcal{P}(X); \mathcal{T}(\mu, \mu) = 0\}$ is non-empty. Assume in addition that for any $\mu, \nu \in \mathcal{P}(X)$, we have*

$$\mathcal{T}(\mu, \nu) = \inf\{\mathcal{T}(\mu, \sigma) + \mathcal{T}(\sigma, \nu); \sigma \in \mathcal{A}\}. \tag{174}$$

Then the following hold:

1. *For any functional $\Phi : \mathcal{A} \to \mathbb{R}$ that is $\mathcal{T}$-Lipschitz, there exists $f \in C(X)$ such that*

$$\Phi(\mu) = \int_X f d\mu \ \text{ for every } \mu \in \mathcal{A}. \tag{175}$$

2. *If $\mathcal{T}$ is also a backward linear transfer, then*

$$\Phi(\mu) = \int_X f d\mu = \int_X T^- f d\mu \ \text{ for every } \mu \in \mathcal{A}. \tag{176}$$

3. *If in addition $\mathcal{T}$ is also a forward linear transfer, then*

$$\Phi(\mu) = \int_X f d\mu = \int_X T^- f d\mu = \int_X T^+ \circ T^- f d\mu \ \text{ for every } \mu \in \mathcal{A}. \tag{177}$$

Note that the functions $\psi_0 := T^- f$ and $\psi_1 := T^+ \circ T^- f$ are conjugate in the sense that $\psi_0 = T^- \psi_1$ and $\psi_1 = T^+ \psi_0$.

4. Moreover, if g is a function in $C(X)$ such that $\int_X g d\mu = \Phi(\mu)$ for all $\mu \in \mathcal{A}$, then

$$\psi_0 \leqslant T^- g \quad \textit{and} \quad \psi_1 \geqslant T^+ g. \tag{178}$$

Proof: Let Φ be such that $\mu \to \Phi(\mu)$ is $\mathcal{T}$-Lipschitz on $\mathcal{A}$ and define

$$\Phi_0(\mu) = \sup_{\sigma \in \mathcal{A}} \{\Phi(\sigma) - \mathcal{T}(\mu, \sigma)\} \quad \text{and} \quad \Phi_1(\nu) = \inf_{\sigma \in \mathcal{A}} \{\Phi(\sigma) + \mathcal{T}(\sigma, \nu)\}.$$

It is clear that

$$\Phi_1(\mu) \leqslant \Phi(\mu) \leqslant \Phi_0(\mu) \quad \text{for all } \mu \in \mathcal{A}. \tag{179}$$

We now show that

$$\Phi_0(\mu) \leqslant \Phi_1(\mu) \quad \text{for all } \mu \in \mathcal{P}(X). \tag{180}$$

For that note that (169) and the fact that $\mu \to \Phi(\mu)$ is $\mathcal{T}$-Lipschitz on $\mathcal{A}$ yield

$$\begin{aligned} \Phi_0(\mu) - \Phi_1(\mu) &= \sup_{\sigma, \tau \in \mathcal{A}} \{\Phi(\sigma) - \mathcal{T}(\mu, \sigma) - \Phi(\tau) - \mathcal{T}(\tau, \mu)\} \\ &\leqslant \sup_{\sigma, \tau \in \mathcal{A}} \{\Phi(\sigma) - \Phi(\tau) - \mathcal{T}(\tau, \sigma)\} \leqslant 0. \end{aligned}$$

This combined with (179) shows that

$$\Phi_1(\mu) = \Phi(\mu) = \Phi_0(\mu) \quad \text{for all } \mu \in \mathcal{A}.$$

We now show that for every $\mu \in \mathcal{P}(X)$,

$$\Phi_0(\mu) = \sup\{\Phi_1(\sigma) - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(X)\}. \tag{181}$$

For every $\mu \in \mathcal{P}(X)$, we have

$$\Phi_0(\mu) = \sup_{\sigma \in \mathcal{A}} \{\Phi(\sigma) - \mathcal{T}(\mu, \sigma)\} = \sup_{\sigma \in \mathcal{A}} \{\Phi_1(\sigma) - \mathcal{T}(\mu, \sigma)\} \leqslant \sup_{\sigma \in \mathcal{P}(X)} \{\Phi_1(\sigma) - \mathcal{T}(\mu, \sigma)\}.$$

On the other hand, for any $\nu, \mu \in \mathcal{P}(X)$, we have

$$\begin{aligned} \Phi_1(\nu) - \Phi_0(\mu) &= \inf_{\sigma, \tau \in \mathcal{A}} \{\Phi(\sigma) + \mathcal{T}(\sigma, \nu) - \Phi(\tau) + \mathcal{T}(\mu, \tau)\} \\ &\leqslant \inf_{\sigma, \tau \in \mathcal{A}} \{\mathcal{T}(\sigma, \nu) + \mathcal{T}(\tau, \sigma) + \mathcal{T}(\mu, \tau)\} \\ &\leqslant \inf_{\sigma \in \mathcal{A}} \{\mathcal{T}(\sigma, \nu) + \mathcal{T}(\mu, \sigma)\} \\ &= \mathcal{T}(\mu, \nu). \end{aligned}$$

This shows (181). The other conjugate formula

$$\Phi_1(\nu) = \inf\{\Phi_0(\sigma) + \mathcal{T}(\sigma, \nu); \sigma \in \mathcal{P}(X)\} \tag{182}$$

can be proved in a similar fashion.

Note now that Φ_0 is a concave weak*-upper semi-continuous function on $\mathcal{P}(X)$, while Φ_1 is a convex weak*-lower semi-continuous. Since $\Phi_0 \leqslant \Phi_1$ on $\mathcal{P}(X)$, there exists $f \in C(X)$ such that

$$\Phi_0(\mu) \leqslant \int_X f d\mu \leqslant \Phi_1(\mu) \quad \text{for all } \mu \in \mathcal{P}(X), \tag{183}$$

hence

$$\Phi_0(\mu) = \int_X f d\mu = \Phi_1(\mu) = \Phi(\mu) \quad \text{for all } \mu \in \mathcal{A}. \tag{184}$$

2) Suppose now $\mathcal{T}$ is also a backward linear transfer with T^- as a Kantorovich operator, then

$$\int_X T^- f d\mu = \sup_{\sigma \in \mathcal{P}(X)} \{\int_X f d\sigma - \mathcal{T}(\mu, \sigma)\} \leqslant \sup_{\sigma \in \mathcal{P}(X)} \{\Phi_1(\sigma) - \mathcal{T}(\mu, \sigma)\} = \Phi_0(\mu).$$

On the other hand, if $\mu \in \mathcal{A}$,

$$\int_X T^- f d\mu \geqslant \sup_{\sigma \in \mathcal{A}} \{\int_X f d\sigma - \mathcal{T}(\mu, \sigma)\} \geqslant \int_X f d\mu - \mathcal{T}(\mu, \mu) = \int_X f d\mu.$$

3) Suppose in addition that $\mathcal{T}$ is a forward linear transfer with T^+ as a Kantorovich operator, then

$$\int_X T^+ \circ T^- f d\mu = \inf_{\sigma \in \mathcal{P}(X)} \{\int_X T^- f d\sigma + \mathcal{T}(\sigma, \mu)\} \leqslant \inf_{\sigma \in \mathcal{P}(X)} \{\Phi_0(\sigma) + \mathcal{T}(\sigma, \mu)\} = \Phi_1(\mu).$$

On the other hand, $T^+ \circ T^- f \geqslant f$ in such a way that

$$\int_X T^+ \circ T^- f d\mu \geqslant \int_X f d\mu \geqslant \Phi_0(\mu).$$

In other words, $T^- f$ and $T^+ \circ T^- f$ are two conjugate functions verifying

$$\int_X T^+ \circ T^- f d\mu = \int_X T^- f d\mu = \Phi(\mu) \quad \text{for all } \mu \in \mathcal{A}.$$

4) To prove (178), first note that

$$\begin{aligned}
\int_X T^- f d\mu &\leqslant \Phi_0(\mu) = \sup\{\Phi(\sigma) - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{A}\} \\
&\leqslant \sup\{\int_X g d\sigma - \mathcal{T}(\mu, \sigma); \sigma \in \mathcal{P}(X)\} \\
&= \int_X T^- g d\mu.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\int_X T^+ \circ T^- f d\mu &= \inf\{\int_X T^- f d\sigma + \mathcal{T}(\sigma,\mu); \sigma \in \mathcal{P}(X)\} \\
&= \inf\{\int_X T^- f d\sigma + \mathcal{T}(\sigma,\lambda) + \mathcal{T}(\lambda,\mu); \lambda \in \mathcal{A}, \sigma \in \mathcal{P}(X)\} \\
&= \inf\{\int_X T^+ \circ T^- f d\lambda + \mathcal{T}(\lambda,\mu); \lambda \in \mathcal{A}\} \\
&= \inf\{\int_X g d\lambda + \mathcal{T}(\lambda,\mu); \lambda \in \mathcal{A}\} \\
&= \inf\{\int_X g d\lambda + \mathcal{T}(\lambda,\mu); \lambda \in \mathcal{A}\} \\
&\geqslant \inf\{\int_X g d\lambda + \mathcal{T}(\lambda,\mu); \lambda \in \mathcal{P}(X)\} \\
&= \int_X T^+ g d\mu,
\end{aligned}$$

which completes the proof of the theorem. □

7.2 Examples of idempotent transfers

In the next sections, we shall associate to any backward or forward linear transfer an idempotent linear transfer. For now, we give a few examples of some transfers that are readily idempotent.

1. If I is any bounded below convex lower semi-continuous functional on Wasserstein space $\mathcal{P}(Y)$, and $m = \inf\{I(\sigma); \sigma \in \mathcal{P}(Y)$, then $\mathcal{T}(\mu,\nu) = I(\nu) - m$ is an idempotent backward linear transfer with an idempotent Kantorovich map $T^- f = I^*(f) + m$.

2. Any transfer induced by a bounded positive linear operator T with $T^2 = T$ and $T1 = 1$, and in particular, any point transformation σ such that $\sigma^2 = \sigma$ as per Example 3. 2.

3. The balayage transfer $\mathcal{B}$ since its Kantorovich map is $Tf = \hat{f}$, where for example in the case of balayage with convex functions, $\hat{f}$ is the concave envelope of f.

4. If $\mathcal{T}_c$ is an optimal mass transport associated to a cost function c, then $\mathcal{T}_c$ is idempotent if $c(x,x) = 0$ for every $x \in X$ and c satisfies the triangular inequality

$$c(x,z) \leqslant c(x,y) + c(y,z) \quad \text{for all } x, y, z \text{ in } X, \tag{185}$$

in which case

$$\mathcal{T}_c(\mu,\nu) = \sup\{\int_X T_c f \, d(\nu - \mu); f \in C(X)\}. \tag{186}$$

A typical example is the Rubinstein-Kantorovich optimal mass transport associated to any metric -such as in the original Monge problem- since the latter satisfies the triangular inequality and is zero on the diagonal. If $c_p(x,y) = |x-y|^p$ and $0 < p \leqslant 1$, then the corresponding optimal mass transport is idempotent since c_p again satisfies the triangular inequality. $c_p \star c_p(x,y) = c_p(x,y)$, so $T^2 f(x) = Tf(x)$, i.e. T is idempotent.

Example 6.7: An idempotent optimal Skorohod embedding

The following transfer was considered in Ghoussoub-Kim-Palmer [35].

$$\mathcal{T}(\mu,\nu) := \inf\Big\{\mathbb{E}\Big[\int_0^\tau L(t,B_t)dt\Big];\ \tau\in\mathcal{S}(\mu,\nu)\Big\}, \tag{187}$$

where $\mathcal{S}(\mu,\nu)$ denotes the set of –possibly randomized– stopping times with finite expectation such that ν is realized by the distribution of B_τ (i.e, $B_\tau\sim\nu$ in our notation), where B_t is Brownian motion starting with μ as a source distribution, i.e., $B_0\sim\mu$. Note that $\mathcal{T}(\mu,\nu)=+\infty$ if $\mathcal{S}(\mu,\nu)=\emptyset$, which is the case if and only if μ and ν are not in subharmonic order. In this case, It has been proved in [35] that under suitable conditions, the backward linear transfer is given by $T^-\psi = J_\psi(0,\cdot)$, where $J_\psi : \mathbb{R}^+\times\mathbb{R}^d\to\mathbb{R}$ is defined via the dynamic programming principle

$$J_\psi(t,x) := \sup_{\tau\in\mathcal{R}^{t,x}}\Big\{\mathbb{E}^{t,x}\Big[\psi(B_\tau)-\int_t^\tau L(s,B_s)ds\Big]\Big\}, \tag{188}$$

where the expectation superscripted with t,x is with respect to the Brownian motions satisfying $B_t=x$, and the minimization is over all finite-expectation stopping times $\mathcal{R}^{t,x}$ on this restricted probability space such that $\tau\geq t$. $J_\psi(t,x)$ is actually a "variational solution" for the quasi-variational Hamilton-Jacobi-Bellman equation:

$$\min\left\{\begin{array}{r} J(t,x)-\psi(x) \\ -\frac{\partial}{\partial t}J(t,x)-\frac{1}{2}\Delta J(t,x)+L(t,x)\end{array}\right\}=0. \tag{189}$$

Note that $J_\psi(t,x)\geqslant\psi(x)$, that is $T^-\psi\geqslant\psi$ for every ψ.

Assume now $t\to L(t,x)$ is decreasing, which yields that $t\to J(t,x)$ is increasing (see [35]). if $\psi(x)=T^-\varphi=J_\varphi(0,x)$ for some φ, then for each $\epsilon>0$, there is τ such that

$$\begin{aligned} J_\psi(0,x) &\leqslant \mathbb{E}^{t,x}\Big[\psi(B_\tau)-\int_t^\tau L(s,B_s)ds\Big]+\epsilon \\ &\leqslant \mathbb{E}^{t,x}\Big[J_\varphi(t,B_\tau)-\int_t^\tau L(s,B_s)ds\Big]+\epsilon \\ &\leqslant J_\varphi(0,x)+\epsilon. \end{aligned}$$

where the last inequality uses the supermartingale property of the process $t\to J_\varphi(t,B_\tau)-\int_t^\tau L(s,B_s)ds$. It follows that

$$T^-\varphi(x)\leqslant (T^-)^2\varphi(x)=J_\psi(0,x)\leqslant J_\varphi(0,x)=T^-\varphi(x),$$

and T^- is therefore idempotent.

8 Ergodic properties of equicontinuous semigroups of transfers

Let X be a compact space. Our main purpose is to associate to any backward linear transfer $\mathcal{T}$ on $\mathcal{P}(X) \times \mathcal{P}(X)$, an idempotent backward linear transfer $\mathcal{T}_\infty$ with the properties listed in Theorem 8.1 below. For that, we shall associate to $\mathcal{T}$, the semi-group of transfers $(\mathcal{T}_n)_n$ defined for each $n \in \mathbb{N}$, as $\mathcal{T}_n = \mathcal{T} \star \mathcal{T} \star .... \star \mathcal{T}$ n-times and study its limit as $n \to \infty$. This section deals with the case where $\mathcal{T}$ is continuous, hence the sequence of transfers $(\mathcal{T}_n)_n$ is equicontinuous for the Wasserstein metric. We shall prove the following.

Theorem 8.1 (Fixed point of weak* continuous backward linear transfers)**.** *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ that is weak*-continuous on $\mathcal{M}(X)$, and let T^- be the corresponding backward Kantorovich operator that maps $C(X)$ into $C(X)$. Then, there exists a constant $c = c(\mathcal{T}) \in \mathbb{R}$, an idempotent backward linear transfer $\mathcal{T}_\infty^-$ on $\mathcal{P}(X) \times \mathcal{P}(X)$ with Kantorovich operator $T_\infty^- : C(X) \to C(X)$ such that,*

1. *The constant $c(\mathcal{T}) = \inf\{\mathcal{T}(\mu, \mu); \mu \in \mathcal{P}(X)\}$;*
2. *For every $f \in C(X)$ and $x \in X$, $\lim\limits_{n\to+\infty} \frac{(T^-)^n f(x)}{n} = -c$:*
3. *$\mathcal{T}_\infty = (\mathcal{T} - c) \star \mathcal{T}_\infty$ and $T^- \circ T_\infty^- f + c = T_\infty^- f$ for all $f \in C(X)$;*
4. *The set $\mathcal{A} := \{\mu \in \mathcal{P}(X); \mathcal{T}_\infty(\mu, \mu) = 0\}$ is non-empty and for every $\mu, \nu \in \mathcal{P}(X)$, we have*
$$\mathcal{T}_\infty(\mu, \nu) = \inf\{\mathcal{T}_\infty(\mu, \sigma) + \mathcal{T}_\infty(\sigma, \nu), \sigma \in \mathcal{A}\}. \tag{190}$$

This will follow from the following more general result. But first, we mention that there is an analogous result for the case when $\mathcal{T}$ is a forward linear transfer with operator T^+. The same statements hold as above, the only difference being that

$$\lim_{n\to+\infty} \frac{(T^+)^n f(x)}{n} = c \quad \text{for every } f \in C(X) \text{ and } x \in X, \tag{191}$$

and

$$T^+ \circ T_\infty^+ f - c = T_\infty^+ f \qquad \text{for all } f \in C(X). \tag{192}$$

If now $\mathcal{T}$ is simultaneously a backward and forward transfer, then we have the following,

Corollary 8.2. *Suppose $\mathcal{T}$ is a backward and forward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ that is continuous for the Wasserstein metric, then the associated effective transfer $\mathcal{T}_\infty$ is also a backward and forward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$, with T_∞^- (resp., T_∞^+) as corresponding backward (resp., forward) effective Kantorovich operator. Moreover, The associated effective transfer $\mathcal{T}_\infty$ can be expressed as*

$$\mathcal{T}_\infty(\mu, \nu) = \sup\Big\{ \int_X f^+ \, d\nu - \int_X f^- \, d\mu;\ (f^-, f^+) \in \mathcal{I}\Big\}, \tag{193}$$

where

$$\mathcal{I} = \big\{(f^-, f^+);\ f^- \text{ (resp., } f^+\text{) is a backward (resp., forward) solution and } f^- = f^+ \text{ on } \mathcal{A}\big\}.$$

Proof: Since T^+_∞ and T^-_∞ are the Kantorovich opeartors for $\mathcal{T}_\infty$, we can use (174) of Proposition 4.6 to write

$$\mathcal{T}_\infty(\mu,\nu) = \sup\Big\{\int_X T^+_\infty \circ T^-_\infty g\, d\nu - \int_X T^-_\infty g\, d\mu;\ g \in C(X)\Big\} \tag{194}$$

$$= \sup\Big\{\int_X T^+_\infty f\, d\nu - \int_X T^-_\infty \circ T^+_\infty f\, d\mu;\ f \in C(X)\Big\}. \tag{195}$$

Note now that $f^- = T^-_\infty g$ (resp., $f^+ = T^+_\infty f^-$) is a backward (resp., forward) weak KAM solution for $\mathcal{T}$ and in view of Proposition 7.1, $\int_X f^-\, d\mu = \int_X f^+\, d\mu$ for every $\mu \in \mathcal{A}$. It follows that

$$\mathcal{T}_\infty(\mu,\nu) \leqslant \sup\Big\{\int_X f^+\, d\nu - \int_X f^-\, d\mu;\ (f^-, f^+) \in \mathcal{I}\Big\}, \tag{196}$$

For the reverse inequality, note first that if $f^-, f^+ \in \mathcal{I}$, then since $f^- = T^- f^-$ and $f^+ = T^+ f^+$, the functions $\mu \to \int_X f^- d\mu$ and $\mu \to \int_X f^+ d\mu$ are $\mathcal{T}_\infty$-Lipschitz on the set $\mathcal{A}$. Hence Theorem 7.2 applies and we get a function χ such that $T^-\chi \leqslant T^- f^- = f^-$ and $T^+ \circ T^-\chi \geqslant T^+ f^+ = f^+$. This readily implies the reverse inequality, hence that (193) hold.

8.1 Effective Kantorovich operator associated to a semi-group of linear transfers

Let $\{\mathcal{T}_t\}_{t\geqslant 0}$ be a family of backward linear transfers on $\mathcal{P}(X) \times \mathcal{P}(X)$ with associated Kantorovich operators $\{T_t\}_{t\geqslant 0}$, where $\mathcal{T}_0$ is the identity transfer,

$$\mathcal{T}_0(\mu,\nu) = \begin{cases} 0 & \text{if } \mu = \nu \in \mathcal{P}(X) \\ +\infty & \text{otherwise.} \end{cases}$$

We make the following assumptions:

(H0) The family $\{\mathcal{T}_t\}_{t\geqslant 0}$ is a semi-group under inf-convolution: $\mathcal{T}_{t+s} = \mathcal{T}_t \star \mathcal{T}_s$ for all $s, t \geqslant 0$.

(H1) For every $t > 0$, the transfer $\mathcal{T}_t$ is weak*-continuous, and the Dirac measures are contained in $D_1(\mathcal{T}_t)$.

(H2) For any $\epsilon > 0$, $\{\mathcal{T}_t\}_{t\geqslant \epsilon}$ has common modulus of continuity δ (possibly depending on ϵ).

The hypotheses (H1) and (H2) amount to an equi-continuity assumption for the family $\{T_t f\}_{t\geqslant 0}$ for each f, and is an artifact to ensure that we remain within the class of continuous functions in the limit $t \to +\infty$ (thanks to Arzela-Ascoli). It is likely these hypotheses can be weakened. Note in relation to (H2) that the semi-group property (H0) implies that a modulus of continuity for $\mathcal{T}_t$ is also one for $\mathcal{T}_{Nt}$, $N \in \mathbb{N}$. In the following, where we will be concerned with taking limits as $t \to +\infty$, it suffices to take $\epsilon = 1$.

Proposition 8.3. *Under condition (H0), there exists a finite constant c and a positive constant $C > 0$ such that*

$$|\mathcal{T}_t(\mu,\nu) - tc| \leqslant C, \quad \textit{for every } t \geqslant 1 \textit{ and all } \mu,\nu \in \mathcal{P}(X).$$

In particular,

$$c = \lim_{t\to+\infty} \frac{\inf\{\mathcal{T}_t(\mu,\nu)\,;\,\mu,\nu\in\mathcal{P}(X)\}}{t}.$$

We shall call the constant $c(\mathcal{T})$ in Proposition 8.3 the *Mañé critical value*, while the solutions $u \in C(X)$ of the functional equation $T_t u + ct = u$ for all $t \geqslant 0$, will be called *backward weak KAM solutions*.

Proof: Define $M_t := \max_{\mu,\nu} \mathcal{T}_t(\mu,\nu)$ and $M := \inf_{t\geqslant 1}\{\frac{M_t}{t}\} > -\infty$. The sequence $\{M_t\}_{t\geqslant 1}$ is subadditive, that is $M_{t+s} \leqslant M_t + M_s$, hence it is well known (see e.g. [12]) that $\{\frac{M_t}{t}\}_{t\geqslant 1}$ decreases to its infimum M as $t \to \infty$. Indeed, fix $t > 0$ and write for any s, the decomposition $s = nt + r$, where $0 \leqslant r < t$. The subadditivity of M_t implies

$$\frac{M_s}{s} = \frac{M_{nt+r}}{nt+r} \leqslant \frac{M_{nt}}{nt} + \frac{M_r}{nt} \leqslant \frac{M_t}{t} + \frac{M_r}{nt}.$$

It follows that $\limsup_{s\to\infty} \frac{M_s}{s} \leqslant \frac{M_t}{t}$. On the other hand, $\inf_{t\geqslant 1} \frac{M_t}{t} \leqslant \liminf_{t\to\infty} \frac{M_t}{t}$. Therefore, $\frac{M_t}{t}$ converges to M as $t \to \infty$.
On the other hand, if $m_t := \min_{\mu,\nu} \mathcal{T}_t(\mu,\nu)$, then the above applied to $-m_t$ yields that $\lim_{t\to\infty} \frac{m_t}{t} = m$.
We now show that $m = M$. The uniform modulus of continuity δ implies the existence of a constant $C > 0$, such that $M_t - m_t \leqslant C$ for every $t > 0$. Then, we obtain the string of inequalities

$$tM - C \leqslant M_t - C \leqslant m_t \leqslant \mathcal{T}_t(\mu,\nu) \leqslant M_t \leqslant m_t + C \leqslant tm + C.$$

The left-most and right-most inequalities imply $M \leqslant m$ upon sending $t \to \infty$, hence $m = M$. □

From Property 1) of Kantorovich operators and Proposition 8.3, we can deduce the following.

Lemma 8.4. *Under conditions* $(H0)$, $(H1)$, *and* $(H2)$, *and with the notation of Proposition 8.3, the following properties hold.*

1. *For any* $f \in C(X)$, *we have* $|T_t f(x) + ct - \sup_X f| \leqslant C$ *for all* $t \geqslant 1$ *and all* $x \in X$.
2. *The semi-group of operators* $\{T_t\}_{t\geqslant 1}$ *has the same modulus of continuity* δ *as* $\{\mathcal{T}_t\}_{t\geqslant 1}$.
3. *If* $k < c$, *then* $T_t f + kt \to -\infty$, *while if* $k > c$, $T_t f + kt \to +\infty$, *as* $t \to \infty$, *for any* $f \in C(X)$.

Proof: 1) By Proposition 8.3 and since $T_t f(x) + ct = \sup_\sigma\{\int f\,\mathrm{d}\sigma - (\mathcal{T}_t(\delta_x,\sigma) - ct)\}$, we have $\sup_X f - C \leqslant T_t f(x) + ct \leqslant \sup_X f + C$.

For 2) we note that

$$\begin{aligned} T_t f(x) &= \sup_\sigma\{\int f\,\mathrm{d}\sigma - \mathcal{T}_t(\delta_x,\sigma)\} \\ &\leqslant \sup_\sigma\{\int f\,\mathrm{d}\sigma - \mathcal{T}_t(\delta_y,\sigma)\} + \sup_\sigma\{\mathcal{T}_t(\delta_y,\sigma) - \mathcal{T}_t(\delta_x,\sigma)\} \\ &= T_t f(y) + \delta(d(x,y)). \end{aligned}$$

We now interchange x and y to obtain the reverse inequality.

3) follows from 1) since $\sup_X f - C + (k-c)t \leqslant T_t f(x) + kt \leqslant \sup_X f + C + (k-c)t$.

Theorem 8.5. *Given a semi-group of backward linear transfers $(\mathcal{T}_t)_{t\geqslant 0}$ satisfying conditions (H0), (H1), and (H2), there exist a backward linear transfer $\mathcal{T}_\infty$, an associated Kantorovich operator $T_\infty : C(X) \to C(X)$ and a constant $c \in \mathbb{R}$ such that:*

1. *For every $f \in C(X)$, $T_\infty f$ is a backward weak KAM solution, and T_∞ is idempotent. In particular, backward weak KAM solutions are fixed points of T_∞.*
2. *The backward linear transfer $\mathcal{T}_\infty$ satisfies,*
$$\mathcal{T}_\infty = (\mathcal{T}_t - ct) * \mathcal{T}_\infty \text{ for every } t \geqslant 0, \quad \text{and} \quad \mathcal{T}_\infty = \mathcal{T}_\infty * \mathcal{T}_\infty. \tag{197}$$
3. *For every $\mu, \nu \in \mathcal{P}(X)$, we have*
$$\sup\left\{\int T_\infty f \,\mathrm{d}(\nu - \mu)\,;\, f \in C(X)\right\} \leqslant \mathcal{T}_\infty(\mu,\nu) \leqslant \liminf_{t\to\infty}(\mathcal{T}_t(\mu,\nu) - ct). \tag{198}$$
4. *The set $\mathcal{A} := \{\sigma \in \mathcal{P}(X); \mathcal{T}_\infty(\sigma,\sigma) = 0\}$ is non-empty, and for every $\mu, \nu \in \mathcal{P}(X)$, we have*
$$\mathcal{T}_\infty(\mu,\nu) = \inf\{\mathcal{T}_\infty(\mu,\sigma) + \mathcal{T}_\infty(\sigma,\nu), \sigma \in \mathcal{A}\}, \tag{199}$$
and the infimum on $\mathcal{A}$ is attained.
5. *We also have*
$$c = \inf\{\mathcal{T}_1(\mu,\mu); \mu \in \mathcal{P}(X)\}, \tag{200}$$
and the infimum is attained by a measure $\bar\mu \in \mathcal{A}$ such that
$$(\bar\mu,\bar\mu) \in \mathcal{D} := \{(\mu,\nu) \in \mathcal{P}(X)\times\mathcal{P}(X) \,:\, \mathcal{T}_1(\mu,\nu) + \mathcal{T}_\infty(\nu,\mu) = c\}. \tag{201}$$
Moreover, every measure which attains the infimum in (200) belongs to $\mathcal{A}$.

The backward linear transfer $\mathcal{T}_\infty$ is an analog of the *Peierls barrier*, and the set $\mathcal{A}$ is an analog of the *projected Aubry set*.

Proof: 1) Given $f \in C(X)$, define $\bar T f(x) := \limsup_{t\to\infty}(T_t f(x) + ct)$. By (H2), $\bar T f$ has modulus of continuity δ, and $\|\bar T f\|_\infty \leqslant \sup_X f + C$.

Noting that $\sup_{s\geqslant t}\{T_s f(x) + cs\}$ is a sequence of continuous functions that decrease monotonically to $\bar T f(x)$ as $t \to \infty$, we may apply Lemma 4.2 to deduce for any $t' \geqslant 0$,

$$\begin{aligned}
T_{t'}\bar T f(x) &= \lim_{t\to\infty} T_{t'}\left[\sup_{s\geqslant t}\{T_s f(x) + cs\}\right] \\
&\geqslant \lim_{t\to\infty}\sup_{s\geqslant t}\{T_{t'+s} f(x) + cs\} \\
&= \lim_{t\to\infty}\sup_{s\geqslant t}\{T_{t'+s} f(x) + c(t'+s)\} - ct' \\
&= \bar T f(x) - ct'.
\end{aligned}$$

Therefore, $T_{t'}\bar{T}f(x) + ct' \geqslant \bar{T}f(x)$. By monotonicity of the operators T_t, this inequality implies

$$T_t\bar{T}f(x) + ct \geqslant T_s\bar{T}f(x) + cs$$

whenever $t \geqslant s \geqslant 0$, i.e. $\{T_t\bar{T}f + ct\}_{t\geqslant 1}$ is a monotone increasing sequence of continuous functions. In addition, we have from Corollary 8.4 the uniform in time bound

$$\|T_t\bar{T}f(x) + ct\|_\infty \leqslant \|\bar{T}f\|_\infty + C \leqslant \|f\|_\infty + 2C.$$

We may therefore define $T_\infty : C(X) \to C(X)$ via the formula,

$$T_\infty f(x) := \lim_{t\to\infty} T_t\bar{T}f(x) + ct,$$

and from Lemma 4.2 deduce

$$\begin{aligned} T_t T_\infty f(x) + ct &= \lim_{s\to\infty} T_t\left[T_s\bar{T}f(x) + cs\right] + ct \\ &= \lim_{s\to\infty}\left\{T_{t+s}\bar{T}f(x) + c(t+s)\right\} \\ &= T_\infty f(x). \end{aligned}$$

This further implies that $T_\infty T_\infty f(x) = T_\infty f(x)$ so T_∞ is idempotent. It is straightforward to see that in the construction of T_∞, properties 1)-4) of Proposition **??** are preserved, and hence T_∞ is a Kantorovich operator.

Finally we note that if u satisfies $T_t u + ct = u$, then $T_\infty u = u$ from the defintion of T_∞.
2) T_∞ is a Kantorovich operator, thus we may define

$$\mathcal{T}_\infty(\mu,\nu) := \sup\left\{\int f\,\mathrm{d}\nu - \int T_\infty f\,\mathrm{d}\mu\,;\, f \in C(X)\right\}$$

and it is a backward linear transfer; from $T_t T_\infty f + ct = T_\infty f$, it satisfies

$$\mathcal{T}_\infty(\mu,\nu) = (\mathcal{T}_t - ct) \star \mathcal{T}_\infty(\mu,\nu), \quad \text{for all } t \geqslant 0,$$

and from $T_\infty T_\infty u(x) = T_\infty u(x)$, it satisfies

$$\mathcal{T}_\infty(\mu,\nu) = \mathcal{T}_\infty \star \mathcal{T}_\infty(\mu,\nu), \text{for all } \mu,\nu.$$

3) Note from 1 that $T_\infty f(x) \geqslant \limsup_{t\to\infty}(T_t f(x) + ct)$, so

$$\begin{aligned} \int_X T_\infty f\,\mathrm{d}\mu &\geqslant \int_X \limsup_{t\to\infty}(T_t f(x) + ct)\,\mathrm{d}\mu \\ &\geqslant \limsup_{t\to\infty}\int_X (T_t f(x) + ct)\,\mathrm{d}\mu. \end{aligned}$$

Hence

$$\begin{aligned} \mathcal{T}_\infty(\mu,\nu) &\leqslant \sup\liminf_{t\to\infty}\left\{\int_X f\,\mathrm{d}\nu - \int_X T_t f\,\mathrm{d}\mu - ct\,;\, f \in C(X)\right\} \\ &\leqslant \liminf_{t\to\infty}\sup\left\{\int_X f\,\mathrm{d}\nu - \int_X T_t f\,\mathrm{d}\mu - ct\,;\, f \in C(X)\right\} \\ &= \liminf_{t\to\infty}(\mathcal{T}_t(\mu,\nu) - ct). \end{aligned}$$

On the other hand, from $T_\infty \circ T_\infty f = T_\infty f$,

$$\begin{aligned}\mathcal{T}_\infty(\mu,\nu) &= \sup\left\{\int_X f\,\mathrm{d}\nu - \int_X T_\infty f\,\mathrm{d}\mu\,;\ f\in C(X)\right\}\\ &\geqslant \sup\left\{\int_X T_\infty f\,\mathrm{d}(\nu-\mu)\,;\ f\in C(X)\right\}.\end{aligned}$$

4) The proof of this result relies solely on the property $\mathcal{T}_\infty = \mathcal{T}_\infty \star \mathcal{T}_\infty$, and the argument is a minor adaption of the one given in [6].

Fix $\mu,\nu\in\mathcal{P}(X)$. From $\mathcal{T}_\infty = \mathcal{T}_\infty \star \mathcal{T}_\infty$, there exists $\sigma_1\in\mathcal{P}(X)$ such that

$$\mathcal{T}_\infty(\mu,\nu) = \mathcal{T}_\infty(\mu,\sigma_1) + \mathcal{T}_\infty(\sigma_1,\nu).$$

Similarly, there exists a σ_2 such that

$$\mathcal{T}_\infty(\sigma_1,\nu) = \mathcal{T}_\infty(\sigma_1,\sigma_2) + \mathcal{T}_\infty(\sigma_2,\nu).$$

Combining the above two equalities, we obtain

$$\mathcal{T}_\infty(\mu,\nu) = \mathcal{T}_\infty(\mu,\sigma_1) + \mathcal{T}_\infty(\sigma_1,\sigma_2) + \mathcal{T}_\infty(\sigma_2,\nu).$$

Note also that

$$\mathcal{T}_\infty(\mu,\sigma_1) + \mathcal{T}_\infty(\sigma_1,\sigma_2) = \mathcal{T}_\infty(\mu,\sigma_2). \tag{202}$$

This follows from

$$\begin{aligned}\mathcal{T}_\infty(\mu,\nu) &= \mathcal{T}_\infty(\mu,\sigma_1) + \mathcal{T}_\infty(\sigma_1,\sigma_2) + \mathcal{T}_\infty(\sigma_2,\nu)\\ &\geqslant \mathcal{T}_\infty\star\mathcal{T}_\infty(\mu,\sigma_2) + \mathcal{T}_\infty(\sigma_2,\nu)\\ &= \mathcal{T}_\infty(\mu,\sigma_2) + \mathcal{T}_\infty(\sigma_2,\nu)\\ &\geqslant \mathcal{T}_\infty\star\mathcal{T}_\infty(\mu,\nu)\\ &= \mathcal{T}_\infty(\mu,\nu).\end{aligned}$$

Hence all the inequalities are equalities; in particular (202).

After k times we have

$$\mathcal{T}_\infty(\mu,\nu) = \sum_{i=0}^{k} \mathcal{T}_\infty(\sigma_i,\sigma_{i+1})$$

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

$$\sum_{i=\ell}^{m} \mathcal{T}_\infty(\sigma_i,\sigma_{i+1}) = \mathcal{T}_\infty(\sigma_\ell,\sigma_{m+1})$$

whenever $0\leqslant \ell < m \leqslant k$. In particular, for any subsequence σ_{k_j}, we have

$$\mathcal{T}_\infty(\mu,\sigma_{k_1}) + \sum_{j=1}^{m} \mathcal{T}_\infty(\sigma_{k_j},\sigma_{k_{j+1}}) + \mathcal{T}_\infty(\sigma_{k_{m+1}},\nu) = \mathcal{T}_\infty(\mu,\nu). \tag{203}$$

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

$$\liminf_j \mathcal{T}_\infty(\sigma_{k_j}, \sigma_{k_{j+1}}) \geqslant \mathcal{T}_\infty(\bar{\sigma}, \bar{\sigma}).$$

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

$$\mathcal{T}_\infty(\sigma_{k_j}, \sigma_{k_{j+1}}) \geqslant \mathcal{T}_\infty(\bar{\sigma}, \bar{\sigma}) - \epsilon. \tag{204}$$

Therefore, by refining to a further (non-relabeled) subsequence if necessary, we obtain a subsequence $\{\sigma_{k_j}\}$ satisfying (204) for all j. By further refinement, we may also assume,

$$\mathcal{T}_\infty(\mu, \sigma_{k_1}) \geqslant \mathcal{T}_\infty(\mu, \bar{\sigma}) - \epsilon. \tag{205}$$

Therefore, by refining to a further (non-relabeled) subsequence if necessary, we obtain a subsequence $\{\sigma_{k_j}\}$ with properties (203), (204), and (205).

Moreover, for all m large enough (depending on ϵ), we have

$$\mathcal{T}_\infty(\sigma_{k_{m+1}}, \nu) \geqslant \mathcal{T}_\infty(\bar{\sigma}, \nu) - \epsilon \tag{206}$$

Applying the inequalities of (204), (205), and (206), to (203), we obtain

$$\mathcal{T}_\infty(\mu, \nu) \geqslant \mathcal{T}_\infty(\mu, \bar{\sigma}) + m\mathcal{T}_\infty(\bar{\sigma}, \bar{\sigma}) + \mathcal{T}_\infty(\bar{\sigma}, \nu) - (m+2)\epsilon$$

for large enough m. From the fact that $\mathcal{T}_\infty = \mathcal{T}_\infty * \mathcal{T}_\infty$, the above inequality is only possible if

$$\mathcal{T}_\infty(\bar{\sigma}, \bar{\sigma}) \leqslant \frac{m+2}{m}\epsilon \leqslant 2\epsilon.$$

As ϵ is arbitrary, we obtain $\mathcal{T}_\infty(\bar{\sigma}, \bar{\sigma}) \leqslant 0$, and consequently $\mathcal{T}_\infty(\bar{\sigma}, \bar{\sigma}) = 0$ (the reverse inequality following from $\mathcal{T}_\infty = \mathcal{T}_\infty \star \mathcal{T}_\infty$).

Finally, we note that $\mathcal{T}_\infty(\mu, \nu) = \mathcal{T}_\infty(\mu, \sigma_{k_j}) + \mathcal{T}_\infty(\sigma_{k_j}, \nu)$ for all j, so at the $\liminf$, we find

$$\mathcal{T}_\infty(\mu, \nu) \geqslant \mathcal{T}_\infty(\mu, \bar{\sigma}) + \mathcal{T}_\infty(\bar{\sigma}, \nu).$$

The reverse inequality is immediate from $\mathcal{T}_\infty = \mathcal{T}_\infty \star \mathcal{T}_\infty$.

5) First, we observe that $\mathcal{T}_1(\mu, \mu) \geqslant c$ for all μ. This follows from

$$c = \lim_{t\to\infty} \min_{\mu,\nu} \frac{\mathcal{T}_t(\mu, \nu)}{t} = \lim_{n\to\infty} \min_{\mu,\nu} \frac{\mathcal{T}_n(\mu, \nu)}{n} \leqslant \mathcal{T}_1(\mu, \mu).$$

To achieve the reverse inequality, we construct inductively 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 $\mathcal{T}_1$ and $\mathcal{T}_\infty$. Therefore, the Cesaro averages belong to $\mathcal{D}$,

$$(\frac{1}{n}\sum_{k=1}^n \mu_k, \frac{1}{n}\sum_{k=1}^n \mu_{k+1}) \in \mathcal{D}.$$

Denoting $\nu_n := \frac{1}{n}\sum_{k=1}^n \mu_k$, we have

$$\mathcal{T}_1(\nu_n, \nu_n + \frac{1}{n}(\mu_{n+1} - \mu_1)) + \mathcal{T}_\infty(\nu_n, \nu_n + \frac{1}{n}(\mu_{n+1} - \mu_1)) = c. \tag{207}$$

Extract a weak*-convergent subsequence ν_{n_j} with limit $\bar{\mu} \in \mathcal{A}$. Then by weak-$*$ lower semi-continuity of $\mathcal{T}_1$ (resp. $\mathcal{T}_\infty$), (207) yields at the limit,

$$\mathcal{T}_1(\bar{\mu}, \bar{\mu}) = \mathcal{T}_1(\bar{\mu}, \bar{\mu}) + \mathcal{T}_\infty(\bar{\mu}, \bar{\mu}) \leqslant c.$$

Hence, $c = \mathcal{T}_1(\bar{\mu}, \bar{\mu})$, and $(\bar{\mu}, \bar{\mu}) \in \mathcal{D}$.

Conversely, if μ is a measure which realises $c = \mathcal{T}_1(\mu, \mu)$, then by Property 3 and 4, we have

$$0 \leqslant \mathcal{T}_\infty(\mu, \mu) \leqslant \liminf_{t\to\infty}(\mathcal{T}_t(\mu, \mu) - ct) \leqslant \liminf_{n\to\infty}(\mathcal{T}_n(\mu, \mu) - cn) \leqslant 0,$$

so $\mu \in \mathcal{A}$.

Similar results hold with appropriate changes for forward linear transfers.

8.2 Optimal transports corresponding to a semi-group of cost functionals

We now identify the effective transfer and Kantorovich map associated to a semi-group of linear transfers given by mass transports.

Proposition 8.6. *Suppose $c_t(x, y)$ is a semi-group of equicontinuous cost functions on a compact space $X \times X$, that is*

$$c_{t+s}(x, y) = c_t \star c_s(x, y) := \inf\{c_t(x, z) + c_s(z, y); z \in X\}, \tag{208}$$

and consider the associated optimal mass transports

$$\mathcal{T}_t(\mu, \nu) = \inf\{\int_{X\times X} c_t(x, y)\, d\pi(x, y)\, ;\, \pi \in \mathcal{K}(\mu, \nu)\}. \tag{209}$$

1. *The family $(\mathcal{T}_t)_t$ then forms a semi-group of linear transfers for the convolution operation i.e., $\mathcal{T}_{t+s} = \mathcal{T}_t \star \mathcal{T}_s$ for any $s, t \geqslant 0$ that is equicontinuous on $\mathcal{P}(X) \times \mathcal{P}(X)$, hence one can associate its effective transfer $\mathcal{T}_\infty$ and the corresponding Kantorovich operator T_∞.*

2. *The following holds for the constant c defined in the previous section Theorem 8.5:*

$$c = \inf\{\mathcal{T}_1(\mu, \mu); \mu \in \mathcal{P}(X)\} = \min\{\int_{X\times X} c_1(x, y)\, d\pi; \pi \in \mathcal{P}(X \times X), \pi_1 = \pi_2\} \tag{210}$$

3. *Letting $c_\infty(x, y) := \liminf_{t\to\infty}(c_t(x, y) - ct)$, then :*

$$\mathcal{T}_\infty(\mu, \nu) = \mathcal{T}_{c_\infty}(\mu, \nu) := \inf\{\int_{X\times X} c_\infty(x, y)\, d\pi(x, y)\, ;\, \pi \in \mathcal{K}(\mu, \nu)\}, \tag{211}$$

$$T^-_\infty f(x) = \sup\{f(y) - c_\infty(x, y)\, ;\, y \in X\} \text{ and } T^+_\infty f(y) = \inf\{f(x) + c_\infty(x, y)\, ;\, x \in X\}$$

4. *The set $\mathcal{A} := \{\sigma \in \mathcal{P}(X); \mathcal{T}_\infty(\sigma, \sigma) = 0\}$ consists of those $\sigma \in \mathcal{P}(X)$ supported on the set $A = \{x \in X; c_\infty(x, x) = 0\}$.*

5. *The minimizing measures in (210) are all supported on the set*

$$D := \{(x, y) \in X \times X\, ;\, c_1(x, y) + c_\infty(y, x) = c\}.$$

Proof: The Kantorovich operator for $\mathcal{T}_t$ is given by $T_t f(x) = \sup\{f(y) - c_t(x,y)\,;\, y \in X\}$ and as shown in Proposition 6.6, we have $\mathcal{T}_{s+t} = \mathcal{T}_{c_t \star c_s} = \mathcal{T}_{c_t} \star \mathcal{T}_{c_s} = \mathcal{T}_t \star \mathcal{T}_s$, and $T_{t+s} = T_t \circ T_s$ for every s, t. It remains to show that the effective Kantorovich map T_∞ associated to $(T_t)_t$ is equal to $T_{c_\infty} f := \sup\{f(y) - c_\infty(x,y)\,;\, y \in X\}$. For that, we first note that

$$\limsup_t (T_t f(x) + ct) \geqslant \sup_y \{f(y) - c_\infty(x,y)\} = T_{c_\infty} f(x). \tag{212}$$

On the other hand, let y_n achieve the supremum for $T_n f(x) = \sup\{f(y) - c_n(x,y)\,;\, y \in X\}$, and let $(n_j)_j$ be a subsequence such that $\lim_{j\to\infty}(T_{n_j} f(x) + cn_j) = \limsup_n (T_n f(x) + cn)$. By refining to a further subsequence, we may assume by compactness of X, that $y_{n_j} \to \bar{y}$ as $j \to \infty$. Then by equi-continuity of the c_n's, we deduce that

$$\limsup_n (T_n f(x) + cn) = \lim_{j\to\infty} (T_{n_j} f(x) + cn_j) = f(\bar{y}) - \liminf_j (c_{n_j}(x,\bar{y}) - cn_j). \tag{213}$$

As $\liminf_j (c_{n_j}(x,\bar{y}) - cn_j) \geqslant \liminf_n (c_n(x,\bar{y}) - cn) = c_\infty(x,\bar{y})$, we obtain

$$\limsup_n (T_n f(x) + nc) \leqslant f(\bar{y}) - c_\infty(x,\bar{y}) \leqslant \sup_y \{f(y) - c_\infty(x,y)\} = T_{c_\infty} f(x). \tag{214}$$

The inequality (214) is true for every sequence $(n_k)_k$ going to ∞, so we deduce that $\limsup_t (T_t f(x) + ct) \leqslant T_{c_\infty} f(x)$, and hence combining this with (212) gives equality: $\limsup_t (T_t f(x) + ct) = T_{c_\infty} f(x)$.

Finally, we note that $T_s(\limsup_t (T_f f + ct))(x) + cs = T_s T_{c_\infty} f(x) + cs = T_{c_\infty} f(x)$ thanks to the fact that $c_s \star c_\infty = c_\infty$. This implies from the definition of T_∞ as the limit as $s \to \infty$ (see Thereom 8.5) that $T_\infty f(x) = T_{c_\infty} f(x)$.

Properties (1), (2) and (3) follow then immediately. Properties (4) and (5) now follow from an adaptation of the results of Bernard-Buffoni [6].

8.3 Fathi-Mather weak KAM theory

Let L be a time-independent *Tonelli Lagrangian* on a compact Riemanian manifold M, and consider $\mathcal{T}_t$ to be the cost minimizing transport

$$\mathcal{T}_t(\mu,\nu) = \inf\{\int_{M\times M} c_t(x,y)\,\mathrm{d}\pi(x,y)\,;\, \pi \in \mathcal{K}(\mu,\nu)\},$$

where

$$c_t(x,y) := \inf\{\int_0^t L(\gamma(s),\dot{\gamma}(s))\,\mathrm{d}s\,;\, \gamma \in C^1([0,t];M); \gamma(0) = x, \gamma(t) = y\}.$$

As mentioned in the introduction, the Lax-Oleinik semi-group S_t^-, $t > 0$ is defined by the formula

$$S_t^- u(x) := \inf\{u(\gamma(0)) + \int_0^t L(\gamma(s),\dot{\gamma}(s))\,\mathrm{d}s\,;\, \gamma \in C^1([0,t];M), \gamma(t) = x\},$$

and a function $u \in C(M)$ is said to be a *negative weak KAM solution* if $S_t^- u - ct = u$ for all $t \geqslant 0$.

Another semigroup S_t^+ is defined in terms of S_t^- via the formula $S_t^+ u = -\hat{S}_t^-(-u)$, where $\hat{S}_t^-$ is the Lax-Oleinik semi-group of the Lagrangian $\hat{L}(x,v) := L(x,-v)$. It turns out that

$$S_t^+ u(x) = \sup\{u(\gamma(t)) - \int_0^t L(\gamma(s),\dot{\gamma}(s))\,\mathrm{d}s\,;\,\gamma \in C^1([0,t];M), \gamma(0)=x\}.$$

Analogous to the negative weak KAM solutions, *positive weak KAM solutions* are those u satisfying $S_t^+ u + ct = u$ for all $t \geqslant 0$. The semi-groups S_t^- and S_t^+ are intimately connected with Hamilton-Jacobi equations, and Aubry-Mather theory.

Theorem 8.7. *Under the above conditions on L, there exists a unique constant $c \in \mathbb{R}$ such that the following hold:*

1. (Fathi [25]) *There exists a function $u_- : M \to \mathbb{R}$ (resp. u_+) such that $S_t^- u_- - ct = u_-$ (resp. $S_t^+ u_- + ct = u_-$) for each $t \geqslant 0$.*

2. (Bernard-Buffoni [6]) *Let $c_\infty(x,y) := \liminf_{t\to\infty} c_t(x,y)$ denote the Peierls barrier function. The following duality then holds:*

$$\inf\{\int_{M\times M} c_\infty(x,y)\,\mathrm{d}\pi(x,y)\,;\,\pi\in\mathcal{K}(\mu,\nu)\} = \sup_{u_+,u_-}\{\int_M u_+\,\mathrm{d}\nu - \int_M u_-\,\mathrm{d}\mu\},$$

where the supremum ranges over all $u_+, u_- \in C(M)$ such that u_+ (resp. u_-) is a positive (resp. negative) weak KAM solution, and such that $u_+ = u_-$ on the set $\mathcal{A} := \{x \in M\,;\, c_\infty(x,x) = 0\}$. Moreover, $c_\infty(x,y) = \min_{z\in\mathcal{A}}\{c_\infty(x,z) + c_\infty(z,y)\}$.

3. (Bernard-Buffoni [7]) *The constant c satisfies*

$$c = \min_\pi \int_{M\times M} c_1(x,y)\,\mathrm{d}\pi(x,y),$$

where the minimum is taken over all $\pi \in \mathcal{P}(M\times M)$ with equal first and second marginals. The minimizing measures are all supported on $\mathcal{D} := \{(x,y) \in M\times M\,;\, c_1(x,y) + c_\infty(y,x) = c\}$.

4. (Mather [49]) *The constant $c = \inf_m \int_{TM} L(x,v)\,\mathrm{d}m(x,v)$ where the infimum is taken over all measures $m \in \mathcal{P}(TM)$ which are invariant under the Euler-Lagrange flow (generated by L).*

5. (Fathi [25]) *A continuous function $u : M \to \mathbb{R}$ is a viscosity solution of $H(x,\nabla u(x)) = c[0]$ if and only if it is Lipschitz and u is a negative weak KAM solution (i.e. $T_t^- u + c[0]t = u$). In particular, the statement is false if $c[0]$ is replaced with any other constant c.*

In the language of transfers, the cost-minimizing transport is both a forward and backward linear transfer, with forward (resp. backward) Kantorovich operators given by $T_t^+ f(x) = V_f(t,x)$ and $T_t^- g(y) = W_g^t(0,y)$, where

$$V_f(t',x) = \inf\{f(\gamma(0)) + \int_0^{t'} L(\gamma(s),\dot{\gamma}(s))\,\mathrm{d}s\,;\,\gamma\in C^1([0,t'),M), \gamma(t')=x\}$$

and $W_g^t(t', y)$ the value functional

$$W_g^t(t', y) = \sup\{g(\gamma(t')) - \int_{t'}^{t} L(\gamma(s), \dot\gamma(s))\,\mathrm{d}s\,;\, \gamma \in C^1([0,t'), M), \gamma(0) = x\}.$$

Observe that $V_f(t,x) = S^- f(x)$, while $W_g^t(0,y) = S^+ g(y)$. Hence (with unfortunate signs), $T_t^+ f = S_t^- f(x)$, while $T_t^- f(x) = S_t^+ f(x)$. Note also the translation of terminology in this setting: Our backward weak KAM solutions are Fathi's positive weak KAM solutions, while the analogous forward weak KAM solutions are Fathi's negative weak KAM solutions.

One can proceed with the construction outlined above to construct the negative (resp. positive) weak KAM solutions as the image of the Kantorovich operators T_∞^+ (resp. T_∞^-), and they will be given by

$$T_\infty^- f(x) = \sup\{f(y) - c_\infty(x,y)\,;\, y \in M\} \quad \text{and} \quad T_\infty^+ f(y) = \inf\{f(x) + c_\infty(x,y)\,;\, x \in M\}$$

where $c_\infty(x,y) := \liminf_{t\to\infty} c_t(x,y)$.

The backward (resp. forward) generalised Peierls barrier associated to T_∞^- (resp. T_∞^+) are the same and is the cost-minimizing transport with cost c_∞, which by duality we can write as

$$\inf\{\int_{M\times M} c_\infty(x,y)\,\mathrm{d}\pi(x,y)\,;\, \pi \in \mathcal{K}(\mu,\nu)\} = \sup\{\int_M T_\infty^+ f\,\mathrm{d}\nu - \int_M T_\infty^- \circ T_\infty^+ f\,\mathrm{d}\mu\,;\, f \in C(M)\}.$$

It can be checked this is exactly the statement 2 in the above theorem.

8.4 The Schrödinger semigroup

Recall the Schrödinger bridge of Example 4.5. Let M be a compact Riemannian manifold and fix some reference non-negative measure R on path space $\Omega = C([0,\infty], M)$. Let $(X_t)_t$ be a random process on M whose law is R, and denote by R_{0t} the joint law of the initial position X_0 and the position X_t at time t, that is $R_{0t} = (X_0, X_t)_\# R$. Assume R is the reversible Kolmogorov continuous Markov process associated with the generator $\frac{1}{2}(\Delta - \nabla V \cdot \nabla)$ and the initial probability measure $m = e^{-V(x)}dx$ for some function V.

For probability measures μ and ν on M, define

$$\mathcal{T}_t(\mu,\nu) := \inf\{\int_M \mathcal{H}(r_t^x, \pi_x) d\mu(x)\,;\, \pi \in \mathcal{K}(\mu,\nu),\, d\pi(x,y) = d\mu(x) d\pi_x(y)\} \tag{215}$$

where $dR_{0t}(x,y) = dm(x) dr_t^x(y)$ is the disintegration of R_{0t} with respect to its initial measure m.

Proposition 8.8. *The collection $\{\mathcal{T}_t\}_{t\geqslant 0}$ is a semigroup of backward linear transfers with Kantorovich operators $T_t f(x) := \log S_t e^f(x)$ where $(S_t)_t$ is the semi-group associated to R; in particular,*

$$\mathcal{T}_t(\mu,\nu) = \sup\left\{\int_M f d\nu - \int_M \log S_t e^f d\mu\,;\, f \in C(M)\right\}. \tag{216}$$

The corresponding idempotent backward linear transfer is $\mathcal{T}_\infty(\mu,\nu) = \mathcal{H}(m,\nu)$, and its effective Kantorovich map is $T_\infty f(x) := \log S_\infty e^f$, where $S_\infty g := \int g\,\mathrm{d}m$.

Proof: It is easy to see that for each t, T_t is monotone, 1-Lipschitz and convex, and also satisfies $T_t(f+c) = T_t f + c$ for any constant c. It follows that $\mathcal{T}^*_{t,\mu}(f) = \int_M T_t f \, d\mu$ for each t by Proposition **??**. The semigroup property then follows from the semigroup $(S_t)_t$ and the property that $\mathcal{T}_t \star \mathcal{T}_s$ is a backward linear transfer with Kantorovich operator $T_t \circ T_s f(x) = \log S_t S_s e^f(x) = \log S_{s+t} e^f(x) = T_{t+s} f(x)$ by Proposition 6.3.

Now we remark that it is a standard property of the semigroup $(S_t)_t$ on a compact Riemannian manifold, that under suitable conditions on V, $S_t e^f \to S_\infty e^f$, uniformly on M, as $t \to \infty$, for any $f \in C(M)$. This immediately implies by definition of T_t, that $T_t f \to T_\infty f$ uniformly as $t \to \infty$ for any $f \in C(M)$. We then deduce from the 1-Lipschitz property, that $T_t \circ T_\infty f(x) = T_\infty f(x)$. We conclude that T_∞ is a Kantorovich operator from Theorem 8.5. Finally we see that $\mathcal{T}_\infty(\mu, \nu)$ is

$$\begin{aligned}\mathcal{T}_\infty(\mu,\nu) &:= \sup\{\int f \, \mathrm{d}\nu - \int T_\infty f \, \mathrm{d}\mu \, ; \, f \in C(M)\} \\ &= \sup\{\int f \, \mathrm{d}\nu - \log \int e^f \, \mathrm{d}m \, ; \, f \in C(M)\} \\ &= \mathcal{H}(m,\nu),\end{aligned}$$

(see Section 9, for the last equality).

9 Weak KAM solutions for non-continuous transfers

We now deal with cases where $\mathcal{T}$ is not necessarily weak*-continuous on $\mathcal{M}(X)$.

9.1 The case of non-continuous transfers with bounded oscillation

We now consider situations where $\mathcal{T}$ is not equicontinuous, but there is some control on the oscillation of the transfers $\mathcal{T}^n$.

Lemma 9.1. *Let X be a compact space and let $\mathcal{T}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains the Dirac measures. Assume that*

$$\mathcal{T}(\mu_0, \mu_0) < +\infty \text{ for some } \mu_0 \in \mathcal{P}(X). \tag{217}$$

Then, the following properties hold:

1. $c(\mathcal{T}) := \sup_n \frac{\inf\{\mathcal{T}_n(\mu,\nu)\,;\,\mu,\nu \in \mathcal{P}(X)\}}{n} \leqslant \inf\{\mathcal{T}(\mu,\mu)\,;\,\mu \in \mathcal{P}(X)\} < +\infty.$
2. *For each $f \in C(X)$ and $x \in X$, we have*
$$\limsup_n \frac{T^n f(x)}{n} \leqslant -c(\mathcal{T}). \tag{218}$$
3. *For each $f \in C(X)$ and $\mu \in \mathcal{P}(X)$, we have*
$$\liminf_n \frac{1}{n} \int_X T^n f \, d\mu \geqslant -\mathcal{T}(\mu,\mu). \tag{219}$$

4. For each $n \in N$, we have

$$\sup_{\mu \in \mathcal{P}(X)} \int (T^n f(x) + nc)\, d\mu \geqslant \inf_{y \in X} f(y). \tag{220}$$

5. If for some $K > 0$, we have

$$\liminf_n \{ \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}^n(\mu, \mu) - \inf_{\mu,\nu \in \mathcal{P}(X)} \mathcal{T}^n(\mu, \nu) \} \leqslant K, \tag{221}$$

then,

$$\sup_{x \in X} \liminf_n T^n f(x) + nc \leqslant \sup_{y \in X} f(y) + K. \tag{222}$$

In the next section, we shall prove that actually,

$$c(\mathcal{T}) = \inf\{\mathcal{T}(\mu, \mu)\,;\, \mu \in \mathcal{P}(X)\}.$$

Proof: 1) Let $\mathcal{T}$ be a backward linear transfer and consider for each $n \in \mathbb{N}$, $\mathcal{T}_n = \mathcal{T} \star \mathcal{T} \star .... \star \mathcal{T}$ the backward linear transfer obtained by iterating its convolution n times. The sequence $m_n := \inf\{\mathcal{T}_n(\mu, \nu)\,;\, \mu, \nu \in \mathcal{P}(X)\}$ is superadditive, that is $m_{n+k} \geqslant m_n + m_k$ for all positive integers, n, k. Since

$$\frac{m_n}{n} \leqslant \sup_n \frac{1}{n} \mathcal{T}_n(\mu_0, \mu_0) \leqslant \mathcal{T}(\mu_0, \mu_0) < +\infty,$$

it follows that there exists a number $c(\mathcal{T}) \in \mathbb{R}$ such that

$$\lim_n \frac{m_n}{n} := \lim_n \frac{1}{n} \inf\{\mathcal{T}_n(\mu, \nu)\,;\, \mu, \nu \in \mathcal{P}(X)\} = \sup_n \inf_{\mu,\nu} \frac{\mathcal{T}_n(\mu, \nu)}{n} = c(\mathcal{T}) < +\infty. \tag{223}$$

2) follows from 1) since

$$\begin{aligned} T^n f(x) &= \sup\{\int_X f \; d\sigma - \mathcal{T}_n(x, \sigma)\,;\, \sigma \in \mathcal{P}(X)\} \\ &\leqslant \sup f - \inf\{\mathcal{T}_n(x, \sigma)\,;\, \sigma \in \mathcal{P}(X)\} \\ &\leqslant \sup f - \inf\{\mathcal{T}_n(\mu, \sigma)\,;\, \mu, \sigma \in \mathcal{P}(X)\}. \end{aligned}$$

For 3) note that

$$\begin{aligned} \int_X T^n f \, \mathrm{d}\mu &= \sup\{\int_X f \; d\sigma - \mathcal{T}_n(\mu, \sigma)\,;\, \sigma \in \mathcal{P}(X)\} \\ &\geqslant \int_X f \, d\mu - \mathcal{T}_n(\mu, \mu) \\ &\geqslant \int_X f \, d\mu - n\mathcal{T}(\mu, \mu). \end{aligned}$$

4) Write

$$\begin{aligned} \sup_{\mu \in \mathcal{P}(X)} \int (T^n f(x) + nc)\, d\mu &= \sup_{\mu \in \mathcal{P}(X)} \sup_{\sigma \in \mathcal{P}(X)} \{\int f d\sigma - \mathcal{T}_n(\mu, \sigma) + nc\} \\ &= \inf_X f - \inf_{\sigma,\mu} \mathcal{T}_n(\mu, \sigma) + nc \\ &\geqslant \inf_X f. \end{aligned}$$

The latter inequality follows from Lemma 11.1 since $\inf_{\mu,\sigma}\mathcal{T}_n(\mu,\sigma)\leqslant nc$.
For 5), write

$$\begin{aligned}\sup_{x\in X}\liminf_n T^n f(x)+nc &\leqslant \liminf_n \sup_{x\in X} T^n f(x)+kc\\ &=\liminf_n \sup_{x\in X}\sup_{\sigma\in\mathcal{P}(X)}\{\int_X f d\sigma-\mathcal{T}^n(x,\sigma)+nc\}\\ &\leqslant \liminf_n \sup_{x\in X}\sup_{\sigma\in\mathcal{P}(X)}\{\sup f-\mathcal{T}^k(x,\sigma)+kc\}\\ &=\sup f-\limsup_n \inf_{x\in X}\inf_{\sigma\in\mathcal{P}(X)}\{\mathcal{T}^n(x,\sigma)-nc\}\\ &\leqslant \sup f-\limsup_n \inf_{\mu,\sigma\in\mathcal{P}(X)}\{\mathcal{T}^n(\mu,\sigma)-nc\}\\ &\leqslant \sup f-\limsup_n \inf_{\mu,\sigma\in\mathcal{P}(X)}\{\mathcal{T}^n(\mu,\sigma)-nc\}\\ &\leqslant \sup f+K.\end{aligned}$$

Now we can prove the following.

Theorem 9.2. *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(X)$ such that $\mathcal{D}_1(\mathcal{T})$ contains all Dirac measures. Assume (217), (221) and*

$$\sup_{x\in X}\inf_{\sigma\in\mathcal{P}(X)}\mathcal{T}(x,\sigma)<+\infty. \tag{224}$$

If $T: C(X)\to USC(X)$, where T is the backward Kantorovich operator associated to $\mathcal{T}$, then there exists $h\in USC_\sigma(X)$ such that $Th+c=h$ on X.

Proof: Note that condition (224) means that $T^i f$ is bounded below for any $f\in C(X)$ and any $i\in N$. We distinguish two cases:

Case 1: Assume the following:

$$\text{There is } f\in C(X) \text{ so that } \forall x\in X, \text{ there exists } n\in\mathbb{N} \text{ with } T^n f(x)+nc<f(x). \tag{225}$$

Since $T^n f$ is in $USC(X)$, then for each $x\in X$, there exists $n\in\mathbb{N}$ such that $T^n f+nc<f$ on a neighborhood of x, and since X is compact, there is a finite number r of iterates of T such that $\inf_{0\leqslant i\leqslant r}(T^i f+ic)<f$.
Set $g_r=\inf_{1\leqslant i\leqslant r}(T^i f+ic)$ and note that $g_r\in USC(X)$, $\inf g_r>-\infty$ because of (224), and $g_r<f$. Note now that

$$Tg_r+c\leqslant \inf_{2\leqslant i\leqslant r+1}\{T^i f+ic\}.$$

On the other hand, $Tg_r+c\leqslant Tf+c$, hence

$$Tg_r+c\leqslant \inf_{1\leqslant i\leqslant r}\{T^i f+ic\}=g_r.$$

It follows that the sequence $\{T^n g_r+nc\}_n$ is decreasing to some function $h\in USC(X)$. Note that $h\leqslant g_r$ hence is bounded above.

Now we show that h is proper, that is not identically $-\infty$. Indeed, if it was, then for every x, the sequence $g_n(x)=T^n g_r(x)+nc$ will be decreasing to $-\infty$. It follows that for each $x \in X$, there exists i, such that $g_i(x) < \inf g_r - 1$, hence on a neighborhood of x since g_i is in $USC(X)$. By compactness and since the $(g_n)_n$ is decreasing, we get a function g_N such that $g_N < \inf g_r - 1$ on X. On the other hand, the preceding lemma yields that $\sup_{\mu\in\mathcal{P}(X)} \int g_N\, d\mu \geqslant \inf g_r$. It follows that

$$\inf g_r \leqslant \sup_{\mu\in\mathcal{P}(X)} \int g_N\, d\mu \leqslant \inf g_r - 1,$$

which is a contradiction, hence h is proper.

Finally, note that by Lemma 4.2, we have

$$Th + c = T(\lim_n T^n g_r + nc) + c = \lim_n T^{n+1} g_r + (n+1)c = h.$$

Case 2: We now assume that for any $f \in C(X)$, there exists $x \in X$ such that

$$T^n f(x) + nc \geqslant f(x) \quad \text{for all } n \in \mathbb{N}. \tag{226}$$

We now consider for each $f \in C(X)$, the function $\tilde{f} := \liminf_n T^n f + nc$. It is clear that $\tilde{f} \in USC_\sigma(X)$, and by our assumption, there exists $x \in X$ such that $\tilde{f}(x) \geqslant f(x) > -\infty$, and hence it is proper. On the other, we have by Lemma 9.1, that $\sup_{x\in X} \tilde{f} \leqslant \sup_{x\in X} f(x) + K$. Moreover, by Lemma 4.2,

$$T\tilde{f} + c = T(\liminf_n T^n f + nc) + c \geqslant \liminf_n T^{n+1} f + (n+1)c = \tilde{f}.$$

It follows that the sequence $\{T^n \tilde{f} + nc\}_n$ is increasing to a function $h \in USC_\sigma(X)$. Note that $h \geqslant \tilde{f}$, hence it is proper. On the other hand, by Lemma 9.1, we have $h \leqslant \sup \tilde{f} \leqslant \sup f + K < +\infty$ and we are done.

Corollary 9.3. *Let X is a compact space and let $\mathcal{T}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains the Dirac measures. If $\mathcal{T}$ is bounded above on $\mathcal{P}(X) \times \mathcal{P}(X)$, then*

$$\frac{\mathcal{T}_n(\mu,\nu)}{n} \to c \quad \textit{uniformly on } \mathcal{P}(X) \times \mathcal{P}(X). \tag{227}$$

Moreover, there exists an idempotent operator $T_\infty : C(X) \to USC_\sigma(X)$ such that for each $f \in C(X)$, $T_\infty f$ is a backward weak KAM solution.

Proof: Note that conditions (217) and (224) are readily satisfied. To prove (221), one can easily see that for any $\mu, \nu \in \mathcal{P}(X)$,

$$\inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n + 2 \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T} \leqslant \mathcal{T}_{n+2}(\mu,\nu) \leqslant 2 \sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T} + \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n,$$

from which follows that

$$\begin{aligned}\sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_{n+2} - \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_{n+2} &\leqslant 2 \sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T} + \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n - \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n - 2 \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T} \\ &= 2 \sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T} - 2 \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T} \\ &=: K < \infty.\end{aligned}$$

Theorem 9.2 then applies to get a weak KAM solution h in $USC_\sigma(X)$.

Note now that since h is bounded above and is proper, i.e., $h(x_0) > -\infty$ for some $x_0 \in X$, we have

$$h(x_0) - \sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n \leqslant T_n h(y) = \sup_{\sigma\in\mathcal{P}}\{\int_X h\, d\sigma - \mathcal{T}_n(y,\sigma)\} \leqslant \sup_X h - \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n,$$

hence

$$\inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n - \sup_X h \leqslant nc - h(x_0) \leqslant \sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n - h(x_0) \leqslant \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n + K - h(x_0),$$

and

$$-K \leqslant \mathcal{T}_n(\mu,\nu) - nc \leqslant K + \sup_X h - h(x_0), \tag{228}$$

from which follows that

$$\frac{\mathcal{T}_n(\mu,\nu)}{n} \to c \quad \text{uniformly on } \mathcal{P}(X)\times\mathcal{P}(X). \tag{229}$$

Note now that (228) yields that for every $f \in C(X)$, there is $C > 0$ such that

$$\|T^n f + nc\|_\infty \leqslant \|f\|_\infty + C, \tag{230}$$

from which follows that $\hat{T}f := \liminf_n T^n f + nc$ is bounded, belongs to $USC(X)$ and satisfies $T(\hat{T}f) + c \geqslant \hat{T}f$. The sequence $(T^n(\hat{T}f) + nc)_n$ is therefore increasing to a function $T_\infty f$ in $USC_\sigma(X)$ such that $T \circ T_\infty f + c = T_\infty f$. □

Here is another situation where we can obtain weak KAM solutions. It will be relevant for the stochastic Mather theory.

Proposition 9.4. *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(X)$ such that $\mathcal{D}_1(\mathcal{T})$ contains all Dirac measures and that (217), (224) hold. If there exists $u, v \in USC(X)$ that are bounded below such that*

$$T^n u + nv = u \quad \text{for all } n \in \mathbb{N}, \tag{231}$$

then there exists $h \in USC(X)$ such that $Th + c = h$ on X, where T is the backward Kantorovich operator associated to $\mathcal{T}$.

Proof: Note that (231) and Lemma 9.1 yield that necessarily $-v(x) \leqslant -c$, from which follows that

$$T^n u + nc \leqslant u \quad \text{for all } n \in \mathbb{N}.$$

Applying T^m and using the linearity of T^m with respect to constants, we find $T^{m+n}u + cn \leqslant T^m u$, and hence

$$T^{m+n}u + c(m+n) \leqslant T^m u + cm$$

So $n \mapsto T^n u + cn$ is decreasing. The same reasoning as in Case (1) of the proof of Theorem 9.2 yields that $(T^n u + cn)_n$ decreases to a proper function $h \in USC(X)$ such that $Th + c = h$ on X.

9.2 Weak KAM solutions associated to non-continuous optimal mass transports

The following extends a result established by Bernard-Buffoni [7] in the case where the cost function $c(x, y)$ is continuous.

Corollary 9.5. *Let c be a bounded lower semi-continuous cost functional on a compact space X and consider the associated optimal mass transport*

$$\mathcal{T}(\mu, \nu) = \inf\{\int_{X\times X} c(x, y)\, \mathrm{d}\pi(x, y)\,;\, \pi \in \mathcal{K}(\mu, \nu)\}. \tag{232}$$

Let $c_\infty(x, y) := \liminf_{n\to\infty} c_n(x, y)$, where for each $n \in \mathbb{N}$,

$$c_n(y, x) = \inf\left\{c(y, x_1) + c(x_1, x_2) .... + c(x_{n-1}, x);\ x_1, x_2, x_{n-1} \in X\right\}.$$

Then, the following hold:

1. *The corresponding effective transfer is given by*

$$\mathcal{T}_\infty(\mu, \nu) = \mathcal{T}_{c_\infty}(\mu, \nu) := \inf\{\int_{X\times X} c_\infty(x, y)\, \mathrm{d}\pi(x, y)\,;\, \pi \in \mathcal{K}(\mu, \nu)\}, \tag{233}$$

   and the associated effective Kantorovich maps are given by

$$T^-_\infty f(x) = \sup\{f(y) - c_\infty(x, y)\,;\, y \in X\} \text{ and } T^+_\infty f(y) = \inf\{f(x) + c_\infty(x, y)\,;\, x \in X\}. \tag{234}$$

2. *The following also holds*

$$c(\mathcal{T}) = \inf\{\mathcal{T}(\mu, \mu); \mu \in \mathcal{P}(X)\} = \min_{\mu\in\mathcal{P}(X)} \int_{X\times X} c(x, y)\, \mathrm{d}\pi(x, y); \pi \in \mathcal{K}(\mu, \nu)\}.. \tag{235}$$

3. *The set $\mathcal{A} := \{\sigma \in \mathcal{P}(X); \mathcal{T}_\infty(\sigma, \sigma) = 0\}$ consists of those $\sigma \in \mathcal{P}(X)$ supported on the set $A = \{x \in X; c_\infty(x, x) = 0\}$.*

4. *The minimizing measures in (235) are all supported on the set*

$$D := \{(x, y) \in X \times X\,;\, c(x, y) + c_\infty(y, x) = c(\mathcal{T})\}.$$

Example 7.1: Iterates of power costs: Let $c_p(x, y) = |x - y|^p$ for $p > 0$, then, $c_p \star c_p(x, y) = \inf\{|x - z|^p + |z - y|^p\,;\, z \in X\}$ is minimised at some point $z = (1 - \lambda)x + \lambda y$ on the line between x and y, so that $c_p \star c_p(x, y) = (\lambda^p + (1 - \lambda)^p)\,|x - y|^p$. For $p > 1$, the optimal λ is $\frac{1}{2}$. Hence, by considering $\mathcal{T}_p$ to be the optimal mass transport associated to c_p with its corresponding Kantorovich operator, $T_p f(x) = \sup\{f(y) - c_p(x, y)\,;\, y \in X\}$, we then have

$$(T_p)^n f(x) = \sup\{f(y) - \frac{1}{n^{p-1}}|x - y|^p\}.$$

Hence when $n \to \infty$, $(T_p)^n f(x) \to \sup_x f(x) = T_\infty f(x)$ and $c(\mathcal{T}_p) = 0$.

10 Linear transfers and ergodic optimization

This section was developed jointly with Dorian Martino [46]. We shall consider here linear transfers where the associated Kantorovich maps are affine operators that is of the form $T^- f(x) = Tf(x) - A(x)$, where T is a Markov operator and A is a given function (observable). We have already noted that if $A \equiv 0$, then the general Aubry theory reduces to standard ergodic theory. In this section, we shall see that the presence of A allows the theory of transfers to incorporate ergodic optimization for expanding dynamical systems. For simplicity, we shall focus here on the case where the linear Markov operator is given by a point transformation σ.

Proposition 10.1. *Let $\sigma : X \to X$ a continuous onto map on a compact space X, and assume there is a compact space Y such that for each $y \in Y$, there exists a compact subset X_y of X and a continuous map $\tau_y : X_y \to X$ such that $\sigma \circ \tau_y(x) = x$ for all $x \in X_y$. Let $A \in C(Y \times X)$ be a continuous function and consider the lower semi-continuous cost function $c : X \times X \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$c(z, x) := \begin{cases} \inf\{A(y, x)\,;\, y \in Y_x, \tau_y(x) = z\} & \text{if } \sigma(z) = x \\ +\infty & \text{otherwise,} \end{cases}$$

where for each $x \in X$, $Y_x := \{y \in Y\,;\, x \in X_y\}$. Assume that $x \mapsto \bar{A}(x) := c(x, \sigma(x))$ is continuous and $\sup_{x \in X} \bar{A}(x) < +\infty$. Then,

1. *The optimal mass transport $\mathcal{T}$ associated to the cost c has a backward (resp. forward) Kantorovich operator given by*

$$T^- g(x) = g(\sigma(x)) - \bar{A}(x), \quad (\textit{resp.}, \quad T^+ f(x) = \inf_{y \in Y_x} \{f(\tau_y(x)) + A(y, x)\}).$$

2. *The following duality formulae holds:*

$$\begin{aligned} c(\mathcal{T}) := \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu) &= \inf\{\int_{\hat{X}} A(y, x)\, \mathrm{d}\hat{\mu}(y, x)\,;\, \hat{\mu} \in \mathcal{M}_0\} \\ &= \inf\{\int_X \bar{A}(x)\, d\mu(x); \mu \in \mathcal{P}_\sigma(X)\} \\ &= \sup_{f \in C(X)} \inf_{x \in X} \{f(x) - f(\sigma(x)) + \bar{A}(x)\} \\ &= \sup_{f \in C(X)} \inf_{x \in X} \inf_{y \in Y_x} \{f(\tau_y(x)) - f(x) + A(y, x)\}. \end{aligned}$$

3. *Moreover, there exists $h \in USC_\sigma(X)$ such that*

$$h(\sigma(x)) - \bar{A}(x) + c(\mathcal{T}) = h(x) \quad \text{for all } x \in X, \tag{236}$$

equivalently,

$$\inf_{y \in Y_x} \{h(\tau_y(x)) + A(y, x)\} - c(\mathcal{T}) = h(x) \quad \text{for all } x \in X. \tag{237}$$

Remark 10.2. *The assumption that σ is surjective ensures that c is lower semi-continuous.*

PROOF. (1) is straightforward, For (2) note first that the dualities

$$\inf\{\int_X \bar{A}(x)\,d\mu(x); \mu \in \mathcal{P}_\sigma(X)\} = \sup_{f\in C(X)} \inf_{x\in X} \{f(x) - f(\sigma(x)) + \bar{A}(x)\}$$

and

$$\inf\{\int_{\hat{X}} A(y,x)\,\mathrm{d}\hat{\mu}(y,x)\,;\, \hat{\mu} \in \mathcal{M}_0\} = \sup_{f\in C(X)} \inf_{x\in X} \inf_{y\in Y_x} \{f(\tau_y(x)) - f(x) + A(y,x)\}$$

are established by a standard application of the Fenchel-Rockafeller duality formula. Indeed, for the first, let $h_1, h_2 : C(X) \to \mathbb{R}$ be defined by $h_1(\varphi) = \sup_x \varphi(x) - \bar{A}(x)$ and

$$h_2(\varphi) = \begin{cases} 0 & \text{if } \varphi \text{ is in the closure of } \{f \circ \sigma - f; f \in C(X)\} \\ -\infty & \text{otherwise} \end{cases}$$

then their respective Legendre transforms are given by

$$h_1^*(\mu) = \begin{cases} \int_X \bar{A}\,d\mu & \text{if } \mu \in \mathcal{P}(X) \\ +\infty & \text{otherwise} \end{cases}$$

and

$$h_2^*(\mu) = \begin{cases} 0 & \text{if } \mu \in \mathcal{P}_\sigma(X) \\ -\infty & \text{otherwise} \end{cases}$$

where $h_1^*(\mu) := \sup_{\varphi\in C(X}\{\int_X \varphi(x)\,\mathrm{d}\mu(x) - h_1(\varphi)\}$ and $h_2^*(\mu) := \inf_{\varphi\in C(X)}\{\int_X \varphi(x)\,\mathrm{d}\mu(x) - h_2(\varphi)\}$. It now suffices to apply the formula

$$\inf_{\mu\in\mathcal{M}(X)} \{h_1^*(\mu) - h_2^*(\mu)\} = \sup_{\varphi\in C(X)} \{h_2(\varphi) - h_1(\varphi)\}.$$

Similarly, let $h_1, h_2 : C(X) \to \mathbb{R}$ be defined by $h_1(\varphi) = \sup_{(y,x)\in\hat{X}} \varphi(y,x) - A(y,x)$, where $\hat{X} := \{(y,x)\,;\, x \in X, y \in X_y\}$ and

$$h_2(\varphi) = \begin{cases} 0 & \text{if } \varphi \text{ is in the closure of } \{f - f \circ \tau_y; f \in C(X)\} \\ -\infty & \text{otherwise.} \end{cases}$$

Their respective Legendre transforms are then given by

$$h_1^*(\hat{\mu}) = \begin{cases} \int_{\hat{X}} A\,\mathrm{d}\hat{\mu} & \text{if } \hat{\mu} \in \mathcal{P}(\hat{X}) \\ +\infty & \text{otherwise,} \end{cases}$$

and

$$h_2^*(\hat{\mu}) = \begin{cases} 0 & \text{if } \hat{\mu} \in \mathcal{M}_0 \\ -\infty & \text{otherwise,} \end{cases}$$

where $\mathcal{M}_0 := \Big\{\mu \in \mathcal{P}(\hat{X}) \mid \int_{\hat{X}} f(\tau_y(x)) - f(x)\; d\mu(y,x) = 0\Big\}$. It now suffices to apply again the formula $\inf\limits_{\hat{\mu}\in\mathcal{M}(\hat{X})} \{h_1^*(\hat{\mu}) - h_2^*(\hat{\mu})\} = \sup\limits_{\varphi\in C(\hat{X})} \{h_2(\varphi) - h_1(\varphi)\}$.

To equate the two duality statements, we observe by definition of $\bar{A}$ that

$$\begin{aligned}\inf_{x\in X}\{f(x) - f(\sigma(x)) + \bar{A}(x)\} &= \inf_{x\in X} \inf_{y\in Y_{\sigma(x)}, \tau_y(\sigma(x))=x} \{f(x) - f(\sigma(x)) + A(y,\sigma(x))\} \\ &= \inf_{z\in X} \inf_{y\in Y_z} \{f(\tau_y(z)) - f(z) + A(y,z)\}\end{aligned}$$

where the last equality holds by making the change of variable $z := \sigma(x)$, along with the fact that σ is assumed to be surjective.
To establish further equality with $\inf_{\mu\in\mathcal{P}_\sigma(X)} \mathcal{T}(\mu,\mu)$, we just note that the optimal plan for $\mathcal{T}(\mu,\mu)$ must be supported on the graph of σ since c is finite only on these points. Hence $\pi = (id\times\sigma)\#\mu$, so that $\mathcal{T}(\mu,\mu) = \int_X c(x,\sigma(x))\,\mathrm{d}\mu(x) = \int_X \bar{A}\,\mathrm{d}\mu$.

(3) To establish the existence of a function satisfying (236), note that this is equivalent to having a function h such that $T^- h(x) + c(\mathcal{T}) = h(x)$ and $T^+ h(x) - c(\mathcal{T}) = h(x)$, hence it suffices to show that the assumptions of Theorem 9.2 are satisfied (see also Remark 10.3). For that, first note that by a theorem of Bogolyubov and Krylov, σ has an invariant measure $\bar{\mu}$, hence $\mathcal{T}(\bar{\mu},\bar{\mu}) < +\infty$ and condition (217) is satisfied. On the other hand, we have for each $x\in X$,

$$\sup_{x\in X}\inf_{\nu\in\mathcal{P}(\Sigma)} \mathcal{T}(\delta_x,\nu) \leqslant \sup_{x\in X}\bar{A}(x) < +\infty,$$

hence Condition (221) is also satisfied. Finally, in order to show (224), we let for each $n\in\mathbb{N}$, $\mu_n\in\mathcal{P}(X)$ be such that $\mathcal{T}_n(\mu_n,(\sigma^n)_\sharp\mu_n) = \inf_{\mu,\nu}\mathcal{T}_n(\mu,\nu)$. Up to extraction, we can assume that $\left(\frac{1}{n}\sum_{k=0}^{n-1}(\sigma^k)_\sharp\mu_n\right)_{n\in\mathbb{N}}$ converges to some $\bar{\mu}\in\mathcal{P}(X)$. Then $\bar{\mu}$ is σ-invariant. Indeed,

$$\sigma_\sharp\bar{\mu} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\sigma^k)_\sharp\mu_n = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=0}^{n-1}(\sigma^k)_\sharp\mu_n + \frac{1}{n}((\sigma^n)_\sharp\mu_n - \mu_n)\right) = \bar{\mu}.$$

Up to extraction again, one can assume that there exists $K>0$ such that

$$\forall n\geqslant 1,\quad \left|\int_X \bar{A}\,\mathrm{d}\bar{\mu} - \frac{1}{n}\sum_{k=0}^{n-1}\int_X \bar{A}\,\mathrm{d}((\sigma^k)_\sharp\mu_n)\right| \leqslant \frac{K}{n}.$$

Finally, we obtain

$$\begin{aligned}
&\liminf_{n\to\infty}\left(\inf_{\mu\in\mathcal{P}(X)}\mathcal{T}_n(\mu,\mu) - \inf_{\mu,\nu\in\mathcal{P}(X)}\mathcal{T}_n(\mu,\nu)\right)\\
&\quad\leqslant \liminf_{n\to\infty}\mathcal{T}_n(\bar{\mu},\bar{\mu}) - \mathcal{T}_n(\mu_n,(\sigma^n)_\sharp\mu_n)\\
&\quad\leqslant \liminf_{n\to\infty}\sum_{k=0}^{n-1}\mathcal{T}(\bar{\mu},\bar{\mu}) - \mathcal{T}((\sigma^k)_\sharp\mu_n,(\sigma^{k+1})_\sharp\mu_n)\\
&\quad\leqslant \liminf_{n\to\infty} n\left(\int_X \bar{A}\,\mathrm{d}\bar{\mu} - \frac{1}{n}\sum_{k=0}^{n-1}\int_X \bar{A}\,\mathrm{d}((\sigma^k)_\sharp\mu_n)\right)\\
&\quad\leqslant K.
\end{aligned}$$

Theorem 9.2 now applies to get the existence of $h\in USC_\sigma(X)$ such that $T^-h + c(\mathcal{T}) = h$.

Remark 10.3. *In fact, for any $g\in C(X)$,*

$$T^-_\infty g(x) := \limsup_{n\to\infty}\{g(\sigma^n(x)) + n(c(\mathcal{T}) - \bar{A}(x))\} = \begin{cases}\limsup_{n\to\infty} g(\sigma^n(x)), & \text{if } \bar{A}(x) = c(\mathcal{T})\\ +\infty, & \text{if } \bar{A}(x) > c(\mathcal{T})\\ -\infty, & \text{if } \bar{A}(x) < c(\mathcal{T})\end{cases}$$

solves (236).

10.1 Ergodic optimization in the deterministic holonomic setting

Fix $r \in \mathbb{N}$, and let M be an $r \times r$ transition matrix. Denote by

$$\Sigma = \{x \in \{1, ..., r\}^{\mathbb{N}} \mid \forall i \geqslant 0,\ M(x_i, x_{i+1}) = 1\}$$

the set of admissible words, its dual

$$\Sigma^* = \{y \in \{1, ..., r\}^{\mathbb{N}} \mid \forall i \geqslant 0,\ M(y_{i+1}, y_i) = 1\},$$

and consider the space

$$\hat{\Sigma} = \{(y, x) \in \Sigma^* \times \Sigma \mid M(y_0, x_0) = 1\}.$$

For each $x \in \Sigma$, we let $\Sigma^*_x = \{y \in \Sigma^* \mid (y, x) \in \hat{\Sigma}\}$ and assume that $\forall x,\ \Sigma^*_x \neq \emptyset$.
We will denote the words of Σ with their starting letters, i.e., $(x_0, x_1, ...)$ while the words in Σ^* will be identified with their ending letters, i.e., $(..., y_1, y_0)$. We consider Σ and Σ^* as metric spaces with the distance $d(x, \bar{x}) = 2^{-\min\{j \in \mathbb{N};\ x_j \neq \bar{x}_j\}}$. In particular, all these sets are compact.
Consider now the two continuous maps $\sigma : \Sigma \to \Sigma$ and $\tau : \hat{\Sigma} \to \Sigma$ defined as

$$\sigma(x_0, x_1, ...) = (x_1, x_2, ...) \quad \text{and} \quad \tau(y, x) = (y_0, x_0, x_1, ...).$$

We will denote $\tau(y, x)$ by $\tau_y(x)$ and consider the set of holonomic probability measures

$$\mathcal{M}_0 := \left\{\mu \in \mathcal{P}(\hat{\Sigma}) \mid \int_{\hat{\Sigma}} f(\tau_y(x)) - f(x)\ d\mu(y, x) = 0\right\}.$$

An application of the previous proposition yields the following results of E. Garibaldi and A. O. Lopes related to the Aubry-Mather theory for symbolic dynamics [**?**].

Proposition 10.4. *Given $A \in C(\hat{\Sigma})$, then the following hold:*

$$c(A) := \inf_{\mu \in \mathcal{M}_0} \int_{\hat{\Sigma}} A d\mu = \sup_{f \in C(\Sigma)} \inf_{(y,x) \in \hat{\Sigma}} f(\tau_y(x)) - f(x) + A(y, x). \tag{238}$$

Moreover, there exists $h \in USC_\sigma(\Sigma)$ such that

$$\inf_{y \in \Sigma^*_x} h(\tau_y(x)) + A(y, x) - c(A) = h(x) \quad \forall x \in \Sigma. \tag{239}$$

Remark 10.5. *If we iterate T^+, we obtain $\forall k,\ \forall f \in C(\Sigma),\ \forall x \in \Sigma$,*

$$(T^+)^k f(x) = \inf\left\{f(\tau_{y^{k-1}}(x^{k-1})) + \sum_{i=0}^{k-1} A(y^i, x^i) \mid x^0 = x,\ \forall i,\ y^i \in \Sigma^*_{x^i},\ x^{i+1} = \tau_{y^i}(x^i)\right\},$$

*which correspond to the non-regularized Mané functional S^ϵ_A Garibaldi and Lopes [**?**] where we limit the number of steps to k and with $\epsilon = 0$.*

10.2 Ergodic optimization in the stochastic holonomic setting

We now propose the following model: Given $X^0 \in \Sigma$ a random word, we consider a random "noise" $B^0 \in \Sigma^*_{X^0}$ and let $\bar{X}^0 := \tau_{B^0}(X^0)$. We then use a random "control" $Y^0 \in \Sigma^*_{\bar{X}^0}$ and consider $X^1 := \tau_{Y^0}(\bar{X}^0)$. We assume that B^0, Y^0, satisfy the following "martingale-type" property:

$$\mathbb{E}[f(Y^0, \tau_{B^0}(X^0))|X^0 = \sigma(x)] = \mathbb{E}[f(Y^0, x)] \quad \text{for any } f \in C(\hat{\Sigma}). \tag{240}$$

Iterating this process, an entire random past trajectory of X^0 is represented via the random family $(X^n)_n \in \Sigma^{\mathbb{N}}$. The goal is to minimise the long time average cost,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\sum_{i=0}^{n-1} A(Y^i, \bar{X}^i)]$$

among all possible such choices.

Remark 10.6. *Note that $(B^i)_i$ also depends on such a choice, one should define a brownian motion on each Σ^*_x or fix a probability measure on each Σ^*_x and choose B following this law. Moreover, the choice of $Y^i \in \Sigma^*_{\bar{x}^i}$ depends only on B^i_0, given the definition of $\hat{\Sigma}$ ($Y^i \in \Sigma^*_{\bar{X}^i} \Leftrightarrow 1 = M(Y^i_0, \bar{X}^i_0) = M(Y^i_0, B^i_0)$).*

Given now a "strategy" $(Y^n)_n \in (\Sigma^*)^{\mathbb{N}}$, we consider for each $n \in \mathbb{N}$, the measure $\mu_n \in \mathcal{P}(\hat{\Sigma})$ defined as

$$\forall \varphi \in C(\hat{\Sigma}), \quad \int_{\hat{\Sigma}} \varphi d\mu_n := \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[\varphi(Y^i, \bar{X}^i)].$$

From $(\mu_n)_n$, one can extract a subsequence converging to some measure $\mu^{(Y^i)_i}$. We denote

$$\mathcal{M}_0 = \overline{\{\mu^{(Y^i)_i} \mid (Y^i)_i \text{ is a strategy}\}} \subset \mathcal{M}(\hat{\Sigma}).$$

For $f \in C(\Sigma)$ and $(y, x) \in \hat{\Sigma}$, denote

$$\frac{1}{2}D^y f(x) := f(\tau_y(x)) - f(x) - \frac{f(\tau_y(x)) - 2f(x) + f(\sigma(x))}{2} = \frac{f(\tau_y(x)) - f(\sigma(x))}{2}.$$

Note that the assumption made on the random noise B^0 yields an Itô-type formula: For all $f \in C(\Sigma)$, $x \in \Sigma$, with $B^0 \in \Sigma^*_{\sigma(x)}$ $Y^0 \in \Sigma^*_{\tau_{B^0}(x)}$,

$$\mathbb{E}[f(\tau_{Y^0}(\tau_{B^0}(\sigma(x)))) - f(\sigma(x))] = \mathbb{E}[D^{Y^0} f(x)].$$

Let

$$\mathcal{N}_0 = \{\mu \in \mathcal{M}(\hat{\Sigma}) \mid \forall f \in C(\Sigma), \int_{\hat{\Sigma}} D^y f(x)\, \mathrm{d}\mu(y, x) = 0\},$$

which is closed in $\mathcal{M}(\hat{\Sigma})$ as a kernel of a continuous linear map.

Lemma 10.7. *We have $\mathcal{M}_0 \subset \mathcal{N}_0$.*

PROOF. Each measure $\mu^{(Y^i)_i} \in \mathcal{M}_0$,

$$\int_{\hat{\Sigma}} D^y f(x)\,\mathrm{d}\mu_n(y,x) = \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1} f(X^{i+1}) - f(X^i)\right] = \frac{1}{n}\mathbb{E}\left[f(X^n) - f(X^0)\right] \xrightarrow[n\to\infty]{} 0.$$

Proposition 10.8. *With the above notation, we have the following*

$$c(A) := \inf_{\hat{\mu}\in\mathcal{M}_0\cap\mathcal{P}(\Sigma)} \int_{\hat{\Sigma}} A\,\mathrm{d}\hat{\mu} = \inf_{\mu\in\mathcal{N}_0\cap\mathcal{P}(\Sigma)} \int_{\hat{\Sigma}} A\,\mathrm{d}\hat{\mu} = \sup_{f\in C(\Sigma)} \inf_{(y,x)\in\hat{\Sigma}} D^y f(x) + A(y,x). \quad (241)$$

Moreover, there exists an $h \in USC_\sigma(\Sigma)$ such that

$$c(A) = \inf_{y\in\Sigma_x^*} D^y h(x) + A(y,x). \quad (242)$$

PROOF. The last equality in (241), i.e.,

$$\inf_{\hat{\mu}\in\mathcal{N}_0\cap\mathcal{P}(\hat{\Sigma})} \int_{\hat{\Sigma}} A\,\mathrm{d}\hat{\mu} = \sup_{f\in C(\Sigma)} \inf_{(y,x)\in\hat{\Sigma}} D^y f(x) + A(y,x) \quad (243)$$

is again an application of the Rockafellar-Fenchel duality. Indeed, consider the functions

$$h_1 \;:\; \begin{cases} C(\hat{\Sigma}) & \to & \mathbb{R}\cup\{+\infty\} \\ \varphi & \mapsto & \sup_{z\in\hat{\Sigma}}\{\varphi(z) - A(z)\} \end{cases}$$

$$h_2 \;:\; \begin{cases} C(\hat{\Sigma}) & \to & \mathbb{R}\cup\{+\infty\} \\ \varphi & \mapsto & \begin{cases} 0 & \text{if } \varphi \in \overline{\{Df \mid f\in C(\Sigma)\}} \\ -\infty & \text{otherwise,} \end{cases} \end{cases}$$

and note that h_1 is convex lower semicontinuous and h_2 is concave upper semicontinuous. Their Legendre transform $h_1^*(\hat{\mu}) = \sup_{\varphi\in C(\hat{\Sigma})} \{\int \varphi d\hat{\mu} - h_1(\varphi)\}$ and $h_2^*(\hat{\mu}) = \inf_\varphi \{\int_{\hat{\Sigma}} \varphi\,\mathrm{d}\hat{\mu} - h_2(\varphi)\}$ are given by

$$h_1^*(\hat{\mu}) = \begin{cases} \int_{\hat{\Sigma}} A\;d\hat{\mu} & \text{if } \hat{\mu}\in\mathcal{P}(\hat{\Sigma}) \\ +\infty & \text{otherwise.} \end{cases}$$

If $\hat{\mu} \in \mathcal{M}(\hat{\Sigma})$ is nonpositive, there exists $(\varphi_n)_n \in C(\hat{\Sigma})^{\mathbb{N}}$ such that $\forall n,\ \varphi_n \leqslant 0$ and $\int_{\hat{\Sigma}} \varphi_n\,\mathrm{d}\hat{\mu} \xrightarrow[n\to\infty]{} +\infty$. Thus, $h_1(\varphi_n) \xrightarrow[n\to\infty]{} -\infty$ and $h_1^*(\hat{\mu}) = +\infty$.
If $\hat{\mu}$ is positive, then

$$\begin{aligned} h_1^*(\hat{\mu}) &= \sup_{\varphi\in C(\hat{\Sigma})} \inf_{z\in\hat{\Sigma}} \int_{\hat{\Sigma}} \varphi\,\mathrm{d}\hat{\mu} - \varphi(z) + A(z) \\ &= \int_{\hat{\Sigma}} A\,\mathrm{d}\hat{\mu} + \sup_{\varphi\in C(\hat{\Sigma})} \inf_{z\in\hat{\Sigma}} \left(\int_{\hat{\Sigma}} (\varphi - A)\,\mathrm{d}\hat{\mu} - \varphi(z) + A(z)\right) \\ &= \int_{\hat{\Sigma}} A\,\mathrm{d}\hat{\mu} + \sup_{\psi\in C(\hat{\Sigma})} \left(\int_{\hat{\Sigma}} \psi\,\mathrm{d}\hat{\mu} - \sup_{z\in\hat{\Sigma}} \psi(z)\right) \quad \text{with } \psi = \varphi - A \\ &\geqslant \int_{\hat{\Sigma}} A\,\mathrm{d}\hat{\mu} + \sup_{\alpha\in\mathbb{R}} \alpha(\hat{\mu}(\hat{\Sigma}) - 1). \end{aligned}$$

Hence, $h_1^*(\hat\mu) = +\infty$ if $\hat\mu \notin \mathcal{P}(\hat\Sigma)$. If $\hat\mu \in \mathcal{P}(\hat\Sigma)$, $\int_{\hat\Sigma} \psi \,\mathrm{d}\hat\mu \leqslant \sup_{z\in\hat\Sigma} \psi(z)$ and $h_1^*(\hat\mu) = \int_{\hat\Sigma} A \,\mathrm{d}\hat\mu$. Similarly,

$$\forall \hat\mu \in \mathcal{M}(\hat\Sigma), \quad h_2^*(\hat\mu) = \begin{cases} 0 & \text{if } \hat\mu \in \mathcal{N}_0 \\ -\infty & \text{otherwise.} \end{cases}$$

Indeed, if $\hat\mu \notin \mathcal{N}_0$, there exists $\bar f \in C(\hat\Sigma)$ such that $\int_{\hat\Sigma} D^y \bar f(x) \,\mathrm{d}\hat\mu(y,x) \neq 0$. Hence,

$$h_2^*(\hat\mu) = \inf_{f\in C(\Sigma)} \int_{\hat\Sigma} D^y f(x) \,\mathrm{d}\hat\mu(y,x) \leqslant \inf_{\alpha\in\mathbb{R}} \alpha \int_{\hat\Sigma} D^y \bar f(x) \,\mathrm{d}\hat\mu(y,x) = -\infty.$$

If $\hat\mu \in \mathcal{N}_0$, by definition, $h_2^*(\hat\mu) = 0$.
The claimed equality is then a consequence of the Fenchel-Rockafellar duality $\inf_{\hat\mu} h_1^*(\hat\mu) - h_2^*(\hat\mu) = \sup_\varphi h_2(\varphi) - h_1(\varphi)$.

We now provide the proof of (242): Consider the functional: $\forall \mu, \nu \in \mathcal{P}(\Sigma)$,

$$\mathcal{T}(\mu,\nu) := \inf \left\{ \mathbb{E}[A(Y^0, \bar X^0)] \;\middle|\; \begin{array}{ll} X^0 \sim \nu & \\ \bar X^0 = \tau_{B^0}(X^0) & B^0 \in \Sigma^*_{X_0} \\ X^1 = \tau_{Y^0}(\bar X^0) \sim \mu & Y^0 \in \Sigma^*_{\bar X^0} \end{array} \right\}$$

Note first that if $\mathcal{T}(\mu,\nu) < +\infty$, then necessarily $\nu = \sigma^2 \# \mu$. We then have $X^0 = \sigma^2(X_1)$ and $\bar X^0 = \tau_{B^0}(X^0) = \sigma(X^1)$. So

$$\mathcal{T}(\mu,\nu) = \inf\{\mathbb{E}[A(Y^0, \sigma(X^1))] \,;\, X^1 \sim \mu, \tau_{Y^0}(\sigma(X^1)) = X^1, \tau_{B^0}(\sigma^2(X^1)) = \sigma(X^1)\}.$$

If we condition on X^1 in the expectation above, then

$$\mathcal{T}(\mu,\nu) = \int \inf\{A(y, \sigma(x)) \,;\, \tau_y(\sigma(x)) = x\} \,\mathrm{d}\mu(x) = \int \bar A \,\mathrm{d}\mu,$$

where $\bar A$ is as defined in the deterministic section above. $\mathcal{T}$ is then a forward linear transfer with forward Kantorovich operator

$$\forall g \in C(\Sigma),\ \forall x \in \Sigma, \quad T^+ g(x) := \inf\{\mathbb{E}[g(\tau_{Y^0}(\tau_{B^0}(x)) + A(Y^0, \tau_{B^0}(x))] \,;\, Y^0 \in \Sigma^*_{\tau_{B^0}(x)}\}.$$

$\mathcal{T}$ is also a backward linear transfer with backward Kantorovich operator

$$T^- f(x) := f(\sigma^2(x)) - \inf_{y\in\Sigma^*_{\sigma(x)}} A(y, \sigma(x)).$$

Indeed, let $\nu \in \mathcal{P}(\Sigma)$, $g \in C(\Sigma)$, then

$$\begin{aligned}
(\mathcal{T}_\nu)^*(g) &= \sup_{\mu\in\mathcal{P}(\Sigma)} \int_\Sigma g \,\mathrm{d}\mu - \mathcal{T}(\mu,\nu) \\
&= \sup_{\mu} \sup_{X^0\sim\nu, Y^0, B^0, X^1\sim\mu} \mathbb{E}[g(\tau_{Y^0}(\bar X^0)) - A(Y^0, \bar X^0)] \\
&= \sup_{X^0\sim\nu} \sup_{Y^0\in\Sigma^*_{\tau_{B^0}(X^0)}, B^0\in\Sigma^*_{X_0}} \mathbb{E}[g(\tau_{Y^0}(\tau_{B^0}(X^0))) - A(Y^0, \tau_{B^0}(X^0))] \\
&= \int \sup_{Y^0\in\Sigma^*_{\tau_{B^0}(x)}} \mathbb{E}[g(\tau_{Y^0}(\tau_{B^0}(x))) - A(Y^0, \tau_{B^0}(x))] \,\mathrm{d}\nu(x) \\
&:= -\int T^+(-g)(x) d\nu(x).
\end{aligned}$$

At the same time, for $\mu \in \mathcal{P}(\Sigma)$, $f \in C(\Sigma)$, then

$$
\begin{aligned}
(\mathcal{T}_\mu)^*(f) = \sup_\nu \{ \int f \, \mathrm{d}\nu - \mathcal{T}(\mu,\nu)\} &= \int f \, \mathrm{d}\sigma^2\#\mu - \mathcal{T}(\mu, \sigma^2\#\mu) \\
&= \int f \circ \sigma^2 \, \mathrm{d}\mu - \mathcal{T}(\mu, \sigma^2\#\mu) \\
&= \int f \circ \sigma^2 \, \mathrm{d}\mu - \inf_{X^0 \sim \sigma^2\#\mu, X^1 \sim \mu} A(Y^0, \sigma(X^1)) \\
&= \int f \circ \sigma^2 \, \mathrm{d}\mu - \int \inf_{y \in \Sigma^*_{\sigma(x)}} A(y, \sigma(x)) \, \mathrm{d}\mu(x) \\
&=: \int T^- f(x) \, \mathrm{d}\mu(x).
\end{aligned}
$$

We now show that the hypotheses for application of Theorem 9.2 to the backward linear transfer $\tilde{\mathcal{T}}(\mu,\nu) := \mathcal{T}(\nu,\mu)$, are satisfied.
First, it is easy to see that $\sup_{x\in\Sigma} \inf_{\nu\in\mathcal{P}(\Sigma)} \tilde{\mathcal{T}}(\delta_x,\nu) < +\infty$. Indeed, for a fixed $x \in \Sigma$, take any random noise $B^0 \in \Sigma^*_x$ and random strategy $Y^0 \in \Sigma^*_{\tau_{B^0}(x)}$, and denote the law of $\tau_{Y^0}(\tau_{B^0}(x))$ by $\bar{\nu}_x$. Then

$$
\sup_{x\in\Sigma} \inf_{\nu\in\mathcal{P}(\Sigma)} \tilde{\mathcal{T}}(\delta_x,\nu) \leqslant \sup_{x\in\Sigma} \mathcal{T}(\bar{\nu}_x, \delta_x) \leqslant \sup_{x\in\Sigma} \mathbb{E}[A(Y^0, \tau_{B^0}(x))] \leqslant \sup_{\hat{\Sigma}} A < +\infty.
$$

For the hypothesis, $\exists \mu \in \mathcal{P}(\Sigma)$, $\tilde{\mathcal{T}}(\mu,\mu) < +\infty$. The verification that (224) holds follows similarly as in the proof of Proposition 10.1. Therefore we obtain the existence of a $h \in USC_\sigma(\Sigma)$ such that

$$
\tilde{T}^- h(x) + c(\tilde{\mathcal{T}}) = h(x), \quad \forall x \in \Sigma.
$$

With $g := -h$, this is equivalent to

$$
T^+ g(x) - c(\tilde{\mathcal{T}}) = g(x). \tag{244}
$$

The corresponding Mané constant is given by

$$
\begin{aligned}
c(\tilde{\mathcal{T}}) &= \lim_{n\to\infty} \frac{1}{n} \inf_{\mu,\nu} \mathcal{T}_n(\mu,\nu) \\
&= \lim_{n\to\infty} \inf_{\mu,\nu} \frac{1}{n} \sum_{i=0}^{n-1} \inf\{\mathbb{E}[A(Y^i, \bar{X}^i)]\} \\
&= \lim_{n\to\infty} \inf_{\mu,\nu} \inf_{X^0\sim\mu, X^1\sim\nu} \int_{\hat{\Sigma}} A \, \mathrm{d}\mu_n^{(Y^i)_i} \\
&= \inf_{\hat{\mu}\in\mathcal{M}_0\cap\mathcal{P}(\Sigma)} \int_{\hat{\Sigma}} A \, \mathrm{d}\hat{\mu} \\
&= c(A).
\end{aligned}
$$

Replacing x with $\sigma(x)$ for $x \in \Sigma$ in equation (244), we have

$$
T^+ g(\sigma(x)) - g(\sigma(x)) = c(A)
$$

Recalling the definition of T^+ and the martingale assumption (240), we can write

$$\begin{aligned}
c(A) = T^+ g(\sigma(x)) - g(\sigma(x)) &= \inf_{Y^0}\{\mathbb{E}[g(\tau_{Y^0}(\tau_{B^0}(\sigma(x))) + A(Y^0, \tau_{B^0}(\sigma(x)))]\} - g(\sigma(x)) \\
&= \inf_{Y^0}\{\mathbb{E}[g(\tau_{Y^0}(x)) + A(Y^0, x)]\} - g(\sigma(x)) \\
&= \inf_{y\in\Sigma_x^*}\{g(\tau_y(x)) - g(\sigma(x)) + A(y,x)\} \\
&= \inf_{y\in\Sigma_x^*}\{D^y g(x) + A(y,x)\}
\end{aligned}$$

In view of the duality (243), this implies that (241) holds and concludes the proof.

11 Regularizations of linear transfers and applications

We continue to deal with cases where $\mathcal{T}$ is not necessarily weak*-continuous on $\mathcal{M}(X)$ and may even have infinite values. The strategy now is to reduce the situation to the bounded and continuous case via a regularization procedure.

11.1 Regularization and weak KAM solutions for unbounded transfers

Lemma 11.1 (Regularisation of a backward linear transfer). *Let (X,d) be a complete metric space and let $\mathcal{W}_d(\mu,\nu)$ be the cost minimising optimal transport associated to the cost $d(x,y)$. For a given backward linear transfer $\mathcal{T} : \mathcal{P}(X)\times\mathcal{P}(X) \to \mathbb{R}\cup\{+\infty\}$, we associate for each $\epsilon > 0$ the functional*

$$\mathcal{T}_\epsilon(\mu,\nu) := \inf\{\frac{1}{\epsilon}W_1(\mu,\sigma_1) + \mathcal{T}(\sigma_1,\sigma_2) + \frac{1}{\epsilon}W_1(\sigma_2,\nu)\,;\, \sigma_1,\sigma_2 \in \mathcal{P}(X)\}.$$

Then, $\mathcal{T}_\epsilon$ has the following properties:

1. *$\mathcal{T}_\epsilon$ is a weak* continuous backward linear transfer.*
2. $\inf\{\mathcal{T}_\epsilon(\mu,\nu)\,;\,\mu,\nu\in\mathcal{P}(X)\} = \inf\{\mathcal{T}(\mu,\nu)\,;\,\mu,\nu\in\mathcal{P}(X)\}$.
3. *$\mathcal{T}_\epsilon(\mu,\nu) \leqslant \mathcal{T}(\mu,\nu)$ and $\mathcal{T}_\epsilon(\mu,\nu)\uparrow\mathcal{T}(\mu,\nu)$ as $\epsilon\to 0$.*
4. *$\mathcal{T}_\epsilon$ Γ-converges to $\mathcal{T}$ as $\epsilon\to 0$.*
5. *If T_ϵ, T, denote the backward Kantorovich operators associated to $\mathcal{T}_\epsilon$, $\mathcal{T}$, respectively, then for any $f\in USC(X)$, $T_\epsilon f(x) \searrow Tf(x)$ as $\epsilon\to 0$.*

Proof. First note that since d is continuous, the linear transfer $\mathcal{W}_d$ is weak-$*$ continuous on $\mathcal{P}(X)$ (See e.g., [55], Theorem 1.51, p.40).

1. We know that for each fixed $\epsilon > 0$, $\mathcal{T}_\epsilon$ is a weak* lower semi-continuous linear backward transfer. To prove that it is continuous, assume $\mu_n\to\mu$ and $\nu_n\to\nu$. By the lower semi-continuity, we have $\liminf_n \mathcal{T}_\epsilon(\mu_n,\nu_n) \geqslant \mathcal{T}_\epsilon(\mu,\nu)$. On the other hand, from the fact that $\limsup_n \inf_{\sigma_1,\sigma_2} \leqslant \inf_{\sigma_1,\sigma_2}\limsup_n$, we have

$$\begin{aligned}
\limsup_n \mathcal{T}_\epsilon(\mu_n,\nu_n) &\leqslant \inf\{\limsup_n \frac{1}{\epsilon}W_1(\mu_n,\sigma_1) + \mathcal{T}(\sigma_1,\sigma_2) + \limsup_n \frac{1}{\epsilon}W_1(\sigma_2,\nu_n)\,;\,\sigma_1,\sigma_2\in\mathcal{P}(X)\} \\
&= \inf\left\{\frac{1}{\epsilon}W_1(\mu,\sigma_1) + \mathcal{T}(\sigma_1,\sigma_2) + \frac{1}{\epsilon}W_1(\sigma_2,\nu)\,;\,\sigma_1,\sigma_2\in\mathcal{P}(X)\right\} \\
&= \mathcal{T}_\epsilon(\mu,\nu),
\end{aligned}$$

which shows that $\mathcal{T}_\epsilon(\mu_n,\nu_n)\to\mathcal{T}_\epsilon(\mu,\nu)$ as $n\to\infty$.

2. Observe from the definition of $\mathcal{T}_\epsilon$, that

$$\inf_{\mu,\nu}\{\mathcal{T}_\epsilon(\mu,\nu)\} = \inf_{\sigma,\sigma',\mu,\nu}\{\frac{1}{\epsilon}W_1(\mu,\sigma)+\mathcal{T}(\sigma,\sigma')+\frac{1}{\epsilon}W_1(\sigma',\nu)\} \tag{245}$$

and it is clear that for fixed σ,σ', the minimal μ is σ, and ν is σ', and the transport cost $W_1(\mu,\sigma)=0=W_1(\sigma',\nu)$.

3. The inequality $\mathcal{T}_\epsilon(\mu,\nu)\leqslant\mathcal{T}(\mu,\nu)$ holds by selecting $\sigma_1=\mu$ and $\sigma_2=\nu$ and noting that $\mathcal{W}_d(\sigma,\sigma)=0$ for every $\sigma\in\mathcal{P}(X)$. The monotone property of $\epsilon\mapsto\mathcal{T}_\epsilon(\mu,\nu)$ is immediate by definition. Let now σ_1^ϵ, σ_2^ϵ realise the infimum

$$\mathcal{T}_\epsilon(\mu,\nu)=\frac{1}{\epsilon}W_1(\mu,\sigma_\epsilon^1)+\mathcal{T}(\sigma_\epsilon^1,\sigma_\epsilon^2)+\frac{1}{\epsilon}W_1(\sigma_\epsilon^2,\nu). \tag{246}$$

By refining if necessary, we may assume that $\sigma_\epsilon^1\to\overline{\sigma}_1$ and $\sigma_\epsilon^2\to\overline{\sigma}_2$ as $\epsilon\to 0$. If $\sup_{\epsilon>0}\mathcal{T}_\epsilon(\mu,\nu)<\infty$), then $W_1(\mu,\sigma_\epsilon^1)\to 0$ and $W_1(\sigma_\epsilon^2,\nu)\to 0$ as $\epsilon\to 0$, hence $\overline{\sigma}_1=\mu$ and $\overline{\sigma}_2=\nu$. Then (246) and weak-$*$ lower semi-continuity of $\mathcal{T}$ implies

$$\liminf_{\epsilon\to 0}\mathcal{T}_\epsilon(\mu,\nu)\geqslant\liminf_{\epsilon\to 0}\mathcal{T}(\sigma_\epsilon^1,\sigma_\epsilon^2)\geqslant\mathcal{T}(\mu,\nu).$$

4. First recall that for Γ-convergence, one needs to prove the Γ-lim inf inequality: For every sequence $(\mu^\epsilon,\nu^\epsilon)\to(\mu,\nu)$, it holds that $\liminf_{\epsilon\to 0}\mathcal{T}_\epsilon(\mu^\epsilon,\nu^\epsilon)\geqslant\mathcal{T}(\mu,\nu)$, and the Γ-lim sup inequality: There exists a sequence $(\mu^\epsilon,\nu^\epsilon)\to(\mu,\nu)$ such that $\limsup_{\epsilon\to 0}\mathcal{T}_\epsilon(\mu^\epsilon,\nu^\epsilon)\leqslant\mathcal{T}(\mu,\nu)$.
The Γ-lim sup inequality is immediate: Take $(\mu^\epsilon,\nu^\epsilon)=(\mu,\nu)$, and the inequality follows from $\mathcal{T}_\epsilon\leqslant\mathcal{T}$.
For the Γ-lim inf inequality, we can assume without loss that $\liminf_{\epsilon\to 0}\mathcal{T}_\epsilon(\mu^\epsilon,\nu^\epsilon)<+\infty$, since otherwise there is nothing to prove. Now by monotonicity, we have $\mathcal{T}_\epsilon(\mu^\epsilon,\nu^\epsilon)\geqslant\mathcal{T}_{\epsilon'}(\mu^\epsilon,\nu^\epsilon)$ for $\epsilon\leqslant\epsilon'$. The weak-$*$ lower semi-continuity of $\mathcal{T}_{\epsilon'}$ therefore implies

$$\liminf_{\epsilon\to 0}\mathcal{T}_\epsilon(\mu^\epsilon,\nu^\epsilon)\geqslant\liminf_{\epsilon\to 0}\mathcal{T}_{\epsilon'}(\mu^\epsilon,\nu^\epsilon)\geqslant\mathcal{T}_{\epsilon'}(\mu,\nu).$$

By 3) and letting $\epsilon'\to 0$, we obtain $\liminf_{\epsilon\to 0}\mathcal{T}_\epsilon(\mu^\epsilon,\nu^\epsilon)\geqslant\mathcal{T}(\mu,\nu)$.

5. First note that the monotonicity of $T_\epsilon f(x)$ is immediate from the expression

$$T_\epsilon f(x)=\sup\left\{\int f\,\mathrm{d}\sigma-\mathcal{T}_\epsilon(\delta_x,\sigma)\right\},$$

and the monotonicity of $\mathcal{T}_\epsilon$. We immediately have $\liminf_{\epsilon\to 0}T_\epsilon f(x)\geqslant Tf(x)$. On the other hand, let ϵ_j be a sequence such that $T_{\epsilon_j}f(x)\to\limsup_{\epsilon\to 0}T_\epsilon f(x)$. Then

$$T_{\epsilon_j}f(x)=\sup_\sigma\left\{\int f\,\mathrm{d}\sigma-\mathcal{T}_\epsilon(\delta_x,\sigma)\right\}=\int f\,\mathrm{d}\sigma^{\epsilon_j}-\mathcal{T}_{\epsilon_j}(\delta_x,\sigma^{\epsilon_j}).$$

By refining to a further subsequence if necessary, we may assume $\sigma^{\epsilon_j}\to\sigma^*$. Then we obtain with $j\to\infty$,

$$\begin{aligned}\limsup_{\epsilon\to 0}T_\epsilon f(x)&\leqslant\int f\,\mathrm{d}\sigma^*-\liminf_{j\to\infty}\mathcal{T}_{\epsilon_j}(\delta_x,\sigma^{\epsilon_j})\\&\leqslant\int f\,\mathrm{d}\sigma^*-\mathcal{T}(\delta_x,\sigma^*)\\&\leqslant\sup_\sigma\left\{\int f\,\mathrm{d}\sigma-\mathcal{T}(\delta_x,\sigma)\right\}=Tf(x),\end{aligned}$$

where the second inequality was obtained from the Γ-convergence.

Lemma 11.2. *Let X be a compact metric space and let $\mathcal{T}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains the Dirac measures. Assume hypothesis (217) and let $\mathcal{T}_\epsilon$ be the regularisation of $\mathcal{T}$ according to Lemma 11.1. Then, the following properties hold:*

1. *$c(\mathcal{T}_\epsilon)$ is the unique constant such that $|(\mathcal{T}_\epsilon)_n(\mu,\nu) - nc_\epsilon| \leqslant C_\epsilon$, for all n and all μ,ν.*
2. *$c(\mathcal{T}_\epsilon) \uparrow c(\mathcal{T})$ as $\epsilon \to 0$.*
3. *$c(\mathcal{T}) = \inf\{\mathcal{T}(\mu,\mu)\,;\, \mu \in \mathcal{P}(X)\}$.*

Proof: Use Lemma 11.1 to regularise $\mathcal{T}$ to $\mathcal{T}_\epsilon$, and let δ_ϵ be the modulus of continuity for $\mathcal{T}_\epsilon$, which is also the modulus of continuity for $(\mathcal{T}_\epsilon)_n$, the n-fold inf-convolution of $\mathcal{T}_\epsilon$. Use now Corollary 8.1 for each ϵ to find $c_\epsilon = c(\mathcal{T}_\epsilon)$ with the properties stated there. Note, in particular that $c(\mathcal{T}_\epsilon) \leqslant c(\mathcal{T})$. It follows that $c(\mathcal{T}_\epsilon)$ converges as $\epsilon \to 0$. We let $K(\mathcal{T})$ be this limit. Note that $K(\mathcal{T}) \leqslant c(\mathcal{T})$. We shall prove that

$$c(\mathcal{T}) = \inf\{\mathcal{T}(\mu,\mu); \mu \in \mathcal{P}(X)\} = K(\mathcal{T}). \tag{247}$$

This follows from the Γ-convergence, since

$$c(\mathcal{T}_\epsilon) = \inf\{\mathcal{T}_\epsilon(\mu,\mu); \mu \in \mathcal{P}(X)\} = \mathcal{T}_\epsilon(\mu_\epsilon,\mu_\epsilon)$$

for some μ_ϵ, then if $\bar{\mu}$ is a cluster point for (μ_ϵ) as $\epsilon \to 0$, the Γ-convergence of $\mathcal{T}_\epsilon$ implies that $c(\mathcal{T}_\epsilon) = \mathcal{T}_\epsilon(\mu_\epsilon,\mu_\epsilon) \to \mathcal{T}(\bar{\mu},\bar{\mu})$. If now ν is any other probability measure, then $\mathcal{T}(\nu,\nu) \geqslant \mathcal{T}_\epsilon(\nu,\nu) \geqslant \mathcal{T}_\epsilon(\mu_\epsilon,\mu_\epsilon)$, hence $\mathcal{T}(\nu,\nu) \geqslant \mathcal{T}(\bar{\mu},\bar{\mu}) = K(\mathcal{T})$ and

$$K(\mathcal{T}) = \inf\{\mathcal{T}(\mu,\mu); \mu \in \mathcal{P}(X)\}.$$

On the other hand, for every μ,

$$c(\mathcal{T}) = \sup_n \inf_{\sigma,\nu} \frac{\mathcal{T}_n(\sigma,\nu)}{n} \leqslant \sup_n \frac{\mathcal{T}_n(\mu,\mu)}{n} \leqslant \mathcal{T}(\mu,\mu),$$

since $\mathcal{T}_n(\mu,\mu)$ is subadditive on the diagonal, hence $c(\mathcal{T}) \leqslant K(\mathcal{T})$ and (247) follows. □

The above theorem has the following useful corollary, which implies the uniqueness of the level c, where weak KAM solutions occur.

Corollary 11.3. *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(X)$ such that $\mathcal{D}_1(\mathcal{T})$ contains all Dirac measures and that (217), (224) hold. Then,*

1. *If $Tu + d \leqslant u$ for some $d \in \mathbb{R}$ and some $u \in USC(X)$, then $d \leqslant c$.*
2. *If $Tv + d \geqslant v$ for some $d \in \mathbb{R}$ and some $v \in USC(X)$, then $d \geqslant c$.*

Proof: If $Tu + d \leqslant u$ for some $d \in \mathbb{R}$ and $u \in USC(X)$, then $T^n u + nd \leqslant u$ for all $n \in \mathbb{N}$. Applying T^m and using the linearity of T^m with respect to constants, we find $T^{m+n}u + dn \leqslant T^m u$, and hence

$$T^{m+n}u + d(m+n) \leqslant T^m u + dm$$

So $n \mapsto T^n u + dn$ is decreasing to a function $\tilde{u}$. But if $d > c$, then $T^n u + dn \geqslant T^n u + cn$ and $\tilde{u}$ is proper by the first part of the proof of Theorem 9.2 and $T\tilde{u} + d = \tilde{u}$ on X. It follows that for any $\mu \in \mathcal{P}(X)$,

$$\int_X T^n \tilde{u}\, d\mu = -nd + \int_X \tilde{u}\, d\mu.$$

On the other hand, let $\bar{\mu}$ be such that $\mathcal{T}(\bar{\mu}, \bar{\mu}) = \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu) = c$. Then, from Lemma 9.1 we have

$$\liminf_n \frac{1}{n} \int_X T^n f\, d\bar{\mu} \geqslant -\mathcal{T}(\bar{\mu}, \bar{\mu}) = -c.$$

It follows that $-c \leqslant -d$, which is a contradiction. □

Here is a case where we can associate to $\mathcal{T}$ an effective Kantorovich operator without the equicontinuity assumption.

Theorem 11.4. *Let X is a compact metric space and let $\mathcal{T}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains the Dirac measures. Assume (217) and that for some $\epsilon > 0$, we have*

$$c(\mathcal{T}_\epsilon) = c(\mathcal{T}), \tag{248}$$

where $\mathcal{T}_\epsilon$ be the regularisation of $\mathcal{T}$ according to Lemma 11.1. Then, there exists an idempotent backward linear transfer $\mathcal{T}_\infty$ on $\mathcal{P}(X) \times \mathcal{P}(X)$, with a Kantorovich operator $T^\infty : C(X) \to USC(X)$ such that $T \circ T_\infty f + c = T_\infty f$ for all $f \in C(X)$.

Proof: Consider the regularisation $\mathcal{T}_\epsilon$ of $\mathcal{T}$. By Corollary 8.1, there exists a Kantorovich operator $T_\epsilon^\infty : C(X) \to C(X)$, such that $T_\epsilon \circ T_\epsilon^\infty f + c_\epsilon = T_\epsilon^\infty f$ for all $f \in C(X)$, and an idempotent transfer $\mathcal{T}_{\epsilon,\infty}$. We have the following properties: Under the assumption that $c_\epsilon = c$,

1. $T_\epsilon^\infty f \leqslant T_{\epsilon'}^\infty f$ for all f, whenever $\epsilon < \epsilon'$.
2. There exists a $\bar{\mu} \in \mathcal{P}(X)$ such that $\mathcal{T}(\bar{\mu}, \bar{\mu}) = c$, and $\int_X T^\infty f\, \mathrm{d}\bar{\mu} \geqslant \int_X f\, \mathrm{d}\bar{\mu}$.

To see that property 1 holds, observe that from monotonicity in ϵ for T_ϵ (see Lemma 11.1), we obtain monotonicity in ϵ for $\bar{T}_\epsilon f(x) := \limsup_n (T_\epsilon^n f(x) + nc_\epsilon)$ under the assumption that $c_\epsilon = c$. Hence by definition of $T_\epsilon^\infty f(x) = \lim_{n\to\infty} (T_\epsilon^n \circ \bar{T}_\epsilon f(x) + nc_\epsilon)$, we deduce monotonicity for T_ϵ^∞.

For Property 2, let $\bar{\mu}_\epsilon$ achieve $c_\epsilon = \mathcal{T}_\epsilon(\bar{\mu}_\epsilon, \bar{\mu}_\epsilon)$. By Theorem 8.5, we have $\mathcal{T}_{\epsilon,\infty}(\bar{\mu}_\epsilon, \bar{\mu}_\epsilon) = 0$, which implies $\int_X f\, \mathrm{d}\bar{\mu}_\epsilon \leqslant \int_X T_\epsilon^\infty f\, \mathrm{d}\bar{\mu}_\epsilon$.

On the other hand, extract a subsequence ϵ_j of the $\bar{\mu}_\epsilon$ so that $\bar{\mu}_{\epsilon_j} \to \bar{\mu}$. Then for any $\epsilon > 0$, eventually, $\epsilon_j < \epsilon$. It then follows the monotonicity of Property 1 that $\int_X T_{\epsilon_j}^\infty f\, \mathrm{d}\bar{\mu}_{\epsilon_j} \leqslant \int_X T_\epsilon^\infty f\, \mathrm{d}\bar{\mu}_{\epsilon_j}$. Let $j \to \infty$ to obtain

$$\int_X f\, \mathrm{d}\bar{\mu} \leqslant \int_X T_\epsilon^\infty f\, \mathrm{d}\bar{\mu}.$$

The monotonicity of $T_\epsilon^\infty f$ and the above lower bound ensures that for $\bar{\mu}$-a.e. x, the limit $\lim_{\epsilon\to 0} T_\epsilon^\infty f(x)$ exists as a real number and is not $-\infty$. In particular, we deduce that $T^\infty f(x) := \lim_{\epsilon\to 0} T_\epsilon^\infty f(x)$, satisfies $\int_X T^\infty f \,\mathrm{d}\bar\mu \geqslant \int_X f \,\mathrm{d}\bar\mu$; in particular, it is not identically $-\infty$, and belongs to $USC(X)$.

By Lemma 4.2, we have

$$T\circ T^\infty f(x)+c=\lim_{\epsilon\to 0} T\circ T_\epsilon^\infty f(x)+c_\epsilon \leqslant \lim_{\epsilon\to 0} T_\epsilon\circ T_\epsilon^\infty f(x)+c_\epsilon=\lim_{\epsilon\to 0} T_\epsilon^\infty f(x)=T^\infty f.$$

On the other hand, the monotonicity in ϵ gives

$$T_\epsilon^\infty f = T_\epsilon\circ T_\epsilon^\infty f + c_\epsilon \leqslant T_{\epsilon'}\circ T_\epsilon^\infty f + c$$

for any $\epsilon' > \epsilon$. By Lemma 4.2 applied to $T_{\epsilon'}$ and the sequence $T_\epsilon^\infty f$, we can pass the limit in ϵ through T_ϵ' to obtain

$$T^\infty f(x)=\lim_{\epsilon\to 0} T_\epsilon^\infty f(x)\leqslant \lim_{\epsilon\to 0} T_{\epsilon'}\circ T_\epsilon^\infty f(x)+c=T_{\epsilon'}\circ T^\infty f(x)+c.$$

Now we let $\epsilon'\to 0$ and use Property 4 of Lemma 11.1 to obtain $T_\infty f(x)\leqslant T\circ T^\infty f(x)+c$, and thus obtaining equality.

11.2 Weak KAM solutions for unbounded transfers

The following lemma shows that the above hypothesis $c(\mathcal{T}_\epsilon)=c(\mathcal{T})$ is not vacuous as it occurs in many examples.

Proposition 11.5. *Let $\mathcal{T}$ be a backward linear transfer with Kantorovich operator T. In any of the following cases,*

1. $\inf\{\mathcal{T}(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}=\inf\{\mathcal{T}(\mu,\mu)\,;\,\mu\in\mathcal{P}(X)\}$,
2. *$\mathcal{T}$ is symmetric and for some $\epsilon_0>0$, T^- maps every continuous function to a $1/\epsilon_0$-Lipschitz function,*

we have $c(\mathcal{T}_\epsilon)=c(\mathcal{T})$ for all small enough $\epsilon>0$.

Proof: To see 1) note that $\inf\{\mathcal{T}(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}=\inf\{\mathcal{T}_\epsilon(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}$ for every $\epsilon>0$. By property 2 of Lemma 11.1, and property 3 of Lemma 11.2, we get

$$c(\mathcal{T}_\epsilon)\leqslant c(\mathcal{T})=\inf_{\mu\in\mathcal{P}(X)}\mathcal{T}(\mu,\mu)=\inf\{\mathcal{T}(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}=\inf\{\mathcal{T}_\epsilon(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}\leqslant c(\mathcal{T}_\epsilon).$$

For 2) write

$$\begin{aligned}
\mathcal{T}_\epsilon(\mu,\mu)&=\inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\frac{1}{\epsilon}W(\sigma_2,\mu)\}\\
&\geqslant\inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W(\mu,\sigma_1)-\mathcal{T}(\sigma_1,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\frac{1}{\epsilon}W(\sigma_2,\mu)\}+\inf\{\mathcal{T}(\sigma_1,\sigma_1)\,;\,\sigma_1\}\\
&\geqslant\inf_{\sigma_1,\sigma_2,\sigma_3}\{\frac{1}{\epsilon}W(\mu,\sigma_1)-\mathcal{T}(\sigma_1,\sigma_3)+\mathcal{T}(\sigma_3,\sigma_2)+\frac{1}{\epsilon}W(\sigma_2,\mu)\}+\inf\{\mathcal{T}(\sigma_1,\sigma_1)\,;\,\sigma_1\}\\
&=(\frac{1}{\epsilon}W)\star(-\mathcal{T})\star\mathcal{T}\star(\frac{1}{\epsilon}W)(\mu,\mu)+c.
\end{aligned}$$

It suffices to show that $(\frac{1}{\epsilon}W) \star (-\mathcal{T}) \star \mathcal{T} \star (\frac{1}{\epsilon}W)(\mu,\mu) \geqslant 0$. Note that we can write

$$\begin{aligned}(-\mathcal{T}) \star \mathcal{T}(\mu,\nu) &= \inf_\sigma\{-\mathcal{T}(\mu,\sigma) + \mathcal{T}(\sigma,\nu)\} \\ &= \inf_f \inf_\sigma \{\int_X Tf \,\mathrm{d}\mu - \int_X f \,\mathrm{d}\sigma + \mathcal{T}(\sigma,\nu)\} \\ &= \inf_f\{\int_X T^- f\,\mathrm{d}\mu + \int_X T^+(-f)\,\mathrm{d}\nu\}.\end{aligned}$$

Then with the notation that S^-_ϵ is the backward Kantorovich operator for $\frac{1}{\epsilon}W$, we arrive at

$$\begin{aligned}(\frac{1}{\epsilon}W) \star (-\mathcal{T}) \star \mathcal{T} \star (\frac{1}{\epsilon}W)(\mu,\mu) &= \inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W(\mu,\sigma_1) + (-\mathcal{T})\star\mathcal{T}(\sigma_1,\sigma_2) + \frac{1}{\epsilon}W(\sigma_2,\mu)\} \\ &= \inf_f \inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W(\mu,\sigma_1) + \int_X T^- f\,\mathrm{d}\sigma_1 + \int_X T^+(-f)\,\mathrm{d}\sigma_2 + \frac{1}{\epsilon}W(\sigma_2,\mu)\} \\ &= \inf_f\{-\int_X S^-_\epsilon(-T^- f)\,\mathrm{d}\mu - \int_X S^-_\epsilon(-T^+(-f))\,\mathrm{d}\mu\} \\ &= \inf_f\{-\int_X S^-_\epsilon(-T^- f)\,\mathrm{d}\mu - \int_X S^-_\epsilon(T^- f)\,\mathrm{d}\mu\} \\ &= \inf_f\{-\int_X (-T^- f)\,\mathrm{d}\mu - \int_X T^- f\,\mathrm{d}\mu\} \\ &= 0,\end{aligned}$$

where the second-last equality follows from the fact that whenever g is $\frac{1}{\epsilon_0}$-Lipschitz, then $S^-_\epsilon g = g$.

Proposition 11.6. *Let $S : C(X) \to C(Y)$ be a Markov operator (i.e., a bounded linear positive operator such that $T1 = 1$) and let $S^* : \mathcal{M}(X) \to \mathcal{M}(Y)$ be its adjoint. Given a backward linear transfer $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y)$ and $\lambda \in (0,1)$, define*

$$\tilde{\mathcal{T}}(\mu,\nu) := \mathcal{T}(\mu, \lambda S^*\mu + (1-\lambda)\nu). \tag{249}$$

Then, $\tilde{\mathcal{T}}$ is a backward linear transfer with Kantorovich operator

$$\tilde{T}^- f(x) := T^-\left(\frac{1}{1-\lambda}f\right)(x) - \frac{\lambda}{1-\lambda}Sf(x) \tag{250}$$

Proof: Write

$$\begin{aligned}(\tilde{\mathcal{T}}_\mu)^*(f) &= \sup_\sigma\{\int_X f\,\mathrm{d}\sigma - \tilde{\mathcal{T}}(\mu,\sigma)\} \\ &= \sup_\sigma\{\int_X f\,\mathrm{d}\sigma - \mathcal{T}(\mu, \lambda S^*\mu + (1-\lambda)\sigma)\}\end{aligned}$$

with $\tilde{\sigma} := \lambda S^*\mu + (1-\lambda)\sigma$, we obtain $\sigma = \frac{1}{1-\lambda}\tilde{\sigma} - \frac{\lambda}{1-\lambda}S^*\mu$. Hence after substitution we obtain

$$\begin{aligned}(\tilde{\mathcal{T}}_\mu)^*(f) &= \sup_{\tilde{\sigma}}\{\int_X \frac{1}{1-\lambda}f\,\mathrm{d}\tilde{\sigma} - \mathcal{T}(\mu,\tilde{\sigma})\} - \frac{\lambda}{1-\lambda}\int_Y Sf\,\mathrm{d}\mu \\ &= \int_X \left[T^-\left(\frac{1}{1-\lambda}f\right) - \frac{\lambda}{1-\lambda}Sf\right]\mathrm{d}\mu.\end{aligned}$$

Theorem 11.7. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$. Then, for every $\lambda \in (0,1)$, there exists a convex function φ_λ such that the linear transfer given by*

$$\tilde{\mathcal{T}}_\lambda(\mu,\nu) := \mathcal{T}(\mu, \lambda(\nabla\varphi_\lambda)_{\#}\mu + (1-\lambda)\nu) \tag{251}$$

is such that

$$\inf_{\mu.\nu\in\mathcal{P}(X)} \tilde{\mathcal{T}}_\lambda(\mu,\nu) = \inf_{\mu\in\mathcal{P}(X)} \tilde{\mathcal{T}}_\lambda(\mu,\mu).$$

In particular, $\tilde{\mathcal{T}}_\lambda$ admits weak KAM solutions, that is there exists $g \in USC(X)$ and $c \in \mathbb{R}$ such that

$$T^- g + c = \lambda\, g(\nabla\varphi) + (1-\lambda)\, g. \tag{252}$$

Proof: Let $\mathcal{T}(\mu_0,\nu_0) = \inf_{\mu,\nu} \mathcal{T}(\mu,\nu) < +\infty$ for some μ_0 and ν_0, and use Brenier's theorem to find a convex function φ such that $\nabla\varphi_{\#}\mu_0 = (1-\frac{1}{\lambda})\mu_0 + \frac{1}{\lambda}\nu_0$.

Consider now the backward linear transfer $\tilde{\mathcal{T}}(\mu,\nu) := \mathcal{T}(\mu, \lambda\nabla\varphi\#\mu + (1-\lambda)\nu)$ and note that $\mathcal{T}(\mu_0,\mu_0) = \mathcal{T}(\mu_0,\nu_0) < +\infty$. Moreover,

$$\inf_{\mu,\nu} \tilde{\mathcal{T}}(\mu,\nu) \geqslant \inf_{\mu,\nu} \mathcal{T}(\mu,\nu) = \mathcal{T}(\mu_0,\nu_0) = \tilde{\mathcal{T}}(\mu_0,\mu_0) \geqslant \inf_{\mu} \tilde{\mathcal{T}}(\mu,\mu),$$

hence $\inf_{\mu,\nu} \tilde{\mathcal{T}}(\mu,\nu) = \inf_\mu \tilde{\mathcal{T}}(\mu,\mu)$., and in particular, $\tilde{\mathcal{T}}$ satisfies the hypotheses of Theorem 11.4, and admits weak KAM solutions for its Kantorovich operator, which is given by $\tilde{T}^- f = T^-(\frac{1}{1-\lambda} f) - \frac{\lambda}{1-\lambda} f \circ \nabla\varphi$. In other words, by setting $g := \frac{1}{1-\lambda} f$, we have

$$T^- g + c = \lambda\, g(\nabla\varphi) + (1-\lambda)\, g.$$

11.3 The heat semi-group and other examples

Assumption (248) is actually satisfied by a large number of our transfer examples.

1) Let $\mathcal{T}$ be the backward transfer associated to a convex lower semi-continuous functional I on Wasserstein space, that is $\mathcal{T}(\mu,\nu) := I(\nu)$. Assumption (248) then holds trivially as $c(\mathcal{T}) = \inf I$ in this case, and the associated idempotent transfer is $\mathcal{T}_\infty(\mu,\nu) := I(\nu) - c$, while the corresponding idempotent operator is $T_\infty f = I^*(f) + c$, where I^* is the Legendre transform of I.

2) Assumption (248) clearly holds for any transfer that is $\{0,+\infty\}$-valued provided (217) is satisfied. Note that if T is a Markov operator, then assumption (217) means that T has an invariant measure. In this case, $c(\mathcal{T}) = 0$, and for every f, $T_\infty f$ is an invariant function under f.

3) If T is induced by a continuous point transformation, i.e., $Tf(x) = f(\sigma(x)$ for a continuous map $\sigma : X \to X$, then by a Theorem of Bogolyubov and Krylov, T has an invariant measure and the above applies. The operator T_∞ is then given by

$$T^\infty f(x) = f(\limsup_{m\to\infty} \sigma^m(x)) := f(\sigma^\infty(x)).$$

However, the regularity of the invariant functions $T_\infty f$ can vary widely. For example,

- If one takes $X = [0,1]$ and $\sigma(x) = x^2$, then $\sigma^\infty(x) = 0$, if $x \in [0,1)$ and 1 if $x = 1$. In this case, $T^\infty f$ only belongs to $USC_\sigma(X)$.

- On the other hand, if $\sigma(x) = 1 - x$, then $\sigma^\infty(x) = \max\{x, 1-x\}$ is continuous.

4) *The heat semi-group:* Recall the Skorokhod transfer of Example 3. 4. Instead of considering a stopping time τ, we let $\tau = t > 0$ to be deterministic, and define, for measures μ, ν, on a compact Riemannian manifold M, the linear transfer,

$$\mathcal{T}_t(\mu,\nu) = \begin{cases} 0 & \text{if } B_0 \sim \mu \text{ and } B_t \sim \nu \\ +\infty & \text{otherwise.} \end{cases}$$

Then $T_t f(x) = \mathbb{E}^x[f(B_t)] = P_t f(x)$, where P_t is the heat semigroup. Note that since the volume measure λ_M, i.e., the uniform probability measure on M, is invariant, we have $\mathcal{T}_t(\lambda_M, \lambda_M) = 0$, hence Condition (217) is satisfied. We now have the following easy proposition, which puts our asymptotic result in the following classical context.

Proposition 11.8. *The collection $\{\mathcal{T}_t\}_{t>0}$ is a semigroup of backward linear transfers with Kantorovich operators $\{T_t\}_{t>0}$. The corresponding idempotent backward linear transfer $\mathcal{T}_\infty(\mu,\nu) = \sup\{\int_M f d\nu - \int_M f d\lambda_M\,;\, f \in C(M)\}$ with Kantorovich operator $T_\infty f(x) = \int_M f \,\mathrm{d}\lambda_M$.*

Proof: It is immediate to verify that

$$(\mathcal{T}_{t,\mu})^*(f) = \sup\{\int_M f d\nu\,;\, \nu \text{ such that } B_t \sim \nu, B_0 \sim \mu\} = \int_M P_t f(x) d\mu(x).$$

Moreover, it is a standard property of the heat semigroup, that $P_t f \to P_\infty f = \int_M f \,\mathrm{d}\lambda_M$, uniformly on M, as $t \to \infty$, for any $f \in C(M)$. By the 1-Lipschitz property of T_t, we conclude $T_t \circ T_\infty f = T_\infty f$.

12 Stochastic weak KAM on the Torus

In this section, we are interested in making the connection between our general notion of linear transfers, stochastic mass transports, and existing work on stochastic weak KAM theory, in particular, by Gomes [36]. We shall therefore restrict our setting to $M = \mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$, the d-dimensional flat torus. Note that, unlike the deterministic Mather theory, this does not fall under the Monge-Kantorovich setting.

First, we introduce the stochastic mass transport of a probability measure μ to a probability measure ν on $\mathcal{P}(M)$ in time $t > 0$ (see e.g. [52] when the space is $\mathbb{R}^d$ and $t = 1$). Define $\mathcal{T}_t(\mu,\nu) : \mathcal{P}(M) \times \mathcal{P}(M) \to \mathbb{R} \cup \{+\infty\}$ via the formula,

$$\mathcal{T}_t(\mu,\nu) := \inf\left\{\mathbb{E}\int_0^t L(X(s), \beta_X(s,X))\,\mathrm{d}s\,;\, X(0) \sim \mu, X(t) \sim \nu, X \in \mathcal{A}_{[0,t]}\right\}, \tag{253}$$

where $L : TM \to [0,\infty)$ is a given Lagrangian function which we detail below, and X is a continuous semi-martingale with an associated drift β_X, belonging to a class of stochastic processes $\mathcal{A}_{[0,t]}$ defined below.

Stochastic transport has a dual formulation (first proven in Mikami-Thieullin [52] for the space $\mathbb{R}^d$) that permits it to be realised as a backward linear transfer. In fact, by introducing the operator $T_t : C(M) \to USC(M)$ via the formula

$$T_t f(x) := \sup_{X \in \mathcal{A}_{[0,t]}} \left\{ \mathbb{E}\left[f(X(t))|X(0)=x\right] - \mathbb{E}\left[\int_0^t L(X(s), \beta_X(s,X))\, \mathrm{d}s | X(0)=x\right]\right\}, \tag{254}$$

the duality relation between $\mathcal{T}_t$ and T_t can be readily detailed, see Proposition 12.1. The operator T_t connects to the work of Gomes via a Hamilton-Jacobi-Bellman equation (255),

Concerning the assumptions on L, we make the following hypotheses.

(A0) L is continuous, non-negative, $L(x,0)=0$, and $D_v^2 L(x,v)$ is positive definite for all $(x,v) \in TM$ (in particular $v \mapsto L(x,v)$ is convex).

(A1) There exists a function $\gamma = \gamma(|v|) : \mathbb{R}^n \to [0,\infty)$ such that $\lim_{|v|\to\infty} \frac{L(x,v)}{\gamma(v)} = +\infty$ and $\lim_{|v|\to\infty} \frac{|v|}{\gamma(v)} = 0$.

To complete the definition for $\mathcal{T}_t$, we need to define the set of processes $\mathcal{A}_{[0,t]}$. As in [52], let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with normal filtration $\{\mathcal{F}_t\}_{t \geqslant 0}$, and define $\mathcal{A}_{[0,t]}$ to be the set of continuous semi-martingales $X : \Omega \times [0,t] \to M$ such that there exists a Borel measurable drift $\beta_X : [0,t] \times C([0,t]) \to \mathbb{R}^d$ for which

1. $\omega \mapsto \beta_X(s,\omega)$ is $\mathcal{B}(C([0,s]))_+$-measurable for all $s \in [0,t]$, where $\mathcal{B}(C([0,s]))$ is the Borel σ-algbera of $C[0,s]$.
2. $W_X(s) := X(s) - X(0) - \int_0^s \beta_X(s', X)\, \mathrm{d}s'$ is a $\sigma(X(s)\,;\, 0 \leqslant s \leqslant t)$ M-valued Brownian motion.

An adaptation of their proofs to the case of a compact torus yields the following.

Proposition 12.1. *Under the above hypothesis on L, the following assertions hold:*

1. *For each $t > 0$, $\mathcal{T}_t$ is a backward linear transfer with Kantorovich operator T_t, and the family $\{\mathcal{T}_t\}_{t>0}$ is a semi-group of transfers under convolutions.*
2. *For any $\mu, \nu \in \mathcal{P}(M)$ for which $\mathcal{T}_t(\mu,\nu) < \infty$, there exists a minimiser $\bar{X} \in \mathcal{A}_{[0,t]}$ for $\mathcal{T}_t(\mu,\nu)$. For every $f \in C(M)$ and $x \in M$, there exists a maximiser for $T_t f(x)$.*
3. *Fix $t_1 > 0$, and $u \in C(M)$, the function $U(t,x) := T_{t_1 - t} u(x)$ defined for $0 \leqslant t \leqslant t_1$ is the unique viscosity solution of*

$$\frac{\partial U}{\partial t}(t,x) + \frac{1}{2}\Delta_x U(t,x) + H(x, \nabla_x U(t,x)) = 0, \quad (t,x) \in [0,t_1) \times M, \tag{255}$$

with $U(t_1, x) = u(x)$.

4. *If $f \in C^\infty(M)$ and $t > 0$, $U(t',x) := T_{t-t'} f \in C^{1,2}([0,t] \times M)$ and U is a classical solution to the Hamilton-Jacobi-Bellman equation (255). The maximiser $\bar{X}$ satisfies*

$$\beta_{\bar{X}}(s, \bar{X}) = D_p H(\bar{X}(s), D_x U(s, \bar{X}(s))).$$

In order to define the Mané constant $c(\mathcal{T})$ and develop a corresponding Mather theory, we need to establish that there exists a probability measure $\mu \in \mathcal{P}(M)$ such that $\mathcal{T}_1(\mu,\mu) < +\infty$. Such a measure can be obtained as the first marginal of a probability measure m on phase space TM that is flow invariant, that is one that satisfies

$$\int_{TM} A^v\varphi(x)\,\mathrm{d}m(x,v) = 0 \text{ for all } \varphi \in C^2(M) \text{ where } A^v\varphi := \tfrac{1}{2}\Delta\varphi + v\cdot\nabla\varphi. \tag{256}$$

To this end, let $\mathcal{P}_\gamma(M)$ denote the set of probability measures on TM such that

$$\int_{TM} \gamma(v)\,\mathrm{d}m(x,v) < +\infty,$$

and denote by $\mathcal{N}_0$ the class of such probability measures m, that is,

$$\mathcal{N}_0 := \{m \in \mathcal{P}_\gamma(TM)\,;\, \int_{TM} A^v\varphi(x)\,\mathrm{d}m(x,v) = 0 \text{ for all } \varphi \in C^2(M)\}.$$

Proposition 12.2. *The set $\mathcal{N}_0$ of 'flow-invariant' probability measures m on TM is non-empty and*

$$c := \inf\{\mathcal{T}_1(\mu,\mu)\,;\, \mu \in \mathcal{P}(M)\} = \inf\{\int_{TM} L(x,v)\,\mathrm{d}m(x,v);\, m \in \mathcal{N}_0\}. \tag{257}$$

Moreover, the infimum over $\mathcal{N}_0$ is attained by a measure $\bar{m}$, that we call a stochastic Mather measure. Its projection $\mu_{\bar{m}}$ on $\mathcal{P}(M)$ is a minimiser for $\mathcal{T}_1$.

Conversely, every minimizing measure $\bar{\mu}$ of $\mathcal{T}_1(\mu,\mu)$ induces a stochastic Mather measure $m_{\bar{\mu}}$.

Proof: Given $\mu \in \mathcal{P}(M)$, consider $X \in \mathcal{A}_{[0,1]}$ that realises the infimum for $\mathcal{T}_1(\mu,\mu)$, that is

$$\mathcal{T}_1(\mu,\mu) = \mathbb{E}\int_0^1 L(X(s),\beta_X(s,X))\,\mathrm{d}s.$$

Define a probability measure $m = m_\mu \in \mathcal{P}_\gamma(TM)$ via its action on the subset of continuous functions $\psi : TM \to \mathbb{R}$ with $\sup_{(x,v)\in TM}\left|\frac{\psi(x,v)}{\gamma(v)}\right| < +\infty$ and $\lim_{|(x,v)|\to\infty}\frac{\psi(x,v)}{\gamma(v)} \to 0$ via the formula

$$\int_{TM} \psi(x,v)\,\mathrm{d}m(x,v) := \mathbb{E}\int_0^1 \psi(X(s),\beta_X(s,X))\,\mathrm{d}s. \tag{258}$$

We claim that $\int_{TM} A^v\varphi(x)\,\mathrm{d}m(x,v) = 0$ for every $\varphi \in C^2(M)$. Indeed, by the definition of m,

$$\begin{aligned}
\int_{TM} A^v\varphi(x)\,\mathrm{d}m(x,v) &= \mathbb{E}\int_0^1 A^{\beta_X(s,X)}\varphi(X(s))\,\mathrm{d}s\\
&= \mathbb{E}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s}[\varphi(X(s))]\,\mathrm{d}s \quad \text{(Itô's lemma)}\\
&= \mathbb{E}\varphi(X(1)) - \mathbb{E}\varphi(X(0))\\
&= 0, \qquad (X(0)\sim\mu\sim X(1)).
\end{aligned}$$

This implies that $m \in \mathcal{N}_0$, so that

$$\mathcal{T}_1(\mu,\mu) = \mathbb{E}\int_0^1 L(X(s), \beta_X(s,X))\,\mathrm{d}s = \int_{TM} L(x,v)\,\mathrm{d}m(x,v) \geqslant \inf_{m\in\mathcal{N}_0}\int_{TM} L(x,v)\,\mathrm{d}m(x,v), \tag{259}$$

hence

$$\inf_{\mu\in\mathcal{P}(M)} \mathcal{T}_1(\mu,\mu) \geqslant \inf_{m\in\mathcal{N}_0}\int_{TM} L(x,v)\,\mathrm{d}m(x,v).$$

Conversely, suppose $m \in \mathcal{N}_0$, and let $\varphi(x,t)$ be a smooth solution to the Hamilton-Jacobi-Bellman equation. Since $\int_{TM} A^v\varphi(x,t)\,\mathrm{d}m(x,v) = 0$ for every t, it follows that $\mu_m := \pi_M\#m$ satisfies

$$\begin{aligned}
\int_M [\varphi(x,1) - \varphi(x,0)]\,\mathrm{d}\mu_m(x) &= \int_{[0,1]} \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_M \varphi(x,t)\,\mathrm{d}\mu_m\right]\mathrm{d}t \\
&= \int_0^1\int_{TM} \partial_t\varphi(x,t)\,\mathrm{d}m(x,v)\,\mathrm{d}t \\
&= \int_0^1\int_{TM} [v\cdot\nabla\varphi(x,t) - H(x,\nabla_x\varphi(x,t))]\,\mathrm{d}m(x,v)\,\mathrm{d}t.
\end{aligned}$$

Since $H(x,p) := \sup_v\{\langle p,v\rangle - L(x,v)\}$,

$$v\cdot\nabla\varphi(x,t) - H(x,\nabla_x\varphi(x,t)) \leqslant L(x,v),$$

hence combining the above two displays implies

$$\int_M [\varphi(x,1)-\varphi(x,0)]\,\mathrm{d}\mu_m(x) \leqslant \int_{TM} L(x,v)\,\mathrm{d}m(x,v)$$

for every Hamilton-Jacobi-Bellman solution φ on $[0,1)\times M$ with $\varphi(\cdot,1)\in C^\infty(M)$. Taking the supremum over all such solutions φ yields

$$\sup\left\{\int_M [\varphi(x,1)-\varphi(x,0)]\,\mathrm{d}\mu_m(x)\,;\,\varphi(\cdot,1)\in C^\infty(M)\right\} \leqslant \int_{TM} L(x,v)\,\mathrm{d}m(x,v).$$

By duality, $\mathcal{T}_1(\mu_m,\mu_m) = \sup\left\{\int_M[\varphi(x,1)-\varphi(x,0)]\,\mathrm{d}\mu_m(x)\,;\,\varphi(\cdot,1)\in C^\infty(M)\right\}$, so that

$$\mathcal{T}_1(\mu_m,\mu_m) \leqslant \int_{TM} L(x,v)\,\mathrm{d}m(x,v) \tag{260}$$

and therefore $\inf_{\mu\in\mathcal{P}(M)}\mathcal{T}_1(\mu,\mu) \leqslant \int_{TM} L(x,v)\,\mathrm{d}m(x,v)$, and we are done. □

The following summarizes the main asymptotic properties of $\{\mathcal{T}_t\}_{t>0}$.

Proposition 12.3. *Let $\{\mathcal{T}_t\}_{t\geqslant 0}$ be the family of stochastic transfers defined via (253) with associated backward Kantorovich operators $\{T_t\}_{t\geqslant 0}$ given by (254). Let c be the critical value obtained in the last proposition. Then,*

1. *The equation*
$$T_t u + kt = u, \quad t \geqslant 0, \quad u \in C(M),$$
has solutions (the backward weak KAM solutions) if and only if $k = c$.

2. *The backward weak KAM solutions are exactly the viscosity solutions of the stationary Hamilton-Jacobi-Bellman equation*
$$\frac{1}{2}\Delta u + H(x, D_x u) = -c. \tag{261}$$

PROOF. The fact that there are solutions for (261) was established by Gomes [36]. We give a proof based on Proposition 9.4 that clarifies the relationship between such solutions and the notion of backward weak KAM solutions.

Let $\alpha > 0$ and consider
$$u_\alpha(x) := \inf\left\{\mathbb{E}\int_0^{+\infty} e^{-s} L(X(s), \beta_X(s, X))\,\mathrm{d}s\,;\ X \in \mathcal{A}_{[0,t]},\ X(0) = x\right\}.$$
It is well known that one then has
$$u_\alpha(x) = \inf\left\{\mathbb{E}\int_0^{t} e^{-s} L(X(s), \beta_X(s, X))\,\mathrm{d}s + e^{-\alpha t} u_\alpha(X(t))\,;\ X \in \mathcal{A}_{[0,t]},\ X(0) = x\right\},$$
and
$$\alpha u_\alpha - \frac{1}{2}\Delta u_\alpha + H(x, D_x u_\alpha) = 0.$$
It is straightforward to check that this implies that
$$T_t u_\alpha + t\alpha u_\alpha = u_\alpha.$$
Proposition 9.4 applies to get the result with $t = n$. Note that for constructing a viscosity solution for (261), it suffices to find a weak KAM solution for T_1. Indeed, suppose there exists a function $f \in C(M)$ such that $T_1 f(x) + c = f(x)$ for all $x \in M$, we need to show that $T_t f(x) + ct = f(x)$ for all $t > 0$. But note that from the semi-group property, the claim is true for $t = n \in \mathbb{N}$. For other $t > 0$, by writing uniquely $t = n + \alpha$ where $n \in \mathbb{N}$ and $0 \leqslant \alpha < 1$, it then suffices to prove that $T_\alpha f(x) + \alpha c = f(x)$. Note that the function $U(t, x) := T_{1-t} f(x) + c(1 - t)$ satisfies the Hamilton-Jacobi-Bellman equation
$$\begin{cases} \frac{\partial U}{\partial t}(t, x) + \frac{1}{2}\Delta U(t, x) + H_c(x, \nabla U(t, x)) = 0, \quad t \in [0, 1), x \in M \\ U(1, t) = f(x). \end{cases} \tag{262}$$
where $H_c(x, p) := H(x, p) + c$, with the additional property that $U(0, x) = U(1, x)$. We may then apply a comparison result for Hamilton-Jacobi-Bellman (see e.g. [26], Section V.8, Theorem 8.1) to deduce that in fact the condition $U(0, x) = U(1, x)$ implies $U(t, x) = U(1, x)$ for every $t \in [0, 1]$. In particular, at $t = 1 - \alpha$, we deduce that $T_\alpha f(x) + c\alpha = f(x)$.

As to the relationship between 1) and 2) observe that if u is a backward weak KAM solution, then $U(t, x) := T_{t_1 - t} u(x) + c(t_1 - t) = u(x)$ is a viscosity solution to (262) where the final time is t_1. Hence u is a viscosity solution of (261).

Conversely, suppose u is a viscosity solution to (261).Then, $(x, t) \mapsto u(x)$ is a viscosity solution to (262). On the other hand, $T_{t_1 - t} u + c(t_1 - t)$ is also a viscosity solution of (262). By the uniqueness of such solutions, it follows that $T_{t_1 - t} u(x) + c(t_1 - t) = u(x)$. As $t_1 > 0$ is arbitrary, this shows that u is a backward weak KAM solution.

We finish this section with the following characterization of the Mané value, motivated by the work of Fathi [25] in the deterministic case. Let $u \in C(M)$ and $k \in \mathbb{R}$, and say that *u is dominated by* $L - k$ and write $u \prec L - k$ if for every $t > 0$, it holds for every $X \in \mathcal{A}_{[0,t]}$ and every $x \in M$,

$$\mathbb{E}[u(X(t))|X(0) = x] - u(x) \leqslant \mathbb{E}\left[\int_0^t L(X(s), \beta_X(s, X)\,\mathrm{d}s|X(0) = x\right] - kt. \tag{263}$$

Proposition 12.4. *The Mañé critical value satisfies*

$$c = \sup\left\{k \in \mathbb{R} : \exists u \quad \textit{such that} \quad u \prec L - k\right\}.$$

Proof. By the above, there exists a u such that $T_t u + ct = u$, so that by definition of T_t,

$$u(x) - ct \geqslant \mathbb{E}[u(X(t))|X(0) = x] - \mathbb{E}\left[\int_0^t L(X(s), \beta_X(s, X)\,\mathrm{d}s|X(0) = x\right]$$

for every $X \in \mathcal{A}_{[0,t]}$. This shows that $u \prec L - c$, so c is itself admissible in the supremum.

On the other hand, if $k \in \mathbb{R}$ is such that $u \prec L - k$, then it is easy to see that $T_t u(x) \leqslant u(x) - kt$ for all t. In particular, applying T_s and using the linearity of T_s with respect to constants, we find $T_{s+t}u + kt \leqslant T_s u$, and hence

$$T_{s+t}u + k(t + s) \leqslant T_s u + ks$$

So $t \mapsto T_t u + kt$ is decreasing and the result follows from Corollary 11.3.

13 Convex couplings and convex and Entropic 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

$$\alpha^{\oplus}(t) = \sup\{ts - \alpha(s); s \geqslant 0\} \quad \text{resp., } \beta^{\ominus}(t) = \sup\{-ts - \beta(s); s > 0\} \tag{264}$$

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

$$\alpha(t) = \sup\{ts - \alpha^{\oplus}(s); s \geqslant 0\}. \tag{265}$$

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

$$\beta(t) = \sup\{-ts - \beta^{\ominus}(s); s \geqslant 0\}. \tag{266}$$

13.1 Convex couplings

We now give a few examples of convex couplings, which are not necessarily convex transfers.

Proposition 13.1. *Let $\alpha : \mathbb{R}^+ \to \mathbb{R}$ (resp., $\beta : \mathbb{R}^+ \setminus \{0\} \to \mathbb{R}$) be a convex (resp., concave) increasing functions. If $\mathcal{T}$ is a linear backward (resp., forward) transfer with Kantorovich operator T^- (resp., Kantorovich operator T^+), then $\alpha(\mathcal{T})$ is a backward convex (resp., forward convex) coupling associated to a family of (Kantorovich) operators $(T_s^-)_{s\geqslant 0}$ (resp.,$(T_s^+)_{s\geqslant 0}$, where*

$$T_s^- f = sT^-(\frac{f}{s}) - \alpha^{\oplus}(s) \quad (resp.,\ T_s^+ f = sT^+(\tfrac{f}{s}) - \alpha^{\oplus}(s). \tag{267}$$

In particular, for any $p \geqslant 1$, $\mathcal{T}^p$ is a forward (resp., backward) convex coupling.

Proof: It suffices to write

$$\begin{aligned}\alpha(\mathcal{T}(\mu,\nu)) &= \sup\big\{s\int_Y T^+ f\, d\nu - s\int_X f\, d\mu - \alpha^{\oplus}(s);\ s \in \mathbb{R}^+, f \in C(X)\big\}\\ &= \sup\big\{\int_Y sT^+(\frac{h}{s})\, d\nu - \alpha^{\oplus}(s) - \int_X h\, d\mu;\ s \in \mathbb{R}^+, h \in C(X)\big\},\end{aligned}$$

which means that $\alpha(\mathcal{T})$ is a forward convex coupling corresponding to the family of (Kantorvich) operators $T_s^+ f = sT^+(\frac{h}{s}) - \alpha^{\oplus}(s)$.

Example 11.1): A mean-field planning problem (Orrieri-Porretta-Savaré [58])

Let $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a Tonelli Lagrangian and $F : \mathbb{R}^d \times L^\infty([0,T];\mathcal{P}(\mathbb{R}^d)) \to \mathbb{R}$ be a functional that is convex in the second variable, and consider the following mean-field planning problem between two probability measures μ and ν,

$$\mathcal{T}(\mu,\nu) =: \min\left\{\int_0^T \int_{\mathbb{R}^d} L(x,\mathbf{v})\, \rho(t,dx)\, dt + \int_0^T F(x,\rho(t,dx))\ \mathrm{d}t;\ \mathbf{v} \in L^2(\rho(t,dx)\, dt)\right\}, \tag{268}$$

subject to ρ and v satisfying

$$\partial_t \rho + \nabla\cdot(\rho\, \mathbf{v}) = 0, \quad \rho(0,\cdot) = \mu\, , \rho(T,\cdot) = \nu. \tag{269}$$

Then, $\mathcal{T}$ is both a forward and backward convex coupling.

Indeed, following Orrieri-Porretta-Savaré [58], we consider for each $\ell \in C([0,T],\mathbb{R}^d)$ the Kantorovich operator defined on $C(\mathbb{R}^d)$ via

$$T_\ell(u) = u_\ell(T,x) - \iint_Q F^*(x,\ell(t,x))\ \mathrm{d}x$$

where $u_\ell(t,x)$ is a solution of the Hamilton-Jacobi equation

$$\begin{aligned}-\partial_t u + H(x,Du) &= \ell \quad \text{in } Q := (0,T)\times\mathbb{R}^d, &(270)\\ u(0,x) &= u(x). &(271)\end{aligned}$$

and $F^*(x,\ell) = \sup\big\{\langle \ell,\rho\rangle - F(x,\rho); m \in L^\infty([0,T];\mathcal{P}(\mathbb{R}^d))\big\}$.

A standard min-max argument then yields that

$$\begin{aligned}\mathcal{T}(\mu,\nu) &= \sup\left\{\int_{\mathbb{R}^d} T_\ell u\, d\nu - \int_{\mathbb{R}^d} u\, d\mu; u \in C(\mathbb{R}^d), \ell \in C([0,T],\mathbb{R}^d)\right\}\\ &= \sup\left\{\int_{\mathbb{R}^d} u_\ell(T,x) d\nu - \int_{\mathbb{R}^d} u_\ell(0,x)\, d\mu(x) - \iint_Q F^*(x,\ell(t,x))\ \mathrm{d}x;\ u_\ell \text{ solves } (270)\right\}.\end{aligned}$$

Remark 13.2. *Another convex -but only backward- coupling can be defined as*

$$\mathcal{T}(\mu,\nu) =: \min \int_0^T \int_{\mathbb{R}^d} L(x,\mathbf{v})\,\rho(t,dx)\,dt + \int_0^T F(x,\rho(t,dx))\ \mathrm{d}t;\ \mathbf{v} \in L^2(\rho(t,dx)\,dt), \quad (272)$$

subject to ρ and v satisfying

$$\partial_t \rho - \Delta\rho + \nabla\cdot(\rho\,\mathbf{v}) = 0,\ \rho(0,\cdot) = \mu\,,\rho(T,\cdot) = \nu. \quad (273)$$

We do not know whether $\mathcal{T}$ is a convex transfer. This is equivalent to the question whether $\ell \to T_\ell$ is concave, or equivalently whether the map $\ell \to u_\ell(T,x)$ is concave.

Example 11.2: A backward convex coupling which is not a convex transfer

Let $\Omega \subset \mathbb{R}^d$ be a Borel measurable subset with $1 < |\Omega| < \infty$, $\lambda := \frac{1}{|\Omega|}$, and define for any two given probability measures μ, ν on Ω, the correlation,

$$\mathcal{T}_\lambda(\mu,\nu) = \begin{cases} 0 & \text{if } \nu \in \mathcal{C}_\lambda(\mu) \\ +\infty & \text{otherwise,} \end{cases} \quad (274)$$

where $\mathcal{C}_\lambda(\mu) := \{\nu \in \mathcal{P}(\Omega)\,;\, \lambda \left|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right| \leqslant 1\,\mu\text{-a.e.}\}$. Note that when $\mu = \lambda\,\mathrm{d}x|_\Omega$ (the uniform measure on Ω),

$$\mathcal{T}_\lambda(\lambda\,\mathrm{d}x|_\Omega,\nu) = \begin{cases} 0 & \text{if } \left|\frac{\mathrm{d}\nu}{\mathrm{d}x}\right| \leqslant 1 \text{ Lebesgue-a.e.} \\ +\infty & \text{otherwise.} \end{cases} \quad (275)$$

We claim that $\mathcal{T}_\lambda$ is a backward convex coupling but not a convex transfer. Indeed, for the first claim, consider $\alpha_m(t) := (\lambda t)^m \log(\lambda t)$ for $m \geqslant 1$ and $t \geqslant 0$, and define

$$\mathcal{T}_m(\mu,\nu) := \begin{cases} \int_\Omega \alpha_m\left(\left|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right|\right)\,\mathrm{d}\mu, & \text{if } \nu << \mu, \\ +\infty & \text{otherwise.} \end{cases} \quad (276)$$

By Example 11.1, $\mathcal{T}_m$ is a backward convex transfer and

$$(\mathcal{T}_{m,\mu})^*(f) = \inf\{\int_\Omega [\alpha_m^\oplus(f(x)+t) - t]\,\mathrm{d}\mu(x)\,;\, t \in \mathbb{R}\}. \quad (277)$$

The function $\alpha_m^\oplus$ can be explicitly computed as

$$\alpha_m^\oplus(t) = \begin{cases} e^{-1+\frac{1}{m-1}W(\beta_m t)}\left[\beta_m t + \frac{1}{m} e^{W(\beta_m t)}\right] & \text{if } t \geqslant -\frac{\lambda}{m-1}e^{-1}, \\ 0 & \text{if } t < -\frac{\lambda}{m-1}e^{-1}. \end{cases} \quad (278)$$

where $\beta_m := \frac{m-1}{\lambda m} e^{\frac{m-1}{m}}$, and W is the *Lambert-W* function. It is easy to see that $\mathcal{T}_\lambda(\mu,\nu) = \sup_m \mathcal{T}_m(\mu,\nu)$; hence it is a backward convex coupling (as a supremum of backward convex transfers).

However, $\mathcal{T}_\lambda$ is not a backward convex transfer, since

$$(\mathcal{T}_{\lambda,\mu})^*(f) = (\sup_m \mathcal{T}_{m,\mu})^*(f) \leqslant \inf_m \mathcal{T}^*_{m,\mu}(f) = \int \frac{f}{\lambda}\,\mathrm{d}\mu,$$

with the inequality being in general strict.

Note that this also implies that the Wasserstein projection on the set $\mathcal{C}_\mu$, that is

$$W_2^2(P_1[\nu],\nu)=\inf\{W_2^2(\sigma,\nu);\left|\frac{\mathrm{d}\sigma}{\mathrm{d}x}\right|\leqslant 1\}=\inf\{\mathcal{T}_\lambda(\lambda\,\mathrm{d}x|_\Omega,\sigma)+W_2^2(\sigma,\nu)\,;\,\sigma\in\mathcal{P}(\Omega)\}\quad(279)$$

is in fact an inf-convolution of a backward convex coupling $\mathcal{T}_\lambda$ with the linear transfer W_2^2, and no duality formula can then be extracted.

13.2 Convex and entropic transfers

Proposition 13.3. *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 $\mathcal{E}$ is a β-entropic backward transfer with Kantorovich operator E^-, then it is a backward convex transfer with Kantorovich family $(T_s^-)_{s>0}$ given by*

$$T_s^-f=sT^-f+(-\beta)^\ominus(s).\quad(280)$$

2. *Similarly, if $\mathcal{E}$ is an α-entropic forward transfer with Kantorovich operator E^+, then it is a forward convex transfer with Kantorovich family $(T_s^+)_{s>0}$ given by*

$$T_s^+f=sT^+f-\alpha^\oplus(s).\quad(281)$$

Proof: Use the fact that $(-\beta)$ is convex decreasing to write that for any $g\in C(Y)$,

$$\beta(\int_X T^-g\,d\mu)=\inf\{s\int_X T^-g\,d\mu+(-\beta)^\ominus(s);s>0\},$$

hence $\mathbb{E}$ is a backward convex transfer with Kantorovich family given by $T_s^-f=sT^-f+(-\beta)^\ominus(s)$.

Example 11.4: General entropic functionals are convex transfers

Consider the following *generalized entropy,*

$$\mathcal{E}_\alpha(\mu,\nu)=\int_X\alpha(|\frac{d\nu}{d\mu}|)\,d\mu,\quad\text{if }\nu<<\mu\text{ and }+\infty\text{ otherwise,}\quad(282)$$

where α is any strictly convex lower semi-continuous superlinear (i.e., $\lim\limits_{t\to+\infty}\frac{\alpha(t)}{t}=+\infty$) real-valued function on $\mathbb{R}^+$. It is then easy to show [39] that

$$(\mathcal{E}_\alpha)^*_\mu(f)=\inf\{\int_X[\alpha^\oplus(f(x)+t)-t]\,d\mu(x);t\in\mathbb{R}\},\quad(283)$$

In other words, $\mathcal{E}_\alpha$ is a backward convex transfer with Kantorovich family

$$T_t^-f(x)=\alpha^\ominus(f(x)+t)-t.$$

Example 11.5: The logarithmic entropy is a log-entropic backward transfer

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

$$\mathcal{H}(\mu,\nu) := \int_X \log(\tfrac{d\nu}{d\mu})\, d\nu \text{ if } \nu << \mu \text{ and } +\infty \text{ otherwise.}$$

It can also be written as

$$\mathcal{H}(\mu,\nu) := \int_X h(\tfrac{d\nu}{d\mu})\, d\mu \text{ if } \nu << \mu \text{ and } +\infty \text{ otherwise,}$$

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

$$\mathcal{H}^*_\mu(f) = \inf\{\int_X (e^t e^{f(x)} - 1 - t)\, d\mu(x); t \in \mathbb{R}\} = \log \int_X e^f\, d\mu.$$

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

$$\log \int_X e^f\, d\mu = \inf\{s \int_X e^f\, d\mu + \beta^\ominus(s); s > 0\}.$$

In other words, it is a backward convex transfer with Kantorovich family $T^-_s f = se^f + \beta^\ominus(s)$ where $s > 0$.

Example 11.6: The Fisher-Donsker-Varadhan information is a backward convex transfer [23]

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

$$\mathcal{I}(\mu|\nu) := \begin{cases} \mathcal{E}(\sqrt{f}, \sqrt{f}), & \text{if } \nu = f\mu, \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty, & \text{otherwise.} \end{cases} \tag{284}$$

Note that when (P_t) is μ-symmetric, $\nu \mapsto I(\mu|\nu)$ is exactly the Donsker-Varadhan entropy i.e. the rate function governing the large deviation principle of the empirical measure $L_t := \frac{1}{t}\int_0^t \delta_{X_s} ds$ for large time t. The corresponding Feynman-Kac semigroup on $L^2(\mu)$

$$P^u_t g(x) := \mathbb{E}^x g(X_t) \exp\left(\int_0^t u(X_s)\, ds\right). \tag{285}$$

It has been proved in [65] that $\mathcal{I}^*_\mu(f) = \log \|P^f_1\|_{L^2(\mu)}$, which yields that $\mathcal{I}$ is a backward convex transfer.

$$\mathcal{I}^*_\mu(f) = \log \|P^f_1\|_{L^2(\mu)} = \frac{1}{2}\log \|P^f_1\|^2_{L^2(\mu)} = \frac{1}{2}\log \sup\{\int |P^f_1 g|^2\, d\mu; \|g\|_{L^2(\mu)} \leqslant 1\}.$$

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

$$\begin{aligned}
\mathcal{I}(\mu,\nu) &= \sup\{\int_Y f\,d\nu - \frac{1}{2}\log\sup\{\int_X |P_1^f g|^2\,d\mu; \|g\|_{L^2(\mu)} \leqslant 1\}; f \in C(X)\} \\
&= \sup\{\int_Y f\,d\nu + \sup_{s>0}\sup_{\|g\|_{L^2(\mu)}\leqslant 1}\frac{1}{2}\{\int_X (-s|P_1^f g|^2 - \beta^{\ominus}(s))\,d\mu\}; f \in C(X)\} \\
&= \sup\{\int_Y f\,d\nu - \inf_{s>0}\inf_{\|g\|_{L^2(\mu)}\leqslant 1}\frac{1}{2}\{\int_X (s|P_1^f g|^2 + \beta^{\ominus}(s))\,d\mu\}; f \in C(X)\} \\
&= \sup\{\int_Y f\,d\nu - \int_X T_{s,g}^- f\,d\mu\,; s \in \mathbb{R}^+, \|g\|_{L^2(\mu)} \leqslant 1, f \in C(X)\}.
\end{aligned}$$

Hence, it is a backward convex transfer, with Kantorovich family $(T_{s,g}^-)_{s,g}$ defined by $T_{s,g}^- f = \frac{s}{2}|P_1^f g|^2 + \frac{1}{2}\beta^{\ominus}(s)$.

13.3 Operations on convex and entropic transfers

The class of backward (resp., forward) convex couplings and transfers 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 13.4. *Let $\mathcal{F}$ be a backward convex coupling (resp., transfer) with Kantorovich family $(F)_i^-$, Then,*

1. *If $a \in \mathbb{R}^+ \setminus \{0\}$, then $a\mathcal{F}$ is a backward convex coupling (resp., transfer) with Kantorovich family given by $F_{a,i}^-(f) = aF_i^-(\frac{f}{a})$.*
2. *If $\mathcal{T}$ is a backward linear transport on $Y \times Z$ with Kantorovich operator T^-, and $\mathcal{F}$ is a backward convex transfer, then $\mathcal{F}\star\mathcal{T}$ is a backward convex transfer with Kantorovich family given by $F_i^- \circ T^-$.*

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

$$\begin{aligned}
(\mathcal{F}\star T)_\mu^*(g) &= \sup_{\nu\in\mathcal{P}(Z)}\sup_{\sigma\in\mathcal{P}(Y)}\left\{\int_Z g\,d\nu - \mathcal{F}(\mu,\sigma) - \mathcal{T}(\sigma,\nu)\right\} \\
&= \sup_{\sigma\in\mathcal{P}(Y)}\{\mathcal{T}_\sigma^*(g) - \mathcal{F}(\mu,\sigma)\} \\
&= \sup_{\sigma\in\mathcal{P}(Y)}\left\{\int_Y T^- g\,d\sigma - \mathcal{F}(\mu,\sigma)\right\} \\
&= (\mathcal{F})_\mu^*(T^-(g)) \\
&= \inf_{i\in I}\int_X F_i^- \circ T^- g(x))\,d\mu(x).
\end{aligned}$$

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

Proposition 13.5. Let $\beta : \mathbb{R} \to \mathbb{R}$ be a concave increasing function and let $\mathcal{E}$ be a backward β-entropic transfer with Kantorovich operator E^-. Then,

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

2. $\tilde{\mathcal{E}}$ is a forward $((-\beta)^\ominus)^\oplus$-entropic transfer with Kantorovich operator $\tilde{E}^+h = -E^-(-h)$.

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

$$\mathcal{E} \star \mathcal{T}\,(\mu, \nu) = \sup\big\{ \int_Z g(y)\, d\nu(y) - \beta(\int_X E^- \circ T^- g(x))\, d\mu(x));\ g \in C(Z)\big\}. \quad (286)$$

Proof: 1) is trivial. For 2) note that since β is concave and increasing, then

$$\begin{aligned}
\tilde{\mathcal{T}}(\nu, \mu)) &= \mathcal{T}(\mu, \nu)) \\
&= \sup\{\int_Y g\, d\nu - \beta\big(\int_X T^- g\, d\mu); g \in C(Y)\} \\
&= \sup\{\int_Y g\, d\nu + \sup_{s>0}\{\int_X -sT^- g\, d\mu - (-\beta)^\ominus(s)\}; g \in C(X)\} \\
&= \sup\{\int_Y g\, d\nu - s\int_X T^- g\, d\mu - (-\beta)^\ominus(s); s > 0, g \in C(X)\} \\
&= \sup\{s\int_X -T^-(-h)\, d\mu - (-\beta)^\ominus(s) - \int_Y h\, d\nu; s > 0, g \in C(X) \\
&= \sup\{((-\beta)^\ominus)^\oplus(\int_X -T^-(-h)\, d\mu) - \int_Y h\, d\nu; s > 0, h \in C(X)\}.
\end{aligned}$$

In other words, $\tilde{\mathcal{T}}$ is a $(\beta^\ominus)^\oplus$-entropic forward transfer.
For 3) we calculate the Legendre dual of $(\mathcal{E} \star T)_\mu$ at $g \in C(Z)$ and obtain,

$$\begin{aligned}
(\mathcal{E} \star T)^*_\mu(g) &= \sup_{\nu\in\mathcal{P}(Z)} \sup_{\sigma\in\mathcal{P}(Y)} \Big\{\int_Z g\, d\nu - \mathcal{E}(\mu, \sigma) - \mathcal{T}(\sigma, \nu)\Big\} \\
&= \sup_{\sigma\in\mathcal{P}(Y)} \{\mathcal{T}^*_\sigma(g) - \mathcal{E}(\mu, \sigma)\} \\
&= \sup_{\sigma\in\mathcal{P}(Y)} \Big\{\int_Y T^- g\, d\sigma - \mathcal{E}(\mu, \sigma)\Big\} \\
&= (\mathcal{E})^*_\mu(T^-(g)) \\
&= \beta\Big(\int_X E^- \circ T^- g(x))\, d\mu(x)).
\end{aligned}$$

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

$$\mathcal{T} \star \mathcal{E}\,(\mu, \nu) = \sup\big\{\alpha\big(\int_Y E^+ \circ T^+ f(y))\, d\nu(y)\big) - \int_X f(x)\, d\mu(x);\ f \in C(X)\big\}. \quad (287)$$

13.4 Subdifferentials of linear and convex transfers

If $\mathcal{T}$ is a linear transfer, then both $\mathcal{T}_\mu$ and $\mathcal{T}_\nu$ are convex weak* lower semi-continuous and one can therefore consider their (weak*) subdifferential $\partial\mathcal{T}_\mu$ (resp., $\partial\mathcal{T}_\nu$) in the sense of convex analysis. In other words,

$$g \in \partial\mathcal{T}_\mu(\nu) \text{ if and only if } \mathcal{T}(\mu,\nu') \geqslant \mathcal{T}(\mu,\nu) + \textstyle\int_Y g\,\mathrm{d}(\nu'-\nu) \quad \text{for any } \nu' \in \mathcal{P}(Y).$$

In other words, $g \in \partial\mathcal{T}_\mu(\nu)$ if and only if $\mathcal{T}_\mu(\nu) + \mathcal{T}_\mu^*(g) = \langle g, \nu\rangle$. Since $\mathcal{T}_\mu(\nu) = \mathcal{T}(\mu,\nu)$ and $\mathcal{T}_\mu^*(g) = \int T^- g\,\mathrm{d}\mu$, we then obtain the following characterization of the subdifferentials.

Proposition 13.6. *Let $\mathcal{T}$ be a backward (resp., forward) linear transfer. Then the subdifferential of $\mathcal{T}_\mu : \mathcal{P}(Y) \to \mathbb{R}\cup\{+\infty\}$ at $\nu \in \mathcal{P}(Y)$ (resp., $\mathcal{T}_\nu : \mathcal{P}(X) \to \mathbb{R}\cup\{+\infty\}$ at $\mu \in \mathcal{P}(X)$) is given by*

$$\partial\mathcal{T}_\mu(\nu) = \left\{ g \in C(Y) \,:\, \int_Y g(y)\,\mathrm{d}\nu(y) - \int_X T^- g(x)\,\mathrm{d}\mu(x) = \mathcal{T}(\mu,\nu)\right\} \tag{288}$$

respectively,

$$\partial\mathcal{T}_\nu(\mu) = \left\{ f \in C(X) \,:\, \int_Y T^+ f(y)\,\mathrm{d}\nu(y) - \int_X f(x)\,\mathrm{d}\mu(x) = \mathcal{T}(\mu,\nu)\right\} \tag{289}$$

In other words, the subdifferential of $\mathcal{T}_\mu$ at ν (resp., $\mathcal{T}_\nu$ at μ) is exactly the set of maximisers for the dual formulation of $\mathcal{T}(\mu,\nu)$.

It is easy to see that the same expressions hold - with the necessary modifications - for backward convex (resp., forward) transfers, as well as backward β-entropic (resp., forward α-entropic) transfers.

In the following, we observe some elementary consequences for elements in the subdifferential.

Proposition 13.7. *Suppose $\mathcal{T}$ is a linear backward transfer such that the Dirac masses are contained in $D_1(\mathcal{T})$. Fix $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Then, there exists $\bar\pi \in \mathcal{K}(\mu,\nu)$ such that for each $\bar f \in \partial\mathcal{T}_\mu(\nu)$, we have*

$$T^-\bar f(x) = \int_Y \bar f(y)\,\mathrm{d}\bar\pi_x(y) - \mathcal{T}(x,\bar\pi_x), \quad \text{for } \mu\text{-a.e. } x \in X, \tag{290}$$

where $\bar\pi_x$ is a disintegration of $\bar\pi$ with w.r.t. μ.

Conversely, if $\nu \mapsto \mathcal{T}(\mu,\nu)$ is strictly convex and $\bar f \in \partial\mathcal{T}_\mu(\nu)$ for some $\nu \in \mathcal{P}(Y)$. If $x \to \sigma_x$ is any selection such that

$$T^-\bar f(x) = \sup_\sigma\left\{\int \bar f\,\mathrm{d}\sigma - \mathcal{T}(\delta_x,\sigma)\right\} = \int_Y \bar f\,\mathrm{d}\sigma_x - \mathcal{T}(\delta_x,\sigma_x),$$

then $\mathcal{T}(\mu,\nu)$ is attained by the measure $\bar\pi = \int_X \sigma_x d\mu(x)$.

Proof: By a recent result [5], there exists $\bar\pi \in \mathcal{K}(\mu,\nu)$ such that

$$\mathcal{T}(\mu,\nu) = \int_X \mathcal{T}(x,\bar\pi_x)\,\mathrm{d}\mu(x).$$

If $\bar{f} \in \partial \mathcal{T}_\mu(\nu)$, then by definition

$$\int_Y \bar{f}(y)\,\mathrm{d}\nu(y) - \int_X T^- \bar{f}(x)\,\mathrm{d}\mu(x) = \mathcal{T}(\mu,\nu) = \int_X \mathcal{T}(x, \bar{\pi}_x)\,\mathrm{d}\mu(x),$$

that is $\int_X \left[T^-\bar{f}(x) - \int_Y \bar{f}(y)\,\mathrm{d}\bar{\pi}_x(y) + \mathcal{T}(x,\bar{\pi}_x)\right]\,\mathrm{d}\mu = 0$. Since $T^-\bar{f}(x) = \sup_\sigma \left\{\int \bar{f}\,\mathrm{d}\sigma - \mathcal{T}(x,\sigma)\right\}$, the quantity in the brackets is non-negative and we get our claim.

Conversely, If $\bar{f} \in \partial\mathcal{T}_\mu(\nu)$ is non-empty for some $\nu \in \mathcal{P}(Y)$, then $\int \bar{f}\,\mathrm{d}\nu - \int T^-\bar{f}\,\mathrm{d}\mu = \mathcal{T}(\mu,\nu)$. From the expression $T^-\bar{f}(x) = \sup_\sigma \left\{\int_Y \bar{f}\,\mathrm{d}\sigma - \mathcal{T}(\delta_x,\sigma)\right\}$, we know the supremum will be achieved by some σ_x. Defining $\tilde{\pi}$ by $\mathrm{d}\tilde{\pi}(x,y) = \mathrm{d}\mu(x)\,\mathrm{d}\sigma_x(y)$, and the right marginal of $\tilde{\pi}$ by $\tilde{\nu}$, we integrate against μ to achieve

$$\int T^-\bar{f}\,\mathrm{d}\mu = \int \bar{f}\,\mathrm{d}\tilde{\nu} - \int \mathcal{T}(\delta_x,\sigma_x)\,\mathrm{d}\mu.$$

This shows that $\mathcal{T}(\mu,\tilde{\nu}) = \inf_{\pi\in\Gamma(\mu,\tilde{\nu})} \int \mathcal{T}(\delta_x,\pi_x)\,\mathrm{d}\mu = \int \mathcal{T}(\delta_x,\sigma_x)\,\mathrm{d}\mu$, and consequently, $\bar{f} \in \partial\mathcal{T}_\mu(\tilde{\nu})$. But by strict convexity, this can only be true if $\tilde{\nu} = \nu$. □

While the attainment in the primal problem $\mathcal{T}(\mu,\nu)$ holds in full generality as shown in [5], the attainment in the dual problem depends heavily on the problem at hand [31]. However, since this is equivalent to the sub-differentiability of the partial functional $\mathcal{T}_\mu$, we can use general existence results such as the Brondsted-Rockafellar theorem [54], to state that $\partial\mathcal{T}_\mu(\nu)$ exist for a weak*-dense set of $\nu \in \mathcal{P}(Y)$, and therefore the dual problem is generically attained.

Corollary 13.8. *Suppose $\mathcal{T}$ is a linear backward transfer on $\mathcal{P}(X)\times\mathcal{P}(Y)$ such that the Dirac masses are contained in $D_1(\mathcal{T})$. Assume Y is metrizable. Fix $\mu \in \mathcal{P}(X)$, then for every $\nu \in \mathcal{P}(Y)$ and every $\epsilon > 0$, there exists $\nu_\epsilon \in \mathcal{P}(Y)$ such that $W_2(\nu,\nu_\epsilon) < \epsilon$ and the dual problem for $\mathcal{T}(\mu,\nu_\epsilon)$ is attained.*

The following can be seen as Euler-Lagrange equations for variational problems on spaces of measures, and follows closely [27].

Proposition 13.9. *Let $\mathcal{T}_\alpha(\mu,\nu) := \int_X \alpha\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)\mathrm{d}\mu$ be the generalised entropy transfer considered in Example 11.1, and let $\mathcal{T}$ be any linear backward transfer. For a fixed μ, consider the functional $I_\mu(\nu) := \mathcal{T}_\alpha(\mu,\nu) - \mathcal{T}(\mu,\nu)$, and assume $\bar{\nu}$ realises $\inf_{\nu\in\mathcal{P}(X)} I_\mu(\nu)$. Then, there exists $\bar{f} \in \partial\mathcal{T}_\mu(\bar{\nu})$ such that the following Euler-Lagrange equation holds for $\bar{\nu}-$a.e. $x \in X$,*

$$\alpha'\left(\frac{\mathrm{d}\bar{\nu}}{\mathrm{d}\mu}\right) = \bar{f} + C,$$

where C is a constant.

If $\mathcal{T}_\alpha$ is replaced with the logarithmic entropic transfer $\mathcal{H}(\mu,\nu) = \int \log(\frac{\mathrm{d}\nu}{\mathrm{d}\mu})\,\mathrm{d}\nu$, then

$$\log\left(\frac{\mathrm{d}\bar{\nu}}{\mathrm{d}\mu}\right) = \bar{f} + C.$$

Proof: Recall that $\mathcal{T}_\alpha(\mu,\nu) := \int_X \alpha(|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}|)\,\mathrm{d}\mu$ if $\nu << \mu$ (and $+\infty$ otherwise) is a backward convex transferwith

$$\mathcal{T}_\mu^*(f) = \inf\left\{\int_X [\alpha^\oplus(f(x)+t) - t]\,\mathrm{d}\mu(x)\,;\, t \in \mathbb{R}\right\},$$

where $T_t^- f(x) := \alpha^{\ominus}(f(x)+t) - t$. are the corresponding Kantorovich transfers. Here $\alpha \in C^1$, is strictly convex and superlinear. It follows that

$$\alpha'(|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}|) \in \partial \mathcal{T}_\mu(\nu).$$

We can see this either directly from the subdifferential definition, or from observing

$$\alpha^{\oplus}(\alpha'(|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}|)) = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\alpha'(|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}|) - \alpha(|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}|).$$

In particular,

$$\mathcal{T}_\mu^*\left(\alpha'(|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}|)\right) = \int_X \alpha^{\oplus}\left(\alpha'(|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}|)\right)\mathrm{d}\mu.$$

The rest is an easy adaptation of Theorem 2.2 in [27].

14 Inequalities between transfers

Let $\mathcal{T}$ be a linear or convex coupling, 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

$$\mathcal{T}(\sigma,\mu) \leqslant \lambda_1 \mathcal{E}_1(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{291}$$

$$\mathcal{T}(\mu,\sigma) \leqslant \lambda_2 \mathcal{E}_2(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{292}$$

$$\mathcal{T}(\sigma_1,\sigma_2) \leqslant \lambda_1 \mathcal{E}_1(\sigma_1,\mu) + \lambda_2 \mathcal{E}_2(\sigma_2,\mu) \quad \text{for all } \sigma_1,\sigma_2 \in \mathcal{P}(X), \tag{293}$$

where μ is a fixed measure, and λ_1, λ_2 are two positive reals. In our terminology, Problem 291 (resp., 292), (resp., 293) amount to find μ, λ_1, and λ_2 such that

$$(\lambda_1 \mathcal{E}_1) \star (-\mathcal{T})\,(\mu,\mu) \geqslant 0, \tag{294}$$

$$(\lambda_2 \mathcal{E}_2) \star (-\tilde{\mathcal{T}})\,(\mu,\mu) \geqslant 0, \tag{295}$$

$$(\lambda_1 \tilde{\mathcal{E}}_1) \star (-\mathcal{T}) \star (\lambda_2 \mathcal{E}_2)\,(\mu,\mu) \geqslant 0, \tag{296}$$

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

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

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

14.1 Backward convex coupling to backward convex transfer inequalities

We would like to prove inequalities such as

$$\mathcal{F}_2(\sigma,\mu) \leqslant \mathcal{F}_1(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{297}$$

where $\mathcal{F}_1$ is a backward convex transfer and $\mathcal{F}_2$ is a backward convex coupling. We then apply it to Transport-Entropy inequalities of the form

$$\mathcal{F}(\sigma,\mu) \leqslant \lambda\mathcal{E} \star \mathcal{T}(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{298}$$

where $\mathcal{F}$ is a backward convex coupling, while $\mathcal{E}$ is a β-entropic transfer and $\mathcal{T}$ is a backward linear transfer.

Proposition 14.1. *Let $\mathcal{F}_1$ be a backward 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 coupling on $X_2 \times X_3$ with Kantorovich operator $(F^-_{2,j})_{j\in J}$.*

1. *The following duality formula hold:*

$$\mathcal{F}_1 \star -\mathcal{F}_2\,(\mu,\nu) = \inf_{f\in C(X_3)} \inf_{j\in J} \sup_{i\in I} \left\{ -\int_{X_1} F^-_{1,i} \circ F^-_{2,j} f\, d\mu - \int_{X_3} f\, d\nu \right\}. \tag{299}$$

2. *If $\mathcal{F}_1$ is a backward β-entropic transfer on $X_1 \times X_2$ with Kantorovich operator E^-_1, then*

$$\mathcal{F}_1 \star -\mathcal{F}_2\,(\mu,\nu) = \inf_{f\in C(X_3)} \inf_{j\in J} \left\{ -\beta(\int_{X_1} E^-_1 \circ F^-_{2,j} f\, d\mu) - \int_{X_3} f\, d\nu \right\}. \tag{300}$$

Proof: Write

$$\begin{aligned}
\mathcal{F}_1 \star -\mathcal{F}_2\,(\mu,\nu) &= \inf\{\mathcal{F}_1(\mu,\sigma) - \mathcal{F}_2(\sigma,\nu);\, \sigma \in \mathcal{P}(X_2)\} \\
&= \inf_{\sigma\in\mathcal{P}(X_2)} \left\{ \mathcal{F}_1(\mu,\sigma) - \sup_{f\in C(X_3)} \sup_{j\in J} \left\{ \int_{X_3} f\, d\nu - \int_{X_2} F^-_{2,j} f\, d\sigma \right\} \right\} \\
&= \inf_{\sigma\in\mathcal{P}(X_2)} \inf_{f\in C(X_3)} \inf_{j\in J} \left\{ \mathcal{F}_1(\mu,\sigma) - \int_{X_3} f\, d\nu + \int_{X_2} F^-_{2,j} f\, d\sigma \right\} \\
&= \inf_{f\in C(X_3)} \inf_{j\in J} \left\{ -\sup_{\sigma\in\mathcal{P}(X_2)} \{-\int_{X_2} F^-_{2,j} f\, d\sigma - \mathcal{F}_1(\mu,\sigma)\} - \int_{X_3} f\, d\nu \right\} \\
&= \inf_{f\in C(X_3)} \inf_{j\in J} \left\{ -(\mathcal{F}_1)^*_\mu(-F^-_{2,j} f) - \int_{X_3} f\, d\nu \right\} \\
&= \inf_{f\in C(X_3)} \inf_{j\in J} \left\{ -\inf_{i\in I} \int_{X_1} F^-_{1,i} \circ -F^-_{2,j} f\, d\mu - \int_{X_3} f\, d\nu \right\} \\
&= \inf_{f\in C(X_3)} \inf_{j\in J} \sup_{i\in I} \left\{ -\int_{X_1} F^-_{1,i} \circ -F^-_{2,j} f\, d\mu - \int_{X_3} f\, d\nu \right\}.
\end{aligned}$$

2) If $\mathcal{F}_1$ is a backward β-entropic transfer on $X_1 \times X_2$ with Kantorovich operator E^-_1, then use in the above calculation that $(\mathcal{F}_1)^*_\mu(-F^-_{2,j} f) = \beta(\int_{X_1} E^-_1 \circ -F^-_{2,j} f\, d\mu)$.

Corollary 14.2. *Let $\mathcal{F}$ be a backward convex coupling on $Y_2 \times X_2$ with Kantorovich family $(F_i^-)_{i\in I}$ and let $\mathcal{E}$ be a backward β-entropic 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 $\mathcal{F}(\sigma,\nu) \leqslant \lambda\, \mathcal{E} \star \mathcal{T}\,(\mu,\sigma)$.*
2. *For all $g \in C(X_2)$ and $i \in I$, we have $\beta\big(\int_{X_1} E^- \circ T^- \circ \frac{-1}{\lambda} F_i^-(\lambda g)\, d\mu\big) + \int_{X_2} g\, d\nu \leqslant 0$.*

In particular, if we apply the above in the case where $\mathcal{E}$ is the logarithmic entropy, that is

$$\mathcal{H}(\mu,\nu) = \int_X \log(\tfrac{d\nu}{d\mu})\, d\nu \text{ if } \nu << \mu \text{ and } +\infty \text{ otherwise,} \tag{301}$$

which is a backward β-entropic transfer with $\beta(t) = \log t$ and $E^- f = e^f$ as a backward Kantorovich operator.

Corollary 14.3. *Let $\mathcal{F}$ be a backward convex coupling on $X_2 \times Y_2$ with Kantorovich family $(F_i^-)_{i\in I}$ and let $\mathcal{E}$ be a backward β-entropic 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 $\sigma \in \mathcal{P}(Y)$, we have $\mathcal{F}(\sigma,\nu) \leqslant \lambda\, \mathcal{H} \star \mathcal{T}\,(\mu,\sigma)$.*
2. *For all $g \in C(X_2)$, we have $\sup\limits_{i\in I} \int_{X_1} e^{T^- \circ \frac{-1}{\lambda} F_i^-(\lambda g)}\, d\mu \leqslant e^{-\int_{X_2} g\, d\nu}$.*

In particular, if $\mathcal{T}$ is the identity transfer and $\mathcal{F}$ is a backward linear transfer, then the following are equivalent:

1. *$\mathcal{F}(\sigma,\nu) \leqslant \lambda\, \mathcal{H}\,(\sigma,\mu)$ for all $\sigma \in \mathcal{P}(Y)$*
2. *$\int_{X_1} e^{-F^-(\lambda g)}\, d\mu \leqslant e^{-\frac{1}{\lambda}} e^{-\int_{X_2} g\, d\nu}$ for all $g \in C(X_2)$.*

14.2 Forward convex coupling to backward convex transfer inequalities

We are now interested in inequalities such as

$$\mathcal{F}_2(\nu,\sigma) \leqslant \mathcal{F}_1(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{302}$$

where both $\mathcal{F}_1$ and $\mathcal{F}_2$ are convex backward transfers, and in particular, Transport-Entropy inequalities of the form

$$\mathcal{F}(\nu,\sigma) \leqslant \lambda \mathcal{E} \star \mathcal{T}(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{303}$$

where $\mathcal{E}$ is a β-entropic transfer and $\mathcal{T}$ is a backward linear transfer. But we can write (304) as

$$\tilde{\mathcal{F}}_2(\sigma,\nu) \leqslant \mathcal{F}_1(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{304}$$

where now $\tilde{\mathcal{F}}_2(\sigma,\nu) = \mathcal{F}_2(\nu,\sigma)$ is a convex forward transfer. So, we need to establish the following type of duality.

Proposition 14.4. *Let $\mathcal{F}_1$ be a backward convex transfer with Kantorovich operator $(F^-_{1,i})_{i\in I}$ on $X_1\times X_2$, and let $\mathcal{F}_2$ be a convex forward coupling on $X_2\times X_3$ with Kantorovich operator $(F^+_{2,j})_{j\in J}$.*

1. *The following duality formula then holds:*

$$\mathcal{F}_1\star -\mathcal{F}_2\,(\mu,\nu)=\inf_{g\in C(X_2)}\inf_{j\in J}\sup_{i\in I}\Big\{-\int_{X_1}F^-_{1,i}(-g)\,d\nu-\int_{X_3}F^+_{2,j}(g)\,d\nu\Big\}. \tag{305}$$

2. *If $\mathcal{F}_1$ is a backward β-entropic transfer on $X_1\times X_2$ with Kantorovich operator E_1^-, then*

$$\mathcal{F}_1\star -\mathcal{F}_2\,(\mu,\nu)=\inf_{g\in C(X_2)}\inf_{j\in J}\Big\{-\beta(\int_{X_1}E_1^-(-g)\,d\mu)-\int_{X_3}F_j^+(g)\,d\nu\Big\}. \tag{306}$$

3. *If $\mathcal{F}_1$ is a backward β-entropic transfer with Kantorovich operator E_1^-, and $\mathcal{F}_2$ is a forward α-entropic transfer with Kantorovich operator E_2^+, then*

$$\mathcal{F}_1\star -\mathcal{F}_2\,(\mu,\nu)=\inf_{g\in C(X_2)}\Big\{-\beta(\int_{X_1}E_1^-(-g)\,d\mu)-\alpha(\int_{X_3}E_2^+g\,d\nu)\Big\}. \tag{307}$$

4. *In particular, if $\mathcal{E}$ is a backward β-entropic transfer with Kantorovich operator E^-, and $\mathcal{T}$ is a forward linear transfer with Kantorovich operator T^+, then*

$$\mathcal{E}\star -\mathcal{T}\,(\mu,\nu)=\inf_{g\in C(X_2)}\Big\{-\beta(\int_{X_1}E^-(-g)\,d\mu)-\int_{X_3}T^+g\,d\nu\Big\}. \tag{308}$$

Proof: 1) Assume $\mathcal{F}_1$ is a backward convex transfer with Kantorovich operator $F^-_{1,i}$, and $\mathcal{F}_2$ is a forward convex coupling with Kantorovich operator $F^+_{2,j}$, then

$$\begin{aligned}
\mathcal{F}_1\star -\mathcal{F}_2\,(\mu,\nu) &= \inf\{\mathcal{F}_1(\mu,\sigma)-\mathcal{F}_2(\sigma,\nu);\,\sigma\in\mathcal{P}(X_2)\}\\
&= \inf_{\sigma\in\mathcal{P}(X_2)}\Big\{\mathcal{F}_1(\mu,\sigma)-\sup_{g\in C(X_2)}\Big\{\sup_{j\in J}(\int_{X_3}F^+_{2,j}g\,d\nu)-\int_{X_2}g\,d\sigma\Big\}\Big\}\\
&= \inf_{\sigma\in\mathcal{P}(X_2)}\inf_{g\in C(X_2)}\inf_{j\in J}\Big\{\mathcal{F}_1(\mu,\sigma)-\int_{X_3}F^+_{2,j}g\,d\nu+\int_{X_2}g\,d\sigma\Big\}\\
&= \inf_{g\in C(X_2)}\inf_{j\in J}\Big\{-\sup_{\sigma\in\mathcal{P}(X_2)}\{-\int_{X_2}g\,d\sigma-\mathcal{F}_1(\mu,\sigma)\}-\int_{X_3}F^+_{2,j}g\,d\nu)\Big\}\\
&= \inf_{g\in C(X_2)}\inf_{j\in J}\Big\{-(\mathcal{F}_1)^*_\mu(-g)-\int_{X_3}F^+_{2,j}(g)\,d\nu\Big\}\\
&= \inf_{g\in C(X_2)}\inf_{j\in J}\Big\{-(\inf_{i\in I}\int_{X_1}F^-_{1,i}(-g)\,d\nu-\int_{X_3}F^+_{2,j}(g)\,d\nu\Big\}\\
&= \inf_{g\in C(X_2)}\inf_{j\in J}\sup_{i\in I}\Big\{-\int_{X_1}F^-_{1,i}(-g)\,d\nu-\int_{X_3}F^+_{2,j}(g)\,d\nu\Big\}
\end{aligned}$$

2) If $\mathcal{F}_1$ is a backward β-entropic transfer with Kantorovich operator E^-, it suffices to note in the above proof that $(\mathcal{F}_1)^*_\mu(g)=\beta(\int_X E_1^-(-g)\,d\mu)$.

3) If now $\mathcal{F}_2$ is a forward α-entropic transfer with Kantorovich operator E_2^+, then it suffices to note in the above proof that $(\mathcal{F}_2)^*_\nu(g) = \alpha(\int_X E_2^+ g\, d\nu)$.
4) corresponds to when $\alpha(t) = t$.

Corollary 14.5. *Let $\mathcal{F}$ be a convex backward coupling on $X_2 \times Y_2$ with Kantorovich family $(F_i^-)_{i\in I}$ and let $\mathcal{E}$ be a backward β-entropic 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 $\mathcal{F}(\nu, \sigma) \leqslant \lambda\, \mathcal{E} \star \mathcal{T}\,(\mu, \sigma)$.*
2. *For all $g \in C(X_2)$, we have $\beta\big(\int_{X_1} E^- \circ T^- g)\, d\mu\big) \leqslant \inf\limits_{i\in I} \frac{1}{\lambda} \int_{X_2} F_i^-(\lambda g) d\nu$.*

In particular, if $\mathcal{E}_2$ is a backward β_2-entropic transfer on $X_2 \times Y_2$ with Kantorovich operator E_2^-, and $\mathcal{E}_1$ is a backward β_1-entropic 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 $\mathcal{E}_2(\nu, \sigma) \leqslant \lambda\, \mathcal{E}_1 \star \mathcal{T}\,(\mu, \sigma)$.*
2. *For all $g \in C(X_2)$ and $i \in I$, we have $\beta_1\big(\int_{X_1} E_1^- \circ T^- g)\, d\mu\big) \leqslant \frac{1}{\lambda}\beta_2(\int_{X_2} E_2^-(\lambda g) d\nu)$.*

Proof: Note that here, we need the formula for $(\mathcal{E} \star \mathcal{T}) \star (-\tilde{\mathcal{F}})(\mu, \nu)$. Since $\tilde{\mathcal{F}}$ is now a convex forward transfer with Kantorovich operators equal to $\tilde{F}_i^+(g) = -F_i^-(-g)$, we can apply Part 2) of Proposition 14.4 to $\mathcal{F}_2 = \frac{1}{\lambda}\tilde{\mathcal{F}}$ and $\mathcal{F}_1 = \mathcal{E} \star \mathcal{T}$, which is a backward β-entropic transfer with Kantorovich operator $E^- \circ T^-$, to obtain

$$(\mathcal{E} \star \mathcal{T}) \star (-\tilde{\mathcal{F}})(\mu, \nu) = \inf_{g\in C(X_2)} \inf_{j\in J} \left\{ -\beta(\int_{X_1} E^- \circ T^- g\, d\mu) + \frac{1}{\lambda} \int_{X_3} F_j^-(\lambda g)\, d\nu \right\}.$$

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 14.6. *Let $\mathcal{F}$ be a backward convex transfer on $X_2 \times Y_2$ with Kantorovich family $(F_i^-)_{i\in I}$, and 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 $\mathcal{F}(\nu, \sigma) \leqslant \lambda\, \mathcal{H} \star \mathcal{T}\,(\mu, \sigma)$*
2. *For all $g \in C(X_2)$, we have $\log\big(\int_{X_1} e^{T^- g}\, d\mu\big) \leqslant \inf\limits_{i\in I} \frac{1}{\lambda} \int_{X_2} F_i^-(\lambda g) d\nu$.*

Remark 14.7. An immediate application of (4) in Proposition 14.4 is the following result in [22]

$$\inf\{\overline{\mathcal{W}}_2(\mu, \sigma) + \mathcal{H}(dx, \sigma); \sigma \in \mathcal{P}(\mathbb{R}^d)\} = \inf\{-\log \int e^{-f^*}\, dx + \int f\, d\mu; f \in \mathcal{C}(\mathbb{R}^d)\}, \quad (309)$$

where $\mathcal{C}onv(\mathbb{R}^d)$ is the cone of convex functions on $\mathbb{R}^d$, and $\overline{\mathcal{W}}_2(\mu, \sigma) = -\mathcal{W}_2(\sigma, \bar{\mu})$, the latter being the Brenier transfer of Example 3.12 and $\bar{\mu}$ is defined as $\int f(x) d\bar{\mu}(x) = \int f(-x) d\mu(x)$.

Note that in this case, $T^+f(x) = -f^*(-x)$, $E^-f = e^f$ and $\beta(t) = \log t$, and since $g^{**} \leqslant g$,

$$
\begin{aligned}
\inf\{\overline{\mathcal{W}}_2(\mu,\sigma) + \mathcal{H}(dx,\sigma); \sigma \in \mathcal{P}(\mathbb{R}^d)\} &= \mathcal{H} \star (-\mathcal{W}_2)(dx, \bar{\mu}) \\
&= \inf\{-\log \int e^{-g}\, dx + \int g^*(x)\, d\mu; g \in C(\mathbb{R}^d)\} \\
&= \inf\{-\log \int e^{-f^*}\, dx + \int f\, d\mu; f \in \mathcal{C}onv(\mathbb{R}^d)\}.
\end{aligned}
$$

What is remarkable in the result of Cordero-Erausquin and Klartag [22] is the characterization of those measures μ (the moment measures) for which there is attainment in both minimization problems.

14.3 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

$$
\mathcal{F}(\sigma_1, \sigma_2) \leqslant \lambda_1 \mathcal{T}_1 \star \mathcal{H}_1(\sigma_1, \mu) + \lambda_2 \mathcal{T}_2 \star \mathcal{H}_2(\sigma_2, \nu). \tag{310}
$$

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

Theorem 14.8. *Assume $\mathcal{F}$ is a backward convex coupling on $Y_1 \times Y_2$ with Kantorovich family $(F_i^-)_{i\in I}$, $\mathcal{E}_1$ (resp., $\mathcal{E}_2$) is a forward α_1-entropic transfer on $Y_1 \times X_1$ (resp., a forward α_2-entropic transfer on $Y_2 \times X_2$) with Kantorovich operator E_1^+ (resp., E_2^+), then for any $(\mu, \nu) \in \mathcal{P}(X_1) \times \mathcal{P}(X_2)$, we have*

$$
\tilde{\mathcal{E}}_1 \star (-\mathcal{F}) \star \mathcal{E}_2\, (\mu, \nu) = \inf_{i\in I} \inf_{f\in C(X_3)} \left\{ \alpha_1\Big(\int_{X_1} E_1^+ \circ F_i^- f\, d\mu\Big) + \alpha_2\Big(\int_{X_2} E_2^+(f)\, d\nu\Big) \right\}. \tag{311}
$$

Proof: If $\mathcal{E}_1$ a forward α_1-entropic transfer on $Y_1 \times X_1$, then $\tilde{\mathcal{E}}_1$ is a backward $-(\alpha_1^{\oplus})^{\ominus}$-entropic transfer on $X_1 \times Y_1$ with Kantorovich operator $\tilde{E_1^-} g = -E_1^+(-g)$. Apply Proposition 14.1 with $\mathcal{F}_1 = \tilde{\mathcal{E}}_1$, and $\mathcal{F}_2 = \mathcal{F}$ to get

$$
\begin{aligned}
\tilde{\mathcal{E}}_1 \star (-\mathcal{F})\, (\mu, \nu) &= \inf_{f\in C(X_3)} \inf_{i\in I} \left\{ (\alpha_1^{\oplus})^{\ominus}\Big(\int_{X_1} -E_1^+ \circ F_i^- f\, d\mu\Big) - \int_{X_3} f\, d\nu \right\} \\
&= \inf_{f\in C(X_3)} \inf_{i\in I} \left\{ \alpha_1\Big(\int_{X_1} E_1^+ \circ F_i^- f\, d\mu\Big) - \int_{X_3} f\, d\nu \right\}.
\end{aligned}
$$

Write now,

$$
\begin{aligned}
\tilde{\mathcal{E}}_1 \star (-\mathcal{F}) \star \mathcal{E}_2\, (\mu, \nu) &= \inf \left\{ \tilde{\mathcal{E}}_1 \star (-\mathcal{F})(\mu, \sigma) + \mathcal{E}_2(\sigma, \nu);\ \sigma \in \mathcal{P}(Y_2) \right\} \\
&= \inf_{\sigma\in\mathcal{P}(Y_2)} \inf_{f\in C(X_3)} \inf_{i\in I} \left\{ \alpha_1\Big(\int_{X_1} E_1^+ \circ F_i^- f\, d\mu\Big) - \int_{X_3} f\, d\sigma + \mathcal{E}_2(\sigma, \nu) \right\} \\
&= \inf_{i\in I} \inf_{f\in C(X_3)} \left\{ \alpha_1\Big(\int_{X_1} E_1^+ \circ F_i^- f\, d\mu\Big) - \sup_{\sigma\in\mathcal{P}(Y_2)} \{ \int_{Y_2} f\, d\sigma - \mathcal{E}_2(\sigma, \nu) \} \right\} \\
&= \inf_{i\in I} \inf_{f\in C(X_3)} \left\{ \alpha_1\Big(\int_{X_1} E_1^+ \circ F_i^- f\, d\mu\Big) + \alpha_2\Big(\int_{X_2} E_2^+(-f)\, d\nu\Big) \right\}.
\end{aligned}
$$

Corollary 14.9. *Assume $\mathcal{E}_1$ (resp., $\mathcal{E}_2$) is a forward α_1-entropic transfer on $Z_1 \times X_1$ (resp., α_2-entropic transfer on $Z_2 \times X_2$) with Kantorovich operator E_1^+ (resp., E_2^+). 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_1^+ (resp., T_2^+), and let $\mathcal{F}$ be a backward convex coupling on $Y_1 \times Y_2$ with Kantorovich family $(F_i^-)_i$. 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*

$$\mathcal{F}(\sigma_1, \sigma_2) \leqslant \lambda_1 \mathcal{T}_1 \star \mathcal{E}_1(\sigma_1, \mu) + \lambda_2 \mathcal{T}_2 \star \mathcal{E}_2(\sigma_2, \nu). \tag{312}$$

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

$$\lambda_1 \alpha_1 \Big(\int_{X_1} E_1^+ \circ T_1^+ \circ \big(\frac{1}{\lambda_1} F_i^- g\big)\, d\mu \Big) + \lambda_2 \alpha_2 \Big(\int_{X_2} E_2^+ \circ T_2^+ \big(\frac{-1}{\lambda_2} g\big)\, d\nu \Big) \geqslant 0. \tag{313}$$

Proof: It suffices to apply the above with the forward $\lambda_i \alpha_i$-transfers $\mathcal{F}_i := \lambda_i \mathcal{T}_i \star \mathcal{E}_i$, whose Kantorovich operators are $F_i(g) = E_i^+ \circ T_i^+ (\frac{g}{\lambda_i})$ 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 [50].

Corollary 14.10. *Assume $\mathcal{F}$ is a convex backward coupling on $Y_1 \times Y_2$ with Kantorovich family $(F_i^-)_{i \in I}$, and let $\mathcal{T}_1$ (resp., $\mathcal{T}_2$) 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*

$$\mathcal{F}(\sigma_1, \sigma_2) \leqslant \lambda_1 \mathcal{T}_1 \star \mathcal{H}(\sigma_1, \mu) + \lambda_2 \mathcal{T}_2 \star \mathcal{H}(\sigma_2, \nu). \tag{314}$$

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

$$\Big(\int_{X_1} e^{-T_1^+ \circ \frac{1}{\lambda_1} F_i^- g}\, d\mu \Big)^{\lambda_1} \Big(\int_{X_2} e^{-T_2^+ (\frac{1}{-\lambda_2} g)}\, d\nu \Big)^{\lambda_2} \leqslant 1. \tag{315}$$

If $\mathcal{T}_1 = \mathcal{T}_2$ 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

$$\Big(\int_{X_1} e^{\frac{-1}{\lambda_1} F_i^- g}\, d\mu \Big)^{\lambda_1} \Big(\int_{X_2} e^{\frac{1}{\lambda_2} g}\, d\nu \Big)^{\lambda_2} \leqslant 1. \tag{316}$$

References

[1] Aubry, S.: *The twist map, the extended Frenkel-Kontorova model and the devil?s staircase,* Phys. D 7, 240?258 (1983)

[2] Alibert, J., Bouchité, G., Champion, T., *A new class of costs for optimal transport planning*, European Journal of Applied Mathematics, 1-35. doi:10.1017/S0956792518000669

[3] Alvarez, L., Guichard, F., Lions, P.L. et al. , *Axioms and fundamental equations of image processing*, Arch. Rational Mech. Anal., Vol. 123, Issue 3, pp 199?257 (1993).

[4] L. Ambrosio, N. Gigli & G. Savaré, *Gradient flows in metric spaces and in the Wasserstein space of probability measures.* Lectures in Mathematics, Birkhäuser, ETH Zurich (2005).

[5] J. Backhoff Veraguas, M. Beiglböck and G. Pammer, *Existence, Duality, and Cyclic Monotonicity for Weak Transport Costs,* https://arxiv.org/abs/1809.05893 (Nov. 7, 2018)

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

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

[8] A. Barton, N. Ghoussoub, *Dynamic and Stochastic Propagation of Brenier's Optimal Mass Transport*, European Journal of Applied Mathematics (Published online: 20 March 2019) https://doi.org/10.1017/S0956792519000032

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

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

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

[12] R. Bowen. *Equilibrium States and Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Mathematics 470, Springer-Verlag (1975)

[13] M. Bowles, N. Ghoussoub, *A Theory of Transfers: Duality and convolution*, https://arxiv.org/abs/1804.08563 (April 16, 2018, Revised October 24, 2018) 41 pp.

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

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

[16] Y. Brenier, *The least action principle and the related concept of generalized flows for incompressible perfect fluids*, J. Amer. Math. Soc., 2 (1989), 225-255.

[17] Y. Brenier, *The dual least action problem for an ideal, incompressible fluid,* Arch. Rational Mech. Anal., 122 (1993), 323-351.

[18] Y. Brenier, *Minimal geodesics on groups of volume-preserving maps and general- ized solutions of the Euler equations,* Comm. Pure Appl. Math., 52 (1999), 411-452.

[19] Gustave Choquet *Lectures on Anaysis* W. A. Benjamin, Inc. 1968

[20] L. C. Evans, *Three singular variational problems*, Viscosity Solutions of Differential Equations and Related Topics, Research Institute for the Math. Sciences, Kokyuroku (2003), 13-23.

[21] L. C. Evans, *A survey of partial differential equations methods in weak KAM theory*, Comm. in Pure and Applied Mathematics 57 (2004), 445-480.

[22] D. Cordero-Erausquin, B. Klartag, *Moment measures*, J. Funct. Anal. 268, no. 12, (2015), 3834-3866.

[23] 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.

[24] S. Dweik, N. Ghoussoub, Y-H. Kim and A. Z. Palmer, *Optimal stochastic transport with free end time,* In preparation (2019).

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

[26] W. H. Fleming and H.M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag, vol. 25, New York, 1993.

[27] J. Fontbona, N. Gozlan, and J-F. Jabir, *A variational approach to some transport inequalities,* Ann. Inst. H. Poincaré Prob. Stat. Vol. 53, No. 4 (2017), 1719-1746.

[28] E. Garibaldi, A. O. Lopes, *On the Aubry-Mather theory for symbolic dynamics*, Ergodic Theory and Dynamical Systems 28(3), 791?815 (2008)

[29] I. Gentil, C. Léonard, L. Ripani, *About the analogy between optimal transport and minimal entropy,* Ann. de la Faculté des Sciences de Toulouse, Mathématiques. Série 6, Université Paul Sabatier 26, 3 (2017) pp. 569-600.

[30] N. Ghoussoub, B. Maurey, *Plurisubharmonic martingales and barriers in complex quasi- Banach spaces. Annales Inst. Fourier,* Tome 39-Fasc 4 (1989) p. 1007-1060

[31] N. Ghoussoub, Y-H. Kim, and T. Lim, *Structure of optimal martingale transport in general dimensions,* The Annals of Probability, Vol. 47, No. 1, (2019) p. 109-164.

[32] N. Ghoussoub, Y. Kim, and T. Lim, *Optimal Brownian Stopping When the Source and Target are Radially Symmetric Distributions*, (June 25, 2019) 29 pp.

[33] N. Ghoussoub, Y-H. Kim and A. Z. Palmer, *Optimal Transport With Controlled Dynamics and Free End Times*, SIAM Journal on Control and Optimization, Vol. 56, No. 5 (2018) p. 3239-3259.

[34] N. Ghoussoub, Y-H. Kim and A. Z. Palmer, *PDE Methods for Optimal Skorokhod Embeddings*, Calculus of Variations and PDE, in press (Submitted August 14, 2018) 31 pp.

[35] N. Ghoussoub, Y-H. Kim and A. Z. Palmer, *A solution to the Monge transport problem for Brownian martingales*, submitted (March 2019) 25 pp.

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

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

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

[39] 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.

[40] N. Gozlan, C. Roberto, P.M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal., preprint (2014), `arXiv:1412.7480v4`.

[41] N. Gozlan, N. Juillet, *On a mixture of Brenier and Strassen theorems,* arXiv:1808.02681, (August 2018) 23 pp.

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

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

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

[45] R. Mané, *Lagrangian flows: The dynamics of globally minimizing orbits,* Bol. Soc. Bras. Mat, 28 (1997) 141-153.

[46] D. Martino, Personal communication, August 2019.

[47] K. Marton, *Bounding $\bar{d}$-distance by informational divergence: a way to prove measure concentration,* Annals of Probability, 24:857–866 (1996).

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

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

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

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

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

[53] G. Mokobodzki, *Cônes de potentiels et noyaux subordonnés,* C.I.M.E., 1. Ciclo Stresa 1969, Potential Theory, 207-248 (1970).

[54] R. R. Phelps, *Convex functions, monotone operators and differentiability.* Lecture notes in mathematics, **1364** (1989).

[55] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Springer International Publishing, 87 (2015).

[56] 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.

[57] 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.

[58] C. Orrieri, A. Porretta, G. Savaré, *A variational approach to the mean field planning problem,* https://arxiv.org/abs/1807.09874 (July 2018) 52 pp.

[59] Skorokhod, A. V., *Studies in the theory of random processes*, Translated from the Russian by Scripta Technica, Inc, Addison-Wesley Publishing Co., Inc., Reading, Mass., viii+199 (1965).

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

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

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

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

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

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