Linear transfers, Kantorovich operators, and their ergodic properties

by

Malcolm Bowles

B. Sc. University of Victoria, 2012

M. Sc, University of Victoria, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Mathematics)

The University of British Columbia
(Vancouver)

August 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Linear transfers, Kantorovich operators, and their ergodic properties

submitted by **Malcolm Bowles** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Mathematics**.

Examining Committee:

Nassif Ghoussoub, Mathematics
Supervisor

Young-Heon Kim, Mathematics
Supervisor

Philip Loewen, Mathematics
Supervisory Committee Member

Stephen Gustafson, Mathematics
University Examiner

Gordon Semenoff, Physics
University Examiner

Abstract

In this work, we introduce and study a class of convex functionals on pairs of probability measures, the *linear transfers*, which have a structure that commonly arises in the dual formulations of many well-studied variational problems. We show that examples of linear transfers include a large number of well-known transport problems, including the *weak*, *stochastic*, *martingale*, and *cost-minimising* transports. Further examples include the *balayage of measures*, and *ergodic optimisation of expanding dynamical systems*, among others. We also introduce an extension of the linear transfers, the *convex transfers*, and show that they include the *relative entropy* functional and p-powers ($p \geq 1$) of linear transfers.

We study the properties of linear and convex transfers and show that the *inf-convolution* operation preserves their structure. This allows dual formulations of transport-entropy and other related inequalities, to be computed in a systematic fashion.

Motivated by connections of optimal transport to the theory of Aubry-Mather and weak KAM for Hamiltonian systems, we develop an analog in the setting of linear transfers. We prove the existence of an idempotent operator which maps into the set of weak KAM solutions, an idempotent linear transfer that plays the role of the Peierls barrier, and we identify analogous objects in this setting such as the Mather measures and the Aubry set. We apply this to the framework of ergodic optimisation in the holonomic case.

Lay Summary

A probability measure specifies the chance of a particular event occurring. Many mathematical problems are concerned with computing similarities/differences between two probability measures. Among such problems, a large number of them have a common structure; we isolate and define this common formalism as a "linear transfer", and more generally, a "convex transfer". We demonstrate linear and convex transfers encompass a wide range of well-known mathematical problems and study their properties.

A Hamiltonian system is a system whose state over time is determined by a function describing its total energy. In trying to investigate the long-time dynamics of this system, one is led to the search for functions which solve certain equations involving certain constants; this is collectively known as Aubry-Mather/weak KAM theory. We generalise aspects of this theory to the setting of linear transfers.

Preface

This thesis is based on [10], which is currently in preparation to be submitted for publication.

The material presented in Chapters 2 and 3 is based on initial ideas of Nassif Ghoussoub, which we subsequently developed and expanded jointly. The material in Chapter 4 is an overview of classical Aubry-Mather theory, which motivated the present work, and was written by myself. The material presented in Chapter 5 is work developed in collaboration with Nassif Ghoussoub. Section 5.6.1 is based on a key contribution by Dorian Martino, which I modified and extended.

Table of Contents

Acknowledgments

First, I would like to thank Nassif Ghoussoub for this work would not have been possible without his support and collaboration. I also want to express my gratitude to Young-Heon Kim for his valuable advice, guidance, and generosity, throughout my time at UBC.

I would also like to thank Philip Loewen for his careful reading and valuable suggestions to improve the thesis, along with Stephen Gustafson and Gordon Semenoff for their time as the university examiners.

A special thanks to all the people I met at UBC, especially my friends and fellow graduate students, Sarafa Iyaniwura, Kelsey Everard, and Virginie Sonnet, who have substantially enriched my time here and am so happy to have met.

I especially appreciate all the discussions, both mathematical, and non-mathematical, with my fellow graduate students Guillermo Martinez-Dibene, Sophie Macdonald, Alperen Bulut, Igor Pinheiro, Li Wang, and Reza Sadoughian. I wish them success in all their endeavours. I also would like to thank Aaron Palmer, Samer Dweik, and Hugo Lavenant, for their helpful discussions.

Finally, I save until the end, my thanks to my Mum, who has always been a constant source of encouragement, and my sisters Alison and Rachel. I would not be writing this today without their support.

Chapter 1

Introduction

This work is devoted to an axiomatic study of certain convex functionals on pairs of probability measures, whose structure commonly arises in the dual formulations of many well-studied variational problems. Typical examples include, the *weak*, *stochastic*, *martingale*, and *cost-minimising transport* problems, the *relative entropy*, the *balayage of measures*, and *ergodic optimisation of expanding dynamical systems*, among others.

The overall broad aim is to unify the settings of these various problems together, with the goal to add clarity and understanding, and to derive non-trivial extensions, including in this work, an analog of the Aubry-Mather and weak KAM theory for Lagrangian systems. We begin by introducing in Chapter 2 the notions of *backward/forward linear couplings*, as those convex functionals on pairs of probability measures (μ, ν) which arise as the supremum of linear functionals $(\mu, \nu) \mapsto \int_Y g d\nu - \int_X f d\mu$ where the supremum is taken over pairs of continuous functions (g, f) which *lie on the graph of an operator*; either $(g, T^- g)$, or $(T^+ f, f)$ for operators T^-, T^+. Among these backward/forward linear couplings, there exists a distinguished subset that we call *linear transfers*. These possess an additional structure that arises as a consequence of being defined on pairs of measures (μ, ν) and therefore allow us to consider one-variable convex maps by fixing μ, or fixing ν. A *backward/forward linear transfer* is a backward/forward linear coupling for which, when fixing μ or ν, and considering the resulting convex function in ν or μ, its *Legendre transform* precisely coincides with the function $g \mapsto \int_X T^- g d\mu$ or $f \mapsto \int_Y -T^+(-f) d\nu$. In other words, the operator T^- or T^+ completely characterises the Legendre duality. This leads to the study of such operators T^-, T^+, which we call *backward/forward Kantorovich operators* who, conversely, define a linear transfer. The Kantorovich operators are the main focus for an "Aubry-Mather and weak KAM theory for linear transfers" later in Chapter 5.

We proceed by exhibiting a number of examples in Section 2.5 which can be realised as backward/forward linear transfers. These include all convex energies $\nu \mapsto I(\nu)$ of one variable, and any transfer whose Kantorovich operator is given by a point transformation $\sigma : X \to X$ on the underlying space X, or more generally, by a positive bounded linear operator T on continuous functions; this is connected to ergodic theory for expanding dynamical systems which we discuss later in Chapter 5. Interesting examples also include the balayage of measures, which is concerned with pairs of measures (μ, ν) that are in *partial order* with respect to convex cones $\mathcal{A}$ of continuous functions, of which lies the work of Strassen, the important theory developed by Choquet for convex functions when $\mathcal{A}$ is the cone of convex functions, and the work of Skorokhod for Brownian motion when the cone is subharmonic functions. In Section 2.6, we show that cost-minimising transport is a linear transfer. This is especially important for us as a connection made by Bernard-Buffoni between cost-minimising transport, and Aubry-Mather and weak KAM theory, provides an inspiration for an analog in the setting of linear transfers in Chapter 5. In Section 2.7, we show that, under certain assumptions, linear transfers actually coincide with weak transports studied by Gozlan et. al. and permits us to introduce the notion of a "linear transfer envelope" in an analogy to classical convex envelopes. In Section 2.8, we introduce further examples of linear transfers, including martingale transport, the Schrödinger bridge, stochastic transport, and related problems.

We next introduce a natural extension of linear transfers in Section 2.9, the *convex transfers*. As with linear couplings and linear transfers, we define *convex couplings* as those functionals $\mathcal{T}(\mu, \nu)$ which are *supremums* of linear couplings, and convex transfers as the distinguished subset of convex couplings for which, when restricting to the one variable convex function $\nu \mapsto \mathcal{T}(\mu, \nu)$ (similarly for $\mu \mapsto \mathcal{T}(\mu, \nu)$), their Legendre transform is given by $g \mapsto \inf_{i \in I} \int_X T_i^- g d\mu$, an infimum over a family of operators T_i^-. An important example of a convex, but not linear, transfer is the relative entropy. We then introduce operations that preserve linear and convex transfers in Section 2.10, the most important of which is the notion of *inf-convolution* which we shall extensively use in Chapter 5. Finally, considered as one variable convex functions $\nu \mapsto \mathcal{T}(\mu, \nu)$ or $\mu \mapsto \mathcal{T}(\mu, \nu)$, we discuss their subdifferentials in Section 2.11.

In Chapter 3, we use the structure of linear and convex transfers to write equivalent statements for "transport-entropy" type inequalities and various generalisations. A typical transport-entropy inequality is of the form $\mathcal{T}_c(\nu, \mu) \leq \mathcal{H}(\mu, \nu)$ where $\mathcal{T}_c(\mu, \nu)$ is an optimal transport for some chosen cost function c between ν and μ, and $\mathcal{H}(\mu, \nu)$ is the relative entropy of ν with respect to μ. Here μ is a fixed reference measure, and the inequality should hold for all ν. The inequality is equivalent to the positivity of $\inf_{\nu \in \mathcal{P}(Y)} \{\mathcal{H}(\mu, \nu) - \mathcal{T}_c(\nu, \mu)\}$. By using the structure of linear and convex transfers, we can write in a systematic way a dual formula for $\inf_{\nu \in \mathcal{P}(Y)} \{\mathcal{H}(\mu, \nu) - \mathcal{T}_c(\nu, \mu)\}$, as well as various other types of inequalities, for which the positivity could be obtained.

Chapter 4 is a preliminary to Chapter 5, which consists of an overview of the theory of weak KAM developed by Fathi. In Section 4.2, we begin by introducing integrable Hamiltonian systems and the role the Hamilton-Jacobi equation plays in the search for a change of coordinates that transform a non-integrable Hamiltonian to an integrable one. We then proceed in Section 4.3 to introduce the *Peierls barrier* from which we may construct the important *Aubry set* where viscosity subsolutions of Hamilton-Jacobi are differentiable, and on which the *Mather measures* are supported. The Lax-Oleinik semi-group of operators and weak KAM solutions of Fathi are discussed in Section 4.4 which connects solutions of Hamilton-Jacobi with the Aubry set and Mather measures. An important characterisation established by Bernard-Buffoni is then presented in Section 4.5, where they characterise the Peierls barrier, Mather measures, Aubry set, and weak KAM solutions, via optimal transport.

In Chapter 5, which consists of the main application in this work, we are interested in developing an analog of the weak KAM and Aubry-Mather theory of Chapter 4, for linear transfers. Here the Kantorovich operators T^-, T^+, introduced in Chapter 2, play the role of the Lax-Oleinik semi-group of the Chapter 4 in defining *weak KAM solutions for linear transfers.* These are the functions f that satisfy $T^+f - c = f$, or those g satisfying $T^-g + c = g$ for a specific (critical) constant c. Through Theorems 5.4.3, 5.5.3, and 5.6.3, in Sections 5.4, 5.5, and 5.6, we show that for most linear transfers, there exists an *idempotent* Kantorovich operator T_∞ which maps into the set of weak KAM solutions.

Theorem 1.0.1 (Theorem 5.4.3)**.** *Let $\mathcal{T}$ be a weak* continuous backward linear transfer on $\mathcal{M}(X) \times \mathcal{M}(X)$ with modulus of continuity δ, and with backward Kantorovich operator $T^- : C(X) \to C(X)$. Then $c(\mathcal{T}) := \inf_\mu \mathcal{T}(\mu, \mu)$ is the unique constant for which there exists a backward Kantorovich operator $T^-_\infty : C(X) \to C(X)$, together with its induced backward linear transfer $\mathcal{T}_\infty$, satisfying:*

1. *T^-_∞ maps every $g \in C(X)$ to a backward weak KAM solution for T^-, i.e.,*

$$T^-_n \circ T^-_\infty g + nc(\mathcal{T}) = T^-_\infty g \quad \text{for all } g \in C(X) \text{ and all } n \in \mathbb{N}.$$

2. *T^-_∞ is idempotent, i.e. $T^-_\infty \circ T^-_\infty g = T^-_\infty g$ for all $g \in C(X)$.*

3. *$\mathcal{T}_\infty$ satisfies,*

$$(\mathcal{T}_n - nc(\mathcal{T})) \star \mathcal{T}_\infty(\mu, \nu) = \mathcal{T}_\infty(\mu, \nu) = \mathcal{T}_\infty \star (\mathcal{T}_n - nc(\mathcal{T}))(\mu, \nu) \quad \text{for every } n \in \mathbb{N}.$$

4. *$\mathcal{T}_\infty$ is idempotent and therefore $\mathcal{A}$-factorisable, i.e. 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*

$$\mathcal{T}_\infty(\mu, \nu) = \inf\{\mathcal{T}_\infty(\mu, \sigma) + \mathcal{T}_\infty(\sigma, \nu), \sigma \in \mathcal{A}\},$$

and the infimum on $\mathcal{A}$ is attained.

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

$$\sup\left\{\int T_\infty^- g d(\nu-\mu)\,;\, g \in C(X)\right\} \leq \mathcal{T}_\infty(\mu,\nu) \leq \liminf_{n\to\infty}(\mathcal{T}_n(\mu,\nu) - nc(\mathcal{T})).$$

6. *If $\mathcal{T}(\mu,\mu) = c(\mathcal{T})$, then $\mu \in \mathcal{A}$. Additionally, $(\mu,\mu) \in \mathcal{D}$ if and only if $\mu \in \mathcal{A}$ and $\mathcal{T}(\mu,\mu) = c(\mathcal{T})$, where*

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

Inspired by the optimal transport analysis provided by Bernard-Buffoni in the previous chapter, we can identify analogs of the objects from Aubry-Mather and weak KAM. The Peierls barrier in this setting is the *idempotent linear transfer* associated to the Kantorovich operator T_∞^-. The Mather measures consist of those minimising transport plans for the *weak transport* associated to the linear transfer $\mathcal{T}$, and the Aubry set is the set $\mathcal{D}$.

Moving from weak* continuity to weak* lower semi-continuity, which is the natural setting for linear transfers, yields the following.

Theorem 1.0.2 (Theorem 5.5.3)**.** *Let $\mathcal{T}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains the Dirac measures. Assume $c(\mathcal{T}) := \inf_\mu \mathcal{T}(\mu,\mu) < +\infty$, and that $\inf_{(\mu,\nu)} \mathcal{T}(\mu,\nu) = \inf_\mu \mathcal{T}(\mu,\mu)$.*

Then $c(\mathcal{T})$ is the unique constant for which there exists an idempotent Kantorovich operator $T_\infty^- : C(X) \to USC(X)$ mapping into the set of backward weak KAM solutions, i.e.

$$T_n^- \circ T_\infty^- g + nc(\mathcal{T}) = T_\infty^- g \quad \text{for all } g \in C(X) \text{ and all } n \in \mathbb{N}.$$

The corresponding backward linear transfer $\mathcal{T}_\infty$ is idempotent, the set $\mathcal{A} := \{\sigma \in \mathcal{P}(X); \mathcal{T}_\infty(\sigma,\sigma) = 0\}$ is non-empty, and for every (μ,ν), satisfies

$$\mathcal{T}_\infty(\mu,\nu) = \inf\{\mathcal{T}_\infty(\mu,\sigma) + \mathcal{T}_\infty(\sigma,\nu), \sigma \in \mathcal{A}\},$$

and the infimum on $\mathcal{A}$ is attained.

Finally, the existence of an idempotent Kantorovich operator mapping into the set of backward weak KAM solutions breaks down when we weaken the hypotheses on the linear transfer.

Theorem 1.0.3 (Theorem 5.6.3)**.** *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. Assume:*

1. $c(\mathcal{T}) := \inf_\mu \mathcal{T}(\mu, \mu) < +\infty$,
2. $\sup_{x \in X} \inf_{\nu \in \mathcal{P}(X)} \mathcal{T}(\delta_x, \nu) < +\infty$,
3. *there exists $K > 0$ such that*

$$\limsup_{n \to \infty} \{nc(\mathcal{T}) - \inf_{\mu, \nu \in \mathcal{P}(X)} \mathcal{T}_n(\mu, \nu)\} \leq K,$$

4. *there exists $g \in C(X)$, such that the function*

$$x \mapsto \liminf_{n \to \infty} (T_n^- g(x) + nc(\mathcal{T}))$$

   belongs to $USC_b(X)$.

Then there exists $h \in USC(X)$ such that $T^- h + c(\mathcal{T}) = h$ on X.

In the last Section 5.6.1, we apply our results to ergodic optimization, in particular, to the setting of symbolic dynamics.

Chapter 2

Linear and convex transfers

2.1 Introduction

Throughout this chapter and entire thesis, X shall denote a compact metric space, equipped with its Borel σ-algebra. Where necessary, we will write the metric on X by d_X. The collection of Borel signed measures will be denoted by $\mathcal{M}(X)$, which is equipped with the weak* topology in duality with the real-valued continuous functions $C(X)$; the subset of Borel probability measures on X will be written as $\mathcal{P}(X)$. The set of functions $f : X \to \mathbb{R} \cup \{-\infty\}$ which are upper semi-continuous shall be denoted $USC(X)$; the subset of *bounded* functions in $USC(X)$ shall be denoted by $USC_b(X)$. The notation $LSC(X)$ is the set of lower semi-continuous functions, i.e. those functions f for which $-f \in USC(X)$; similarly for $LSC_b(X)$. The set of functions $f : X \to \mathbb{R} \cup \{\pm\infty\}$ which are Borel-measurable will be denoted by $\mathcal{B}(X)$; the subset of $\mathcal{B}(X)$ which are bounded above is $\mathcal{B}^b(X)$; those which are bounded below is $\mathcal{B}_b(X)$. The same conventions above hold for a second space Y in place of X.

Throughout, $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ will denote a proper (i.e. not identically $+\infty$), bounded below, convex, and weak* lower semi-continuous functional. By convex, we will always intend to mean *joint convexity*:

If $\mu_0, \mu_1 \in \mathcal{P}(X)$, and $\nu_0, \nu_1 \in \mathcal{P}(Y)$, then

$$\mathcal{T}(\mu_\lambda, \nu_\lambda) \leq (1-\lambda)\mathcal{T}(\mu_0, \nu_0) + \lambda \mathcal{T}(\mu_1, \nu_1), \quad \text{for all } \lambda \in [0,1],$$

where $\mu_\lambda := (1-\lambda)\mu_0 + \lambda\mu_1$ and $\nu_\lambda := (1-\lambda)\nu_0 + \lambda\nu_1$.

To make precise the weak* lower semi-continuity of $\mathcal{T}$, we note that the weak* topology on $\mathcal{P}(X)$ is metrizable when X is compact, and the metric can be taken to be the *Wasserstein distance* of optimal transport with distance cost given by the metric d_X on X (see e.g. [60], Chapter 7, and also Section 2.6 of this thesis for more on optimal transport). Therefore, weak* lower semi-continuity is equivalent in this case to *sequential* weak* lower semi-continuity, which we recall is the following property:

If $(\mu_n) \subset \mathcal{P}(X)$, $\mu_\infty \in \mathcal{P}(X)$, such that $\mu_n \to \mu_\infty$ as $n \to \infty$ in the weak* sense (and similarly for a sequence $(\nu_n) \subset \mathcal{P}(Y)$, and $\nu_\infty \in \mathcal{P}(Y)$), then

$$\mathcal{T}(\mu_\infty, \nu_\infty) \leq \liminf_{n\to\infty} \mathcal{T}(\mu_n, \nu_n).$$

Throughout, $\mathcal{T}$ shall also be viewed implicitly as a function on $\mathcal{M}(X) \times \mathcal{M}(Y)$ by defining $\mathcal{T} \equiv +\infty$ on $\mathcal{M}(X) \times \mathcal{M}(Y) \backslash \mathcal{P}(X) \times \mathcal{P}(Y)$. Its first partial domain $D_1(\mathcal{T})$ is the set of all $\mu \in \mathcal{P}(X)$ such that $\mathcal{T}(\mu, \nu) < +\infty$ for some $\nu \in \mathcal{P}(Y)$; similarly its second partial domain is $D_2(\mathcal{T})$ the set of all $\nu \in \mathcal{P}(Y)$ such that $\mathcal{T}(\mu, \nu) < +\infty$ for some $\mu \in \mathcal{P}(X)$.

2.2 The Legendre-Fenchel transform and duality for convex functions

We briefly recall the Legendre-Fenchel transform and the Fenchel-Moreau-Rockafellar theorem for convex functions (see e.g. [2], Section 9.3).

Definition 2.2.1. Let V be a normed linear space and $f : V \to \mathbb{R} \cup \{+\infty\}$ a proper function. The **Legendre-Fenchel transform (or conjugate)** of f is the function $f^* : V^* \to \mathbb{R} \cup \{+\infty\}$ defined on the topological dual space V^* via

$$f^*(v^*) := \sup\{\langle v^*, v\rangle - f(v)\,;\, v \in V\}$$

The collection of convex functions f on V which are proper and lower semi-continuous are distinguished by the fact that f can be recovered from f^*; this is the following Fenchel-Moreau theorem.

Theorem 2.2.2. *Let V be a normed linear space and V^* its topological dual space, equipped with the weak* topology. Then the following hold:*

1. *If $f : V \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex, and lower semi-continuous function, then*

$$f(v) = \sup\{\langle v^*, v\rangle - f^*(v^*)\,;\, v^* \in V^*\}$$

where $f^*(v^*) := \sup\{\langle v^*, v\rangle - f(v)\,;\, v \in V\}$.

2. *If $g : V^* \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex, and weak* lower semi-continuous function, then*

$$g(v^*) = \sup\{\langle v^*, v\rangle - g^*(v)\,;\, v \in V\}.$$

where $g^*(v) := \sup\{\langle v^*, v\rangle - g(v^*)\,;\, v^* \in V^*\}$ for $v \in V$.

2.3 Linear transfers

When the normed linear space V of the last section is the space of continuous functions $C(Y)$ equipped with the supremum norm, then Riesz' theorem says that V^* can be identified with $\mathcal{M}(Y)$. Therefore according to Fenchel-Moreau, any functional $F : \mathcal{M}(Y) \to \mathbb{R} \cup \{+\infty\}$ which is proper, convex, and weak* lower semi-continuous, satisfies

$$F(\nu) = \sup_{g \in C(Y)} \{\int_Y g d\nu - F^*(g)\} \tag{2.1}$$

where $F^*(g) = \sup_{\nu \in \mathcal{M}(Y)}\{\int_Y g d\nu - F(\nu)\}$.

For a proper, convex, bounded below, and weak* lower semi-continuous functional $\mathcal{T}$ on $\mathcal{P}(X) \times \mathcal{P}(Y)$, viewed as a function on $\mathcal{M}(X) \times \mathcal{M}(Y)$ (recall the notation in Section 2.1), we may then consider the partial map,

$$\mathcal{T}_\mu : \nu \mapsto \mathcal{T}(\mu, \nu)$$

for $\mu \in D_1(\mathcal{T})$. This is a proper, convex, bounded below, and weak* lower semi-continuous function on $\mathcal{M}(Y)$, hence identifying $F = \mathcal{T}_\mu$ in (2.1)

$$\mathcal{T}(\mu,\nu) = \mathcal{T}_\mu(\nu) = \sup_{g \in C(Y)} \{\int_Y g d\nu - \mathcal{T}_\mu^*(g)\}, \quad \mu \in D_1(\mathcal{T}),\, \nu \in \mathcal{P}(Y). \tag{2.2}$$

In exactly the same way, the other partial map $\mathcal{T}_\nu : \mu \mapsto \mathcal{T}(\mu,\nu)$ on $\mathcal{M}(X)$ satisfies

$$\mathcal{T}(\mu,\nu) = \mathcal{T}_\nu(\mu) = \sup_{f \in C(X)} \{\int_X f d\mu - \mathcal{T}_\nu^*(f)\}, \quad \mu \in \mathcal{P}(X),\, \nu \in D_2(\mathcal{T}). \tag{2.3}$$

The **linear transfers** are then those $\mathcal{T}$ whose partial maps $\mathcal{T}_\mu$, $\mathcal{T}_\nu$ have Legendre-Fenchel transforms which are *linear* with respect to μ and ν, respectively.

Definition 2.3.1 (Linear transfers)**.** Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex, bounded below, and weak* lower semi-continuous function.

1. We say that $\mathcal{T}$ is a **backward linear transfer**, if there exists an operator $T^- : C(Y) \to F^b(X)$ such that for all $\mu \in D_1(\mathcal{T})$ and all $g \in C(Y)$,
$$\mathcal{T}_\mu^*(g) = \int_X T^- g d\mu. \tag{2.4}$$

2. We say $\mathcal{T}$ is a **forward linear transfer**, if there exists an operator $T^+ : C(X) \to F_b(Y)$, such that for all $\nu \in D_2(\mathcal{T})$ and all $f \in C(X)$,
$$\mathcal{T}_\nu^*(f) = \int_Y -T^+(-f) d\nu.$$

Remark 2.3.2. It is important to stress that the operator T^- (resp., T^+) is *non-linear* in general, and not the usual operators from linear functional analysis; the term 'linear' refers to the linearity of the Legendre transform $\mathcal{T}_\mu^*$ with respect to μ.

A consequence of the above definition, is that the expression (2.2) for a backward linear transfer then becomes

$$\mathcal{T}(\mu,\nu) = \sup_{g\in C(Y)} \{\int_Y g d\nu - \int_X T^- g d\mu\}, \quad \mu \in \mathcal{D}_1(\mathcal{T}),\, \nu \in \mathcal{P}(Y),$$

and (2.3) for a forward linear transfer,

$$\mathcal{T}(\mu,\nu) = \sup_{f\in C(X)} \{\int_Y T^+ f d\nu - \int_X f d\mu\}, \quad \mu \in \mathcal{P}(X),\, \nu \in D_2(\mathcal{T}).$$

An important observation is that if $\mathcal{T}$ is a forward linear transfer, then $\tilde{\mathcal{T}}(\nu,\mu) := \mathcal{T}(\mu,\nu)$ is a backward linear transfer on $\mathcal{P}(Y)\times\mathcal{P}(X)$, and vice versa. Indeed, if $\mathcal{T}$ is a forward linear transfer, then

$$\tilde{\mathcal{T}}_\nu^*(f) = \int_X -T^+(-f) d\nu$$

so if we define $\tilde{T}^- f := -T^+(-f)$, then we conclude $\tilde{\mathcal{T}}$ is a backward linear transfer on $\mathcal{P}(Y)\times\mathcal{P}(X)$.

It therefore suffices to focus either on those $\mathcal{T}$ which are backward linear transfers, or on those $\mathcal{T}$ which are forward linear transfers; it will be convenient to focus and state results for the backward linear transfers. However, occasionally, we shall see that it is sometimes useful to consider $\mathcal{T}$ which is *both* a forward, and a backward, linear transfer, hence why we introduce both notions.

We shall often say "$\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(Y)$ (or $\mathcal{P}(X)\times\mathcal{P}(X)$, $\mathcal{P}(Y)\times\mathcal{P}(X)$, etc.)" for which we mean that $\mathcal{T}$ is a backward linear transfer with domain $\mathcal{P}(X)\times\mathcal{P}(Y)$ (or $\mathcal{P}(X)\times\mathcal{P}(X)$, $\mathcal{P}(Y)\times\mathcal{P}(X)$, etc.).

Remark 2.3.3. If $\mathcal{T}$ happens to be both a forward, and backward, linear transfer on $\mathcal{P}(X)\times\mathcal{P}(X)$, and is *symmetric* (i.e. $\mathcal{T}(\mu,\nu) = \mathcal{T}(\nu,\mu)$), then $T^+ f = -T^-(-f)$. This in particular, the case for optimal transport (see Section 2.6).

Remark 2.3.4. If $\mathcal{T}$ is a backward linear transfer, the definition implies the existence of an operator T^- satisfying $\mathcal{T}_\mu^*(g) = \int_X T^- g d\mu$. Regarding uniqueness, if there are two such operators T_1^-, then $\int_X T_1^- g d\mu = \int_X T_2^- g d\mu$ for all $g \in C(Y)$ and all $\mu \in D_1(\mathcal{T})$. In particular if $\{\delta_x\,;\, x \in X\} \subset D_1(\mathcal{T})$ (i.e. the first partial domain of $\mathcal{T}$ contains all the Dirac measures), then in fact $T_1^- g(x) = T_2^- g(x)$ for all g and all x. We therefore by abuse of language shall call in all cases T^- *the* operator associated to $\mathcal{T}$.

2.4 Kantorovich operators

We now focus our attention on the non-linear operators T^-. If $\mathcal{T}$ is a backward linear transfer such that $D_1(\mathcal{T})$ contains the Dirac measures, then its operator T^- satisfies $(T^-g)(x) = \mathcal{T}_{\delta_x}^*(g)$. Consequently, it inherits certain properties from the Legendre transform. To this end, we find it convenient to make the following definition. In the following, $USC(X)$ is the collection of upper semi-continuous functions $f : X \to \mathbb{R} \cup \{-\infty\}$; similarly $LSC(X)$ the collection of $g : X \to \mathbb{R} \cup \{+\infty\}$ such that $-g \in USC(X)$ (the same notation holds for Y in place of X).

Definition 2.4.1. A **backward Kantorovich operator** is a map $T^- : C(Y) \to USC(X)$ which is proper (i.e. $T^-g \not\equiv -\infty$ for every $g \in C(Y)$) and verifies the following properties:

1. T^- is *monotone*, i.e., if $g_1 \leq g_2$ in $C(Y)$, then $T^-g_1 \leq T^-g_2$.

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

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

3. T^- is *a convex operator*, that is for any $\lambda \in [0, 1]$, g_1, g_2 in $C(Y)$, we have

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

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

For completeness we also define the forward counterpart, although it can be derived from the backward definition (see the earlier remark on forward linear transfers). In the following, $LSC(Y)$ denotes the lower semi-continuous functions on Y, i.e. $g \in LSC(Y)$ if and only if $-g \in USC(Y)$.

Definition 2.4.2. A **forward Kantorovich operator** is a map $T^+ : C(X) \to LSC(Y)$ which is proper and verifies

1. T^+ is *monotone*, i.e., $f_1 \leq f_2$ in $C(X)$, then $T^+ f_1 \leq T^+ f_2$.

2. T^+ is *affine on the constants*, i.e., for any $c \in \mathbb{R}$ and $f \in C(X)$,

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

3. T^+ is *a concave operator*, that is for any $\lambda \in [0, 1]$, f_1, f_2 in $C(X)$, we have

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

4. T^+ is *upper semi-continuous*, in the sense that if $f_n \to f$ in $C(X)$ for the sup norm, then $\limsup_{n\to\infty} T^+ f_n \leq T^+ f$.

Remark 2.4.3. We call these operators 'Kantorovich' due to their connection with optimal transport (see Section 2.6). The notion of 'backward' is the interpretation that the operator T^- maps a function on Y 'backward' to a function on X (similarly for the notion of 'forward' and T^+). The forward/backward terminology coincides with Hamilton-Jacobi equations that are solved forward/backward in time (see Example 2.6.6) □

We shall see in Section 2.7 that under certain assumptions, backward linear transfers turn out to be the dual problems of the weak transports studied by Gozlan et al. [36]; for a precise statement, see Theorem 2.7.2. We also mention the recent paper by Alibert, Bouchitte, and Champion [1] which we learned of while preparing [10], in which they study Gozlan's weak transport costs and their Legendre transform, that we will see correspond to Kantorovich operators mapping into the set of bounded upper semi-continuous functions. We also mention the work of Roos [47] where they introduce the notion of *viscosity operators* as an axiomatic characterisation for viscosity solutions to the evolutive Hamilton-Jacobi equation.

We have the following which makes the connection between the backward linear transfers and backward Kantorovich operators.

Proposition 2.4.4. *1. Suppose T^- is a backward Kantorovich operator. Then the function defined by*

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

is a backward linear transfer provided the expression on the right-hand side is finite for at least one $(\mu,\nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. Note that the Dirac measures may not be in $D_1(\mathcal{T})$.

2. Conversely, suppose $\mathcal{T}$ is a backward linear transfer such that the Dirac measures $\{\delta_x\,;\, x \in X\}$ belong to $D_1(\mathcal{T})$. Then its associated operator T^- in Definition 2.3.1 is a backward Kantorovich operator.

In particular $T^- g(x)$ is given by

$$T^- g(x) = \sup_{\nu \in \mathcal{P}(Y)} \{\int_Y g d\nu - \mathcal{T}(\delta_x,\nu)\},$$

and if g_1, g_2 are such that $T^- g_1$, $T^- g_2$ are bounded, then

$$\|T^- g_2 - T^- g_2\|_\infty \leq \|g_1 - g_2\|_\infty,$$

where $\|h\|_\infty := \sup_{x \in X} |h(x)|$.

Proof. 1. Define for $\mu \in \mathcal{D}_1(\mathcal{T})$, $F_\mu(g) := \int_X T^- g d\mu$, which is necessarily finite for all g by choice of μ. Since $g \mapsto T^- g$ is convex and lower semi-continuous on $C(Y)$, it is convex and weakly lower semi-continuous; in particular, this implies F_μ is convex and lower semi-continuous on $C(Y)$. Therefore, by Fenchel-Moreau, $F_\mu(g) = \sup_{\nu \in \mathcal{M}(Y)}\{\int_Y g d\nu - F_\mu^*(\nu)\}$, where

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

We will be done if we can show that $F_\mu^*(\nu) = \mathcal{T}(\mu,\nu)$ for all $\nu \in \mathcal{P}(Y)$. The equality holds when $\nu \in \mathcal{P}(Y)$ by definition of $\mathcal{T}$, so suppose $\nu \in \mathcal{M}(Y)$ with $\nu(Y) = \lambda \neq 1$. Taking $g(x) \equiv n \in \mathbb{Z}$, we have $T^-(g) = T^-(0+n) = n + T^-(0)$, and we obtain

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

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

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

2. First we note that if $\{\delta_x\,; x \in X\} \subset D_1(\mathcal{T})$, then in fact $T^- g \in USC(X)$. Indeed, if $x_n \to x$ in X, extract a subsequence so that

$$\limsup_{n\to\infty} T^- g(x_n) = \lim_{j\to\infty} T^- g(x_{n_j}).$$

From the expression

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

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

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

The remaining properties of monotonicity, convexity, and affine on constants in Definition 2.3.1 follow immediately from the expression (2.5). For the lower semi-continuity property, suppose $g_n \to g \in C(Y)$ for the sup norm.

Then from (2.5), we have

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

hence

$$T^{-}g \le \liminf_{n\to\infty} T^{-}g_n(x).$$

Finally, if $g_1, g_2 \in C(Y)$ are such that $T^{-}g_1$, $T^{-}g_2$ are bounded, then we can repeat exactly the same estimate as above to have

$$T^{-}g_1(x) \le T^{-}g_2(x) + \|g_1-g_2\|_\infty,$$

which, with the assumption that $T^{-}g_1$ and $T^{-}g_2$ are bounded, means we may subtract and write

$$T^{-}g_1(x) - T^{-}g_2(x) \le \|g_1-g_2\|_\infty.$$

Interchanging g_1 and g_2 yields the 1-Lipschitz estimate for T^{-}. □

With a further assumption on $\mathcal{T}$, we can ensure that T^{-} maps into $USC_b(X)$ rather than $USC(X)$ (which in particular, implies $T^{-}g$ is bounded).

Proposition 2.4.5. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(Y)$ such that $D_1(\mathcal{T})$ contains all the Dirac measures. If*

$$\sup_{x\in X}\inf_{\nu\in\mathcal{P}(Y)} \mathcal{T}(\delta_x,\nu) < +\infty,$$

then

1. *The associated backward Kantorovich operator T^{-} maps $C(X)$ to $USC_b(X)$.*
2. *$D_1(\mathcal{T}) = \mathcal{P}(X)$.*

Proof. 1. The only thing which we need to show is that $T^{-}g$ is bounded below, since we already know $T^{-}g$ belongs to $USC(X)$ for all $g \in C(Y)$ by the above proposition. From Proposition 2.4.4, we have

$$T^{-}g(x) = \sup_{\nu \in \mathcal{P}(Y)} \{ \int_Y g d\nu - \mathcal{T}(\delta_x, \nu) \}$$

so that we immediately have

$$\begin{aligned} \inf_{x \in X} T^{-}g(x) &= \inf_{x \in X} \sup_{\nu \in \mathcal{P}(Y)} \{ \int_Y g d\nu - \mathcal{T}(\delta_x, \nu) \} \\ &\geq \inf(g) - \sup_{x \in X} \inf_{\nu \in \mathcal{P}(Y)} \mathcal{T}(\delta_x, \nu) > -\infty. \end{aligned}$$

2. If $D_1(\mathcal{T})$ contains the Dirac measures, then for each $x \in X$, there exists $\nu_x \in \mathcal{P}(Y)$, such that $\mathcal{T}(\delta_x, \nu_x) = \inf_{\nu \in \mathcal{P}(Y)} \mathcal{T}(\delta_x, \nu)$ by compactness of Y and weak* lower semi-continuity of $\mathcal{T}$. Consequently, since $\mathcal{T}$ is jointly convex then any empirical measure of the form $\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ is in $D_1(\mathcal{T})$ with

$$\begin{aligned} \mathcal{T}(\mu_n, \nu_n) \leq \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(\delta_{x_i}, \nu_{x_i}) &= \frac{1}{n} \sum_{i=1}^{n} \inf_{\nu \in \mathcal{P}(Y)} \mathcal{T}(\delta_{x_i}, \nu) \\ &\leq \sup_{x \in X} \inf_{\nu \in \mathcal{P}(Y)} \mathcal{T}(\delta_x, \nu) < +\infty, \end{aligned}$$

where $\nu_n := \frac{1}{n} \sum_{i=1}^{n} \nu_{x_i}$.

Now let $\mu \in \mathcal{P}(X)$. Since the Dirac measures are the extreme points of $\mathcal{P}(X)$, the measure μ is a weak* limit of empirical measures $\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ as $n \to \infty$. The measures μ_n have associated ν_n such that $\mathcal{T}(\mu_n, \nu_n) \leq \sup_{x \in X} \inf_{\nu \in \mathcal{P}(Y)} \mathcal{T}(\delta_x, \nu) < +\infty$ as above. Then by weak* compactness of $\mathcal{P}(Y)$, we may extract a subsequence so that $\nu_{n_j} \to \nu$ for some $\nu \in \mathcal{P}(Y)$. Consequently by weak* lower semi-continuity of $\mathcal{T}$, we conclude that

$$\mathcal{T}(\mu, \nu) \leq \liminf_{j \to \infty} \mathcal{T}(\mu_{n_j}, \nu_{n_j}) \leq \sup_{x \in X} \inf_{\nu \in \mathcal{P}(Y)} \mathcal{T}(\delta_x, \nu) < +\infty.$$

$\square$

Proposition 2.4.4 suggests that for an arbitrary map T^- from $C(Y)$ to Borel-measurable functions on X that does not satisfy the conditions for a backward Kantorovich operator, the functional

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

may not define a backward linear transfer.

Indeed, consider $T^- g := e^g$. Then $\mathcal{T}(\mu,\nu) := \sup_{g\in C(X)}\{\int_X g d\nu - \int_X e^g d\mu\}$ can be checked to be proper, and fixing $\mu \in D_1(\mathcal{T})$, a property of the Legendre transform, is that $\mathcal{T}^*_\mu(g+c) = \mathcal{T}^*_\mu(g)+c$ for any constant $c \in \mathbb{R}$ and all $g \in C(Y)$. Therefore the equality $\mathcal{T}^*_\mu(g) = \int_X e^g d\mu$ cannot hold, since that would imply (taking $g \equiv 0$ and $c = 1$), $1+\mathcal{T}^*_\mu(0) = \mathcal{T}^*_\mu(0+1) = e$ while at the same time (taking $g \equiv 0$ and $c = 2$), it must hold also that $2 + \mathcal{T}^*_\mu(0) = e^2$, which is impossible. Therefore, we make the following definition.

Definition 2.4.6 (Backward linear coupling)**.** $\mathcal{T}$ is a **backward linear coupling** if there exists a map $T^- : C(Y) \to \mathcal{B}^b(X)$, such that

$$\mathcal{T}(\mu,\nu) = \sup_{g\in C(Y)}\{\int_Y g d\nu - \int_X T^- g d\mu\}, \quad \mu \in D_1(\mathcal{T}), \nu \in \mathcal{P}(Y).$$

Remark 2.4.7. Every backward linear transfer is a backward linear coupling. On the other hand, note that a backward linear coupling $\mathcal{T}$ has $\mathcal{T}^*_\mu(g) \leq \int_X T^- g d\mu$ for all $\mu \in D_1(\mathcal{T})$ and $g \in C(X)$.

2.5 First examples of linear transfers

So far we have defined the notions of backward (resp., forward) linear transfers, but have not provided any examples to suggest the large variety of functionals which belong to this class. The aim of this section is to highlight a few of the basic ones. As usual, we will present examples on $\mathcal{P}(X)\times\mathcal{P}(Y)$, with the implicit assumption that they should always be taken identically $+\infty$ outside $\mathcal{P}(X)\times\mathcal{P}(Y)$ when viewed as a function on $\mathcal{M}(X)\times\mathcal{M}(Y)$.

2.5.1 Convex energies

A class of examples of linear transfers is the *convex energies* - functionals whose dependence on the pair (μ, ν) is trivially only through ν.

Let $I : \mathcal{P}(Y) \to \mathbb{R}$ be a bounded below, convex, weak*-lower semi-continuous function on $\mathcal{P}(Y)$. Consider

$$\mathcal{T}(\mu, \nu) := I(\nu) \quad \text{for all } (\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y). \tag{2.6}$$

Then $\mathcal{T}_\mu^*(g) = I^*(g) = \int_Y I^*(g) d\mu$, so $\mathcal{T}$ is a backward linear transfer with corresponding backward Kantorovich operator $T^- g \equiv I^*(g)$ (a constant function of x). Some choices of I include the following.

Example 2.5.1 (The null transfer)**.** This is simply the trivial mapping $\mathcal{N}(\mu, \nu) = 0$ for all probability measures μ on X and ν on Y, which achieved by taking $I \equiv 0$ in (2.6). One can then compute $I^*(g) = \sup_{y \in Y} g(y)$, and it is easy to see that it is both a backward and forward linear transfer with Kantorovich operators,

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

This is also a particular case of a result from optimal mass transport (which we discuss in Section 2.6) when the cost is identically zero.

Example 2.5.2 (Potential energy)**.** If I is the linear functional $I(\nu) = \int_Y V(y)\, d\nu(y)$, where V is a bounded below lower semi-continuous potential on Y, then for every $x \in X$,

$$T^- g \equiv I^*(g) = \sup\{\int_Y (g - V) d\nu\,;\, \nu \in \mathcal{P}(Y)\} = \sup_{y \in Y}\{g(y) - V(y)\}.$$

Example 2.5.3 (Relative entropy)**.** Fix any reference measure $\nu_0 \in \mathcal{P}(Y)$, and define I as the relative entropy with respect to ν_0, that is

$$I_{\nu_0}(\nu) := \begin{cases} \int_Y \frac{d\nu}{d\nu_0} \log(\frac{d\nu}{d\nu_0}) d\nu_0 & \text{if } \nu \text{ is absolutely continuous with respect to } \nu_0 \\ +\infty & \text{otherwise,} \end{cases}$$

where $\frac{d\nu}{d\nu_0}$ is the Lebesgue-Radon-Nikodym derivative of ν with respect to ν_0, i.e. $d\nu = \frac{d\nu}{d\nu_0} d\nu_0$. Then

$$T^- g \equiv I^*(g) = \sup\{\int_Y [g\lambda - \lambda \log(\lambda)]\, d\nu_0\,;\ \lambda \geq 0,\ \int_Y \lambda d\nu_0 = 1\}. \tag{2.7}$$

We can actually compute exactly the value of the supremum in (2.7), which we state in the following proposition (the following proof is provided in [35], Proposition 2.9 for a more general context).

Proposition 2.5.4.

$$T^- g \equiv \sup\{\int_Y [g\lambda - \lambda \log(\lambda)]\, d\nu_0\,;\ \lambda \geq 0,\ \int_Y \lambda d\nu_0 = 1\} = \log \int_Y e^g d\nu_0$$

Proof. Consider for fixed $s \in \mathbb{R}$, the function $t \mapsto st - t\log(t)$ for $t \geq 0$. This achieves a maximum $\sup_{t\geq 0}\{st - t\log(t)\} = e^{s-1}$. Identify $s = g(y)$ and $t = \lambda(y)$, we therefore have

$$\lambda g - \lambda \log \lambda \leq e^{g-1}. \tag{2.8}$$

Hence using this inequality (2.8) in (2.7), we have $T^- g(x) \leq \int_Y e^{g-1} d\nu_0$. From the fact that $T^- g = T^-(g+t) - t$ for all $t \in \mathbb{R}$, it must be the case that

$$T^- g \leq \inf_{t\in\mathbb{R}} \int_Y (e^{g-1+t} - t) d\nu_0.$$

We can now compute the infimum in the inequality: Let $F(t) := \int_Y (e^{g+t-1} - t) d\nu_0$ for $t \in \mathbb{R}$. Since F is convex in t, the minimal value occurs exactly where $F'(t_0) = 0$, i.e. where $\int_Y e^{g+t_0-1} d\nu_0 = 1$, which implies that $t_0 = 1 - \log \int_Y e^g d\nu_0$. This means that

$$T^- g \leq \inf_{t\in\mathbb{R}} F(t) = F(t_0) = \log \int_Y e^g d\nu_0.$$

On the other hand, consider the measure ν with density $\lambda = \frac{e^g}{\int_Y e^g d\nu_0}$ with respect to ν_0. Then one can check for this λ

$$\int_Y [g\lambda - \lambda \log(\lambda)]\, d\nu_0 = \log \int_Y e^g d\nu_0$$

giving the reverse inequality $T^- g \geq \log \int_Y e^g d\nu_0$. $\square$

Remark 2.5.5. The above shows that the relative entropy functional, defined on $\mathcal{P}(X) \times \mathcal{P}(X)$ via

$$\mathcal{H}(\mu, \nu) = \begin{cases} \int_Y \frac{d\nu}{d\mu} \log(\frac{d\nu}{d\mu}) d\mu & \text{if } \nu \text{ is absolutely continuous with respect to } \mu \\ +\infty & \text{otherwise} \end{cases}$$

is *not* a backward linear transfer, since $\mathcal{H}^*_\mu(g) = \log \int_X e^g d\mu$. This motivates us to introduce a further extension of linear transfers, *convex transfers*, which we will discuss in Section 2.9.

Example 2.5.6 (Variance)**.** For the (negative) of the variance,

$$I(\nu) := -\text{var}(\nu) := \left| \int_Y y\ d\nu(y) \right|^2 - \int_Y |y|^2\, d\nu(y),$$

the associated Kantorovich map is given by

$$T^- g \equiv \sup_{y \in Y} \{ \widehat{g + q}(y) - |y|^2 \},$$

where q is the quadratic function $q(x) = \frac{1}{2}|x|^2$ and $\hat{h}$ is the *concave envelope* of the function h. The details for this example are given in Example 2.8.4 below.

2.5.2 Linear transfers induced from positively homogenous operators

Let $T : C(Y) \to C(X)$ be a bounded linear positive operator such that $T(1) = 1$; we call such an operator a Markov operator. One can associate a backward linear transfer in the following way:

$$\mathcal{T}_T(\mu,\nu) := \begin{cases} 0 & \text{if } \nu = T^*(\mu) \\ +\infty & \text{otherwise,} \end{cases} \tag{2.9}$$

where $T^* : \mathcal{M}(X) \to \mathcal{M}(Y)$ is the adjoint operator. It is immediate to check that $\nu \mapsto \mathcal{T}_T(\mu,\nu)$ is convex and weak* lower semi-continuous, and

$$(\mathcal{T}_T)^*_\mu(g) = \int_Y g dT^*(\mu) = \int_Y Tg d\mu.$$

Therefore $T^- g := Tg$ is the corresponding backward Kantorovich operator.

Defining $\pi_x := T^*(\delta_x)$, we have that $T^- f(x) = \int_Y f(y) d\,\pi_x(y)$ and that

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

Example 2.5.7 (The prescribed push-forward transfer)**.** Let F be a continuous map from X to Y. Given $\mu \in \mathcal{P}(X)$, define the push-forward measure

$$\nu := F_\# \mu \in \mathcal{P}(Y) \quad \text{by } \nu(B) = \mu(F^{-1}(B)) \text{ for all Borel sets } B \subset Y.$$

The operator $T^- g := g \circ F$ is readily seen to be a Markov operator and $T^*(\mu) = F_\# \mu$; hence (2.9) becomes in this case

$$\mathcal{T}_F(\mu,\nu) := \begin{cases} 0 & \text{if } \nu = F_\# \mu, \\ +\infty & \text{otherwise} \end{cases} \tag{2.10}$$

and the backward Kantorovich operator is given by $T^- g = g \circ F$. In the case when $X = Y$ and $F(x) = x$, $\mathcal{T}_F$ is both a forward and backward linear transfer, with $T^- f = f = T^+ f$, which we call the *identity transfer*.

As a natural extension of the previous example, let $A \in LSC(X)$, $A \not\equiv +\infty$, and consider the Kantorovich operator,

$$T^- g(x) := g \circ F(x) - A(x).$$

Then the induced backward linear transfer is

$$\mathcal{T}(\mu,\nu) = \begin{cases} \int_X A d\mu & \text{if } \nu = F_{\#}\mu,\ \int_X A d\mu < +\infty, \\ +\infty & \text{otherwise.} \end{cases}$$

We shall see in Section 5.6.1 that this backward linear transfer is related to *ergodic optimisation in the expanding case* [25], where it is of interest to minimise the action $\mu \mapsto \int_X A d\mu$, for a given potential A, among all F-invariant measures μ (and other related generalisations).

Example 2.5.8 (The prescribed Skorokhod transfer)**.** Let $X = Y$ be a smooth compact Riemannian manifold without boundary, and $(B_t)_{t\geq 0}$ denote Brownian motion on X, and S the corresponding class of (possibly randomized) stopping times. For a fixed $\tau \in S$, define

$$\mathcal{T}_\tau(\mu,\nu) := \begin{cases} 0 & \text{if } B_0 \sim \mu \text{ and } B_\tau \sim \nu. \\ +\infty & \text{otherwise,} \end{cases} \tag{2.11}$$

where $Z \sim \rho$ if Z is a random variable with distribution ρ. It is easy to verify $\nu \mapsto \mathcal{T}_\tau(\mu,\nu)$ is convex and weak* lower semi-continuous, and we have

$$(\mathcal{T}_\tau)^*_\mu(g) = \int_Y g d\nu = \mathbb{E}[g(B_\tau)] = \int_Y \mathbb{E}\left[g(B_\tau)|B_0 = x\right] d\mu(x)$$

so it is in fact a backward linear transfer with backward Kantorovich operator $T^- g(x) := \mathbb{E}\left[g(B_\tau)|B_0 = x\right]$.

Example 2.5.9 (Prescribed marginals)**.** Let π be a probability measure on $X \times Y$, and denote $\pi_1 := \mathrm{proj}_{X\#}\pi$ and $\pi_2 := \mathrm{proj}_{Y\#}\pi$ as its *marginal* on X (resp., on Y). Here we are using the push-forward measure notation of Example 2.5.7, which in this context, means $\pi_1(A) := \pi(\mathrm{proj}_X^{-1}(A)) = \pi(A \times Y)$ (similarly for $\mathrm{proj}_{Y\#}\pi$). Define

$$\mathcal{T}_\pi(\mu,\nu) = \begin{cases} 0 & \text{if } \mu = \pi_1 \text{ and } \nu = \pi_2. \\ +\infty & \text{otherwise,} \end{cases} \tag{2.12}$$

In this case, π_1 is the only member of $D_1(\mathcal{T}_\pi)$, and

$$(\mathcal{T}_\pi)^*_{\pi_1}(g) = \int_Y g d\pi_2 = \int_{X\times Y} g d\pi = \int_X \left[\int_Y g(y) d\pi_x(y)\right] d\pi_1(x)$$

where $(\pi_x)_x$ is the disintegration of π with respect to π_1. A similar expression holds for $(\mathcal{T}_\pi)^*_{\pi_2}(f)$. Therefore

$$T^- g(x) := \int_Y g(y) d\,\pi_x(y) \quad \text{and} \quad T^+ f(y) = \textstyle\int_X f(x) d\,\pi_y(x).$$

Note in this example that $T^- g$ is merely a (bounded above and below) Borel measurable function.

2.5.3 Balayage of measures

Given a closed convex cone $\mathcal{A} \subset C(X)$ that contains the non-negative constant functions and is closed under maxima, define a partial order relation between probability measures μ, ν, which we call an $\mathcal{A}$-*order*, via

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

Define the *balayage of measures* $\mathcal{B}$ on $\mathcal{P}(X) \times \mathcal{P}(X)$ via

$$\mathcal{B}(\mu, \nu) = \begin{cases} 0 & \text{if } \mu \prec_{\mathcal{A}} \nu \\ +\infty & \text{otherwise.} \end{cases} \tag{2.13}$$

It is easy to see that $\nu \mapsto \mathcal{B}(\mu, \nu)$ is convex and weak* lower semi-continuous, and that

$$\mathcal{B}^*_\mu(g) = \sup\left\{\int_X g d\nu\,;\ \nu \in \mathcal{P}(X),\, \mu \prec_{\mathcal{A}} \nu\right\}.$$

Proposition 2.5.10. *We have*

$$\sup\left\{\int_X g d\nu\,;\ \nu \in \mathcal{P}(X),\, \mu \prec_{\mathcal{A}} \nu\right\} = \int_X \left[\sup_{\delta_x \prec_{\mathcal{A}} \sigma} \int_X g(y) d\sigma(y)\right] d\mu(x)$$

and hence $\mathcal{B}$ is a backward linear transfer.

For the proof, we state the following result that has previously been established in [17] and based on the classical results of Strassen [56]. We call a *Markov kernel* a real-valued function $P : \mathcal{P}(X) \times X \to \mathbb{R}$, such that $P(\cdot, x)$ belongs to $\mathcal{P}(X)$, and $x \mapsto P(A, x)$ is Borel-measurable for any Borel set $A \subset X$.

Theorem 2.5.11 ([17], Theorem 2)**.** *Let $\mathcal{A}$ be a closed convex cone in $C(X)$ which contains the non-negative constant functions and is closed under maxima. If $\mu \prec_{\mathcal{A}} \nu$, then there exists a Markov kernel P, such that $\nu(A) = \int_X P(A, x) d\mu(x)$ for all Borel $A \subset X$, and $\delta_x \prec_{\mathcal{A}} P(\cdot, x)$ for all $x \in X$.*

In addition, the set of extreme points of $\{(\mu, \nu)\,;\, \mu \prec_{\mathcal{A}} \nu\}$ is contained in the set of measures (δ_x, σ), where $\delta_x \prec_{\mathcal{A}} \sigma$.

We also state the following measurable selection theorem, which shall be of use here, and also in later sections.

Proposition 2.5.12 ([7], Proposition 7.3.3)**.** *Let X be a metrizable space, Y a compact metrizable space, $D \subset X \times Y$ a closed subset, and $f : D \to \mathbb{R} \cup \{\pm\infty\}$ lower semi-continuous. Let $f^* : proj_X(D) \to \mathbb{R} \cup \{\pm\infty\}$, where $proj_X(D) := \{x \in X\,;\,(x, y) \in D$ for some $y \in Y\}$, be defined by*

$$f^*(x) = \inf_{y \in D_x} f(x, y)$$

where $D_x := \{y \in Y\,;\,(x, y) \in D\}$. The f^* is lower semi-continuous, and there exists a Borel-measurable function $\varphi : Proj_X(D) \to Y$ such that $Graph(\varphi) \subset D$ and $f(x, \varphi(x)) = f^*(x)$ for all $x \in Proj_X(D)$.

The above results readily leads to the proof of Proposition 2.5.10.

Proof. Of Proposition 2.5.10. By Theorem 2.5.11, we conclude that whenever $\mu \prec_{\mathcal{A}} \nu$, we have a Markov kernel P with the specified properties, so that

$$\int_X g(y)d\nu(y) = \int_X \left[\int_X g(y)P(dy,x)\right] d\mu(x)$$
$$\leq \int_X \left(\sup_{\delta_x \prec_{\mathcal{A}} \sigma} \int_X g(y)d\sigma(y)\right) d\mu(x).$$

On the other hand, let σ_x achieve the supremum for each $x \in X$ (the supremum is achieved since the set of $\delta_x \prec_{\mathcal{A}} \sigma$ is weak* closed in $\mathcal{P}(X)$ and therefore weak* compact). Then by the measurable selection Proposition 2.5.12, $P(A,x) := \sigma_x(A)$ is a Markov kernel, and $\nu(A) := \int_X P(A,x)d\mu(x)$ satisfies $\mu \prec_{\mathcal{A}} \nu$. □

Defining $T^- g(x) := \sup_{\delta_x \prec_{\mathcal{A}} \sigma} \int_X g(y)d\sigma(y)$, we conclude that $\mathcal{B}$ is a backward linear transfer. We have the following representation for T^-:

Proposition 2.5.13.

$$T^- g(x) = \hat{g}(x) := \inf\{h(x)\,;\, h \in -\mathcal{A},\, h \geq g \text{ on } X\}.$$

Proof. For any admissible h, we have $\int_X g d\sigma \leq \int_X h d\sigma$, hence

$$T^- g(x) = \sup_{\delta_x \prec_{\mathcal{A}} \sigma} \int_X g(y)d\sigma(y) \leq \sup_{\delta_x \prec_{\mathcal{A}} \sigma} \int_X h(y)d\sigma(y) \leq h(x),$$

the last inequality following since $h \in -\mathcal{A}$. Hence $T^- g(x) \leq \hat{g}(x)$.

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

$$h_\epsilon(x) := \min\{h_{\epsilon,x_1}(x), \ldots, h_{\epsilon,x_n}(x)\}$$

since $\mathcal{A}$ is closed under maxima, it follows that $h_\epsilon \in -\mathcal{A}$. Moreover, each

$x \in X$ belongs to $B_{r_{x_i}}(x_i)$ for some $i \in \{1, \ldots, n\}$, hence

$$g(x) \leq h_\epsilon(x) \leq g(x) + \epsilon \quad \text{for all } x \in X.$$

Now for any σ such that $\delta_x \prec_{\mathcal{A}} \sigma$, we have

$$\epsilon + \int_X g d\sigma \geq \int_X h_\epsilon d\sigma \geq h_\epsilon(x) \quad \text{since } h_\epsilon \in -\mathcal{A},$$

and since $h_\epsilon(x) \geq \hat{g}(x)$, we conclude,

$$\epsilon + T^- g(x) \geq \epsilon + \int_X g(y) d\sigma(y) \geq \hat{g}(x).$$

As ϵ is arbitrary, we obtain the reverse inequality, which concludes the proof. $\square$

By a similar argument to that given above, one can show that $\mathcal{B}$ is also a *forward* linear transfer with a forward Kantorovich operator $T^+ f := \check{f}$, where

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

We now discuss particular choices of $\mathcal{A}$.

Example 2.5.14 (Convex order)**.** If X is a convex compact space in a locally convex topological vector space, then $\mathcal{A}$ can be taken to be the cone of continuous convex functions. In this case, $T^- g = \hat{g}$ is the *concave envelope* of g (i.e. the smallest concave function above g), since an infimum of concave functions is concave. This is a classical result of Choquet (see e.g. [16] Proposition 26.13).

Example 2.5.15 (Subharmonic order)**.** Consider now when $X \subset \mathbb{R}^n$ and $\mathcal{A}$ is the cone of Lipschitz subharmonic functions on some open set O containing X. In this case, we can identify $\mathcal{B}$ in another way using the following characterisation of Skorokhod.

Proposition 2.5.16 (Skorokhod, e.g. [30])**.** *Let $\mathcal{A}$ be the cone of Lips-*

chitz subharmonic functions on a domain X in $\mathbb{R}^n$. Then, the following are equivalent for two probability measures μ and ν on X.

1. *$\mu \prec_{\mathcal{A}} \nu$.*

2. *There exists a stopping time $\tau \in S$ with $\mathbb{E}[\tau] < \infty$ such that $B_0 \sim \mu$ and $B_\tau \sim \nu$, where S denote the collection of (random) Brownian stopping times with finite expectation.*

Therefore,

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

This identification of $\mathcal{B}$ provides various ways to write T^- in this case:

1. $T^- g = \hat{g}$ is the smallest Lipschitz superharmonic function above g.

2. $T^- g = \sup\limits_{\tau \in S} \mathbb{E}\left[g(B_\tau) | B_0 = x\right]$.

3. $T^- g = J_g$, where $J_g(x)$ is a viscosity solution for the heat variational inequality $\max\left\{g(x) - J(x), \Delta J(x)\right\} = 0$.

Note also that $\mathcal{B}(\mu, \nu) = \inf\limits_{\tau \in S} \mathcal{T}_\tau(\mu, \nu)$ where $\mathcal{T}_\tau$ is the "prescribed Skorokhod transfer" of Example 2.5.8.

2.6 Linear transfers and optimal mass transportation

In the Monge-Kantorovich theory of optimal transport (see e.g. [59, 60], or [48]), given a lower semi-continuous cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ one associates a map $\mathcal{T}_c$ on $\mathcal{P}(X) \times \mathcal{P}(Y)$ to be the optimal total transport cost between $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, via

$$\mathcal{T}_c(\mu, \nu) := \inf_{\pi \in \mathcal{K}(\mu, \nu)} \int_{X \times Y} c(x, y)\, d\pi(x, y), \tag{2.14}$$

where $\mathcal{K}(\mu,\nu)$ is the set of probability measures π on $X\times Y$ whose marginal $\mathrm{proj}_{X\sharp}\pi$ on X is μ, and marginal $\mathrm{proj}_{Y\sharp}\pi$ on Y is ν, (recall proj_X is the projection $(x,y)\mapsto x$ onto X; similarly for Y). Also recall the push-forward of a measure by a map ϕ is the measure $(\phi_\#)\mu(A) := \mu(\phi^{-1}(A))$.

The measures $\pi\in\mathcal{K}(\mu,\nu)$ are interpreted as *transport plans*, with $d\pi(x,y)$ representing the amount of mass which is sent from x to y, and costing $c(x,y)d\pi(x,y)$ to do so.

The optimal transport problem (2.14) is an infinite dimensional linear programming problem; as in the classical finite dimensional theory, it has a dual formulation first studied by Kantorovich (see e.g. [59], Theorem 1.3, and references therein).

Theorem 2.6.1 (Kantorovich duality [59]). *Let $c : X\times Y\to\mathbb{R}\cup\{+\infty\}$ be a lower semi-continuous function. Then*

$$\begin{aligned}\inf_{\pi\in\mathcal{K}(\mu,\nu)}\int_{X\times Y} c(x,y)\,d\pi(x,y) &= \sup\{\int_Y g d\nu - \int_X T^- g d\mu\,;\, g\in C(Y)\}\\ &= \sup\{\int_Y T^+ f d\nu - \int_X f d\mu\,;\, f\in C(X)\}\end{aligned}$$

where

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

Remark 2.6.2. If X and Y are not compact, then c should also be assumed to be bounded below; this additional hypothesis is not necessary when X and Y are compact as c is automatically bounded from below by the assumption of lower semi-continuity.

Proof. Note first that $\nu\mapsto\mathcal{T}_c(\mu,\nu)$ is convex and weak* lower semi-continuous. Indeed, for $\nu_1,\nu_2\in\mathcal{P}(Y)$, and fixed $\mu\in\mathcal{P}(X)$, the infimum is achieved for $\mathcal{T}_c(\mu,\nu_i)$ at some $\pi_i\in\mathcal{K}(\mu,\nu_i)$, $i=1,2$. It immediately follows that $\pi_\lambda := (1-\lambda)\pi_1+\lambda\pi_2$ is admissible for $\mathcal{T}_c(\mu,(1-\lambda)\nu_1+\lambda\nu_2)$. In addition, if $\nu_n\to\nu$, then for any $\pi_n\in\mathcal{K}(\mu,\nu_n)$, one may extract a subsequence converging to $\pi\in\mathcal{P}(X\times Y)$ by weak* compactness of $\mathcal{P}(X\times Y)$, and moreover,

$\pi \in \mathcal{K}(\mu, \nu)$. The lower semi-continuity then follows. Finally,

$$\begin{aligned}
(\mathcal{T}_c)^*_\mu(g) &= \sup_{\nu \in \mathcal{P}(Y)} \{ \int_Y g d\nu - \inf_{\pi \in \mathcal{K}(\mu,\nu)} \int_{X \times Y} c(x,y)\, d\pi(x,y) \} \\
&= \sup_{\nu \in \mathcal{P}(Y)} \sup_{\pi \in \mathcal{K}(\mu,\nu)} \{ \int_{X \times Y} (g(y) - c(x,y)) d\pi(x,y) \} \\
&= \sup_{\pi \in \mathcal{K}(\mu,\cdot)} \{ \int_{X \times Y} (g(y) - c(x,y)) d\pi(x,y) \}.
\end{aligned}$$

It is clear that the supremum is achieved by a π whose disintegration with respect to μ is given by $d\pi(x,y) = d\mu(x)\delta_{y_x}(y)$ where y_x is a value of y achieving $\sup_{y \in Y}\{g(y) - c(x,y)\}$. A similar argument holds for showing $\mathcal{T}_c$ is a forward linear transfer. □

Remark 2.6.3. In fact one can refine the duality further and show that

$$\inf_{\pi \in \mathcal{K}(\mu,\nu)} \int_{X \times Y} c(x,y)\, d\pi(x,y) = \sup\{ \int_Y \phi_1 d\nu - \int_X \phi_0 d\mu \}$$

where the supremum is over all pairs (ϕ_0, ϕ_1) such that $\phi_1 = T^- \phi_0$ and $\phi_0 = T^+ \phi_1$. This will be relevant for Bernard-Buffoni's work [5, 6] discussed in Chapter 4 on optimal transport and Aubry-Mather and weak KAM theory.

In the following examples, we highlight the backward/forward Kantorovich operators for a few special cases for the cost function.

Example 2.6.4 (The trivial Kantorovich transfer)**.** Let $c_1 \in USC_b(X)$, $c_2 \in LSC(Y)$, and define $c(x,y) := c_2(y) - c_1(x)$. Then c is lower semi-continuous, and

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

The Kantorovich operators are $T^+ f = c_2 + \inf_{x \in X}\{f(x) - c_1(x)\}$ and $T^- g = c_1 + \sup_{y \in Y}\{g(y) - c_2(y)\}$.

Example 2.6.5 (The Kantorovich-Rubinstein transport)**.** For the cost $c(x,y) = d_X(x,y)$ (recall d_X is the metric on X), the Kantorovich operators are given by

$$T^+f(y) = \inf_{x\in X}\{f(x)+d_X(x,y)\} \quad \text{and} \quad T^-g(x) = \sup_{y\in X}\{g(y)-d_X(x,y); y \in X\}.$$

A computation shows that $T^- \circ T^+ = T^+$ and $T^+ \circ T^- = T^-$. Hence by taking $f = T^-h$ for $h \in C(X)$, we have

$$\mathcal{T}_c(\mu,\nu) = \sup_{f\in C(X)}\{\int_X T^+f d\nu - \int_X f d\mu\} \geq \sup_{h\in C(X)}\{\int_X T^-h\, d(\nu-\mu)\}$$

while at the same time, from $g \leq T^-g$ we have

$$\mathcal{T}_c(\mu,\nu) = \sup_{g\in C(X)}\{\int_X g d\nu - \int_X T^-g d\mu\} \leq \sup_{g\in C(X)}\{\int_X T^-g\, d(\nu-\mu)\}.$$

Hence, $\mathcal{T}_c(\mu,\nu) = \sup_{g\in C(X)}\{\int_X T^-g\, d(\nu-\mu)\}$. It is readily checked that T^-g is 1-Lipschitz for any $g \in C(X)$, and moreover, every g which is λ-Lipschitz, $\lambda \leq 1$, satisfies $T^-g = g$. Therefore,

$$\mathcal{T}(\mu,\nu) = \sup\left\{\int_X g\, d(\nu-\mu); g \in C(X), \|g\|_{\text{Lip}} \leq 1\right\}$$

where $\|g\|_{\text{Lip}} := \sup_{x\neq y}\frac{|g(y)-g(x)|}{d_X(x,y)}$. (The above discussion holds also for any other lower semi-continuous metric d, not necessarily the one defining the topology.)

Example 2.6.6 (Lagrangian cost [5])**.** This example links the Kantorovich backward and forward operators with the respective backward and forward Hopf-Lax operators that solve first order Hamilton-Jacobi equations, which we will see is intimately connected to weak KAM theory (see Chapter 4).

On a given smooth compact Riemannian manifold M without boundary (take e.g. $M = \mathbb{T}^n$ the flat torus), with tangent bundle TM, let $L : TM \to \mathbb{R} \cup \{+\infty\}$ be a given *Tonelli Lagrangian* (see Chapter 4 for a definition), and consider the cost

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

For μ and ν two probability measures on M, consider the optimal transport with cost c_L.

Then, $T^+f(x) = V_f(1,x)$, where $V_f(t,x)$ is the "value functional"

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

V_f is in fact a viscosity solution for the Hamilton-Jacobi equation

$$\left\{\begin{array}{rcl} \partial_t V + H(x,\nabla_x V) &=& 0 \ \text{ on } \ (0,+\infty)\times M, \\ V(0,x) &=& f(x). \end{array}\right.$$

(see e.g. [20], Theorem 7.2.8) where H is the Hamiltonian (see Chapter 4). Similarly, the backward Kantorovich operator is given by $T^-g(y) = W_g(0,y)$, where $W_g(t,y)$ is the value functional

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

which is a viscosity solution for the backward Hamilton-Jacobi equation

$$\left\{\begin{array}{rcl} \partial_t W + H(x,\nabla_x W) &=& 0 \ \text{ on } \ [0,1)\times M, \\ W(1,y) &=& g(y). \end{array}\right.$$

Example 2.6.7 (The Brenier-Wasserstein distance [12])**.** 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

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

Here, the Kantorovich operators are

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

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

2.7 Representation and envelopes of linear transfers

2.7.1 Linear transfers and optimal weak transports

A natural generalisation of optimal transport is *optimal weak transport*, introduced by Gozlan et. al. [36]. Weak transport is concerned with cost functions $c(x,\sigma)$ between points x and *distributions* $\sigma \in \mathcal{P}(Y)$, and thus is a generalisation of the cost-minimizing optimal transport of the last section which concerns cost functions between points x and y.

Definition 2.7.1. Let $c : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a bounded below, lower semi-continuous function with $\sigma \mapsto c(x,\sigma)$ convex. The **optimal weak transport problem with cost** c from $\mu \in \mathcal{P}(X)$ to $\nu \in \mathcal{P}(Y)$ is

$$V(\mu,\nu) := \inf_{\pi}\{\int_X c(x,\pi_x)\, d\mu(x); \pi \in \mathcal{K}(\mu,\nu)\} \tag{2.15}$$

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

This problem has been the study of several recent papers [1], [3], and [36], where existence, duality, and certain properties of optimisers of (2.15) have been established. In the next theorem we show that weak transports are backward linear transfers, and, conversely, appropriate backward linear transfers can be represented as weak transports. The implication of this is that, roughly speaking, backward linear transfers are the dual formulation of weak transports.

Theorem 2.7.2. *1. Any optimal weak transport cost $c(x,\sigma)$ satisfying the properties in Definition 2.7.1 defines a backward linear transfer via*

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

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

2. *Conversely, if $\mathcal{T}$ is a backward linear transfer such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$, then $\mathcal{T}$ is an optimal weak transport with cost function $c(x,\sigma) := \mathcal{T}(\delta_x, \sigma)$.*

In order to prove that a weak transport $V(\mu,\nu)$ is a backward linear transfer, we wish to show convexity and weak* lower semi-continuity for $\nu \mapsto V(\mu,\nu)$. It turns out that the weak* lower semi-continuity is more delicate to establish than standard optimal transport due to the disintegration π_x of π with respect to its first marginal; two recent papers, [1] (see also [3]) have closed this gap.

Theorem 2.7.3 ([1], Theorem 2.9, Lemma 3.5)**.** *Let c satisfy the conditions given in Definition 2.7.1. Then the map $\pi \mapsto I(\pi) := \int_X c(x,\pi_x)d\mu(x)$ is weak* lower semi-continuous, where $\mu := Proj_{X\#}\pi$ is the first marginal of π. Consequently, the problem (2.15) admits a minimiser.*

With these properties, we give the rest of the details for the proof of Theorem 2.7.2.

Proof. (Of Theorem 2.7.2)

1. We show that $\nu \mapsto \mathcal{T}(\mu,\nu)$ is convex. This is essentially a consequence of the assumption that $\sigma \mapsto c(x,\sigma)$ is convex. Indeed, fix $\nu_1, \nu_2 \in \mathcal{P}(Y)$, and find for a fixed $\epsilon > 0$, $\pi^1 \in \mathcal{K}(\mu,\nu_1)$ and $\pi^2 \in \mathcal{K}(\mu,\nu_2)$ such that
$$\int_X c(x,\pi^i_x)\,d\mu(x) \le \mathcal{T}_c(\mu,\nu_i) + \epsilon \quad \text{for } i = 1,2.$$
Consider $\pi \in \mathcal{P}(X \times Y)$ which is defined via
$$d\pi(x,y) := (\lambda d\pi^1_x(y) + (1-\lambda)d\pi^2_x(y))d\mu(x).$$
With $\nu_\lambda := \lambda\nu_1 + (1-\lambda)\nu_2$, we have $\pi \in \mathcal{K}(\mu,\nu_\lambda)$, and therefore, by convexity of c in the second variable, we have
$$\begin{aligned}\mathcal{T}_c(\mu,\nu_\lambda) &\le \int_X c(x,\pi_x)\,d\mu(x) \le \int_X \lambda c(x,\pi^i_x)\,d\mu(x) + \int_X (1-\lambda)c(x,\pi^i_x)\,d\mu(x)\\ &\le \lambda\mathcal{T}_c(\mu,\nu_1) + (1-\lambda)\mathcal{T}_c(\mu,\nu_2) + \epsilon,\end{aligned}$$

which implies convexity of $\nu \mapsto \mathcal{T}_c(\mu, \nu)$.

We now show that $\nu \mapsto \mathcal{T}(\mu, \nu)$ is weak* lower semi-continuous. This is almost immediate from Theorem 2.7.3. Indeed, let $\nu_n \to \nu$ and select a subsequence so that $\lim_{j\to\infty} \mathcal{T}(\mu, \nu_{n_j}) = \liminf_{n\to\infty} \mathcal{T}(\mu, \nu_n)$. For notational ease, relabel the subsequence to ν_n. For each n, let $\pi^n \in \mathcal{K}(\mu, \nu_n)$ be optimal for the weak transport $\mathcal{T}(\mu, \nu_n)$, which we know exists by Theorem 2.7.3. By compactness of $\mathcal{P}(X \times Y)$, there is a further subsequence (which we again relabel to n) so that $\pi^n \to \pi$ for some $\pi \in \mathcal{K}(\mu, \nu)$. This π is admissible for $\mathcal{T}(\mu, \nu)$ so

$$\mathcal{T}(\mu, \nu) \leq \int_X c(x, \pi_x)\, d\mu(x)$$

while at the same time by Theorem 2.7.3, we have

$$\int_X c(x, \pi_x)\, d\mu(x) \leq \liminf_{n\to\infty} \int_X c(x, \pi_x^n)\, d\mu(x) = \liminf_{n\to\infty} \mathcal{T}(\mu, \nu_n)$$

which concludes the proof of lower semi-continuity. Finally, we compute the Legendre transform,

$$\begin{aligned}
\mathcal{T}_\mu^*(g) &= \sup_{\nu\in\mathcal{P}(Y)} \{\int_Y g\, d\nu - \mathcal{T}(\mu, \nu)\} \\
&= \sup_{\nu\in\mathcal{P}(Y)} \sup_{\pi\in\mathcal{K}(\mu,\nu)} \{\int_Y g(y)\, d\nu(y) - \int_X c(x, \pi_x)\, d\mu(x)\} \\
&= \sup_{\pi\in\mathcal{K}(\mu,\cdot)} \{\int_X \left[\int_Y g(y) d\pi_x(y) - c(x, \pi_x)\right] d\mu(x)\} && (2.16) \\
&\leq \int_X \sup_{\sigma\in\mathcal{P}(Y)} \{\int_Y g d\sigma - c(x, \sigma)\} d\mu(x) && (2.17) \\
&= \int_X T^- g d\mu.
\end{aligned}$$

On the other hand, by compactness of $\mathcal{P}(Y)$ and lower semi-continuity of c, for each x the supremum in (2.17) is achieved by some σ_x in a measurable way (see the measurable selection proposition 2.5.12) so that

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

Define $\pi \in \mathcal{P}(X \times Y)$ via $d\pi(x,y) := d\sigma_x(y) d\mu(x)$. Denoting $\nu := \mathrm{Proj}_{Y\#}\pi$, we have $\pi \in \mathcal{K}(\mu, \nu)$. Hence, this π is admissible in the supremum (2.16), so that

$$\begin{aligned}
\mathcal{T}_\mu^*(g) &\geq \int_Y \left[\int_X g(y) d\sigma_x(y)\ d\mu(x) - \int_X c(x, \sigma_x) \right] d\mu(x) \\
&= \int_X T^- g(x) d\mu(x),
\end{aligned}$$

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

2. Let $\mathcal{T}$ be a backward linear transfer with backward Kantorovich operator T^-. Define $c(x, \sigma) := \mathcal{T}(\delta_x, \sigma)$. Then c is bounded below, proper, lower semi-continuous, and $\sigma \mapsto c(x, \sigma)$ is convex, so we may define the weak transport

$$\tilde{\mathcal{T}}(\mu, \nu) := \inf_{\pi \in \mathcal{K}(\mu,\nu)} \int_X c(x, \pi_x)\, d\mu(x)$$

for $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ (and $+\infty$ otherwise on $\mathcal{M}(X) \times \mathcal{M}(Y)$). In an identical way to the proof of item 1, we have for any μ,

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

where $\tilde{T}^- g(x) := \sup_{\sigma \in \mathcal{P}(Y)} \{ \int_Y g d\sigma - c(x, \sigma) \}$. On the other hand, since $c(x, \sigma) = \mathcal{T}(\delta_x, \sigma)$ and $\{\delta_x\,;\, x \in X\} \subset D_1(\mathcal{T})$, we also have that $\tilde{T}^- g(x) = \mathcal{T}_{\delta_x}^*(g) = T^- g(x)$ for all $x \in X$. Hence $\tilde{\mathcal{T}}_\mu^*(g) = \int_X T^- g d\mu = \mathcal{T}_\mu^*(g)$ for all $\mu \in \mathcal{P}(X)$ and consequently, $\mathcal{T}(\mu, \nu) = \tilde{\mathcal{T}}(\mu, \nu)$.

$\square$

2.7.2 Linear transfer envelopes

Suppose $\mathcal{T}$ is merely a proper, convex, bounded below, weak* lower semi-continuous functional, but is not a backward linear transfer, or even a backward linear coupling. Is there a "canonical" backward linear transfer associated to $\mathcal{T}$? We therefore consider in the following linear transfer "envelopes", in an analogy to the convex/concave envelopes of functions.

Proposition 2.7.4. *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex, bounded below, weak* lower semi-continuous functional such that $\{\delta_x\,;\, x \in X\} \subset D_1(\mathcal{T})$. Denote $\overline{\mathcal{T}}$ to be the weak transport defined by (2.15) associated to the weak transport cost $c(x,\sigma) := \mathcal{T}(\delta_x, \sigma)$, and let*

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

Then, $\mathcal{T} \leq \overline{\mathcal{T}} \leq \tilde{\mathcal{T}}$, and $\overline{\mathcal{T}}$ is the only weak transport between $\mathcal{T}$ and $\tilde{\mathcal{T}}$.

In particular, the following holds: For any backward linear transfer $\mathcal{S}$ with $\{\delta_x\,;\, x \in X\} \subset D_1(\mathcal{S})$,

1. *If $\mathcal{T} \leq \mathcal{S}$, then $\overline{\mathcal{T}} \leq \mathcal{S}$.*
2. *If $\mathcal{S} \leq \tilde{\mathcal{T}}$, then $\mathcal{S} \leq \overline{\mathcal{T}}$.*

Proof. Note that by the first part of Theorem 2.7.2, $\overline{\mathcal{T}}$ is a backward linear transfer with backward Kantorovich operator

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

Moreover, as $\overline{\mathcal{T}}$ is a weak transport, an admissible $\pi \in \mathcal{K}(\mu,\nu)$ is the product measure $\mu \otimes \nu$. Hence

$$\overline{\mathcal{T}}(\mu,\nu) \leq \int_X c(x,\nu)d\mu(x) = \tilde{\mathcal{T}}(\mu,\nu).$$

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

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

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

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

hence

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

To see that $\overline{\mathcal{T}}$ is the smallest backward linear transfer greater than $\mathcal{T}$, note that if $\mathcal{S}$ is a backward linear transfer and $\mathcal{T} \leq \mathcal{S}$, then $\overline{\mathcal{T}} \leq \overline{\mathcal{S}}$, and $\overline{\mathcal{S}} = \mathcal{S}$ by the previous Theorem 2.7.2.

To see $\overline{\mathcal{T}}$ is the greatest backward linear transfer smaller than $\tilde{\mathcal{T}}$, suppose $\mathcal{S}$ is a backward linear transfer with $\mathcal{S} \leq \tilde{\mathcal{T}}$ and S^- is its backward Kantorovich operator. Note that $\overline{\mathcal{T}}(\delta_x, \nu) = \tilde{\mathcal{T}}(\delta_x, \nu)$. Therefore

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

and therefore $\mathcal{S} \leq \overline{\mathcal{T}}$.

$\square$

Remark 2.7.5. Suppose $\mathcal{T}$ is any convex, bounded below, and weak* lower semi-continuous functional on $\mathcal{P}(X) \times \mathcal{P}(Y)$ that is finite on the set of Dirac measures $\{(\delta_x, \delta_y)\,;\, x \in X, y \in Y\}$. One can then define a cost function $c(x, y) = \mathcal{T}(\delta_x, \delta_y)$, and the associated optimal transport $\mathcal{T}_c(\mu, \nu)$. To compare $\mathcal{T}$ with $\mathcal{T}_c$, note that

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

and so

$$\mathcal{T}(\mu,\nu) \le \overline{\mathcal{T}}(\mu,\nu) \le \mathcal{T}_c(\mu,\nu).$$

In many cases, it is not possible to define a proper cost $c(x,y) = \mathcal{T}(\delta_x,\delta_y)$, i.e. c is identically $+\infty$. This is the case for many stochastic transport problems where transport via Brownian motion makes it impossible for a Dirac measure to be transported to another Dirac measure; see in particular the stochastic mass transport of Example 2.8.7. □

In the next proposition, we describe dually a "Kantorovich operator envelope" for operators $T: C(Y) \to USC(X)$. It will be necessary first (and later in this thesis) to state a classical minimax theorem that will allow us to interchange sup and inf in appropriate cases.

Theorem 2.7.6 (Sion's minimax theorem [52], Theorem 3.4)**.** *Let V be a compact convex subset of a topological vector space, and W a convex set in a (possibly different) topological vector space. Let $f: V \times W \to \mathbb{R}$ be such that*

1. *For all $w \in W$, $v \mapsto f(v,w)$ is: (i) lower semi-continuous on V and (ii) quasiconvex on V, i.e. $\{v \in V\,;\, f(v,w) \le \lambda\}$ is convex or empty for all $\lambda \in \mathbb{R}$,*

2. *For all $v \in V$, $w \mapsto f(v,w)$ is: (i) upper semi-continuous on W and (ii) quasiconcave on W, i.e. $\{w \in W\,;\, f(v,w) \ge \lambda\}$ is convex or empty for all $\lambda \in \mathbb{R}$,*

Then $\inf_{v\in V}\sup_{w\in W} f(v,w) = \sup_{w\in W}\inf_{v\in V} f(v,w)$.

Proposition 2.7.7. *Let $T: C(Y) \to USC_b(X)$ be any map such that for every $x \in X$,*

$$\sup_{x\in X}\inf_{\nu\in\mathcal{P}(Y)}\sup_{g\in C(Y)}\Big\{\int_Y g\,d\nu - Tg(x)\Big\} < +\infty. \tag{2.18}$$

Then the operator defined by

$$\overline{T}^{-}g(x) := \sup_{\sigma\in\mathcal{P}(Y)}\inf_{h\in C(Y)}\{\int_Y (g-h)\,d\sigma + Th(x)\}$$

is the backward Kantorovich operator of the weak transport $\bar{\mathcal{T}}$ associated to $\mathcal{T}$ of Proposition 2.7.4 and maps $C(Y)$ to $USC_b(X)$. It satisfies $\overline{T}^- \leq T$, with the property that for any other backward Kantorovich operator S^- : $C(Y) \to USC_b(X)$ such that $S^- \leq T$, then $S^- \leq \overline{T}^-$. Consequently if T is itself a backward Kantorovich operator, then $\overline{T}^- = T$.

Proof. Consider the backward linear coupling on $\mathcal{M}(X) \times \mathcal{M}(Y)$ given by

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

Note that $\mathcal{T}$ is a bounded below, convex, lower semi-continuous functional and the assumption (2.18) means that $\{\delta_x\, ;\, x \in X\} \subset D_1(\mathcal{T})$. Hence Proposition 2.7.4 applies to yield a weak transport $\overline{\mathcal{T}}$ with a corresponding backward Kantorovich operator defined as

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

We can then provide an upper estimate for $\overline{T}^- g(x)$ by substituting the expression for $\mathcal{T}(\delta_x, \sigma)$:

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

If S is a backward Kantorovich operator such that $S \leq T$, then

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

where the last two steps used the fact that S is a Kantorovich operator. $\square$

2.7.3 Recessions of linear transfers

We recall in classical convex analysis, a *direction of recession* for a convex set $C \subset X$ is a direction d such that $x + \lambda d \in C$ for all $x \in C$ and all $\lambda \geq 0$. The recession cone is the collection of all such d. The *recession cone for a convex function* f is the intersection of all recession cones for its sublevel sets. One can introduce the *recession function of* f as

$$r_f(d) := \lim_{\lambda \to +\infty} \frac{f(x + \lambda d) - f(x)}{\lambda}$$

which characterises the recession cone for f as the collection d where $r_f(d) \leq 0$.

In an analogy, we have the following recession operator for linear transfers, which was introduced in [1].

Proposition 2.7.8. *Let $\mathcal{T}$ be a backward linear transfer such that $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$, with backward Kantorovich operator T^-. Then, the functional*

$$\mathcal{T}_{\mathit{finite}}(\mu, \nu) = \begin{cases} 0 & \textit{if } \mathcal{T}(\mu, \nu) < +\infty \\ +\infty & \text{otherwise,} \end{cases}$$

is a backward linear transfer with Kantorovich operator

$$T_r^- g(x) := \lim_{\lambda\to+\infty} \frac{T^-(\lambda g)(x)}{\lambda}$$
$$= \sup\{\int_Y g d\nu\,;\, \nu \in \mathcal{P}(Y),\, \mathcal{T}(\delta_x, \nu) < +\infty\}. \tag{2.19}$$

Proof. Without loss of generality, we may assume $\mathcal{T} \geq 0$ (otherwise consider $\mathcal{T} - C$ where $C \in \mathbb{R}$ is a lower bound for $\mathcal{T}$). It is immediate that $\mathcal{T}_{\text{finite}}$ is a proper, bounded below, convex, and weak* lower semi-continuous function with $D_1(\mathcal{T}_{\text{finite}}) = D_1(\mathcal{T})$. We have for $\mu \in D_1(\mathcal{T}_{\text{finite}})$,

$$(\mathcal{T}_{\text{finite}})^*_\mu(g) = \sup\{\int_Y g d\nu\,;\, \nu \in \mathcal{P}(Y),\, \mathcal{T}(\mu, \nu) < +\infty\}. \tag{2.20}$$

At the same time, we know by Proposition 2.4.4 that T^- is given by

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

so for $\lambda > 0$, we have

$$\frac{T^-(\lambda g)(x)}{\lambda} = \sup_{\sigma\in\mathcal{P}(Y)} \{\int_Y g\, d\sigma - \frac{1}{\lambda}\mathcal{T}(\delta_x, \sigma)\}.$$

We then have $\liminf_{\lambda\to+\infty} \frac{T^-(\lambda g)(x)}{\lambda} \geq \sup\{\int_Y g d\sigma\,;\, \mathcal{T}(\delta_x, \sigma) < +\infty\}$. On the other hand, since $\mathcal{T}$ is non-negative, we have

$$\frac{T^-(\lambda g)(x)}{\lambda} \leq \sup\{\int_Y g\, d\sigma\,;\, \sigma \in \mathcal{P}(Y),\, \mathcal{T}(\delta_x, \sigma) < +\infty\}$$

and we conclude that

$$\lim_{\lambda\to+\infty} \frac{T^-(\lambda g)(x)}{\lambda} = \sup\{\int_Y g\, d\sigma\,;\, \sigma \in \mathcal{P}(Y),\, \mathcal{T}(\delta_x, \sigma) < +\infty\}. \tag{2.21}$$

Exactly the same argument also shows that for any $\mu \in D_1(\mathcal{T})$,

$$\lim_{\lambda\to+\infty} \int_X \frac{T^-(\lambda g)}{\lambda} d\mu = \sup\{\int_Y g d\nu\,;\, \nu \in \mathcal{P}(Y),\, \mathcal{T}(\mu, \nu) < +\infty\}. \tag{2.22}$$

To conclude that $\mathcal{T}_{\text{finite}}$ is a backward linear transfer, it remains then to justify $\lim_{\lambda\to+\infty}\int_X \frac{T^-(\lambda g)}{\lambda}d\mu = \int_X \lim_{\lambda\to+\infty}\frac{T^-(\lambda g)}{\lambda}d\mu$. This is simply by monotone convergence. Indeed, as $\mathcal{T}$ is non-negative, then $\frac{T^-(\lambda g)(x)}{\lambda}$ is monotone increasing in λ; if $\lambda_2 \geq \lambda_1$,

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

We therefore conclude by (2.20), (2.21), and (2.22), that

$$(\mathcal{T}_{\text{finite}})^*_\mu(g) = \int_X T_r^- g\, d\mu$$

where $T_r^- g(x)$ is given by (2.19).

$\square$

Remark 2.7.9. Note that the above implies

$$\mathcal{T}(\mu,\nu) < +\infty \quad \text{if and only if} \quad \textstyle\int_X T_r^- g d\mu \geq \int_Y g d\nu \text{ for every } g \in C(Y).$$

The latter condition can be seen as a *generalized order condition* between μ and ν. For example, if $\mathcal{T}$ is the balayage transfer of Section 2.5.3 on $\mathcal{P}(X)\times\mathcal{P}(X)$ with $\mathcal{A} = \mathcal{C}$ the cone of convex functions, then $T_r^- g = \hat{g}$, the concave envelope of g. In this case, if $\mathcal{T}(\mu,\nu) = 0$, then $\mu \preceq_{\mathcal{C}} \nu$, i.e. $\int_X \varphi d\mu \leq \int_X \varphi d\nu$ for all convex functions φ. Hence for all $g \in C(X)$,

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

On the other hand, if $\int_X \hat{g} d\mu = \int_X T_r^- g d\mu \geq \int_X g d\nu$ for every $g \in C(X)$, the inequality holds for all *concave* η, and in this case $\hat{\eta} = \eta$. This implies $\mu \preceq_{\mathcal{C}} \nu$ and consequently $\mathcal{T}(\mu,\nu) = 0$.

2.8 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 of Section 2.6.

2.8.1 Linear transfers associated to weak mass transports

Example 2.8.1 (Marton transports [41, 42] are backward linear transfers)**.** Consider the weak cost

$$c(x,\sigma) := \gamma\left(\int_Y d(x,y)d\sigma(y)\right)$$

where γ is a convex function on $\mathbb{R}^+$ and $d : X \times Y \to \mathbb{R}$ is a lower semi-continuous function. The associated weak transport is a backward linear transfer with Kantorovich operator

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

Proposition 2.8.2. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(Y)$, where Y is convex and compact. Suppose for some lower semi-continuous functional $c : X \times Y \to \mathbb{R}$, we have*

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

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

$$T^- g(x) = \sup_{y \in Y} \{\hat{g}(y) - c(x,y)\},$$

where $\hat{g}$ is the concave envelope of g (see Example 2.5.14).

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

$$\begin{aligned}
T^- g(x) = & \sup_{\sigma \in \mathcal{P}(Y)} \{\int_Y g\, d\sigma - c(x, \int_Y y\, d\sigma(y))\} \\
&= \sup_{z \in Y} \left\{ \sup\{\int g\, d\sigma; \sigma \in \mathcal{P}(Y), \delta_z \prec_{\mathcal{C}} \sigma\} - c(x,z) \right\} \\
&= \sup_{z \in Y} \{\hat{g}(z) - c(x,z)\}
\end{aligned}$$

where the last equality we have used Proposition 2.5.13. $\square$

Example 2.8.3 (A barycentric cost function (Gozlan et al. [36]))**.** Consider the weak transport

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

where $\|\cdot\|$ is the norm on X induced from d_X. This is a backward linear transfer, with Kantorovich operator given by

$$T^- gf(x) = \sup\{\hat{g}(y) - \|y - x\|; y \in Y\},$$

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

Example 2.8.4 (The variance functional)**.** Consider the variance functional

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

Denoting $q(y) = |y|^2$ as the quadratic function, 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 (backward) linear transfers.

Example 2.8.5 (Martingale transport)**.** Let $c : X \times X \to \mathbb{R}$ be a continuous cost function and $\mathcal{C}$ the cone of convex functions. Martingale transport (see e.g. [29], [37]) is defined by

$$\mathcal{T}(\mu,\nu) = \begin{cases} \inf\{\int_{X\times X} c(x,y) d\pi(x,y)\,;\, \pi \in \mathcal{M}(\mu,\nu)\} & \text{if } \mu \prec_{\mathcal{C}} \nu \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mathcal{M}(\mu,\nu)$ consists of all transport plans $\pi \in \mathcal{K}(\mu,\nu)$ such that their disintegration π_x with respect to μ, $d\pi(x,y) = d\mu(x) d\pi_x(y)$, has barycentre x.

Note that π_x has barycentre x, if and only if, $\delta_x \prec_{\mathcal{C}} \pi_x$. Therefore, defining the weak cost

$$\tilde{c}(x,\sigma) = \begin{cases} \int_X c(x,y)\, d\sigma(y) & \text{if } \delta_x \prec_{\mathcal{C}} \sigma, \\ +\infty & \text{if not,} \end{cases}$$

martingale transport can be written as

$$\mathcal{T}(\mu,\nu) = \begin{cases} \inf_{\pi \in \mathcal{K}(\mu,\nu)}\{\int_{X\times X} \tilde{c}(x,\pi_x)\, d\ \mu\} & \text{if } \mu \prec_{\mathcal{C}} \nu, \\ +\infty & \text{otherwise.} \end{cases}$$

Since it is a weak transport, we know from Section 2.7 that it is a backward linear transfer with

$$
\begin{aligned}
T^- g(x) &= \sup_{\sigma \in \mathcal{P}(X)} \{ \int_X g d\sigma - \tilde{c}(x,\sigma) \} \\
&= \sup \{ \int_Y (g(y) - c(x,y)) d\sigma(y) \, ; \, \sigma \in \mathcal{P}(Y), \, \delta_x \preceq_{\mathcal{C}} \sigma \} \\
&= \hat{g}_{c,x}(x),
\end{aligned}
$$

where $\hat{g}_{c,x}$ is the concave envelope of the function $y \mapsto g(y) - c(x,y)$. The last equality follows from Proposition 2.5.13.

Example 2.8.6 (Schrödinger bridge (Gentil-Leonard-Ripani [27]))**.** This example is in $\mathbb{R}^d$, but we include it even though we have not formally defined linear transfers on non-compact spaces. Fix some reference non-negative measure R on path space $\Omega = C([0,1], \mathbb{R}^d)$. Let $X = (X_t)_t$ be a random process on M whose law is R, i.e. $R = (\Phi_X)_{\#}(\mathbb{P})$, where $(\Phi_X(\omega))(t) := X_t(\omega)$. Denote by R_{01} the joint law of the initial position X_0 and the final position X_1. For example (see [27]), 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

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

where $dR_{01}(x,y) = dm(x) dr_1^x(y)$ is the disintegration of R_{01} with respect to its initial measure m. We see that $\mathcal{T}_{R_{01}}$ is a weak transport corresponding to the weak cost $c(x,p) = \mathcal{H}(r_1^x, p)$ and hence, by an appropriate extension of Theorem 2.7.2 to non-compact spaces, it is a backward linear transfer. Its backward 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 (2.8.6) 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

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

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.

2.8.2 One-sided transfers associated to stochastic mass transport

Let $M = \mathbb{T}^n$ be the flat torus, and consider a Lagrangian on phase space $L : TM \to [0,\infty)$. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ be a complete filtered probability space and define $\mathcal{A}_{[0,1]}$ to be the set of continuous semi-martingales $X : \Omega\times[0,1] \to M$ such that there exists a Borel measurable drift $\beta_X : [0,1]\times C([0,1]) \to \mathbb{R}^d$ for which

1. $\omega \mapsto \beta_X(t,\omega)$ is $\mathcal{B}(C([0,t]))_+$-measurable for all $t \in [0,1]$, where $\mathcal{B}(C([0,t]))$ is the Borel σ-algebra of $C[0,t]$.
2. $W_X(t) := X(t) - X(0) - \int_0^t \beta_X(s,X)ds$ is a $\sigma(X(t)\,;\, 0 \le t \le 1)$ M-valued Brownian motion.

Example 2.8.7 (Stochastic mass transport between two probability measures)**.** Consider the functional $\mathcal{T} : \mathcal{P}(M) \times \mathcal{P}(M) \to \mathbb{R}\cup\{+\infty\}$ defined by

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

where we use the notation $Z \sim \sigma$ to denote that the random variable Z has distribution σ.

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 enjoys a dual formulation (first proven by Mikami-Thieullin [45] for the space $\mathbb{R}^n$) that permits it to be realised as a backward linear transfer. An adaptation of the proofs of Mikami-Thieullin to $M = \mathbb{T}^n$ (see [45] Lemma 3.1, 3.2, and 3.3, and Theorem 2.2), yield the following.

Proposition 2.8.8. *Under suitable conditions on L (for example, if $L(x,\beta) = \frac{1}{2}|\beta|^2$; see in [45] the assumptions (A1), (A2), (A3)(i - v), and (A4)), $\nu \mapsto \mathcal{T}(\mu,\nu)$ is convex and weak* lower semi-continuous, and*

$$\mathcal{T}(\mu,\nu) = \sup_{g \in C(M)} \{ \int_M g(x) d\nu(x) - \int_M u_g(0,x) d\mu(x) \}$$

where $u_g : [0,1] \times M \to \mathbb{R}$ is a viscosity solution of

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

with $u(1,x) = g(x)$. In addition $T^- g(x) := u_g(0,x)$ can be written as

$$T^- g(x) = \sup_{X \in \mathcal{A}_{[0,1]}} \left\{ \mathbb{E}\left[g(X(1)) | X(0) = x \right] - \mathbb{E}\left[\int_0^1 L(X(s), \beta_X(s,X)) ds | X(0) = x \right] \right\}.$$

From this expression, we can readily check that T^- is a backward Kantorovich operator (and consequently, $\mathcal{T}$ is a backward linear transfer).

2.8.3 Transfers associated to optimally stopped stochastic transports

In dimensions greater than one, there are many different types of martingales. If one chooses those that essentially follow a Brownian path, then we have the following examples of linear transfers.

Example 2.8.9 (Optimal subharmonic martingale transfers (Ghoussoub-Kim–Palmer [31]))**.** Let O be a convex bounded domain in $\mathbb{R}^d$. 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,

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

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 $d\pi(x,y) = d\pi_x(y)d\mu(x)$. Otherwise, set $\mathcal{P}_c(\mu,\nu) = +\infty$.

By a theorem of Skorokhod [53], 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 [30]. The above problem associated to a cost c can then be formulated as

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

where $(B_t)_t$ is Brownian motion starting with distribution μ and ending at a stopping time τ such that B_τ realises the distribution ν.

In [31] it is shown that $\mathcal{P}_c$ is a backward linear transfer with a backward Kantorovich operator given by $T^- g(x) = J_g(x,x)$, where

$$J_g(x,y) = \sup_{\tau \le \tau_O} \mathbb{E}\big[g(B_\tau^y) - c(x, B_\tau^y)\big],$$

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

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

as well as the unique minimiser of the variational problem

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

Example 2.8.10 (Optimally stopped stochastic transport [28, 31])**.** Given a Lagrangian $L : [0,1]\times\mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$, consider the optimal stopping problem

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

where S is the set of possibly randomized stopping times, and $\mathcal{A}$ is the class of processes defined in Section 2.8.2. In this case, $\mathcal{T}_L$ is a backward linear transfer with Kantorovich operator given by $T_L^- f = \hat{V}_f(0,\cdot)$, where

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

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

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

2.9 Convex transfers

We saw in Example 2.5.3 that the relative entropy functional $\mathcal{H}(\mu,\nu)$ on $\mathcal{P}(X)\times\mathcal{P}(X)$, which we recall is defined via

$$\mathcal{H}(\mu,\nu) := \begin{cases} \int_X \frac{d\nu}{d\mu}\log(\frac{d\nu}{d\mu})d\mu & \text{if } \nu \ll \mu,\\ +\infty & \text{otherwise,}\end{cases}$$

satisfies $\mathcal{H}^*_\mu(g) = \log\int_X e^g d\mu$ and thus is not a linear transfer. We saw in Proposition 2.5.4 that we can also express it as

$$\mathcal{H}^*_\mu(g) = \inf_{t\in\mathbb{R}}\int_X T_t(g)d\mu, \quad T_t g := e^{g+t-1} - t. \tag{2.23}$$

This means that

$$
\begin{aligned}
\mathcal{H}(\mu,\nu) &= \sup_{g\in C(X)} \sup_{t\in\mathbb{R}} \{\int_X g d\nu - \int_X T_t g d\mu\} \\
&= \sup_{t\in\mathbb{R}} \sup_{g\in C(X)} \{\int_X g d\nu - \int_X T_t g d\mu\} \\
&= \sup_{t\in\mathbb{R}} \mathcal{T}_t(\mu,\nu) \qquad (2.24)
\end{aligned}
$$

where $\mathcal{T}_t(\mu,\nu) := \sup_{g\in C(X)}\{\int_X g d\nu - \int_X T_t g d\mu\}$ is a backward linear coupling, but *not* a backward linear transfer. The expression (2.24) motivates the following definition.

Definition 2.9.1 (Convex coupling)**.** Let $\mathcal{T} : \mathcal{P}(X)\times\mathcal{P}(Y) \to \mathbb{R}\cup\{+\infty\}$ be a proper, bounded below, weak* lower semi-continuous, and jointly convex functional. We say $\mathcal{T}$ is a **backward convex coupling** provided it is the supremum of a family of backward linear couplings,

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

where I is some index set and for each i, $\mathcal{T}_i(\mu,\nu) := \sup_{g\in C(Y)}\{\int_Y g d\nu - \int_X T_i^- g d\mu\}$ is a backward linear coupling.

In the same way as for forward linear couplings, we shall define $\mathcal{T}$ as a **forward convex coupling** if $\tilde{\mathcal{T}}(\nu,\mu) := \mathcal{T}(\mu,\nu)$ is a backward convex coupling on $\mathcal{P}(Y)\times\mathcal{P}(X)$.

We shall refer to $\{T_i^-\}_{i\in I}$ as the family of operators associated to $\mathcal{T}$. Note they are not in general Kantorovich operators as per Definition 2.4.1.

Recalling from Remark (2.4.7) that a backward linear coupling $\mathcal{T}$ has $\mathcal{T}_\mu^*(g) \leq \int_X T^- g d\mu$, the backward linear transfers were defined as those linear couplings where we have equality. In an analogous way, we see that in general, a convex coupling $\mathcal{T}$ satisfies $\mathcal{T}_\mu^*(g) \leq \inf_{i\in I}\int_X T_i^- g d\mu$. This motivates the next definition.

Definition 2.9.2 (Convex transfer)**.** Let $\mathcal{T}$ be a backward convex coupling. We say $\mathcal{T}$ is a **backward convex transfer** if for all $\mu \in D_1(\mathcal{T})$, for all

$g \in C(Y)$, it holds that

$$\mathcal{T}_\mu^*(g) = \inf_{i\in I} \int_X T_i^- g d\mu.$$

Note that when I is a singleton, then $\mathcal{T}$ is a backward linear transfer, so all linear transfers are convex transfers. A first example of a backward convex transfer is of course the relative entropy via (2.23).

Also note that if $\{\delta_x\,;\, x \in X\} \subset D_1(\mathcal{T})$, then $\mathcal{T}_{\delta_x}^*(g) = \inf_{i\in I} T_i^- g(x)$, which implies that $g \mapsto \inf_{i\in I} T_i^- g$ is a backward Kantorovich operator. More generally, for backward convex couplings, there is the inequality

$$\mathcal{T}_\mu^*(g) \le \int_X \inf_{i\in I} T_i^- g d\mu,$$

thus $\mathcal{T}(\mu,\nu)$ is bounded above by the backward linear coupling

$$\sup_{g\in C(Y)} \{\int_Y g d\nu - \int_X \inf_{i\in I} T_i^- g d\mu\}. \tag{2.25}$$

If it happens that $g \mapsto \inf_{i\in I} T_i^- g$ is a backward Kantorovich operator (which in general it is not, but guaranteed if $D_1(\mathcal{T})$ contains the Dirac masses), then this linear coupling (2.25) should in fact be the linear transfer envelope $\overline{\mathcal{T}}$ of $\mathcal{T}$ provided by Proposition 2.7.4. The next proposition provides the details for this.

Proposition 2.9.3. *Let $\mathcal{T}$ be a convex coupling on $\mathcal{P}(X) \times \mathcal{P}(Y)$ of the form*

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

where for each $i \in I$,

$$\mathcal{T}_i(\mu,\nu) = \sup_{g\in C(Y)} \{\int_Y f d\nu - \int_X T_i^- g d\mu\}$$

for some map $T_i^- : C(Y) \to USC_b(X)$. Assume $\{\delta_x; x \in X\} \subset D_1(\mathcal{T})$ and $\sup_{x\in X} \inf_{\nu\in\mathcal{P}(Y)} \mathcal{T}(\delta_x,\nu) < +\infty$. Consider the weak transport (or "transfer envelope") $\overline{\mathcal{T}}$ of $\mathcal{T}$ provided by Proposition 2.7.4 and the corresponding

Kantorovich operator $\overline{T}^-$ *. Then,*

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

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

   and therefore satisfies $\overline{T}^- g \leq \inf_i T_i^- g$ *on* $C(Y)$*.*

2. *If in addition each* T_i^- *is a backward Kantorovich operator, then* $\overline{T}^- g = \inf_{i \in I} T_i^- g$ *if and only if* $g \mapsto \inf_{i \in I} T_i^- g(x)$ *is convex and lower semi-continuous.*

Proof. 1. We show that the hypotheses of Proposition 2.7.7 are satisfied. We have that the map $g \mapsto Tg := \inf_{i \in I} T_i^- g$ takes $C(Y)$ to $USC_b(X)$, and,

$$\sup_{x \in X} \inf_{\nu \in \mathcal{P}(Y)} \sup_{g \in C(Y)} \{ \int_Y g d\nu - Tg(x) \} = \sup_{x \in X} \inf_{\nu \in \mathcal{P}(Y)} \mathcal{T}(\delta_x, \nu) < +\infty$$

by assumption. Therefore the map $Tg = \inf_{i \in I} T_i^- g$ satisfies the hypotheses of Proposition 2.7.7, and we conclude that the operator defined by

$$\overline{T}^- g(x) = \sup_{\sigma \in \mathcal{P}(Y)} \inf_{h \in C(Y)} \{ \int_Y (g-h) d\sigma + \inf_{i \in I} T_i^- g(x) \}$$

is the Kantorovich operator of the weak transport $\overline{\mathcal{T}}$ associated to $\mathcal{T}$, and satisfies $\overline{T}^- g \leq \inf_{i \in I} T_i^- g$.

2. Suppose each T_i^- is a backward Kantorovich operator. If $g \to \inf_{i \in I} T_i^- g(x)$ is convex and lower semi-continuous for any $x \in X$, then it is itself a backward Kantorovich operator. Therefore the additional consequence of Proposition 2.7.7 implies $\inf_{i \in I} T_i^- g = \overline{T}^- g$. $\square$

Example 2.9.4 (Linear transfer envelope of relative entropy)**.** Recall from the beginning of this section or Example 2.5.3, that $\mathcal{H}_\mu^*(g) = \inf_{t \in \mathbb{R}} \int_X T_t(g) d\mu$, where $T_t g := e^{g+t-1} - t$. It is easy to compute that

$$\inf_{t \in \mathbb{R}} \{ e^{g(x)+t-1} - t \} = g(x)$$

and consequently $\overline{T}^- g = g$ is the Kantorovich operator for the linear transfer envelope. Therefore $\overline{\mathcal{T}}$ is nothing but the trivial linear backward transfer of Example 2.5.7

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

□

A natural question to investigate is whether convex functions of backward linear transfers, are backward convex transfers. The next proposition says this is indeed the case.

Proposition 2.9.5. *Let $\alpha : [0,\infty) \to \mathbb{R}$ be a convex increasing function, and $\mathcal{T}$ a backward linear transfer with corresponding operator T^-. Then $\alpha(\mathcal{T})$ is a backward convex transfer, and its envelope $\overline{\alpha(\mathcal{T})}$ as given in Proposition 2.7.4 has Kantorovich operator given by*

$$\overline{T}_\alpha^- g = \inf_{s>0} \{sT^-(\frac{g}{s}) + \alpha^\oplus(s)\}, \tag{2.26}$$

where $\alpha^\oplus(t) := \sup\{ts - \alpha(s); s \geq 0\}$ corresponds to the Legendre transform of the extended real-valued function $\tilde{\alpha}(t) := \begin{cases} \alpha(t) & t \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$

Proof. Note that $\tilde{\alpha}$ is convex and lower semi-continuous, hence $\tilde{\alpha}^{**}(t) = \tilde{\alpha}(t)$ and so $\alpha(t) = \sup\{ts - \alpha^\oplus(s); s \geq 0\}$. Therefore,

$$\begin{aligned} \alpha(\mathcal{T}(\mu,\nu)) &= \sup_{s\geq 0}\{s\mathcal{T}(\mu,\nu) - \alpha^\oplus(s)\} \\ &= \sup_{s\geq 0} \sup_{g\in C(Y)} \{\int_Y g d\nu - \int_X (sT^-(\frac{g}{s}) + \alpha^\oplus(s)) d\mu\} \end{aligned}$$

so we see that $\alpha(\mathcal{T})$ is a backward convex coupling.

To show that $\alpha(\mathcal{T})$ is a backward convex transfer, we have that $\nu \mapsto \alpha(\mathcal{T}(\mu,\nu))$ is convex and weak* lower semi-continuous, thus we just need to compute the Legendre transform for $\mu \in D_1(\mathcal{T})$,

$$\alpha(\mathcal{T})^*_\mu(g) = \sup_{\nu \in \mathcal{P}(Y)} \{\int_Y g d\nu - \alpha(\mathcal{T}(\mu,\nu))\},$$

where we recall the notation $\alpha(\mathcal{T})^*_\mu$ is the Legendre transform of $\nu \mapsto \alpha(\mathcal{T}(\mu,\nu))$. Substituting $\alpha(t) = \sup\{ts - \alpha^{\oplus}(s); s \geq 0\}$ for $\alpha(\mathcal{T}(\mu,\nu))$ in the above expression, we obtain

$$\begin{aligned}\alpha(\mathcal{T})^*_\mu(g) &= \sup_{\nu \in \mathcal{P}(Y)} \inf_{s\geq 0}\{\int_Y g d\nu - s\mathcal{T}(\mu,\nu) + \alpha^{\oplus}(s)\} \\ &= \sup_{\nu\colon \mathcal{T}(\mu,\nu)<+\infty} \inf_{s\geq 0}\{\int_Y g d\nu - s\mathcal{T}(\mu,\nu) + \alpha^{\oplus}(s)\}\end{aligned}$$

Notice now that $\nu \mapsto f(\nu,s) := \int_Y g d\nu - s\mathcal{T}(\mu,\nu) + \alpha^{\oplus}(s)$, as a function on the convex and weak* closed subset $V := \{\nu \in \mathcal{P}(Y)\,;\, \mathcal{T}(\mu,\nu) < +\infty\}$, is weak* upper semi-continuous, and also quasi-concave: $\{\nu\,;\, f(\nu,s) \geq \lambda\}$ is convex or empty for $\lambda \in \mathbb{R}$. At the same time, $s \mapsto f(\nu,s)$ on $W := [0,\infty)$ is lower semi-continuous on W and also quasi-convex: $\{s\,;\, f(\nu,s) \leq \lambda\}$ is convex or empty for $\lambda \in \mathbb{R}$. Then by Sion's minimax theorem (Theorem 2.7.6), we may interchange sup and inf to obtain

$$\begin{aligned}\alpha(\mathcal{T})^*_\mu(g) &= \inf_{s\geq 0} \sup_{\nu \in \mathcal{P}(Y)} \{\int_Y g d\nu - s\mathcal{T}(\mu,\nu) + \alpha^{\oplus}(s)\} \\ &= \inf_{s\geq 0} \int_Y [sT^-(\frac{g}{s}) + \alpha^{\oplus}(s)] d\mu.\end{aligned}$$

where we have used the fact that $s\mathcal{T}$ is a backward linear transfer with operator $g \mapsto sT^-(\frac{g}{s})$ (see Section 2.10).

Regarding the envelope $\overline{\alpha(\mathcal{T})}$ we note that for each $s > 0$, $T^-_s g := sT^-(\frac{g}{s}) + \alpha^{\oplus}(s)$ is a backward Kantorovich operator, and, moreover, the function $(s,g) \mapsto sT^-(\frac{g}{s}) + \alpha^{\oplus}(s)$ is jointly convex on $\mathbb{R}^+ \times C(Y)$. Hence, the infimum in s is convex, and therefore $g \mapsto \inf_{s\geq 0}\{sT^-(\frac{g}{s}) + \alpha^{\oplus}(s)\}$ is convex. We conclude by Proposition 2.9.3. $\square$

Example 2.9.6 (Generalised entropy functional)**.** Consider for a strictly convex function $\alpha : [0, \infty) \to \mathbb{R}$, the generalised entropy functional on $\mathcal{P}(X) \times \mathcal{P}(X)$, given by,

$$\mathcal{T}_\alpha(\mu, \nu) = \begin{cases} \int_X \alpha\left(\frac{d\nu}{d\mu}\right) d\mu & \text{if } \nu \ll \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

By a similar argument as given in Example 2.5.3 (see [35], Proposition 2.9), we have

$$(\mathcal{T}_\alpha)^*_\mu(g) = \inf_{t\in\mathbb{R}} \int_X [\alpha^\oplus(t + g(x)) - t] d\mu(x)$$

where $\alpha^\oplus$ is as in Proposition 2.9.5, so that $\mathcal{T}_\alpha$ is a backward convex transfer.

Example 2.9.7 (A backward convex coupling which is not a convex transfer)**.** Let $\Omega \subset \mathbb{R}^d$ be compact with $1 < |\Omega| < \infty$, $\lambda := \frac{1}{|\Omega|}$, and define for any two given probability measures μ, ν on Ω,

$$\mathcal{T}_\lambda(\mu, \nu) = \begin{cases} 0 & \text{if } \lambda\frac{d\nu}{d\mu} \leq 1 \ \mu\text{-a.e.} \\ +\infty & \text{otherwise,} \end{cases}$$

Note that when $\mu = \lambda dx|_\Omega$ (the uniform measure on Ω),

$$\mathcal{T}_\lambda(\lambda dx|_\Omega, \nu) = \begin{cases} 0 & \text{if } \frac{d\nu}{dx} \leq 1 \text{ Lebesgue-a.e.} \\ +\infty & \text{otherwise.} \end{cases}$$

We claim that $\mathcal{T}_\lambda$ is a backward convex coupling but not a convex transfer. Indeed, for the first claim, consider $\alpha_{m,\lambda}(t) := (\lambda t)^m \log(\lambda t)$ for $m \geq 1$ and $t \geq 0$, and define

$$\mathcal{T}_{m,\lambda}(\mu, \nu) := \begin{cases} \int_\Omega \alpha_{m,\lambda}\left(\frac{d\nu}{d\mu}\right) d\mu, & \text{if } \nu << \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

By Proposition 2.9.5, $\mathcal{T}_{m,\lambda}$ is a backward convex transfer and for $g \in C(\Omega)$,

$$(\mathcal{T}_{m,\lambda})^*_\mu(g) = \inf\{\int_\Omega [\alpha^\oplus_{m,\lambda}(g(x) + t) - t] d\mu(x)\,;\, t \in \mathbb{R}\}$$

where we recall $(\mathcal{T}_{m,\lambda})^*_\mu$ is the Legendre transform of the map $\nu \mapsto \mathcal{T}_{m,\lambda}(\mu,\nu)$. The function $\alpha^{\oplus}_{m,\lambda}$ can be explicitly computed as

$$\alpha^{\oplus}_{m,\lambda}(t) = \begin{cases} e^{-1+\frac{1}{m-1}W(\beta_m t)}\left[\beta_m t + \frac{1}{m}e^{W(\beta_m t)}\right] & \text{if } t \geq -\frac{\lambda}{m-1}e^{-1}, \\ 0 & \text{if } t < -\frac{\lambda}{m-1}e^{-1}. \end{cases}$$

where $\beta_m := \frac{m-1}{\lambda m}e^{\frac{m-1}{m}}$, and W is the *Lambert-W* function (the multi-valued inverse of $w \mapsto we^w$). It is easy to see that $\mathcal{T}_\lambda(\mu,\nu) = \sup_m \mathcal{T}_{m,\lambda}(\mu,\nu)$; hence it is a backward convex coupling (as a supremum of backward convex transfers).

However, $\mathcal{T}_\lambda$ is not a backward convex transfer itself, since

$$(\mathcal{T}_\lambda)^*_\mu(g) = (\sup_m \mathcal{T}_{m,\lambda})^*_\mu(g) \leq \inf_m (\mathcal{T}_{m,\lambda})^*_\mu(g) = \int_\Omega \frac{g}{\lambda}d\mu,$$

with the inequality being in general strict.

One well-studied problem for which this is relevant is the *quadratic Wasserstein projection*. Denoting $W_2^2(\mu,\nu)$ as the optimal transport with cost function $c(x,y) = |x-y|^2$, the quadratic Wasserstein projection of a measure ν onto the set of Lebesgue densities bounded above by 1, is the minimiser of the following variational problem,

$$\inf\{W_2^2(\sigma,\nu)\,;\, \sigma \in \mathcal{P}(\Omega),\, \frac{d\sigma}{dx} \leq 1\}.$$

This problem has received a lot of attention in the context of congested crowd motion (see, e.g. [49]). By writing

$$\inf\{W_2^2(\sigma,\nu); \frac{d\sigma}{dx} \leq 1\} = \inf\{\mathcal{T}_\lambda(\lambda dx|_\Omega, \sigma) + W_2^2(\sigma,\nu)\,;\, \sigma \in \mathcal{P}(\Omega)\}$$

we see that it consists of an inf-convolution of the backward convex coupling $\mathcal{T}_\lambda$ with the backward linear transfer W_2^2. Therefore we cannot take advantage of the inf-convolution property for backward convex *transfers* and backward linear transfers (see Section 2.10.2) to write a dual formulation.

2.9.1 Backward entropic transfers

We saw previously that the relative entropy functional was a backward convex transfer. However, there is further structure: not only $\mathcal{H}_\mu^*(g) = \inf_{t\in\mathbb{R}} \int_Y T_t^- g d\mu$ (see (2.23)), but in fact $\mathcal{H}_\mu^*(g) = \log\left(\int_Y e^g d\mu\right)$. We thus isolate a special subclass of backward convex transfers.

Definition 2.9.8. Let $\beta : \mathbb{R} \to \mathbb{R}$ be concave and increasing, and let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ be a bounded below, proper, weak* lower semi-continuous and convex functional. We say that $\mathcal{T}$ is a **backward β-entropic transfer** if there exists a map $T^- : C(Y) \to USC(X)$ such that for every $\mu \in \mathcal{D}_1(\mathcal{T})$,

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

We note from the definition that a backward β-entropic transfer is a backward convex transfer. This follows from the concavity of β, so that $\beta(t) = \inf_{s\in\mathbb{R}}\{-st + (-\beta)^*(s)\}$ where $(-\beta)^*$ is the Legendre transform of $-\beta$.

2.10 Operations on convex and linear transfers

In this section, we highlight a few basic operations on convex transfers. In particular, the notion of *inf-convolution* (see below in Section 2.10.2) will be, in subsequent chapters, the operation which is of most interest.

2.10.1 Linear transfers are a convex cone

In this section, we highlight some basic operations for convex and linear transfers, the most important for our purposes in Chapter 5, is inf-convolution.

Proposition 2.10.1. *1.* **(Scalar multiplication for convex transfers)** *If $a \in \mathbb{R}^+ \setminus \{0\}$ and $\mathcal{T}$ is a backward convex transfer, then $(a\mathcal{T})$ given by*

$$(a\mathcal{T})(\mu,\nu) := a\mathcal{T}(\mu,\nu)$$

is also a backward convex transfer with

$$(a\mathcal{T})^*_\mu(g) = \inf_{i\in I}\int_X aT_i^-(\frac{g}{a})d\mu. \tag{2.27}$$

2. **(Addition for linear transfers)** *If $\mathcal{T}_1$ and $\mathcal{T}_2$ are backward linear transfers on $X \times Y$ such that $\{\delta_x\,;\, x \in X\} \subset D(\mathcal{T}_1) \cap D(\mathcal{T}_2)$ with Kantorovich operator T_1^-, T_2^- respectively, then the sum defined as*

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

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

$$\begin{aligned} T^- g(x) &= \sup_{\sigma\in\mathcal{P}(Y)}\{\int_Y g\,d\sigma - \mathcal{T}_1(\delta_x,\sigma) - \mathcal{T}_2(\delta_x,\sigma)\} \\ &= \inf_{h\in C(Y)}\{T_1^- h(x) + T_2^-(g-h)(x)\}. \end{aligned}$$

2.10.2 Inf-convolution

An important operation from the perspective of ergodic theory, is the notion of inf-convolution. We will see shortly that this will allow us to iterate a linear transfer which we shall investigate and go into greater detail in Chapter 5.

Definition 2.10.2 (Inf-convolution)**.** Let X_1, X_2, X_3 be 3 compact, complete, and separable metric 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 **inf-convolution** of $\mathcal{T}_1$ and $\mathcal{T}_2$ is the functional on $\mathcal{P}(X_1)\times\mathcal{P}(X_3)$ given by

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

By induction, one can then consider inf-convolution of n functionals in the obvious way. Note that inf-convolution preserves convexity and weak* lower semi-continuity. This leads to the following important stability property.

Proposition 2.10.3. *If $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(Y)\times\mathcal{P}(Z)$ with Kantorovich operator $T^-: C(Z)\to C(Y)$, and $\mathcal{F}$ is a backward convex transfer on $\mathcal{P}(X)\times\mathcal{P}(Y)$, then $\mathcal{F}\star\mathcal{T}$ is a backward convex transfer on $\mathcal{P}(X)\times\mathcal{P}(Z)$, with*

$$(\mathcal{F}\star\mathcal{T})^*_\mu(h)=\inf_{i\in I}\int_X F_i^-\circ T^- h\,d\mu.$$

Proof. We calculate the Legendre dual of $(\mathcal{F}\star T)_\mu$ at $g\in C(Z)$ and obtain,

$$\begin{aligned}
(\mathcal{F}\star T)^*_\mu(g) &= \sup_{\nu\in\mathcal{P}(Z)}\sup_{\sigma\in\mathcal{P}(Y)}\left\{\int_Z g\,d\nu-\mathcal{F}(\mu,\sigma)-\mathcal{T}(\sigma,\nu)\right\}\\
&= \sup_{\sigma\in\mathcal{P}(Y)}\{\mathcal{T}^*_\sigma(g)-\mathcal{F}(\mu,\sigma)\}\\
&= \sup_{\sigma\in\mathcal{P}(Y)}\left\{\int_Y T^-g\,d\sigma-\mathcal{F}(\mu,\sigma)\right\}\\
&= (\mathcal{F})^*_\mu(T^-(g))\\
&= \inf_{i\in I}\int_X F_i^-\circ T^-g(x))\,d\mu(x).
\end{aligned}$$

□

Corollary 2.10.4. *If $\mathcal{T}_1$ (resp., $\mathcal{T}_2$) is a backward linear transfer on $\mathcal{P}(X_1)\times\mathcal{P}(X_2)$ (resp., on $\mathcal{P}(X_2)\times\mathcal{P}(X_3)$) with Kantorovich operator $T_1^-: C(X_2)\to USC(X_1)$ (resp., $T_2^-: C(X_3)\to C(X_2)$), then $\mathcal{T}_1\star\mathcal{T}_2$ is also a backward linear transfer on $\mathcal{P}(X_1)\times\mathcal{P}(X_3)$ with Kantorovich operator equal to $T_1^-\circ T_2^-$.*

Remark 2.10.5. The assumption that T_2^- maps continuous functions to continuous functions is too restrictive; the appropriate setting is for upper semi-continuous functions, but strictly speaking in that case $T_1^-\circ T_2^- g$ is not defined. This is a non-issue, but requires us to spend some words in

Chapter 5 on extension of the operators to upper semi-continuous functions, in which case the above proposition still holds in the more general case when $T_2^- : C(X_3) \to USC(X_2)$

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

Corollary 2.10.6. *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 and ν on X_n, define*

$$\begin{aligned}\mathcal{T}(\mu, \nu) &:= \mathcal{T}_1 \star \mathcal{T}_2 \star \ldots \star \mathcal{T}_n(\mu, \nu)\\ &= \inf\{\mathcal{T}_1(\mu, \nu_1) + \mathcal{T}_2(\nu_1, \nu_2)... + \mathcal{T}_n(\nu_{n-1}, \nu);\ \nu_i \in \mathcal{P}(X_i), i = 1, ..., n-1\}.\end{aligned}$$

If each $\mathcal{T}_i$ is a backward (resp., forward) linear 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 backward (resp., forward) linear transfer with a Kantorovich operator given by

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

In other words, the following duality formula holds:

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

respectively,

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

We will study in Chapter 5, iterates of a single linear transfer $\mathcal{T} \star \mathcal{T} \star \ldots \star \mathcal{T}$.

2.10.3 Tensor products and dual sums of linear transfers

We highlight other operations for linear transfers.

Definition 2.10.7. 1. (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 $\{\delta_{x_1}\,;\,x_1\in X_1\}\subset D(\mathcal{T}_1)$ and $\{\delta_{x_2}\,;\,x_2\in X_2\}\subset D(\mathcal{T}_2)$, 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:

$$\mathcal{T}_1\otimes\mathcal{T}_2(\mu,\nu):=\inf\left\{\int_{X_1\times X_2}\big(\mathcal{T}_1(\delta_{x_1},\pi^1_{x_1,x_2})+\mathcal{T}_2(\delta_{x_2},\pi^2_{x_1,x_2})\big)\,d\mu(x_1,x_2);\pi\in\mathcal{K}(\mu,\nu)\right\}$$

where $d\pi(x_1,x_2,y_1,y_2)=d\pi_{x_1x_2}(y_1,y_2)d\mu(x_1,x_2)$, and $\pi^i_{x_1x_2}=\mathrm{Proj}_{Y_i\#}\pi_{x_1x_2}$ is the projection of $\pi_{x_1x_2}$ onto Y_i, $i=1,2$.

2. (Dual sum) If $\mathcal{T}_1$ and $\mathcal{T}_2$ are backward linear transfers on $\mathcal{P}(X)\times\mathcal{P}(Y)$ such that $\{\delta_x\,;\,x\in X\}\subset D_1(\mathcal{T}_1)\cap D_1(\mathcal{T}_2)$, with Kantorovich operator T_1^-, T_2^- respectively, then the **dual sum**, $\mathcal{T}_1\Box\mathcal{T}_2$, is defined as the transfer whose Kantorovich operator is $T_1^-+T_2^-$, that is

$$\mathcal{T}_1\Box\mathcal{T}_2(\mu,\nu):=\sup\{\int_Y g d\nu-\int_X(T_1^-g+T_2^-g)d\mu;g\in C(Y)\}$$

The definition of tensorization is via a weak transport, so it is not surprising that we have the following stability property.

Proposition 2.10.8. *If $\mathcal{T}_1$ (resp., $\mathcal{T}_2$) is a backward linear transfer on $\mathcal{P}(X_1)\times\mathcal{P}(Y_1)$ (resp., $\mathcal{P}(X_2)\times\mathcal{P}(Y_2)$) such that $\{\delta_x\,;\,x\in X_1\}\subset D_1(\mathcal{T}_1)$ and $\{\delta_x\,;\,x\in X_2\}\subset D_1(\mathcal{T}_2)$, then $\mathcal{T}_1\otimes\mathcal{T}_2$ is a backward linear transfer on $\mathcal{P}(X_1\times X_2)\times\mathcal{P}(Y_1\times Y_2)$, with Kantorovich operator given by*

$$T^-g(x_1,x_2)=\sup_{\sigma\in\mathcal{K}(\sigma_1,\sigma_2)}\{\int_{Y_1\times Y_2}g d\sigma-\mathcal{T}_1(\delta_{x_1},\sigma_1)-\mathcal{T}_2(\delta_{x_2},\sigma_2)\}. \tag{2.28}$$

Moreover,

$$\mathcal{T}_1\otimes\mathcal{T}_2(\mu,\nu_1\otimes\nu_2)\le\mathcal{T}_1(\mu_1,\nu_1)+\int_{X_2}\mathcal{T}_2(\delta_{x_2},\nu_2)\,d\mu_2(x_2), \tag{2.29}$$

where $\mu_2=Proj_{X_2\#}\mu$.

Proof. The tensor product as defined above is a weak transport, where the cost on $X_1 \times X_2 \times \mathcal{P}(Y_1 \times Y_2)$ is simply,

$$c((x_1, x_2), \sigma) := \mathcal{T}_1(\delta_{x_1}, \sigma_1) + \mathcal{T}_2(\delta_{x_2}, \sigma_2),$$

where $\sigma_i = \mathrm{proj}_{Y_i\#}\sigma$, $i = 1, 2$, are the marginals of σ on Y_1 and Y_2 respectively. The cost c is bounded below, lower semi-contiuous, and $\sigma \mapsto c((x_1, x_2), \sigma)$ is convex. Therefore by Theorem 2.7.2, $\mathcal{T}_1 \otimes \mathcal{T}_2$ is a backward linear transfer with

$$T^- g(x_1, x_2) := \sup_{\sigma \in \mathcal{P}(Y_1 \times Y_2)} \{\int_{Y_1 \times Y_2} g d\sigma - c((x_1, x_2), \sigma)\}.$$

For the upper bound (2.29) for $\mathcal{T}_1 \otimes \mathcal{T}_2$, since $\pi = \mu \otimes (\nu_1 \otimes \nu_2)$ is admissible in the infimum, we have

$$\mathcal{T}_1 \otimes \mathcal{T}_2(\mu, \nu_1 \otimes \nu_2) \leq \mathcal{T}_1(\mu_1, \nu_1) + \int_{X_1 \times X_2} \mathcal{T}_2(\delta_{x_2}, \nu_2)\, d\mu(x_1, x_2)$$

$\square$

2.10.4 Projections and Hopf-Lax formulae

We now investigate linear transfers which arise via inf-convolution.

Example 2.10.9 (Projection onto the set of balayées of a given measure)**.** Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(Y)$ and consider the following "projection" problem

$$\mathcal{S}(\mu, \nu) := \inf\{\mathcal{T}(\mu, \sigma); \sigma \in K_\nu\}, \tag{2.30}$$

where K_ν is a closed convex subset of $\mathcal{P}(Y)$ that depends on ν. Define on $\mathcal{P}(Y) \times \mathcal{P}(Y)$,

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

so that we may write $\mathcal{S}$ as

$$\mathcal{S}(\mu,\nu) = \mathcal{T} \star \mathcal{I}(\mu,\nu) = \inf_{\sigma\in\mathcal{P}(Y)}\{\mathcal{T}(\mu,\sigma) + \mathcal{I}(\sigma,\nu)\}.$$

In general, $\mathcal{S}$ will not be a backward linear transfer. However, consider the case when $K_\nu = \{\sigma\,;\,\sigma \preceq_{\mathcal{A}} \nu\}$ is the set of probability measures in $\mathcal{A}$-order with ν, for a convex cone $\mathcal{A}$ of continuous functions (see Section 2.5.3). Then $\mathcal{I} = \mathcal{B}$ where $\mathcal{B}$ is the balayage of measures from Section 2.5.3, and the inf-convolution property of Section 2.10.2 implies that $\mathcal{S}$ is a backward linear transfer with backward Kantorovich operator

$$S^- g(x) := T^-(\hat{g})(x),$$

where we recall, again from Section 2.5.3, that

$$\hat{g}(x) = \inf\{h(x)\,;\,h \in -\mathcal{A}, h \geq g\}.$$

A particular setting that has been studied in, e.g. [33], [22], is when $\mathcal{T}$ is the optimal transport with quadratic cost $c(x,y) = |x-y|^2$ and $\mathcal{A}$ is the collection of convex functions on some compact subset of $\mathbb{R}^n$. In that case, we have seen from Section 2.5.3 that $\hat{g}$ is the concave envelope of g, hence we obtain the duality

$$\inf\{\mathcal{T}(\mu,\sigma); \sigma \preceq_{\mathcal{A}} \nu\} = \sup_{g\in C(X)}\{\int_X \hat{g}\,d\nu - \int_X T^- g\,d\mu\}$$

where $T^- g(x) = \sup_{y\in X}\{g(y) - |x-y|^2\}$ is the Kantorovich operator of optimal transport (see Section 2.6). By noting $\hat{g} \geq g$, it follows that $T^-(\hat{g}) \geq T^-(g)$, so it actually suffices to restrict the supremum to g which are *concave*, in which case $\hat{g} = g$. The conclusion is

$$\inf\{\mathcal{T}(\mu,\sigma); \sigma \preceq_{\mathcal{A}} \nu\} = \sup_{g \text{ concave}}\{\int_X g\,d\nu - \int_X T^- g\,d\mu\}$$

which is the duality obtained by Gozlan et. al. (see [36], Theorem 2.11 (3)).

It is easy to see that the convolution of two 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 $c_1 \star c_2$. We use the inf-convolution property of linear transfers to establish this easy result.

Proposition 2.10.10 (Lifting convolutions to Wasserstein space). *Let X_i, $i = 0, 1, 2$, be compact spaces, and, $c_i : X_i \times X_{i+1} \to \mathbb{R} \cup \{+\infty\}$, $i = 0, 1$, lower semi-continuous cost functions. Define the corresponding optimal mass transports*

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

with corresponding forward and backward Kantorovich operators $T^+_{c_i}$ and $T^-_{c_i}$ as in Section 2.6. Consider the cost function on $X_0 \times X_2$ which is the convolution of the two cost functions,

$$c(x, x') := c_1 \star c_2(x, x') = \inf \left\{c_1(x, x_1) + c_2(x_1, x');\, x_i \in X_i\right\}.$$

Then

$$\begin{aligned}
\mathcal{T}_c(\mu, \nu) &= \mathcal{T}_{c_1} \star \mathcal{T}_{c_2}(\mu, \nu) \\
&= \sup_{f \in C(X_0)} \left\{\int_{X_2} T^+_{c_1} \circ T^+_{c_2} f\, d\nu - \int_{X_0} f\, d\mu\right\} \\
&= \sup_{g \in C(X_2)} \left\{\int_{X_2} g\, d\nu - \int_{X_0} T^-_{c_1} \circ T^-_{c_2} g d\mu\right\}.
\end{aligned}$$

Proof. By the inf-convolution property of Section 2.10.2, $\mathcal{T}_{c_1 \star c_2}$ has backward

Kantorovich operator equal to

$$\begin{aligned} T_{c_1}^- \circ T_{c_2}^- g(x) &= \sup_{x_1 \in X_1} \{T_{c_2}^- g(x_1) - c_1(x, x_1)\} \\ &= \sup_{x_1 \in X_1, x_2 \in X_2} \{g(x_2) - c_2(x_1, x_2) - c_1(x, x_1)\} \\ &= \sup_{x_2 \in X_2} \{g(x_2) - c(x, x_2)\} \\ &= T_c^- g(x). \end{aligned}$$

□

The next two problems are technically defined on $\mathbb{R}^d$ and not on a compact space, so strictly speaking, they do not fit into our compact framework. One then needs to change to the space $\mathcal{M}_1(\mathbb{R}^d)$ of finite Borel measures with finite first moments, in duality with $\mathrm{Lip}(M)$ the bounded uniformly Lipschitz functions on M.

Example 2.10.11 (The ballistic transfer (Barton-Ghoussoub [4])). Let L be a Tonelli Lagrangian, which we do not make precise here but instead define later in Chapter 4, which satisfies appropriate assumptions (see [4], (A0)). For $M = \mathbb{R}^d = M^*$ (M^* the dual space), the *deterministic ballistic mass transport* is defined as

$$\underline{\mathcal{B}}(\mu, \nu) := \inf\{\int_{M^* \times M} b(v, x) d\pi(v, x)\,;\, \pi \in \mathcal{K}(\mu, \nu)\},$$

where

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

Note that we can write b in the form

$$b(v, x) = \inf\{\langle v, y\rangle + c_L(y, x);\ y \in M\},$$

where the cost c_L is given by the generating function associated to L in Example 2.6.6, which means that b is the convolution of the Brenier cost of

Example 2.6.7 with the cost induced by the Lagrangian L.

By the inf-convolution property 2.10.3, it is both a forward, and backward, linear transfer. In particular, its backward Kantorovich operator is then the composition of the two corresponding backward Kantorovich operators, and we conclude

$$\underline{\mathcal{B}}(\mu,\nu) = \sup\{\int_M g(x)d\nu(x) - \int_{M^*} (-W_g(0,\cdot))^*(-v)d\nu(v)\,;\, g \in \mathrm{Lip}(M)\}$$

where W_g solves the Hamilton-Jacobi equation of Example 2.6.6 with $W_g(1,x) = g(x)$, and where $h^*(v) := \sup_{x\in M}\{\langle v,x\rangle - h(x)\}$ is the Legendre transform of a function h on M (see Theorem 2 in [4]).

Example 2.10.12 (Stochastic ballistic transfer (Barton-Ghoussoub [4]))**.** For the *stochastic ballistic transport problem*,

$$\underline{\mathcal{B}}^s(\mu,\nu) := \inf\left\{\mathbb{E}\left[\langle V, X(0)\rangle + \int_0^1 L(t,X,\beta_X(t,X))\,dt\right];\, V\sim\mu,\ X(\cdot)\in\mathcal{A}_{[0,1]},\ X(1)\sim\nu\right\},$$

where we are using the same notation for the stochastic processes as in Section 2.8.2, this is a convolution of the Brenier-Wasserstein transfer of Example 2.6.7 with now the stochastic transport of Example 2.8.7. Under suitable conditions on L (see [4], (A1), (A2), (A3)), we have, similarly to the deterministic version,

$$\underline{\mathcal{B}}^s(\mu,\nu) = \sup\left\{\int_M g(x)\,d\nu(x) - \int_{M^*} (-\psi_g(0,\cdot))^*(-v)\,d\mu(v); g\in \mathrm{Lip}(M)\right\},$$

only now ψ_g is the solution to the Hamilton-Jacobi-Bellman equation

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

In other words, $\underline{\mathcal{B}}^s$ is a backward linear transform with Kantorovich operator $T^- g(v) = (-\psi_g(0,\cdot))^*(-v)$.

2.11 Subdifferentials of convex transfers

If $\mathcal{T}$ is a backward linear transfer, then $\nu \mapsto \mathcal{T}(\mu,\nu) =: \mathcal{T}_\mu(\nu)$ is convex and weak* lower semi-continuous. Therefore one can consider its (weak*) subdifferential $\partial\mathcal{T}_\mu$ given by

$g \in \partial\mathcal{T}_\mu(\nu)$ if and only if $\mathcal{T}(\mu,\nu') \geq \mathcal{T}(\mu,\nu) + \int_Y g d(\nu'-\nu)$ 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 $\mathcal{T}_\mu^*(g) = \int T^- g d\mu$, we then obtain the following characterization of the subdifferentials, which says that the subdifferential $\partial\mathcal{T}_\mu(\nu)$ is simply all those g for which dual attainment holds for $\mathcal{T}(\mu,\nu)$.

Proposition 2.11.1. *Let $\mathcal{T}$ be a backward linear transfer. For any $\mu \in D_1(\mathcal{T})$, the subdifferential of $\mathcal{T}_\mu : \mathcal{P}(Y) \to \mathbb{R}\cup\{+\infty\}$ at $\nu \in \mathcal{P}(Y)$ is given by*

$$\partial\mathcal{T}_\mu(\nu) = \left\{ g \in C(Y) \,:\, \int_Y g(y)d\nu(y) - \int_X T^- g(x)d\mu(x) = \mathcal{T}(\mu,\nu) \right\}$$

It is easy to see that the same expressions hold (with the necessary modifications) for backward convex transfers.

In the following, we observe some elementary consequences for elements in the subdifferential.

Proposition 2.11.2. *Suppose $\mathcal{T}$ is a linear backward transfer such that $\{\delta_x\,;\, x \in X\} \subset D_1(\mathcal{T})$. Fix $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ such that $\mathcal{T}(\mu,\nu) < +\infty$. Then, there exists $\bar\pi \in \mathcal{K}(\mu,\nu)$ such that for each $\bar g \in \partial\mathcal{T}_\mu(\nu)$, we have*

$$T^-\bar g(x) = \int_Y \bar g(y)d\bar\pi_x(y) - \mathcal{T}(\delta_x, \bar\pi_x), \quad \text{for } \mu\text{-a.e. } x \in X,$$

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 $\bar g \in \partial\mathcal{T}_\mu(\nu)$ for some $\nu \in \mathcal{P}(Y)$. If $x \to \sigma_x$ is any Borel-measurable selection such that

$$T^-\bar g(x) = \sup_\sigma \left\{ \int \bar g d\sigma - \mathcal{T}(\delta_x,\sigma) \right\} = \int_Y \bar g d\sigma_x - \mathcal{T}(\delta_x,\sigma_x),$$

then $\mathcal{T}(\mu,\nu)$ is attained by the measure $\bar\pi$ defined via $d\bar\pi(x,y) = d\sigma_x(y)d\mu(x)$.

Proof. By a recent result [3], there exists $\bar\pi \in \mathcal{K}(\mu,\nu)$ such that

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

If $\bar f \in \partial\mathcal{T}_\mu(\nu)$, then by definition

$$\int_Y \bar f(y)d\nu(y) - \int_X T^- \bar f(x)d\mu(x) = \mathcal{T}(\mu,\nu) = \int_X \mathcal{T}(x,\bar\pi_x)d\mu(x),$$

that is $\int_X \left[T^-\bar f(x) - \int_Y \bar f(y)d\bar\pi_x(y) + \mathcal{T}(x,\bar\pi_x)\right]d\mu = 0$. Since $T^-\bar f(x) = \sup_\sigma \left\{\int \bar f d\sigma - \mathcal{T}(x,\sigma)\right\}$, 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 $\int \bar f d\nu - \int T^-\bar f d\mu = \mathcal{T}(\mu,\nu)$. From the expression $T^-\bar f(x) = \sup_\sigma \left\{\int_Y \bar f d\sigma - \mathcal{T}(\delta_x,\sigma)\right\}$, we know the supremum will be achieved by some σ_x. Defining $\tilde\pi$ by $d\tilde\pi(x,y) = d\mu(x)d\sigma_x(y)$, and the right marginal of $\tilde\pi$ by $\tilde\nu$, we integrate against μ to achieve

$$\int T^-\bar f d\mu = \int \bar f d\tilde\nu - \int \mathcal{T}(\delta_x,\sigma_x)d\mu.$$

This shows that $\mathcal{T}(\mu,\tilde\nu) = \inf_{\pi\in\mathcal{K}(\mu,\tilde\nu)} \int \mathcal{T}(\delta_x,\pi_x)d\mu = \int \mathcal{T}(\delta_x,\sigma_x)d\mu$, and consequently, $\bar f \in \partial\mathcal{T}_\mu(\tilde\nu)$. But by strict convexity, this can only be true if $\tilde\nu = \nu$. □

Note that while we can use general existence results such as the Brondsted-Rockafellar theorem [46], to state that $\partial\mathcal{T}_\mu(\nu)$ exists for a weak*-dense set of $\nu \in \mathcal{P}(Y)$ (and therefore the dual problem is "generically" attained for this weak*-dense set), proving attainment in the dual problem is, in general, a difficult problem.

Corollary 2.11.3. *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(Y)$ such that the Dirac masses are in $D_1(\mathcal{T})$, and 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.*

We briefly mention the following Euler-Lagrange equation for variational problems on spaces of measures, which follows closely [24] (see in particular Theorem 2.2 there).

Proposition 2.11.4. *Let $\mathcal{T}_\alpha(\mu,\nu) := \int_X \alpha\left(\frac{d\nu}{d\mu}\right) d\mu$ be the generalised entropy functional considered in Example 2.9.6, where $\alpha : [0,\infty) \to \mathbb{R}$ is in $C^1([0,\infty))$, strictly convex, and superlinear. Let $\mathcal{T}$ be any backward linear 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. $x \in X$,*

$$\alpha'\left(\frac{d\bar{\nu}}{d\mu}\right) = \bar{f} + C,$$

where C is a constant.

Proof. Recall from Example 2.9.6 that $\mathcal{T}_\alpha$ is a backward convex transfer with

$$\mathcal{T}_\mu^*(g) = \inf_{t\in\mathbb{R}} \left\{ \int_X [\alpha^\oplus(g(x)+t) - t] d\mu(x) \right\},$$

where $\alpha^\oplus(t) := \sup_{s\geq 0}\{st - \alpha(t)\}$. It follows that

$$\alpha'\left(\frac{d\nu}{d\mu}\right) \in \partial\mathcal{T}_\mu(\nu).$$

We can see this either directly from the subdifferential definition, or from observing

$$\alpha^\oplus\left(\alpha'\left(\frac{d\nu}{d\mu}\right)\right) = \frac{d\nu}{d\mu}\alpha'\left(\frac{d\nu}{d\mu}\right) - \alpha\left(\frac{d\nu}{d\mu}\right).$$

In particular,

$$\mathcal{T}_\mu^*\left(\alpha'\left(\frac{d\nu}{d\mu}\right)\right) = \int_X \alpha^\oplus\left(\alpha'\left(\frac{d\nu}{d\mu}\right)\right) d\mu.$$

The rest is a straightforward adaptation of Theorem 2.2 in [24]. □

.

Chapter 3

Dualities for transfer inequalities

3.1 Introduction

Let $\mathcal{T}_c$ be an optimal transport associated to a cost function c on $X \times X$, and let J be a functional on $\mathcal{P}(X) \times \mathcal{P}(X)$. A family of inequalities, the *transport inequalities*, compare the cost $\mathcal{T}_c(\mu, \nu)$ of transporting a measure $\nu \in \mathcal{P}(X)$ to a fixed "reference measure" $\mu \in \mathcal{P}(X)$, in the form (see, e.g. the survey article [34])

$$\alpha(\mathcal{T}_c(\nu, \mu)) \leq J(\mu, \nu) \quad \text{or} \quad \alpha(\mathcal{T}_c(\mu, \nu)) \leq J(\mu, \nu),$$

where $\alpha : [0, \infty) \to [0, \infty)$ is a given increasing function with $\alpha(0) = 0$. A classical choice for J is the relative entropy functional

$$\mathcal{H}(\mu, \nu) = \textstyle\int_X \log(\frac{d\nu}{d\mu})\, d\nu \text{ if } \nu \ll \mu \text{ and } +\infty \text{ otherwise,}$$

in which case they are known as the *transport-entropy* inequalities, including Talagrand's transport inequality for the Gaussian measure [57]. One interest of transport-entropy inequalities is that they imply *concentration of measure* phenomena (see e.g. Theorem 1.7 in [34] originally due to Marton [41, 42], or the survey [38]).

In order to prove a transport-entropy inequality, one can take advantage of the fact that both $\mathcal{T}_c$ and $\mathcal{H}$ have dualities, which in the present context, is to exploit the fact that $\mathcal{T}_c$ is a linear transfer, and $\mathcal{H}$ is a convex transfer. As was originally proved by Bobkov and Götze [8], one can show (see [34]) the following equivalence.

Theorem 3.1.1 ([34], Theorem 3.2, Corollary 3.3)**.** *Let c be a lower semi-continuous cost function, $\alpha : [0,\infty) \to [0,\infty)$ convex, increasing, and $\alpha(0) = 0$, and fix a reference measure $\mu \in \mathcal{P}(X)$. Then the following are equivalent:*

1. *$\alpha(\mathcal{T}_c(\nu,\mu)) \leq \mathcal{H}(\mu,\nu)$ for all $\nu \in \mathcal{P}(X)$,*
2. *For all $g \in C(X)$,*

$$\int_X e^{-sT_c^- g} d\mu \leq \int_X e^{-sgd\mu + \alpha^{\oplus}(s)}, \qquad s \geq 0$$

where we recall that $\alpha^{\oplus}(s) := \sup_{t\geq 0}\{st - \alpha(t)\}$ for $s \geq 0$, and $T_c^- g(x) = \sup_{y\in X}\{g(y) - c(x,y)\}$.

Proof. The proof is illustrative so we shall provide it here: We have as a backward linear transfer,

$$\mathcal{T}_c(\nu,\mu) = \sup_{g\in C(X)} \{\int_X g d\mu - \int_X T_c^- g d\nu\}$$

as well as $\alpha(t) = \sup_{s\geq 0}\{st - \alpha^{\oplus}(s)\}$ (this follows by extending α to $+\infty$ for $t < 0$ (denoting this by $\tilde{\alpha}$) and noting α is convex and lower semi-continuous so $\alpha(t) = \sup_{s\in\mathbb{R}}\{st - \tilde{\alpha}^*(s)\} = \sup_{s\geq 0}\{st - \alpha^{\oplus}(s)\}$). Then for all $g \in C(X)$ and all $s \geq 0$, it follows that

$$\alpha(\mathcal{T}_c(\nu,\mu)) = \sup_{s\geq 0} \sup_{g\in C(X)} \{\int_X sg d\mu - \int_X sT_c^- g d\nu - \alpha^{\oplus}(s)\}.$$

Then we write

$$\alpha(\mathcal{T}_c(\nu,\mu)) \leq \mathcal{H}(\mu,\nu) \quad \text{for all } \nu \in \mathcal{P}(X) \iff 0 \leq \inf_{\nu\in\mathcal{P}(X)} \{\mathcal{H}(\mu,\nu) - \alpha(\mathcal{T}_c(\nu,\mu))\}.$$

Substituting directly the above expression for $\alpha(\mathcal{T}_c(\nu,\mu))$ yields

$$\begin{aligned}
0 &\leq \inf_{\nu\in\mathcal{P}(X)} \inf_{g\in C(X)} \inf_{s\geq 0} \{\mathcal{H}(\mu,\nu) - \int_X sg d\mu + \int_X sT_c^- g d\nu + \alpha^{\oplus}(s)\} \\
&= \inf_{g\in C(X)} \inf_{s\geq 0} \left\{ \int_X (-sg+\alpha^{\oplus}(s))d\mu - \sup_{\nu\in\mathcal{P}(X)} \{\int_X (-T_c^- g)d\nu - \mathcal{H}(\mu,\nu)\} \right\} \\
&= \inf_{g\in C(X)} \inf_{s\geq 0} \{ \int_X (-sg+\alpha^{\oplus}(s))d\mu - \mathcal{H}_\mu^*(-sT_c^- g)\} \\
&= \inf_{g\in C(X)} \inf_{s\geq 0} \{ \int_X (-sg+\alpha^{\oplus}(s))d\mu - \log \int_X e^{-sT_c^- g} d\mu\}.
\end{aligned}$$

Hence by rearranging and taking the exponential, we obtain

$$\int_X e^{-sT_c^- g} d\mu \leq e^{\int_X -sgd\mu + \alpha^{\oplus}(s)} \quad \text{for all } g\in C(X) \text{ and all } s\geq 0.$$

□

The goal now is to deduce analogous duality statements when $\alpha(\mathcal{T}_c)$ is replaced by $\mathcal{T}$ which is a linear or convex coupling, and when the relative entropy $\mathcal{H}$ is replaced by a entropic, or more generally, a convex transfer (recall Definition 2.9.8). The main interest for computing the dualities is that they transform inequalities between measures, to inequalities on functions. Therefore if one wishes to prove that a particular transfer inequality holds, they can instead prove the equivalent functional inequality, for which one may hope to take advantage of the extensive literature that has been developed for functional inequalities. As an example, see [8].

We shall assume throughout that the operators associated to the couplings/transfers discussed in this chapter, map continuous functions into at least the space of upper semi-continuous functions. In that case, we use, in advance, the inf-convolution property Corollary 5.1.5 to justify certain Legendre transform computations, which is the statement of Proposition 2.10.3 and Corollary 2.10.4 extended to upper semi-continuous functions (note Chapter 5 does not depend on Chapter 3).

3.2 Backward convex coupling and backward linear transfer

Fix $\mu \in \mathcal{P}(X)$. As a generalization of the transport-entropy inequality of type

$$\alpha(\mathcal{T}_c(\sigma,\mu)) \leq \mathcal{H}(\mu,\sigma) \quad \text{for all } \sigma,$$

we would like to prove inequalities such as,

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

where $\mathcal{F}_1$ is a backward convex transfer and $\mathcal{F}_2$ is a backward convex coupling. In this section, we provide the dual characterization for such an inequality, and refine it further to the special case when $\mathcal{F}_1 = \lambda\mathcal{E} \star \mathcal{T}$. To find the dual characterization, we compute the duality for the $\inf_\sigma\{\mathcal{F}_1(\mu,\sigma) - \mathcal{F}_2(\sigma,\nu)\}$ where μ and ν are some fixed probability measures.

Proposition 3.2.1. *Let $\mathcal{F}_1$ be a backward convex transfer on $\mathcal{P}(X_1)\times\mathcal{P}(Y_1)$ with operators $(F^-_{1,i})_{i\in I}$, and $\mathcal{F}_2$ a backward convex coupling on $\mathcal{P}(Y_1)\times\mathcal{P}(X_2)$ with operators $(F^-_{2,j})_{j\in J}$. Then for fixed $\mu \in \mathcal{P}(X_1)$ and $\nu \in \mathcal{P}(X_2)$, we have*

$$\inf_{\sigma\in\mathcal{P}(Y_1)}\{\mathcal{F}_1(\mu,\sigma)-\mathcal{F}_2(\sigma,\nu)\} = \inf_{g\in C(X_2)}\inf_{j\in J}\left\{-\inf_{i\in I}\int_{X_1} F^-_{1,i}(-F^-_{2,j}g)d\mu - \int_{X_2} gd\nu\right\}.$$

Proof. As $\mathcal{F}_2$ is a backward convex coupling, we may write $\mathcal{F}_2(\sigma,\nu) = \sup_{g\in C(X_2)}\sup_{j\in J}\{\int_{X_2} gd\nu - \int_{Y_1} F^-_{2,j}gd\sigma\}$. We may then substitute this into the expression

$$\begin{aligned}\inf_{\sigma\in\mathcal{P}(Y_1)}\{\mathcal{F}_1(\mu,\sigma) - \mathcal{F}_2(\sigma,\nu)\} &= \inf_{\sigma\in\mathcal{P}(Y_1)}\inf_{g\in C(X_2)}\inf_{j\in J}\left\{\mathcal{F}_1(\mu,\sigma) - \int_{X_2} gd\nu + \int_{Y_1} F^-_{2,j}gd\sigma\right\}\\ &= \inf_{g\in C(X_2)}\inf_{j\in J}\left\{-\sup_{\sigma\in\mathcal{P}(Y_1)}\{\int_{Y_1}(-F^-_{2,j}g)d\sigma - \mathcal{F}_1(\mu,\sigma)\} - \int_{X_2} gd\nu\right\}\\ &= \inf_{g\in C(X_2)}\inf_{j\in J}\left\{-\inf_{i\in I}\int_{X_1} F^-_{1,i}(-F^-_{2,j}g)d\mu - \int_{X_2} gd\nu\right\}\end{aligned}$$

where the last equality follows since $\mathcal{F}_1$ is a backward convex transfer and thus has the specified Legendre transform. $\square$

Corollary 3.2.2. *Under the same hypotheses as the above Proposition 3.2.1, the following are equivalent:*

1. *For all $\sigma \in \mathcal{P}(Y_1)$, we have*

$$\mathcal{F}_2(\sigma, \nu) \leq \mathcal{F}_1(\mu, \sigma).$$

2. *For all $g \in C(X_2)$, it holds*

$$\sup_{j \in J} \inf_{i \in I} \int_{X_1} F_{1,i}^{-}(-F_{2,j}^{-}(g)) d\mu + \int_{X_2} g d\nu \leq 0.$$

Proof. By Proposition 3.2.1 it follows immediately that

$$0 \leq \inf_{\sigma \in \mathcal{P}(Y_1)} \{\mathcal{F}_1(\mu, \sigma) - \mathcal{F}_2(\sigma, \nu)\} \quad \text{if and only if} \quad 0 \leq \inf_{g \in C(X_2)} \inf_{j \in J} \left\{ -\inf_{i \in I} \int_{X_1} F_{1,i}^{-}(-F_{2,j}^{-} g) d\mu - \int_{X_2} g d\nu \right\}$$

This equivalence is easily seen by rearrangement to be the statement of equivalence between Item 1 and 2 in the above statement. $\square$

Corollary 3.2.3. *Specializing Corollary 3.2.2 to the case where $\mathcal{F}_1 := \lambda \mathcal{E} \star \mathcal{T}$ (and relabeling $\mathcal{F}_2$ to $\mathcal{F}$) for $\mathcal{E}$ a backward β-entropic transfer on $\mathcal{P}(X_1) \times \mathcal{P}(Y_1)$ with operator E^-, $\mathcal{T}$ a backward linear transfer on $\mathcal{P}(Y_1) \times \mathcal{P}(Y_2)$ with Kantorovich operator T^-, and $\lambda > 0$, the following are equivalent:*

1. *For all $\sigma \in \mathcal{P}(Y_2)$, we have*

$$\mathcal{F}(\sigma, \nu) \leq \lambda \, \mathcal{E} \star \mathcal{T} \, (\mu, \sigma).$$

2. *For all $g \in C(X_2)$,*

$$\sup_{j \in J} \beta \left(\int_{X_1} E^- \circ T^- \left(\frac{-F_j^-(\lambda g)}{\lambda} \right) d\mu \right) + \int_{X_2} g \, d\nu \leq 0.$$

Proof. Note that the inf-convolution $\lambda \mathcal{E} \star \mathcal{T}$ is a backward $(\lambda\beta)$-entropic transfer (in particular, a backward convex transfer) with operator $g \mapsto$

$E^- \circ T^-(\frac{g}{\lambda})$, therefore the above proposition applies and we obtain the dual inequality

$$\sup_{j\in J} \lambda\beta\left(\int_{X_1} E^- \circ T^-\left(\frac{-F_j^-(g)}{\lambda}\right) d\mu\right) + \int_{X_2} g\, d\nu \leq 0 \quad \text{for all } g \in C(X_2).$$

Dividing by λ and relabeling $h := \frac{g}{\lambda}$, the inequality has to hold for all $h \in C(X_2)$, which completes the equivalence. □

Corollary 3.2.4. *Under the same setting as the above Corollary* **??**, *but with $\mathcal{E} = \mathcal{H}$ the relative entropy, the following are equivalent.*

1. *For all $\sigma \in \mathcal{P}(Y_2)$, we have $\mathcal{F}(\sigma,\nu) \leq \lambda\,\mathcal{H} \star \mathcal{T}\,(\mu,\sigma)$.*
2. *For all $g \in C(X_2)$,*

$$\sup_{j\in J} \int_{X_1} e^{T^-\left(\frac{-F_j^-(\lambda g)}{\lambda}\right)} d\mu \leq e^{-\int_{X_2} g\, d\nu}.$$

In particular, if $\mathcal{T}$ is the identity transfer and $\mathcal{F}$ is a backward linear transfer, then the following are equivalent:

1. *For all $\sigma \in \mathcal{P}(Y_2)$, we have $\mathcal{F}(\sigma,\nu) \leq \lambda\,\mathcal{H}\,(\mu,\sigma)$*
2. *For all $g \in C(X_2)$, we have $\int_{X_1} e^{-\frac{F^-(\lambda g)}{\lambda}}\, d\mu \leq e^{-\int_{X_2} g\, d\nu}$.*

Proof. Taking in the above Corollary $\mathcal{E} = \mathcal{H}$ the relative entropy, then $\beta(t) = \log(t)$ and $E^- g = e^g$, and we deduce that for all $g \in C(X_2$ and all $j \in J$,

$$\log \int_{X_1} e^{T^-\left(\frac{-F_j^-(\lambda g)}{\lambda}\right)} d\mu + \int_{X_2} g\, d\nu \leq 0,$$

which gives the asserted equivalence. If further $\mathcal{T}$ is the identity transfer (i.e. $\mathcal{T}(\mu,\nu) = 0$ if $\mu = \nu$, $+\infty$ otherwise), then $T^- g = g$. □

3.3 Forward convex coupling and backward linear transfer

We are now interested in inequalities such as

$$\mathcal{F}_2(\nu,\sigma) \le \mathcal{F}_1(\mu,\sigma) \quad \text{for all } \sigma \in \mathcal{P}(X), \tag{3.1}$$

where $\mathcal{F}_1$ is a backward convex transfer and $\mathcal{F}_2$ is a backward convex coupling (which, if we define $\tilde{F}_2(\sigma,\nu) := \mathcal{F}_2(\nu,\sigma)$, we can view as an inequality between a forward convex coupling and backward linear transfer, hence the name of the section). As in the previous section, to compute the dual inequality to (3.1), we find the dual expression for the inf-convolution $\inf_{\sigma\in\mathcal{P}(X)}\{\mathcal{F}_1(\mu,\sigma) - \mathcal{F}_2(\nu,\sigma)\}$.

Proposition 3.3.1. *Let $\mathcal{F}_1$ be a backward convex transfer on $\mathcal{P}(X_1) \times \mathcal{P}(Y_1)$ with operators $(F^-_{1,i})_{i\in I}$, and let $\mathcal{F}_2$ be a backward convex coupling on $\mathcal{P}(X_2) \times \mathcal{P}(Y_1)$ with operators $(F^-_{2,j})_{j\in J}$. Then*

$$\inf_{\sigma\in\mathcal{P}(Y_1)}\{\mathcal{F}_1(\mu,\sigma) - \mathcal{F}_2(\nu,\sigma)\} = \inf_{g\in C(X_2)}\left\{-\inf_{i\in I}\int_{X_1} F^-_{1,i}g\,d\mu + \inf_{j\in J}\int_{X_2} F^-_{2,j}g\,d\nu\right\}$$

Proof. We have from the expression for $\mathcal{F}_2$ as a backward convex coupling,

$$\begin{aligned}
\inf_{\sigma\in\mathcal{P}(Y_1)}\{\mathcal{F}_1(\mu,\sigma) - \mathcal{F}_2(\nu,\sigma)\} &= \inf_{\sigma\in\mathcal{P}(Y_1)}\inf_{g\in C(X_2)}\inf_{j\in J}\left\{\mathcal{F}_1(\mu,\sigma) - \int_{Y_1} g\,d\sigma + \int_{X_2} F^-_{2,j}g\,d\nu\right\} \\
&= \inf_{g\in C(X_2)}\left\{-\sup_{\sigma\in\mathcal{P}(Y_1)}\{\int_{Y_1} g\,d\sigma - \mathcal{F}_1(\mu,\sigma)\} + \inf_{j\in J}\int_{X_2} F^-_{2,j}g\,d\nu\right\} \\
&= \inf_{g\in C(X_2)}\left\{-\inf_{i\in I}\int_{X_1} F^-_{1,i}g\,d\mu + \inf_{j\in J}\int_{X_2} F^-_{2,j}g\,d\nu\right\}
\end{aligned}$$

where the last equality follows since $\mathcal{F}_1$ is a backward convex transfer. □

Corollary 3.3.2. *Let $\mathcal{F}_1$ be a backward convex transfer on $\mathcal{P}(X_1) \times \mathcal{P}(Y_1)$ with operators $(F^-_{1,i})_{i\in I}$, and let $\mathcal{F}_2$ be a backward convex coupling on $\mathcal{P}(X_2) \times \mathcal{P}(Y_1)$ with operators $(F^-_{2,j})_{j\in J}$.*

Then the following are equivalent.

1. *For all $\sigma \in \mathcal{P}(Y_1)$, we have $\mathcal{F}_2(\nu,\sigma) \le \mathcal{F}_1(\mu,\sigma)$.*

2. *For all $g \in C(Y_1)$, we have $\inf_{i\in I}\int_{X_1} F^-_{1,i}(g)d\mu \le \inf_{j\in J}\int_{X_2} F^-_{2,j}(g)d\nu$.*

Proof. From Proposition 3.3.1, we have

$$0 \le \inf_{\sigma\in\mathcal{P}(Y_1)}\{\mathcal{F}_1(\mu,\sigma) - \mathcal{F}_2(\nu,\sigma)\}$$

if and only if

$$0 \le \inf_{g\in C(X_2)}\left\{-\inf_{i\in I}\int_{X_1} F^-_{1,i} g d\mu + \inf_{j\in J}\int_{X_2} F^-_{2,j} g d\nu\right\}.$$

This equivalence then reduces to the equivalence between 1. and 2. □

Corollary 3.3.3. *Specializing Corollary 3.3.2 to the case where $\mathcal{F}_1 = \lambda\mathcal{E}\star\mathcal{T}$ (and relabeling $\mathcal{F}_2$ to $\mathcal{F}$), for $\mathcal{E}$ a backward β-entropic transfer on $\mathcal{P}(X_1)\times\mathcal{P}(Y_1)$ with operator E^-, $\mathcal{T}$ a backward linear transfer on $\mathcal{P}(Y_1)\times\mathcal{P}(Y_2)$ with 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 $\mathcal{F}(\nu,\sigma) \le \lambda\,\mathcal{E}\star\mathcal{T}\,(\mu,\sigma)$.*

2. *For all $g \in C(X_2)$, we have $\beta\big(\int_{X_1} E^-\circ T^- g\, d\mu\big) \le \inf\limits_{i\in I}\frac{1}{\lambda}\int_{X_2} F^-_i(\lambda g)d\nu$.*

Proof. Note that the inf-convolution $\lambda\mathcal{E}\star\mathcal{T}$ is a backward $(\lambda\beta)$-entropic transfer (in particular, a backward convex transfer) with operator $g \mapsto E^-\circ T^-(\frac{g}{\lambda})$, therefore the above Corollary 3.3.2 applies and we obtain the dual inequality

$$\lambda\beta\left(\int_{X_1} E^-\circ T^-\left(\frac{g}{\lambda}\right)\, d\mu\right) \le \inf_{i\in I}\int_{X_2} F^-_i(g)d\nu \quad \text{for all } g \in C(X_2).$$

Dividing by λ and relabeling $h := \frac{g}{\lambda}$, the inequality has to hold for all $h \in C(X_2)$, which completes the equivalence. □

We now apply the above to the case where $\mathcal{E}$ is the relative entropy $\mathcal{H}$, in which case $\beta(t) = \log(t)$ and $E^- g = e^g$.

Corollary 3.3.4. *In the setting of the above corollary, if we take $\mathcal{E} = \mathcal{H}$ the relative entropy, then, for any fixed pair of probability measures $\mu \in \mathcal{P}(X_1)$ and $\nu \in \mathcal{P}(X_2)$, the following are equivalent:*

1. *For all $\sigma \in \mathcal{P}(Y_2)$, we have $\mathcal{F}(\nu, \sigma) \leq \lambda \mathcal{H} \star \mathcal{T}(\mu, \sigma)$*

2. *For all $g \in C(X_2)$, we have $\log\big(\int_{X_1} e^{T^- g}\, d\mu\big) \leq \inf\limits_{i \in I} \frac{1}{\lambda} \int_{X_2} F_i^-(\lambda g) d\nu$.*

Corollary 3.3.5. *In the setting of Corollary 3.3.3, taking $\mathcal{F} = \mathcal{E}_2$ to be a backward β_2-entropic transfer on $\mathcal{P}(X_2) \times \mathcal{P}(Y_2)$ with operator E_2^-, the following are equivalent:*

1. *For all $\sigma \in \mathcal{P}(Y_2)$, we have $\mathcal{E}_2(\nu, \sigma) \leq \lambda \mathcal{E}_1 \star \mathcal{T}(\mu, \sigma)$.*

2. *For all $g \in C(X_2)$,*

$$\beta_1\Big(\int_{X_1} E_1^- \circ T^- g\, d\mu\Big) \leq \frac{1}{\lambda} \beta_2\Big(\int_{X_2} E_2^-(\lambda g) d\nu\Big).$$

3.3.1 Moment measures

The equivalences stated in Proposition 3.3.2 and following corollaries, was the writing of the duality for $\inf_{\sigma \in \mathcal{P}(Y_1)}\{\mathcal{F}_1(\mu, \sigma) - \mathcal{F}_2(\nu, \sigma)\}$. It turns out that for a particular choice of $\mathcal{F}_1$ and $\mathcal{F}_2$, this duality is related to the study of *moment measures*, which we briefly highlight here.

Given a convex function $u : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ with $\int_{\mathbb{R}^d} e^{-u(x)} dx < +\infty$, its *moment measure* is defined as the measure $\mu := (\nabla u)_\# \rho$, where $d\rho(x) = e^{-u(x)} dx$. Conversely, we say μ is a moment measure if there exists a convex function u such that $\mu = (\nabla u)_\# \rho$. In [18], the authors provide sufficient and necessary conditions for a probability measure μ to be the moment measure of some convex function.

To prove the existence of a convex function u whose moment measure is the given measure μ, one can introduce, as was done by Santambrogio, a variational problem involving quadratic optimal transport and relative entropy. Indeed, the connection to quadratic optimal transport is via Brenier's theorem, namely that $\mu = (\nabla u)_\# \rho$ implies ∇u is the optimal transport map for the transport of ρ to μ with quadratic cost $c(x,y) = |x-y|^2$. At the same time, in light of the exponential form of the density $d\rho = e^{-u(x)}dx$, one is also not surprised that the logarithmic entropy also plays a role.

To be precise, let $\mathcal{H}(dx,\nu) := \int_{\mathbb{R}^d} \log(\frac{d\nu}{dx})d\nu$ if $\nu \in \mathcal{P}(\mathbb{R}^d)$ is absolutely continuous with respect to Lebesgue measure, and $+\infty$ otherwise. Recalling from Example 2.6.7, the Brenier transfer

$$\mathcal{T}(\mu,\nu) := \inf_{\pi\in\mathcal{K}(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} -\langle x,y\rangle d\pi(x,y),$$

consider the problem

$$\inf_{\sigma\in\mathcal{P}_1(\mathbb{R}^d)} \{\mathcal{H}(dx,\sigma) - \mathcal{T}(\sigma,\mu)\}.$$

If we allow ourselves the non-compact setting of $\mathbb{R}^d$, an immediate application of (4) in Proposition **??** is the following result in [18]

$$\inf_{\sigma\in\mathcal{P}(\mathbb{R}^d)} \{\mathcal{H}(dx,\sigma) - \mathcal{W}_2(\sigma,\bar{\mu})\} = \inf_{f\in\mathcal{C}(\mathbb{R}^d)} \{-\log \int e^{-f^*}\, dx + \int f\, d\mu\},$$

where $\mathcal{C}(\mathbb{R}^d)$ is the cone of convex functions on $\mathbb{R}^d$, $\mathcal{W}_2(\sigma,\bar{\mu})$ is the Brenier transfer of Example 2.6.7, and $\bar{\mu}$ is defined as $\int_{\mathbb{R}^d} f(x)d\bar{\mu}(x) = \int_{\mathbb{R}^d} f(-x)d\mu(x)$. Note that in this case, $T^+f(x) = -f^*(-x)$, $E^-f = e^f$ and $\beta(t) = \log t$, and since $g^{**} \leq g$,

$$\begin{aligned}
\inf_{\sigma\in\mathcal{P}(\mathbb{R}^d)} \{\mathcal{H}(dx,\sigma) - \mathcal{W}_2(\sigma,\bar{\mu})\} &= \mathcal{H}\star(-\mathcal{W}_2)(dx,\bar{\mu}) \\
&= \inf\{-\log \int e^{-g}\, dx + \int g^*(x)\, d\mu; g \in C(\mathbb{R}^d)\} \\
&= \inf\{-\log \int e^{-f^*}\, dx + \int f\, d\mu; f \in \mathcal{C}(\mathbb{R}^d)\}.
\end{aligned}$$

We remark that a particular interest is the characterization of those measures μ (the *moment measures*) for which there is attainment in both minimisation problems (see Cordero-Erausquin and Klartag [18]).

3.4 Maurey-type inequalities

Proposition 3.4.1. *Let $\mathcal{F}_1$ be a backward convex transfer on $\mathcal{P}(X_1)\times\mathcal{P}(Y_1)$ with operators $(F^-_{1,i})_{i\in I}$, $\mathcal{F}_2$ a backward convex transfer on $\mathcal{P}(X_2)\times\mathcal{P}(Y_2)$ with operators $(F^-_{2,j})_{j\in J}$, and $\mathcal{F}_3$ a backward convex coupling on $\mathcal{P}(Y_1)\times\mathcal{P}(Y_2)$ with operators $(F^-_{3,k})_{k\in K}$. Then for fixed $\mu\in\mathcal{P}(X_1)$ and $\nu\in\mathcal{P}(X_2)$, the following holds:*

$$\inf_{\sigma_1\in\mathcal{P}(Y_1),\sigma_2\in\mathcal{P}(Y_2)}\{\mathcal{F}_1(\mu,\sigma_1)-\mathcal{F}_3(\sigma_1,\sigma_2)+\mathcal{F}_2(\nu,\sigma_2)\}$$
$$=\inf_{g\in C(X_2)}\{-\sup_{k\in K}\inf_{i\in I}\int_{X_1}F^-_{1,i}(-F^-_{3,k}(g))\,d\mu-\inf_{j\in J}\int_{X_2}F^-_{2,j}(g)\,d\nu\}.$$

Proof. Since as a convex coupling,

$$\mathcal{F}_3(\sigma,\sigma_2)=\sup_{g\in C(Y_2)}\{\int_{Y_2}gd\sigma_2-\inf_{k\in K}\int_{Y_1}F^-_{3,k}gd\sigma_1,$$

we can substitute this expression into the following inf-convolution,

$$\begin{aligned}
&\inf_{\sigma_1\in\mathcal{P}(Y_1),\sigma_2\in\mathcal{P}(Y_2)}\{\mathcal{F}_1(\mu,\sigma_1)-\mathcal{F}_3(\sigma_1,\sigma_2)+\mathcal{F}_2(\nu,\sigma_2)\}\\
&=\inf_{\sigma_1\in\mathcal{P}(Y_1),\sigma_2\in\mathcal{P}(Y_2)}\inf_{g\in C(Y_2)}\left\{\mathcal{F}_1(\mu,\sigma_1)-\int_{Y_2}gd\sigma_2+\inf_{k\in K}\int_{Y_1}F^-_{3,k}gd\sigma_1+\mathcal{F}_2(\nu,\sigma_2)\right\}\\
&=\inf_{g\in C(Y_2)}\left\{-\sup_{k\in K}\sup_{\sigma_1\in\mathcal{P}(Y_1)}\{\int_{Y_1}(-F^-_{3,k}g)d\sigma_1-\mathcal{F}_1(\mu,\sigma_1)\}-\sup_{\sigma\in\mathcal{P}(Y_2)}\{\int_{Y_2}gd\sigma_2-\mathcal{F}_2(\nu,\sigma_2)\}\right\}\\
&=\inf_{g\in C(Y_2)}\left\{-\sup_{k\in K}\inf_{i\in I}\int_{X_1}F^-_{1,i}(-F^-_{3,k}g)d\nu-\inf_{j\in J}\int_{X_2}F^-_{2,j}(g)d\nu\right\}
\end{aligned}$$

□

Corollary 3.4.2. *Let $\mathcal{F}_1$ be a backward convex transfer on $\mathcal{P}(X_1)\times\mathcal{P}(Y_1)$ with operators $(F^-_{1,i})_{i\in I}$, $\mathcal{F}_2$ a backward convex transfer on $\mathcal{P}(X_2)\times\mathcal{P}(Y_2)$ with operators $(F^-_{2,j})_{j\in J}$, and $\mathcal{F}_3$ a backward convex coupling on $\mathcal{P}(Y_1)\times\mathcal{P}(Y_2)$ with operators $(F^-_{3,k})_{k\in K}$. Then for fixed $\mu\in\mathcal{P}(X_1)$ and $\nu\in\mathcal{P}(X_2)$, the following are equivalent:*

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

$$\mathcal{F}_3(\sigma_1, \sigma_2) \leq \mathcal{F}_1(\mu, \sigma_1) + \mathcal{F}_2(\nu, \sigma_2).$$

2. *For all $g \in C(Y_2)$*

$$\sup_{k \in K} \inf_{i \in I} \int_{X_1} F_{1,i}^-(-F_{3,k}^-(g))\, d\mu + \inf_{j \in J} \int_{X_2} F_{2,j}^-(g)\, d\nu \leq 0.$$

Proof. By the previous Proposition 3.4.1, we have

$$0 \leq \inf_{\sigma_1 \in \mathcal{P}(Y_1), \sigma_2 \in \mathcal{P}(Y_2)} \{\mathcal{F}_1(\mu, \sigma_1) - \mathcal{F}_3(\sigma_1, \sigma_2) + \mathcal{F}_2(\nu, \sigma_2)\}$$

if and only if

$$0 \leq \inf_{g \in C(X_2)} \{- \sup_{k \in K} \inf_{i \in I} \int_{X_1} F_{1,i}^-(-F_{3,k}^-(g))\, d\mu - \inf_{j \in J} \int_{X_2} F_{2,j}^-(g)\, d\nu\}.$$

□

Corollary 3.4.3. *Specializing Corollary 3.4.2 to the case where $\mathcal{F}_1 = \mathcal{E}_1 \star \mathcal{T}_1$ and $\mathcal{F}_2 = \mathcal{E}_2 \star \mathcal{T}_2$ (and relabeling $\mathcal{F}_3$ to $\mathcal{F}$), where*

- *$\mathcal{E}_1$ is a backward β_1-entropic transfer on $\mathcal{P}(X_1) \times \mathcal{P}(Z_1)$ with operator E_1^-,*
- *$\mathcal{E}_2$ is a backward β_2-entropic transfer on $\mathcal{P}(X_2) \times \mathcal{P}(Z_2)$ with operator E_2^-,*
- *$\mathcal{T}_1$ is a backward linear transfer on $\mathcal{P}(Z_1) \times \mathcal{P}(Y_1)$ with operator T_1^-, and*
- *$\mathcal{T}_2$ is a backward linear transfer on $\mathcal{P}(Z_2) \times \mathcal{P}(Y_2)$ with operator 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 $\sigma_1 \in \mathcal{P}(Y_1), \sigma_2 \in \mathcal{P}(Y_2)$, we have*

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

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

$$\lambda_1 \sup_{k\in K} \beta_1 \left(\int_{X_1} E_1^- \circ T_1^- \circ (-\frac{1}{\lambda_1} F_k^- g)\, d\mu \right) + \lambda_2 \beta_2 \left(\int_{X_2} E_2^- \circ T_2^- (\frac{1}{\lambda_2} g)\, d\nu \right) \leq 0.$$

By applying the above to $\mathcal{E}_i = \mathcal{H}$ the logarithmic entropy (i.e. where $\beta_i(t) = \log(t)$ and operator $E_i^- f = e^f$), we get the following extension of a celebrated result of Maurey [44].

Corollary 3.4.4. *Let $\mathcal{E}_1$ and $\mathcal{E}_2$ in the above Corollary 3.4.3 be equal to the logarithmic entropy $\mathcal{H}$. Then the following are equivalent:*

1. For all $\sigma_1 \in \mathcal{P}(X_1), \sigma_2 \in \mathcal{P}(X_2)$, we have

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

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

$$\sup_{k\in K} \left(\int_{X_1} e^{T_1^- (-\frac{1}{\lambda_1} F_k^- g)}\, d\mu \right)^{\lambda_1} \left(\int_{X_2} e^{T_2^- (\frac{1}{\lambda_2} g)}\, d\nu \right)^{\lambda_2} \leq 1.$$

If $\mathcal{T}_1 = \mathcal{T}_2$ are the identity transfer, then item 1 in the above is equivalent to saying that for all $g \in C(Y_2)$, we have

$$\sup_{k\in K} \left(\int_{X_1} e^{-\frac{1}{\lambda_1} F_k^- g}\, d\mu \right)^{\lambda_1} \left(\int_{X_2} e^{\frac{1}{\lambda_2} g}\, d\nu \right)^{\lambda_2} \leq 1.$$

Chapter 4

Hamiltonian/Lagrangian dynamics: Weak KAM, Aubry-Mather, and optimal transport

4.1 Introduction

In this chapter, we provide an overview of *weak KAM theory* and *Aubry-Mather theory* for a *Hamiltonian/Lagrangian system*, drawing from [15, 19–21, 23, 32, 54, 55]. We shall see that the work of Bernard-Buffoni [5, 6] demonstrates that in fact much of the analysis can be obtained by studying an optimal transport problem with the cost function of Example 2.6.6,

$$c_L(x,y) = \inf\{\int_0^1 L(\gamma(t),\dot\gamma(t))dt\,;\,\gamma\in C^1([0,1];M),\gamma(0)=x,\gamma(1)=y\}.$$

Inspired by the theory presented in this chapter, we shall then endeavour in the next chapter to develop a "weak KAM/Aubry-Mather" theory for general linear transfers. Thus, the main purpose here is to describe the connections with the aim of generalising them. For brevity, we shall not attempt to provide complete proofs of results from the literature, and for pedagogical reasons, describe the results for the flat n-dimensional torus $\mathbb{T}^n$, although a smooth compact connected Riemannian manifold without boundary would suffice (with the appropriate changes).

4.2 Hamiltonian systems

Consider a dynamical system on $\mathbb{T}^n := \mathbb{R}^n/\mathbb{Z}^n$, the n-dimensional flat torus, whose energy is described by a Hamiltonian function $H : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$, $(x, p) \mapsto H(x, p)$, where $\mathbb{R}^n$ is Euclidean space. The dynamics of the system are obtained by solving *Hamilton's system of equations*; a coupled, non-linear, system of ODE's, given (at least in local coordinates) by

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial p}(x, p) \\ \dot{p} = -\frac{\partial H}{\partial x}(x, p), \end{cases} \tag{4.1}$$

together with some initial conditions $x(0) = x_0$ and $p(0) = p_0$.

Solving (4.1) explicitly is not, in general, possible, except in certain special cases. One such case, the case of an *integrable system*, occurs if, in the coordinates (x, p), the Hamiltonian H does not depend on x, i.e. $H = H(p)$. Then the system reduces to

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial p}(p) \\ \dot{p} = 0, \end{cases} \tag{4.2}$$

which is easily solved and the dynamics is completely known: it corresponds to a uniform rotation on $\mathbb{T}^n$ given by

$$x(t) = (x_0 + \rho(p_0)t) \,(\mathrm{mod}\ \mathbb{Z}^n),$$

with *rotation vector* $\rho(p_0) := \frac{\partial H}{\partial p}(p_0)$. In particular, the phase space $\mathbb{T}^n \times \mathbb{R}^n$ is foliated by invariant tori $\mathbb{T}^n \times \{p_0\}$ which describe stable motions of the system, i.e. trajectories starting in an invariant tori will remain there for all time. Different initial conditions of (4.2) will therefore trace out different invariant tori in phase space. In addition, if the rotation vector $\rho(p_0)$ has *irrational* coordinates (resp., rational), then the trajectory $t \mapsto x(t)$ on $\mathbb{T}^n$ is *quasi-periodic* (resp., periodic).

It is then natural to ask about the dynamics of an integrable system if it is slightly perturbed. What happens to the invariant tori and the stable motions? Are they destroyed, or might some still persist?

These questions are related to the search for *a smooth and invertible change of coordinates* $\Phi : (X,P) \mapsto (x,p)$ so that the Hamiltonian system (4.1) reduces to an integrable system in the new coordinates (X,P):

$$\begin{cases} \dot{X} = \frac{\partial \bar{H}}{\partial P}(P) \\ \dot{P} = 0, \end{cases} \tag{4.3}$$

where the *effective Hamiltonian* $\bar{H}(P) := H \circ \Phi(X,P)$ depends only on P.

In general, of course, such a coordinate change does not exist, for the following reason: By comparing (4.1) and (4.3), a necessary condition relating the coordinates, is

$$\frac{\partial X}{\partial x} = \frac{\partial p}{\partial P}.$$

The above condition is satisfied if there exists a *generating function* u so that $X = \frac{\partial u}{\partial p}(x,P)$ and $p = \frac{\partial u}{\partial x}(x,P)$. The relation $\bar{H}(P) = H(x,p)$ then yields an equation for this function u, the (stationary) *Hamilton-Jacobi equation*,

$$H(x, \nabla_x u(x,P)) = \bar{H}(P). \tag{4.4}$$

Thus the strategy becomes: Given $P \in \mathbb{R}^n$, find a constant $\bar{H}(P)$ and a function $u(\cdot,P)$ which satisfies (4.4), and define a change of coordinates $(x,p) \to (X,P)$ to satisfy

$$\begin{aligned} X &= \nabla_P u(x,P) \\ p &= \nabla_x u(x,P). \end{aligned} \tag{4.5}$$

There are two main issues which prevent this procedure from being carried out in general: The first is that the Hamilton-Jacobi equation (4.4) may not admit any smooth solutions u, and second, even if it does, it may not be possible to invert (4.5) globally to solve for P (and X) in terms of x and p.

In the *nearly integrable* case, that is, when H is "close" to some integrable H_0, the theory contributed by Kolmogorov, Arnold, and Moser, collectively known as *KAM theory* (see e.g. [58], [61], and references therein) provides results which say that such a coordinate change does in fact exist. Some of the invariant tori of the unperturbed system are "deformed" and survive the perturbation, while others are destroyed. The surviving tori are those of the unperturbed system which satisfy a "non-resonance condition", made precise in the following (pseudo-) theorem.

Theorem 4.2.1 (KAM theorem, [58], Theorem 2.1)**.** *Let $H(x,p,\epsilon) := H_0(p) + \epsilon H_1(x,p,\epsilon)$ be a real-analytic Hamiltonian, where $(x,p) \in \mathbb{T}^n \times \mathbb{R}^n$ and $\epsilon \in \mathbb{R}$. Suppose that the unperturbed system H_0 has invertible Hessian at a point p_0, where p_0 is such that its corresponding rotation vector $\rho(p_0) = \frac{\partial H_0}{\partial p}(p_0)$ is diophantine, i.e. there exist $c, \gamma > 0$ such that $|\langle k, \rho(p_0)\rangle| \geq \frac{1}{c|k|^\gamma}$ for all non-zero $k \in \mathbb{Z}^n$.*

Then for small ϵ the torus $\mathbb{T}^n \times \{p_0\}$ of the unperturbed system survives and is slightly deformed by the perturbation. It carries trajectories with rotation vector $\rho(p_0)$.

For more general H which are "far" from being integrable, one cannot hope to obtain similar results in this non-perturbative setting; in other words, the Hamiltonian dynamics are not conjugate to a rigid rotation and there may not exist any quasi-periodic solutions. However, although there may not exist classical C^1 solutions to the stationary Hamilton-Jacobi (4.4), Lions, Papanicolaou, and Varadhan [39], have shown the existence of *viscosity* solutions to an equivalent formulation (sometimes called the *cell problem*),

$$H(x, P + \nabla_x w(x,P)) = \bar{H}(P). \tag{4.6}$$

The equivalence of (4.6) and (4.4) is via $u(x,P) = x \cdot P + w(x,P)$.

Theorem 4.2.2 ([39], Theorem 1)**.** *Suppose H is superlinear, i.e.* $\lim_{|p|\to\infty} H(x,p) =$

$+\infty$, uniformly in x. Then for each $P \in \mathbb{R}^n$, there exists a unique $\bar{H}(P)$ for which there exists a viscosity solution w of (4.6). In addition, $P \mapsto \bar{H}(P)$ is continuous.

It is often sufficient to consider in the next sections only the case $P = 0$, the general case for any P obtained by replacing H with $H_P(x,p) := H(x, P+p)$.

Let us recall here the definition of a viscosity solution in this context.

Definition 4.2.3 (Viscosity solution)**.** Let $k \in \mathbb{R}$.

1. $u \in C(\mathbb{T}^n)$ is a **viscosity subsolution** of $H(x, \nabla u(x)) = k$ iff for every $\varphi \in C^1(\mathbb{T}^n)$ such that $\varphi \geq u$ and every $x_0 \in \mathbb{T}^n$ such that $\varphi(x_0) = u(x_0)$, it holds that

$$H(x_0, \nabla\varphi(x_0)) \leq k.$$

2. $u \in C(\mathbb{T}^n)$ is a **viscosity supersolution** of $H(x, \nabla u(x)) = k$ iff for every $\varphi \in C^1(\mathbb{T}^n)$ such that $\varphi \leq u$ and every $x_0 \in \mathbb{T}^n$ such that $\varphi(x_0) = u(x_0)$, it holds that

$$H(x_0, \nabla\varphi(x_0)) \geq k.$$

3. $u \in C(\mathbb{T}^n)$ is a **viscosity solution** iff u is both a visocity subsolution, and a viscosity supersolution.

How is the effective Hamiltonian $\bar{H}(P)$ related to the original system (4.1) when there exist only viscosity solutions? This is the *weak KAM theory* of Fathi [20] that we discuss in the next section. It connects $\bar{H}(P)$ and the associated viscosity solutions, to the theory of invariant sets of *Aubry-Mather* from dynamical systems.

Indeed, the lack of smooth solutions to Hamilton-Jacobi, implies non-existence of invariant tori. But we shall see in the next section that one can still speak of more general sets which are invariant under the Hamiltonian flow; this is Aubry-Mather theory. On the other hand, a result of Hamilton-Jacobi implies that solutions to the Hamilton-Jacobi equation give invariant sets.

Theorem 4.2.4 (Hamilton-Jacobi)**.** *Let $u : \mathbb{T}^n \to \mathbb{R}$ be C^2. Then for every $P \in \mathbb{R}^n$, Graph$(P + \nabla u)$ is invariant under the Hamiltonian flow, if and only if $H(x, P + \nabla u(x)) = c$ for a constant c independent of x.*

The basic premise of weak KAM theory is therefore to build the Aubry-Mather sets from viscosity solutions to the Hamilton-Jacobi equation.

In the sections that follow the appropriate Hamiltonians are the *Tonelli* Hamiltonians; that is, those $H : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying

1. $(x, p) \mapsto H(x, p)$ is C^k smooth for some $k \geq 2$,
2. (Strict Convexity) $p \mapsto H(x, p)$ is strictly convex for every fixed x,
3. (Superlinearity) $\lim_{|p|\to+\infty} \frac{H(x,p)}{|p|} = +\infty$ uniformly in x.

A Tonelli Hamiltonian defines, dually, a *Tonelli Lagrangian* $L : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ via the *Fenchel-Legendre transform* of H:

$$L(x, v) := \sup_{p\in\mathbb{R}^n} \{\langle p, v\rangle - H(x, p)\}. \tag{4.7}$$

The Lagrangian L and Hamiltonian H are completely equivalent in the sense that H can be recovered from L via

$$H(x, p) = \sup_{v\in\mathbb{R}^n} \{\langle p, v\rangle - L(x, v)\},$$

and the statements "H is Tonelli", and "L is Tonelli", are equivalent, where a *Tonelli Lagrangian* is a function $L : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying

1. $(x, v) \mapsto L(x, v)$ is C^k smooth for some $k \geq 2$,
2. (Strict Convexity) $v \mapsto L(x, v)$ is strictly convex for every fixed x,
3. (Superlinearity) $\lim_{|v|\to+\infty} \frac{L(x,v)}{|v|} = +\infty$ uniformly in x.

The transformation $\mathcal{L}(x,v) := (x, \frac{\partial L}{\partial v}(x,v))$ is a global C^{k-1} diffeomorphism between the Lagrangian coordinates (x,v) and the Hamiltonian coordinates (x,p), so that

$$H \circ \mathcal{L}(x,v) = \langle \frac{\partial L}{\partial v}(x,v), v\rangle - L(x,v),$$

with inverse given by $\mathcal{L}^{-1}(x,p) = (x, \frac{\partial H}{\partial p}(x,p))$. The Hamiltonian system (4.1), yields in Lagrangian coordinates, the Euler-Lagrange equations,

$$\begin{cases} \dot{x} = v, \\ \frac{d}{dt}\left(\frac{\partial L}{\partial v}(x,v)\right) = \frac{\partial L}{\partial x}(x,v). \end{cases} \tag{4.8}$$

The Hamiltonian flow ϕ_t^H associated to the Hamiltonian system (4.1) is conjugated to the Lagrangian flow ϕ_t^L, via $\phi_t^H = \mathcal{L} \circ \phi_t^L \circ \mathcal{L}^{-1}$.

4.3 Aubry-Mather theory

4.3.1 Mañé critical value

Definition 4.3.1. The **Mañé critical value**, denoted by c, is the infimum of all $k \in \mathbb{R}$ such that there exists a continuous function $u : \mathbb{T}^n \to \mathbb{R}$ which is a viscosity subsolution to $H(x, \nabla u(x)) = k$.

Note that $c \in \mathbb{R}$, since by superlinearity, there exists $K > 0$ and $C > 0$, such that $L(x,v) \geq K|v| - C$, in which case

$$H(x,0) = \sup_{v \in \mathbb{R}^n} \{-L(x,v)\} \leq \sup_{v \in \mathbb{R}^n} \{-K|v| + C\} = C$$

hence $u \equiv 0$ is a viscosity subsolution for any $k \leq C$. At the same time $c > -\infty$ since for any $u \in C^1(\mathbb{T}^n)$,

$$H(x, \nabla u(x)) = \sup_{v \in \mathbb{R}^n} \{\langle \nabla u(x), v\rangle - L(x,v)\} \geq -L(x,0) \geq -\sup_{x \in \mathbb{T}^n} L(x,0) > -\infty,$$

thus there exists no viscosity subsolution of $H(x, \nabla u(x)) = k$, for $k < -\sup_{x \in \mathbb{T}^n} L(x,0)$. We shall see that in fact $c = \bar{H}(0)$, so the Mañé critical value is exactly the effective Hamiltonian (recall it suffices to consider

$P = 0$). The Legendre transform yields an important variational characterisation of viscosity subsolutions to the Hamilton-Jacobi equation which is recorded in the following proposition.

Proposition 4.3.2. *$u \in C(\mathbb{T}^n)$ is a viscosity subsolution to $H(x, \nabla u(x)) = c$ if and only if for any $a < b \in \mathbb{R}$ and every Lipschitz curve $\gamma : [a, b] \to \mathbb{T}^n$, it holds that*

$$u(\gamma(b)) - u(\gamma(a)) \leq \int_a^b L(\gamma, \dot{\gamma}) ds + c(b - a). \tag{4.9}$$

Note in particular that this implies every viscosity subsolution is Lipschitz by taking a curve γ with constant speed. The next proposition shows that in fact if there is a γ for which there is equality in (4.9), then u is actually a viscosity solution and not just a subsolution.

Proposition 4.3.3. *If u is a viscosity subsolution to $H(x, \nabla u(x)) = c$, and for every $x \in \mathbb{T}^n$, there exists a Lipschitz curve $\gamma : (-\infty, 0] \to \mathbb{T}^n$, with $\gamma(0) = x$, satisfying*

$$u(\gamma(b)) - u(\gamma(a)) = \int_a^b L(\gamma, \dot{\gamma}) ds + c(b - a)$$

for all $a < b \leq 0$, then u is a viscosity solution to $H(x, \nabla u(x)) = c$ (and consequently, $c = \bar{H}(0)$ is the unique constant for which this is true).

The previous two propositions suggest the introduction of the *Lax-Oleinik semi-group*

$$S_t^- u(x) := \inf_{\gamma} \{u(\gamma(-t)) + \int_{-t}^0 L(\gamma, \dot{\gamma}) ds\}, \quad t > 0, \tag{4.10}$$

where the infimum is over all Lipschitz curves γ such that $\gamma(0) = x$, and consider those functions u for which

$$S_t^- u(x) + ct = u(x).$$

Definition 4.3.4. A function $u \in C(\mathbb{T}^n)$ is a **negative weak KAM solution** if $S_t^- u(x) + ct = u(x)$ for all $x \in \mathbb{T}^n$ and all $t > 0$.

The term "negative" refers to the fact that the Lipschitz curve in Proposition 4.3.3 is defined for all negative time. In view of Proposition 4.3.3, negative weak KAM solutions are in fact exactly the viscosity solutions for the Hamilton-Jacobi equation.

Proposition 4.3.5. *$u \in C(\mathbb{T}^n)$ is a negative weak KAM solution if and only if it is a viscosity solution to $H(x, \nabla u(x)) = c$.*

Remark 4.3.6. In a similar way to above, we may introduce another Lax-Oleinik semi-group (see (4.9) above),

$$S_t^+ u(x) := \sup_\gamma \{u(\gamma(t)) - \int_0^t L(\gamma, \dot\gamma) ds\}, \quad t > 0,$$

where the supremum is over all Lipschitz curves $\gamma : [0, t] \to \mathbb{R}$ such that $\gamma(0) = x$, and define the *positive weak KAM solutions* as those u satisfying

$$S_t^+ u(x) - ct = u(x) \quad \text{for all } x \in \mathbb{T}^n \text{ and all } t > 0.$$

The positive weak KAM solutions are also viscosity solutions to $H(x, \nabla u(x)) = c$. We shall see the connection between positive/negative weak KAM solutions in the next sections.

4.3.2 Peierls Barrier and the Aubry set

In this section, we will construct an invariant set under the Hamiltonian flow, the *Aubry set*, via the *Peierls Barrier*.

Related to the Lax-Oleinik semi-group (see (4.10)),
consider the *minimal action*,

$$h_t(x, y) := \inf_\gamma \left\{ \int_0^t L(\gamma, \dot\gamma) ds \,;\, \gamma(0) = x, \gamma(t) = y \right\} \tag{4.11}$$

where the infimum is over all Lipschitz curves $\gamma : [0, t] \to \mathbb{T}^n$ with the specified end values. It is known that a minimiser of (4.11) exists, and that it is as regular as L itself; this is Tonelli's theorem. Moreover, by the Calculus of Variations, a minimiser of (4.11) satisfies the Euler-Lagrange equation (4.8), which is nothing but the first variation of the action (4.11).

Definition 4.3.7. The **Peierls Barrier** $h_\infty(x, y)$ is defined via

$$h_\infty(x, y) = \liminf_{t \to +\infty} (h_t(x, y) + ct),$$

where $c = \bar{H}(0)$ is Mañé's constant of the previous section.

Proposition 4.3.8. *The Peierls Barrier is finite for all $x, y \in \mathbb{T}^n$, Lipschitz, and satisfies the following properties:*

1. *Every viscosity subsolution u satisfies $u(x) - u(y) \leq h_\infty(x, y)$,*
2. $h_\infty(x, z) \leq h_\infty(x, y) + h_\infty(y, z)$,
3. $h_\infty(x, y) + h_\infty(y, x) \geq 0$.

The Peierls Barrier is important for a number of reasons, the first being that it provides negative weak KAM solutions.

Proposition 4.3.9. *For every fixed $x \in \mathbb{T}^n$, the map $y \mapsto h_\infty(x, y)$ is a negative weak KAM solution.*

The collection of points where the Peierls Barrier vanishes is particularly important as we will see next.

Definition 4.3.10. The **projected Aubry set** is defined as the collection of points where the Peierls barrier vanishes:

$$\mathcal{A} := \{x \in \mathbb{T}^n \, ; \, h_\infty(x, x) = 0\}.$$

Proposition 4.3.11. *The projected Aubry set $\mathcal{A}$ is compact and non-empty.*

Proposition 4.3.12. *The projected Aubry set satisfies*

$$\mathcal{A} = \bigcap_u (u - S_1^- u)^{-1}(c)$$

where the intersection is over all viscosity subsolutions u to $H(x, \nabla u(x)) = c$.

The projected Aubry set is distinguished for a number of reasons. The first we shall note is that it a set of "uniqueness" for weak KAM solutions.

Proposition 4.3.13. *1. If u is a negative weak KAM solution, then there exists a unique positive weak KAM solution v such that $u = v$ on $\mathcal{A}$.*

2. If u is a viscosity subsolution of $H(x, \nabla u(x)) = c$, then there exists a negative weak KAM solution v such that $u = v$ on $\mathcal{A}$.

3. If u_1, u_2 are negative weak KAM solutions, and $u_1 \leq u_2$ on $\mathcal{A}$, then $u_1 \leq u_2$ everywhere.

The second we shall observe is that it is exactly the differentiability set for viscosity subsolutions of the Hamilton-Jacobi equation.

Proposition 4.3.14. *For any $x \in \mathcal{A}$, there exists a C^2 curve $\gamma_x : \mathbb{R} \to \mathcal{A}$ with $\gamma_x(0) = x$, that solves the Euler-Lagrange equation,*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v}(\gamma_x, \dot{\gamma_x})\right) = \frac{\partial L}{\partial x}(\gamma_x, \dot{\gamma_x}), \quad \text{for all } t \in \mathbb{R}$$

for which any viscosity subsolution u satisfies

$$u(\gamma_x(b)) - u(\gamma_x(a)) = \int_a^b L(\gamma_x, \dot{\gamma_x})ds + c(b-a), \quad a < b \in \mathbb{R}. \tag{4.12}$$

In addition, u is differentiable at x with $\nabla u(x) = \frac{\partial L}{\partial v}(x, \dot{\gamma_x}(0))$ and satisfies $H(x, \nabla u(x)) = c$ at that point.

The curve γ is unique in the sense that if $\tilde{\gamma_x} : [a,b] \to \mathbb{T}^n$ is defined on an interval $[a,b]$ containing 0, satisfies $\tilde{\gamma_x}(0) = x$ and (4.12), then $\gamma_x = \tilde{\gamma_x}$ on $[a,b]$.

The above proposition implies that the map $x \mapsto \frac{\partial L}{\partial v}(x, \dot{\gamma_x}(0))$ is well-defined on $\mathcal{A}$. We define the graph of this function over $\mathcal{A}$ to be the *Aubry set*.

Definition 4.3.15. The **Aubry set** $\tilde{A}$ is the subset of $\mathbb{T}^n \times \mathbb{R}^n$ defined via

$$\tilde{\mathcal{A}} := \{(x, \frac{\partial L}{\partial v}(x, \dot{\gamma_x}(0)))\,;\, x \in \mathcal{A}\}$$

where γ is as in Proposition 4.3.14.

Proposition 4.3.16. *The Aubry set is non-empty, compact, and invariant under the Hamiltonian flow. It is a Lipschitz graph over the projected Aubry set.*

4.3.3 Mather measures

Recall the minimal action,

$$h_t(x, y) := \inf_{\gamma}\{\int_0^t L(\gamma, \dot{\gamma})ds\}$$

where the infimum is over Lipschitz curves $\gamma : [0, t] \to \mathbb{T}^n$ with $\gamma(0) = x$ and $\gamma(t) = y$.

In the previous section, the Peierls Barrier was defined as a long-time limit of the minimal action, and the Aubry set was constructed. However, one alternative to studying minimising trajectories of (4.11), is instead look at the *average action* of a "collection of orbits", in the sense of a Lagrangian action of a *flow-invariant probability distribution*:

$$A_L[\mu] := \int_{\mathbb{T}^n \times \mathbb{R}^n} L(x, v)d\mu(x, v),$$

and introduce the analogous minimisation problem,

$$\inf\{A_L[\mu]\,;\, \mu \in \mathcal{D}\},$$

where $\mathcal{D} := \left\{\mu \in \mathcal{P}(\mathbb{T}^n \times \mathbb{R}^n)\,;\, (\phi_t^L)_{\#}\mu = \mu\right\}$ is the set of probability measures invariant under the Euler-Lagrange flow determined by L.

Since the constraint set $\mathcal{D}$ depends on L itself, Mañé introduced a larger set of probability measures, the so-called set of *holonomic measures* $\mathcal{F}$, which contains $\mathcal{D}$, and is defined by

$$\mathcal{F} := \left\{ \mu \in \mathcal{P}(\mathbb{T}^n \times \mathbb{R}^n) \,;\, \int_{\mathbb{T}^n \times \mathbb{R}^n} v \cdot \nabla\varphi(x) d\mu(x,v) = 0,\, \forall \varphi \in C^1(\mathbb{T}^n) \right\}.$$

The corresponding minimisation problem,

$$\inf \{A_L[\mu] \,;\, \mu \in \mathcal{F}\} \tag{4.13}$$

is a relaxation of (4.11).

How do minimisers of (4.13) relate to trajectories of (4.1)? Consider the integrable Hamiltonian system (4.2). In this case, $L(x,v) = L(v)$, and by Fenchel-Legendre (4.7), we have for any $p_0 \in \mathbb{R}^n$,

$$L(v) \geq \langle p_0, v \rangle - H(p_0)$$

with equality when $v = \frac{\partial H}{\partial p}(p_0) = \rho(p_0)$. Integrating with respect to $\mu \in \mathcal{F}$, we obtain

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} L(v) d\mu(x,v) \geq \langle p_0, \int_{\mathbb{T}^n \times \mathbb{R}^n} v d\mu(x,v) \rangle - H(p_0) \tag{4.14}$$

which can be rearranged to yield the inequality

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} [L(v) - \langle p_0, v \rangle] d\mu(x,v) \geq -H(p_0). \tag{4.15}$$

This linear shift of the Lagrangian, $L_{p_0}(v) := L(v) - \langle p_0, v \rangle$, is again a Tonelli Lagrangian with the same Euler-Lagrange equation as L.

We recall the invariant tori $\mathbb{T}^n \times \{p_0\}$ for the integrable Hamiltonian system. These translate via the Fenchel-Legendre transform (4.7) into invariant tori $\mathbb{T}^n \times \{\rho(p_0)\}$ for the corresponding Euler-Lagrange flow. If $\mu \in \mathcal{F}$ is now supported on $\mathbb{T}^n \times \{\rho(p_0)\}$, then the left-hand side of (4.15) becomes

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} [L(\rho(p_0)) - \langle p_0, \rho(p_0) \rangle] d\mu(x,v) = -H(p_0).$$

The conclusion is the following: Any invariant measure $\mu \in \mathcal{F}$ supported on

$\mathbb{T}^n \times \{\rho(p_0)\}$ minimises the Lagrangian L_{p_0} among all measures $\mu \in \mathcal{F}$, i.e.,

$$\inf\{A_{L_{p_0}}[\mu]\,;\,\mu \in \mathcal{F}\} = -H(p_0), \tag{4.16}$$

and, in addition,

$$\mathbb{T}^n \times \{\rho(p_0)\} = \bigcup_{\{\mu\in\mathcal{F}\,;\,\mu \text{ minimises } A_{L_{p_0}}\}} \operatorname{spt}(\mu) \tag{4.17}$$

The action-minimising measures μ thus provide a characterisation for the invariant tori in the case of an integrable system.

Still considering for the moment the case of an integrable system, define an *average rotation vector* for μ, by $\rho(\mu) := \int_{\mathbb{T}^n\times\mathbb{R}^n} v d\mu(x,v) \in \mathbb{R}^n$, which can be interpreted as a kind of average over the rotation vectors of a collection of orbits. From (4.14), we have for measures $\mu \in \mathcal{F}$ with $\rho(\mu) = \rho(p_0)$,

$$\begin{aligned}\int_{\mathbb{T}^n\times\mathbb{R}^n} L(v)d\mu(x,v) &\geq \langle p_0, \rho(p_0)\rangle - H(p_0)\\ &= L(\rho(p_0)),\end{aligned}$$

with equality if μ is supported in $\mathbb{T}^n \times \{\rho(p_0)\}$. Therefore,

$$\inf\{A_L[\mu]\,;\,\mu \in \mathcal{F},\ \rho(\mu) = \rho(p_0)\} = L(\rho(p_0)), \tag{4.18}$$

and moreover,

$$\mathbb{T}^n \times \{\rho(p_0)\} = \bigcup_{\{\mu\in\mathcal{F}\,;\,\mu \text{ minimises } A_L \text{ and } \rho(\mu)=\rho(p_0)\}} \operatorname{spt}(\mu). \tag{4.19}$$

Motivated by the integrable case (4.16), for general Hamiltonian systems the so-called *Mather's α-function* is then defined via

$$\alpha(c) := -\inf\{A_{L_c}[\mu]\,;\,\mu \in \mathcal{F}\}, \quad c \in \mathbb{R}^n, \tag{4.20}$$

where we recall that $L_c(x,v) := L(x,v) - \langle v, c\rangle$. In the case of an integrable system, Mather's α-function is exactly the Hamiltonian (4.16); hence α is also known as the *effective Hamiltonian* in the general case. In fact, we shall see in Section 4.4 that a result of weak KAM says in fact that $\alpha(P) = \bar{H}(P)$ where we recall $\bar{H}(P)$ is the effective Hamiltonian of Lions, Papanicolaou, and Varadhan.

Let $\mathcal{M}_c$ denote the measures $\mu \in \mathcal{F}$ whose action $A_{L_c}[\mu]$ realise the minimal value $-\alpha(c)$. As motivated by (4.17), define the *Mather set* $\tilde{M}_c$ corresponding to c as the union of the support of all measures in $\mathcal{M}_c$:

$$\tilde{M}_c := \bigcup_{\mu \in \mathcal{M}_c} \operatorname{spt}(\mu) \quad \subset \mathbb{T}^n \times \mathbb{R}^n \tag{4.21}$$

and the *projected Mather set* M_c as the projection of $\tilde{M}_c$ onto the base $\mathbb{T}^n$:

$$M_c := \pi(\tilde{M}_c) \quad \subset \mathbb{T}^n.$$

Theorem 4.3.17 (Mather)**.** *$\tilde{M}_c$ is a compact, non-empty, and invariant set under the Euler-Lagrange flow. Moreover, it is a graph over M_c.*

Analogous to Mather's α-function, and again motivated by the integrable case (4.18), *Mather's β-function* is defined via,

$$\beta(h) := \inf\{A_L[\mu]\,;\, \mu \in \mathcal{F},\, \rho(\mu) = h\}, \quad h \in \mathbb{R}^n, \tag{4.22}$$

where $\rho(\mu)$ is the rotation vector of μ. Since β coincides with the Lagrangian when the system is integrable, it is also termed the *effective Lagrangian*. The terminology is suggestive that β should be the Fenchel-Legendre transform of the effective Hamiltonian, i.e., that $\beta = \alpha^*$ (see Theorem 4.3.18 below).

Similarly to Mather's α-function, let $\mathcal{M}^h$ denote the measures $\mu \in \mathcal{F}$ whose rotation vector $\rho(\mu)$ is equal to h, and whose action $A_L[\mu]$ realises the minimum $\beta(h)$ in (4.22). Define the Mather set $\tilde{M}^h$ as the union of the support of all measures in $\mathcal{M}^h$:

$$\tilde{M}^h := \bigcup_{\mu \in \mathcal{M}^h} \operatorname{spt}(\mu) \quad \subset \mathbb{T}^n \times \mathbb{R}^n,$$

and its projection M^h as the projection of $\tilde{M}^h$ onto the base $\mathbb{T}^n$:

$$M^h := \pi(\tilde{M}^h) \quad \subset \mathbb{T}^n.$$

Mather again proved that $\tilde{M}^h$ is a compact, non-empty, and invariant set under the Euler-Lagrange flow. Moreover, it is a graph over M^h.

Theorem 4.3.18. *Mather's functions α and β are convex conjugates of each other, i.e. $\alpha = \beta^*$ and $\beta^* = \alpha$.*

In addition, if $\mu \in \mathcal{F}$ is an invariant probability measure, then

$$\beta(\rho(\mu)) = A_L[\mu] \quad \Longleftrightarrow \quad \textit{there exists } c \in \mathbb{R}^n \textit{ such that } -\alpha(c) = A_{L_c}[\mu].$$

Furthermore, if $\beta(\rho(\mu)) = A_L[\mu]$, then

$$-\alpha(c) = A_{L_c}[\mu] \quad \Longleftrightarrow \quad c \in \partial\beta(\rho(\mu)).$$

These results imply that the two collections of minimising measures are the same,

$$\bigcup_{c\in\mathbb{R}^n} \mathcal{M}_c = \bigcup_{h\in\mathbb{R}^n} \mathcal{M}^h,$$

and moreover, the Mather sets are related via

$$\tilde{M}_c = \bigcup_{h\in\partial\alpha(c)} \tilde{M}^h.$$

Proposition 4.3.19. *The Mather set is contained in the Aubry set, i.e., $\tilde{M}_0 \subset \tilde{A}$, where $\tilde{M}_0$ is given by (4.21).*

Proposition 4.3.20. *A probability measure μ belongs to $\mathcal{M}_0$ if and only if its support is contained in the Aubry set $\tilde{\mathcal{A}}$.*

4.4 Weak KAM theory

As mentioned in the introduction, the goal of weak KAM theory is to connect viscosity solutions of the stationary Hamilton-Jacobi equation to Aubry-Mather; specifically the Mather measures, Mather set, and Aubry set.

The Aubry set was defined above via the Peierls Barrier. Here we show that the Aubry set can in fact be constructed from the negative weak KAM solutions.

First, we have the following connection between the various functions.

Proposition 4.4.1. *Mather's α-function, the effective Hamiltonian $\bar{H}$ of Lions, Papanicolaou, and Varadhan, and Mañé's critical constant c, are the same, i.e. $\alpha(0) = \bar{H}(0) = c$.*

Proposition 4.4.2. *For every negative weak KAM solution u, the set*

$$Graph(\nabla u) := \{(x, \nabla u(x))\,;\, \nabla u(x) \text{ exists }\},$$

under backwards Hamiltonian flow satisfies,

$$\phi^H_{-t}(\overline{Graph(\nabla u)}) \subset Graph(\nabla u) \quad \text{for all } t \geq 0,$$

and therefore

$$\tilde{\mathcal{I}}(u) := \bigcap_{t \geq 0} \phi^H_{-t}(\overline{Graph(\nabla u)})$$

is invariant under the Hamiltonian flow. It is non-empty and compact, and known as the Aubry set associated to u.

Proposition 4.4.3. *The Aubry set $\tilde{\mathcal{A}}$ as defined in Section 4.3.2 is equal to the intersection over all $\tilde{\mathcal{I}}(u)$ for all negative weak KAM solutions u,*

$$\tilde{\mathcal{A}} = \bigcap_u \tilde{\mathcal{I}}(u).$$

4.5 Connections of weak KAM and Aubry-Mather to optimal transport

Recall the minimal action $h_t(x, y)$ between two points x and y defined in Section 4.3.2,

$$h_t(x, y) := \inf_\gamma \left\{ \int_0^t L(\gamma, \dot{\gamma})ds\,;\, \gamma(0) = x, \gamma(t) = y \right\} \tag{4.23}$$

where the infimum is over Lipschitz curves $\gamma : [0,t] \to \mathbb{T}^n$. Bernard and Buffoni [5, 6] noted that a lot of the analysis in the previous sections does not really depend on the explicit form of $h_t(x,y)$.

To emphasise this, they develop an abstract theory based on a continuous function $A : X \times X \to \mathbb{R}$ on a compact connected metric space X, of which h_t for $t = 1$ in (4.23) is an example. They build an analogue of the Peierls Barrier h_∞ of the previous sections via inf-convolution of A with itself:

$$\begin{aligned} A_n(x,y) &= A \star A \star \ldots \star A(x,y) \\ &= \inf\{A(x,x_1) + A(x_2,x_3) + \ldots + A(x_{n-1},y)\,;\, x_i \in X\}. \end{aligned}$$

The cost $A_n(x,y)$ between two points x and y, does not deviate too far from the long time average cost

$$c := \lim_{n\to\infty} \frac{\inf\{A_n(x,y)\,;\, x,y \in X\}}{n} = \lim_{n\to\infty} \frac{\sup\{A_n(x,y)\,;\, x,y \in X\}}{n} \tag{4.24}$$

in a uniform way, in the sense that $|A_n(x,y) - nc| \leq C$ for a constant C independent of x,y, and n.

Lemma 4.5.1 ([6], Lemma 9)**.** *The functions A_n are equi-continuous, and there exists a constant $C > 0$, such that*

$$|A_n(x,y) - nc| \leq C \quad \textit{for all } x,y \in X, \textit{ and } n \in \mathbb{N}.$$

Similar to Section 4.3.2, one can define a notion of the Peierls Barrier in this setting.

Definition 4.5.2. The **Peierls Barrier** is the function

$$A_\infty(x,y) := \liminf_{n\to\infty}(A_n(x,y) - nc).$$

The function A_∞ is continuous and real-valued as the family $(A_n)_{n\in\mathbb{N}}$ is equi-continuous and $A_n - cn$ is uniformly bounded in x, y, and n. The Peierls Barrier A_∞ has the same properties as h_∞ (cf. Proposition 4.3.8), which are recorded in the following.

Proposition 4.5.3 ([6], Lemma 11)**.** *The function A_∞ satisfies*

$$A_\infty(x,z) \le A_\infty(x,y) + A_\infty(y,z) \quad \text{for all } x,y,z \in X,$$

and

$$A_\infty(x,z) = \inf_{y\in\mathcal{A}}\{A_\infty(x,y) + A_\infty(y,z)\}$$

where $\mathcal{A} := \{x\,;\, A_\infty(x,x) = 0\}$ is non-empty.

Definition 4.5.4. The **projected Aubry set** is the set $\mathcal{A} = \{x\,;\, A_\infty(x,x) = 0\}$.

Recall from Proposition 4.3.8 that for any viscosity subsolution u, it held that $u(y) - u(x) \le h_\infty(x,y)$, i.e., u is h_∞-*Lipschitz*. In the abstract setting for A_∞, functions u which are A_∞-Lipschitz are exactly the admissible functions for the Kantorovich dual problem corresponding to an optimal transport with cost A_∞ (see Section 2.6).

Definition 4.5.5. Let $\phi_0, \phi_1 \in C(X)$. We say (ϕ_0, ϕ_1) is a **conjugate pair** (for A_∞) if

$$\phi_1(y) = T^-_{A_\infty}\phi_0(y) \quad \text{and} \quad \phi_0(x) = T^+_{A_\infty}\phi_1(x)$$

where

$$T^-_{A_\infty}\phi_0(x) := \sup_{y\in X}\{\phi_0(y) - A_\infty(x,y)\} \quad \text{and} \quad T^+_{A_\infty}\phi_1(y) := \inf_{x\in X}\{\phi_1(x) + A_\infty(x,y)\}.$$

We have the following proposition.

Proposition 4.5.6 ([6], Proposition 8)**.** *If (ϕ_0, ϕ_1) are a conjugate pair for A_∞, then there exists a A_∞-Lipschitz function $\phi : \mathcal{A} \to \mathbb{R}$ such that*

$$\phi_0 = \phi_1 = \phi \quad \text{on } \mathcal{A},$$

and ϕ_0, ϕ_1 can recovered from ϕ via

$$\phi_1(y) = \inf_{x\in\mathcal{A}}\{\phi(x) + A_\infty(x,y)\} \quad \text{and} \quad \phi_0(x) = \sup_{y\in\mathcal{A}}\{\phi(y) - A_\infty(x,y)\}. \tag{4.25}$$

Conversely, for every A_∞-Lipschitz function φ on $\mathcal{A}$, the functions ϕ_0, ϕ_1 defined by (4.25) are a conjugate pair for A_∞.

Remark 4.5.7. The above proposition is stated for A_∞, but in fact it is true for any $c \in C(X \times X)$ which satisfies the properties listed in Proposition 4.5.3.

As was the case in the previous sections, the projected Aubry set is particularly important.

Theorem 4.5.8 ([6], Theorem 12)**.** *(ϕ_0, ϕ_1) are a conjugate pair for A_∞ if and only if, $T_A^-\phi_0 + c = \phi_0$, $T_A^+\phi_1 - c = \phi_1$, and $\phi_0 = \phi_1$ on $\mathcal{A}$, where*

$$T_A^-\phi_0(x) := \sup_{y\in X}\{\phi_0(y) - A(x,y)\} \quad \text{and} \quad T_A^+\phi_1(y) := \inf_{x\in X}\{\phi_1(x) + A(x,y)\}$$

and c is defined as in (4.24).

Remark 4.5.9. This theorem is the abstract analogue of Proposition 4.3.13, item 1.

Recall from Section 2.6 that conjugate pairs are connected to an optimal transport via

$$\inf_{\pi\in\mathcal{K}(\mu,\nu)} \int_{X\times Y} A_\infty(x,y)\, d\pi(x,y) = \sup\{\int_Y \phi_1 d\nu - \int_X \phi_0 d\mu\}$$

where the supremum is over (ϕ_0, ϕ_1) which are are conjugate pairs for A_∞. Therefore the interpretation of this theorem is that taking the Peierls Barrier A_∞ as the cost function for an optimal transport, the admissible pairs for the Kantorovich duality are exactly the negative/positive weak KAM solutions which agree on the projected Aubry set. Indeed, those functions which satisfy

$$T_A^- \phi_0 + c = \phi_0 \quad \text{and} \quad T_A^+ \phi_1 - c = \phi_1$$

are the analogues of Fathi's positive (resp., negative) weak KAM solutions in this abstract setting (note here that one should think of T_A^- as corresponding to S_1^+, while T_A^+ as corresponding to S_1^-, where $S_1^{\pm}$ are the Lax-Oleinik operators of the previous section; the apparent discrepancy in $+$, $-$, notation is chosen to be consistent with backward/forward Kantorovich operators for the next chapter) since the above relations imply

$$T_{A,n}^- \phi_0 + nc = \phi_0 \quad \text{and} \quad T_{A,n}^+ \phi_1 - nc = \phi_1, \quad \text{for all } n \in \mathbb{N}$$

where $T_{A,n}^{\pm}$ denotes the n-fold composition of $T_A^{\pm}$. The main distinction is that the "time" index here is integer-valued (compare with Definition 4.3.4).

Finally the Mather measures are interpreted in this setting.

Theorem 4.5.10 ([6], Theorem 13)**.** *Denote the optimal transport with cost function A by*

$$\mathcal{T}_A(\mu, \nu) := \inf_{\pi \in \mathcal{K}(\mu,\nu)} \int_{X \times X} A(x,y) d\pi(x,y).$$

Then

$$c = \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}_A(\mu, \mu),$$

and a measure $\pi \in \mathcal{P}(X \times X)$ satisfies $\int_{X \times X} A(x,y) d\pi(x,y) = \mathcal{T}_A(\mu,\mu) = c$ if and only if π is supported on the set

$$\mathcal{D} := \{(x,y)\,;\, A(x,y) + A_\infty(y,x) = c\}.$$

Remark 4.5.11. In the setting of the previous sections when $X = \mathbb{T}^n$, there is a bijection between the Mather measures $\mathcal{M}_0$ and the measures π, which is given by the mapping $(\mathrm{proj}_{\mathbb{T}^n}, \mathrm{proj}_{\mathbb{T}^n} \circ \varphi_1^L)_\#$, where $\mathrm{proj}_{\mathbb{T}^n}$ is the projection $\mathbb{T}^n \times \mathbb{R}^n \to \mathbb{T}^n$, and φ_1^L is the time 1 Lagrangian flow.

Chapter 5

Weak KAM and Aubry-Mather for linear transfers

In this chapter, our aim is the development of a "weak-KAM/Aubry-Mather" theory for a linear transfer in an analogy to the previous chapter. In particular, for a backward linear transfer $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ with backward Kantorovich operator $T^- : C(X) \to USC(X)$, it is natural in view of the last chapter to define *backward weak KAM solutions* as those functions $g \in USC(X)$ satisfying $T^-g(x)+c = g(x)$ for a particular constant c (essentially the fixed points for T^- up to an additive constant). An immediate technical issue is that T^-g is not strictly defined when $g \in USC(X)$; we therefore discuss its extension from $C(X)$ in the next section, Section 5.1. The backward weak KAM solutions can be viewed as a generalisation of Fathi's weak KAM solutions from the previous chapter, and indeed reduce to them when the linear transfer $\mathcal{T}$ is optimal transport generated from a Lagrangian. If $\mathcal{T}$ is also a forward linear transfer with forward Kantorovich operator T^+, one can speak of *forward weak KAM solutions* $T^+f(x) - c = f(x)$, but since a backward linear transfer is not, in general, also a forward linear transfer, we shall focus mainly on backward weak KAM solutions.

We shall see that for many backward linear transfers, we can construct an idempotent backward Kantorovich operator T_∞^-, *which maps* $C(X)$ *into the set of backward weak KAM solutions for* T^-. We can also associate to T_∞^- an idempotent backward linear transfer $\mathcal{T}_\infty$, which in the case of optimal transport generated from a Lagrangian, is related to the Peierls barrier from the previous chapter.

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

We recall that $USC(X)$ is the set of all extended real-valued upper semi-continuous functions $f : X \to \mathbb{R} \cup \{-\infty\}$, and $USC_\sigma(X)$ the closure of $USC(X)$ with respect to monotone increasing limits (i.e. $f \in USC_\sigma(X)$ if and only if there exists a monotone increasing sequence $f_n \in USC(X)$ with $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in X$). We also denote $USC_b(X)$ as those $USC(X)$ functions which are bounded below (and therefore bounded above and below by compactness of X). We shall also denote by $USC_{\sigma,b}(X)$ those functions in $USC_\sigma(X)$ which are bounded above *and* below. The same conventions also hold for the space Y in place of X.

As mentioned in the introduction of this chapter, we wish to find backward weak KAM solutions that are achieved by iterating T^- with itself n times in the limit as $n \to \infty$. Thus it is necessary to extend the domain of the Kantorovich operator T^-, which as defined is only on $C(X)$, to these larger classes of functions. We note that the extension of T^- to $USC_b(X)$ coincides with the discussion of Section 4.2 in the independent work [1].

Lemma 5.1.1. *Let* $\mathcal{T}$ *be a backward linear transfer such that* $\{\delta_x; x \in X\} \subset \mathcal{D}_1(\mathcal{T})$*, and let* $T^- : C(Y) \to USC(X)$ *be the associated backward Kantorovich operator.*

1. *For* $g \in USC(Y)$*, define*

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

Then T^- maps $USC(Y)$ into $USC(X)$, and has representation

$$T^-g(x) = \sup\{\int_Y g d\nu - \mathcal{T}(\delta_x, \nu);\, \nu \in \mathcal{P}(Y)\}. \tag{5.1}$$

Moreover, T^- maps $USC_b(Y)$ into $USC_b(X)$ if and only if

$$\sup_{x\in X} \inf_{\nu\in\mathcal{P}(Y)} \mathcal{T}(\delta_x, \nu) < +\infty.$$

2. *For $g \in USC_\sigma(Y)$, define*

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

where T^-h is defined as in 1. Then T^- maps $USC_\sigma(Y)$ to $USC_\sigma(X)$. If, in addition, $g \in USC_\sigma(Y)$ is bounded above, then (5.1) also holds.

T^- therefore extends to an operator from $USC_\sigma(Y) \to USC_\sigma(X)$ satisfying the monotonicity, convexity, and affine constant properties of Definition 2.4.1, only on $USC_\sigma(Y)$ rather than $C(Y)$.

Proof. 1. The fact that T^-g belongs to $USC(X)$ follows from the fact that an infimum of a family of upper semi-continuous functions, is also upper semi-continuous.

Suppose now that $g \in USC(Y)$. Let $h_n \searrow g$ be a decreasing sequence of continuous functions converging to g. Then,

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

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

By weak* compactness of $\mathcal{P}(Y)$, extract an increasing subsequence n_k so that $\sigma_{n_k} \to \overline{\sigma}$. Then for any $j \le k$, $T^-g(x) \le \int_Y h_{n_j} d\sigma_{n_k} - \mathcal{T}(\delta_x, \sigma_{n_k})$ where we have used the fact that $h_{n_k} \le h_{n_j}$ whenever $j \le k$. For this fixed j, we have that $h_{n_j} \in C(Y)$ and so $\int h_{n_j} d\sigma_{n_k} \to \int h_{n_j} d\overline{\sigma}$ as $k \to \infty$. Hence we obtain

$$T^-g(x) \le \lim_{k\to\infty}\int_Y h_{n_j}d\sigma_{n_k} - \liminf_{k\to\infty}\mathcal{T}(\delta_x,\sigma_{n_k}) \le \int_Y h_{n_j}d\overline{\sigma} - \mathcal{T}(\delta_x,\overline{\sigma}).$$

Finally $\int_Y h_{n_j}d\overline{\sigma} \to \int_Y gd\overline{\sigma}$ by monotone convergence, and we obtain

$$T^-g(x) \le \sup_{\sigma\in\mathcal{P}(Y)}\{\int_Y gd\sigma - \mathcal{T}(\delta_x,\sigma)\}.$$

On the other hand, for any $h \in C(Y)$, $h \ge g$,

$$\sup_{\sigma\in\mathcal{P}(Y)}\{\int_Y gd\sigma - \mathcal{T}(\delta_x,\sigma)\} \le \sup_{\sigma\in\mathcal{P}(Y)}\{\int_Y hd\sigma - \mathcal{T}(\delta_x,\sigma)\} = T^-h(x).$$

Therefore we obtain the reverse inequality

$$\sup_{\sigma\in\mathcal{P}(Y)}\{\int_Y gd\sigma - \mathcal{T}(\delta_x,\sigma)\} \le \inf\{T^-h(x)\,;\, h\in C(Y), h\ge g\} = T^-g(x).$$

Suppose now that $T^-g \in USC_b(X)$. Then using (5.1), we have

$$\begin{aligned} -\infty < \inf_{x\in X} T^-g(x) &= \inf_{x\in X}\sup_{\nu\in\mathcal{P}(Y)}\{\int_Y gd\nu - \mathcal{T}(\delta_x,\nu)\} \\ &\le \sup_{y\in Y} g(y) - \sup_{x\in X}\inf_{\nu\in\mathcal{P}(Y)}\mathcal{T}(\delta_x,\nu) \end{aligned}$$

which implies $\sup_{x\in X}\inf_{\nu\in\mathcal{P}(Y)}\mathcal{T}(\delta_x,\nu) < +\infty$.

On the other hand, if $\sup_{x\in X}\inf_{\nu\in\mathcal{P}(Y)}\mathcal{T}(\delta_x,\nu) < +\infty$, then we easily see using (5.1) again that

$$\inf_{x\in X} T^-g(x) \ge \inf_{y\in Y} g(y) - \sup_{x\in X}\inf_{\nu\in\mathcal{P}(Y)}\mathcal{T}(\delta_x,\nu) > -\infty$$

from which we deduce that $T^-g \in USC_b(X)$.

Note that T^-g is bounded above since $T^-g(x) \le \sup_{y\in Y} g(y) - m_{\mathcal{T}}$, where $m_{\mathcal{T}}$ is a lower bound for $\mathcal{T}$.

2. For $g \in USC_\sigma(Y)$, we have $T^-g \in USC_\sigma(X)$ as it is by definition the supremum of a family of upper semi-continuous functions. Now suppose g

is, in addition, bounded above. Then the expression

$$\int_Y g d\sigma - \mathcal{T}(\delta_x, \sigma)$$

is well-defined for all $\sigma \in \mathcal{P}(Y)$ and takes values in $\mathbb{R} \cup \{-\infty\}$. Use now the first part to write for any $h \in USC(Y)$, $h \le g$,

$$\sup_{\sigma \in \mathcal{P}(Y)} \{\int_Y g d\sigma - \mathcal{T}(\delta_x, \sigma)\} \ge \sup_{\sigma \in \mathcal{P}(Y)} \{\int_Y h d\sigma - \mathcal{T}(\delta_x, \sigma)\} = T^- h(x)$$

which implies by definition of $T^- g$ that

$$\sup_{\sigma \in \mathcal{P}(Y)} \{\int_Y g d\sigma - \mathcal{T}(\delta_x, \sigma)\} \ge T^- g(x). \tag{5.2}$$

On the other hand, for an increasing $h_n \nearrow g$, $h_n \in USC(Y)$,

$$T^- g(x) \ge T^- h_n(x) \ge \int_Y h_n d\sigma - \mathcal{T}(\delta_x, \sigma), \quad \text{for any } \sigma \in \mathcal{P}(Y).$$

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

$$T^- g(x) \ge \int g d\sigma - \mathcal{T}(\delta_x, \sigma) \quad \text{for any } \sigma \in \mathcal{P}(Y),$$

whereby taking the supremum in σ yields the reverse inequality of (5.2) and gives the desired equality 5.1.

$\square$

Remark 5.1.2. Note that even though for $g \in USC_\sigma(Y)$ which is bounded above, we have the expression $T^- g(x) = \sup_{\sigma \in \mathcal{P}(Y)} \{\int_Y g d\sigma - \mathcal{T}(\delta_x, \sigma)\}$, it does not necessarily imply that $T^- g \in USC(X)$. Indeed, the map $\sigma \mapsto \int_Y g d\sigma$ has no weak* semi-continuity property in general when $g \in USC_\sigma(Y)$, so the supremum above may not be achieved.

Also note that it may happen that both terms in the expression $\int_Y g d\sigma - \mathcal{T}(\delta_x, \sigma)$ are simultaneously $+\infty$ for a given σ when $g \in USC_\sigma(Y)$.

Lemma 5.1.3 (Monotone limits)**.** *Let $\mathcal{T}$ be a backward linear transfer as in Lemma 5.1.1, and let T^- denote its corresponding Kantorovich operator, extended to $USC_\sigma(Y)$.*

1. *If $g_n, g \in USC(Y)$ and $g_n \searrow g$, then $T^- g_n \searrow T^- g$.*
2. *If $g_n, g \in USC_\sigma(Y)$ are bounded above with $g_n \nearrow g$, then $T^- g_n \nearrow T^- g$.*

Proof. 1) We have by the monotonicity property of T^-, that $T^- g(x) \leq \liminf_n T^- g_n(x)$. If for some n, $T^- g_n(x) = -\infty$, then $T^- g(x) = -\infty$ and there is nothing to prove. Otherwise, $T^- g_n(x) > -\infty$ for all n, in which case the expression (5.1) is finite. The map

$$\sigma \mapsto \int_Y g_n d\sigma - \mathcal{T}(\delta_x, \sigma)$$

is weak* upper semi-continuous (since $\sigma \mapsto \int_Y h d\sigma$ is weak* upper semi-continuous for any $h \in USC(Y)$) , so it achieves its supremum at some σ_n, i.e.,

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

Extract an increasing subsequence n_k so that $\limsup_n T^- g_n(x) = \lim_k T^- g_{n_k}(x)$ and $\sigma_{n_k} \to \bar{\sigma}$. Similarly to the proof of Lemma 5.1.1, by monotonicity of g_n, we have

$$T^- g_{n_k}(x) \leq \int g_{n_j} d\sigma_{n_k} - \mathcal{T}(\delta_x, \sigma_{n_k}) \quad \text{for fixed } j \leq k. \tag{5.3}$$

As $g_{n_j} \in USC(Y)$ and $\sigma_{n_k} \to \bar{\sigma}$, it follows that $\limsup_{k\to\infty} \int g_{n_j} d\sigma_{n_k} \leq \int g_{n_j} d\bar{\sigma}$. Hence upon taking $\limsup_{k\to\infty}$ in (5.3), we conclude

$$\limsup_n T^- g_n(x) \leq \int g_{n_j} d\bar{\sigma} - \mathcal{T}(\delta_x, \bar{\sigma}).$$

Then we let $j \to \infty$ and use monotone convergence to conclude that

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

2) Again, by the monotoncity property for T^-, $T^-g \geq \limsup_n T^-g_n(x)$. On the other hand, we know by Lemma 5.1.1 that

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

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

Remark 5.1.4. We note that the given proof of item (1) of the above lemma fails if we allow sequences $g_n \in USC_\sigma(Y)$. This is because there is no weak* semi-continuity property for $\nu \mapsto \int_Y g d\nu$ when g merely belongs to $USC_\sigma(Y)$ and not $USC(Y)$.

Corollary 5.1.5. *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*

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

2. *The Legendre transform formula (2.4) for $\mathcal{T}_\mu$, which holds for $C(Y)$, extends to $USC(Y)$; that is, for $\mu \in \mathcal{P}(X)$ and all $g \in USC(Y)$, we define*

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

and then we have

$$\mathcal{T}_\mu^*(g) = \int_X T^-g\,d\mu. \tag{5.4}$$

Proof. 1. For $g \in USC_b(Y)$, take a monotone decreasing sequence $g_n \in C(Y)$ with $g_n \to g$. By Lemma 5.1.3, and monotone convergence, we infer that

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

from which we conclude

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

The reverse inequality is immediate because $C(Y) \subset USC_b(Y)$.

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

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

so that by Lemma 5.1.3 upon taking $n \to \infty$, we obtain

$$\int_Y g d\sigma - \mathcal{T}(\mu,\sigma) \leq \int_X T^- g\, d\mu$$

and consequently

$$\sup\{\int_Y g d\sigma - \mathcal{T}(\mu,\sigma); \sigma \in \mathcal{P}(Y)\} \leq \int_X T^- g\, d\mu. \tag{5.5}$$

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

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

Since $\mu \in D_1(\mathcal{T})$, the supremum on the right-hand side is achieved by some σ_n. Extract an increasing subsequence n_j so that $\sigma_{n_j} \to \sigma$ for some $\sigma \in \mathcal{P}(Y)$. Then if $i \leq j$, we have $g_{n_j} \leq g_{n_i}$, so that

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

where upon sending $j \to \infty$ yields

$$\int_X T^- g\, d\mu \leq \int_Y g_{n_i} d\sigma - \mathcal{T}(\mu,\sigma)$$

and finally $i \to \infty$ by monotone convergence yields the reverse inequality of (5.5). □

Remark 5.1.6. For similar reasons as noted in a previous remark, the proof of item 1 in the above corollary fails if $USC_b(Y)$ is replaced with $USC(Y)$, (similarly item 2 to $USC_\sigma(Y)$), since it may happen that both $\int_Y g d\nu = -\infty$ and $\int_X T^- g d\mu = -\infty$ simultaneously.

The following is the extension of the inf-convolution stability of Corollary 2.10.4 as mentioned in Remark 2.10.5.

Corollary 5.1.7. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ with $\{\delta_x\,;\, x \in X\} \subset D_1(\mathcal{T})$ and $D_1(\mathcal{T}) \cap D_2(\mathcal{T}) \neq \emptyset$. Then $\mathcal{T} \star \mathcal{T}$ is a backward linear transfer with backward Kantorovich operator $T^- \circ T^-$.*

Proof. The key point here is that $T^- g \in USC(X)$, so by Corollary 5.1.5 (in particular, (5.4); replace g there by $T^- g$ for $g \in C(X)$), we conclude that when $\mu \in D_1(\mathcal{T} \star \mathcal{T})$,

$$
\begin{aligned}
(\mathcal{T} \star \mathcal{T})^*_\mu(g) &= \sup_{\nu \in \mathcal{P}(X)} \{\int_X g d\nu - \mathcal{T} \star \mathcal{T}(\mu, \nu)\} \\
&= \sup_{\sigma \in \mathcal{P}(X)} \{\int_X T^- g d\sigma - \mathcal{T}(\mu, \sigma)\} \\
&= \int_X T^- \circ T^- g d\mu, \quad \text{for all } g \in C(X).
\end{aligned}
$$

□

5.1.1 Conjugate functions for forward and backward linear transfers

We have the following notion motivated by the theory of mass transport (see Section 2.6 and also 4.5).

Definition 5.1.8. Let $\mathcal{T}$ be both a backward and forward transfer with Kantorovich operators T^- and T^+. A pair $(f, g) \in USC_\sigma(X) \times USC_\sigma(Y)$ are a **conjugate pair** if

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

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 5.1.9. *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* $\{(\delta_x, \delta_y)\,;\,(x,y) \in X \times Y\} \subset D(\mathcal{T})$ *such that* $T^+ : C(X) \to C(Y)$ *and* $T^- : C(Y) \to C(X)$*. Then for any* $g \in C(Y)$ *, (resp.,*$f \in C(X)$*)*

$$T^+ \circ T^- g \geq g \qquad T^- \circ T^+ f \leq f,$$

and

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

In particular,

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

Proof. We can write for any $g \in C(Y)$,

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

Since by assumption $T^+ f \in C(Y)$, we have with $g = T^+ f$,

$$T^- \circ T^+ f(x) = \sup_{\sigma \in \mathcal{P}(Y)} \{\int_Y T^+ f d\sigma - \mathcal{T}(\delta_x, \sigma)\}.$$

At the same time, we have $\int_Y T^+ f d\sigma = \inf_{\mu \in \mathcal{P}(X)} \{\int_X f d\mu + \mathcal{T}(\mu, \sigma)\}$ since $\mathcal{T}$ is a forward linear transfer. Substituting into the above yields

$$\begin{aligned}T^- \circ T^+ f(x) &= \sup_{\sigma \in \mathcal{P}(Y)} \inf_{\mu \in \mathcal{P}(X)} \{\int_X f d\mu + \mathcal{T}(\mu, \sigma) - \mathcal{T}(\delta_x, \sigma)\}\\ &\leq f(x).\end{aligned}$$

By a similar argument, we also deduce that $T^+ \circ T^- g(y) \geq g(y)$ for all

$g \in C(Y)$.

By replacing $f \in C(X)$ in the inequality $T^- \circ T^+ f(x) \leq f(x)$ with $T^- g$, we obtain, $T^- \circ T^+ \circ T^- g \leq T^- g$. At the same time applying T^- to the inequality $T^+ \circ T^- g \geq g$, implies $T^- \circ T^+ \circ T^- g \geq T^- g$. Therefore we have equality: $T^- \circ T^+ \circ T^- g = T^- g$. Similarly, we have $T^+ \circ T^- \circ T^+ f = T^+ f$.

Finally, since $\mathcal{T}(\mu, \nu) = \sup_{g \in C(Y)} \{\int_Y g d\nu - \int_X T^- g d\mu\}$, we can replace g with $T^+ \circ T^- g \in C(Y)$ in the supremum since $T^+ \circ T^- g \geq g$.

□

5.2 Mañé constant and weak KAM solutions for backward linear transfers

In this section, we introduce the analogous notions of subsolutions, the Mañé constant, and weak KAM solutions of Section 4.3.1. Throughout this section, let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be a backward linear transfer such that $\{\delta_x ; x \in X\} \subset D_1(\mathcal{T})$ and assume $c(\mathcal{T}) := \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu) < +\infty$. Note that $c(\mathcal{T}) > -\infty$ by the standing assumption that $\mathcal{T}$ is bounded below (recall this is part of the definition of a backward linear transfer). It is always possible to assume without loss of generality that $c(\mathcal{T}) = 0$ by simply considering the transfer $\mathcal{T} - c(\mathcal{T})$; however we prefer to explicitly write the constant in the following. In addition, by compactness and lower semi-continuity, there always exists at least one minimizer $\mu \in \mathcal{P}(X)$ so that $\mathcal{T}(\mu, \mu) = c(\mathcal{T})$.

We first begin by showing that $c(\mathcal{T})$ can be expressed in alternative ways.

Proposition 5.2.1. *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be a backward linear transfer such that $c(\mathcal{T}) := \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu) < +\infty$. Then*

$$c(\mathcal{T}) = \lim_{n \to \infty} \frac{\inf_{(\mu,\nu)} \mathcal{T}_n(\mu, \nu)}{n}.$$

Proof. First note that

$$\inf_{(\mu,\nu)} \mathcal{T}_n(\mu, \nu) \leq \mathcal{T}_n(\mu, \mu) \leq n\mathcal{T}(\mu, \mu)$$

hence

$$\limsup_{n\to\infty} \frac{\inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu)}{n} \le \inf_{\mu} \mathcal{T}(\mu,\mu) = c(\mathcal{T}).$$

On the other hand, let μ_1^n, μ_{n+1}^n be such that

$$\inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu) = \mathcal{T}_n(\mu_1^n, \mu_{n+1}^n). \tag{5.6}$$

By definition of $\mathcal{T}_n$ as an inf-convolution of $\mathcal{T}$ with itself n-times, we may write

$$\mathcal{T}_n(\mu_1^n, \mu_{n+1}^n) = \sum_{j=1}^{n} \mathcal{T}(\mu_j^n, \mu_{j+1}^n)$$

for some $\mu_j^n \in \mathcal{P}(X)$, $j = 1, \ldots, n+1$ (the infimum is achieved by weak* lower semi-continuity). Hence by joint convexity,

$$\frac{1}{n}\mathcal{T}_n(\mu_1^n, \mu_{n+1}^n) = \frac{1}{n}\sum_{j=1}^{n} \mathcal{T}(\mu_j^n, \mu_{j+1}^n) \ge \mathcal{T}(\frac{1}{n}\sum_{j=1}^{n} \mu_j^n, \frac{1}{n}\sum_{j=1}^{n} \mu_{j+1}^n). \tag{5.7}$$

Define $\nu_n := \frac{1}{n}\sum_{j=1}^n \mu_j^n$. Then

$$\mathcal{T}(\frac{1}{n}\sum_{j=1}^{n} \mu_j^n, \frac{1}{n}\sum_{j=1}^{n} \mu_{j+1}^n) = \mathcal{T}(\nu_n, \nu_n + \frac{1}{n}(\mu_{n+1}^n - \mu_1^n)) \tag{5.8}$$

Now let n_k be a subsequence such that

$$\liminf_{n\to\infty} \frac{\inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu)}{n} = \lim_{k\to\infty} \frac{\inf_{(\mu,\nu)} \mathcal{T}_{n_k}(\mu,\nu)}{n_k}.$$

Up to extracting a further subsequence, we may assume that $\nu_{n_k} \to \bar{\nu}$ for some $\bar{\nu} \in \mathcal{P}(X)$. It then follows from (5.6), (5.7), and (5.8), together with weak* lower semi-continuity of $\mathcal{T}$, that

$$\liminf_{n\to\infty} \frac{1}{n} \inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu) \ge \mathcal{T}(\bar{\nu},\bar{\nu}) \ge c(\mathcal{T}),$$

which concludes the proof. $\square$

Proposition 5.2.2. *If $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$*

such that $\sup_{x\in X}\inf_{\nu\in\mathcal{P}(X)}\mathcal{T}(\delta_x,\nu)<+\infty$*, then*

$$c(\mathcal{T})=\sup_{g\in C(X)}\inf_{x\in X}\{g(x)-T^-g(x)\}.$$

Proof. If $\sup_{x\in X}\inf_{\nu\in\mathcal{P}(X)}\mathcal{T}(\delta_x,\nu)<+\infty$, then by Lemma 5.1.1, $T^-g\in USC_b(X)$ for all $g\in C(X)$. Consider now the function

$$f:\mathcal{P}(X)\times C(X)\to\mathbb{R},\quad f(\mu,g):=\int_X(g-T^-g)d\mu.$$

Note that f is real-valued since $T^-g\in USC_b(X)$. We have that $g\mapsto f(\mu,g)$ is upper semi-continuous on $C(X)$ since T^- is convex and weakly lower semi-continuous on $C(X)$, and $\mu\mapsto f(\mu,g)$ is lower semi-continuous on $\mathcal{P}(X)$ since $T^-g\in USC_b(X)$. Moreover, $\mu\mapsto f(\mu,g)$ is quasi-convex, i.e. $\{\mu\in\mathcal{P}(X)\,;\,f(\mu,g)\le\lambda\}$ is convex or empty for $\lambda\in\mathbb{R}$, and $g\mapsto f(\mu,g)$ is quasi-concave, i.e. $\{g\in C(X)\,;\,f(\mu,g)\ge\lambda\}$ is convex or empty for $\lambda\in\mathbb{R}$. Therefore by Sion's minimax theorem (see Theorem 2.7.6), we have

$$\inf_{\mu\in\mathcal{P}(X)}\sup_{g\in C(X)}f(\mu,g)=\sup_{g\in C(X)}\inf_{\mu\in\mathcal{P}(X)}f(\mu,g).$$

Therefore we have

$$\begin{aligned}c(\mathcal{T})=\inf_{\mu\in\mathcal{P}(X)}\mathcal{T}(\mu,\mu)&=\sup_{g\in C(X)}\{\int_X(g-T^-g)d\mu\}\\&=\inf_{\mu\in\mathcal{P}(X)}\sup_{g\in C(X)}f(\mu,g)\\&=\sup_{g\in C(X)}\inf_{\mu\in\mathcal{P}(X)}f(\mu,g)\\&=\sup_{g\in C(X)}\inf_{\mu\in\mathcal{P}(X)}\{\int_X(g-T^-g)d\mu\}.\end{aligned}$$

Since $g-T^-g$ is a lower semi-continuous function bounded below, it achieves a minimum on the compact space X, so that $\inf_{\mu\in\mathcal{P}(X)}\{\int_X(g-T^-g)d\mu\}=$

$\min_{x\in X}\{g(x) - T^- g(x)\}$. Consequently,

$$c(\mathcal{T}) = \sup_{g\in C(X)} \inf_{x\in X} \{g(x) - T^- g(x)\}.$$

$\square$

Definition 5.2.3. Let $k \in \mathbb{R}$. A function $g \in USC(X)$ is a **subsolution** for T^- at level k iff

1. $T^- g(x) + k \leq g(x)$ for all $x \in X$, and
2. $\int_X g d\mu > -\infty$ for some $\mu \in \mathcal{P}(X)$ which achieves the minimum $\mathcal{T}(\mu, \mu) = c(\mathcal{T})$.

Regarding the role of item 2 in the definition of a subsolution, we wish to be able to compare the constant k with $c(\mathcal{T})$ (see below in this section); note that by definition of $c(\mathcal{T})$, we have $\int_X g d\mu \leq \int_X T^- g d\mu + c(\mathcal{T})$. The condition is essentially telling us that a subsolution must be proper in a particular way: that it is finite on the support of some minimising measure μ.

Lemma 5.2.4. *For every $k < \sup_{g\in C(X)} \inf_{x\in X}\{g(x)-T^- g(x)\}$, there exists a subsolution $g \in C(X)$.*

Proof. This is immediate, since by definition, there exists a sequence $(g_j) \subset C(X)$ such that

$$\inf_{x\in X}\{g_j(x) - T^- g_j(x)\} \nearrow \sup_{g\in C(X)} \inf_{x\in X}\{g(x) - T^- g(x)\}.$$

$\square$

Definition 5.2.5. The **Mañé constant** c_0 is the supremum over all $k \in \mathbb{R}$ such that there exists a subsolution g for T^- at level k.

Lemma 5.2.6. *The Mañé constant c_0 satisfies*

$$\sup_{g\in C(X)} \inf_{x\in X}\{g(x) - T^- g(x)\} \leq c_0 \leq c(\mathcal{T}).$$

In particular, if $\sup_{x\in X}\inf_{\nu\in\mathcal{P}(X)}\mathcal{T}(\delta_x,\nu)<+\infty$, *then* $c_0=c(\mathcal{T})$.

Proof. It is immediate from the previous lemma that

$$c_0 \geq \sup_{g\in C(X)}\inf_{x\in X}\{g(x)-T^-g(x)\}$$

since there is a subsolution at every level k strictly below this value. On the other hand, take $\mu\in\mathcal{P}(X)$ for which $c(\mathcal{T})=\mathcal{T}(\mu,\mu)$. Then from $\mathcal{T}(\mu,\mu)=\sup_{h\in C(X)}\{\int_X hd\mu-\int_X T^-hd\mu\}$, we have for any $h\in C(X)$,

$$\int_X(h-T^-h)d\mu\leq c(\mathcal{T}).$$

For any $g\in USC(X)$, take a decreasing sequence $(h_j)\subset C(X)$ with $h_j\searrow g$. Then by Lemma 5.1.3, we have $T^-h_j\searrow T^-g$, so by monotone convergence

$$\int_X gd\mu\leq\int_X T^-gd\mu+c(\mathcal{T}).$$

Therefore for any level k for which there exists $g\in USC(X)$ which is a subsolution, we have

$$\int_X gd\mu\leq\int_X T^-gd\mu+c(\mathcal{T})\leq\int_X gd\mu+c(\mathcal{T})-k$$

hence since $\int_X gd\mu>-\infty$, we may subtract from both sides and deduce $k\leq c(\mathcal{T})$. Consequently $c_0\leq c(\mathcal{T})$.

□

Definition 5.2.7. A function $g\in USC(X)$ is a **backward weak KAM solution at level** k for $\mathcal{T}$ if

1. $T^-g(x)+k=g(x)$ for all $x\in X$, and

2. $\int_X gd\mu>-\infty$ for some μ such that $\mathcal{T}(\mu,\mu)=c(\mathcal{T})$.

Proposition 5.2.8. *If* $g\in USC(X)$ *is a backward weak KAM solution at level* k*, then necessarily,* $k=c(\mathcal{T})$.

Proof. Clearly, if g is a backward weak KAM solution at level k, then it is a subsolution, so $k \leq c_0 \leq c(\mathcal{T})$. On the other hand, g also satisfies $T_n^- g + nk = g$, where T_n^- is the n-fold composition of T^- with itself. This means that

$$\int_X g d\mu - nk = \int_X T_n^- g d\mu = \sup_{\nu \in \mathcal{P}(X)} \{\int_X g d\nu - \mathcal{T}_n(\mu,\nu)\} \leq \sup_{x \in X} g(x) - \inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu)$$

where for the second equality in the above, we have used Corollary 5.1.7. Dividing by n and letting $n \to \infty$, we have

$$-k \leq -\lim_{n\to\infty} \frac{\inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu)}{n} = -c(\mathcal{T}) \tag{5.9}$$

where the latter equality is by Proposition 5.2.1. The above inequality (5.9) then implies $c(\mathcal{T}) \leq k$, which concludes the proof. □

We shall therefore say g is a *backward weak KAM solution* if and only if g is a backward weak KAM solution at level $c(\mathcal{T})$.

Corollary 5.2.9. *If there exists a backward weak KAM solution g, then the Mañé constant c_0 is equal to $c(\mathcal{T})$.*

5.3 Idempotent linear transfers

We will in subsequent sections construct an idempotent backward Kantorovich operator that maps into the set of backward weak KAM solutions; its induced backward linear transfer will be idempotent. Therefore in this section, we study idempotent backward linear transfers.

Definition 5.3.1. Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be a given functional.

1. We say $\mathcal{T}$ is **idempotent** if

$$\mathcal{T}(\mu,\nu) = \mathcal{T} \star \mathcal{T}(\mu,\nu) \quad \text{for all } \mu,\nu \in \mathcal{P}(X)$$

where

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

2. Define $\mathcal{A} := \{\mu \in \mathcal{P}(X)\,;\, \mathcal{T}(\mu,\mu) = 0\}$. We say that $\mathcal{T}$ is **$\mathcal{A}$-factorisable**, if $\mathcal{A} \neq \emptyset$, and $\mathcal{T}$ satisfies

$$\mathcal{T}(\mu,\nu) = \inf\{\mathcal{T}(\mu,\sigma) + \mathcal{T}(\sigma,\nu)\,;\, \sigma \in \mathcal{A}\} \quad \text{ for all } \mu,\nu \in \mathcal{P}(X).$$

3. $\mathcal{T}$ is **distance-like**, if it satisfies

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

The following lemma and proposition is based on ([6], Lemma 11) where we weaken the continuity hypothesis to lower semi-continuity.

Lemma 5.3.2. *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be weak* lower semi-continuous and idempotent. Fix $\mu,\nu \in \mathcal{P}(X)$. Let σ_1, σ_2 be such that*

$$\mathcal{T}(\mu,\nu) = \mathcal{T}(\mu,\sigma_1) + \mathcal{T}(\sigma_1,\nu) \quad \textit{and} \quad \mathcal{T}(\sigma_1,\nu) = \mathcal{T}(\sigma_1,\sigma_2) + \mathcal{T}(\sigma_2,\nu).$$

Then $\mathcal{T}(\mu,\sigma_1) + \mathcal{T}(\sigma_1,\sigma_2) = \mathcal{T}(\mu,\sigma_2)$.

Proof. This simply a consequence of $\mathcal{T}$ being idempotent. Indeed

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

so all the inequalities are in fact equalities. This means in particular comparing the first and third line that $\mathcal{T}(\mu,\sigma_1) + \mathcal{T}(\sigma_1,\sigma_2) = \mathcal{T}(\mu,\sigma_2)$. □

Proposition 5.3.3. *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be weak* lower semi-continuous and idempotent. Then $\mathcal{T}$ is distance-like and $\mathcal{A}$-factorisable.*

Proof. It is immediate that $\mathcal{T}$ is distance-like if it is idempotent; indeed, this is from

$$\begin{aligned}\mathcal{T}(\mu,\nu) = \mathcal{T}\star\mathcal{T}(\mu,\nu) &= \inf_{\sigma\in\mathcal{P}(X)}\{\mathcal{T}(\mu,\sigma)+\mathcal{T}(\sigma,\nu)\}\\ &\leq \mathcal{T}(\mu,\sigma)+\mathcal{T}(\sigma,\nu) \quad \text{for all } \mu,\nu,\sigma.\end{aligned}$$

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

$$\mathcal{T}(\mu,\nu) = \mathcal{T}(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\nu); \tag{5.10}$$

similarly, there exists a σ_2 such that

$$\mathcal{T}(\sigma_1,\nu) = \mathcal{T}(\sigma_1,\sigma_2)+\mathcal{T}(\sigma_2,\nu). \tag{5.11}$$

We conclude from Lemma 5.3.2 that

$$\mathcal{T}(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2) = \mathcal{T}(\mu,\sigma_2). \tag{5.12}$$

In addition, we if we add (5.10) and (5.11), we have

$$\mathcal{T}(\mu,\nu) = \mathcal{T}(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\mathcal{T}(\sigma_2,\nu). \tag{5.13}$$

We therefore have (σ_1,σ_2) such that (5.10), (5.11), and (5.12) (and, consequently, (5.13)) hold.

Continue this process with $\mathcal{T}(\sigma_2,\nu)$, to find a σ_3 satisfying $\mathcal{T}(\sigma_2,\nu) = \mathcal{T}(\sigma_2,\sigma_3)+\mathcal{T}(\sigma_3,\nu)$. In this way we inductively create a sequence $(\sigma_k)_{k\in\mathbb{N}}$ with the property that for $(\sigma_1,\sigma_2,\ldots,\sigma_k)$, we have

$$\mathcal{T}(\mu,\nu) = \mathcal{T}(\mu,\sigma_1)+\sum_{i=1}^{k-1}\mathcal{T}(\sigma_i,\sigma_{i+1})+\mathcal{T}(\sigma_k,\nu) \tag{5.14}$$

and also

$$\mathcal{T}(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2) = \mathcal{T}(\mu,\sigma_2) \tag{5.15}$$

$$\mathcal{T}(\sigma_{k-1},\sigma_k)+\mathcal{T}(\sigma_k,\nu) = \mathcal{T}(\sigma_{k-1},\nu), \tag{5.16}$$

as well as

$$\sum_{i=\ell}^{m} \mathcal{T}(\sigma_i, \sigma_{i+1}) = \mathcal{T}(\sigma_\ell, \sigma_{m+1}) \tag{5.17}$$

whenever $1 \leq \ell < m \leq k-1$.

In particular, the same properties (5.14), (5.15), (5.16), (5.17) above hold if we take instead $m+1$ terms $(\sigma_{k_1}, \ldots, \sigma_{k_{m+1}})$ of any increasing subsequence k_j of k. In particular, this means

$$\mathcal{T}(\mu, \sigma_{k_1}) + \sum_{j=1}^{m} \mathcal{T}(\sigma_{k_j}, \sigma_{k_{j+1}}) + \mathcal{T}(\sigma_{k_{m+1}}, \nu) = \mathcal{T}(\mu, \nu). \tag{5.18}$$

We take now a subsequence σ_{k_j} of σ_k converging to some $\sigma \in \mathcal{P}(X)$. By weak* lower semi-continuity of $\mathcal{T}$, we have

$$\begin{aligned} \liminf_{j\to\infty} \mathcal{T}(\sigma_{k_j}, \sigma_{k_{j+1}}) &\geq \mathcal{T}(\sigma, \sigma) \\ \liminf_{j\to\infty} \mathcal{T}(\mu, \sigma_{k_j}) &\geq \mathcal{T}(\mu, \sigma) \\ \liminf_{j\to\infty} \mathcal{T}(\sigma_{k_j}, \nu) &\geq \mathcal{T}(\sigma, \nu). \end{aligned}$$

In particular, given $\epsilon > 0$, for all but finitely many j, it must hold that

$$\mathcal{T}(\sigma_{k_j}, \sigma_{k_{j+1}}) \geq \mathcal{T}(\sigma, \sigma) - \epsilon \tag{5.19}$$

$$\mathcal{T}(\mu, \sigma_{k_j}) \geq \mathcal{T}(\mu, \sigma) - \epsilon \tag{5.20}$$

$$\mathcal{T}(\sigma_{k_j}, \nu) \geq \mathcal{T}(\sigma, \nu) - \epsilon \tag{5.21}$$

so up to removing the first N terms for a finite $N = N_\epsilon$, we may assume we have a subsequence σ_{k_j} satisfying (5.19), (5.20), and (5.21) for all j, as well as (5.18) .

Applying the inequalities of (5.19), (5.20), and (5.21), to (5.18), we obtain

$$\mathcal{T}(\mu, \nu) \geq \mathcal{T}(\mu, \sigma) + m\mathcal{T}(\sigma, \sigma) + \mathcal{T}(\sigma, \nu) - (m+2)\epsilon$$

for $m \geq 1$. From the fact that $\mathcal{T} = \mathcal{T} \star \mathcal{T}$, we have $\mathcal{T}(\mu, \sigma) + \mathcal{T}(\sigma, \nu) \geq$

$\mathcal{T} \star \mathcal{T}(\mu,\nu) = \mathcal{T}(\mu,\nu)$, so the above inequality implies

$$\mathcal{T}(\sigma,\sigma) \leq \frac{m+2}{m}\epsilon \leq 2\epsilon.$$

As $\epsilon > 0$ is arbitrary, we obtain $\mathcal{T}(\sigma,\sigma) \leq 0$, and consequently $\mathcal{T}(\sigma,\sigma) = 0$ (the reverse inequality following from $\mathcal{T} = \mathcal{T} \star \mathcal{T}$).

Finally, we note that $\mathcal{T}(\mu,\nu) = \mathcal{T}(\mu,\sigma_{k_j}) + \mathcal{T}(\sigma_{k_j},\nu)$ for all j, so at the $\liminf$, we find

$$\mathcal{T}(\mu,\nu) \geq \mathcal{T}(\mu,\sigma) + \mathcal{T}(\sigma,\nu).$$

The reverse inequality is immediate from $\mathcal{T} = \mathcal{T} \star \mathcal{T}$. □

Definition 5.3.4. Suppose $T^- : USC(X) \to USC(X)$ is some given map. We say that T^- is **idempotent** if $T^- \circ T^- g = T^- g$ for all $g \in C(X)$.

Remark 5.3.5. We observe that if $T^- : USC(X) \to USC(X)$ is an idempotent backward Kantorovich operator, then the induced backward linear transfer $\mathcal{T}(\mu,\nu) := \sup_{g\in C(X)}\{\int_X g d\nu - \int_X T^- g d\mu\}$ is idempotent thanks to Corollary 5.1.7.

5.3.1 Examples of idempotent linear transfers

We present some examples of idempotent linear transfers.

Example 5.3.6 (Convex energy)**.** Recall the convex energy example of Section 2.5.1. If I is any bounded below, convex, and weak* lower semicontinuous functional, 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 $T^- g := I^*(g) + m$. Indeed, note that

$$\begin{aligned} T^- \circ T^- g = I^*(T^- g) + m &= \sup_{\nu\in\mathcal{P}(Y)}\{\int_Y T^- g d\nu - I(\nu)\} + m \\ &= \sup_{\nu\in\mathcal{P}(Y)}\{\int_Y (I^*(g)+m) d\nu - I(\nu)\} + m \\ &= I^*(g) + 2m - \inf_{\nu\in\mathcal{P}(Y)} I(\nu) \\ &= I^*(g) + m = T^- g. \end{aligned}$$

(See also Example 5.5.5 later in Section 5.5.3.)

Example 5.3.7 (Markov operator)**.** 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 in Example 2.5.7.

Example 5.3.8 (Balayage transfer)**.** The balayage transfer $\mathcal{B}$ of Example 2.5.3 is idempotent since its Kantorovich map is $T^- g = \hat{g}$, where, for example, in the case of balayage with convex functions, $\hat{g}$ is the concave envelope of g.

Example 5.3.9 (Optimal Mass Transport)**.** If $\mathcal{T}_c$ is an optimal mass transport associated to a bounded below lower semi-continuous cost function $c : X \times X \to \mathbb{R} \cup \{+\infty\}$, then $\mathcal{T}_c$ is idempotent if $c(x,x) = 0$ for every $x \in X$ and c satisfies the triangular inequality

$$c(x,z) \leq c(x,y) + c(y,z) \quad \text{for all } x,y,z \text{ in } X. \tag{5.22}$$

In particular, this implies

$$\mathcal{T}_c(\mu,\nu) = \sup\{\int_X T_c^- g\, d(\nu - \mu); g \in C(X)\}. \tag{5.23}$$

For example, if $c(x,y) = d_X(x,y)^p$ for $0 < p \leq 1$, then the corresponding optimal mass transport is idempotent since c satisfies the reverse triangle inequality, and therefore $c \star c(x,y) = \inf_{z \in X}\{c(x,z) + c(z,y)\} = c(x,y)$.

Example 5.3.10 (An idempotent optimal Skorohod embedding)**.** The following transfer was considered in Ghoussoub-Kim-Palmer [31].

$$\mathcal{T}(\mu,\nu) := \inf\Big\{\mathbb{E}\Big[\int_0^\tau L(t,B_t)dt\Big];\ \tau \in S(\mu,\nu)\Big\}, \tag{5.24}$$

where $S(\mu,\nu)$ 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 $S(\mu,\nu) = \emptyset$, which is the case if and only if μ and ν are not in subharmonic order. In this case, it has been proved in [31] that under suitable conditions, the backward linear transfer is given by $T^- g = J_g(0,\cdot)$, where $J_g : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ is defined via the dynamic programming principle

$$J_g(t,x) := \sup_{\tau \in \mathcal{R}^{t,x}} \Big\{ \mathbb{E}^{t,x}\Big[g(B_\tau) - \int_t^\tau L(s,B_s)ds \Big] \Big\}, \tag{5.25}$$

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 $\tau \geq t$. $J_g(t,x)$ is actually a "variational solution" for the quasi-variational Hamilton-Jacobi-Bellman equation:

$$\min \left\{ \begin{array}{r} J(t,x) - \psi(x) \\ -\frac{\partial}{\partial t} J(t,x) - \frac{1}{2}\Delta J(t,x) + L(t,x) \end{array} \right\} = 0. \tag{5.26}$$

Note that $J_g(t,x) \geq g(x)$, that is $T^- g \geq g$ for every g, hence $(T^-)^2 g \geq T^- g$.

Assume now $t \to L(t,x)$ is decreasing, which yields that $t \to J(t,x)$ is increasing (see [31]). Given $\epsilon > 0$, fix τ_ϵ such that

$$(T^-)^2 g(x) = J_{T^- g}(0,x) \leq \mathbb{E}^{t,x}\Big[T^- g(B_{\tau_\epsilon}) - \int_t^{\tau_\epsilon} L(s,B_s)ds \Big] + \epsilon,$$

hence since $T^- g(B_{\tau_\epsilon}) = J_g(0, B_{\tau_\epsilon}) \leq J_g(t, B_{\tau_\epsilon})$,

$$\begin{aligned} (T^-)^2 g(x) &\leq \mathbb{E}^{t,x}\Big[J_g(0,B_{\tau_\epsilon}) - \int_t^{\tau} L(s,B_s)ds \Big] + \epsilon \\ &\leq \mathbb{E}^{t,x}\Big[J_g(t,B_{\tau_\epsilon}) - \int_t^{\tau_\epsilon} L(s,B_s)ds \Big] + \epsilon \\ &\leq J_g(0,x) + \epsilon = T^- g(x) + \epsilon, \end{aligned}$$

where the last inequality uses the supermartingale property of the process $t \to J_g(t,B_{\tau_\epsilon}) - \int_t^{\tau_\epsilon} L(s,B_s)ds$. It follows that we obtain the reverse inequality $(T^-)^2 g \leq T^- g$, and T^- is therefore idempotent.

5.3.2 $\mathcal{T}$-Lipschitz functionals

Proposition 5.3.11. *Suppose $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ is a weak* lower semi-continuous, bounded below, jointly convex functional such that $\mathcal{T}$ is $\mathcal{A}$-factorisable. Then the following hold:*

1. *For any bounded functional $\Phi : \mathcal{A} \to \mathbb{R}$ that is $\mathcal{T}$-Lipschitz, that is, for any Φ satisfying*

$$\Phi(\nu) - \Phi(\mu) \leq \mathcal{T}(\mu, \nu), \quad \text{for all } \mu, \nu \in \mathcal{A},$$

   there exists $\overline{f} \in C(X)$ such that

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

2. *If $\mathcal{T}$ is, in addition, a backward linear transfer, and*

$$\sup_{\nu \in \mathcal{P}(X)} \inf_{\mu \in \mathcal{A}} \mathcal{T}(\mu, \nu) < +\infty,$$

   then

$$\Phi(\mu) = \int_X T^- \overline{f} d\mu = \int_X \overline{f} d\mu \text{ for every } \mu \in \mathcal{A}. \tag{5.27}$$

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

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

   Moreover, the functions $\psi_0 := T^- \overline{f}$ and $\psi_1 := T^+ \circ T^- \overline{f}$ 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*

$$\psi_0 \leq T^- g \quad \text{and} \quad \psi_1 \geq T^+ g. \tag{5.28}$$

Proof. 1. Let Φ be such that $\mu \to \Phi(\mu)$ is $\mathcal{T}$-Lipschitz on $\mathcal{A}$ and define

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

Note that Φ_0 and Φ_1 are both finite on $\mathcal{A}$, but in general Φ_0 may be $-\infty$ (resp., Φ_1 may be $+\infty$) on certain subsets of $\mathcal{P}(X)$, depending on the effective domain of $\mathcal{T}$.

We now show that $\Phi_0 \leq \Phi_1$ on $\mathcal{P}(X)$. This is trivally true if one of Φ_0, Φ_1 is not finite, so assume $\mu \in \mathcal{P}(X)$ is such that $\Phi_0(\mu), \Phi_1(\mu) \in \mathbb{R}$. Then, by definition of Φ_0, Φ_1, and the fact that Φ is $\mathcal{T}$-Lipschitz on $\mathcal{A}$, and $\mathcal{T}$ is $\mathcal{A}$-factorisable, we may write

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

Note now that Φ_0 is concave and weak* upper semi-continuous, while Φ_1 is convex and weak* lower semi-continuous, on the convex subset $\mathcal{P}(X)$ of the real vector space $\mathcal{M}(X)$. Thus by Hahn-Banach (see, e.g. [50] p.319), there exists $\bar{f} \in C(X)$ such that $\mu \mapsto \int_X \overline{f} d\mu$ on $\mathcal{P}(X)$ lies between Φ_0 and Φ_1:

$$\Phi_0(\mu) \leq \int_X \overline{f} d\mu \leq \Phi_1(\mu) \quad \text{for all } \mu \in \mathcal{P}(X). \tag{5.29}$$

On the other hand, it also holds by the definition of Φ_0 and Φ_1, that

$$\Phi_1(\mu) \leq \Phi(\mu) \leq \Phi_0(\mu) \quad \text{for all } \mu \in \mathcal{A} \tag{5.30}$$

so that (5.29) and (5.30) together shows

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

2. Suppose now $\mathcal{T}$ is additionally a backward linear transfer with T^- as

its Kantorovich operator. If $\mu \in \mathcal{A}$,

$$\int_X T^- \overline{f} d\mu = \sup_{\nu \in \mathcal{P}(X)} \{ \int_X \overline{f} d\nu - \mathcal{T}(\mu, \nu)\} \geq \int_X \overline{f} d\mu - \mathcal{T}(\mu, \mu) = \int_X \overline{f} d\mu.$$

Assume now that $\sup_{\nu \in \mathcal{P}(X)} \inf_{\mu \in \mathcal{A}} \mathcal{T}(\mu, \nu) < +\infty$. It can be checked that this implies Φ_0, Φ_1 are bounded. Then it holds that

$$\sup_{\nu \in \mathcal{P}(X)} \{\Phi_1(\nu) - \mathcal{T}(\mu, \nu)\} \leq \sup_{\nu \in \mathcal{P}(X)} \Phi_1(\nu) - C < +\infty$$

where C is a lower bound for $\mathcal{T}$. This means that for any $\mu \in \mathcal{A}$, by (5.29), we can write

$$-\infty < \int_X T^- \overline{f} d\mu = \sup_{\nu \in \mathcal{P}(X)} \{ \int_X \overline{f} d\nu - \mathcal{T}(\mu, \nu)\} \leq \sup_{\nu \in \mathcal{P}(X)} \{\Phi_1(\nu) - \mathcal{T}(\mu, \nu)\}. \tag{5.31}$$

In view of item 1, we will be done if we show that we have the conjugate formula,

$$\sup_{\sigma \in \mathcal{P}(X)} \{\Phi_1(\sigma) - \mathcal{T}(\mu, \sigma)\} = \Phi_0(\mu). \tag{5.32}$$

To this end, for every $\mu \in \mathcal{P}(X)$, we have

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

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

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

This shows (5.32).

3. Suppose in addition that $\mathcal{T}$ is a forward linear transfer with T^+ as a

Kantorovich operator. Recalling the property $T^+ \circ T^- f \geq f$, we have

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

On the other hand, by (5.31) and (5.32),

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

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 (5.28), first note that

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

On the other hand,

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

which completes the proof. $\square$

Corollary 5.3.12. *For every idempotent backward linear transfer $\mathcal{T}$ such that $\{\delta_x\,;\, x \in X\} \subset \mathcal{A}$ and $\sup_{\nu\in\mathcal{P}(X)} \inf_{\mu\in\mathcal{A}} \mathcal{T}(\mu,\nu) < +\infty$, there is a function $\overline{f} \in C(X)$ which is fixed by T^-, i.e. $T^-\overline{f}(x) = \overline{f}(x)$.*

5.4 Ergodic properties of equicontinuous semigroups of transfers

We now proceed to the construction of the backward weak KAM solutions first introduced in Section 5.2. Given a backward linear transfer $\mathcal{T}$ with backward Kantorovich operator T^-, we shall construct a backward Kantorovich operator T^-_∞ which maps $C(X)$ into the set of backward weak KAM solutions for T^-. This operator will be idempotent, and its induced backward linear transfer $\mathcal{T}_\infty$ is the analog of the Peierls barrier of standard Aubry-Mather theory in Chapter 4.

We begin by assuming $\mathcal{T}$ is weak* continuous on $\mathcal{P}(X) \times \mathcal{P}(X)$. We use the notation $\mathcal{T}_n$ to denote, for each $n \in \mathbb{N}$, the inf-convolution $\mathcal{T}_n = \mathcal{T} \star \mathcal{T} \star \ldots. \star \mathcal{T}$ n-times, and T^-_n the n-fold composition of T^- with itself.

Recall that the *quadratic Wasserstein distance*, denoted by $W_2(\mu,\nu) := \sqrt{\mathcal{T}_2(\mu,\nu)}$ where $\mathcal{T}_2$ is optimal transport with quadratic cost $c(x,x') := d_X(x,x')^2$, metrizes the topology of weak* convergence on $\mathcal{P}(X)$ when X is compact. Therefore, if $\mathcal{T}$ is weak* continuous on $\mathcal{P}(X) \times \mathcal{P}(X)$, then it is weak* uniformly continuous since $\mathcal{P}(X) \times \mathcal{P}(X)$ is weak* compact. Therefore, there exists a *modulus of continuity* $\delta : [0,\infty) \to [0,\infty)$, $\delta(0) = 0$, such that

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

We start with the following lemma which can be found in Bernard-Buffoni [6] and adapted to our setting at hand.

Lemma 5.4.1 (Bernard-Buffoni [6], Lemma 9)**.** *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}$ be a weak* continuous functional with a modulus of continuity δ. Then the family $\{\mathcal{T}_n\}_{n\in\mathbb{N}}$ are equicontinuous with the same modulus of continuity δ, and there exists a positive constant $C > 0$ such that*

$$|\mathcal{T}_n(\mu,\nu) - nc(\mathcal{T})| \le C, \quad \textit{for every } n \ge 1 \textit{ and all } \mu,\nu \in \mathcal{P}(X),$$

where we recall from Section 5.2 that

$$c(\mathcal{T}) = \inf_{\mu} \mathcal{T}(\mu,\mu) = \lim_{n\to\infty} \frac{\inf\{\mathcal{T}_n(\mu,\nu)\,;\, \mu,\nu \in \mathcal{P}(X)\}}{n}.$$

Proof. Define $M_n := \max_{\mu,\nu} \mathcal{T}_n(\mu,\nu)$ and $M := \inf_{n\ge1}\{\frac{M_n}{n}\} > -\infty$. The sequence $\{M_n\}_{n\ge1}$ is subadditive, that is $M_{n+m} \le M_n + M_m$, hence $\{\frac{M_n}{n}\}_{n\ge1}$ decreases to its infimum M as $n \to \infty$. Indeed (see e.g. [9] Lemma 1.18), fix $n > 0$ and write for any m, the decomposition $m = nq + r$, where $0 \le r < n$. The subadditivity of M_n implies

$$\frac{M_m}{m} = \frac{M_{nq+r}}{nq+r} \le \frac{M_{nq}}{nq} + \frac{M_r}{nq} \le \frac{M_n}{n} + \frac{M_r}{nq}.$$

We therefore obtain $\limsup_{m\to\infty} \frac{M_m}{m} \le \frac{M_n}{n}$. On the other hand, $\inf_{n\ge1} \frac{M_n}{n} \le \liminf_{n\to\infty} \frac{M_n}{n}$. Therefore, $\frac{M_n}{n}$ converges to M as $n \to \infty$.
In addition, if $m_n := \min_{\mu,\nu} \mathcal{T}_n(\mu,\nu)$, then the above applied to $-m_n$ yields that $\lim_{n\to\infty} \frac{m_n}{n} = m$, where $m := \sup_n \frac{m_n}{n}$.
We now show that $m = M$. Note that the family $\mathcal{T}_n$ all have the same modulus of continuity; this follows exactly because $(\mu,\nu) \mapsto \mathcal{T}(\mu,\sigma_1) + \mathcal{T}(\sigma_1,\sigma_2) + \ldots + \mathcal{T}(\sigma_n,\nu)$ has modulus of continuity δ for each choice of $(\sigma_1,\ldots,\sigma_n) \in \mathcal{P}(X)\times\ldots\times\mathcal{P}(X)$, so the infimum (i.e. the function $\mathcal{T}_n$) also has modulus of continuity δ. The uniform modulus of continuity δ implies the existence of a constant $C > 0$, such that $M_n - m_n \le C$ for every $n \ge 1$. Then, we obtain the string of inequalities

$$nM - C \le M_n - C \le m_n \le \mathcal{T}_n(\mu,\nu) \le M_n \le m_n + C \le nm + C.$$

The left-most and right-most inequalities imply $M \le m$ upon sending $n \to \infty$, hence $m = M$. $\square$

Lemma 5.4.2. *1. For any $g \in C(X)$, there is a constant $C > 0$ such* *that*

$$|T_n^- g(x) + nc(\mathcal{T}) - \sup_X g| \le C \quad \textit{for all } n \in \mathbb{N} \textit{ and all } x \in X.$$

2. *The semi-group of operators* $\{T_n^-\}_{n\ge 1}$ *has the same modulus of continuity as* $\mathcal{T}$.

3. *The constant* $c(\mathcal{T})$ *is critical in the sense that* $T_n^- g + kn \to \pm\infty$ *as* $n \to \infty$ *if* $k \ne c(\mathcal{T})$, *depending on if* $k < c(\mathcal{T})$ *or* $k > c(\mathcal{T})$, *for any* $g \in C(X)$.

Proof. 1) By Lemma 5.4.1 and since $T_n^- g(x) + nc(\mathcal{T}) = \sup_\sigma\{\int g d\sigma - (\mathcal{T}_n(\delta_x, \sigma) - nc(\mathcal{T}))\}$, we have $\sup_X(g) - C \le T_n^- g(x) + nc(\mathcal{T}) \le \sup_X(g) + C$.

For 2) we note that

$$\begin{aligned}
T_n^- g(x) &= \sup_\sigma\{\int g d\sigma - \mathcal{T}_n(\delta_x, \sigma)\} \\
&\le \sup_\sigma\{\int g d\sigma - \mathcal{T}_n(\delta_y, \sigma)\} + \sup_\sigma\{\mathcal{T}_n(\delta_y, \sigma) - \mathcal{T}_n(\delta_x, \sigma)\} \\
&= T_n^- g(y) + \delta(d(x,y)).
\end{aligned}$$

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

3) follows from 1) since

$$\sup_X(g) - C + (k-c)n \le T_n^- g(x) + kn \le \sup_X(g) + C + (k-c)n.$$

□

We shall prove the following, which is an analogue of Bernard-Buffoni [5, 6].

Theorem 5.4.3 (Ergodic properties of weak* continuous linear transfers)**.** *Let* $\mathcal{T}$ *be a weak* continuous backward linear transfer on* $\mathcal{M}(X) \times \mathcal{M}(X)$ *with modulus of continuity* δ, *and with backward Kantorovich operator* $T^- : C(X) \to C(X)$. *Then, there exists a backward Kantorovich operator* T_∞^- :

$C(X) \to C(X)$, which may be taken to be

$$T^-_\infty g(x) := \lim_{n\to\infty}(T^-_n \circ \overline{T}^- g(x) + nc(\mathcal{T}))$$

where $\overline{T}^- g(x) := \limsup_{n\to\infty}(T^-_n g(x) + nc(\mathcal{T}))$. Together with its corresponding backward linear transfer $\mathcal{T}_\infty$, they satisfy:

1. *T^-_∞ maps every $g \in C(X)$ to a backward weak KAM solution for T^-, i.e.,*

   $$T^-_n \circ T^-_\infty g + nc(\mathcal{T}) = T^-_\infty g \quad \text{and} \quad T^-_\infty \circ T^-_n g + nc(\mathcal{T}) = T^-_\infty g$$

   for all $n \in \mathbb{N}$, where $c(\mathcal{T}) = \inf_\mu \mathcal{T}(\mu,\mu) = \lim_{n\to\infty} \frac{\inf\{\mathcal{T}_n(\mu,\nu)\,;\,\mu,\nu\in\mathcal{P}(X)\}}{n}$.

2. *T^-_∞ is idempotent, i.e. $T^-_\infty \circ T^-_\infty g = T^-_\infty g$ for all $g \in C(X)$.*

3. *$\mathcal{T}_\infty$ satisfies,*

   $$(\mathcal{T}_n - nc(\mathcal{T}))\star\mathcal{T}_\infty(\mu,\nu) = \mathcal{T}_\infty(\mu,\nu) = \mathcal{T}_\infty\star(\mathcal{T}_n - nc(\mathcal{T}))(\mu,\nu) \quad \text{for every } n \in \mathbb{N}.$$

4. *$\mathcal{T}_\infty$ is idempotent and therefore $\mathcal{A}$-factorisable, i.e. 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*

   $$\mathcal{T}_\infty(\mu,\nu) = \inf\{\mathcal{T}_\infty(\mu,\sigma) + \mathcal{T}_\infty(\sigma,\nu), \sigma \in \mathcal{A}\},$$

   and the infimum on $\mathcal{A}$ is attained.

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

   $$\sup\left\{\int T^-_\infty g d(\nu-\mu)\,;\, g \in C(X)\right\} \le \mathcal{T}_\infty(\mu,\nu) \le \liminf_{n\to\infty}(\mathcal{T}_n(\mu,\nu) - nc(\mathcal{T})).$$

6. *If $\mathcal{T}(\mu,\mu) = c(\mathcal{T})$, then $\mu \in \mathcal{A}$. Additionally, $(\mu,\mu) \in \mathcal{D}$ if and only if $\mu \in \mathcal{A}$ and $\mathcal{T}(\mu,\mu) = c(\mathcal{T})$, where*

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

Remark 5.4.4. The backward linear transfer $\mathcal{T}_\infty$ is an analog of the *Peierls barrier.*

Proof. 1) Given $g \in C(X)$, define

$$\overline{T}^- g(x) := \limsup_{n\to\infty}(T_n^- g(x) + nc(\mathcal{T})).$$

The first observation we make is that $\overline{T}^-$ is a backward Kantorovich operator. Indeed, note that the monotonicity, convexity, and affine with respect to constants, properties of T^- are inherited by $\overline{T}^-$. The remaining property to check is the lower semi-continuity. For this, suppose $g_k \to g$ in $C(X)$ for the sup norm. We have by Proposition 2.4.4 that

$$T_n^- g(x) \leq T_n^- g_k(x) + \|g - g_k\|_\infty,$$

so that

$$\overline{T}^- g(x) \leq \overline{T}^- g_k(x) + \|g - g_k\|_\infty$$

where $k \to \infty$ then yields $\overline{T}^- g(x) \leq \liminf_{k\to\infty} \overline{T}^- g_k(x)$. By Lemma 5.4.2, $\overline{T}^- g$ is continuous and has a modulus of continuity δ, and $\|\overline{T}^- g\|_\infty \leq \sup_X(g) + C$.

Now the sequence $\sup_{m\geq n}\{T_m^- g + mc(\mathcal{T})\}$ is a sequence of continuous functions that decrease monotonically to $\overline{T}^- g$ pointwise as $n \to \infty$, we may apply Lemma 5.1.3 to deduce,

$$\begin{aligned}
T^- \circ \overline{T}^- g(x) &= \lim_{n\to\infty} T^- \circ \left[\sup_{m\geq n}\{T_m^- g + c(\mathcal{T})m\}\right](x) \\
&\geq \lim_{n\to\infty} \sup_{m\geq n}\{T_{m+1}^- g(x) + c(\mathcal{T})m\} \\
&= \lim_{n\to\infty} \sup_{m\geq n}\{T_{m+1}^- g(x) + (m+1)c(\mathcal{T})\} - c(\mathcal{T}) \\
&= \overline{T}^- g(x) - c(\mathcal{T}).
\end{aligned}$$

Therefore, $T^- \circ \overline{T}^- g(x) + c(\mathcal{T}) \geq \overline{T}^- g(x)$. By monotonicity of the operator

T^-, this inequality implies

$$T_n^- \circ \overline{T}^- g(x) + nc(\mathcal{T}) \geq T_m^- \circ \overline{T}^- g(x) + mc(\mathcal{T})$$

whenever $n \geq m$, i.e. $\{T_n^- \circ \overline{T}^- g + nc(\mathcal{T})\}_{n\geq 1}$ is a monotone increasing sequence of continuous functions, with

$$|T_n^- \circ \overline{T}^- g(x) - T_n^- \circ \overline{T}^- g(x')| \leq \delta(d_X(x, x')).$$

In addition, we have from Lemma 5.4.2 the uniform in n bound

$$\|T_n^- \circ \overline{T}^- g(x) + nc(\mathcal{T})\|_\infty \leq \|\overline{T}^- g\|_\infty + C \leq \|g\|_\infty + 2C.$$

We therefore have a monotone increasing sequence of equicontinuous functions $\{T_n^- \circ \overline{T}^- g + nc(\mathcal{T})\}_{n\in\mathbb{N}}$ bounded above uniformly in n and x. It therefore converges pointwise to a continuous function, and we can define the pointwise limit via the operator $T_\infty^- : C(X) \to C(X)$ defined via the formula,

$$T_\infty^- g(x) := \lim_{n\to\infty} T_n^- \circ \overline{T}^- g(x) + nc(\mathcal{T}).$$

We now claim that T_∞^- is a backward Kantorovich operator. Indeed, since T_n^- and $\overline{T}^-$ are backward Kantorovich operators, the monotonicity, convexity, and affine on constants properties hold in the pointwise limit $n \to \infty$, and therefore hold for T_∞^-. Regarding lower semi-continuity, note that we can repeat a similar argument as was given for $\overline{T}^-$ to deduce that

$$T_n^- \circ \overline{T}^- g(x) \leq T_n^- \circ \overline{T}^- g_k + \|g - g_k\|_\infty. \tag{5.33}$$

which gives

$$T_\infty^- g(x) \leq \liminf_{k\to\infty} T_\infty^- g_k(x).$$

Since the convergence is monotone (even uniform, by Dini's theorem) we can apply Lemma 5.1.3 to deduce that

$$\begin{aligned} T^- \circ T^-_\infty g(x) + c(\mathcal{T}) &= \lim_{m\to\infty} T^- \circ \left[T^-_m \circ \overline{T}^- g(x) + mc(\mathcal{T})\right] + c(\mathcal{T}) \\ &= \lim_{m\to\infty} \left\{T^-_{m+1} \circ \overline{T}^- g(x) + (m+1)c(\mathcal{T})\right\} \\ &= T^-_\infty g(x) \end{aligned}$$

so repeated application of T^- yields $T^-_n \circ T^-_\infty g(x) + nc(\mathcal{T}) = T^-_\infty g(x)$.

We can also consider the composition

$$T^-_\infty \circ T^- g(x) = \lim_{n\to\infty} (T^-_n \circ \overline{T}^- \circ T^- g(x) + nc(\mathcal{T})) \tag{5.34}$$

and we have

$$\begin{aligned} \overline{T}^- T^- g(x) &= \limsup_{n\to\infty} (T^-_n T^- g(x) + nc(\mathcal{T})) \\ &= \limsup_{n\to\infty} (T^-_{n+1} g(x) + (n+1)c(\mathcal{T})) - c(\mathcal{T}) \\ &= \overline{T}^- g(x) - c(\mathcal{T}) \end{aligned}$$

which means from (5.34) that

$$T^-_\infty \circ T^- g(x) + c(\mathcal{T}) = T^-_\infty g(x).$$

2) From $T^-_n \circ T^-_\infty g(x) + nc(\mathcal{T}) = T^-_\infty g(x)$, we obtain $\overline{T}^- \circ T^-_\infty g(x) = T^-_\infty g(x)$ Consequently by definition of T^-_∞, this further implies that $T^-_\infty \circ T^-_\infty g(x) = T^-_\infty g(x)$ so T^-_∞ is idempotent.

3) T^-_∞ is a Kantorovich operator, thus we may define

$$\mathcal{T}_\infty(\mu, \nu) := \sup_{g\in C(X)} \left\{\int_X g d\nu - \int_X T^-_\infty g d\mu\right\}$$

and it is a backward linear transfer; from $T^-_n \circ T^-_\infty g + nc(\mathcal{T}) = T^-_\infty g$ and $T^-_\infty \circ T^-_n g + nc(\mathcal{T}) = T^-_\infty g$, it therefore satisfies

$$\mathcal{T}_\infty(\mu, \nu) = (\mathcal{T}_n - nc(\mathcal{T})) \star \mathcal{T}_\infty(\mu, \nu) = \mathcal{T}_\infty \star (\mathcal{T}_n - nc(\mathcal{T}))(\mu, \nu), \quad \text{for all } n \geq 1.$$

4) From $T_\infty^- \circ T_\infty^- g(x) = T_\infty^- g(x)$, it satisfies

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

so $\mathcal{T}_\infty$ is idempotent. Therefore by Proposition 5.3.3, $\mathcal{T}_\infty$ is $\mathcal{A}$-factorisable.

5) Note from 1) that $T_\infty^- g(x) \geq \limsup_{n\to\infty}(T_n^- g(x) + nc(\mathcal{T}))$, so

$$\begin{aligned}\int_X T_\infty^- g d\mu &\geq \int_X \limsup_{n\to\infty}(T_n^- g(x) + nc(\mathcal{T}))d\mu \\ &\geq \limsup_{n\to\infty} \int_X (T_n^- g(x) + nc(\mathcal{T}))d\mu.\end{aligned}$$

Hence

$$\begin{aligned}\mathcal{T}_\infty(\mu,\nu) &\leq \sup \liminf_{n\to\infty} \left\{ \int_X g d\nu - \int_X T_n^- g d\mu - nc(\mathcal{T})\, ;\, g \in C(X) \right\} \\ &\leq \liminf_{n\to\infty} \sup \left\{ \int_X g d\nu - \int_X T_n^- g d\mu - nc(\mathcal{T})\, ;\, g \in C(X) \right\} \\ &= \liminf_{n\to\infty} (\mathcal{T}_n(\mu,\nu) - nc(\mathcal{T})).\end{aligned}$$

On the other hand, from $T_\infty^- \circ T_\infty^- g = T_\infty^- g$,

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

6) Suppose μ is a measure which realises $c(\mathcal{T}) = \mathcal{T}(\mu,\mu)$. Then by 6), we have

$$\begin{aligned}0 \leq \mathcal{T}_\infty(\mu,\mu) &\leq \liminf_{n\to\infty} (\mathcal{T}_n(\mu,\mu) - nc(\mathcal{T})) \\ &\leq \liminf_{n\to\infty} (n\mathcal{T}(\mu,\mu) - nc(\mathcal{T})) = 0,\end{aligned}$$

so $\mu \in \mathcal{A}$. If $(\mu,\mu) \in \mathcal{D}$, then we immediately see $\mu \in \mathcal{A}$ and $\mathcal{T}(\mu,\mu) = c(\mathcal{T})$; the converse is also immediate.

□

Remark 5.4.5. The above theorem also holds when we have an appropriate semi-group of backward linear transfers $(\mathcal{T}_t)_{t>0}$. In this case, we do not build a discrete-time semi-group $(\mathcal{T}_n)_{n\in\mathbb{N}}$ from one linear transfer $\mathcal{T}$, but instead already have in hand an appropriate semi-group $(\mathcal{T}_t)$.

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

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

then under the following assumptions

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

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

the results of Theorem 5.4.3 and their proofs hold (with appropriate changes from n to t). For clarity we decided to present Theorem 5.4.3 for the discrete-time case.

5.4.1 The case of optimal transport for continuous cost functions

We now identify T_∞^- and $\mathcal{T}_\infty$ associated to a semi-group of linear transfers which are given by mass transports.

Proposition 5.4.6. *Suppose $c_t(x,y)$ is a semi-group of equicontinuous cost functions on $X\times X$, that is*

$$c_{t+s}(x,y)=c_t\star c_s(x,y):=\inf\{c_t(x,z)+c_s(z,y);z\in X\},$$

and suppose $\inf_{(x,y)} c_t(x,y) = \inf_x c_t(x,x)$. *Without loss of generality, assume* $\inf_x c_t(x,x) = 0$, *and consider the associated optimal mass transports*

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

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* $s,t \geq 0$ *that is equicontinuous on* $\mathcal{P}(X)\times\mathcal{P}(X)$, *hence one can associate an idempotent Kantorovich operator* T_∞^- *and associated backward linear transfer* $\mathcal{T}_\infty$.

2. *The following holds for the constant* $c := \lim_{t\to\infty}\frac{\inf_{\mu,\nu\in\mathcal{P}(X)}\mathcal{T}_t(\mu,\mu)}{t}$ *:*

$$c = \min\{\int_{X\times X} c_1(x,y)d\pi; \pi\in\mathcal{P}(X\times X), \pi_1=\pi_2\} \tag{5.35}$$

3. *Letting* $c_\infty(x,y) := \liminf_{t\to\infty}(c_t(x,y)-ct)$, *then* c_∞ *is continuous on* $\mathcal{P}(X)\times\mathcal{P}(X)$, *and:*

$$\mathcal{T}_\infty(\mu,\nu) = \mathcal{T}_{c_\infty}(\mu,\nu) := \inf\{\int_{X\times X} c_\infty(x,y)d\pi(x,y)\,;\,\pi\in\mathcal{K}(\mu,\nu)\},$$

$T_\infty^- f(x) = \sup\{f(y)-c_\infty(x,y)\,;\,y\in X\}$ *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 (5.35) are all supported on the set*

$$D := \{(x,y)\in X\times X\,;\, c_1(x,y)+c_\infty(y,x)=c\}.$$

Proof. We apply Theorem 5.4.3 in this context: The Kantorovich operator for $\mathcal{T}_t$ is given by $T_t^- g(x) = \sup\{g(y)-c_t(x,y)\,;\,y\in X\}$ and as shown in Proposition 2.10.10, 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. For the idempotent Kantorovich operator T_∞^- associated to $(T_t)_t$, we recall that in the proof of Theorem 5.4.3 we had the operator

$$\overline{T}^- g(x) := \limsup_{t\to\infty}(T_t^- g(x) + ct).$$

We claim that $\overline{T}^- g(x) = T_{c_\infty}^- g(x)$ where $T_{c_\infty}^- g(x) := \sup_{y\in X}\{g(y) - c_\infty(x,y)\}$ is the backward Kantorovich operator for the optimal transport with cost c_∞, so that in fact $T_\infty^- = T_{c_\infty}^-$. Indeed, first note that

$$\limsup_t (T_t^- g(x) + ct) \geq \sup_{y\in X}\{g(y) - c_\infty(x,y)\} = T_{c_\infty}^- g(x). \tag{5.36}$$

On the other hand, let y_n achieve the supremum for $T_n^- g(x) = \sup\{g(y) - c_n(x,y)\,;\, y \in X\}$, and let $(n_j)_j$ be a subsequence such that $\lim_{j\to\infty}(T_{n_j}^- g(x) + cn_j) = \limsup_n (T_n f(x) + cn)$. 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 equicontinuity of the c_n's, we deduce that

$$\limsup_n (T_n^- g(x) + cn) = \lim_{j\to\infty}(T_{n_j}^- g(x) + cn_j) = g(\bar{y}) - \liminf_j (c_{n_j}(x,\bar{y}) - cn_j). \tag{5.37}$$

As $\liminf_j (c_{n_j}(x,\bar{y}) - cn_j) \geq \liminf_n (c_n(x,\bar{y}) - cn) = c_\infty(x,\bar{y})$, we obtain

$$\limsup_n (T_n^- g(x) + nc) \leq g(\bar{y}) - c_\infty(x,\bar{y}) \leq \sup_y \{g(y) - c_\infty(x,y)\} = T_{c_\infty}^- g(x). \tag{5.38}$$

The inequality (5.38) is true for every sequence $(n_k)_k$ going to ∞, so we deduce that $\limsup_t (T_t^- g(x) + ct) \leq T_{c_\infty}^- g(x)$, and hence combining this with (5.36) gives equality: $\limsup_t (T_t^- g(x) + ct) = T_{c_\infty}^- g(x)$.

Finally, we note that $T_s^-(\limsup_t (T_t^- g + ct))(x) + cs = T_s^- T_{c_\infty}^- g(x) + cs = T_{c_\infty}^- g(x)$ 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 Theorem 5.4.3) that $T_\infty^- g(x) = T_{c_\infty}^- g(x)$.

Properties (1), (2) and (3) follow then immediately. Properties (4) and (5) now follow since $c_t(x,y)$ is minimised (by assumption) on the diagonal $c_t(x,x)$, so that $\mathcal{T}_\infty(\mu,\mu) = 0 = \mathcal{T}_{c_\infty}(\mu,\mu)$ implies μ is supported on set where $c_\infty(x,x) = 0$. $\square$

Example 5.4.7 (Iterates of power costs)**.** Let $X \subset \mathbb{R}^n$ be compact and convex, and $c(x,y) := |x-y|^p$ for $p > 0$ and $x, y \in X$. Let $\mathcal{T}_c$ denote the optimal transport with cost c, with corresponding backward Kantorovich operator $T_c^- g(x) = \sup_{y\in X}\{g(y) - c(x,y)\}$.

If $0 < p \le 1$, then c satisfies the reverse triangle inequality $c(x,z) + c(z,y) \ge c(x,y)$, hence $c \star c(x,y) = \inf\{|x-z|^p + |z-y|^p\,;\, z \in X\} = c(x,y)$. Therefore $T_\infty^- g(x) = T_c^- g(x)$ (i.e. T_c^- is itself idempotent).

If $p > 1$, then $c \star c(x,y) = \inf\{|x-z|^p + |z-y|^p\,;\, z \in X\}$ is minimised at some point $z = (1-\lambda)x + \lambda y$ on the line between x and y, so that $c_p \star c_p(x,y) = (\lambda^p + (1-\lambda)^p)\,|x-y|^p$. The optimal λ is $\frac{1}{2}$. Hence,

$$(T_c^-)^n g(x) = \sup_{y\in X}\{g(y) - \frac{1}{n^{p-1}}|x-y|^p\}.$$

Therefore, when $n \to \infty$, $(T_c^-)^n g(x) \to \sup_{x\in X} g(x)$, and it follows that $T_\infty^- g(x) := \sup_{x\in X} g(x)$, with the corresponding backward linear transfer $\mathcal{T}_\infty$ being the null transfer of Example 2.5.1.

5.4.2 Aubry-Mather theory and weak KAM theory for Lagrangian systems

In this section, we briefly mention how, when the cost semi-group of the previous section is generated from a Lagrangian, how Theorem 5.4.3 and Proposition 5.4.6 fits in the context of the Aubry-Mather theory, weak KAM, and Bernard-Buffoni connection to optimal transport discussed in Chapter 4.

Let L be a time-independent *Tonelli Lagrangian* on $M = \mathbb{T}^n$ the flat torus, and consider $\mathcal{T}_t$ to be the cost minimizing transport

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

where

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

Recall from Chapter 4 (see Section 4.3), the Lax-Oleinik semi-group S_t^-,

$t > 0$ defined by the formula

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

as well as the semi-group

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

Theorem 5.4.8. *There exists a unique constant $c \in \mathbb{R}$ such that the following hold:*

1. (Fathi [20]) *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 $t \geq 0$.*

2. (Bernard-Buffoni [5]) *Let $c_\infty(x,y) := \liminf_{t\to\infty} c_t(x,y)$ denote the Peierls barrier function. The following duality then holds:*

$$\inf\{\int_{M\times M} c_\infty(x,y)d\pi(x,y)\,;\, \pi \in \mathcal{K}(\mu,\nu)\} = \sup_{u_+,u_-}\{\int_M u_+ d\nu - \int_M u_- d\mu\},$$

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 [6]) *The constant c satisfies*

$$c = \min_\pi \int_{M\times M} c_1(x,y)d\pi(x,y),$$

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 [43]) *The constant $c = \inf_m \int_{TM} L(x,v)dm(x,v)$ 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 [20]) *A continuous function* $u : M \to \mathbb{R}$ *is a viscosity solution of* $H(x, \nabla u(x)) = -c$ *if and only if it is Lipschitz and* u *is a negative weak KAM solution (i.e.* $S_t^- u - ct = u$*). In particular, the statement is false if* c *is replaced with any other constant.*

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

$$V_f(t', x) = \inf\{f(\gamma(0)) + \int_0^{t'} L(\gamma(s), \dot{\gamma}(s))ds\, ;\, \gamma \in C^1([0, t'), M), \gamma(t') = x\}$$

and $W_g^t(t', y)$ the value functional

$$W_g^t(t', y) = \sup\{g(\gamma(t')) - \int_{t'}^{t} L(\gamma(s), \dot{\gamma}(s))ds\, ;\, \gamma \in C^1([0, t'), M), \gamma(0) = x\}.$$

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.

As mentioned in Section 5.4.1 above, the negative (resp. positive) weak KAM solutions are the image of the Kantorovich operators T_∞^+ (resp. T_∞^-), and are 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 $c_\infty(x, y) := \liminf_{t\to\infty}(c_t(x, y) - ct)$, with $c = \lim_{t\to\infty} \frac{\inf_{\mu,\nu\in\mathcal{P}(M)} \mathcal{T}_t(\mu,\nu)}{t}$.

The cost-minimizing transport $\mathcal{T}_{c_\infty}$ with cost c_∞, is then the idempotent backward (and forward) linear transfer associated to T_∞^- (or T_∞^+) which by duality we can write as

$$\inf\{\int_{M\times M} c_\infty(x, y) d\pi(x, y)\, ;\, \pi \in \mathcal{K}(\mu, \nu)\} = \sup\{\int_M T_\infty^+ g d\nu - \int_M T_\infty^- \circ T_\infty^+ g d\mu\, ;\, g \in C(M)\}.$$

It can be checked this is exactly statement 2 in Theorem 5.4.8 above. In particular, Theorem 5.4.3 provides us with statements 1,2, and 3, in the above theorem.

5.4.3 The Schrödinger semigroup

Recall the Schrödinger bridge of Example 2.8.6. 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

$$\mathcal{T}_t(\mu,\nu) := \inf\{\int_M \mathcal{H}(r_t^x, \pi_x)d\mu(x)\,;\, \pi \in \mathcal{K}(\mu,\nu),\, d\pi(x,y) = d\mu(x)d\pi_x(y)\} \tag{5.39}$$

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 5.4.9. *The collection $\{\mathcal{T}_t\}_{t\geq 0}$ 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,*

$$\mathcal{T}_t(\mu,\nu) = \sup\left\{\int_M f d\nu - \int_M \log S_t e^f d\mu\,;\, f \in C(M)\right\}. \tag{5.40}$$

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 $S_\infty g := \int g dm$.

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 2.4.4. 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 Corollary 5.1.5.

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 5.4.3. Finally we see that $\mathcal{T}_\infty(\mu, \nu)$ is

$$\begin{aligned}
\mathcal{T}_\infty(\mu, \nu) &:= \sup\{\int f d\nu - \int T_\infty f d\mu \,;\, f \in C(M)\} \\
&= \sup\{\int f d\nu - \log \int e^f dm \,;\, f \in C(M)\} \\
&= \mathcal{H}(m, \nu).
\end{aligned}$$

□

5.5 Regularization and ergodic properties of non-continuous linear transfers

The assumption of weak* continuity in the last section was important for taking limits of the iterates $T_n^- g + nc$. When $\mathcal{T}$ is not necessarily weak*-continuous on $\mathcal{M}(X)$ and may even have infinite values for some $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X)$, we can no longer guarantee that taking limits of iterates will yield a proper function (i.e. one that is real-valued at least somewhere). Thus we turn to alternative strategies. The first and most natural one is to reduce the situation to the bounded and continuous case via a regularization procedure.

5.5.1 Regularization

Lemma 5.5.1 (Regularisation of a backward linear transfer). *Let $W_1(\mu, \nu)$ be the cost minimising optimal transport associated to the cost $d_X(x, y)$. For a given backward linear transfer $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$, define for*

$\epsilon > 0$,

$$\begin{aligned}\mathcal{T}_\epsilon(\mu,\nu) &:= \left(\frac{1}{\epsilon}W_1\right)\star\mathcal{T}\star\left(\frac{1}{\epsilon}W_1\right)\\ &= \inf\{\frac{1}{\epsilon}W_1(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\frac{1}{\epsilon}W_1(\sigma_2,\nu)\,;\,\sigma_1,\sigma_2\in\mathcal{P}(X)\}.\end{aligned}$$

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. *$\mathcal{T}_\epsilon(\mu,\nu)\le\mathcal{T}(\mu,\nu)$ and $\mathcal{T}_\epsilon(\mu,\nu)\nearrow\mathcal{T}(\mu,\nu)$ as $\epsilon\searrow 0$.*
4. *$\mathcal{T}_\epsilon$ Γ-converges to $\mathcal{T}$ as $\epsilon\searrow 0$.*
5. *If $\{\delta_x\,;\,x\in X\}\subset D_1(\mathcal{T})$, and T_ϵ^-, T^-, denote the backward Kantorovich operators associated to $\mathcal{T}_\epsilon$, $\mathcal{T}$, respectively, then for any $g\in USC(X)$, $T_\epsilon^- g(x)\searrow Tg(x)$ as $\epsilon\searrow 0$.*

Proof. First note that since d_X is continuous, the linear transfer W_1 is weak-$*$ continuous on $\mathcal{P}(X)$ (see e.g., [48], 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 $\liminf_n\mathcal{T}_\epsilon(\mu_n,\nu_n)\ge\mathcal{T}_\epsilon(\mu,\nu)$. On the other hand, from the fact that $\limsup_n\inf_{\sigma_1,\sigma_2}\le\inf_{\sigma_1,\sigma_2}\limsup_n$, we have

$$\begin{aligned}\limsup_n\mathcal{T}_\epsilon(\mu_n,\nu_n) &\le \inf_{\sigma_1,\sigma_2\in\mathcal{P}(X)}\left\{\limsup_n\frac{1}{\epsilon}W_1(\mu_n,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\limsup_n\frac{1}{\epsilon}W_1(\sigma_2,\nu_n)\right\}\\ &= \inf_{\sigma_1,\sigma_2\in\mathcal{P}(X)}\left\{\frac{1}{\epsilon}W_1(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\frac{1}{\epsilon}W_1(\sigma_2,\nu)\right\}\\ &= \mathcal{T}_\epsilon(\mu,\nu),\end{aligned}$$

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

$$\inf_{\mu,\nu}\{\mathcal{T}_\epsilon(\mu,\nu)\} = \inf_{\sigma,\sigma',\mu,\nu}\{\frac{1}{\epsilon}W_1(\mu,\sigma)+\mathcal{T}(\sigma,\sigma')+\frac{1}{\epsilon}W_1(\sigma',\nu)\}.$$

Since $W_1 \geq 0$, it follows that for fixed σ, σ', the minimal value when minimising over all μ, ν is to take $\mu = \sigma$, and $\nu = \sigma'$, in which case the transport cost vanishes $W_1(\mu, \sigma) = 0 = W_1(\sigma', \nu)$.

3. The inequality $\mathcal{T}_\epsilon(\mu, \nu) \leq \mathcal{T}(\mu, \nu)$ holds by selecting $\sigma_1 = \mu$ and $\sigma_2 = \nu$ and noting that $W_1(\sigma, \sigma) = 0$ 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

$$\mathcal{T}_\epsilon(\mu, \nu) = \frac{1}{\epsilon} W_1(\mu, \sigma_\epsilon^1) + \mathcal{T}(\sigma_\epsilon^1, \sigma_\epsilon^2) + \frac{1}{\epsilon} W_1(\sigma_\epsilon^2, \nu). \tag{5.41}$$

By selecting a further subsequence if necessary, we may assume that $\sigma_\epsilon^1 \to \overline{\sigma}_1$ and $\sigma_\epsilon^2 \to \overline{\sigma}_2$ as $\epsilon \to 0$. Suppose now $\liminf_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu, \nu) < \infty$. Then it must be the case that $W_1(\mu, \sigma_\epsilon^1) \to 0$ and $W_1(\sigma_\epsilon^2, \nu) \to 0$ as $\epsilon \to 0$. This implies $\overline{\sigma}_1 = \mu$ and $\overline{\sigma}_2 = \nu$ since W_1 is a metric on $\mathcal{P}(X)$. Then (5.41) and weak* lower semi-continuity of $\mathcal{T}$ implies

$$\mathcal{T}(\mu, \nu) \geq \liminf_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu, \nu) \geq \liminf_{\epsilon \to 0} \mathcal{T}(\sigma_\epsilon^1, \sigma_\epsilon^2) \geq \mathcal{T}(\mu, \nu).$$

If $\liminf_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu, \nu) = \infty$, then necessarily $\mathcal{T}(\mu, \nu) = +\infty$ since $\mathcal{T}(\mu, \nu) \geq \mathcal{T}_\epsilon(\mu, \nu)$ for all ϵ. In either case, we deduce that $\lim_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu, \nu) = \mathcal{T}(\mu, \nu)$.

4. First recall that for Γ-convergence, one needs to prove

(i) the Γ-$\liminf$ inequality: For every sequence $(\mu^\epsilon, \nu^\epsilon) \to (\mu, \nu)$, it holds that $\liminf_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu_\epsilon, \nu_\epsilon) \geq \mathcal{T}(\mu, \nu)$, and

(ii) the Γ-$\limsup$ inequality: There exists a sequence $(\mu_\epsilon, \nu_\epsilon) \to (\mu, \nu)$ such that $\limsup_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu_\epsilon, \nu_\epsilon) \leq \mathcal{T}(\mu, \nu)$.

The Γ-$\limsup$ inequality is immediate: Take $(\mu_\epsilon, \nu_\epsilon) = (\mu, \nu)$, and the inequality follows from $\mathcal{T}_\epsilon \leq \mathcal{T}$. For the Γ-$\liminf$ inequality, we have by monotonicity that $\mathcal{T}_\epsilon(\mu^\epsilon, \nu^\epsilon) \geq \mathcal{T}_{\epsilon'}(\mu^\epsilon, \nu^\epsilon)$ for $\epsilon \leq \epsilon'$. The weak* lower semi-continuity of $\mathcal{T}_{\epsilon'}$ therefore implies

$$\liminf_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu^\epsilon, \nu^\epsilon) \geq \liminf_{\epsilon \to 0} \mathcal{T}_{\epsilon'}(\mu^\epsilon, \nu^\epsilon) \geq \mathcal{T}_{\epsilon'}(\mu, \nu).$$

By 3) and letting $\epsilon' \to 0$, we obtain $\liminf_{\epsilon \to 0} \mathcal{T}_\epsilon(\mu^\epsilon, \nu^\epsilon) \geq \mathcal{T}(\mu, \nu)$.

5. First note that the monotonicity of $T_\epsilon^- g(x)$ is immediate from the

expression

$$T_\epsilon^- g(x) = \sup_{\sigma\in\mathcal{P}(X)} \left\{ \int_X g d\sigma - \mathcal{T}_\epsilon(\delta_x, \sigma) \right\},$$

together with the monotonicity of $\mathcal{T}_\epsilon$. From

$$\begin{aligned}
\limsup_{\epsilon\to 0} \sup_{\sigma\in\mathcal{P}(X)} \left\{ \int_X g d\sigma - \mathcal{T}_\epsilon(\delta_x, \sigma) \right\} &\geq \sup_{\sigma\in\mathcal{P}(X)} \left\{ \int_X g d\sigma - \liminf_{\epsilon\to 0} \mathcal{T}_\epsilon(\delta_x, \sigma) \right\} \\
&= \sup_{\sigma\in\mathcal{P}(X)} \left\{ \int_X g d\sigma - \mathcal{T}(\delta_x, \sigma) \right\}
\end{aligned}$$

we immediately have $\liminf_{\epsilon\to 0} T_\epsilon^- g(x) \geq T^- g(x)$. On the other hand, let ϵ_j be a sequence such that $T_{\epsilon_j}^- g(x) \to \limsup_{\epsilon\to 0} T_\epsilon^- g(x)$. Then there is σ_{ϵ_j} such that

$$T_{\epsilon_j}^- g(x) = \sup_{\sigma\in\mathcal{P}(X)} \left\{ \int_X g d\sigma - \mathcal{T}_\epsilon(\delta_x, \sigma) \right\} = \int_X g d\sigma_{\epsilon_j} - \mathcal{T}_{\epsilon_j}(\delta_x, \sigma_{\epsilon_j}).$$

By selecting a further subsequence if necessary, we may assume $\sigma_{\epsilon_j} \to \sigma^*$. Then we obtain with $j \to \infty$,

$$\begin{aligned}
\limsup_{\epsilon\to 0} T_\epsilon^- g(x) = \lim_{j\to\infty} T_{\epsilon_j}^- g(x) &\leq \int_X g d\sigma^* - \liminf_{j\to\infty} \mathcal{T}_{\epsilon_j}(\delta_x, \sigma_{\epsilon_j}) \\
&\leq \int_X g d\sigma^* - \mathcal{T}(\delta_x, \sigma^*) \\
&\leq \sup_{\sigma\in\mathcal{P}(X)} \left\{ \int_X g d\sigma - \mathcal{T}(\delta_x, \sigma) \right\} = T^- g(x),
\end{aligned}$$

where the second inequality was obtained from the Γ-convergence. □

Lemma 5.5.2. *Let $\mathcal{T}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains all the Dirac measures. Assume $c(\mathcal{T}) := \inf_\mu \mathcal{T}(\mu, \mu) < +\infty$. and let $\mathcal{T}_\epsilon$ be the regularisation of $\mathcal{T}$ according to Lemma 5.5.1. Then, the following properties hold:*

1. *$c(\mathcal{T}_\epsilon) := \inf_{\mu,\mu} \mathcal{T}_\epsilon(\mu, \mu) = \lim_{n\to\infty} \frac{\inf_{\mu,\nu\in\mathcal{P}(X)} \mathcal{T}_\epsilon(\mu,\nu)}{n}$ is the unique constant such that $|(\mathcal{T}_\epsilon)_n(\mu, \nu) - nc(\mathcal{T}_\epsilon)| \leq C_\epsilon$, for all n and all μ, ν.*

2. *$c(\mathcal{T}_\epsilon) \nearrow c(\mathcal{T})$ as $\epsilon \searrow 0$.*

Proof. Use Lemma 5.5.1 to regularise $\mathcal{T}$ to $\mathcal{T}_\epsilon$. Then item 1 is simply Lemma 5.4.1 for $\mathcal{T}_\epsilon$. For 2), note that since $\mathcal{T}_\epsilon(\mu,\mu) \le \mathcal{T}(\mu,\mu)$, then $\limsup_{\epsilon\to 0} c(\mathcal{T}_\epsilon) \le c(\mathcal{T})$. On the other hand, let

$$c(\mathcal{T}_\epsilon) = \inf\{\mathcal{T}_\epsilon(\mu,\mu); \mu \in \mathcal{P}(X)\} = \mathcal{T}_\epsilon(\mu_\epsilon, \mu_\epsilon)$$

for some μ_ϵ. If $\bar{\mu}$ is a cluster point for (μ_ϵ) as $\epsilon \to 0$, the Γ-convergence of $\mathcal{T}_\epsilon$ to $\mathcal{T}$ implies

$$\liminf_{\epsilon\to 0} c(\mathcal{T}_\epsilon) = \liminf_{\epsilon\to 0} \mathcal{T}_\epsilon(\mu_\epsilon, \mu_\epsilon) \ge \mathcal{T}(\bar{\mu}, \bar{\mu}) \ge c(\mathcal{T})$$

which concludes the proof. □

5.5.2 Ergodic properties for non-continuous linear transfers via regularisation

We now present some results for backward linear transfers which are only weak* lower semi-continuous.

Theorem 5.5.3 (Ergodic properties for non-continuous linear transfers)**.** *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains all the Dirac measures. Assume $c(\mathcal{T}) := \inf_{\mu\in\mathcal{P}(X)} \mathcal{T}(\mu,\mu) < +\infty$, and that for some $\epsilon' > 0$ (and thus for all $0 < \epsilon \le \epsilon'$), we have*

$$c(\mathcal{T}_\epsilon) = c(\mathcal{T}), \tag{5.42}$$

where $\mathcal{T}_\epsilon$ is the regularisation of $\mathcal{T}$ according to Lemma 5.5.1.

Then, there exists an idempotent backward Kantorovich operator $T^-_\infty : C(X) \to USC(X)$ such that $T^- \circ T^-_\infty g + c(\mathcal{T}) = T^-_\infty g$ for all $g \in C(X)$. Consequently, its associated backward linear transfer $\mathcal{T}_\infty$ is idempotent and therefore $\mathcal{A}$-factorisable.

Proof. Consider the regularisation $\mathcal{T}_\epsilon$ of $\mathcal{T}$. By Theorem 5.4.3, there exists a Kantorovich operator $T^-_{\infty,\epsilon} : C(X) \to C(X)$, such that $T^-_\epsilon \circ T^-_{\infty,\epsilon} g + c(\mathcal{T}_\epsilon) = T^-_{\infty,\epsilon} g$ for all $g \in C(X)$, and an idempotent transfer $\mathcal{T}_{\infty,\epsilon}$. In fact, from Theorem 5.4.3 we may take $T^-_{\infty,\epsilon} g = \lim_{n\to\infty}[(T^-_\epsilon)_n \circ \bar{T}^-_\epsilon g + nc(\mathcal{T}_\epsilon)]$, where

$\bar{T}_\epsilon^- g(x) := \limsup_n [(T_\epsilon^-)_n g(x) + nc(\mathcal{T}_\epsilon)]$. Under the assumption that $c(\mathcal{T}_\epsilon) = c(\mathcal{T})$, the monotonicity in ϵ for T_ϵ^- (see Lemma 5.5.1) actually extends to monotonicity for $T_{\infty,\epsilon}^-$, in the following sense,

$$T_{\infty,\epsilon}^- g \leq T_{\infty,\epsilon'}^- g \quad \text{for all } g, \text{ whenever } \epsilon < \epsilon'.$$

Define now

$$T_\infty^- g(x) := \inf_{\epsilon>0} T_{\infty,\epsilon}^- g(x) = \lim_{\epsilon\to 0} T_{\infty,\epsilon}^- g(x).$$

As an infimum of continuous functions, $T_\infty^- g$ is upper semi-continuous, but we have to check that it is proper (i.e. not identically $-\infty$).

To see this, recall the monotonicity $\mathcal{T}_{\epsilon'}(\mu,\mu) \leq \mathcal{T}_\epsilon(\mu,\mu) \leq \mathcal{T}(\mu,\mu)$ for $\epsilon \leq \epsilon'$, for all μ. When ϵ' is small enough, the hypothesis implies $c(\mathcal{T}_{\epsilon'}) = c(\mathcal{T}_\epsilon) = c(\mathcal{T})$.

Let now $\bar{\mu}_\epsilon$ achieve $c(\mathcal{T}_\epsilon) = c(\mathcal{T})$. From the monotonicity,

$$\begin{aligned} c(\mathcal{T}) = c(\mathcal{T}_{\epsilon'}) &\leq \mathcal{T}_{\epsilon'}(\bar{\mu}_\epsilon, \bar{\mu}_\epsilon) \\ &\leq \mathcal{T}_\epsilon(\bar{\mu}_\epsilon, \bar{\mu}_\epsilon) = c(\mathcal{T}_\epsilon) = c(\mathcal{T}), \end{aligned}$$

hence $\mathcal{T}_{\epsilon'}(\bar{\mu}_\epsilon, \bar{\mu}_\epsilon) = c(\mathcal{T})$. By Theorem 5.4.3, we have $\mathcal{T}_{\infty,\epsilon'}(\bar{\mu}_\epsilon, \bar{\mu}_\epsilon) = 0$, which implies

$$\int_X g d\bar{\mu}_\epsilon \leq \int_X T_{\infty,\epsilon'}^- g d\bar{\mu}_\epsilon. \tag{5.43}$$

Extract a subsequence ϵ_j of the $\bar{\mu}_\epsilon$ so that $\bar{\mu}_{\epsilon_j} \to \bar{\mu}$. We know by Γ-convergence that $\mathcal{T}(\bar{\mu}, \bar{\mu}) = c(\mathcal{T})$, and (5.43) implies that

$$\int_X g d\bar{\mu} \leq \int_X T_{\infty,\epsilon'}^- g d\bar{\mu}.$$

Now let $\epsilon' \to 0$ to obtain by monotone convergence

$$\int_X g d\bar{\mu} \leq \int_X T_\infty^- g d\bar{\mu}. \tag{5.44}$$

In particular, we deduce that $T_\infty^- g$ is finite for $\bar{\mu}$-a.e. x, so $T_\infty^- g \in USC(X)$

for every $g \in C(X)$.

Regarding the claim that T_∞^- is a backward Kantorovich operator, we have from Theorem 5.4.3 that $T_{\infty,\epsilon}^-$ is a backward Kantorovich operator for every $\epsilon > 0$. Therefore, the monotonicity, convexity, and affine on constants, properties hold in the monotone limit as $\epsilon \to 0$. For the lower semi-continuity, we can return to the proof of Theorem 5.4.3 and write an equivalent inequality as was given in 5.33,

$$T_{\infty,\epsilon}^- g(x) \leq T_{\infty,\epsilon}^- g_k(x) + \|g - g_k\|_\infty.$$

Now let $\epsilon \to 0$ to obtain

$$T_\infty^- g(x) \leq T_\infty^- g_k(x) + \|g - g_k\|_\infty$$

which gives the desired lower semi-continuity property for T_∞^- and shows that T_∞^- is a backward Kantorovich operator.

Regarding the idempotent property for T_∞^-, we can exploit the monotonicity and idempotent properties for the Kantorovich operator $T_{\infty,\epsilon}^-$. This yields

$$T_{\infty,\epsilon}^- g = T_{\infty,\epsilon}^- \circ T_{\infty,\epsilon}^- g \geq T_\infty^- \circ T_\infty^- g$$

so that with $\epsilon \to 0$ we obtain $T_\infty^- g \geq T_\infty^- \circ T_\infty^- g$. For the reverse inequality, we again use montonicity in ϵ to have for $\epsilon \leq \epsilon'$,

$$T_{\infty,\epsilon}^- g = T_{\infty,\epsilon}^- \circ T_{\infty,\epsilon}^- g \leq T_{\infty,\epsilon'}^- \circ T_{\infty,\epsilon}^- g$$

We let $\epsilon \to 0$ and use the monotone convergence Lemma 5.1.3 applied to $T_{\epsilon'}^\infty$, and then let $\epsilon' \to 0$ to achieve

$$T_\infty^- g \leq T_\infty^- \circ T_\infty^- g$$

so that $T_\infty^- g = T_\infty^- \circ T_\infty^- g$.

Define

$$\mathcal{T}_\infty(\mu, \nu) := \sup\{\int_X g d\nu - \int_X T_\infty^- g d\mu \,;\, g \in C(X)\}$$

as the induced linear transfer from T^-_∞. Note that $\mathcal{T}_\infty(\bar\mu,\bar\mu)=0$ by (5.44), so $\mathcal{T}_\infty$ is proper, and $\bar\mu\in\mathcal{A}$. In addition, $\mathcal{T}_\infty$ is idempotent. By the monotone convergence Lemma 5.1.3, and the monotonicity $T^-h\le T^-_\epsilon h$, we have

$$\begin{aligned}T^-\circ T^-_\infty g(x)+c(\mathcal{T})&=\lim_{\epsilon\to 0}T^-\circ T^-_{\infty,\epsilon}g(x)+c(\mathcal{T})\\&\le\lim_{\epsilon\to 0}T^-_\epsilon\circ T^-_{\infty,\epsilon}g(x)+c(\mathcal{T}_\epsilon)=\lim_{\epsilon\to 0}T^-_{\infty,\epsilon}g(x)=T^-_\infty f(x).\end{aligned}$$

On the other hand, the monotonicity again $T^-_\epsilon\le T^-_{\epsilon'}$ for $\epsilon<\epsilon'$ gives

$$T^-_{\infty,\epsilon}g=T^-_\epsilon\circ T^-_{\infty,\epsilon}g+c(\mathcal{T}_\epsilon)\le T^-_{\epsilon'}\circ T^-_{\infty,\epsilon}g+c(\mathcal{T})$$

By Lemma 5.1.3 applied to $T^-_{\epsilon'}$ and the sequence $T^-_{\infty,\epsilon}g$, we can pass the limit in ϵ through T'^-_ϵ to obtain

$$T^-_\infty g(x)=\lim_{\epsilon\to 0}T^-_{\infty,\epsilon}g(x)\le\lim_{\epsilon\to 0}T^-_{\epsilon'}\circ T^-_{\infty,\epsilon}g(x)+c(\mathcal{T})=T^-_{\epsilon'}\circ T^-_\infty g(x)+c(\mathcal{T}).$$

Now we let $\epsilon'\to 0$ and use Property 5 of Lemma 5.5.1 to obtain $T^-_\infty g(x)\le T^-\circ T^-_\infty g(x)+c(\mathcal{T})$, and thus obtaining equality.

□

The hypothesis $c(\mathcal{T}_\epsilon)=c(\mathcal{T})$ in Theorem 5.5.3 seems possibly difficult to check. In the next proposition we give more easily checkable sufficient conditions (see in particular item 1 below) for when this equality holds.

Proposition 5.5.4. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X)\times\mathcal{P}(X)$. 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. *If $\mathcal{T}$ is both a backward and forward linear transfer, is symmetric, and the operator T^- associated to $\mathcal{T}$ maps every continuous function to a L-Lipschitz function ($L>0$),*

we have $c(\mathcal{T}_\epsilon)=c(\mathcal{T})$ for all $\epsilon>0$ small enough; in particular, the equality holds for all $\epsilon>0$ in the first case, and for all $\frac{1}{L}>\epsilon>0$ in the second.

Proof. 1) Note that property 2 of Lemma 5.5.1 says

$$\inf\{\mathcal{T}(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}=\inf\{\mathcal{T}_\epsilon(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}.$$

At the same time, since $c(\mathcal{T}_\epsilon)\le c(\mathcal{T})$ (see property 3 of Lemma 5.5.2), together with the assumption $\inf\{\mathcal{T}(\mu,\nu);\mu,\nu\in\mathcal{P}(X)\}=\inf\{\mathcal{T}(\mu,\mu)\,;\,\mu\in\mathcal{P}(X)\}$, we get

$$\begin{aligned}\inf_{\mu\in\mathcal{P}(X)}\mathcal{T}_\epsilon(\mu,\mu)=c(\mathcal{T}_\epsilon)\le c(\mathcal{T})=\inf_{\mu\in\mathcal{P}(X)}\mathcal{T}(\mu,\mu)&=\inf_{\mu,\nu\in\mathcal{P}(X)}\mathcal{T}(\mu,\nu)\\&=\inf_{\mu,\nu\in\mathcal{P}(X)}\mathcal{T}_\epsilon(\mu,\nu)\\&\le\inf_{\mu\in\mathcal{P}(X)}\mathcal{T}_\epsilon(\mu,\mu)\end{aligned}$$

so all the inequalities are in fact equalities, and therefore $c(\mathcal{T}_\epsilon)=c(\mathcal{T})$.

2) Write

$$\begin{aligned}\mathcal{T}_\epsilon(\mu,\mu)&=\inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W_1(\mu,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\frac{1}{\epsilon}W_1(\sigma_2,\mu)\}\\&\ge\inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W_1(\mu,\sigma_1)-\mathcal{T}(\sigma_1,\sigma_1)+\mathcal{T}(\sigma_1,\sigma_2)+\frac{1}{\epsilon}W_1(\sigma_2,\mu)\}+\inf_{\sigma_1}\{\mathcal{T}(\sigma_1,\sigma_1)\}\\&\ge\inf_{\sigma_1,\sigma_2,\sigma_3}\{\frac{1}{\epsilon}W_1(\mu,\sigma_1)-\mathcal{T}(\sigma_1,\sigma_3)+\mathcal{T}(\sigma_3,\sigma_2)+\frac{1}{\epsilon}W_1(\sigma_2,\mu)\}+\inf_{\sigma_1}\{\mathcal{T}(\sigma_1,\sigma_1)\}\\&=(\frac{1}{\epsilon}W_1)\star(-\mathcal{T})\star\mathcal{T}\star(\frac{1}{\epsilon}W_1)(\mu,\mu)+c(\mathcal{T}).\end{aligned}$$

It suffices to show that $\inf_\mu(\frac{1}{\epsilon}W_1)\star(-\mathcal{T})\star\mathcal{T}\star(\frac{1}{\epsilon}W_1)(\mu,\mu)\ge 0$, since that would imply based on the above inequalities, that $c(\mathcal{T}_\epsilon)\ge c(\mathcal{T})$ (the reverse inequality we already have by Lemma 5.5.2). To this end, first we can write

$$\begin{aligned}(-\mathcal{T})\star\mathcal{T}(\mu,\nu)&=\inf_{\sigma\in\mathcal{P}(X)}\{-\mathcal{T}(\mu,\sigma)+\mathcal{T}(\sigma,\nu)\}\\&=\inf_{g\in C(X)}\inf_{\sigma\in\mathcal{P}(X)}\{\int_X T^-g\,d\mu-\int_X g\,d\sigma+\mathcal{T}(\sigma,\nu)\}\\&=\inf_{g\in C(X)}\{\int_X T^-g\,d\mu+\int_X T^+(-g)\,d\nu\},\end{aligned}$$

where we have used $\inf_{\sigma\in\mathcal{P}(X)}\{\int_X(-g)d\sigma+\mathcal{T}(\sigma,\nu)\}=\int_X T^+(-g)d\nu$ since

$\mathcal{T}$ is a forward linear transfer. Then with the notation that $S_\epsilon^- g(x) := \sup_{y\in X}\{g(y) - \frac{1}{\epsilon}d_X(x,y)\}$ is the backward Kantorovich operator for $\frac{1}{\epsilon}W$, we arrive at

$$
\begin{aligned}
&(\frac{1}{\epsilon}W)\star(-\mathcal{T})\star\mathcal{T}\star(\frac{1}{\epsilon}W)(\mu,\mu)\\
&\qquad = \inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W(\mu,\sigma_1) + (-\mathcal{T})\star\mathcal{T}(\sigma_1,\sigma_2) + \frac{1}{\epsilon}W(\sigma_2,\mu)\}\\
&\qquad = \inf_{g}\inf_{\sigma_1,\sigma_2}\{\frac{1}{\epsilon}W(\mu,\sigma_1) + \int_X T^- g d\sigma_1 + \int_X T^+(-g)d\sigma_2 + \frac{1}{\epsilon}W(\sigma_2,\mu)\}\\
&\qquad = \inf_{g}\{-\int_X S_\epsilon^-(-T^-g)d\mu - \int_X S_\epsilon^-(-T^+(-g))d\mu\}\\
&\qquad = \inf_{g}\{-\int_X S_\epsilon^-(-T^-g)d\mu - \int_X S_\epsilon^-(T^-g)d\mu\} \qquad (5.45)
\end{aligned}
$$

where we have used the fact that $T^-g = -T^+(-g)$ since $\mathcal{T}$ is symmetric. Finally, we make use of the following property: if $h \in C(X)$ is L-Lipschitz for $L < \frac{1}{\epsilon}$, then $S_\epsilon^-(h) = h$. Indeed, to see this, write

$$
\begin{aligned}
S_\epsilon^- h(x) &= \sup_{y\in X}\{h(y) - \frac{1}{\epsilon}d_X(x,y)\}\\
&= \sup_{y\in X}\{h(y) - h(x) - \frac{1}{\epsilon}d_X(x,y)\} + h(x)\\
&\le \sup_{y\in X}\{(L - \frac{1}{\epsilon})d_X(x,y)\} + h(x) = h(x)
\end{aligned}
$$

the reverse inequality following similarly. Hence continuing from 5.45, we have

$$
\begin{aligned}
(\frac{1}{\epsilon}W)\star(-\mathcal{T})\star\mathcal{T}\star(\frac{1}{\epsilon}W)(\mu,\mu) &= \inf_{g}\{-\int_X(-T^-g)d\mu - \int_X T^-gd\mu\}\\
&= 0,
\end{aligned}
$$

$\square$

5.5.3 Examples

The assumption $\inf\{\mathcal{T}(\mu,\nu);\mu,\nu \in \mathcal{P}(X)\} = \inf\{\mathcal{T}(\mu,\mu)\,;\,\mu \in \mathcal{P}(X)\}$ is satisfied by a number of examples of which we present a few in this section.

Example 5.5.5. Let $\mathcal{T}$ be the backward linear transfer associated to the convex energy $I(\nu)$, that is $\mathcal{T}(\mu,\nu) := I(\nu)$ and $T^-g(x) = I^*(g)$ (see Section 2.5.1). It is immediate that $\inf_{\mu,\nu}\mathcal{T}(\mu,\nu) = \inf_\mu \mathcal{T}(\mu,\mu)$ since $\mathcal{T}$ only depends on ν, so $\mathcal{T}$ satisfies the conditions of Theorem 5.5.3.

We have $c(\mathcal{T}) = \inf_\mu \mathcal{T}(\mu,\mu) = \inf_\nu I(\nu)$. The operator $T^-_\infty g(x) := I^*(g) + c(\mathcal{T})$, where I^* is the Legendre transform of I, is an idempotent backward linear transfer that maps into the backward weak KAM solutions for $\mathcal{T}$. Indeed, we can verify

$$\begin{aligned}
T^- \circ T^-_\infty g(x) + c(\mathcal{T}) &= I^*(T^-_\infty g) + c(\mathcal{T}) \\
&= \sup_{\nu\in\mathcal{P}(X)} \{\int_X T^-_\infty g d\nu - I(\nu)\} + c(\mathcal{T}) \\
&= T^-_\infty g - \inf_{\nu\in\mathcal{P}(X)} I(\nu) + c(\mathcal{T}) \quad \text{(since } T^-_\infty g \text{ is a constant in } x) \\
&= T^-_\infty g.
\end{aligned}$$

We can also observe that T^-_∞ is idempotent since $T^-_\infty \circ T^-_\infty g = I^*(T^-_\infty g) + c(\mathcal{T}) = T^-_\infty g$ exactly as the computation above.

The associated idempotent backward linear transfer is $\mathcal{T}_\infty(\mu,\nu) = I(\nu) - c(\mathcal{T})$, with $\mathcal{A} = \{\sigma\,;\, I(\sigma) = \inf_\nu I(\nu)\}$.

Example 5.5.6. Consider the continuous point transformation of Example 2.5.7, i.e.,

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

with $T^-g(x) = g(F(x))$ for a continuous map $F : X \to X$. The Krylov–Bogolyubov theorem (see e.g. [51], Theorem 1.1) then says that there exists a measure $\mu \in \mathcal{P}(X)$ such that $F_\#\mu = \mu$; therefore the hypotheses for Theorem 5.5.3 are readily checked to be satisfied.

Suppose for simplicity that $X \subset \mathbb{R}$. We then claim that

$$T_\infty^- g(x) = \widehat{g \circ F^\infty}(x)$$

is an idempotent backward Kantorovich operator mapping into the set of backward weak KAM solutions, where $\widehat{g \circ F^\infty}$ is the upper semi-continuous envelope of $g \circ F^\infty$, and

$$F^\infty(x) := \limsup_{m \to \infty} F^m(x)$$

where F^m denotes composition of F m-times. Indeed we have

$$\begin{aligned} T^- \circ T_\infty^- g(x) &= \widehat{g \circ F^\infty}(F(x)) \\ &= \inf\{h(F(x))\,;\, h \in USC(X),\, h \geq g \circ F^\infty\} \\ &= \inf\{h(x)\,;\, h \in USC(X),\, h \geq g \circ F^\infty\} \\ &= \widehat{g \circ F^\infty}(x) = T_\infty^- g(x) \end{aligned}$$

where we have used the fact that if $h \in USC(X)$ with $h \geq g \circ F^\infty$, then $h \circ F \in USC(X)$ with $h \circ F \geq g \circ F^\infty$ (here we use the fact that $F^\infty \circ F = F^\infty$).

The upper semi-continuous envelope is in general necessary; consider for example $X = [0, 1]$ and $F(x) = x^2$, then

$$g \mapsto g \circ F^\infty(x) = \begin{cases} g(0) & \text{if } x \in [0, 1) \\ g(1) & \text{if } x = 1, \end{cases}$$

although idempotent and satisfying $T^- \circ (g \circ F^\infty) = g \circ F^\infty$, does not in general belong to $USC(X)$, unless $g(0) \leq g(1)$.

Another idempotent operator mapping into backward weak KAM solutions when $F^m(x)$ has more than one subsequential limit, is

$$g \mapsto g(\liminf_{m \to \infty} F^m(x)) =: g(F_\infty(x)).$$

For example, on $X = [0, 1]$ if $F(x) = 1 - x$, then $F^\infty(x) = \max\{x, 1 - x\}$, while $F_\infty(x) = \min\{x, 1 - x\}$.

Example 5.5.7. Recall Example 2.5.7: With $X = Y$, for any lower semi-continuous $A : X \to \mathbb{R}$, consider $T^- g(x) = g(x) - A(x)$ and the corresponding

$$\mathcal{T}(\mu, \nu) = \begin{cases} \int_X A d\mu & \text{if } \nu = \mu \text{ and } \int_X A d\mu < +\infty, \\ +\infty & \text{otherwise.} \end{cases}$$

We have $\inf_{\mu,\nu \in \mathcal{P}(X)} \mathcal{T}(\mu, \nu) = \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu)$, so $\mathcal{T}$ satisfies the hypotheses of Theorem 5.5.3. Then

$$c(\mathcal{T}) = \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu) = \inf_{x \in X} A(x)$$

and

$$T^-_\infty g(x) = \begin{cases} g(x) & \text{if } A(x) = c(\mathcal{T}) \\ -\infty & \text{otherwise.} \end{cases}$$

We know that $T^-_\infty g$ is proper since A achieves its infimum. To see that $T^-_\infty g \in USC(X)$, note that if $x_k \to x$, then $T^-_\infty g(x_k)$ is either $-\infty$ or $g(x_k)$. In either case, we have $\limsup_{k\to\infty} T^-_\infty g(x_k) \leq g(x)$ by continuity of g. For the lower semi-continuity property of T^-_∞, suppose $g_k \to g$ in $C(X)$ for the sup norm. If $A(x) = c(\mathcal{T})$, then $\liminf_{k\to\infty} T^-_\infty g_k(x) = \liminf_{k\to\infty} g_k(x) = g(x) = T^-_\infty g(x)$; otherwise both $T^-_\infty g_k(x)$ and $T^-_\infty g(x)$ are $-\infty$; in either case $\liminf_{k\to\infty} T^-_\infty g_k(x) \geq T^-_\infty g(x)$.

We can check the backward weak KAM solution,

$$\begin{aligned} T^- \circ T^-_\infty g(x) + c(\mathcal{T}) &= T^-_\infty g(x) - A(x) + c(\mathcal{T}) \\ &= \begin{cases} g(x) - A(x) + c(\mathcal{T}) & \text{if } A(x) = c(\mathcal{T}) \\ -\infty & \text{otherwise.} \end{cases} \\ &= \begin{cases} g(x) & \text{if } A(x) = c(\mathcal{T}) \\ -\infty & \text{otherwise.} \end{cases} \\ &= T^-_\infty g(x). \end{aligned}$$

For the idempotent property, note that

$$T_\infty^- \circ T_\infty^- g(x) = \begin{cases} T_\infty^- g(x) & \text{if } A(x) = c(\mathcal{T}) \\ -\infty & \text{otherwise.} \end{cases} = \begin{cases} g(x) & \text{if } A(x) = c(\mathcal{T}) \\ -\infty & \text{otherwise.} \end{cases} = T_\infty^- g(x).$$

The corresponding backward linear transfer $\mathcal{T}_\infty$ is

$$\mathcal{T}_\infty(\mu,\nu) = \begin{cases} 0 & \text{if } \mu = \nu \text{ is supported on the set } \{x \in X\,;\, A(x) = c(\mathcal{T})\} \\ +\infty & \text{otherwise} \end{cases}$$

and $\mathcal{A} = \{\mu \in \mathcal{P}(X)\,;\, \mathrm{spt}(\mu) \subset \{x\,;\, A(x) = c(\mathcal{T})\}\}$.

Note that if $T^- g(x) = g \circ F(x) - A(x)$ for a continuous A, then the resulting linear transfer

$$\mathcal{T}(\mu,\nu) = \begin{cases} \int_X A d\mu & \text{if } \nu = F_\# \mu, \\ +\infty & \text{otherwise,} \end{cases}$$

in general fails the assumption $\inf_{(\mu,\nu)} \mathcal{T}(\mu,\nu) = \inf_\mu \mathcal{T}(\mu,\mu)$. The latter minimisation, $\inf_\mu \mathcal{T}(\mu,\mu)$, is of interest in ergodic optimisation for expanding dynamical systems. We discuss more on this problem in Section 5.6.1.

Example 5.5.8. Recall the Skorokhod transfer of Example 2.5.8 but now we let the stopping time τ be deterministic, and define, for measures μ, ν, on a compact Riemannian manifold X, the linear transfer,

$$\mathcal{T}(\mu,\nu) = \begin{cases} 0 & \text{if } B_0 \sim \mu \text{ and } B_1 \sim \nu \\ +\infty & \text{otherwise.} \end{cases}$$

Then $T^- g(x) = \mathbb{E}\left[g(B_1) | B_0 = x\right] = P_1 g(x)$, where $\{P_t\}_{t>0}$ is the heat semigroup. Note that $\inf_{\mu,\nu} \mathcal{T}(\mu,\nu) = \inf_\mu \mathcal{T}(\mu,\mu) < +\infty$ since the volume measure λ_X (i.e., the uniform probability measure on X), is invariant, so $\mathcal{T}(\lambda_X, \lambda_X) = 0$ (and in particular $c(\mathcal{T}) = 0$). Therefore the conditions of Theorem 5.5.3 are satisfied.

A corresponding idempotent backward Kantorovich operator is given by

$T_\infty^- g(x) = \int_X g d\lambda_X$. Indeed, we can check that

$$T^- \circ T_\infty^- g(x) + c(\mathcal{T}) = P_1(T_\infty^- g) = T_\infty^- g$$

since $T_\infty^- g$ is constant.

The idempotent backward linear transfer is given by

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

and $\mathcal{A} = \{\mu\,;\, \mathcal{T}_\infty(\mu, \mu) = 0\} = \{\lambda_X\}$.

5.5.4 Interpolation and ergodic properties

We saw in the previous sections that the condition $\inf_{\mu,\nu} \mathcal{T}(\mu, \nu) = \inf_\mu \mathcal{T}(\mu, \mu) < +\infty$ is sufficient to apply Theorem 5.5.3. In this section, we apply a transformation to $\mathcal{T}$ to achieve another backward linear transfer $\tilde{\mathcal{T}}$, that has the property, $\inf_{\mu,\nu} \tilde{\mathcal{T}}(\mu, \nu) = \inf_\mu \tilde{\mathcal{T}}(\mu, \mu) < +\infty$.

Lemma 5.5.9. *Let $S : C(X) \to C(X)$ be a Markov operator (i.e., a bounded linear positive operator such that $S1 = 1$) and let $S^* : \mathcal{M}(X) \to \mathcal{M}(X)$ be its adjoint. Given a backward linear transfer $\mathcal{T}$ on $\mathcal{P}(X) \times \mathcal{P}(X)$ and $\lambda \in (0, 1)$, define*

$$\tilde{\mathcal{T}}(\mu, \nu) := \mathcal{T}(\mu, \lambda S^* \mu + (1 - \lambda)\nu).$$

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

$$\tilde{T}^- g(x) := T^- \left(\frac{1}{1 - \lambda} g\right)(x) - \frac{\lambda}{1 - \lambda} Sg(x)$$

Proof. We first note that $\nu \mapsto \tilde{\mathcal{T}}(\mu, \nu)$ is convex, weak* lower semi-continuous, and bounded below. In particular, to verify convexity, we write for $\beta \in [0, 1]$

and $\nu_1, \nu_2 \in \mathcal{P}(X)$,

$$\begin{aligned}
\tilde{\mathcal{T}}(\mu, (1-\beta)\nu_1 + \beta\nu_2) &= \mathcal{T}(\mu, \lambda S^*\mu + (1-\lambda)((1-\beta)\nu_1 + \beta\nu_2)) \\
&= \mathcal{T}(\mu, (1-\beta)[\lambda S^*\mu + (1-\lambda)\nu_1] + \beta[\lambda S^*\mu + (1-\lambda)\nu_2]) \\
&\leq (1-\beta)\mathcal{T}(\mu, \lambda S^*\mu + (1-\lambda)\nu_1) + \beta\mathcal{T}(\mu, \lambda S^*\mu + (1-\lambda)\nu_2).
\end{aligned}$$

Now we compute,

$$\begin{aligned}
(\tilde{\mathcal{T}}_\mu)^*(g) &= \sup_{\sigma \in \mathcal{P}(X)} \{\int_X g d\sigma - \tilde{\mathcal{T}}(\mu, \sigma)\} \\
&= \sup_{\sigma \in \mathcal{P}(X)} \{\int_X g d\sigma - \mathcal{T}(\mu, \lambda S^*\mu + (1-\lambda)\sigma)\}.
\end{aligned}$$

Let now $\nu := \lambda S^*\mu + (1-\lambda)\sigma$; we obtain $\sigma = \frac{1}{1-\lambda}\nu - \frac{\lambda}{1-\lambda}S^*\mu$. Hence after substitution we obtain

$$\begin{aligned}
(\tilde{\mathcal{T}}_\mu)^*(g) &= \sup_{\nu \in \mathcal{P}(X)} \{\int_X \frac{1}{1-\lambda} g d\nu - \mathcal{T}(\mu, \nu)\} - \frac{\lambda}{1-\lambda}\int_Y S g d\mu \\
&= \int_X \left[T^- \left(\frac{1}{1-\lambda} g\right) - \frac{\lambda}{1-\lambda} S g\right] d\mu.
\end{aligned}$$

□

Theorem 5.5.10. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ where $X \subset \mathbb{R}^n$ is compact, and suppose (μ_0, ν_0) is such that $\mathcal{T}(\mu_0, \nu_0) = \inf_{(\mu,\nu)} \mathcal{T}(\mu, \nu)$ and μ_0 has density with respect to Lebesgue. Then, for every $\lambda \in (0,1)$, there exists a convex function ϕ_λ such that the backward linear transfer given by*

$$\tilde{\mathcal{T}}(\mu, \nu) := \mathcal{T}(\mu, \lambda(\nabla\phi_\lambda)_\# \mu + (1-\lambda)\nu) \tag{5.46}$$

is such that

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

In particular, the conclusion of Theorem 5.5.3 holds for $\tilde{\mathcal{T}}$, and there exists *a backward Kantorovich operator $T^-_\infty : C(X) \to USC(X)$ such that*

$$T^- \circ T^-_\infty g + \mathcal{T}(\mu_0, \nu_0) = \lambda\, (T^-_\infty g) \circ \nabla\phi_\lambda + (1-\lambda)\, T^-_\infty g \quad \textit{for all } g \in C(X).$$

Proof. The fact that $\tilde{\mathcal{T}}$ is a backward linear transfer follows by Lemma 5.5.9. By assumption there is $\mu_0, \nu_0 \in \mathcal{P}(X)$ such that $\mathcal{T}(\mu_0, \nu_0) = \inf_{\mu,\nu} \mathcal{T}(\mu, \nu)$ and μ_0 has density with respect to Lebesgue. Therefore, by Brenier's theorem (see e.g. [59], Corollary 2.30) there exists a convex function ϕ_λ such that $\nabla\phi_{\lambda\#}\mu_0 = (1 - \frac{1}{\lambda})\mu_0 + \frac{1}{\lambda}\nu_0$.

Consider now the backward linear transfer

$$\tilde{\mathcal{T}}(\mu, \nu) := \mathcal{T}(\mu, \lambda\nabla\phi_\lambda\#\mu + (1-\lambda)\nu).$$

Since $\lambda\nabla\phi_\lambda\#\mu_0 + (1-\lambda)\mu_0 = \lambda(1-\frac{1}{\lambda})\mu_0 + \nu_0 + (1-\lambda)\mu_0 = \nu_0$, we have $\tilde{\mathcal{T}}(\mu_0, \mu_0) = \mathcal{T}(\mu_0, \nu_0) < +\infty$, and therefore

$$\inf_{\mu} \tilde{\mathcal{T}}(\mu, \mu) \geq \inf_{\mu,\nu} \tilde{\mathcal{T}}(\mu, \nu) \geq \inf_{\mu,\nu} \mathcal{T}(\mu, \nu) = \mathcal{T}(\mu_0, \nu_0) = \tilde{\mathcal{T}}(\mu_0, \mu_0) \geq \inf_{\mu} \tilde{\mathcal{T}}(\mu, \mu),$$

hence $\inf_{\mu,\nu} \tilde{\mathcal{T}}(\mu, \nu) = \inf_\mu \tilde{\mathcal{T}}(\mu, \mu) = \mathcal{T}(\mu_0, \nu_0)$.. The backward linear transfer $\tilde{\mathcal{T}}$ therefore satisfies the hypotheses of Theorem 5.5.3, so there exists an idempotent backward Kantorovich operator $\tilde{T}^-_\infty$ such

$$\tilde{T}^- \circ \tilde{T}^-_\infty g + c(\tilde{\mathcal{T}}) = \tilde{T}^-_\infty g.$$

By Lemma 5.5.9, $\tilde{T}^- g = T^-(\frac{1}{1-\lambda} g) - \frac{\lambda}{1-\lambda} g \circ \nabla\phi_\lambda$, so that

$$T^-(\frac{1}{1-\lambda}\tilde{T}^-_\infty g) - \frac{\lambda}{1-\lambda}(\tilde{T}^-_\infty g) \circ \nabla\phi_\lambda + c(\tilde{\mathcal{T}}) = \tilde{T}^-_\infty g.$$

In other words, by setting $T^-_\infty g := \frac{1}{1-\lambda}\tilde{T}^-_\infty g$, we have

$$T^- \circ T^-_\infty g + c(\tilde{\mathcal{T}}) = \lambda\, (T^-_\infty g) \circ \nabla\phi_\lambda + (1-\lambda)\, T^-_\infty g.$$

□

5.6 Ergodic properties for non-continuous linear transfers with control on convergence to $c(\mathcal{T})$

We saw in Section 5.5, that under the hypothesis $c(\mathcal{T}_\epsilon) = c(\mathcal{T})$ for small enough $\epsilon > 0$, we obtained certain ergodic properties described in Theorem 5.5.3. We are interested in this section in obtaining results for backward linear transfers that do not satisfy this hypothesis. In particular, we shall focus on the existence of functions h, possibly in $USC_\sigma(X)$, that satisfy $T^- h + c(\mathcal{T}) = h$ (we shall not be concerned with trying to find a backward Kantorovich operator T^-_∞ as in the previous sections).

A preliminary idea for constructing such a function h is very simple: For a backward Kantorovich operator T^-, consider $\varphi := \limsup_{n\to\infty}(T^-_n g + nc(\mathcal{T}))$. Then $\sup_{m\geq n}(T^-_n g + nc(\mathcal{T}))$ is a monotone decreasing sequence, so $T^-\varphi + c(\mathcal{T}) \geq \varphi$, and thus $T^-_n\varphi + nc(\mathcal{T})$ is a monotone increasing sequence, so we can define $h := \lim_{n\to\infty}(T^-_n\varphi + nc(\mathcal{T}))$, and we have $T^- h + c(\mathcal{T}) = h$. One can also sketch a similar construction using the liminf instead of the limsup.

The issue with this construction concerns two things: (i) $\limsup_{n\to\infty}(T^-_n g + nc(\mathcal{T}))$ (or $\liminf_{n\to\infty}(T^-_n g + nc(\mathcal{T}))$) may not be proper, and (ii) these functions are usually not in a "good enough space" (usually only in $USC_\sigma(X)$) to pass the monotone limit through T^- (see Lemma 5.1.3). This section therefore is concerned with finding some additional assumptions for which to deal with (i) and (ii).

We begin with a technical lemma concerning mainly with some preliminary estimates which will help us to deal with issue (i) later in this section. Recall that the constant $c(\mathcal{T}) := \inf_\mu \mathcal{T}(\mu,\mu) = \sup_{n\in\mathbb{N}} \frac{\inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu)}{n} < +\infty$.

Lemma 5.6.1. *Let $\mathcal{T} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be a backward linear transfer such that $\mathcal{D}_1(\mathcal{T})$ contains the Dirac measures. Assume that $c(\mathcal{T}) = \inf_{\mu\in\mathcal{P}(X)} \mathcal{T}(\mu,\mu) < +\infty$. Then, the following properties hold:*

1. *For each $g \in C(X)$ and $\mu \in \mathcal{P}(X)$, we have*

$$-\mathcal{T}(\mu,\mu) \leq \liminf_{n\to\infty} \frac{1}{n}\int_X T_n^- g\, d\mu$$
$$\leq \limsup_{n\to\infty} \frac{1}{n}\int_X T_n^- g d\mu \leq -c(\mathcal{T}).$$

   Consequently if $\mathcal{T}(\bar{\mu},\bar{\mu}) = c(\mathcal{T})$, then for each $g \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n}\int_X T_n^- g\, d\bar{\mu} = c(\mathcal{T}).$$

2. *For every $n \in \mathbb{N}$ and $g \in USC_b(X)$, we have*

$$\sup_{x\in X}(T_n^- g(x) + nc(\mathcal{T})) \geq \inf_{y\in X} g(y). \tag{5.47}$$

3. *Suppose there exists $K > 0$ such that*

$$\limsup_{n\to\infty}\{nc(\mathcal{T}) - \inf_{\mu,\nu\in\mathcal{P}(X)} \mathcal{T}_n(\mu,\nu)\} \leq K. \tag{5.48}$$

   Then for all $g \in C(X)$,

$$\sup_{x\in X}\limsup_{n\to\infty}\left(T_n^- g(x) + nc(\mathcal{T})\right) \leq \sup_{y\in X} g(y) + K. \tag{5.49}$$

Proof. 1)Write for the upper bound,

$$\int_X T_n^- g d\mu = \sup\{\int_X g\ d\nu - \mathcal{T}_n(\mu,\nu)\,;\, \nu \in \mathcal{P}(X)\}$$
$$\leq \sup(g) - \inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu).$$

Dividing by n and recalling that $c(\mathcal{T}) = \lim_{n\to\infty} \frac{\inf_{(\mu,\nu)} \mathcal{T}_n(\mu,\nu)}{n}$ we deduce the stated upper bound for $\limsup_{n\to\infty} \frac{1}{n}\int_X T_n^- g d\mu$. For the lower bound, write

$$\begin{aligned}\int_X T_n^- g d\mu &= \sup\{\int_X g\ d\sigma - \mathcal{T}_n(\mu,\sigma)\,;\,\sigma\in\mathcal{P}(X)\}\\ &\geq \int_X g\,d\mu - \mathcal{T}_n(\mu,\mu)\\ &\geq \int_X g\,d\mu - n\mathcal{T}(\mu,\mu).\end{aligned}$$

2) Write

$$\begin{aligned}\sup_{x\in X}(T_n^- g(x)+nc(\mathcal{T})) &= \sup_{\mu\in\mathcal{P}(X)}\int_X (T_n^- g(x)+nc(\mathcal{T}))\,d\mu\\ &= \sup_{\mu\in\mathcal{P}(X)}\sup_{\sigma\in\mathcal{P}(X)}\{\int_X gd\sigma - \mathcal{T}_n(\mu,\sigma)+nc(\mathcal{T})\}\\ &\geq \inf_X g - \inf_{\sigma,\mu}\mathcal{T}_n(\mu,\sigma)+nc(\mathcal{T})\\ &\geq \inf_X g.\end{aligned}$$

The latter inequality follows from $\inf_{\mu,\sigma}\mathcal{T}_n(\mu,\sigma)\leq nc(\mathcal{T})$ (see Proposition 5.2.1).

3) Write

$$\begin{aligned}\sup_{x\in X}\limsup_{n\to\infty}(T_n^- g(x)+nc(\mathcal{T})) &\leq \limsup_n \sup_{x\in X} T_n^- g(x)+nc(\mathcal{T})\\ &= \limsup_n \sup_{x\in X}\sup_{\sigma\in\mathcal{P}(X)}\{\int_X gd\sigma - \mathcal{T}_n(\delta_x,\sigma)+nc(\mathcal{T})\}\\ &\leq \sup(g)+\limsup_n\{nc(\mathcal{T})-\inf_{x\in X}\inf_{\sigma\in\mathcal{P}(X)}\mathcal{T}_n(\delta_x,\sigma)\}\\ &\leq \sup(g)+\limsup_n\{nc(\mathcal{T})-\inf_{(\mu,\nu)}\mathcal{T}_n(\mu,\nu)\}\\ &\leq \sup(g)+K.\end{aligned}$$

□

The following lemma is perhaps not surprising; the key point is that a monotone decreasing sequence is guaranteed to be proper.

Lemma 5.6.2. *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ such that $c(\mathcal{T}) = \inf_{\mu\in\mathcal{P}(X)} \mathcal{T}(\mu,\mu) < +\infty$ and $\sup_{x\in X} \inf_{\nu\in\mathcal{P}(X)} \mathcal{T}(\delta_x,\nu) < +\infty$. If $g \in USC_b(X)$ is such that $\{T_n^- g + nc(\mathcal{T})\}_{n\in\mathbb{N}}$ is a decreasing sequence in n, then $h(x) := \lim_{n\to\infty}(T_n^- g(x) + nc(\mathcal{T}))$ belongs to $USC(X)$ (in particular, it is proper), and $T^- h + c(\mathcal{T}) = h$.*

Proof. The assumption $\sup_{x\in X} \inf_{\nu\in\mathcal{P}(X)} \mathcal{T}(\delta_x,\nu) < +\infty$ implies that T^- maps $USC_b(X)$ to $USC_b(X)$ (see Lemma 5.1.1). Therefore $T_n^- g + nc(\mathcal{T})$ is a decreasing sequence in $USC_b(X)$, hence converges to its infimum h, which is therefore upper semi-continuous. To see that h is proper, suppose for every x, it holds that $T_n^- g(x) + nc(\mathcal{T})$ decreases to $-\infty$. Then for each x_0, there is $n_0 \in \mathbb{N}$ such that $T_{n_0}^- g(x_0) + n_0 c(\mathcal{T}) \leq \inf_{y\in X} g(y) - 1$. Since $T_{n_0}^- g + nc(\mathcal{T})$ is upper semi-continuous, the inequality for $T_{n_0}^- g(x_0) + n_0 c(\mathcal{T})$ must hold in a neighbourhood of x_0. Since X is compact and $T_n^- g + nc(\mathcal{T})$ is decreasing, it follows that there is $M \in \mathbb{N}$ such that $T_M^- g(x) + Mc(\mathcal{T}) \leq \inf_{y\in X} g(y) - 1$ for all $x \in X$. But from Lemma 5.6.1, we also conclude that

$$\sup_{x\in X} T_M^- g(x) + Mc(\mathcal{T}) \geq \inf_{x\in X} g(x)$$

and the two inequalities together imply

$$\inf_{x\in X} g(x) \leq \inf_{x\in X} g(x) - 1,$$

an impossibility since $\inf_{x\in X} g(x) > -\infty$ as $g \in USC_b(X)$. Therefore we conclude that h is proper and therefore belongs to $USC(X)$. Then by the monotonicity convergence Lemma 5.1.3, we conclude that

$$T^- h + c(\mathcal{T}) = \lim_{n\to\infty} T^-(T_n^- g + nc(\mathcal{T})) + c(\mathcal{T}) = \lim_{n\to\infty}(T_{n+1}^- g + (n+1)c(\mathcal{T})) = h.$$

□

The following theorem is then concerned with finding conditions which guarantee the existence of such a $g \in USC_b(X)$ as in Lemma 5.6.2.

Theorem 5.6.3. *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. Assume:*

1. $c(\mathcal{T}) := \inf_\mu \mathcal{T}(\mu,\mu) < +\infty$,

2. $\sup_{x\in X}\inf_{\nu\in\mathcal{P}(X)} \mathcal{T}(\delta_x,\nu) < +\infty$,

3. *there exists* $K > 0$ *such that*

$$\limsup_{n\to\infty}\{nc(\mathcal{T}) - \inf_{\mu,\nu\in\mathcal{P}(X)} \mathcal{T}_n(\mu,\nu)\} \le K,$$

4. *there exists* $g \in C(X)$, *such that the function*

$$x \mapsto \liminf_{n\to\infty}(T_n^- g(x) + nc(\mathcal{T}))$$

belongs to $USC_b(X)$.

Then there exists $h \in USC(X)$ *such that* $T^-h + c(\mathcal{T}) = h$ *on* X.

Proof. Note that the second hypothesis implies that $T_n^- g \in USC_b(X)$ for all $n \in \mathbb{N}$ and all $g \in USC_b(X)$ (see Lemma 5.1.1).

We distinguish two cases:

Case 1: There is $g \in C(X)$ so that $\forall x \in X$, there exists $n \in \mathbb{N}$ with

$$T_n^- g(x) + nc(\mathcal{T}) < g(x).$$

Since $T_n^- g$ is in $USC_b(X)$, it follows that $T_n^- g + nc(\mathcal{T}) < g$ on an open neighborhood $B_x \subset X$ of x. Indeed, otherwise if there was a sequence $x_k \to x$ with $T_n^- g(x_k) + nc(\mathcal{T}) \ge g(x_k)$, then $T_n^- g(x) + nc(\mathcal{T}) \ge \limsup_k T_n^- g(x_k) + nc(\mathcal{T}) \ge g(x)$, a contradiction.

Since X is compact, there exists a finite number $\{x_1,\dots,x_k\} \subset X$ such that $\{B_{x_j}\}_{1\le j\le k}$ cover X. Set $N := \max\{n_{x_1},\dots,n_{x_k}\}$ and define $\varphi_N(x) := \inf_{1\le n\le N}(T_n^- g(x) + nc(\mathcal{T}))$. We have $\varphi_N \in USC_b(X)$, and moreover note by construction that for any $x \in X$, $\varphi_N(x) \le g(x)$. By the monotonicity property of T^-, we have

$$T^-\varphi_N + c(\mathcal{T}) \le T^- g + c(\mathcal{T}) \quad \text{and} \quad T^-\varphi_N + c(\mathcal{T}) \le \inf_{2\le n\le N+1}\{T_n^- g + nc(\mathcal{T})\}.$$

Therefore, combining the two we deduce that

$$T^- \varphi_N + c(\mathcal{T}) \leq \inf_{1 \leq n \leq N} \{T_n^- g + nc(\mathcal{T})\} = \varphi_N.$$

It follows that the sequence $\{T_n^- \varphi_N + nc(\mathcal{T})\}_n$ is decreasing and $\varphi_N \in USC_b(X)$. Hence by Lemma 5.6.2, we conclude the existence of $h \in USC(X)$, such that $T^- h + c(\mathcal{T}) = h$.

Case 2: We now assume that for any $g \in C(X)$, there exists $x \in X$ such that

$$T_n^- g(x) + nc(\mathcal{T}) \geq g(x) \quad \text{for all } n \in \mathbb{N}. \tag{5.50}$$

By assumption, there exists $g \in C(X)$ such that $\tilde{g} := \liminf_{n\to\infty}(T_n^- g + nc(\mathcal{T}))$ belongs to $USC_b(X)$. Then (5.50) implies there exists $x \in X$ such that $\tilde{g}(x) \geq g(x) > -\infty$, and by Lemma 5.6.1, $\sup_{x \in X} \tilde{g}(x) \leq \sup_{x\in X} g(x) + K$.

Since $\inf_{m \geq n}\{T_m^- g + mc(\mathcal{T})\}$ is increasing to $\tilde{g}$ and $\tilde{g}$ is bounded above, we have by Lemma 5.1.3, that

$$\begin{aligned}
T^- \tilde{g} + c(\mathcal{T}) &= \lim_{n\to\infty} T^-\big(\inf_{m \geq n}\{T_m^- g + mc(\mathcal{T})\}\big) + c(\mathcal{T}) \\
&\leq \liminf_{n\to\infty} (T_{n+1}^- g + (n+1)c(\mathcal{T})) \\
&= \tilde{g}.
\end{aligned}$$

It follows that the sequence $\{T_n^- \tilde{g} + nc(\mathcal{T})\}_{n\in\mathbb{N}} \subset USC_b(X)$ is decreasing and $\tilde{g} \in USC_b(X)$. Consequently, we may apply Lemma 5.6.2 to conclude that there exists a function $h \in USC(X)$ such that $T^- h + c(\mathcal{T}) = h$.

□

The following provides a sufficient condition for $\limsup_{n\to\infty}\{nc(\mathcal{T}) - \inf_{\mu,\nu\in\mathcal{P}(X)} \mathcal{T}_n(\mu,\nu)\} \leq K$ of the previous theorem.

Proposition 5.6.4. *Let $\mathcal{T}$ be a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ such that $c(\mathcal{T}) = \inf_{\mu\in\mathcal{P}(X)} \mathcal{T}(\mu,\mu) < +\infty$. If $\mathcal{T}$ is bounded above on $\mathcal{P}(X)\times$*

$\mathcal{P}(X)$, *then*

$$\frac{\mathcal{T}_n(\mu,\nu)}{n} \to c(\mathcal{T}) \quad \textit{uniformly on } \mathcal{P}(X)\times\mathcal{P}(X). \tag{5.51}$$

In particular, we have $\limsup_{n\to\infty}\{nc(\mathcal{T}) - \inf_{\mu,\nu\in\mathcal{P}(X)} \mathcal{T}_n(\mu,\nu)\} \le K$ *for some constant* $K > 0$.

Proof. Note that $\{\frac{\sup_{\mu,\nu}\mathcal{T}_n(\mu,\nu)}{n}\}_{n\in\mathbb{N}}$ is a sub-additive sequence that satisfies,

$$\frac{\sup_{\mu,\nu}\mathcal{T}_n(\mu,\nu)}{n} \ge \inf_{\mu,\nu}\mathcal{T}(\mu,\nu) > -\infty$$

hence converges to its infimum (see e.g. [9] Lemma 1.18). Since

$$c(\mathcal{T}) = \lim_{n\to\infty}\frac{\inf_{\mu,\nu}\mathcal{T}_n(\mu,\nu)}{n} \quad \text{and} \quad \frac{\inf_{\mu,\nu}\mathcal{T}_n(\mu,\nu)}{n} \le \frac{\sup_{\mu,\nu}\mathcal{T}_n(\mu,\nu)}{n}$$

we conclude that

$$c(\mathcal{T}) \le \inf_{n\in\mathbb{N}}\frac{\sup_{\mu,\nu}\mathcal{T}_n(\mu,\nu)}{n}.$$

Therefore, for all $n \in \mathbb{N}$, it holds that

$$\inf_{\mu,\nu}\mathcal{T}_n(\mu,\nu) \le nc(\mathcal{T}) \le \sup_{\mu,\nu}\mathcal{T}_n(\mu,\nu),$$

so

$$|\mathcal{T}_n(\mu,\nu) - nc(\mathcal{T})| \le \sup_{\mu,\nu}\mathcal{T}_n(\mu,\nu) - \inf_{\mu,\nu}\mathcal{T}_n(\mu,\nu). \tag{5.52}$$

At the same time, we have for any $\mu,\nu\in\mathcal{P}(X)$ (writing $\inf\limits_{\mathcal{P}\times\mathcal{P}}\mathcal{T}_n$ for $\inf_{\mu,\nu\in\mathcal{P}(X)}\mathcal{T}_n(\mu,\nu)$ for brevity),

$$\inf_{\mathcal{P}\times\mathcal{P}}\mathcal{T}_{n-2} + 2\inf_{\mathcal{P}\times\mathcal{P}}\mathcal{T} \le \mathcal{T}_n(\mu,\nu) \le 2\sup_{\mathcal{P}\times\mathcal{P}}\mathcal{T} + \inf_{\mathcal{P}\times\mathcal{P}}\mathcal{T}_{n-2},$$

from which follows that

$$\begin{aligned}\sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n - \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_n &\leq 2 \sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T} + \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_{n-2} - \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T}_{n-2} - 2 \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T} \\ &= 2 \sup_{\mathcal{P}\times\mathcal{P}} \mathcal{T} - 2 \inf_{\mathcal{P}\times\mathcal{P}} \mathcal{T} \\ &=: K < \infty. \end{aligned} \tag{5.53}$$

Combining (5.52) and (5.53), we conclude that

$$|\mathcal{T}_n(\mu,\nu) - nc(\mathcal{T})| \leq K \quad \text{for all } \mu,\nu \in \mathcal{P}(X) \text{ and all } n \in \mathbb{N},$$

which implies $\frac{\mathcal{T}_n(\mu,\nu)}{n} \to c(\mathcal{T})$ uniformly on $\mathcal{P}(X) \times \mathcal{P}(X)$.

□

Here is another situation where we can obtain weak KAM solutions.

Proposition 5.6.5. *Suppose $\mathcal{T}$ is a backward linear transfer on $\mathcal{P}(X) \times \mathcal{P}(X)$ such that $c(\mathcal{T}) = \inf_{\mu\in\mathcal{P}(X)} \mathcal{T}(\mu,\mu) < +\infty$, and $\sup_{x\in X} \inf_{\nu\in\mathcal{P}(X)} \mathcal{T}(\delta_x,\nu) < +\infty$. If there exists $u, v \in USC_b(X)$ such that*

$$T_n^- u + nv = u \quad \textit{for all } n \in \mathbb{N}, \tag{5.54}$$

then there exists $h \in USC(X)$ such that $T^- h + c(\mathcal{T}) = h$ on X, where T^- is the backward Kantorovich operator associated to $\mathcal{T}$.

Proof. By (5.54) and Lemma 5.6.1, we have $-v(x) = \limsup_n \frac{T_n^- u}{n} \leq -c(\mathcal{T})$. Therefore,

$$T_n^- u + nc(\mathcal{T}) \leq u \quad \text{for all } n \in \mathbb{N}.$$

Applying T_m^- and using the linearity of T_m^- with respect to constants, we find $T_{m+n}^- u + nc(\mathcal{T}) \leq T_m^- u$, and hence

$$T_{m+n}^- u + (m+n)c(\mathcal{T}) \leq T_m^- u + mc(\mathcal{T}).$$

So $n \mapsto T_n^- u + nc(\mathcal{T})$ is decreasing, and $u \in USC_b(X)$. Consequently by Lemma 5.6.2, there exists $h \in USC(X)$ such that $T^- h + c(\mathcal{T}) = h$.

□

5.6.1 Linear transfers and ergodic optimization

This section was developed jointly with Dorian Martino [40]. 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). In this section, we shall see that the presence of A allows the theory of transfers to incorporate ergodic optimization for expanding dynamical systems (see, for instance, [25]). For simplicity, we shall focus here on the case where the linear Markov operator is given by a point transformation σ.

We will in this section make use of the Fenchel-Rockafellar duality, and thus for convenience we record the statement of this theorem here.

Theorem 5.6.6 (Fenchel-Rockafellar duality (see e.g. [59] Theorem 1.9))**.** *Let E be a normed vector space, E^* its topological dual space, and $h_1, h_2 : E \to \mathbb{R} \cup \{+\infty\}$ two convex functions on E. If there exists $z_0 \in E$ such that*

$$h_1(z_0) < +\infty,\ h_2(z_0) < +\infty, \quad \textit{and } h_1 \textit{ is continuous at } z_0,$$

then

$$\inf_{z \in E}\{h_1(z) + h_2(z)\} = \sup_{z^* \in E^*}\{-h_1^*(-z^*) - h_2(z^*)\}$$

where h_1^*, h_2^*, is the Fenchel-Legendre transform of h_1, h_2, respectively,

$$h_1^*(z^*) := \sup_{z \in E}\{\langle z^*, z\rangle - h_1(z)\}.$$

Proposition 5.6.7. *Let $\sigma : X \to X$ be continuous and surjective, and assume there is a (sequentially) 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 $x \in X_y$.*

Define $Y_x := \{y \in Y\, ;\, y \in X_y\}$, and assume the following continuity properties: If x_k is a sequence in X (resp. y_k a sequence in Y) with $y_k \in Y_{x_k}$, then if $x_k \to x$ and $y_k \to y$, we have $y \in Y_x$, and $\tau_{y_k}(x_k) \to \tau_y(x)$. Moreover, *assume that the maps τ_y are unique, in the sense that if $\tau_y(x) = \tau_{y'}(x)$ for all x, then $y' = y$.*

Let $A \in C(Y \times X)$ be a continuous function and consider the cost function $c : X \times X \to \mathbb{R} \cup \{+\infty\}$ defined by

$$c(z,x) := \begin{cases} \inf\{A(y,x)\,;\, y \in Y_x, \tau_y(x) = z\} & \text{if } \sigma(z) = x \\ +\infty & \text{otherwise.} \end{cases}$$

Then, c is lower semi-continuous, and the optimal mass transport $\mathcal{T}$ associated to the cost c is simply the backward, and forward, linear transfer

$$\mathcal{T}(\mu,\nu) = \begin{cases} \int_X \bar{A} d\mu & \text{if } \nu = \sigma_{\#}\mu \\ +\infty & \text{otherwise,} \end{cases}$$

which has backward (resp. forward) Kantorovich operator given by

$$T^- g(x) = g(\sigma(x)) - \bar{A}(x), \quad (resp., \quad T^+ f(x) = \inf_{y \in Y_x} \{f(\tau_y(x)) + A(y,x)\}).$$

where $\bar{A}(x) := c(x, \sigma(x))$.

Proof. To see that c is lower semi-continuous, suppose $x_k \to x$ and $z_k \to z$. If for all but finitely many (z_k, x_k) we have $c(z_k, x_k) = +\infty$, then there is nothing to prove. Therefore assume $\sigma(z_k) = x_k$, and hence

$$c(z_k, x_k) = \inf\{A(y, x_k)\,;\, y \in Y_{x_k}, \tau_y(x_k) = z_k\}$$

Since A is continuous and $\{y \in Y_{x_k}\,;\, \tau_y(x_k) = z_k\}$ is a closed subset of the compact set Y_{x_k}, the infimum is achieved by some y_k, i.e.

$$c(z_k, x_k) = A(y_k, x_k)$$

The sequence $\{y_k\} \subset Y$ and Y is sequentially compact, so we may select a convergent subsequence (that we relabel back to y_k). This y has the property that $y \in Y_x$ and $\tau_y(x) = z$ by assumption. Consequently,

$$\liminf_{k\to\infty} c(z_k, x_k) = A(y, x) \geq \inf\{A(y, x)\,;\, y \in Y_x, \tau_y(x) = z\} = c(z, x).$$

We now compute the optimal transport

$$\mathcal{T}(\mu, \nu) = \inf_{\pi\in\mathcal{K}(\mu,\nu)} \int_{X\times X} c(z, x)d\pi(z, x).$$

If $\nu \neq \sigma_{\#}\mu$, then $\pi \in \mathcal{K}(\mu, \nu)$ will give non-zero mass to a region $\{(z, x)\}$ where $\sigma(z) \neq x$. Since $c(z, x) = +\infty$ there, we deduce that $\mathcal{T}(\mu, \nu) = +\infty$ in this case. Otherwise assume $\nu = \sigma_{\#}\mu$. Then $\pi \in \mathcal{K}(\mu, \nu)$ is supported on $\{(z, \sigma(z))\,;\, z \in X\}$, so

$$\begin{aligned}
\mathcal{T}(\mu, \nu) &= \inf_{\pi\in\mathcal{K}(\mu,\nu)} \{\int_{X\times X} c(z, \sigma(z))d\pi(z, x)\} \\
&= \int_X c(z, \sigma(z))d\mu \\
&= \int_X \bar{A}d\mu.
\end{aligned}$$

We know from Section 2.6 that the backward (resp. forward) Kantorovich operator is given by

$$T^- g(z) = \sup_{x\in X}\{g(x) - c(z, x)\} = g(\sigma(z)) - c(z, \sigma(z)) = g \circ \sigma(z) - \bar{A}(z).$$

and

$$T^+ f(x) = \inf_{z\in X}\{f(z) + c(z, x)\} = \inf_{y\in Y_x}\{f(\tau_y(x)) + c(\tau_y(x), x)\}.$$

The final observation to make for T^+ is that

$$c(\tau_y(x), x) = \inf\{A(y', x)\,;\, y' \in Y_x\, \tau_{y'}(x) = \tau_y(x)\}$$

hence by assumption $y' = y$, so that $c(\tau_y(x), x) = A(y, x)$. $\square$

Theorem 5.6.8. *In the set-up described in the previous Proposition 5.6.7,* *let $\mathcal{P}_\sigma(X)$ denote the subset of probability measures in $\mathcal{P}(X)$ which are invariant under σ. Let $\hat{X}$ denote the subset of $Y \times X$ consisting of points (y, x) such that $x \in X_y$, and define $\mathcal{M}_0(\hat{X})$ as the subset of "holonomic" probability measures $\hat{\mu} \in \mathcal{P}(\hat{X})$ satisfying $\int_{\hat{X}} [f(\tau_y(x)) - f(x)] d\hat{\mu} = 0$.*

Then, the following duality formulae holds:

$$
\begin{aligned}
c(\mathcal{T}) := \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu) &= \inf\{\int_X \bar{A}(x)\, d\mu(x); \mu \in \mathcal{P}_\sigma(X)\} \\
&= \inf\{\int_{\hat{X}} A(y, x) d\hat{\mu}(y, x)\,;\, \hat{\mu} \in \mathcal{M}_0(\hat{X})\} \\
&= \sup_{f \in C(X)} \inf_{x \in X} \{f(x) - f(\sigma(x)) + \bar{A}(x)\} \\
&= \sup_{f \in C(X)} \inf_{x \in X} \inf_{y \in Y_x} \{f(\tau_y(x)) - f(x) + A(y, x)\}.
\end{aligned}
$$

Remark 5.6.9. The equality of the second and fourth lines has already been established in [26] for the setting of symbolic dynamics (see the next section).

Proof. Let us first demonstrate the equality

$$
\begin{aligned}
\inf\{\int_X \bar{A}(x)\, d\mu(x); \mu \in \mathcal{P}_\sigma(X)\} &= \sup_{f \in C(X)} \inf_{x \in X} \{f(x) - f(\sigma(x)) + \bar{A}(x)\} \\
&= \sup_{f \in C(X)} \inf_{x \in X} \{f(x) - T^- f(x)\}. \qquad (5.55)
\end{aligned}
$$

We can write using the duality of optimal transport,

$$
\begin{aligned}
\inf\{\int_X \bar{A}(x)\, d\mu(x); \mu \in \mathcal{P}_\sigma(X)\} &= \inf_{\mu \in \mathcal{P}(X)} \inf_{\pi \in \mathcal{K}(\mu, \mu)} \int_{X \times X} c(x, y) d\pi(x, y) \\
&= \inf_{\mu \in \mathcal{P}(X)} \sup_{f \in C(X)} \{\int_X (f - T^- f) d\mu\} \quad (5.56)
\end{aligned}
$$

Comparing (5.55) and (5.56), it suffices to show that

$$
\sup_{f \in C(X)} \inf_{x \in X} \{f(x) - T^- f(x)\} = \inf_{\mu \in \mathcal{P}(X)} \sup_{f \in C(X)} \{\int_X (f - T^- f) d\mu\}.
$$

We would therefore like to apply Sion's minimax theorem (Theorem 2.7.6) to the function $F : \mathcal{P}(X) \times C(X)$ defined by $F(\mu, f) := \int_X (f - T^- f) d\mu$, which is real-valued for all $f \in C(X)$ and all $\mu \in \mathcal{P}(X)$.

First, $\mu \mapsto F(\mu, f)$ is weak* lower semi-continuous since $f - T^- f = f - f \circ \sigma + \bar{A}$ is a lower semi-continuous function for each $f \in C(X)$. Moreover $\mu \mapsto F(\mu, f)$ is quasi-convex on $\mathcal{P}(X)$, i.e. $\{\mu \in \mathcal{P}(X)\,;\, F(\mu, f) \leq \lambda\}$ is convex or empty for all $\lambda \in \mathbb{R}$.

On the other hand, $f \mapsto F(\mu, f)$ is upper semi-continuous, since if $f_k \to f$ in $C(X)$, then

$$\begin{aligned}\limsup_{k\to\infty} \int_X (f_k - T^- f_k) d\mu &\leq \int_X (f - \liminf_{k\to\infty} T^- f_k) d\mu \\ &\leq \int_X (f - T^- f) d\mu.\end{aligned}$$

Moreover, $f \mapsto F(\mu, f)$ is quasi-concave on $C(X)$, i.e. $\{f \in C(X)\,;\, F(\mu, f) \geq \lambda\}$ is convex or empty for all $\lambda \in \mathbb{R}$.

We may therefore apply Sion's minimax theorem (Theorem (2.7.6)) to conclude that

$$\begin{aligned}\inf_{\mu\in\mathcal{P}(X)} \sup_{f\in C(X)} \int_X (f - T^- f) d\mu &= \sup_{f\in C(X)} \inf_{\mu\in\mathcal{P}(X)} \int_X (f - T^- f) d\mu \\ &= \sup_{f\in C(X)} \inf_{x\in X} \int_X (f - T^- f) d\mu.\end{aligned}$$

Next let us demonstrate the equality,

$$\inf\{\int_{\hat{X}} A(y, x) d\hat{\mu}(y, x)\,;\, \hat{\mu} \in \mathcal{M}_0(\hat{X})\} = \sup_{f\in C(X)} \inf_{x\in X} \inf_{y\in Y_x} \{f(\tau_y(x)) - f(x) + A(y, x)\}$$

As mentioned, this has already been established in [26] for the setting of symbolic dynamics (see Theorem 1 there). For this, we shall use Fenchel-Rockafellar duality. Recall $\hat{X} := \{(y, x)\,;\, x \in X, y \in X_y\}$, and define $h_1, h_2 : C(\hat{X}) \to \mathbb{R} \cup \{+\infty\}$ be defined by

$$h_1(\phi) := \sup_{(y,x)\in\hat{X}} \{-\phi(y, x) + A(y, x)\},$$

and

$$h_2(\phi) = \begin{cases} 0 & \text{if } \phi \text{ is in the closure of } \{g \in C(\hat{X})\,;\, g(y,x) = f(x) - f \circ \tau_y(x) \text{ for some } f \in C(X)\} \\ +\infty & \text{otherwise.} \end{cases}$$

Note that $h_1(\phi) - h_1(\tilde{\phi}) \leq \|\phi - \tilde{\phi}\|_\infty$, so h_1 is continuous on $C(\hat{X})$. To compute their respective Legendre transforms, we have

$$h_1^*(\hat{\mu}) = \sup_{\phi \in C(\hat{X})} \{\int_{\hat{X}} \phi d\hat{\mu} - h_1(\phi)\}.$$

If $\hat{\mu}(\hat{X}) \neq -1$, then

$$h_1^*(\hat{\mu}) \geq \sup_{\lambda \in \mathbb{R}} \{\lambda(\hat{\mu}(\hat{X}) + 1) + \inf_{(y,x)\in\hat{X}} A(y,x)\}$$

and the supremum is $+\infty$. Suppose now that $\hat{\mu}(\hat{X}) = -1$, but $-\hat{\mu} \notin \mathcal{P}(\hat{X})$. Then there exists a sequence of functions $\phi_n \in C(\hat{X})$ such that $\phi_n \leq 0$ and $\lim_{n\to\infty} \int_{\hat{X}} \phi_n d\hat{\mu} = +\infty$. Then we have

$$\begin{aligned} h_1^*(\hat{\mu}) &\geq \int_{\hat{X}} \phi_n d\hat{\mu} - h_1(\phi_n) \\ &\geq \int_{\hat{X}} \phi_n d\hat{\mu} - \sup_{(y,x)\in\hat{X}} A(y,x) \end{aligned}$$

hence $h_1^*(\hat{\mu}) = +\infty$. Finally suppose $-\hat{\mu} \in \mathcal{P}(\hat{X})$. Then

$$\begin{aligned} h_1^*(\hat{\mu}) &\geq \int_{\hat{X}} A d\hat{\mu} - h_1(A) \\ &= \int_{\hat{X}} A d\hat{\mu}, \end{aligned}$$

while also since $\phi + h_1(\phi) \geq A$, then $\int_{\hat{X}} (\phi + h_1(\phi)) d\hat{\mu} \leq \int_{\hat{X}} A d\hat{\mu}$ (recall

$\hat{\mu}(\hat{X}) = -1$), so we have

$$\begin{aligned} h_1^*(\hat{\mu}) &= \sup_{\phi \in C(\hat{X})} \{ \int_{\hat{X}} (\phi(y,x) + h_1(\phi)) d\hat{\mu} \} \\ &\leq \int_{\hat{X}} A d\hat{\mu}. \end{aligned}$$

Therefore,

$$h_1^*(\hat{\mu}) = \begin{cases} \int_{\hat{X}} A d\hat{\mu} & \text{if } -\hat{\mu} \in \mathcal{P}(\hat{X}) \\ +\infty & \text{otherwise.} \end{cases}$$

We also have

$$\begin{aligned} h_2^*(\hat{\mu}) &= \sup_{\phi \in C(\hat{X})} \{ \int_{\hat{X}} \phi d\hat{\mu} - h_2(\phi) \} \\ &= \sup_{f \in C(X)} \{ \int_{\hat{X}} (f - f \circ \tau_y) d\hat{\mu} \}. \end{aligned}$$

If $\int_{\hat{X}} (f - f \circ \tau_y) d\hat{\mu} \neq 0$ for some f, then substituting f with λf, $\lambda \in \mathbb{R}$, implies this supremum will be $+\infty$. Let $\mathcal{S}_0 := \{\hat{\mu} \in \mathcal{M}(\hat{X})\,;\, \int_{\hat{X}} (f \circ \tau_y(x) - f(x)) d\hat{\mu}(y,x) = 0 \text{ for all } f \in C(X)\}$. Then,

$$h_2^*(\hat{\mu}) = \begin{cases} 0 & \text{if } \hat{\mu} \in \mathcal{S}_0 \\ +\infty & \text{otherwise.} \end{cases}$$

It now suffices to apply Fenchel-Rockafellar:

$$\inf_{\phi \in C(\hat{X})} \{ h_1(\phi) + h_2(\phi) \} = \sup_{\hat{\mu} \in \mathcal{M}(\hat{X})} \{ -h_1^*(-\hat{\mu}) - h_2(\hat{\mu}) \}.$$

Finally, to complete the string of equalities stated in the theorem, the equality $c(\mathcal{T}) = \inf_{\mu \in \mathcal{P}(X)} \mathcal{T}(\mu, \mu) = \inf\{\int_X \bar{A}(x)\, d\mu(x); \mu \in \mathcal{P}_\sigma(X)\}$ is immediate from Proposition 5.6.7. In addition, we observe by definition of

$\bar{A}(x) = c(x, \sigma(x)$ that for any $f \in C(X)$,

$$\begin{aligned}\inf_{x\in X}\{f(x) - f(\sigma(x)) + \bar{A}(x)\} &= \inf_{x\in X} \inf_{y\in Y_{\sigma(x)}, \tau_y(\sigma(x))=x} \{f(x) - f(\sigma(x)) + A(y, \sigma(x))\} \\ &= \inf_{z\in X} \inf_{y\in Y_z} \{f(\tau_y(z)) - f(z) + A(y, z)\},\end{aligned}$$

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. It therefore follows that

$$\sup_{f\in C(X)} \inf_{x\in X}\{f(x) - f(\sigma(x)) + \bar{A}(x)\} = \sup_{f\in C(X)} \inf_{x\in X} \inf_{y\in Y_x} \{f(\tau_y(x)) - f(x) + A(y, x)\}$$

which concludes the demonstration of the equalities stated in the theorem.

□

Theorem 5.6.10. *Suppose that*

$$x \mapsto \liminf_{n\to\infty} \sum_{k=0}^{n-1} \inf_{\mu\in\mathcal{P}_\sigma(X)} \int_X \bar{A} d\mu - \bar{A}(\sigma^k(x))$$

belongs to $C(X)$. Then there exists $h \in USC(X)$ such that

$$h(\sigma(x)) - \bar{A}(x) + c(\mathcal{T}) = h(x) \quad \textit{for all } x \in X, \tag{5.57}$$

equivalently,

$$\inf_{y\in Y_x} \{h(\tau_y(x)) + A(y, x)\} - c(\mathcal{T}) = h(x) \quad \textit{for all } x \in X. \tag{5.58}$$

That is, h satisfies $T^- h + c(\mathcal{T}) = h$, and $T^+ h - c(\mathcal{T}) = h$.

Proof. To establish the existence of a function satisfying (5.57), note that this is equivalent to having a function h such that $T^- h(x) + c(\mathcal{T}) = h(x)$, hence it suffices to show that the assumptions of Theorem 5.6.3 are satisfied. For that, first note that by a theorem of Bogolyubov and Krylov ([51], Theorem 1.1), σ has an invariant measure, hence $c(\mathcal{T}) = \inf_\mu \mathcal{T}(\mu, \mu) < +\infty$.

In addition, we have for each $x \in X$,

$$\sup_{x\in X} \inf_{\nu\in\mathcal{P}(X)} \mathcal{T}(\delta_x, \nu) \le \sup_{x\in X} \bar{A}(x) < +\infty.$$

In order to show the condition (5.48), we let for each $n \in \mathbb{N}$, $\mu_n \in \mathcal{P}(X)$ be such that $\mathcal{T}_n(\mu_n, (\sigma^n)_\sharp\mu_n) = \inf_{\mu,\nu} \mathcal{T}_n(\mu, \nu)$. Select a subsequence (which we relabel back to n) so that $\nu_n := (\sigma^n)_\sharp\mu_n$ converges to some $\bar{\mu} \in \mathcal{P}(X)$. Then the Césaro average $\frac{1}{n}\sum_{k=1}^{n} \nu_k$ converges to $\bar{\mu}$. Moreover, $\bar{\mu}$ is σ-invariant. Indeed,

$$\sigma_\sharp\bar{\mu} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}(\sigma^k)_\sharp\mu_n = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=0}^{n-1}(\sigma^k)_\sharp\mu_n + \frac{1}{n}((\sigma^n)_\sharp\mu_n - \mu_n)\right) = \bar{\mu}.$$

Recall that $\bar{A}(x) = c(x, \sigma(x))$ is lower semi-continuous, so

$$\limsup_{n\to\infty} \int_X \bar{A}d\bar{\mu} - \frac{1}{n}\sum_{k=0}^{n-1}\int_X \bar{A}d((\sigma^k)_\sharp\mu_n) \le 0.$$

Therefore, by selecting a further subsequence (which again we may relabel back to n), we may assume that this subsequence has the property that there exists $K > 0$ such that

$$\int_X \bar{A}d\bar{\mu} - \frac{1}{n}\sum_{k=0}^{n-1}\int_X \bar{A}d((\sigma^k)_\sharp\mu_n) \le \frac{K}{n}, \quad n \in \mathbb{N}.$$

We then obtain the desired estimate,

$$\begin{aligned}
&\limsup_{n\to\infty}\left(nc(\mathcal{T})-\inf_{\mu,\nu\in\mathcal{P}(X)}\mathcal{T}_n(\mu,\nu)\right)\\
&\quad\le \limsup_{n\to\infty}\left(n\mathcal{T}(\bar{\mu},\bar{\mu})-\mathcal{T}_n(\mu_n,(\sigma^n)_\sharp\mu_n)\right)\\
&\quad= \limsup_{n\to\infty}\left(n\mathcal{T}(\bar{\mu},\bar{\mu})-\sum_{k=0}^{n-1}\mathcal{T}((\sigma^k)_\sharp\mu_n,(\sigma^{k+1})_\sharp\mu_n)\right)\\
&\quad= \limsup_{n\to\infty}\left(n\int_X \bar{A}d\bar{\mu}-\sum_{k=0}^{n-1}\int_X \bar{A}d((\sigma^k)_\sharp\mu_n)\right)\\
&\quad\le K.
\end{aligned}$$

Finally note that for $g\equiv 0$,

$$T_n^-(0)+nc(\mathcal{T})=\sum_{k=0}^{n-1}(\inf_{\mu\in\mathcal{P}_\sigma(X)}\int_X \bar{A}d\mu-\bar{A}\circ\sigma^k).$$

Since by assumption,

$$x\mapsto \liminf_{n\to\infty}\sum_{k=0}^{n-1}(\inf_{\mu\in\mathcal{P}_\sigma(X)}\int_X \bar{A}d\mu-\bar{A}\circ\sigma^k(x))$$

belongs to $C(X)$, it follows that $\liminf_{n\to\infty}T_n^-(0)+nc(\mathcal{T})$ belongs to $C(X)$. Hence all the hypotheses for application of Theorem 5.6.3 hold, and we deduce that there exists $h\in USC(X)$ such that

$$T^-h+c(\mathcal{T})=h.$$

$\square$

Remark 5.6.11. A conjecture is that the assumption

$$x\mapsto \liminf_{n\to\infty}\sum_{k=0}^{n-1}\left(\inf_{\mu\in\mathcal{P}_\sigma(X)}\int_X \bar{A}d\mu-\bar{A}(\sigma^k(x))\right)$$

belongs to $C(X)$, holds when A is Hölder continuous. This matches exactly the condition on A required by Garibaldi and Lopes ([26], Theorem 4) - see the next section.

5.6.2 Ergodic optimization in the deterministic holonomic setting

We now apply the result of the previous section to the setting of symbolic dynamics. Fix $r \in \mathbb{N}$, and let M be an $r \times r$ transition matrix, whose entries are either 0 or 1, specifying the allowable transitions. Denote by

$$\Sigma = \{x \in \{1, ..., r\}^{\mathbb{N}} \ ; \ \forall i \geq 0, \ M(x_i, x_{i+1}) = 1\}$$

the set of admissible words, its dual

$$\Sigma^* = \{y \in \{1, ..., r\}^{\mathbb{N}} \ ; \ \forall i \geq 0, \ M(y_{i+1}, y_i) = 1\},$$

and consider the space

$$\hat{\Sigma} = \{(y, x) \in \Sigma^* \times \Sigma \ ; \ M(y_0, x_0) = 1\}.$$

For each $x \in \Sigma$, we let $\Sigma^*_x = \{y \in \Sigma^* \ ; \ (y, x) \in \hat{\Sigma}\}$ 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

$$\sigma(x_0, x_1, ...) = (x_1, x_2, ...) \quad \text{and} \quad \tau(y, x) = (y_0, x_0, x_1, ...). \tag{5.59}$$

The map σ can be considered as the time-evolution operator. We will denote $\tau(y, x)$ by $\tau_y(x)$ and consider the set of holonomic probability measures

$$\mathcal{M}_0(\hat{\Sigma}) := \left\{\mu \in \mathcal{P}(\hat{\Sigma}) \ ; \ \int_{\hat{\Sigma}} f(\tau_y(x)) - f(x) \ d\mu(y, x) = 0\right\}.$$

E. Garibaldi and A. O. Lopes studied an Aubry-Mather theory for symbolic dynamics [26]; in particular, they prove the following results.

Theorem 5.6.12 (Garibali and Lopes [26], Theorem 1, Theorem 4)**.** *Under the set-up described above, given $A \in C(\hat{\Sigma})$, define $\beta_A := \max_{\hat{\mu} \in \mathcal{M}_0(\hat{\Sigma})} \int_{\hat{\Sigma}} A(y,x)d\hat{\mu}(y,x)$. Then*

$$\beta_A = \inf_{f \in C(\Sigma)} \max_{(y,x) \in \hat{\Sigma}} \{A(y,x) + f(x) - f(\tau_y(x))\}.$$

If $A \in C^{0,\theta}(\hat{\Sigma})$ is θ-Hölder continuous, then there exists a function $u \in C^{0,\theta}(\Sigma)$ such that

$$u(x) = \min_{u \in \Sigma_x^*} \{u(\tau_y(x)) - A(y,x) + \beta_A\}.$$

By applying our results of the previous Section 5.6.1 in this symbolic dynamics setting, we obtain the following, using the notation β_A of Theorem 5.6.12.

Proposition 5.6.13. *Under the set-up described at the beginning of this section, given $A \in C(\hat{\Sigma})$, then the following hold:*

$$\begin{aligned}\beta_A &= \sup_{\mu \in \mathcal{P}_\sigma(\Sigma)} \{\int_\Sigma \bar{A} d\mu\} &(5.60)\\ &= \inf_{f \in C(\Sigma)} \sup_{x \in \Sigma} \{f(\sigma(x)) - f(x) + \bar{A}(x)\} &(5.61)\\ &= \inf_{f \in C(\Sigma)} \sup_{(y,x) \in \hat{\Sigma}} \{f(x) - f(\tau_y(x)) + A(y,x)\}, &(5.62)\end{aligned}$$

where $\bar{A}(x) := \sup\{A(y,x)\,;\, y \in \Sigma_x^*, \tau_y(\sigma(x)) = x\}$.

Moreover, there exists $h \in USC(\Sigma)$ such that

$$\inf_{y \in \Sigma_x^*} \{h(\tau_y(x)) - A(y,x) + \beta(A)\} = h(x) \quad \forall x \in \Sigma. \tag{5.63}$$

Proof. We will apply Proposition 5.6.7, Theorem 5.6.8, and Theorem 5.6.10, with $-A$ instead of A, and with the following identifications:

$$X := \Sigma, \quad Y := \Sigma^*, \quad Y_x := \Sigma_x^*, \quad \hat{X} := \hat{\Sigma},$$

and σ, τ_y, as in (5.59). $\square$

5.6.3 Ergodic optimization in the stochastic holonomic setting

We now propose the following model:

Construct inductively, the following sequence: Let $X^0 \in \Sigma$ be a random word, $B^0 \in \Sigma^*_{X^0}$ a random "noise", and let $\bar{X}^0 := \tau_{B^0}(X^0)$. We then let $Y^0 \in \Sigma^*_{\bar{X}^0}$ be a random "control", and consider $X^1 := \tau_{Y^0}(\bar{X}^0)$. We then choose $B^1 \in \Sigma^*_{X^1}$ and let $Y^1 \in \Sigma^*_{\bar{X}^1}$.

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.

We assume that B^0, Y^0, satisfy the following "martingale-type" property:

$$\mathbb{E}[f(Y^0, \tau_{B^0}(X^0))|X^0 = \sigma(x)] = \mathbb{E}[f(Y^0, x)] \quad \text{for any } f \in C(\hat{\Sigma}). \tag{5.64}$$

For each such trajectory, consider for each $n \in \mathbb{N}$, the measure $\mu_n \in \mathcal{P}(\hat{\Sigma})$ defined via

$$\int_{\hat{\Sigma}} \phi d\mu_n := \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[\phi(Y^i, \bar{X}^i)], \quad \phi \in C(\hat{\Sigma}).$$

From $(\mu_n)_n$, one can extract a subsequence converging to some measure $\mu^{(X^i)_i}$ by compactness of $\mathcal{P}(\hat{\Sigma})$. We denote

$$\mathcal{M}_0 = \overline{\{\mu^{(X^i)_i}\}} \subset \mathcal{M}(\hat{\Sigma}),$$

as the closure of all such $\mu^{(X^i)_i}$. 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 also

$$\mathcal{N}_0 = \{\mu \in \mathcal{M}(\hat{\Sigma}) \mid \forall f \in C(\Sigma), \int_{\hat{\Sigma}} D^y f(x) d\mu(y,x) = 0\},$$

which is closed in $\mathcal{M}(\hat{\Sigma})$ as a kernel of a continuous linear map.

Lemma 5.6.14. *We have $\mathcal{M}_0 \subset \mathcal{N}_0$.*

Proof. Each measure $\mu^{(X^i)_i} \in \mathcal{M}_0$,

$$\begin{aligned}\int_{\hat{\Sigma}} D^y f(x) d\mu_n(y,x) &= \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1} f(X^{i+1}) - f(X^i)\right] \\ &= \frac{1}{n}\mathbb{E}\left[f(X^n) - f(X^0)\right]\end{aligned}$$

which tends to zero as $n \to \infty$ since f is bounded. $\square$

Theorem 5.6.15. *With the above notation, we have the following*

$$\inf_{\mu \in \mathcal{N}_0 \cap \mathcal{P}(\hat{\Sigma})} \int_{\hat{\Sigma}} A d\hat{\mu} = \sup_{f \in C(\Sigma)} \inf_{(y,x) \in \hat{\Sigma}} D^y f(x) + A(y,x). \tag{5.65}$$

If

$$x \mapsto \liminf_{n\to\infty} \left(n \inf_{\hat{\mu} \in \mathcal{M}_0 \cap \mathcal{P}(\hat{\Sigma})} \int_{\hat{\Sigma}} A d\hat{\mu} - \inf\{\sum_{k=0}^{n-1} \mathbb{E}[A(Y^k, \tau_{B^k}(\bar{Y}^k)]\,;\, Y^k \in \Sigma^*_{\tau_{B^k}(\bar{Y}^k)}\} \right)$$

belongs to $C(\Sigma)$, where $\bar{Y}^k$ is defined recursively via $\bar{Y}^0 = x$, $\bar{Y}^k = \tau_{Y^{k-1}}(\tau_{B^{k-1}}(\bar{Y}^{k-1}))$, $k \geq 1$, then there exists an $g \in USC(\Sigma)$ such that $h := -g$ satisfies

$$\inf_{\mu \in \mathcal{N}_0 \cap \mathcal{P}(\hat{\Sigma})} \int_{\hat{\Sigma}} A d\hat{\mu} = \inf_{y \in \Sigma^*_x} \{D^y h(x) + A(y,x)\}. \tag{5.66}$$

Proof. The last equality in (5.65), i.e., is again an application of the Fenchel-Rockafellar duality, similar to the previous section. Indeed, consider the functions $h_1, h_2 : C(\hat{\Sigma}) \to \mathbb{R} \cup \{+\infty\}$ defined by

$$h_1(\phi) := \sup_{(y,x)\in\hat{\Sigma}} \{-\phi(y,x) + A(y,x)\}$$

and

$$h_2(\phi) := \begin{cases} 0 & \text{if } \phi \in \overline{\{g \in C(\hat{\Sigma})\,;\, g(y,x) = D^y f(x) \text{ for some } f \in C(\Sigma)\}} \\ +\infty & \text{otherwise.} \end{cases}$$

Note that h_1 and h_2 are convex, with $|h_1(\phi) - h_1(\tilde{\phi})| \le \|\phi - \tilde{\phi}\|_\infty$. To compute their Legendre transform, we have

$$h_1^*(\hat{\mu}) = \sup_{\phi \in C(\hat{\Sigma})} \{\int_{\hat{\Sigma}} \phi d\hat{\mu} - h_1(\phi)\}.$$

We have

$$h_1^*(\hat{\mu}) \ge \sup_{\lambda\in\mathbb{R}} \{(\lambda+1)\hat{\mu}(\hat{\Sigma}) + \inf_{(y,x)\in\hat{\Sigma}} \{A(y,x)\}$$

and therefore if $\hat{\mu}(\hat{\Sigma}) \neq -1$, the supremum is $+\infty$. Suppose now that $\hat{\mu}(\hat{\Sigma}) = -1$, but $-\hat{\mu} \notin \mathcal{P}(\hat{\Sigma})$. Then there exists a sequence of functions $\phi_n \in C(\hat{\Sigma})$ such that $\phi_n \le 0$ and $\lim_{n\to\infty} \int_{\hat{\Sigma}} \phi_n d\hat{\mu} = +\infty$. Then

$$\begin{aligned} h_1^*(\hat{\mu}) &\ge \int_{\hat{\Sigma}} \phi_n d\hat{\mu} - h_1(\phi_n) \\ &\ge \int_{\hat{\Sigma}} \phi_n d\hat{\mu} - \sup_{(y,x)\in\hat{\Sigma}} A(y,x) \end{aligned}$$

hence $h_1^*(\hat{\mu}) = +\infty$. Finally suppose $-\hat{\mu} \in \mathcal{P}(\hat{\Sigma})$. Then we have

$$h_1^*(\phi) \ge \int_{\hat{\Sigma}} A d\hat{\mu} - h_1(A) = \int_{\hat{\Sigma}} A d\hat{\mu}$$

while on the other hand from $\phi + h_1(\phi) \ge A$, we have $\int_{\hat{\Sigma}} (\phi - h_1(\phi)) d\hat{\mu} \le$

$\int_{\hat\Sigma} A d\hat\mu$ (recall $\hat\mu(\hat\Sigma) = -1$), so that

$$\begin{aligned} h_1^*(\phi) &= \sup_{\phi\in C(\hat\Sigma)} \{\int_{\hat\Sigma} (\phi - h_1(\phi)) d\hat\mu\} \\ &\le \sup_{\phi\in C(\hat\Sigma)} \{\int_{\hat\Sigma} (\phi - \phi - A) d\hat\mu\} = \int_{\hat\Sigma} A d\hat\mu. \end{aligned}$$

Consequently,

$$h_1^*(\hat\mu) = \begin{cases} \int_{\hat\Sigma} A d\hat\mu & \text{if } \hat\mu \in \mathcal{P}(\hat\Sigma), \\ +\infty & \text{otherwise.} \end{cases}$$

Similarly, we have

$$h_2^*(\hat\mu) = \begin{cases} 0 & \text{if } \hat\mu \in \mathcal{N}_0 \\ +\infty & \text{otherwise.} \end{cases}$$

Indeed, if $\hat\mu \notin \mathcal{N}_0$, there exists $\bar f \in C(\Sigma)$ such that $\int_{\hat\Sigma} D^y \bar f(x) d\hat\mu(y,x) \neq 0$. Hence, we replacing $\bar f$ with $\lambda \bar f$, $\lambda \in \mathbb{R}$, we see that

$$h_2^*(\hat\mu) = \sup_{f\in C(\Sigma)} \int_{\hat\Sigma} D^y f(x) d\hat\mu(y,x) = +\infty.$$

If $\hat\mu \in \mathcal{N}_0$, by definition, $h_2^*(\hat\mu) = 0$.

Fenchel-Rockafellar then says that

$$\inf_{\phi\in C(\hat\Sigma)} \{h_1(\phi) + h_2(\phi)\} = \sup_{\hat\mu\in\mathcal{M}(\hat\Sigma)} \{-h_1^*(-\hat\mu) - h_2^*(\hat\mu)\},$$

which completes the proof of (5.65).

We now provide the proof of (5.66): Consider the functional $\mathcal{T} : \mathcal{P}(\Sigma) \times \mathcal{P}(\Sigma) \to \mathbb{R} \cup \{+\infty\}$ defined via

$$\mathcal{T}(\mu,\nu) := \inf \left\{ \mathbb{E}[A(Y^0, \bar X^0)] \;\middle|\; \begin{array}{ll} X^0 \sim \nu & \\ \bar X^0 = \tau_{B^0}(X^0) & B^0 \in \Sigma^*_{X_0} \\ X^1 = \tau_{Y^0}(\bar X^0) \sim \mu & Y^0 \in \Sigma^*_{\bar X^0} \end{array} \right\}$$

We claim that $\mathcal{T}$ is a forward linear coupling with

$$T^+f(x) := \inf\{\mathbb{E}[f(\tau_{Y^0}(\tau_{B^0}(x)) + A(Y^0, \tau_{B^0}(x))] \,;\, Y^0 \in \Sigma^*_{\tau_{B^0}(x)}\}.$$

Indeed, we have

$$\begin{aligned}\int_\Sigma T^+f d\nu &= \int_\Sigma \Big(\inf\{\mathbb{E}[f(\tau_{Y^0}(\tau_{B^0}(x)) + A(Y^0, \tau_{B^0}(x))] \,;\, Y^0 \in \Sigma^*_{\tau_{B^0}(x)}\}\Big)\, d\nu \\ &= \inf\{\mathbb{E}[f(\tau_{Y^0}(\tau_{B^0}(X^0)) + A(Y^0, \tau_{B^0}(X^0))] \,;\, Y^0 \in \Sigma^*_{\tau_{B^0}(X^0)},\, X^0 \sim \nu\}.\end{aligned}$$

Note then that $\sup_{f\in C(\Sigma)}\{\int_\Sigma T^+f d\nu - \int_\Sigma f d\mu\}$ will be $+\infty$, unless $\tau_{Y^0}(\tau_{B^0}(X^0)) \sim \mu$, in which case the terms in f cancel, leaving

$$\begin{aligned}\sup_{f\in C(\Sigma)}\{\int_\Sigma T^+f d\nu - \int_\Sigma f d\mu\} &= \inf\{\mathbb{E}[A(Y^0, \tau_{B^0}(X^0))] \,;\, Y^0 \in \Sigma^*_{\tau_{B^0}(X^0)},\, X^0 \sim \nu,\, \tau_{Y^0}(\tau_{B^0}(X^0)) \sim \mu\} \\ &= \mathcal{T}(\mu, \nu).\end{aligned}$$

We now show that the hypotheses for application of Theorem 5.6.3 to the (backward) linear coupling $\tilde{\mathcal{T}}(\mu, \nu) := \mathcal{T}(\nu, \mu)$, 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

$$\sup_{x\in\Sigma} \inf_{\nu\in\mathcal{P}(\Sigma)} \tilde{\mathcal{T}}(\delta_x, \nu) \le \sup_{x\in\Sigma} \mathcal{T}(\bar{\nu}_x, \delta_x) \le \sup_{x\in\Sigma} \mathbb{E}[A(Y^0, \tau_{B^0}(x))] \le \sup_{\hat{\Sigma}} A < +\infty.$$

For the hypothesis, $\tilde{\mathcal{T}}(\mu, \mu) < +\infty$, and the condition 5.48, they follow similarly as in the proof of Theorem 5.6.10. We finally need, in order to satisfy the hypotheses of Theorem 5.6.3, the existence of a function $g \in C(\Sigma)$ such that $\liminf_{n\to\infty}(\tilde{T}^-_n g + nc(\mathcal{T}))$ belongs to $USC_b(\Sigma)$, where we recall that $\tilde{T}^-$ is given by

$$\tilde{T}^- g(x) = -T^+(-g)(x) = \sup\{\mathbb{E}[g(\tau_{Y^0}(\tau_{B^0}(x))) - A(Y^0, \tau_{B^0}(x))] \,;\, Y^0 \in \Sigma^*_{\tau_{B^0}(x)}\}$$

Again, similar to the deterministic case, we can take $g \equiv 0$. Then

$$\tilde{T}_n^-(0)(x) = -\inf\{\sum_{k=0}^{n-1} \mathbb{E}[A(Y^k, \tau_{B^k}(\bar{Y}^k)]\,;\, Y^k \in \Sigma^*_{\tau_{B^k}(\bar{Y}^k)}\}$$

where $\bar{Y}^k$ is defined recursively via $\bar{Y}^0 = x$, $\bar{Y}^k = \tau_{Y^{k-1}}(\tau_{B^{k-1}}(\bar{Y}^{k-1}))$, $k \geq 1$. At the same time,

$$\begin{aligned}
c(\mathcal{T}) = \inf_{\mu} \mathcal{T}(\mu,\mu) &= \lim_{n\to\infty} \frac{1}{n} \inf_{\mu,\nu} \mathcal{T}_n(\mu,\nu) \\
&= \lim_{n\to\infty} \inf_{\mu,\nu} \frac{1}{n} \sum_{i=0}^{n-1} \inf\{\mathbb{E}[A(Y^i, \bar{X}^i)]\} \\
&= \lim_{n\to\infty} \inf_{\mu,\nu} \inf_{X^0\sim\mu, X^1\sim\nu} \int_{\hat{\Sigma}} A d\mu_n^{(Y^i)_i} \\
&= \inf_{\hat{\mu}\in\mathcal{M}_0\cap\mathcal{P}(\hat{\Sigma})} \int_{\hat{\Sigma}} A d\hat{\mu}.
\end{aligned}$$

Hence

$$\tilde{T}_n^-(0)(x) + nc(\mathcal{T}) = n \inf_{\hat{\mu}\in\mathcal{M}_0\cap\mathcal{P}(\hat{\Sigma})} \int_{\hat{\Sigma}} A d\hat{\mu} - \inf\{\sum_{k=0}^{n-1} \mathbb{E}[A(Y^k, \tau_{B^k}(\bar{Y}^k)]\,;\, Y^k \in \Sigma^*_{\tau_{B^k}(\bar{Y}^k)}\}.$$

The hypothesis of the theorem ensures that at the liminf as $n \to \infty$, we get a continuous function. Therefore we have satisfied the hypotheses of Theorem 5.6.3, and conclude the existence of a $h \in USC(\Sigma)$ such that

$$\tilde{T}^- h(x) + c(\tilde{\mathcal{T}}) = h(x), \quad \forall x \in \Sigma,$$

which since $\tilde{T}^- h = -T^+(-h)$ and $c(\tilde{\mathcal{T}}) = c(\mathcal{T})$, implies

$$T^+(-h) - c(\mathcal{T}) = -h,$$

or with $g := -h$,

$$T^+ g(x) - c(\mathcal{T}) = g(x). \tag{5.67}$$

Replacing x with $\sigma(x)$ for $x \in \Sigma$ in equation (5.67), we have

$$(T^+ g)(\sigma(x)) - g(\sigma(x)) = c(\mathcal{T})$$

Recalling the definition of T^+ and the martingale assumption (5.64), we can write

$$\begin{aligned}
c(\mathcal{T}) = (T^+ g)(\sigma(x)) - g(\sigma(x)) &= \inf_{Y^0} \{\mathbb{E}[g(\tau_{Y^0}(\tau_{B^0}(\sigma(x))) + A(Y^0, \tau_{B^0}(\sigma(x)))]\} - g(\sigma(x)) \\
&= \inf_{Y^0} \{\mathbb{E}[g(\tau_{Y^0}(x)) + A(Y^0, x)]\} - g(\sigma(x)) \\
&= \inf_{y \in \Sigma_x^*} \{g(\tau_y(x)) - g(\sigma(x)) + A(y, x)\} \\
&= \inf_{y \in \Sigma_x^*} \{D^y g(x) + A(y, x)\} \qquad (5.68)
\end{aligned}$$

Finally, the corresponding Mañé constant is given by

$$\begin{aligned}
c(\mathcal{T}) = \inf_{\mu} \mathcal{T}(\mu, \mu) &= \lim_{n \to \infty} \frac{1}{n} \inf_{\mu, \nu} \mathcal{T}_n(\mu, \nu) \\
&= \lim_{n \to \infty} \inf_{\mu, \nu} \frac{1}{n} \sum_{i=0}^{n-1} \inf\{\mathbb{E}[A(Y^i, \bar{X}^i)]\} \\
&= \lim_{n \to \infty} \inf_{\mu, \nu} \inf_{X^0 \sim \mu, X^1 \sim \nu} \int_{\hat{\Sigma}} A d\mu_n^{(Y^i)_i} \\
&= \inf_{\hat{\mu} \in \mathcal{M}_0 \cap \mathcal{P}(\hat{\Sigma})} \int_{\hat{\Sigma}} A d\hat{\mu}.
\end{aligned}$$

Since $\mathcal{M}_0 \subset \mathcal{N}_0$, it follows that $c(\mathcal{T}) \geq \inf_{\hat{\mu} \in \mathcal{N}_0 \cap \mathcal{P}(\Sigma)} \int_{\hat{\Sigma}} A d\hat{\mu}$. In view of the duality (5.65) together with (5.68), this implies that this inequality is actually an equality, hence (5.66) holds and concludes the proof.

□

Remark 5.6.16. Similar to the deterministic case, our conjecture is that when A is Hölder continuous, the hypothesis of Theorem 5.6.15 holds.

Bibliography

[1] J. Alibert, G. Bouchitte, and T. Champion. A new class of costs for optimal transport planning. *European Journal of Applied Mathematics*, 30(6):1229–1263, 2019. → pages 13, 33, 34, 41, 107

[2] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization*, volume 17. Society for Industrial and Applied Mathematics, Philadelphia, second edition, 2014. ISBN 1611973473;9781611973471;. → page 8

[3] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):1–28, 2019. → pages 33, 34, 70

[4] A. Barton and N. Ghoussoub. Dynamic and stochastic propagation of the Brenier optimal mass transport. *European Journal of Applied Mathematics*, 30(6):1264–1299, 2019. → pages 67, 68

[5] P. Bernard and B. Buffoni. Optimal mass transportation and Mather theory. *Journal of the European Mathematical Society*, 9(1):85–121, 2007. → pages 30, 31, 85, 102, 134, 144

[6] P. Bernard and B. Buffoni. Weak KAM pairs and Monge-Kantorovich duality. *Advanced Studies in Pure Mathematics*, 47(2):397–420, 2007. → pages 30, 85, 102, 103, 104, 105, 122, 132, 134, 144

[7] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control: the discrete time case*, volume 139. Academic Press, New York, 1978. → page 25

[8] S. G. Bobkov and F. Götze. Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities. *Journal of Functional Analysis*, 163(1):1–28, 1999. → pages 73, 74

[9] R. Bowen and J. Chazottes. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470. Springer, Berlin, 2nd rev. edition, 2008. → pages 133, 170

[10] M. Bowles and N. Ghoussoub. Mather measures and ergodic properties of kantorovich operators associated to general mass transfers, 2019. → pages v, 13

[11] Y. Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.

[12] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991. → page 32

[13] Y. Brenier. The dual least action problem for an ideal, incompressible fluid. *Archive for Rational Mechanics and Analysis*, 122(4):323–351, 1993.

[14] Y. Brenier. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Communications on Pure and Applied Mathematics*, 52(4):411–452, 1999.

[15] F. Cagnetti, D. Gomes, and H. Tran. Aubry-Mather measures in the non convex setting. 2010. → page 85

[16] G. Choquet, J. E. Marsden, T. Lance, and S. S. Gelbart. *Lectures on analysis*. W.A. Benjamin, New York, 1969. ISBN 9780805369601. → page 27

[17] K. J. Ciosmak. Optimal transport and Choquet theory. 2020. → page 25

[18] D. Cordero-Erausquin and B. Klartag. Moment measures. *Journal of Functional Analysis*, 268(12):3834–3866, 2015. → pages 80, 81

[19] L. Evans. A survey of partial differential equations methods in weak KAM theory. *Communications on Pure and Applied Mathematics*, 57 (4):445–480, 2004. → page 85

[20] A. Fathi. Weak KAM Theorem in Lagrangian Dynamics. Version X, 2018. → pages 32, 89, 144, 145

[21] A. Fathi. Weak KAM Theory : The connection between Aubry-Mather theory and viscosity solutions of the Hamilton-Jacobi equation. 2018. → page 85

[22] M. Fathi, N. Gozlan, and M. Prodhomme. A proof of the Caffarelli contraction theorem via entropic regularization, 2019. → page 65

[23] A. Figalli and L. Rifford. Aubry sets, Hamilton-Jacobi equations, and the Mañé Conjecture, 2012. → page 85

[24] J. Fontbona, N. Gozlan, and J. Jabir. A variational approach to some transport inequalities. *Annales de L'Institut Henri Poincaré - Probabilites et Statistiques*, 53(4):1719–1746, 2017. → page 71

[25] E. Garibaldi. *Ergodic Optimization in the Expanding Case: Concepts, Tools and Applications*. Springer International Publishing, Cham, 1st 2017 edition, 2017. ISBN 3319666436. → pages 23, 172

[26] E. Garibaldi and A. Lopes. On the Aubry–Mather theory for symbolic dynamics. *Ergodic Theory and Dynamical Systems*, 28(3):791–815, 2008. → pages 175, 176, 182, 183

[27] I. Gentil, C. Léonard, and L. Ripani. About the analogy between optimal transport and minimal entropy. *Annales de la Faculté des Sciences de Toulouse. Mathématiques*, 3:569–600, 2017. → page 47

[28] N. Ghoussoub, Y.-H. Kim, and A. Z. Palmer. Optimal transport with controlled dynamics and free end times. *SIAM Journal on Control and Optimization*, 56(5):3239–3259, 2018. → page 51

[29] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *Annals of Probability*, 47(1):109–164, 2019. → page 46

[30] N. Ghoussoub, Y.-H. Kim, and T. Lim. Optimal brownian stopping when the source and target are radially symmetric distributions. 2019. → pages 27, 50

[31] N. Ghoussoub, Y.-H. Kim, and A. Z. Palmer. A solution to the Monge transport problem for Brownian martingales. 2019. → pages 49, 50, 51, 126, 127

[32] D. Gomes. Viscosity solution methods and the discrete Aubry-Mather problem. *Discrete and Continuous Dynamical Systems*, 13(1):103–116, 2005. → page 85

[33] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen Theorems. *Proceedings of the London Mathematical Society*, 120(3): 434–463, 2020. → page 65

[34] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Processes and Related Fields*, 16:635–736, 2010. → pages 72, 73

[35] N. Gozlan, C. Roberto, and P.-M. Samson. From concentration to logarithmic Sobolev and Poincaré inequalities. *Journal of Functional Analysis*, 260(5):1491–1522, 2011. → pages 20, 57

[36] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *Journal of Functional Analysis*, 273(11):3327–3405, 2017. → pages 13, 33, 45, 65

[37] P. Henry-Labordere. *Model-free hedging: A martingale optimal transport viewpoint*. Chapman & Hall, London, 1 edition, 2017. ISBN 1351666231. → page 46

[38] M. Ledoux. *The concentration of measure phenomenon*, volume 89;no. 89.;. American Mathematical Society, Providence, R.I, 2001. ISBN 0821828649;9780821828649;. → page 73

[39] P. Lions, G. Papanicolaou, and S. Varadhan. Homogenization of Hamilton-Jacobi equations. 1987. → page 88

[40] D. Martino. personal communication. → page 172

[41] K. Marton. Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *The Annals of Probability*, 24 (2):857–866, 1996. ISSN 00911798. URL http://www.jstor.org/stable/2244952. → pages 44, 72

[42] K. Marton. A measure concentration inequality for contracting Markov chains. *Geometric and Functional Analysis*, 6(3):556–571, 1996. → pages 44, 72

[43] J. N. Mather. Action minimizing invariant measures for positive definite Lagrangian systems. *Mathematische Zeitschrift*, 207(1): 169–207, 1991. → page 144

[44] B. Maurey. Some deviation inequalities. *Geometric and Functional Analysis*, 1(2):188–197, 1991. → page 84

[45] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic Processes and their Applications*, 116(12):1815–1835, 2006. → page 49

[46] R. R. Phelps. *Convex functions, monotone operators and differentiability*, volume 1364;. Springer-Verlag, New York;Berlin;, 2nd edition, 1989;1993;. ISBN 3540507353. → page 70

[47] V. Roos. Variational and viscosity operators for the evolutive Hamilton-Jacobi equation. *Communications in Contemporary Mathematics*, 2017. → page 13

[48] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87. Springer International Publishing, Cham, 1st 2015.;1st 2015; edition, 2015. → pages 28, 148

[49] F. Santambrogio. Crowd motion and evolution PDEs under density constraints. *ESAIM: ProcS*, 64:137–157, 2018. doi:10.1051/proc/201864137. URL https://doi.org/10.1051/proc/201864137. → page 58

[50] E. Schechter. *Handbook of analysis and its foundations*. Academic Press, San Diego, 1997;1996;. ISBN 9780126227604. → page 129

[51] Y. G. Sinai. *Dynamical Systems II: Ergodic Theory with Applications to Dynamical Systems and Statistical Mechanics*, volume 2. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989. ISBN 9783662067888. → pages 157, 179

[52] M. Sion. On general minimax theorems. *Pacific journal of mathematics*, 8(1):171–176, 1958. → page 39

[53] A. V. Skorokhod. *Studies in the theory of random processes*, volume 7021;. Addison-Wesley Pub. Co, Reading, Mass, 1965. → page 50

[54] A. Sorrentino. Lecture notes on Mather's theory for Lagrangian systems. 2010. → page 85

[55] A. Sorrentino. *Action-minimizing Methods in Hamiltonian Dynamics: An Introduction to Aubry-Mather Theory*. Princeton University Press,

2015. ISBN 9780691164502. URL
http://www.jstor.org/stable/j.ctt1b9x11w. → page 85

[56] V. Strassen. The Existence of Probability Measures with Given Marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965. → page 25

[57] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geometric and Functional Analysis*, 6(3):587–600, 1996. → page 72

[58] Treschev, Dmitry and Zubelevich, Oleg. Introduction to the KAM Theory. In *Introduction to the Perturbation Theory of Hamiltonian Systems*, pages 23–58, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-03028-4. doi:10.1007/978-3-642-03028-4_2. URL https://doi.org/10.1007/978-3-642-03028-4_2. → page 88

[59] C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Society, Providence, R.I, 2003. ISBN 9780821833124. → pages 28, 29, 163, 172

[60] C. Villani. *Optimal transport: old and new*, volume 338. Springer, Berlin, 2009;2008;. ISBN 9783540710493. → pages 8, 28

[61] C. E. Wayne. An introduction to KAM theory. In *Dynamical Systems and Probabilistic Methods in Partial Differential Equations*, pages 3–29, 1994. → page 88