MASTERARBEIT | MASTER'S THESIS

Titel | Title

Imposing Convexity on Dual Formulations of Weak Transport Problems

verfasst von | submitted by

dipl. mat. (UN) Filip Pramenković

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von | Supervisor:	emer. o. Univ.-Prof. Mag. Dr. Walter Schachermayer

Zusammenfassung

Das Ziel dieser Arbeit ist es, Bedingungen zu untersuchen, die es uns erlauben, die duale Formulierung von optimalen schwachen Transportproblemen auf konvexe Funktionen zu beschränken.

In Kapitel 1 erinnern wir uns an die klassische Theorie des optimalen Transports und erzählen auch ein paar Resultate aus der deskriptiven Mengenlehre. Wir schließen das Kapitel mit einer kleinen Einführung in den optimalen schwachen Transport ab.

Kapitel 2 beinhaltet eine detaillierte Untersuchung von zwei motivierenden Beispielen. Wir betrachten den Kontext und die Beweise beider.

Der Hauptsatz wird in Kapitel 3 gezeigt, und basiert auf den zwei motivierenden Beispielen. Genauer gesagt, wir verstärken den Beweis des ersten Beispiels genug, dass das zweite Beispiel auch umfasst wird.

Das Ziel Kapitels 4 ist es nicht nur rigoros zu überprüfen, dass die zwei motivierenden Beispiele folgen aus dem Hauptsatz, sondern auch zu zeigen, dass wir ihn woanders verwenden können. Insbesondere, verwenden wir ihn auf ein optimales schwaches Transportproblem mit Varianz und dann auch im Kontext der klassischen Theorie des optimalen Transports, um die Kantorovich Dualität für die quadratische Kostenfunktion zu bekommen.

Im letzten Kapitel 5 diskutieren wir die möglichen Verallgemeinerungen des Hauptsatzes. Insbesondere scheint der Austausch der konvexen Ordnung mit einer allgemeineren Form von Ordnung sehr vielversprechend zu sein.

Abstract

The aim of this thesis[1] is to investigate conditions that allow the dual formulation of weak optimal transport problems to be restricted to convex functions.

In Chapter 1, we recall the classical theory of optimal transport and touch upon a few results from descriptive set theory. We finish the chapter off with a quick primer on weak optimal transport.

Chapter 2 consists of an in-depth analysis of two motivating examples. We consider the context of both of them and cover their proofs.

The main theorem is shown in Chapter 3 and is based on the two motivating examples. More precisely, it sufficiently strengthens the proof of the first one in order to be applicable to the second one.

The aim of Chapter 4 is not only to rigorously check that the two motivating examples follow from the main theorem, but also to apply it elsewhere. In particular, one application is concerned with a weak optimal transport problem involving the variance and another one simply applies the main theorem within the confines of the classical setting to recover the Kantorovich duality for the quadratic cost function.

Finally, in Chapter 5 we discuss the potential extensions of the main theorem. In particular, swapping convex order for a more general type of order seems to be a promising direction.

[1]This research was funded in whole or in part by the Austrian Science Fund (FWF) [P 35519].

Contents

1 Introduction **1**
- 1.1 A Brief History 1
- 1.2 Optimal Transport 2
- 1.3 Wasserstein Spaces 6
- 1.4 Tiny Bit of Descriptive Set Theory 8
- 1.5 Weak Optimal Transport 9

2 Motivating Examples **12**
- 2.1 Maximal Covariance 12
- 2.2 Barycentric Costs 18

3 The Main Result **21**
- 3.1 Auxiliary Lemmas 21
- 3.2 Main Result and Proof 27

4 Special Cases **32**
- 4.1 Maximal Covariance Revisited 32
- 4.2 Barycentric Costs Revisited 34
- 4.3 Variance-Favoring Cost 36
- 4.4 The Classical Case 40

5 Conclusion and Further Research **42**

References **43**

1 Introduction

1.1 A Brief History

Unlike most other areas of mathematics, the inception of optimal transport can be traced back to a single paper. The year was 1781 when the work titled *Mémoire sur la théorie des déblais et des remblais* [18] was made public. Written by the French mathematician, Gaspard Monge[1], the paper describes a mathematical formulation of a real-world problem.

The task we are trying to accomplish entails filling a hole with earth. The exact amount of earth we need is situated someplace further away from the hole. We also imagine that moving a unit of earth from point A to point B requires some effort or costs us something. One would like to *optimally transport* the earth, that is, with the least possible effort or cost incurred.

The mathematical formulation of Monge was important, to say the least, though it did have its drawbacks. Sometimes, depending on the configuration of the mathematical *earth* and *hole*, there wouldn't even be *any* possible way of moving the earth into the hole, let alone optimally! But, as is often the case, in order for the mathematics to become more elegant, it must first become more intricate.

This is where Kantorovich[2] enters the picture. His work during the wartime period involved a variant of the earth-moving problem with the formulation of Monge being merely a special case [25]. The mathematical setting of Kantorovich was more robust, optimal earth-moving *plans* always existed and there was a dual interpretation of the problem. The main feature was that *splitting of mass* was allowed. In 1975, he was awarded the Nobel Prize in Economics for this and related work.

The field experienced a brief period of calm, followed by a booming comeback in 1987, when Yann Brenier[3] published his *Décomposition polaire et réarrangement monotone des champs de vecteurs* [10]. In it, he considered costs of the form distance squared, and not only showed that (under mild assumptions) optimal maps of *Monge flavor* exist, but also characterized them as gradients of convex functions.

The following years saw major breakthroughs within optimal transport in wide-ranging playgrounds such as partial differential equations, fluid mechanics, probability theory, optimization, statistics, and functional analysis. It found applications not only in finance and economics but also in geometry processing, cell biology, data analysis, and machine learning, to name just a few.

[1] Gaspard Monge (1746-1818), French mathematician, physicist, and chemist
[2] Леонид В. Канторович [Leonid V. Kantorovich] (1912-1986), Soviet mathematician and economist
[3] Yann Brenier (1957-), French mathematician

This thesis endeavors to add a modest drop of contribution to the vast ocean that is the study of optimal transport.

1.2 Optimal Transport

In this section, we give a brief overview of the most important results from the classical theory of optimal transport. We omit the proofs and refer the interested reader to [26]. Most of what follows is, in fact, either from [26] or based on the lecture notes [9].

We start with the definition of a central object in the study of optimal transport, a coupling.

Definition 1.1 (Coupling)**.** Let (X, μ) and (Y, ν) be probability spaces. We call a measure $\pi \in \mathcal{P}(X \times Y)$ a *coupling of* μ *and* ν if π has marginals μ and ν, that is,

$$\pi(A \times Y) = \mu(A), \;\; \pi(X \times B) = \nu(B)$$

for all Borel[4] measurable subsets A and B of X and Y, respectively. Denote the set of all couplings of μ and ν as $\Pi(\mu, \nu)$.

One may alternatively understand a coupling in probabilistic terms as the law of the random vector (X_0, X_1) where $X_0 \sim \mu$ and $X_1 \sim \nu$. Note that $\Pi(\mu, \nu)$ is always nonempty, as it contains $\mu \times \nu$, which is probabilistically the independent coupling.

The spaces X and Y will more often than not be assumed to be Polish spaces, that is, separable completely metrizable topological spaces. We will denote the space of probability measures on X by $\mathcal{P}(X)$.

We now define the Kantorovich problem.

Definition 1.2 (Kantorovich problem)**.** Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to [0, \infty]$ be a Borel measurable function. Then the *Kantorovich problem* is the following minimization problem:

$$\inf_{\pi \in \Pi(\mu, \nu)} \int c(x, y) d\pi(x, y).$$

The function c above is called the *cost function*, as we interpret it as being the cost of moving a unit of mass from position $x \in X$ to position $y \in Y$. One may also loosen the nonnegativity assumption. It is common to additionally assume that c is either continuous or at least lower semicontinuous.

Next, in a somewhat antichronological order, we define the Monge problem.

Definition 1.3 (Monge problem)**.** Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to [0, \infty]$ be a Borel measurable function. Then the *Monge problem* is the following minimization problem:

$$\inf_{T} \int c(x, T(x)) d\mu(x),$$

where the infimum is taken over all Borel measurable maps $T : X \to Y$ such that the pushforward measure $T(\mu) := T_\sharp \mu$ is precisely ν. We call such maps *Monge transports.*

[4] Émile Borel (1871-1956), French mathematician and politician

Unlike couplings, Monge transports may fail to exist (e.g. $\mu = \delta_{x_0}$ and ν not a Dirac mass). This, among other reasons, makes them unfavorable to work with. Notice that a Monge transport can be identified with a special kind of coupling. Indeed,

$$T(x) = y \ \pi\text{-a.e.} \iff \pi = (\mathrm{Id} \times T)(\mu).$$

If $\pi \in \Pi(\mu, \nu)$ happens to be of the form given on the right-hand side above, we call it a *Monge coupling.*

We now state the first important theorem from the classical theory of optimal transport.

Theorem 1.4 (Optimal couplings exist)**.** *Let (X, μ) and (Y, ν) be two Polish probability spaces, $a : X \to \mathbb{R} \cup \{-\infty\}$ and $b : Y \to \mathbb{R} \cup \{-\infty\}$ two upper semicontinuous functions in $L^1(\mu)$ and $L^1(\nu)$, respectively. Let $c : X \times Y \to \mathbb{R} \cup \{-\infty\}$ be a lower semicontinuous cost function that satisfies $c(x, y) \geq a(x)+b(y)$ on $X \times Y$. Then the Kantorovich problem admits a minimizer.*

The proof consists of showing two things: the compactness of $\Pi(\mu, \nu)$ in the topology induced by weak convergence (recall that (μ_k) converges to μ weakly if $\int \varphi d\mu_k \to \int \varphi d\mu$ for all φ bounded and continuous) and the lower semicontinuity of $\pi \mapsto \int c d\pi$ in the same topology. These two facts combined yield attainment.

Before discussing duality, we define the crucial concept of c-cyclical monotonicity.

Definition 1.5 (c-cyclically monotone)**.** Let X, Y be arbitrary sets and $c : X \times Y \to (-\infty, \infty]$ be a function. A subset $\Gamma \subseteq X \times Y$ is c-*cyclically monotone* if, for any $N \in \mathbb{N}$, and any family $(x_1, y_1), \ldots, (x_N, y_N)$ of points in Γ, we have the inequality

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{y+1})$$

(with $y_{N+1} = y_1$). We call $\pi \in \Pi(\mu, \nu)$ *c-cyclically monotone* if there is a c-cyclically monotone Γ such that $\pi(\Gamma) = 1$.

The idea underpinning this concept is simple. It says that the transport that maps $x_i \mapsto y_i$ (not necessarily deterministically) cannot be improved upon by shifting the y_i's cyclically. It turns out that this is both a necessary and sufficient condition for the optimality of a plan under certain assumptions.

We now define what it means for a function to be c-convex and c-concave. We also define the c-transforms.

Definition 1.6 (c-convexity)**.** Let X, Y be sets and $c : X \times Y \to (-\infty, \infty]$. We say that $\psi : X \to \mathbb{R} \cup \{\infty\}$ is *c-convex* if it is not identically ∞, and there exists $\zeta : Y \to \mathbb{R} \cup \{\pm\infty\}$ such that

$$\forall x \in X, \ \ \psi(x) = \sup_{y \in Y} (\zeta(y) - c(x, y)).$$

Definition 1.7 (c-concavity)**.** Let X, Y be sets, and $c : X \times Y \to (-\infty, \infty]$. We say that $\phi : Y \to \mathbb{R} \cup \{-\infty\}$ is *c-concave* if it is not identically $-\infty$, and there exists $\psi : X \to \mathbb{R} \cup \{\pm\infty\}$ such that

$$\forall y \in Y,\ \phi(y) = \inf_{x \in X} (\psi(x) + c(x, y)).$$

Definition 1.8 (c-transform)**.** For a function $\psi : X \to \mathbb{R} \cup \{\infty\}$ we define its *c-transform* as

$$\psi^c(y) = \inf_{x \in X} (\psi(x) + c(x, y)),\ y \in Y,$$

and for a function $\phi : Y \to \mathbb{R} \cup \{-\infty\}$ we define its *c-transform* as

$$\phi^c(x) = \sup_{y \in Y} (\phi(y) - c(x, y)),\ x \in X.$$

Definition 1.9 (c-sub- and c-superdifferential)**.** For a c-convex function $\psi : X \to \mathbb{R} \cup \{\infty\}$ we define its *c-subdifferential* as

$$\partial_c \psi := \{(x, y) \in X \times Y \mid \psi^c(y) - \psi(x) = c(x, y)\},$$

and for a c-concave function $\phi : Y \to \mathbb{R} \cup \{-\infty\}$ we define its *c-superdifferential* as

$$\partial^c \phi := \{(x, y) \in X \times Y \mid \phi(y) - \phi^c(x) = c(x, y)\}.$$

We shall use $C_b(X)$ to denote the set of all bounded and continuous real-valued functions on the space X.

Having defined the necessary notions, we state one of the most important theorems in the classical theory of optimal transport, namely the Kantorovich duality.

Theorem 1.10 (Kantorovich Duality)**.** *Let (X, μ) and (Y, ν) be two Polish probability spaces and let $c : X \times Y \to \mathbb{R} \cup \{\infty\}$ be a lower semicontinuous cost function with*

$$c(x, y) \geq a(x) + b(y),\ (x, y) \in X \times Y$$

for real-valued upper semicontinuous functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Then:

(i) Duality holds:

$$\begin{aligned} \min_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} & c(x, y) d\pi(x, y) \\ &= \sup_{\substack{(\psi, \phi) \in C_b(X) \times C_b(Y) \\ \phi - \psi \leq c}} \left(\int_Y \phi(y) d\nu(y) - \int_X \psi(x) d\mu(x) \right) \\ &= \sup_{\substack{(\psi, \phi) \in L^1(\mu) \times L^1(\nu) \\ \phi - \psi \leq c}} \left(\int_Y \phi(y) d\nu(y) - \int_X \psi(x) d\mu(x) \right) \\ &= \sup_{\psi \in L^1(\mu)} \left(\int_Y \psi^c(y) d\nu(y) - \int_X \psi(x) d\mu(x) \right) \\ &= \sup_{\phi \in L^1(\nu)} \left(\int_Y \phi(y) d\nu(y) - \int_X \phi^c(x) d\mu(x) \right) \end{aligned}$$

where the suprema may be assumed to be taken over ψ c-convex and ϕ c-concave.

(ii) If c is real-valued and the optimal cost $C(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int c d\pi$ is finite, then there exists a measurable c-cyclically monotone set $\Gamma \subseteq X \times Y$ (closed for a, b, c continuous) such that for any transference plan $\pi \in \Pi(\mu,\nu)$ the following are equivalent:

(a) π optimal;

(b) π c-cyclically monotone;

(c) $\exists\psi$, c-convex such that $\psi^c(y) - \psi(x) = c(x,y)$ π-a.s.;

(d) $\exists\psi : X \to \mathbb{R}\cup\{\infty\}$ and $\phi : Y \to \mathbb{R}\cup\{-\infty\}$, such that $\phi(y) - \psi(x) \leq c(x,y)$ for all (x,y), with equality π-a.s.;

(e) π concentrated on Γ.

(iii) If c is real-valued, $C(\mu,\nu) < \infty$, and one has the pointwise upper bound

$$c(x,y) \leq c_X(x) + c_Y(y),$$

where c_X and c_Y are in $L^1(\mu)$ and $L^1(\nu)$, respectively, then both the primal and dual Kantorovich problems have optimizers, thus

$$\begin{aligned}
&\min_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y) d\pi(x,y) \\
&\quad = \max_{\substack{(\psi,\phi)\in L^1(\mu)\times L^1(\nu)\\ \phi-\psi\leq c}} \left(\int_Y \phi(y)d\nu(y) - \int_X \psi(x)d\mu(x)\right) \\
&\quad = \max_{\psi\in L^1(\mu)} \left(\int_Y \psi^c(y)d\nu(y) - \int_X \psi(x)d\mu(x)\right),
\end{aligned}$$

where the maxima may be assumed to be taken over ψ c-convex and $\phi = \psi^c$. If a, b and c are continuous, then there is a closed c-cyclically monotone $\Gamma \subseteq X \times Y$ such that for any $\pi \in \Pi(\mu,\nu)$ and for $\psi \in L^1(\mu)$ we have:

- *π is optimal for the Kantorovich problem if and only if $\pi(\Gamma) = 1$,*
- *ψ is optimal for the dual Kantorovich problem if and only if $\Gamma \subseteq \partial_c\psi$.*

Point *(ii)* is sometimes referred to as the fundamental theorem of optimal transport. The theorem stated above is exceptionally powerful and encompasses many interesting special cases.

Corollary 1.11 (Kantorovich–Rubinstein[5] formula)**.** *Let $c(x,y) = d(x,y)$ be a distance on some Polish space X and let μ,ν be probability measures on X with finite first moments. Then*

$$\inf_{\pi\in\Pi(\mu,\nu)} \int d(x,y)d\pi(x,y) = \sup\left\{\int_X \psi d\mu - \int_X \psi d\nu\right\},$$

where the infimum is taken over all 1-Lipschitz[6] functions ψ.

[5] Геннадий Ш. Рубинштейн [Gennadii S. Rubinstein] (1923-2004), Soviet mathematician
[6] Rudolf Lipschitz (1832-1903), German mathematician

Corollary 1.12 (Brenier's Theorem)**.** *Let $X = Y = \mathbb{R}^d$, $c(x,y) = |x-y|^2/2$, and assume that $\mu, \nu \in \mathcal{P}(X)$ have finite second moments and $\mu \ll \mathrm{Leb}$[7]. Then, there is a unique optimal coupling $\pi^* \in \Pi(\mu,\nu)$ and this optimizer is of the form $\pi^* = (\mathrm{Id}, \nabla\bar{\varphi})(\mu)$ for some convex function $\bar{\varphi} : X \to \mathbb{R}$.*

Moreover, there exists a μ-a.e. unique map T of the form $T = \nabla\bar{\varphi}$ such that $T(\mu) = \nu$ and for any convex $\tilde{\varphi}$, the map $\widetilde{T} = \nabla\tilde{\varphi}$ is optimal between μ and $\widetilde{T}(\mu)$.

1.3 Wasserstein Spaces

In this section, we consider the Wasserstein[8] space and the associated Wasserstein distance. We start off with a few definitions, after which we state some useful properties. Like the previous section, this one is almost exclusively based on [26], but also contains a few results from [20].

Definition 1.13 (Wasserstein distance)**.** Let (X,d) be a Polish metric space, and let $p \in [1,\infty)$. For any two probability measures μ, ν on X, the *Wasserstein distance* of order p between μ and ν is defined as

$$\mathcal{W}_p(\mu,\nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_X d(x,y)^p d\pi(x,y) \right)^{1/p}.$$

Definition 1.14 (Wasserstein space)**.** Let (X,d) be a Polish metric space, and let $p \in [1,\infty)$. The *Wasserstein space* of order p is defined as

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) \,\middle|\, \int_X d(x_0,x)^p d\mu(x) < \infty \right\},$$

where $x_0 \in X$ is arbitrary (the choice of x_0 does not affect the space).

Using the knowledge gained in the previous section, one can show that the Wasserstein distance is truly a distance on the Wasserstein space.

Theorem 1.15. *$(\mathcal{P}_p(X), \mathcal{W}_p)$ is a metric space.*

Next, we define weak convergence in $\mathcal{P}_p$ and then consider its connection with convergence in the Wasserstein metric.

Definition 1.16 (Weak convergence in $\mathcal{P}_p$)**.** Let (X,d) be a Polish space, and $p \in [1,\infty)$. Let $(\mu_k)_{k\in\mathbb{N}} \subseteq \mathcal{P}_p(X)$ and let μ be in $\mathcal{P}_p(X)$. Then we say that (μ_k) *converges weakly in* $\mathcal{P}_p(X)$ if any of the following equivalent properties hold for some (and then any) $x_0 \in X$:

(i) $\mu_k \longrightarrow \mu$ weakly and $\int d(x_0,x)^p \, d\mu_k(x) \longrightarrow \int d(x_0,x)^p \, d\mu(x)$,

(ii) $\mu_k \longrightarrow \mu$ weakly and $\limsup_{k\to\infty} \int d(x_0,x)^p \, d\mu_k(x) \le \int d(x_0,x)^p \, d\mu(x)$,

[7] Henri Lebesgue (1875-1941), French mathematician

[8] Леонид Н. Васерштейн [Leonid N. Vaserstein] (1944-), Russian-American mathematician

(iii) $\mu_k \longrightarrow \mu$ weakly and $\lim_{R\to\infty} \limsup_{k\to\infty} \int_{d(x_0,x)\geq R} d\,(x_0,x)^p \, d\mu_k(x) = 0$,

(iv) For all continuous functions φ with $|\varphi(x)| \leq C\,(1 + d\,(x_0,x)^p)$, $C \in \mathbb{R}$, one has

$$\int \varphi(x) d\mu_k(x) \longrightarrow \int \varphi(x) d\mu(x).$$

It turns out that this form of convergence is equivalent to convergence in the Wasserstein metric.

Theorem 1.17. *Let (X,d) be a Polish space, and $p \in [1,\infty)$. Then $\mathcal{W}_p$ metrizes weak convergence in $\mathcal{P}_p$, i.e., if $(\mu_k)_{k\in\mathbb{N}} \subseteq \mathcal{P}_p(X)$ and $\mu \in \mathcal{P}(X)$, then*

$$\mu_k \text{ converges weakly in } \mathcal{P}_p(X) \text{ to } \mu \iff \mathcal{W}_p(\mu_k,\mu) \to 0.$$

We get two important corollaries from this.

Corollary 1.18 (Continuity of $\mathcal{W}_p$)**.** *Let (X,d) be a Polish space, and $p \in [1,\infty)$, then $\mathcal{W}_p$ is continuous on $\mathcal{P}_p(X)$, i.e., if μ_k, ν_k converge weakly in $\mathcal{P}_p(X)$ to μ, ν, respectively, as $k \to \infty$, then*

$$\mathcal{W}_p(\mu_k,\nu_k) \to \mathcal{W}_p(\mu,\nu).$$

Remark 1.19. If the measures converge in the usual weak convergence sense, we still have $\mathcal{W}_p(\mu,\nu) \leq \liminf \mathcal{W}_p(\mu_k,\nu_k)$.

Corollary 1.20 (Metrizing the weak topology)**.** *Let (X,d) be a Polish space. If $\tilde{d}$ is a bounded metric that induces the same topology as d (say, $\tilde{d} = d/(1+d)$), then convergence in the Wasserstein distance for $\tilde{d}$ is equivalent to the usual weak convergence in $\mathcal{P}(X)$.*

We also mention that the Wasserstein space of a Polish space is itself a Polish space.

Theorem 1.21. *Let X be a Polish space and $p \in [1,\infty)$. Then $(\mathcal{P}_p(X),\mathcal{W}_p)$ is also Polish.*

We now list some further properties of the Wasserstein space which can be found in [20].

Theorem 1.22 (Compact sets in $\mathcal{P}_p(X)$)**.** *A tight set $\mathcal{K} \subseteq \mathcal{P}_p(X)$ is relatively compact in $\mathcal{P}_p(X)$ (has compact closure with respect to $\mathcal{W}_p$) if and only if for some (and then any) $x_0 \in X$*

$$\sup_{\mu\in\mathcal{K}} \int_{d(x,x_0)>R} d(x,x_0)^p d\mu(x) \to 0, \text{ as } R \to \infty.$$

Corollary 1.23 (Measures with Common Support)**.** *Let $K \subseteq X$ be a compact set, then*

$$\mathcal{P}(K) = \{\mu \in \mathcal{P}(X) \mid \mu(K) = 1\} \subseteq \mathcal{P}_p(X)$$

is compact.

We then have the following useful theorem about dense subsets of $\mathcal{P}_p(X)$.

Theorem 1.24. *The following collections of measures are dense in $\mathcal{P}_p(X)$:*

1. *finitely supported measures with rational weights,*
2. *compactly supported measures,*
3. *finitely supported measures with rational weights on a dense subset of X,*
4. *when $X = \mathbb{R}^d$, the collection of absolutely continuous and compactly supported measures,*
5. *when $X = \mathbb{R}^d$, the collection of absolutely continuous measures with strictly positive and bounded analytic densities.*

1.4 Tiny Bit of Descriptive Set Theory

In this short section, we mention a few results from the world of descriptive set theory. This particular branch of set theory mostly takes place within Polish spaces and studies their analytic sets (equivalently, projections of Borel sets) and analytic functions. An important question that arises naturally is whether a projection of a certain Borel set is itself Borel. Despite Lebesgue's erroneous proof of this claim [17], it only holds when the Borel set being projected has some additional properties, usually related to the size of its sections. The results mentioned below are from [23] and [24].

Definition 1.25 (Uniformization)**.** Let X, Y be Polish spaces, $B \subseteq X \times Y$. A set $C \subseteq B$ is called a *uniformization* of B if for every $x \in X$, the sections C_x contain at most one point and $\mathrm{proj}_X(C) = \mathrm{proj}_X(B)$.

We will be interested in the case when B is a Borel set. If a Borel set has a Borel uniformization, then its projection is also a Borel set. We present an important result on Borel uniformizations which we later utilize in the proof of the main result.

Theorem 1.26 (Arsenin[9]–Kunugi[10])**.** *Let X, Y be Polish spaces, and $B \subseteq X \times Y$ a Borel set such that the sections B_x are σ-compact for every $x \in X$. Then B admits a Borel uniformization. In particular,* $\mathrm{proj}_X(B)$ *is a Borel set.*

Remark 1.27. There are similar results that replace the σ-compactness by compactness, countableness, or nonmeagerness.

Another important concept from descriptive set theory with wide-ranging applications is the idea of measurable selections.

Definition 1.28 (Multifunction)**.** G is a *multifunction* from X to Y if it maps elements of X to nonempty subsets of Y.

[9] Василий Я. Арсенин [Vasilii Y. Arsenin] (1915-1990), Soviet mathematician

[10] Kinjirō Kunugi (1903-1975), Japanese mathematician

Definition 1.29 (Selection)**.** A *selection* of a multifunction G is a map $s : X \to Y$ such that $s(x) \in G(x)$ for all $x \in X$.

Arbitrary selections are not very interesting, since, by the Axiom of Choice, they always exist. We consider a result that guarantees the existence of a measurable selection, the so-called Kuratowski[11]–Ryll-Nardzewski[12] measurable selection theorem.

Theorem 1.30 (Kuratowski–Ryll-Nardzewski)**.** *Let X be a Polish space, $(\Omega, \mathcal{F})$ a measurable space, and ψ a multifunction on Ω taking values in the nonempty closed subsets of X. If we have*

$$\{\omega \mid \psi(\omega) \cap U \neq \emptyset\} \in \mathcal{F},$$

for all U, open subsets of X, then ψ has a $\mathcal{F}$-$\mathcal{B}(X)$-measurable selection.

1.5 Weak Optimal Transport

This section introduces a fairly recent subfield of optimal transport. Originating in [16], the concept of weak optimal transport is a generalization of the classical theory. Not long after its inception, numerous applications and extensions followed. For example, many results were further developed and strengthened in [4]. We start with a motivating discussion.

Let X, Y be Polish spaces and π be a coupling on $X \times Y$. Recall that the disintegration theorem (see e.g. [12]) allows us to write

$$d\pi(x, y) = d\pi_x(y) d\mu(x),$$

where $(\pi_x)_{x \in X}$ is the disintegration of π with respect to the first marginal.

If we apply this to the classical optimal transport problem for some measurable $c : X \times Y \to \mathbb{R} \cup \{\infty\}$, we get

$$\inf_{\pi \in \Pi(\mu,\nu)} \int c(x, y) d\pi(x, y) = \inf_{\pi \in \Pi(\mu,\nu)} \int \underbrace{\int c(x, y) d\pi_x(y)}_{=:C(x,\pi_x)} d\mu(x) = \inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x) d\mu(x).$$

With this in mind, we define the weak optimal transport problem.

Definition 1.31 (Weak optimal transport problem)**.** Let $\mu \in \mathcal{P}_p(X)$, $\nu \in \mathcal{P}_p(Y)$, and $C : X \times \mathcal{P}_p(Y) \to \mathbb{R} \cup \{\infty\}$ be a measurable function. Then the *weak optimal transport problem* is the following minimization problem:

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x) d\mu(x).$$

[11] Kazimierz Kuratowski (1896-1980), Polish mathematician
[12] Czesław Ryll-Nardzewski (1926-2015), Polish mathematician

Example 1.32 (Martingale Optimal Transport)**.** Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be bounded from below and lower semicontinuous. Consider the cost function

$$C(x, \rho) = \begin{cases} \int c(x, y) d\rho(y), & \mathrm{bary}(\rho) = x \\ \infty, & \text{otherwise}, \end{cases}$$

where $\mathrm{bary}(\rho) := \int y d\rho(y)$. Then one gets

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x) d\mu(x) = \inf_{\pi \in MT(\mu,\nu)} \iint c(x, y) d\pi_x(y) d\mu(x) = \inf_{\pi \in MT(\mu,\nu)} \int c(x, y) d\pi(x, y),$$

where $MT(\mu, \nu)$ denotes the set of martingale couplings, that is,

$$MT(\mu, \nu) = \{\pi \in \Pi(\mu, \nu) \mid \mathrm{bary}(\pi_x) = x \ \ \mu\text{-a.e.}\}.$$

Probabilistically, if one has $(X_0, X_1) \sim \pi$, then $\mathbb{E}[X_1 | X_0 = x] = \mathrm{bary}(\pi_x)$, and if the latter is simply equal to x μ-a.e., we have that the pair (X_0, X_1) is a one step martingale. One should note that unlike $\Pi(\mu, \nu)$, which is always nonempty, the set $MT(\mu, \nu)$ is nonempty if and only if μ and ν are in convex order, or in symbols, $\mu \preceq_c \nu$ (i.e., $\int \phi d\mu \leq \int \phi d\nu$, for all convex real-valued functions ϕ); this is known as Strassen's[13] theorem.

Thus, the framework of weak optimal transport is sufficiently robust to encode a martingale constraint.

Example 1.33 (Entropic Optimal Transport)**.** Recall that the relative entropy (sometimes also referred to as Kullback[14]–Leibler[15] divergence) is defined for measures $\eta, \kappa \in \mathcal{P}(X)$ as

$$H(\eta|\kappa) := \begin{cases} \int log \left(\frac{d\eta}{d\kappa}(x)\right) d\eta(x), & \eta \ll \kappa \\ \infty, & \text{otherwise}. \end{cases}$$

We consider the cost function

$$C(x, \rho) = \int c(x, y) d\rho(y) + H(\rho|\nu).$$

A quick calculation gives us

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x) d\mu(x) = \inf_{\pi \in \Pi(\mu,\nu)} \left[\int c(x, y) d\pi(x, y) + H(\pi|\mu \otimes \nu)\right].$$

The minimization problem on the right-hand side is the entropic optimal transport problem.

[13] Volker Strassen (1936-), German mathematician
[14] Solomon Kullback (1907-1994), American cryptanalyst and mathematician
[15] Richard A. Leibler (1914-2003), American cryptanalyst and mathematician

The special cases presented in the examples assure us that the study of weak optimal transport problems is justified.

We list the assumptions that will hold for the rest of this section. Let X, Y be Polish spaces. We fix a compatible bounded metric d_X on the space X and a compatible metric d_Y on the space Y. Let $p \in [1, \infty)$, and write $\mathcal{P}_{p,d_Y}(Y)$ for the Wasserstein space (as in Definition 1.14), where the d_Y indicates the metric used to define the space.

We state the result on the existence of minimizers for weak optimal transport, as in [4, Theorem 1.2].

Theorem 1.34. *Assume that $\nu \in \mathcal{P}_{p,d_Y}(Y)$ and let $C : X \times \mathcal{P}_{p,d_Y}(Y) \to \mathbb{R} \cup \{\infty\}$ be jointly lower semicontinuous with respect to the product topology on $X \times \mathcal{P}_{p,d_Y}(Y)$, bounded from below, and convex in the second argument. Then the problem*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x) d\mu(x)$$

admits a minimizer.

Remark 1.35. Note that if C is strictly convex in the second argument and the value above is finite, the minimizer is also unique.

To tackle the dual problem, we require the following set:

$$\Phi_{b,p}(Y) := \{\psi : Y \to \mathbb{R} \text{ continuous} \mid \exists a, b, \ell \in \mathbb{R},\ y_0 \in Y,\ \ell \leq \psi(\cdot) \leq a + b d_Y(y_0, \cdot)^p\}.$$

In the absence of any possible confusion, we may simply write $\Phi_{b,p}$.

For each $\psi \in \Phi_{b,p}$, we define

$$R_C\psi(x) := \inf_{\rho \in \mathcal{P}_{p,d_Y}(Y)} \rho(\psi) + C(x, \rho),$$

where $\rho(\psi)$ is shorthand notation for $\int \psi d\rho$.

As noted in [4], $R_C\psi(\cdot)$ is universally measurable if C is measurable and thus the integral $\mu(R_C\psi)$ is well-defined for all $\mu \in \mathcal{P}(X)$ if C is bounded from below.

We recall the duality theorem [4, Theorem 1.3] for weak optimal transport problems.

Theorem 1.36. *Let $C : X \times \mathcal{P}_{p,d_Y}(Y) \to \mathbb{R} \cup \{\infty\}$ be jointly lower semicontinuous with respect to the product topology on $X \times \mathcal{P}_{p,d_Y}(Y)$, bounded from below, and convex in the second argument. Then we have for all $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}_{p,d_Y}(Y)$*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x) d\mu(x) = \sup_{\psi \in \Phi_{b,p}} \mu(R_C\psi) - \nu(\psi).$$

The duality theorem above will serve as one of the backbones of the main result of this thesis. Roughly speaking, in Chapter 3 we will find a condition that allows us to restrict the dual problem to convex functions.

These are but a few results from the rich theory of weak optimal transport. Nonetheless, they will suffice for the purposes of this thesis.

2 Motivating Examples

This chapter aims to investigate two separate examples of a single phenomenon, with the ultimate goal being their unification into a more general result. Both of them are weak transport problems that allow for the dual formulation to be restricted to convex functions.

The first occurrence took place at the very inception of weak transport, in [16, Theorem 2.11(3)], but we prefer to work with the formulation in [15, Theorem 1.1]. Another occurrence can be found in [7, Proposition 3.3], where it is used in the context of martingale optimal transport.

In the two sections that follow, we explain both of these in detail and prove their respective restricted duality principles. For practical purposes, we study the two occurrences in reverse chronological order, i.e., we refer to [7, Proposition 3.3] and [15, Theorem 1.1] as the first and second motivating example, respectively.

2.1 Maximal Covariance

Let us describe the central problem of [7].

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, such that $\mu \preceq_{\mathrm{c}} \nu$. Consider the optimization problem

$$\inf_{\substack{M_0 \sim \mu, M_1 \sim \nu \\ M_t = M_0 + \int_0^t \sigma_s dB_s}} \mathbb{E}\left[\int_0^1 |\sigma_t - \mathrm{Id}|_{\mathrm{HS}}^2 dt\right], \tag{MBB}$$

where B is a d-dimensional Brownian[1] motion and $|\cdot|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm. The fact that (MBB) has a unique optimizer M^* was shown in [6]. This process is the one that maximizes its correlation with Brownian motion subject to the marginal constraints or, as was also shown in [6], it is the process that follows a path of a Brownian motion as closely as possible in the *adapted Wasserstein distance* (see [5]) such that it starts at μ and ends at ν. We will refer to the optimizer M^* as the *stretched Brownian motion* between μ and ν.

One of the main results of [7] links the stretched Brownian motion with the so-called *Bass*[2] *martingale.*

Definition 2.1 (Bass martingale). Let $B = (B_t)_{0 \le t \le 1}$ be a Brownian motion on $\mathbb{R}^d$, where $B_0 \sim \alpha \in \mathcal{P}_2(\mathbb{R}^d)$, and let $v : \mathbb{R}^d \to \mathbb{R}$ be convex such that $\nabla v(B_1)$ is integrable.

[1]Robert Brown (1773-1858), Scottish botanist and paleobotanist

[2]Richard F. Bass (1951-), American mathematician

Then we call

$$M_t := \mathbb{E}[\nabla v(B_1) \mid \sigma(B_s : 0 \le s \le t)] = \mathbb{E}[\nabla v(B_1) \mid B_t], \quad 0 \le t \le 1$$

a *Bass martingale.*

Another concept we will require is *irreducibility.*

Definition 2.2 (Irreducibility)**.** For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ we say that the pair (μ, ν) is *irreducible* if for all measurable sets $A, B \subseteq \mathbb{R}^d$ with positive μ, ν mass, respectively, there is a martingale $X = (X_t)_{0\le t\le 1}$ with $X_0 \sim \mu, X_1 \sim \nu$ such that $\mathbb{P}(X_0 \in A,\ X_1 \in B) > 0$.

We now state [7, Theorem 1.3].

Theorem 2.3. *Let $\mu \preceq_{\mathrm{c}} \nu$ be probabilities on $\mathbb{R}^d$ with finite second moments and suppose that (μ, ν) is irreducible. Then the following are equivalent for a martingale $M = (M_t)_{0\le t\le 1}$ with $M_0 \sim \mu$ and $M_1 \sim \nu$:*

1. *M is a stretched Brownian motion, i.e., the unique optimizer of* (MBB)*,*
2. *M is a Bass martingale.*

We now move on to the dual problem. Instead of working with (MBB) we consider the equivalent problem

$$P(\mu, \nu) := \sup_{\substack{M_0\sim\mu, M_1\sim\nu \\ M_t = M_0 + \int_0^t \sigma_s dB_s}} \mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_t) dt\right].$$

Indeed, by expanding the Hilbert–Schmidt norm and using properties of the quadratic variation one gets

$$\mathbb{E}\left[\int_0^1 |\sigma_t - \mathrm{Id}|^2_{\mathrm{HS}} dt\right] = d + \int |y|^2 d\nu(y) - \int |x|^2 d\mu(x) - 2\mathbb{E}\left[\int_0^1 \mathrm{tr}(\sigma_t) dt\right].$$

The problem $P(\mu, \nu)$ admits a simpler dual formulation [7, Theorem 1.4], which mirrors the classical case of the squared distance. Let γ denote the d-dimensional standard Gaussian distribution. Write $*$ for the convolution of a function with a measure, and ψ^* for the convex conjugate of ψ.

Theorem 2.4. *Assume that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are in convex order. Then the value of $P(\mu, \nu)$ is equal to*

$$D(\mu, \nu) := \inf_{\substack{\psi \in L^1(\nu) \\ \psi \text{ convex}}} \left(\int \psi d\nu - \int (\psi^* * \gamma)^* d\mu\right),$$

and the infimum is attained by a convex function ψ_{opt} if and only if (μ, ν) is irreducible. In this case, the unique optimizer to the primal problem is given by the Bass martingale

$$M_t := \mathbb{E}[\nabla v(B_1) \mid \sigma(B_s : 0 \le s \le t)] = \mathbb{E}[\nabla v(B_1) \mid B_t], \quad 0 \le t \le 1$$

where $v = \psi^*_{opt}$ and $B_0 \sim \nabla(\psi^*_{opt} * \gamma)^*(\mu)$.

Remark 2.5. The dual attainment has to be understood in a relaxed sense since the optimizer will not necessarily be ν integrable. Since this thesis only concerns itself with the first part of Theorem 2.4, we will not delve any further into the details of the second part.

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be such that $\mu \preceq_c \nu$. We consider the discrete version of $P(\mu, \nu)$,

$$\widetilde{P}(\mu,\nu) := \sup_{\pi \in \mathrm{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x, \gamma) d\mu(x),$$

where the *maximal covariance* between two measures $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is defined as

$$\mathrm{MCov}(\rho_1, \rho_2) := \sup_{\tau \in \Pi(\rho_1,\rho_2)} \int x_1 x_2 d\tau(x_1, x_2).$$

By [6, Theorem 2.2], the values of $P(\mu,\nu)$ and $\widetilde{P}(\mu,\nu)$ are finite and equal. Furthermore, their optimizers are related in the sense that π^* is the law of (M_0^*, M_1^*) and conversely, given π^* one can construct M^*.

We now define the dual of the discretized problem. For this, we will require the set

$$\mathcal{C}_2(\mathbb{R}^d) := \left\{ \psi : \mathbb{R}^d \to \mathbb{R} \text{ continuous} \mid \exists a, b, \ell \in \mathbb{R},\ \ell + 1/2|\cdot|^2 \le \psi(\cdot) \le a + b|\cdot|^2 \right\}.$$

Then the dual problem is given by

$$\widetilde{D}(\mu,\nu) := \inf_{\psi \in \mathcal{C}_2(\mathbb{R}^d)} \left(\int \psi d\nu - \int \varphi^{\psi} d\mu \right),$$

where the function $x \mapsto \varphi^{\psi}(x)$ is defined as

$$\varphi^{\psi}(x) := \inf_{\rho \in \mathcal{P}_2^x(\mathbb{R}^d)} \left(\int \psi d\rho - \mathrm{MCov}(\rho, \gamma) \right),$$

and $\mathcal{P}_2^x(\mathbb{R}^d)$ denotes the set of all probability measures on $\mathbb{R}^d$ with finite second moment and mean x.

The following theorem from [7] asserts that there is no duality gap between the discretized primal and dual problem.

Theorem 2.6. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Then $\widetilde{P}(\mu,\nu) = \widetilde{D}(\mu,\nu)$. Moreover, the primal problem is uniquely attained and has a finite value.*

Unique attainment follows from [6], whereas duality can be shown by using Theorem 1.36 since $\widetilde{P}$ is a weak transport problem.

The restriction of the dual problem to convex functions in Theorem 2.4 stems from [7, Proposition 3.3], a result that focuses on the restriction of the discretized dual. This proposition will also serve as our first motivating example. Before we state the result, we define the shorthand notation

$$\mathcal{D}(\psi) := \int \psi d\nu - \int \varphi^{\psi} d\mu, \quad \psi \in \mathcal{C}_2(\mathbb{R}^d).$$

Proposition 2.7 ([7, Proposition 3.3])**.** Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. Then $\mathcal{D}(\text{conv}\psi) \leq \mathcal{D}(\psi)$ and consequently

$$\widetilde{D}(\mu, \nu) = \inf_{\psi \in \mathcal{C}_2(\mathbb{R}^d)} \mathcal{D}(\psi) = \inf_{\substack{\psi \in \mathcal{C}_2(\mathbb{R}^d) \\ \psi \text{ convex}}} \mathcal{D}(\psi).$$

We remark that conv ψ denotes the convex hull of ψ. To prove the proposition, we require the following lemma.

Lemma 2.8. *Let* $\psi \in \mathcal{C}_2(\mathbb{R}^d)$*. Then*

$$(\text{conv}\ \psi)(y) = \inf_{\rho \in \mathcal{P}_2^y(\mathbb{R}^d)} \int \psi d\rho, \quad y \in \mathbb{R}^d$$

and conv $\psi \in \mathcal{C}_2(\mathbb{R}^d)$.

Proof. Fix $\psi \in \mathcal{C}_2(\mathbb{R}^d)$. Denote the infimum on the right-hand side by $f(y)$. We show that f is, indeed, conv ψ.

Step 1. (Convexity of f): Let $y_1, y_2 \in \mathbb{R}^d$, $\epsilon > 0$ arbitrary. Then by definition of f, there are ρ_1, ρ_2 in $\mathcal{P}_2^{y_1}(\mathbb{R}^d)$, $\mathcal{P}_2^{y_2}(\mathbb{R}^d)$, respectively, such that

$$f(y_1) + \epsilon \geq \int \psi d\rho_1, \ f(y_2) + \epsilon \geq \int \psi d\rho_2. \tag{2.1}$$

Let $\alpha \in [0, 1]$. Then

$$\begin{aligned} f(\alpha y_1 + (1-\alpha) y_2) &= \inf_{\rho \in \mathcal{P}_2^{\alpha y_1 + (1-\alpha) y_2}(\mathbb{R}^d)} \int \psi d\rho \\ &\leq \int \psi d(\alpha \rho_1 + (1-\alpha)\rho_2) \\ &= \alpha \int \psi d\rho_1 + (1-\alpha) \int \psi d\rho_2 \\ &\leq \alpha f(y_1) + (1-\alpha) f(y_2) + \epsilon, \end{aligned}$$

where the first inequality follows from the fact that $\alpha\rho_1 + (1-\alpha)\rho_2 \in \mathcal{P}_2^{\alpha y_1 + (1-\alpha) y_2}(\mathbb{R}^d)$, and the second inequality follows from the two inequalities in (2.1). Since ϵ was arbitrary, this proves the convexity of f.

Step 2. (ψ dominates f): Let $y \in \mathbb{R}^d$ be arbitrary. Consider the probability measure δ_y on $\mathbb{R}^d$. Then $\delta_y \in \mathcal{P}_2^y(\mathbb{R}^d)$ and hence

$$f(y) \leq \int \psi d\delta_y = \psi(y).$$

Step 3. (Maximality of f): Fix $y \in \mathbb{R}^d$ and let $g : \mathbb{R}^d \to \mathbb{R}$ be convex and dominated by ψ. Assume that $f \leq g \leq \psi$. Integrating with respect to any $\rho \in \mathcal{P}_2^y(\mathbb{R}^d)$ yields

$$\int f d\rho \leq \int g d\rho \leq \int \psi d\rho.$$

Taking the infimum over all such ρ, we get

$$\inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int f d\rho \leq \inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int g d\rho \leq \inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int \psi d\rho = f(y), \tag{2.2}$$

where the rightmost equality is by definition of f.

On the other hand, for any such ρ, we have the following by Jensen:

$$\int f(z)d\rho(z) \geq f\left(\int z d\rho(z)\right) = f(y) \implies \inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int f d\rho \geq f(y). \tag{2.3}$$

This, combined with (2.2), yields

$$f(y) = \inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int g d\rho.$$

From this, one gathers that

$$f(y) = \inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int g d\rho \geq g(y) \geq f(y),$$

where the first inequality is simply (2.3) with g instead of f and the second inequality is by assumption. So f and g are equal, which proves that f is maximal.

So we have shown that $\mathrm{conv}\ \psi = f$. Let us now show that $\mathrm{conv}\ \psi \in \mathcal{C}_2(\mathbb{R}^d)$ for $\psi \in \mathcal{C}_2(\mathbb{R}^d)$. Since $\psi \in \mathcal{C}_2(\mathbb{R}^d)$, we have $a, b, \ell \in \mathbb{R}$ such that

$$\ell + \frac{1}{2}|\cdot|^2 \leq \psi(\cdot) \leq a + b|\cdot|^2.$$

Fix $y \in \mathbb{R}^d$. Integrating the above inequality and optimizing over $\rho \in \mathcal{P}_2^y(\mathbb{R}^d)$ yields

$$\ell + \frac{1}{2}\underbrace{\inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int |z|^2 d\rho(z)}_{=(\mathrm{conv}\ |\cdot|^2)(y)=|y|^2} \leq \underbrace{\inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int \psi d\rho}_{=(\mathrm{conv}\ \psi)(y)} \leq a + b\underbrace{\inf_{\rho\in\mathcal{P}_2^y(\mathbb{R}^d)}\int |z|^2 d\rho(z)}_{=(\mathrm{conv}\ |\cdot|^2)(y)=|y|^2}.$$

This shows that $\mathrm{conv}\,\psi$ has the appropriate bounds. To show continuity, notice that the inequality above implies $\mathrm{conv}\ \psi$ is finite on $\mathbb{R}^d$. Convex functions are continuous on the relative interior of their domain, which in this case is simply $\mathbb{R}^d$.

So not only is $\mathrm{conv}\ \psi \in \mathcal{C}_2(\mathbb{R}^d)$, but it obeys the same bounding constants as ψ. $\square$

We are now ready to prove Proposition 2.7.

Proof of Proposition 2.7. Let $\epsilon > 0$, $\psi \in \mathcal{C}_2(\mathbb{R}^d)$ and $\{\rho_x\}_{x\in\mathbb{R}^d} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ be a measurable collection of probability measures with $\mathrm{bary}(\rho_x) = x$. It suffices to show that there exists another measurable collection $\{\bar{\rho}_x\}_{x\in\mathbb{R}^d} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ with $\mathrm{bary}(\bar{\rho}_x) = x$ such that

$$\mathrm{MCov}(\rho_x, \gamma) + \int \mathrm{conv}\ \psi d(\nu - \rho_x) \leq \mathrm{MCov}(\bar{\rho}_x, \gamma) + \int \psi d(\nu - \bar{\rho}_x) + \epsilon. \tag{2.4}$$

To construct $\{\bar{\rho}_x\}$, we use Lemma 2.8 and a measurable selection argument to choose a measurable function $\mathbb{R}^d \ni y \mapsto \xi_y \in \mathcal{P}_2^y(\mathbb{R}^d)$ such that

$$\int \psi d\xi_y \leq (\text{conv } \psi)(y) + \epsilon. \tag{2.5}$$

We can now define $\bar{\rho}_x$ by $d\bar{\rho}_x(z) := \int d\xi_y(z) d\rho_x(y)$. Since each ξ_y has mean y, the measure $\bar{\rho}_x$ will have mean x. Now if we integrate (2.5) with respect to ρ_x we get

$$\int \psi d\bar{\rho}_x \leq \int \text{conv } \psi d\rho_x + \epsilon. \tag{2.6}$$

To see that $\bar{\rho}_x \in \mathcal{P}_2^x(\mathbb{R}^d)$, note that ψ, conv $\psi \in \mathcal{C}_2(\mathbb{R}^d)$ and so

$$\ell + \frac{1}{2} \int |y|^2 d\bar{\rho}_x(y) \leq a + b \int |y|^2 d\rho_x(y) + \epsilon < \infty.$$

We now estimate the maximal covariance.

$$\begin{aligned} \text{MCov}(\bar{\rho}_x, \gamma) &= \inf_{\substack{\phi\colon \mathbb{R}^d \to \mathbb{R} \\ \text{convex}}} \left(\iint \phi(z) d\xi_y(z) d\rho_x(y) + \int \phi^* d\gamma \right) \\ &\geq \inf_{\substack{\phi\colon \mathbb{R}^d \to \mathbb{R} \\ \text{convex}}} \left(\int \phi d\rho_x + \int \phi^* d\gamma \right) = \text{MCov}(\rho_x, \gamma), \end{aligned} \tag{2.7}$$

where the equalities are by the classical Kantorovich duality, i.e., Theorem 1.10 and the inequality is due to Jensen.

Now by definition of conv ψ, we know that conv $\psi \leq \psi$, and so

$$\int \text{conv } \psi d\nu \leq \int \psi d\nu. \tag{2.8}$$

Finally, combining (2.6), (2.7), and (2.8) proves (2.4).

□

The techniques employed above will play a major role in inspiring the proof of our main result. Essentially, the main result generalizes the proposition above sufficiently to also include the second motivating example.

2.2 Barycentric Costs

In this section, we discuss the second motivating example. Despite its proof being much simpler than the one we saw in the previous section, it does not generalize well. The result first appeared in [16, Theorem 2.11(3)], but the formulation found in a later paper [15, Theorem 1.1] is slightly more convenient to work with.

Theorem 2.9 ([15, Theorem 1.1]). *Let $\theta : \mathbb{R}^d \to [0, \infty)$ be a convex function and $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, then*

$$\overline{\mathcal{T}}_\theta(\nu|\mu) := \inf_{\pi \in \Pi(\mu,\nu)} \int \theta\left(\int y d\pi_x(y) - x\right) d\mu(x) = \sup_{\psi} \left\{\int Q_\theta \psi d\mu - \int \psi d\nu\right\},$$

where

$$Q_\theta \psi(x) = \inf_{y \in \mathbb{R}^d} \{\psi(y) + \theta(y - x)\}, \quad x \in \mathbb{R}^d,$$

and where the supremum runs over the set of all functions ψ that are convex, Lipschitz, and bounded from below.

Proof. Define the set $\Phi_{b,1}$ as

$$\Phi_{b,1} := \{\psi : \mathbb{R}^d \to \mathbb{R} \text{ continuous} \mid \exists a, b, \ell \in \mathbb{R},\ \ell \le \psi(\cdot) \le a + b|\cdot|\}.$$

Note that assuming that a and b are nonnegative does not affect the set.

By [16, Theorem 2.11(1)], we have the following duality:

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_{\psi \in \Phi_{b,1}} \left\{\int \overline{Q}_\theta \psi(x) d\mu(x) - \int \psi(y) d\nu(y)\right\},$$

where

$$\overline{Q}_\theta \psi(x) = \inf_{\rho \in \mathcal{P}_1(\mathbb{R}^d)} \left\{\int \psi(y) d\rho(y) + \theta\left(\int y d\rho(y) - x\right)\right\}.$$

Notice that the proof of Lemma 2.8 also works for $\psi \in \Phi_{b,1}$, mutatis mutandis. Then we have that

$$\begin{aligned}
\overline{Q}_\theta \psi(x) &= \inf_{\rho \in \mathcal{P}_1(\mathbb{R}^d)} \{\rho(\psi) + \theta(\mathrm{bary}(\rho) - x)\} \\
&= \inf_{z \in \mathbb{R}^d} \left\{\inf_{\rho \in \mathcal{P}_1^z(\mathbb{R}^d)} \rho(\psi) + \theta(z - x)\right\} \\
&= Q_\theta(\mathrm{conv}\, \psi)(x).
\end{aligned}$$

We get the following string of inequalities:

$$\begin{aligned}
\overline{\mathcal{T}}_\theta(\nu|\mu) &= \sup_{\psi\in\Phi_{b,1}} \left\{ \int \overline{Q}_\theta \psi d\mu - \int \psi d\nu \right\} \\
&= \sup_{\psi\in\Phi_{b,1}} \left\{ \int Q_\theta(\text{conv } \psi) d\mu - \int \psi d\nu \right\} \\
&\leq \sup_{\psi\in\Phi_{b,1}} \left\{ \int Q_\theta(\text{conv } \psi) d\mu - \int \text{conv } \psi d\nu \right\} \\
&\leq \sup_{\substack{\psi\in\Phi_{b,1} \\ \text{convex}}} \left\{ \int Q_\theta \psi d\mu - \int \psi d\nu \right\} \\
&\leq \sup_{\psi\in\Phi_{b,1}} \left\{ \int \overline{Q}_\theta \psi d\mu - \int \psi d\nu \right\} = \overline{\mathcal{T}}_\theta(\nu|\mu).
\end{aligned}$$

Note that we have the set equality

$$\Phi_{b,1} \cap \{\psi \text{ convex}\} = \{\psi : \mathbb{R}^d \to \mathbb{R}, \text{ convex, Lipschitz, and bounded from below}\}.$$

To see this, let $\psi \in \Phi_{b,1}$ be convex. Then the convexity of ψ will imply the convexity of the function $f : \mathbb{R} \to \mathbb{R}$, $f(t) = \psi(y + t(x - y))$ for any fixed $x, y \in \mathbb{R}^d$. We know that a convex function on $\mathbb{R}$ has a nondecreasing slope, so

$$\frac{f(1) - f(0)}{1 - 0} \leq \frac{f(t) - f(0)}{t - 0}, \quad \text{for } t \geq 1.$$

This simplifies to

$$\psi(x) - \psi(y) \leq \frac{1}{t}(a + b|y| + bt|x - y| - \psi(y)) \xrightarrow{t\to\infty} b|x - y|.$$

By symmetry, we have

$$|\psi(x) - \psi(y)| \leq b|x - y|.$$

This shows one inclusion. For the other inclusion, simply note that if ψ is convex, Lipschitz, and bounded from below, then it is continuous and

$$\begin{aligned}
\ell \leq \psi(x) \leq |\psi(x)| &\leq |\psi(x) - \psi(0)| + |\psi(0)| \\
&\leq L|x - 0| + |\psi(0)| \\
&= a + b|x|,
\end{aligned}$$

where $a = |\psi(0)|$ and $b = L$, the Lipschitz constant of ψ.

So we have

$$\begin{aligned}
&\inf_{\pi\in\Pi(\mu,\nu)} \int \theta \left(\int y d\pi_x(y) - x \right) d\mu(x) = \\
&\qquad \sup \left\{ \int Q_\theta \psi d\mu - \int \psi d\nu \,\middle|\, \psi \text{ convex, Lipschitz, and bounded from below} \right\}.
\end{aligned}$$

□

A keen observer might spot certain similarities between the two examples. Apart from the main similarity that inspired this thesis, their proofs also exhibit some parallels, a notable one being the dependence on their respective versions of Lemma 2.8.

On the other hand, the two results differ not only in moment assumptions but also in the fact that the first proof requires an explicit construction whilst the second one does not.

3 The Main Result

In this chapter, we present the main result of the thesis, which will unify the two examples from Chapter 2. We devise a condition related to the cost function C that allows us to restrict the dual formulation of a weak transport problem to convex functions. To establish duality in our main result we rely on Theorem 1.36, but instead of having a general Y, we assume that $Y = \mathbb{R}^d$ and d_Y is the Euclidean metric. We denote the associated space of probability measures as $\mathcal{P}_p(\mathbb{R}^d)$ (instead of $\mathcal{P}_{p,d_Y}(Y)$), where $p \in [1, \infty)$. We also require the set

$$\mathcal{C}_p = \mathcal{C}_p(\mathbb{R}^d) := \left\{ \psi : \mathbb{R}^d \to \mathbb{R} \text{ continuous} \,\middle|\, \exists a, b, \ell \in \mathbb{R},\ \ell + 1/2|\cdot|^p \leq \psi(\cdot) \leq a + b|\cdot|^p \right\}.$$

Compare the set $\mathcal{C}_p$ with $\Phi_{b,p}$ (see Section 1.5), which simplifies to the following in the case of $\mathbb{R}^d$ with the Euclidean metric:

$$\Phi_{b,p}(\mathbb{R}^d) = \left\{ \psi : \mathbb{R}^d \to \mathbb{R} \text{ continuous} \,\middle|\, \exists a, b, \ell \in \mathbb{R},\ \ell \leq \psi(\cdot) \leq a + b|\cdot|^p \right\}.$$

3.1 Auxiliary Lemmas

This section is devoted to three lemmas that will prove instrumental in our pursuit of the main result.

Lemma 3.1. *Let $\{\rho_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}_p(\mathbb{R}^d)$ and $\rho \in \mathcal{P}_p(\mathbb{R}^d)$. If either:*

1. *$p > 1$, $\rho_n \to \rho$ weakly, and $\int |x|^p d\rho_n(x)$ bounded, or*
2. *$p = 1$ and $\mathcal{W}_1(\rho_n, \rho) \to 0$,*

then

$$\lim_{n\to\infty} \int x d\rho_n = \int x d\rho.$$

In particular, the conclusion holds whenever $\mathcal{W}_p(\rho_n, \rho) \to 0$ for $p \geq 1$.

Proof. Let us assume that $p > 1$, $\rho_n \to \rho$ weakly, and $\int |x|^p d\rho_n(x)$ is bounded. Then Skorokhod's[1] representation theorem guarantees the existence of $\mathbb{R}^d$-valued random variables $\{X_n\}_{n\in\mathbb{N}}$ and X on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \sim \rho_n$, $X \sim \rho$, and $X_n \xrightarrow{\mathbb{P}\text{-a.s.}} X$.

Then the boundedness assumption translates to

$$\mathbb{E}\left[|X_n|^p\right] < M, \quad \forall n \in \mathbb{N}.$$

[1] Анатолій В. Скороход [Anatoliy V. Skorokhod] (1930-2011), Ukrainian mathematician

So the collection of random variables $\{X_n\}_{n\in\mathbb{N}}$ is bounded in L^p, and since $p > 1$, this implies uniform integrability. But $\mathbb{P}$-a.s. convergence combined with uniform integrability implies that

$$X_n \xrightarrow{L^1} X, \quad \text{i.e.,} \quad \mathbb{E}\left[|X_n - X|\right] \xrightarrow{n\to\infty} 0.$$

Since both X and X_n are integrable, we have

$$\mathbb{E}\left[X_n\right] = \mathbb{E}\left[X_n - X\right] + \mathbb{E}\left[X\right] \xrightarrow{n\to\infty} \mathbb{E}\left[X\right],$$

where the limit follows from

$$\lim_{n\to\infty} \left|\mathbb{E}\left[X_n - X\right]\right| \leq \lim_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] = 0 \implies \lim_{n\to\infty} \mathbb{E}\left[X_n - X\right] = 0.$$

So indeed,

$$\lim_{n\to\infty} \int x d\rho_n = \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[X\right] = \int x d\rho(x).$$

Assume now that $p = 1$ and $\mathcal{W}_1(\rho_n, \rho) \to 0$. Let X and $\{X_n\}_{n\in\mathbb{N}}$ be random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \sim \rho_n$ and $X \sim \rho$. Then by Theorem 1.17, $\mathcal{W}_1(\rho_n, \rho) \to 0$ is equivalent to

$$\mathbb{E}[\phi(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[\phi(X)], \quad \forall \phi : \mathbb{R}^d \to \mathbb{R}, \text{ continuous, } |\phi| \leq A(1 + |\cdot|). \tag{3.1}$$

Then the following holds:

$$\mathbb{E}[X_n] = \begin{bmatrix} \mathbb{E}[X_n^1] \\ \mathbb{E}[X_n^2] \\ \vdots \\ \mathbb{E}[X_n^d] \end{bmatrix} \xrightarrow{n\to\infty} \begin{bmatrix} \mathbb{E}[X^1] \\ \mathbb{E}[X^2] \\ \vdots \\ \mathbb{E}[X^d] \end{bmatrix} = \mathbb{E}[X],$$

with the coordinates of random vectors being indicated by the superscript. To see that the convergence above holds, fix $1 \leq k \leq d$ and take $\phi(x) = \mathrm{proj}_k(x) = x_k$. Then ϕ is continuous and

$$|\phi(x)| = |x_k| = \sqrt{x_k^2} \leq \sqrt{x_1^2 + \cdots + x_k^2 + \cdots + x_d^2} = |x|.$$

So by (3.1) we have

$$\mathbb{E}[X_n^k] = \mathbb{E}[\phi(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[\phi(X)] = \mathbb{E}[X^k].$$

This concludes the second case.

As for the *in particular* statement, if we have $\mathcal{W}_p(\rho_n, \rho) \to 0$ for $p > 1$, then by Theorem 1.17, $\rho_n \to \rho$ weakly and

$$\int |x|^p d\rho_n \to \int |x|^p d\rho.$$

Since the sequence of integrals on the left converges to the finite value on the right, we have that $\int |x|^p d\rho_n$ is bounded. □

Lemma 3.2. *Let $\psi \in \mathcal{C}_p(\mathbb{R}^d)$. Then*

$$(\mathrm{conv}\, \psi)(y) = \inf_{\rho \in \mathcal{P}_p^y(\mathbb{R}^d)} \int \psi d\rho, \quad y \in \mathbb{R}^d$$

and $\mathrm{conv}\, \psi \in \mathcal{C}_p(\mathbb{R}^d)$*, where* $\mathcal{P}_p^y(\mathbb{R}^d)$ *is the set of probability measures* ρ *in* $\mathcal{P}_p(\mathbb{R}^d)$ *such that* $\int z d\rho(z) = y$.

Proof. Write p instead of 2 in the proof of Lemma 2.8. ☐

Lemma 3.3. *Let $\psi \in \mathcal{C}_p(\mathbb{R}^d)$ and $\epsilon > 0$. Then there exits a Borel measurable collection* $\{\xi_y\}_{y\in\mathbb{R}^d}$ *with* $\mathrm{bary}(\xi_y) = y$ *such that*

$$\mathrm{conv}\, \psi(y) + \epsilon \geq \int \psi d\xi_y.$$

Proof. Define the following set-valued function:

$$F : \mathbb{R}^d \to \left\{ A \subseteq \mathcal{P}_p(\mathbb{R}^d) \,\middle|\, A \text{ non-empty and closed} \right\},$$
$$F : y \mapsto \left\{ \rho \in \mathcal{P}_p(\mathbb{R}^d) \,\middle|\, \mathrm{conv}\, \psi(y) + \epsilon \geq \int \psi d\rho, \int z d\rho(z) = y \right\}.$$

Note that the non-emptiness of $F(y)$ follows directly from Lemma 3.2, whereas seeing that it is closed requires some effort. Let $\{\rho_n\}$ be a sequence of measures in $F(y)$ such that for some measure $\rho \in \mathcal{P}_p(\mathbb{R}^d)$ we have

$$\mathcal{W}_p(\rho_n, \rho) \to 0 \text{ as } n \to \infty.$$

By Theorem 1.17, this is equivalent to having

$$\int \phi d\rho_n \to \int \phi d\rho,$$

for any continuous function ϕ with $|\phi(\cdot)| \leq C(1 + |\cdot|^p)$ for some $C \in \mathbb{R}$.

We would like to show that ρ is in $F(y)$, so we check the first condition:

$$\int \psi d\rho = \lim_{n\to\infty} \int \psi d\rho_n \leq \mathrm{conv}\, \psi(y) + \epsilon,$$

where the equality follows from the fact that ψ is continuous and $|\psi| \leq C(1 + |\cdot|^p)$ for $C = \max(a, b, -l, -1/2)$.

By Lemma 3.1, we have

$$y = \lim_{n\to\infty} \int x d\rho_n = \int x d\rho.$$

So ρ is indeed in $F(y)$, which proves that $F(y)$ is closed.

I now claim that for an arbitrary open $U \subseteq \mathcal{P}_p(\mathbb{R}^d)$, the set

$$\left\{ y \in \mathbb{R}^d \,\middle|\, F(y) \cap U \neq \emptyset \right\}$$

is Borel measurable in $\mathbb{R}^d$.

By separability of $\mathcal{P}_p(\mathbb{R}^d)$, any open set can be expressed as a countable union of open balls, so it suffices to show the above for open balls. Fix $\xi \in \mathcal{P}_p(\mathbb{R}^d)$, $\delta > 0$ and set $U = B_\delta(\xi)$. Then the set above becomes

$$A := \left\{ y \in \mathbb{R}^d \,\middle|\, \exists \tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d) : \operatorname{conv} \psi(y) + \epsilon \geq \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta \right\}.$$

Our goal is to show that A is Borel. We do this in three steps. First, we show that the set does not change if the inequality becomes strict, we then show that restricting to $\tilde{\xi}$ with compact support also has no effect on the set, and finally, we apply Theorem 1.26 to conclude that A is indeed Borel.

Step 1. (Reduction to strict inequality): If A is empty, there is nothing to show. Assuming A is not empty, let $y \in A$ be arbitrary. This implies the existence of some $\tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d)$ such that

$$\operatorname{conv} \psi(y) + \epsilon \geq \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta.$$

I claim that one can always find some $\hat{\xi} \in \mathcal{P}_p(\mathbb{R}^d)$ such that

$$\operatorname{conv} \psi(y) + \epsilon > \int \psi d\hat{\xi},\ \operatorname{bary}(\hat{\xi}) = y,\ \mathcal{W}_p(\hat{\xi}, \xi) < \delta. \tag{3.2}$$

To see this, notice that Lemma 3.2 implies the existence of some $\kappa \in \mathcal{P}_p(\mathbb{R}^d)$ such that

$$\operatorname{conv} \psi(y) + \epsilon > \int \psi d\kappa,\ \operatorname{bary}(\kappa) = y.$$

Define a sequence of measures in $\mathcal{P}_p(\mathbb{R}^d)$ by

$$\tilde{\xi}_n := \left(1 - \frac{1}{n}\right) \tilde{\xi} + \frac{1}{n} \kappa.$$

Then $\mathcal{W}_p(\tilde{\xi}_n, \tilde{\xi}) \to 0$, since for any continuous $\phi : \mathbb{R}^d \to \mathbb{R}$ with $|\phi| \leq C(1 + |\cdot|^p)$, we have

$$\int \phi d\tilde{\xi}_n = \left(1 - \frac{1}{n}\right) \int \phi d\tilde{\xi} + \frac{1}{n} \int \phi d\kappa \xrightarrow{n \to \infty} \int \phi d\tilde{\xi}.$$

Take N sufficiently large such that

$$\mathcal{W}_p(\tilde{\xi}_N, \tilde{\xi}) < \underbrace{\delta - \mathcal{W}_p(\tilde{\xi}, \xi)}_{>0}.$$

Then $\hat{\xi} := \tilde{\xi}_N$ will satisfy (3.2).

This shows that

$$A \subseteq \left\{ y \in \mathbb{R}^d \,\middle|\, \exists \tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d) : \operatorname{conv} \psi(y) + \epsilon > \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta \right\}.$$

And since the other inclusion is trivial, we have

$$A = \left\{ y \in \mathbb{R}^d \,\middle|\, \exists \tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d) : \operatorname{conv} \psi(y) + \epsilon > \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta \right\}.$$

Step 2. (Reduction to compact support): Once more, assume that A is nonempty and take an arbitrary $y \in A$. Then we must have a $\tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d)$ such that

$$\operatorname{conv} \psi(y) + \epsilon > \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta.$$

This time, I claim that one can always find some $\hat{\xi} \in \mathcal{P}_p(\mathbb{R}^d)$ such that

$$\hat{\xi} \text{ compactly supported, } \operatorname{conv} \psi(y) + \epsilon > \int \psi d\hat{\xi},\ \operatorname{bary}(\hat{\xi}) = y,\ \mathcal{W}_p(\hat{\xi}, \xi) < \delta. \tag{3.3}$$

In order to construct $\hat{\xi}$, recall that compactly supported probability measures on $\mathbb{R}^d$ form a dense subset of $\mathcal{P}_p(\mathbb{R}^d)$. So let $\{\tilde{\xi}_n\} \subseteq \mathcal{P}_p(\mathbb{R}^d)$ be a sequence of compactly supported measures with $\mathcal{W}_p(\tilde{\xi}_n, \tilde{\xi}) \to 0$. Now define another sequence,

$$\tilde{\tilde{\xi}}_n := \tilde{\xi}_n * \delta_{y - \operatorname{bary}(\tilde{\xi}_n)}.$$

Then this sequence is in $\mathcal{P}_p(\mathbb{R}^d)$ since

$$\int |x|^p d\tilde{\tilde{\xi}}_n = \int |x + y - \operatorname{bary}(\tilde{\xi}_n)|^p d\tilde{\xi}_n \le 2^{p-1} \left[\int |x|^p d\tilde{\xi}_n + |y - \operatorname{bary}(\tilde{\xi}_n)|^p \right] < \infty.$$

I claim that the modified sequence still converges towards $\tilde{\xi}$. To see this, let π_n be a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\tilde{\xi}_n$ and $\tilde{\xi}$ such that

$$\mathcal{W}_p^p(\tilde{\xi}_n, \tilde{\xi}) = \int |x - z|^p d\pi_n(x, z).$$

The existence of such a π_n is classical (see Theorem 1.4). Define the mapping

$$F_n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d,\ F_n(x, z) = (x + y - \operatorname{bary}(\tilde{\xi}_n), z).$$

Then $F_n(\pi_n)$ has marginals $\tilde{\tilde{\xi}}_n$ and $\tilde{\xi}$, hence,

$$\begin{aligned}
\mathcal{W}_p^p(\tilde{\tilde{\xi}}_n, \tilde{\xi}) &\le \int |x - z|^p d(F_n(\pi_n)) \\
&= \int |x + y - \operatorname{bary}(\tilde{\xi}_n) - z|^p d\pi_n \\
&\le 2^{p-1} \left[\int |x - z|^p d\pi_n + |y - \operatorname{bary}(\tilde{\xi}_n)|^p \right] \\
&= 2^{p-1} \big[\underbrace{\mathcal{W}_p^p(\tilde{\xi}_n, \tilde{\xi})}_{\to 0} + \underbrace{|\operatorname{bary}(\tilde{\xi}) - \operatorname{bary}(\tilde{\xi}_n)|^p}_{\to 0, \text{ by Lemma 3.1}} \big] \xrightarrow{n \to \infty} 0.
\end{aligned}$$

So indeed, $\mathcal{W}_p(\tilde{\tilde{\xi}}_n, \tilde{\xi}) \to 0$. By Theorem 1.17, this also implies that

$$\int \psi d\tilde{\tilde{\xi}}_n \to \int \psi d\tilde{\xi}.$$

Define $\hat{\xi} := \tilde{\tilde{\xi}}_N$, where N is sufficiently large to have

$$\left| \int \psi d\tilde{\tilde{\xi}}_N - \int \psi d\tilde{\xi} \right| < \underbrace{\left(\operatorname{conv} \psi(y) + \epsilon - \int \psi d\tilde{\xi} \right)}_{>0} \text{ and } \mathcal{W}_p(\tilde{\tilde{\xi}}_N, \tilde{\xi}) < \underbrace{\delta - \mathcal{W}_p(\tilde{\xi}, \xi)}_{>0}.$$

Then $\hat{\xi}$ will satisfy (3.3).

Thus we have shown

$$A \subseteq \left\{ y \in \mathbb{R}^d \,\middle|\, \exists \tilde{\xi} \in \mathcal{P}_{\text{c.s.}}(\mathbb{R}^d) : \operatorname{conv} \psi(y) + \epsilon > \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta \right\}.$$

The other inclusion is trivial, so

$$A = \left\{ y \in \mathbb{R}^d \,\middle|\, \exists \tilde{\xi} \in \mathcal{P}_{\text{c.s.}}(\mathbb{R}^d) : \operatorname{conv} \psi(y) + \epsilon > \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta \right\}.$$

Step 3. (σ-compactness of the sections): Notice that the set A may be written as

$$\operatorname{proj}_1 \left(\left\{ (y, \tilde{\xi}) \in \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \,\middle|\, \operatorname{conv} \psi(y) + \epsilon > \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta,\ \tilde{\xi} \text{ c.s.} \right\} \right).$$

For the sake of brevity, denote the set being projected as $B \subseteq \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$. In order to use Theorem 1.26, we must show that B is Borel and that its sections are σ-compact.

To see that B is Borel, write

$$B = \bigcup_{R=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{\infty} B_{R,k,j},$$

where $B_{R,k,j}$ is given by

$$\begin{aligned} \Bigg\{ (y, \tilde{\xi}) \in \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \,\Bigg|\, \operatorname{conv} \psi(y) + \epsilon - \frac{1}{j} \geq \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,& \\ \mathcal{W}_p(\tilde{\xi}, \xi) \leq \delta - \frac{1}{k},\ \tilde{\xi}\left(\overline{B_R(0)} \right) = 1 \Bigg\}&. \end{aligned}$$

Then one can write

$$\begin{aligned} B_{R,k,j} = \mathbb{R}^d \times \mathcal{P}\left(\overline{B_R(0)} \right) \cap \left\{ (y, \tilde{\xi}) \,\middle|\, y = \operatorname{bary}(\tilde{\xi}) \right\} \cap \mathbb{R}^d \times \overline{B_{\delta - \frac{1}{k}}(\xi)} \\ \cap \left\{ (y, \tilde{\xi}) \,\middle|\, \operatorname{conv} \psi(y) + \epsilon - \frac{1}{j} \geq \int \psi d\tilde{\xi} \right\}. \end{aligned}$$

The first set is a product of a closed set with a compact one (by Corollary 1.23), thus closed in the product topology. The fact that the second set is closed follows from Lemma 3.1. The third set is a product of two closed sets, thus again closed in the product topology. Finally, to see that the fourth set is closed, one uses continuity of conv ψ and Theorem 1.17.

This shows that $B_{R,k,j}$ is closed, so B is a countable union of closed sets, in particular, B is a Borel set.

We now show that its sections are σ-compact. Consider an arbitrary section of B,

$$B_y = \left\{ \tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d) \,\middle|\, \operatorname{conv} \psi(y) + \epsilon > \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) < \delta,\ \tilde{\xi} \text{ c.s.} \right\}.$$

Then we have

$$B_y = \bigcup_{R=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{\infty} B_y^{R,k,j},$$

where $B_y^{R,k,j}$ is given by

$$\begin{aligned}&\left\{ \tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d) \,\middle|\, \operatorname{conv} \psi(y) + \epsilon - \frac{1}{j} \geq \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) \leq \delta - \frac{1}{k},\ \tilde{\xi}\left(\overline{B_R(0)}\right) = 1 \right\} \\ =&\left\{ \tilde{\xi} \in \mathcal{P}_p(\mathbb{R}^d) \,\middle|\, \operatorname{conv} \psi(y) + \epsilon - \frac{1}{j} \geq \int \psi d\tilde{\xi},\ \operatorname{bary}(\tilde{\xi}) = y,\ \mathcal{W}_p(\tilde{\xi}, \xi) \leq \delta - \frac{1}{k} \right\} \cap \mathcal{P}\left(\overline{B_R(0)}\right).\end{aligned}$$

One can readily check that the first set is closed, whereas the second set is compact by Corollary 1.23, so $B_y^{R,k,j}$ is also compact. This proves that B has σ-compact sections.

Finally, by Theorem 1.26, A is a Borel set. We conclude that F satisfies the assumptions of Theorem 1.30, so we may extract a Borel measurable collection $\{\xi_y\}_{y \in \mathbb{R}^d}$ such that $\xi_y \in F(y)$.

□

3.2 Main Result and Proof

With the preliminaries out of the way, we move on to the central result of the thesis.

Theorem 3.4. *Let $C : X \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R} \cup \{\infty\}$ be jointly lower semicontinuous with respect to the product topology on $X \times \mathcal{P}_p(\mathbb{R}^d)$ and convex in the second argument such that $\widetilde{C}(x, \rho) := C(x, \rho) + \rho\left(|\cdot|^p/2\right)$ is bounded from below. Assume that $C(x, \rho)$ satisfies the following implication whenever $\rho_1, \rho_2 \in \mathcal{P}_p(\mathbb{R}^d)$:*

$$\rho_1 \preceq_{\mathrm{c}} \rho_2 \implies C(x, \rho_1) \geq C(x, \rho_2). \tag{C1}$$

Then we have for all $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}_p(\mathbb{R}^d)$

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_X C(x, \pi_x) d\mu(x) = \sup_{\psi \in \mathcal{C}_p} \mu(R_C \psi) - \nu(\psi) = \sup_{\substack{\psi \in \mathcal{C}_p \\ \psi \text{ convex}}} \mu(R_C \psi) - \nu(\psi).$$

Proof. Let $C : X \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R} \cup \{\infty\}$ be jointly lower semicontinuous with respect to the product topology on $X \times \mathcal{P}_p(\mathbb{R}^d)$, convex in the second argument and $\widetilde{C}$ bounded from below.

Note that $\mu(R_C\psi)$ is well-defined for $\psi \in \mathcal{C}_p$ since

$$\int R_C\psi d\mu = \int \inf_{\rho\in\mathcal{P}_p(\mathbb{R}^d)} \rho(\psi) + C(x,\rho)d\mu(x) = \int \inf_{\rho\in\mathcal{P}_p(\mathbb{R}^d)} \rho(\psi - |\cdot|^p/2) + \widetilde{C}(x,\rho)d\mu(x),$$

where the function being integrated with respect to ρ is bounded from below by ℓ and $\widetilde{C}$ is bounded from below by assumption.

We start the proof by showing that for all $\psi \in \mathcal{C}_p(\mathbb{R}^d)$ we have

$$R_C\psi = R_C \mathrm{conv}\, \psi.$$

The inequality "$\geq$" follows from the fact that ψ dominates $\mathrm{conv}\,\psi$. The other inequality is equivalent to the following statement:

$$\forall \epsilon > 0, \forall \rho \in \mathcal{P}_p(\mathbb{R}^d) \exists \tilde{\rho} \in \mathcal{P}_p(\mathbb{R}^d) : \tilde{\rho}(\psi) + C(x,\tilde{\rho}) \leq \rho(\mathrm{conv}\, \psi) + C(x,\rho) + \epsilon. \tag{3.4}$$

To see that (3.4) holds, we fix $\psi \in \mathcal{C}_p(\mathbb{R}^d)$ and $x \in X$. Let $\epsilon > 0$ and let $\rho \in \mathcal{P}_p(\mathbb{R}^d)$ be arbitrary.

Step 1. (Construction of $\tilde{\rho}$): By Lemma 3.3, there exists a Borel measurable collection $\{\xi_y\}_{y\in\mathbb{R}^d}$ with $\mathrm{bary}(\xi_y) = y$ such that

$$\mathrm{conv}\, \psi(y) + \epsilon \geq \int \psi d\xi_y. \tag{3.5}$$

We define $\tilde{\rho}$ by

$$d\tilde{\rho}(z) := \int d\xi_y(z) d\rho(y).$$

Step 2. (Showing $\tilde{\rho} \in \mathcal{P}_p(\mathbb{R}^d)$) Integrating inequality (3.5) with respect to ρ and recalling the bounds on ψ and $\mathrm{conv}\,\psi$, we get

$$\int \mathrm{conv}\, \psi d\rho + \epsilon \geq \int \psi d\tilde{\rho} \implies a + b\int |z|^p d\rho(z) + \epsilon \geq \ell + \frac{1}{2}\int |z|^p d\tilde{\rho}(z). \tag{3.6}$$

So indeed, $\tilde{\rho} \in \mathcal{P}_p(\mathbb{R}^d)$.

Step 3. (The inequality in (3.4) is satisfied) Notice that from (3.6) we have

$$\int \psi d\tilde{\rho} \leq \int \mathrm{conv}\, \psi d\rho + \epsilon. \tag{3.7}$$

To see that $C(x,\tilde{\rho}) \leq C(x,\rho)$, it suffices to show that $\rho \preceq_{\mathrm{c}} \tilde{\rho}$. So let $f : \mathbb{R}^d \to \mathbb{R}$ be an arbitrary convex function. Then

$$\int f(y)d\rho(y) = \int f\left(\int z d\xi_y(z)\right) d\rho(y) \leq \iint f(z)d\xi_y(z)d\rho(y) = \int f(z)d\tilde{\rho}(z),$$

where the inequality is due to Jensen. So $\rho \preceq_{\mathrm{c}} \tilde{\rho}$ and thus by assumption,

$$C(x, \tilde{\rho}) \leq C(x, \rho). \tag{3.8}$$

Adding up (3.7) with (3.8) yields

$$\tilde{\rho}(\psi) + C(x, \tilde{\rho}) \leq \rho(\text{conv } \psi) + C(x, \rho) + \epsilon.$$

So we have succeeded in showing that $R_C\psi = R_C\text{conv } \psi$. Introduce the following notation:

$$\mathcal{D}(\psi) := \mu(R_C\psi) - \nu(\psi), \quad \psi \in \mathcal{C}_p(\mathbb{R}^d).$$

Then notice that

$$\begin{aligned} \mathcal{D}(\psi) &= \mu(R_C\psi) - \nu(\psi) \\ &= \mu(R_C\text{conv } \psi) - \nu(\psi) \\ &\leq \mu(R_C\text{conv } \psi) - \nu(\text{conv } \psi) \\ &= \mathcal{D}(\text{conv } \psi). \end{aligned} \tag{3.9}$$

Now consider the following string of inequalities:

$$\begin{aligned} \sup_{\psi \in \mathcal{C}_p(\mathbb{R}^d)} \mathcal{D}(\psi) &\leq \sup_{\psi \in \mathcal{C}_p(\mathbb{R}^d)} \mathcal{D}(\text{conv } \psi) && (3.10) \\ &= \sup_{\substack{\phi = \text{conv } \psi \\ \psi \in \mathcal{C}_p(\mathbb{R}^d)}} \mathcal{D}(\phi) && (3.11) \\ &\leq \sup_{\substack{\psi \in \mathcal{C}_p(\mathbb{R}^d) \\ \psi \text{ convex}}} \mathcal{D}(\psi) && (3.12) \\ &\leq \sup_{\psi \in \mathcal{C}_p(\mathbb{R}^d)} \mathcal{D}(\psi), && (3.13) \end{aligned}$$

where (3.10) follows from (3.9). Inequality (3.12) holds since $\{\text{conv } \psi \mid \psi \in \mathcal{C}_p\} \subseteq \{\psi \in \mathcal{C}_p \mid \psi \text{ convex}\}$.

Hence, (3.10)-(3.13) are equalities and in particular

$$\sup_{\psi \in \mathcal{C}_p(\mathbb{R}^d)} \mathcal{D}(\psi) = \sup_{\substack{\psi \in \mathcal{C}_p(\mathbb{R}^d) \\ \psi \text{ convex}}} \mathcal{D}(\psi). \tag{3.14}$$

To complete the proof, we must show that the suprema in (3.14) are equal to the primal problem. Notice that $\widetilde{C}(x, \rho) = C(x, \rho) + \rho(|\cdot|^p/2)$ satisfies the assumptions of Theorem 1.36. Indeed, the integral term is a linear function of ρ, so the convexity of $\widetilde{C}$ follows, since both terms are convex in ρ. The lower semicontinuity follows from the assumed lower semicontinuity of C and the second term's continuity in ρ. Lastly, $\widetilde{C}$ is bounded from below by assumption.

Then we can do the following:

$$
\begin{aligned}
&\inf_{\pi\in\Pi(\mu,\nu)}\int C(x,\pi_x)d\mu(x)\\
&=\inf_{\pi\in\Pi(\mu,\nu)}\left[\int C(x,\pi_x)d\mu(x)+\nu\left(\frac{1}{2}|\cdot|^p\right)\right]-\nu\left(\frac{1}{2}|\cdot|^p\right)\\
&=\inf_{\pi\in\Pi(\mu,\nu)}\left[\int C(x,\pi_x)+\pi_x(|\cdot|^p/2)d\mu(x)\right]-\nu\left(\frac{1}{2}|\cdot|^p\right)\\
&=\inf_{\pi\in\Pi(\mu,\nu)}\int \widetilde{C}(x,\pi_x)d\mu(x)-\nu\left(\frac{1}{2}|\cdot|^p\right)\\
&\overset{*}{=}\sup_{\psi\in\Phi_{b,p}}\left[\mu(R_{\widetilde{C}}\psi)-\nu(\psi)\right]-\nu\left(\frac{1}{2}|\cdot|^p\right)\\
&=\sup_{\psi\in\Phi_{b,p}}\left[\mu\left(R_C\left(\psi+\frac{1}{2}|\cdot|^p\right)\right)-\nu\left(\psi+\frac{1}{2}|\cdot|^p\right)\right]\\
&=\sup_{\psi\in\mathcal{C}_p}\mathcal{D}(\psi)\\
&=\sup_{\substack{\psi\in\mathcal{C}_p\\ \psi\text{ convex}}}\mathcal{D}(\psi),
\end{aligned}
$$

where the starred equality is by Theorem 1.36 and the last equality is by (3.14). This finalizes the proof.

$\square$

Notice that in Theorem 3.4, the auxiliary cost function $\widetilde{C}$ was constructed by adding $\rho(|\cdot|^p/2)$ to the main cost function C. The $1/2$ constant is by no means special. Indeed, one could replace it with any other constant $\kappa>0$ and the proof remains largely the same. So with the appropriate adjustments, one also gets the slightly stronger result stated below.

Theorem 3.5. *Let $C: X\times\mathcal{P}_p(\mathbb{R}^d)\to\mathbb{R}\cup\{\infty\}$ be jointly lower semicontinuous with respect to the product topology on $X\times\mathcal{P}_p(\mathbb{R}^d)$ and convex in the second argument such that $\widetilde{C}(x,\rho):=C(x,\rho)+\rho\left(\kappa|\cdot|^p\right)$ is bounded from below for some fixed $\kappa>0$. Assume that $C(x,\rho)$ satisfies the following implication whenever $\rho_1,\rho_2\in\mathcal{P}_p(\mathbb{R}^d)$:*

$$\rho_1\preceq_{\mathrm{c}}\rho_2\implies C(x,\rho_1)\geq C(x,\rho_2).$$

Then we have for all $\mu\in\mathcal{P}(X)$ and $\nu\in\mathcal{P}_p(\mathbb{R}^d)$

$$\inf_{\pi\in\Pi(\mu,\nu)}\int_X C(x,\pi_x)d\mu(x)=\sup_{\psi\in\mathcal{C}_p^\kappa}\mu(R_C\psi)-\nu(\psi)=\sup_{\substack{\psi\in\mathcal{C}_p^\kappa\\ \psi\text{ convex}}}\mu(R_C\psi)-\nu(\psi),$$

where $\mathcal{C}_p^\kappa$ is given by

$$\mathcal{C}_p^\kappa:=\left\{\psi:\mathbb{R}^d\to\mathbb{R}\text{ continuous}\,\middle|\,\exists a,b,\ell\in\mathbb{R},\ \ell+\kappa|\cdot|^p\leq\psi(\cdot)\leq a+b|\cdot|^p\right\}.$$

In the adjusted proof, one uses the modified version of Lemma 3.2.

Lemma 3.6. *Let* $\psi \in \mathcal{C}_p^\kappa(\mathbb{R}^d)$*. Then*

$$(\operatorname{conv} \psi)(y) = \inf_{\rho \in \mathcal{P}_p^y(\mathbb{R}^d)} \int \psi d\rho, \quad y \in \mathbb{R}^d$$

and $\operatorname{conv} \psi \in \mathcal{C}_p^\kappa(\mathbb{R}^d)$*.*

From now on, we refer to Theorem 3.4 as *the main theorem* and to Theorem 3.5 as *the modified main theorem*.

4 Special Cases

This chapter considers four special cases of Theorem 3.4. Firstly, we recover our motivating examples [7, Proposition 3.3] and [15, Theorem 1.1]. Secondly, we apply the modified main theorem to a variance-favoring weak transport problem found in [1, Theorem 5.6]. Finally, we use the main theorem to recover the classical Kantorovich duality for the quadratic cost function.

4.1 Maximal Covariance Revisited

We consider the setting of [7, Proposition 3.3], i.e., Proposition 2.7, and recover the result as a consequence of Theorem 3.4.

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. To use the main theorem, set $p = 2$, $X = \mathbb{R}^d$ and

$$C(x,\rho) = \begin{cases} \min_{\tau \in \Pi(\rho,\gamma)} \int_{\mathbb{R}^d \times \mathbb{R}^d} -yz d\tau, & x = \mathrm{bary}(\rho) \\ \infty, & \text{otherwise.} \end{cases}$$

We check the assumptions of the main theorem:

- *C jointly lower semicontinuous:* We want to show that for arbitrary $r \in \mathbb{R}$ the set

  $$\left\{ (x,\rho) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \,\middle|\, C(x,\rho) \le r \right\}$$

  is closed. This set simplifies to

  $$S := \left\{ (x,\rho) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \,\middle|\, x = \mathrm{bary}(\rho),\ \min_{\Pi(\rho,\gamma)} \tau(-yz) \le r \right\}.$$

  Let $\{(x_n, \rho_n)\}_{n\in\mathbb{N}} \subseteq S$ be such that $(x_n,\rho_n) \xrightarrow{n\to\infty} (x,\rho)$. In particular, $\rho_n \to \rho$ in $\mathcal{W}_2$. By Lemma 3.1, $x = \mathrm{bary}(\rho)$. To get the other condition, we note that

  $$\begin{aligned} &\min_{\tau\in\Pi(\rho_n,\gamma)} \tau(-yz) \le r \\ \iff &\min_{\tau\in\Pi(\rho_n,\gamma)} \tau(1/2|y-z|^2) - 1/2\rho_n(|y|^2) - 1/2\gamma(|z|^2) \le r \\ \underset{n\to\infty}{\Longrightarrow} &\min_{\tau\in\Pi(\rho,\gamma)} \tau(1/2|y-z|^2) - 1/2\rho(|y|^2) - 1/2\gamma(|z|^2) \le r \\ \iff &\min_{\tau\in\Pi(\rho,\gamma)} \tau(-yz) \le r, \end{aligned}$$

  where the implication is by Corollary 1.18 and Theorem 1.17. So S is closed.

- *C convex in the second argument:* Fix $x \in \mathbb{R}^d$. We want to show that for $0 \le \alpha \le 1$ we have
$$C(x, \alpha\rho_1 + (1-\alpha)\rho_2) \le \alpha C(x, \rho_1) + (1-\alpha)C(x, \rho_2).$$
If either ρ_1 or ρ_2 don't have barycentre x, the result is trivial. If $\mathrm{bary}(\rho_1) = \mathrm{bary}(\rho_2) = x$, the convexity directly follows from the inclusion
$$\{\alpha\tau_1 + (1-\alpha)\tau_2 \mid \tau_1 \in \Pi(\rho_1, \gamma), \tau_2 \in \Pi(\rho_2, \gamma)\} \subseteq \Pi(\alpha\rho_1 + (1-\alpha)\rho_2, \gamma).$$

- *$\widetilde{C}$ bounded from below:* $\widetilde{C}(x, \rho) \ge \min_{\tau \in \Pi(\rho,\gamma)} \tau(-yz + 1/2|y|^2) \ge \gamma(-1/2|z|^2) = -d/2$.

- *C satisfies condition (C1):* Let $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d)$ with $\rho_1 \preceq_{\mathrm{c}} \rho_2$. Then ρ_1 and ρ_2 have the same barycentre, and if it's different from x, the inequality is trivial. So assume that $x = \mathrm{bary}(\rho_1) = \mathrm{bary}(\rho_2)$. Then
$$C(x, \rho_1) = \min_{\tau \in \Pi(\rho_1,\gamma)} \int -yz d\tau = \sup_{\phi \in L^1(\gamma)} \int \phi d\gamma - \int \phi^c d\rho_1,$$
where the equality is by Theorem 1.10 with $c(y, z) = -yz$. Then we have
$$\phi^c(y) = \sup_{z \in \mathbb{R}^d} (\phi(z) - c(y, z)) = \sup_{z \in \mathbb{R}^d} (\phi(z) + yz).$$
So ϕ^c is a supremum of affine functions, hence convex. We use the fact that $\rho_1 \preceq_{\mathrm{c}} \rho_2$ to get:
$$\sup_{\phi \in L^1(\gamma)} \int \phi d\gamma - \int \phi^c d\rho_1 \ge \sup_{\phi \in L^1(\gamma)} \int \phi d\gamma - \int \phi^c d\rho_2 = C(x, \rho_2).$$

So we can use the main theorem. To match the notation, we denote the *maximal covariance* by
$$\mathrm{MCov}(\rho_1, \rho_2) := \sup_{\tau \in \Pi(\rho_1,\rho_2)} \int x_1 x_2 d\tau(x_1, x_2),\ \rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d),$$
and we use the transform
$$\varphi^\psi(x) := \inf_{\rho \in \mathcal{P}_2^x(\mathbb{R}^d)} \left(\int \psi d\rho - \mathrm{MCov}(\rho, \gamma) \right).$$

We then have

$$
\begin{aligned}
&-\inf_{\psi\in\mathcal{C}_2}\int\psi d\nu-\int\varphi^\psi d\mu=\\
&-\inf_{\psi\in\mathcal{C}_2}\int\psi d\nu-\int\inf_{\rho\in\mathcal{P}_2^x(\mathbb{R}^d)}\rho(\psi)-\operatorname{MCov}(\rho,\gamma)d\mu(x)=\\
&\sup_{\psi\in\mathcal{C}_2}\int\inf_{\rho\in\mathcal{P}_2^x(\mathbb{R}^d)}\rho(\psi)-\operatorname{MCov}(\rho,\gamma)d\mu(x)-\int\psi d\nu=\\
&\sup_{\psi\in\mathcal{C}_2}\int\inf_{\rho\in\mathcal{P}_2(\mathbb{R}^d)}\rho(\psi)+C(x,\rho)d\mu(x)-\int\psi d\nu=\\
&\sup_{\psi\in\mathcal{C}_2}\mu(R_C\psi)-\nu(\psi).
\end{aligned}
\tag{4.1}
$$

But by Theorem 3.4,

$$
\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathbb{R}^d}C(x,\pi_x)d\mu(x)=\sup_{\psi\in\mathcal{C}_2}\mu(R_C\psi)-\nu(\psi)=\sup_{\substack{\psi\in\mathcal{C}_2\\ \psi\text{ convex}}}\mu(R_C\psi)-\nu(\psi).
\tag{4.2}
$$

On the other hand,

$$
\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathbb{R}^d}C(x,\pi_x)d\mu(x)=-\sup_{\pi\in\operatorname{MT}(\mu,\nu)}\int_{\mathbb{R}^d}\operatorname{MCov}(\pi_x,\gamma)d\mu(x).
$$

By the same reasoning as in (4.1), we have

$$
\sup_{\substack{\psi\in\mathcal{C}_2\\ \psi\text{ convex}}}\mu(R_C\psi)-\nu(\psi)=-\inf_{\substack{\psi\in\mathcal{C}_2\\ \psi\text{ convex}}}\int\psi d\nu-\int\varphi^\psi d\mu.
$$

So (4.2) amounts to

$$
\sup_{\pi\in\operatorname{MT}(\mu,\nu)}\int_{\mathbb{R}^d}\operatorname{MCov}(\pi_x,\gamma)d\mu(x)=\inf_{\psi\in\mathcal{C}_2}\int\psi d\nu-\int\varphi^\psi d\mu=\inf_{\substack{\psi\in\mathcal{C}_2\\ \psi\text{ convex}}}\int\psi d\nu-\int\varphi^\psi d\mu.
$$

Thus, we have recovered [7, Proposition 3.3], i.e., Proposition 2.7.

4.2 Barycentric Costs Revisited

We now reproduce the result in [15, Theorem 1.1], i.e., Theorem 2.9. Let $\mu,\nu\in\mathcal{P}_p(\mathbb{R}^d)$, and let $\theta:\mathbb{R}^d\to[0,\infty)$ be a convex function. We take $C(x,\rho)=\theta(\operatorname{bary}(\rho)-x)$.

We check the assumptions of the main theorem:

- *C jointly lower semicontinuous:* We want to show that for arbitrary $r\in\mathbb{R}$ the set

$$
S:=\left\{(x,\rho)\in\mathbb{R}^d\times\mathcal{P}_p(\mathbb{R}^d)\,\middle|\,\theta(\operatorname{bary}(\rho)-x)\le r\right\}
$$

is closed. Let $\{(x_n, \rho_n)\}_{n\in\mathbb{N}} \subseteq S$ be such that $(x_n, \rho_n) \xrightarrow{n\to\infty} (x, \rho)$. In particular, $\rho_n \to \rho$ in $\mathcal{W}_p$ and $x_n \to x$. We know that $\theta(\text{bary}(\rho_n) - x_n) \leq r$. Since θ is convex and finite on $\mathbb{R}^d$, it is continuous. Taking the limit yields

$$r \geq \lim_{n\to\infty} \theta(\text{bary}(\rho_n) - x_n) = \theta\left(\lim_{n\to\infty} \text{bary}(\rho_n) - x_n\right) = \theta(\text{bary}(\rho) - x),$$

where the convergence of barycentres is by Lemma 3.1.

- C *convex in the second argument:* This follows directly from the convexity of θ.
- $\widetilde{C}$ *bounded from below:* $\widetilde{C} \geq 0$.
- C *satisfies condition (C1):* Given $\rho_1, \rho_2 \in \mathcal{P}_p(\mathbb{R}^d)$ with $\rho_1 \preceq_c \rho_2$, we know that they have the same barycentre, hence $C(x, \rho_1) = C(x, \rho_2)$.

So using the main theorem is permitted, hence

$$\inf_{\pi\in\Pi(\mu,\nu)} \int \theta(\text{bary}(\pi_x) - x) d\mu(x) = \sup_{\substack{\psi\in\mathcal{C}_p \\ \psi \text{ convex}}} \mu(R_C\psi) - \nu(\psi). \tag{4.3}$$

This result is slightly different than the one in [15, Theorem 1.1]. Let us first show that the exact result from the paper also follows, before closing the section with a remark on the differences between the two duality results.

We start by simplifying $R_C\psi$ for convex $\psi \in \Phi_{b,p}$.

$$\begin{aligned} R_C\psi(x) &= \inf_{\rho\in\mathcal{P}_p(\mathbb{R}^d)} \rho(\psi) + \theta(\text{bary}(\rho) - x) \\ &= \inf_{z\in\mathbb{R}^d} \left[\inf_{\rho\in\mathcal{P}_p^z(\mathbb{R}^d)} \rho(\psi) + \theta(z - x) \right] \\ &= \inf_{z\in\mathbb{R}^d} \text{conv}\, \psi(z) + \theta(z - x) \\ &= \inf_{z\in\mathbb{R}^d} \psi(z) + \theta(z - x) =: Q_\theta\psi(x), \end{aligned}$$

where the penultimate equality is by Lemma 3.2 (more precisely, a variant thereof for $\Phi_{b,p}$), and the final equality is due to the convexity of ψ. For a general $\psi \in \Phi_{b,p}$ we at least have

$$R_C\psi = Q_\theta \text{conv}\, \psi.$$

Note that both of these facts also hold when $\psi \in \mathcal{C}_p \subseteq \Phi_{b,p}$.

So we have

$$\begin{aligned}
\inf_{\pi\in\Pi(\mu,\nu)} \int \theta(\mathrm{bary}(\pi_x)-x)d\mu(x) &= \sup_{\substack{\psi\in\mathcal{C}_p \\ \psi \text{ convex}}} \mu(Q_\theta\psi)-\nu(\psi) \\
&\le \sup_{\substack{\psi\in\Phi_{b,p} \\ \psi \text{ convex}}} \mu(Q_\theta\psi)-\nu(\psi) \\
&= \sup_{\substack{\psi\in\Phi_{b,p} \\ \psi \text{ convex}}} \mu(Q_\theta \mathrm{conv}\,\psi)-\nu(\psi) \\
&\le \sup_{\psi\in\Phi_{b,p}} \mu(Q_\theta \mathrm{conv}\,\psi)-\nu(\psi) \\
&= \sup_{\psi\in\Phi_{b,p}} \mu(R_C\psi)-\nu(\psi) \\
&= \inf_{\pi\in\Pi(\mu,\nu)} \int \theta(\mathrm{bary}(\pi_x)-x)d\mu(x),
\end{aligned}$$

where the last equality is by Theorem 1.36. This implies

$$\inf_{\pi\in\Pi(\mu,\nu)} \int \theta(\mathrm{bary}(\pi_x)-x)d\mu(x) = \sup_{\substack{\psi\in\Phi_{b,p} \\ \psi \text{ convex}}} \mu(Q_\theta\psi)-\nu(\psi). \tag{4.4}$$

In the special case of $p=1$, recall that we have

$$\Phi_{b,1}\cap\{\psi \text{ convex}\} = \{\psi \text{ convex, Lipschitz and bounded from below}\}.$$

Hence for $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^d)$, we retrieve [15, Theorem 1.1], i.e., Theorem 2.9:

$$\begin{aligned}
&\inf_{\pi\in\Pi(\mu,\nu)} \int \theta(\mathrm{bary}(\pi_x)-x)d\mu(x) \\
&\qquad = \sup\left\{\mu(Q_\theta\psi)-\nu(\psi) \,\middle|\, \psi:\mathbb{R}^d\to\mathbb{R} \text{ convex, Lipschitz and bounded from below}\right\}.
\end{aligned}$$

Remark 4.1. The dual formulations in (4.3) and (4.4) differ only in the domain of optimization. Notice that while constant functions are in $\Phi_{b,p}\cap\{\text{convex}\}$, they are not in $\mathcal{C}_p\cap\{\text{convex}\}$, so the latter is a strict subset of the former. We were able to pass back to $\Phi_{b,p}$ by using Theorem 1.36 because C itself was bounded from below.

4.3 Variance-Favoring Cost

This section is dedicated to [1, Theorem 5.6], an example that was kindly suggested by Julio Backhoff.

Theorem 4.2 ([1, Theorem 5.6]). *Let X be the closure of a bounded convex subset of $\mathbb{R}^d$, $\lambda>0$, and μ,ν bounded nonnegative Borel measures on X. Then the primal problem*

$$F_\lambda := \inf_{\pi\in\Pi(\mu,\nu)} \left\{\lambda\int_{X\times X}|x-y|^2d\pi(x,y) - \int_X \mathrm{Var}(\pi_x)d\mu(x)\right\}$$

is related to the dual problem

$$Q_\lambda := \sup \left\{ -\int \varphi \nabla \frac{\lambda}{2} |\cdot|^2 d\mu - \int \varphi^* d\nu \,\middle|\, \varphi \text{ convex, continuous} \right\}$$

by

$$\frac{F_\lambda}{2\lambda} = Q_\lambda + \frac{\mu(|\cdot|^2) + (1 - 1/\lambda)\nu(|\cdot|^2)}{2},$$

where

$$\varphi \nabla \frac{\lambda}{2} |\cdot|^2(x) := \inf_{z \in X} \left\{ \varphi(z) + \frac{\lambda}{2} |x - z|^2 \right\}$$

is the Moreau[1]*-Yosida*[2] *transform and*

$$\varphi^*(y) := \sup_{x \in X} (xy - \varphi(x))$$

denotes the usual Fenchel[3] *transform. Furthermore, a pair* (π, φ) *is optimal if and only if for* μ*-a.e.* x

$$\operatorname{supp}(\pi_x) \subseteq \partial\varphi \left[x - \frac{1}{\lambda} \operatorname{bary}(\pi_x) \right].$$

We will only focus on the duality part of the statement. Instead of using the main theorem, we use the modified version, i.e., Theorem 3.5.

We adjust the setting to fit within our framework. Let $X = Y = \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. Fix $\lambda \in (0, 1]$ (we clarify why later). Then F_λ can be written as

$$\inf_{\pi \in \Pi(\mu,\nu)} \int \int \lambda |x - y|^2 d\pi_x(y) - \operatorname{Var}(\pi_x) d\mu(x).$$

So F_λ is a weak optimal transport problem. One can conveniently write the cost function in the form

$$C(x, \rho) = (\lambda - 1) \operatorname{Var}(\rho) + \lambda |x - \operatorname{bary}(\rho)|^2.$$

In probabilistic terms, we are looking for a coupling (X_0, X_1) with marginals μ, ν that strikes a balance between minimizing $\lambda \mathbb{E}\left[|X_0 - X_1|^2\right]$ and maximizing $\mathbb{E}\left[\operatorname{Var}(X_1 \mid X_0)\right]$. The second goal can intuitively be interpreted to mean that, on average, the variance of X_1 is not greatly diminished when knowledge of X_0 is introduced. This has the effect of favoring multi-valued transports over single-valued maps.

Let us also remark that for $\lambda \neq 1$, this example is not a special case of the barycentric costs from Section 4.2. Even in a more general sense, it is not of the form $\vartheta(x, \operatorname{bary}(\rho))$, since the variance is not a function of the barycentre alone. When we do have $\lambda = 1$, the primal problem F_λ corresponds to taking $\theta(\cdot) = |\cdot|^2$ in the context of Section 4.2.

[1] Jean-Jacques Moreau (1923-2014), French mathematician and mechanician
[2] Kōsaku Yosida (1909-1990), Japanese mathematician
[3] Werner Fenchel (1905-1988), Danish-German mathematician

We also mention that the case $\lambda = 0$ is dismissed as it is trivially optimized by the independent coupling.

We check the assumptions of the modified main theorem with $\kappa = 1$:

- *C jointly lower semicontinuous:* We want to show that for arbitrary $r \in \mathbb{R}$ the set
$$S := \left\{ (x,\rho) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \,\middle|\, (\lambda - 1)\operatorname{Var}(\rho) + \lambda |x - \operatorname{bary}(\rho)|^2 \leq r \right\}$$
is closed. Let $\{(x_n, \rho_n)\}_{n\in\mathbb{N}} \subseteq S$ be such that $(x_n, \rho_n) \xrightarrow{n\to\infty} (x, \rho)$. In particular, $\rho_n \to \rho$ in $\mathcal{W}_2$ and $x_n \to x$. We know that
$$(\lambda - 1)\operatorname{Var}(\rho_n) + \lambda |x_n - \operatorname{bary}(\rho_n)|^2 \leq r.$$
Then taking the limit yields
$$(\lambda - 1)\operatorname{Var}(\rho) + \lambda |x - \operatorname{bary}(\rho)|^2 \leq r,$$
where we recall that $\operatorname{Var}(\rho) = \int |\cdot|^2 d\rho - |\operatorname{bary}(\rho)|^2$ and we apply Lemma 3.1 and Theorem 1.17.

- *C convex in the second argument:* We expand the cost function to get
$$C(x,\rho) = \int (\lambda - 1)|y|^2 - 2\lambda xy + \lambda |x|^2 d\rho + |\operatorname{bary}(\rho)|^2,$$
which is clearly convex in the first term. To see the convexity of the second term, recall that the square function is convex.

- *$\widetilde{C}$ bounded from below:* This is where we see why we need the *modified* version of the main theorem.
$$\begin{aligned}\widetilde{C}(x,\rho) &= (\lambda - 1)\operatorname{Var}(\rho) + \lambda |x - \operatorname{bary}(\rho)|^2 + \int |y|^2 d\rho(y) \\ &= \lambda \operatorname{Var}(\rho) + \lambda |x - \operatorname{bary}(\rho)|^2 + |\operatorname{bary}(\rho)|^2 \geq 0.\end{aligned}$$

- *C satisfies condition (C1):* Having λ in $(0,1]$ is equivalent to condition (C1). Recall that condition (C1) means
$$\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d),\ \rho_1 \preceq_{\mathrm{c}} \rho_2 \implies C(x,\rho_1) \geq C(x,\rho_2).$$
Assume that $\lambda \in (0,1]$ and let $\rho_1 \preceq_{\mathrm{c}} \rho_2$. Then
$$\begin{aligned}C(x,\rho_1) &= (\lambda - 1)\operatorname{Var}(\rho_1) + \lambda |x - \operatorname{bary}(\rho_1)|^2 \\ &= (\lambda - 1)\left(\int |y|^2 d\rho_1(y) - |\operatorname{bary}(\rho_1)|^2 \right) + \lambda |x - \operatorname{bary}(\rho_1)|^2 \\ &= (\lambda - 1)\left(\int |y|^2 d\rho_1(y) - |\operatorname{bary}(\rho_2)|^2 \right) + \lambda |x - \operatorname{bary}(\rho_2)|^2 \\ &\geq (\lambda - 1)\left(\int |y|^2 d\rho_2(y) - |\operatorname{bary}(\rho_2)|^2 \right) + \lambda |x - \operatorname{bary}(\rho_2)|^2 = C(x,\rho_2),\end{aligned}$$

where we use the fact that ρ_1, ρ_2 are in convex order and $\lambda \leq 1$.

For the converse, assume that $\lambda > 1$. Fix $x_0 \in \mathbb{R}^d$ different from zero. Take $\rho_1 = \delta_0$ and $\rho_2 = \frac{1}{2}(\delta_{x_0} + \delta_{-x_0})$. Clearly, $\rho_1 \preceq_c \rho_2$. We evaluate the cost function and get

$$C(x, \rho_2) = (\lambda - 1)\,|x_0|^2 + \lambda |x|^2 > \lambda |x|^2 = C(x, \rho_1).$$

So indeed, (C1) fails in this case.

So we can use the modified main result to get

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x) d\mu(x) = \sup_{\substack{\psi \in \mathcal{C}_2^1 \\ \psi \text{ convex}}} \mu(R_C \psi) - \nu(\psi).$$

We consider $R_C\psi$ for $\psi \in \mathcal{C}_2^1$ and simplify it.

$$\begin{aligned}
R_C\psi(x) &= \inf_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \left[\rho(\psi) + (\lambda - 1)\operatorname{Var}(\rho) + \lambda |x - \operatorname{bary}(\rho)|^2\right] \\
&= \inf_{z \in \mathbb{R}^d} \left[\inf_{\rho \in \mathcal{P}_2^z(\mathbb{R}^d)} \rho(\psi) + (\lambda - 1)\operatorname{Var}(\rho) + \lambda |x - \operatorname{bary}(\rho)|^2\right] \\
&= \inf_{z \in \mathbb{R}^d} \left[\lambda |x - z|^2 - (\lambda - 1)\,|z|^2 + \inf_{\rho \in \mathcal{P}_2^z(\mathbb{R}^d)} \rho(\psi) + (\lambda - 1) \int |y|^2 d\rho(y)\right] \\
&= \inf_{z \in \mathbb{R}^d} \left[\lambda |x - z|^2 - (\lambda - 1)\,|z|^2 + \inf_{\rho \in \mathcal{P}_2^z(\mathbb{R}^d)} \rho\left(\psi + (\lambda - 1)\,|\cdot|^2\right)\right].
\end{aligned}$$

Then the function $\psi + (\lambda - 1)|\cdot|^2$ is in $\mathcal{C}_2^\lambda$, and by Lemma 3.6, we have

$$\begin{aligned}
&= \inf_{z \in \mathbb{R}^d} \left[\lambda |x - z|^2 - (\lambda - 1)\,|z|^2 + \operatorname{conv}\left(\psi + (\lambda - 1)\,|\cdot|^2\right)(z)\right] \\
&= 2 \inf_{z \in \mathbb{R}^d} \left[\operatorname{conv}\left(\frac{\psi + (\lambda - 1)\,|\cdot|^2}{2}\right)(z) - \frac{(\lambda - 1)}{2}|z|^2 + \frac{\lambda}{2}|x - z|^2\right] \\
&= 2 \left[\operatorname{conv}\left(\frac{\psi + (\lambda - 1)\,|\cdot|^2}{2}\right) - \frac{(\lambda - 1)}{2}|\cdot|^2\right] \triangledown \frac{\lambda}{2}|\cdot|^2(x),
\end{aligned}$$

where $\triangledown \frac{\lambda}{2}|\cdot|^2$ is the aforementioned Moreau-Yosida transform. So what we get is

$$\begin{aligned}
&\inf_{\pi \in \Pi(\mu,\nu)} \left\{\lambda \int |x - y|^2 d\pi(x, y) - \int \operatorname{Var}(\pi_x) d\mu(x)\right\} \\
= &\sup_{\substack{\psi \in \mathcal{C}_2^1 \\ \psi \text{ convex}}} \left\{2 \int \left[\operatorname{conv}\left(\frac{\psi + (\lambda - 1)\,|\cdot|^2}{2}\right) - \frac{(\lambda - 1)}{2}|\cdot|^2\right] \triangledown \frac{\lambda}{2}|\cdot|^2 d\mu - \int \psi d\nu\right\}.
\end{aligned}$$

This duality result is related to the one given in Theorem 4.2. Indeed, the dual formulation from the paper can be recovered by passing to functions of the form

$$\varphi := \left(\frac{\psi + (\lambda - 1)|\cdot|^2}{2\lambda} \right)^*,$$

but doing so does not use the fact that ψ is convex. Conversely, our duality result is recovered from the paper by taking

$$\psi := 2\lambda\varphi^* + (1 - \lambda)|\cdot|^2.$$

Further investigation into the relationship between these two results is warranted.

Notice that, unlike Theorem 4.2, we do not require the space to be bounded, but we do restrict the value of λ to $(0, 1]$.

4.4 The Classical Case

In this section, we discuss the implications of the main theorem in the classical context. Since optimal transport can be phrased as a special case of weak transport, we investigate how well the main theorem fares in the classical setting. We study the particular cost function $c(x, y) = |x|^2/2 - xy$, a close relative of the quadratic cost function, but we also remark on the general case throughout.

Let $X = Y = \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, and consider the cost function given by $C(x, \rho) = \int c(x, y) d\rho(y)$.

We check the assumptions of the main theorem:

- *C jointly lower semicontinuous:* We want to show that for arbitrary $r \in \mathbb{R}$ the set

  $$S := \left\{ (x, \rho) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \;\middle|\; |x|^2/2 - x\text{bary}(\rho) \le r \right\}$$

  is closed. Let $\{(x_n, \rho_n)\}_{n \in \mathbb{N}} \subseteq S$ be such that $(x_n, \rho_n) \xrightarrow{n \to \infty} (x, \rho)$. In particular, $\rho_n \to \rho$ in $\mathcal{W}_2$ and $x_n \to x$. We know that $|x_n|^2/2 - x_n\text{bary}(\rho_n) \le r$. Taking the limit yields

  $$r \ge \lim_{n \to \infty} |x_n|^2/2 - x_n\text{bary}(\rho_n) = |x|^2/2 - x\text{bary}(\rho),$$

  where the convergence of the barycentre is by Lemma 3.1.

- *C convex in the second argument:* This always holds for cost functions of the form $C(x, \rho) = \int c(x, y) d\rho(y)$.

- *$\widetilde{C}$ bounded from below:* $\widetilde{C}(x, \rho) = \int |x - y|^2/2 d\rho(y) \ge 0$. In general, $\widetilde{C}$ is bounded from below if and only if $c(x, y) + |y|^2/2$ is bounded from below (or $c(x, y) + \kappa|y|^2$, if one uses the modified main theorem).

- *C satisfies condition (C1):* Let $\rho_1 \preceq_c \rho_2$. We get

$$C(x, \rho_1) = |x|^2/2 - x\mathrm{bary}(\rho_1) = |x|^2/2 - x\mathrm{bary}(\rho_2) = C(x, \rho_2).$$

  For a general c, condition (C1) holds if and only if $c(x, \cdot)$ is concave for all x.

By applying the main theorem, we get the duality

$$\inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_x) d\mu(x) = \sup_{\substack{\psi \in \mathcal{C}_2 \\ \psi \text{ convex}}} \mu(R_C \psi) - \nu(\psi).$$

This simplifies to

$$\inf_{\pi \in \Pi(\mu,\nu)} \int |x|^2/2 - xy d\pi(x, y) = \sup_{\substack{\psi \in \mathcal{C}_2 \\ \psi \text{ convex}}} \mu(|\cdot|^2/2 - \psi^*) - \nu(\psi). \tag{4.5}$$

We cancel the finite second moment of μ and multiply by -1 for the slightly more aesthetically pleasing result

$$\sup_{\pi \in \Pi(\mu,\nu)} \int xy d\pi(x, y) = \inf_{\substack{\psi \in \mathcal{C}_2 \\ \psi \text{ convex}}} \mu(\psi^*) + \nu(\psi).$$

This is precisely the classical Kantorovich duality for $c(x, y) = -xy$.

Remark 4.3. By adding $\nu(|\cdot|^2/2)$ to both sides of (4.5), we get the Kantorovich duality for the celebrated case of the quadratic cost. In fact, the transport problems associated to the cost functions $-xy$ and $|x-y|^2/2$ are essentially the same up to additive constants.

5 Conclusion and Further Research

We successfully crafted Theorem 3.4 to generalize both [7, Proposition 3.3] and [15, Theorem 1.1]. We further extended it to Theorem 3.5, i.e., *the modified main theorem*, which allowed us to consider the primal problem in [1, Theorem 5.6] and produce another form of its dual problem. As a nice bonus, we used the main theorem to recover the classical Kantorovich duality for the quadratic cost function.

We briefly mention the connection to extrapolation in the Wasserstein space [13], which can be formulated as a barycentric transport problem. In the paper, which was kindly brought to my attention by Julio Backhoff, a duality result with the convexity restriction is shown, essentially in line with the proof of [15, Theorem 1.1]. However, the main theorem is not immediately applicable, since the weak transport problem in question is of the form $\int C(y, \pi_y) d\nu(y)$, i.e., disintegrated with respect to the second marginal. If this is circumvented, the paper can serve not only as another example but also as inspiration for further results.

One possible way to extend the main theorem, generously suggested by Mathias Beiglböck, would be to generalize the *convex order* relation. Let D be a convex cone in the space of continuous functions (it should also satisfy some further assumptions, see [2, Section 19.8] for more details), and define the relation $\mu \succeq_D \nu$ to mean that $\mu(f) \leq \nu(f)$ for every $f \in D$. For example, if D is the collection of concave functions, then $\mu \succeq_D \nu$ is equivalent to the familiar $\nu \preceq_c \mu$. So a generalized condition (C1) could look something like

$$\rho_2 \succeq_D \rho_1 \implies C(x, \rho_1) \geq C(x, \rho_2), \tag{D1}$$

and the dual problem could have a form akin to

$$\sup_{(-\mathcal{C}_p) \cap D} \mu\left(R_C(-\phi)\right) + \nu(\phi). \tag{5.1}$$

If this were the case, we could take D to be the set of decreasing concave functions and one would recover the result [8, Corollary 6.2], originally discussed in [11]. Indeed, condition (D1) is satisfied in the multiple-good monopolist problem and the dual proposed in (5.1) coincides with the one given in the paper.

Lastly, another potential direction, also suggested by Mathias Beiglböck, would be to link our results to those of Gangbo and McCann on strictly concave costs in [14]. However, it is not immediately clear how exactly one would approach this. In our main theorem, the central assumption (C1) translates to $c(x, \cdot)$ being concave in the classical setting, and in [14], $c(x, y)$ is of the form $\ell(|x - y|)$ where ℓ is strictly concave. A brief inspection did not yield an immediate connection, but more research is required.

Bibliography

[1] J.-J. Alibert, G. Bouchitté, and T. Champion. A new class of costs for optimal transport planning. *Eur. J. Appl. Math.*, 30(6):1229–1263, December 2019.

[2] C.D. Aliprantis and K.C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 2007.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[4] 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):203, Nov 2019.

[5] Julio Backhoff-Veraguas, Daniel Bartl, Mathias Beiglböck, and Manu Eder. Adapted Wasserstein distances and stability in mathematical finance. *Finance and Stochastics*, 24(3):601–632, Jul 2020.

[6] Julio Backhoff-Veraguas, Mathias Beiglböck, Martin Huesmann, and Sigrid Källblad. Martingale Benamou–Brenier: A probabilistic perspective. *The Annals of Probability*, 48(5):2258 – 2289, 2020.

[7] Julio Backhoff-Veraguas, Mathias Beiglböck, Walter Schachermayer, and Bertram Tschiderer. The structure of martingale Benamou−Brenier in $\mathbb{R}^d$, 2023.

[8] Julio Backhoff-Veraguas and Gudmund Pammer. Applications of weak transport theory. *Bernoulli*, 2020.

[9] Mathias Beiglböck. Stochastic Mass Transport: Lecture Notes, WS 2023/2024.

[10] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. C. R. Acad. Sci. Paris Sér. I Math., 305(19):805–808, 1987.

[11] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Strong duality for a multiple-good monopolist. *Econometrica*, 85(3):735–767, 2017.

[12] C. Dellacherie and P.A. Meyer. *Probabilities and Potential*. North-Holland mathematics studies. North-Holland Publishing Company, 1982.

[13] Thomas O. Gallouët, Andrea Natale, and Gabriele Todeschi. Metric extrapolation in the Wasserstein space, 2024.

[14] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177:113–161, 1996.

[15] Nathael Gozlan and Nicolas Juillet. On a mixture of Brenier and Strassen Theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, March 2020.

[16] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. Kantorovich duality for general transport costs and applications. *Journal of Functional Analysis*, 273(11):3327–3405, 2017.

[17] H. Lebesgue. Sur les fonctions représentables analytiquement. *Journal de Mathématiques Pures et Appliquées*, 1:139–216, 1905.

[18] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. Imprimerie royale, 1781.

[19] Andrea Natale, Gabriele Todeschi, and Thomas O. Gallouët. From geodesic extrapolation to a variational BDF2 scheme for Wasserstein gradient flows. *ArXiv*, abs/2209.14622, 2022.

[20] V.M. Panaretos and Y. Zemel. *An Invitation to Statistics in Wasserstein Space*. SpringerBriefs in Probability and Mathematical Statistics. Springer International Publishing, 2020.

[21] Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176, 2019.

[22] Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: efficient optimal transportation on geometric domains. *ACM Trans. Graph.*, 34(4), jul 2015.

[23] S.M. Srivastava. *A Course on Borel Sets*. Graduate Texts in Mathematics. Springer Berlin Heidelberg, 2013.

[24] В. И. Богачев. *Основы теории меры, т.2.* НИЦ Регулярная и Хаотическая динамика. М.-Ижевск, 2003.

[25] Л. В. Канторович. О перемещении масс. *Докл. АН СССР.*, 37, 7-8:227— 229, 1942.

[26] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.