On the structure of optimal martingale transport in higher dimensions

by

Tongseok Lim

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

The Faculty of Graduate and Postdoctoral Studies
(Mathematics)

The University of British Columbia
(Vancouver)

April 2016

© Tongseok Lim, 2016

Abstract

In the first part of this thesis, we study the structure of solutions to the optimal martingale transport problem, when the marginals lie in higher dimensional Euclidean spaces ($\mathbb{R}^d, d \geq 2$). The problem has been extensively studied in one-dimensional space ($\mathbb{R}$), but few results have been shown in higher dimensions. In this thesis, we propose two conjectures and provide key ideas that lead to solutions in important cases.

In the second part, we study the structure of solutions to the optimal subharmonic martingale transport problem, again when the marginals lie in higher dimensional Euclidean spaces. First, we show that this problem has an equivalent formulation in terms of the celebrated Skorokhod embedding problem in probability theory. We then describe the fine structure of the solution provided the marginals are radially symmetric. The general case remains unsolved, and its potential solution calls for a deeper understanding of harmonic analysis and Brownian motion in higher dimensional spaces.

Preface

Chapter 2 of this thesis is based on the paper [21] prepared in collaboration with Nassif Ghoussoub and Young-Heon Kim, as well as my paper [31].

Chapter 3 is based on paper [20] written in collaboration with Nassif Ghoussoub and Young-Heon Kim. Some text from these papers have been used verbatim in this thesis.

Papers submitted for publication:

[20] N. Ghoussoub, Y-H. Kim, and T. Lim. *Optimal Skorokhod embedding between radially symmetric marginals in general dimensions.*, preprint.
[21] N. Ghoussoub, Y-H. Kim, and T. Lim. *Structure of optimal martingale transport in general dimensions. http://arxiv.org/abs/1508.01806,* 2015. pp 1 - 27.
[31] T. Lim. *Optimal martingale transport between radially symmetric marginals in general dimensions.*, preprint.

Table of Contents

Acknowledgements

I thank my supervisors, Professors Nassif Ghoussoub and Young-Heon Kim, from the UBC Department of Mathematics for their valuable guidance and support throughout my doctoral study. They have oriented me into a beautiful and deep area of research, and they have patiently led me and moulded me into becoming a professional mathematician.

I thank Professor Ali Lazrak from the Sauder School of Business, for his heartfelt advice and encouragement.

I thank Professor Edwin A. Perkins, for his invaluable teaching of probability theory and for generously giving me so much of his time and his ideas.

I thank my mother, Jae-Jeong Lee, for her loving support throughout my studies in Korea and in Canada.

I thank my father, now in heaven, Eui-Taek Lim. He eagerly wanted me to pursue a higher-level education, and for that he supported me with everything he could give. I dedicate this Ph.D. thesis to his memory.

I would also like to thank my uncle Jaekyun Lee, and my aunt Soonok Lee, for their kind hospitality during my stay in Toronto. Chapter 3 could not have been written without their support.

Last but foremost, I thank my God, who las led my way and will lead me now and forever.

Dedication

To my father, who watches over me from heaven, and my mother.

Chapter 1

Introduction

1.1 Optimal transport problem

The research in my thesis has focused on the structure of probability measures which solve certain optimization problems. The prototype is the optimal mass transport problem: for a given cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and two Borel probability measures μ, ν on $\mathbb{R}^d$, we consider:

$$\text{(1.1.1)} \qquad \text{Maximize / Minimize } \operatorname{cost}[\pi] = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi(x, y)$$

over all $\pi \in T(\mu, \nu)$, where $T(\mu, \nu)$ is the set of *mass Transport plans*–i.e. the set of probabilities π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν on $\mathbb{R}^d$. We interpret the transport plan π as follows: for $A, B \subseteq \mathbb{R}^d$, $\pi(A \times B)$ is then the amount of mass transported by the plan π from the resource domain A to the target range B. An equivalent probabilistic formulation is to consider the following problem:

$$\text{(1.1.2)} \qquad \text{Maximize / Minimize} \quad \mathbb{E}_{\mathrm{P}}\, c(X, Y)$$

over all joint random variables $(X, Y) : \Omega \to \mathbb{R}^d \times \mathbb{R}^d$ with given laws $X \sim \mu$ and $Y \sim \nu$ respectively.

In 1781, Gaspard Monge [30] formulated the following question that was relevant to his work in engineering: Given two sets U, V in $\mathbb{R}^d$ of equal volume, find the optimal volume-preserving map between them, where optimality is measured against the cost function $c(x, y)$ of transporting particle x to y. The optimal map should then minimize the total cost of redistributing the mass of U through V. Much later, Kantorovich, a Nobel prize winner in 1975 for related work in economics, generalized the Monge problem and proposed the above formulation.

Many researchers capitalized on Kantorovich's pioneering work, including Brenier, Caffarelli, Evans, McCann, Gangbo, Trudinger, Wang, Ambrosio, Otto, and Villani, to contribute to the development of the fundamental theory of optimal transport. Their contributions have had far-reaching consequences in many areas of mathematics, including differential geometry, isoperimetry, PDE, infinite-dimensional linear programming, functional analysis, economics, probability, and statistics.

In Monge's original problem, the cost was simply the Euclidean distance $c(x, y) = |x - y|$. Even for this seemingly simple case, it took two centuries before Sudakov, Evans, Gangbo, Ambrosio, Caffarelli, McCann, Feldman, Trudinger, Wang and others showed rigourously that an optimal transport map exists. Many deep applications followed since.

More recently, a new direction emerged where the transport plans are assumed to be martingales. This is where our main research contribution lies. In the sequel, I shall describe the problem, its motivation, our contributions, and further directions of research.

1.2 Martingale transport problem

Consider the optimization problem (1.1.1) over $\mathrm{MT}(\mu, \nu)$, where $\mathrm{MT}(\mu, \nu)$ is the set of *Martingale Transport plans*, that is the set of probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, such that for each $x \in \mathbb{R}^d$, its disintegration $(\pi_x)_{x \in \mathbb{R}^d}$ with respect to μ has its barycenter at x; in other words, for any convex function φ on

$\mathbb{R}^d$,

$$\varphi(x) \le \int_{\mathbb{R}^d} \varphi(y)\, d\pi_x(y). \tag{1.2.1}$$

Equivalently, we study (1.1.2) over all martingales (X, Y) (i.e. $E[Y|X] = X$) with given laws $X \sim \mu$, $Y \sim \nu$. We recall that a disintegration is an indexed probability measures $(\pi_x)_{x\in\mathbb{R}^d}$, which can be interpreted as the conditional probability $d\pi_x(y) = \mathbb{P}(Y = y|X = x)$ in such a way that $d\pi(x, y) = d\pi_x(y)d\mu(x)$.

The problem originates in mathematical finance. Indeed, in the process of pricing financial options, the usual approach is to postulate a model and then calculate the price of an option according to the model. However, this assumption is not practical since all what we can observe in the market are the traded option prices and not any specific model. We are therefore led to ask what we could infer about the underlying model, if we can only know vanilla option prices. Mathematically speaking, we are interested in an underlying model (X, Y), where X is the (random) price of the options at the first date of maturity and Y is the price at the second date of maturity. If we know the call and put prices of many exercised prices for these dates of maturity, then we can infer the marginal distribution of the asset price (i.e. $X \sim \mu$ and $Y \sim \nu$). Since there can be many candidate martingales (X, Y) having the same marginals, we can then look at the ones where the two values Min $\mathbb{E}_{\mathrm{P}}\, c(X, Y)$ and Max $\mathbb{E}_{\mathrm{P}}\, c(X, Y)$ is achieved. This will give the lower (resp., upper) bound of the price of the option $c(X, Y)$.

It was D. Hobson who first recognized the importance of Skorokhod embeddings[1] in this "model-free" approach to finance and asset pricing. Since then, much related research has been done in this context, for example, by Beiglböck, Cox, Dolinsky, Galichon, Henry-Labordere, Hobson, Huesmann, Soner, Touzi and others [4], [6], [12], [17], [25], [26]. On the other hand, another long line of research on the general martingale transport problem has been made by Beiglböck, Henry-Labordere, Hobson, Juillet, Klimmek, Neuberger, Penkner, Touzi and others [3], [4], [5], [27], [28].

[1] We will introduce the Skorokhod embedding problem later

Note that in one dimension, the two problems are equivalent, i.e., one can always study martingales induced by stopped Brownian motion. This is not surprising since in $\mathbb{R}^1$, the class of convex functions is the same as the class of subharmonic functions (see section 3.1 for more detail).

But the class of convex functions becomes a strict subset of the class of subharmonic functions in dimension $d \geq 2$, and in this higher-dimensional case, the Skorokhod embedding problem is equivalent to the Subharmonic martingale transport problem, which we define and discuss in chapter 3. We note that the above cited papers are all done in one dimension. Mathematically, this means that the marginals μ, ν are probability measures on the real line, or the random variables X, Y have values in $\mathbb{R}^1$. Financially, this means that the option $c(X, Y)$ depends only on one stock process. It may not be hard to expect that the optimal martingale coupling problem (1.1.1) will be much more difficult when the option depends on arbitrarily many stock processes, simultaneously and nonlinearily. Nevertheless, the higher dimensional problem looks very important, not only mathematically but also financially since many options in the real market indeed depend on many number of assets, e.g. every stock in the S&P 500 index.

It is therefore important to consider the higher dimensional case. In my first paper [31] in this direction, I have shown that the optimal martingale problem has a unique solution in case the marginals μ, ν are radially symmetric, which is satisfied by important distributions such as Gaussians. Very recently, my Prof. Nassif Ghoussoub, Prof. Young-Heon Kim and I have completed a paper [21] that deals with the general case in arbitrarily high dimensions, and which gives a satisfactory description of the optimal solutions. To our knowledge, these are the first such results to be established in arbitrarily high dimensions. We will explain these in detail in chapter 2.

1.2.1 Summary of our result

In Beiglböck-Juillet [5], Hobson-Neuberger [28], Hobson-Klimmek [27], it is shown that in the problem (1.1.1) with $d = 1$ and $c(x, y) = |x - y|$, whenever μ is dispersed

($\mu << \mathcal{L}^1$), an optimal martingale transport π is unique for any given ν, and more interestingly, exhibits an extremal property in that for each $x \in \mathbb{R}$, the disintegration (conditional probability) π_x is concentrated at two boundary points of an interval.

Now the question becomes: what is the higher-dimensional extension of the above result? We conjecture the following:

Problem 1: *Let $c(x,y) = |x-y|$ and assume $\mu << \mathcal{L}^d$. If $\pi \in MT(\mu,\nu)$ solves* (1.1.1)*, then there is a set Γ with $\pi(\Gamma) = 1$ such that for μ-almost every x, the fiber $\Gamma_x := \{y \,|\, (x,y) \in \Gamma\}$ is contained in the Choquet boundary (i.e., the set of extreme points) of the closed convex hull of Γ_x, i.e.,*

$$\Gamma_x \subseteq \mathrm{Ext}\left(\overline{\mathrm{conv}}(\Gamma_x)\right).$$

First, I verifiy that the problem holds true under radial symmetry condition [31]:

Theorem 1.2.1. *Assume that μ and ν are radially symmetric on $\mathbb{R}^d$ and μ is absolutely continuous with respect to Lebesgue measure ($\mu << \mathcal{L}^d$). Let $c(x,y) = |x-y|^p$ for some $0 < p \leq 1$ and let π be a minimizer for the problem* (1.1.1)*. Then for μ a.e. x, the disintegration π_x is concentrated on two points which lie on the one-dimensional subspace spanned by x. Furthermore, π is unique.*

In [31], ideas on the calculus of variations which nicely exploit the radial structure is presented.

However, if the marginals μ, ν do not exhibit any symmetry on $\mathbb{R}^d$, such a two-point splitting type result cannot be expected. For example, if μ is concentrated around the origin in $\mathbb{R}^2$ and ν is concentrated around the vertices of a triangle, then obviously every π_x must be split to 3 points. The variational calculus in [31] is not applicable here as it used the radial symmetry of the marginals.

Instead, we try to seek the dual formulation. In [3], it is shown that if the cost c is lower semi-continuous, then for the minimization problem, the "no-duality gap" result holds:

(1.2.2)

$$\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,d\pi;\ \pi\in MT(\mu,\nu)\right\} = \sup\left\{\int_{\mathbb{R}^d}\alpha d\mu + \int_{\mathbb{R}^d}\beta d\nu\ ;\ (\alpha,\beta)\in D_m(\mu,\nu,c)\right\},$$

where $D_m(\mu,\nu,c)$ is the set of all $(\alpha,\beta)\in L^1(\mu)\times L^1(\nu)$ such that for some bounded $\gamma:\mathbb{R}^d\to\mathbb{R}^d$,

$$\alpha(x)+\beta(y)+\gamma(x)\cdot(y-x)\le c(x,y) \text{ for all } (x,y)\in\mathbb{R}^d\times\mathbb{R}^d. \tag{1.2.3}$$

We note that without the linear term $\gamma(x)\cdot(y-x)$, (2.1.4) is the dual formulation in the classical optimal transport, and this new term makes the problem completely different than the classical case.

By Prokhorov's compactness theorem, it can be shown that the "primal" problem (which is the minimization problem in (1.2.2)) attains a solution. Now our first major contribution in chapter 2 shows:

Theorem 1.2.2. *If the "dual" problem (that is, the supremum problem in* (1.2.2)*) attains a solution* (α,β,γ)*, then the problem 1 holds true.*

Note that even when an optimal dual triplet (α,β,γ) is attained, it does not necessarily have any regularity, making the analysis of the dual problem difficult. This is reminiscent of PDE theory where solutions may first be found in the generalized spaces. In chapter 2, the following regularity theorem is developed, and one of its consequences is that the variational proof of Theorem 1.2.2 can be activated:

Theorem 1.2.3. *Suppose* $x\mapsto c(x,y)$ *is locally Lipschitz (e.g.* $c(x,y)=|x-y|$*) and assume that the dual problem attains a solution* (α,β,γ)*. Then, one can improve their regularity and get the extended dual triple* $(\overline{\alpha},\overline{\beta},\overline{\gamma})$*, such that* $\overline{\alpha}$ *is locally Lipschitz (thus, differentiable a.e), and a certain natural orthogonal projection of* $\overline{\gamma}$ *is also differentiable a.e.*

Now the issue is that it is not clear whether the dual optimization problem attains a solution or not. In the classical transport problem, it is essentially always attained, thus we could study the dual side of the problem. Unfortunately, it is not the case in the martingale transport problem; an example is given in chapter 2. The newly introduced linear them $\gamma(x)(y-x)$ causes trouble in dual attainment problem, and it makes the martingale transport problem hard to deal with.

Since duality is not attained in general, we investigate the possibility of decomposing an optimal martingale transport into "irreducible components", on which duality of the "restricted" problem is attained. For this, we establish a canonical decomposition of a martinagale transport set, and we prove that each component indeed admits a dual. This is another main accomplishment of this thesis:

Theorem 1.2.4 (Decomposition of martingale plans). *Let $\pi \in MT(\mu,\nu)$ and let Γ be a martingale-supporting set with $\pi(\Gamma) = 1$. Then, one can define a natural decomposition map $\Xi : \Gamma \to \mathcal{K}(\mathbb{R}^d)$ where $\mathcal{K}(\mathbb{R}^d)$ is the space of convex closed subsets of $\mathbb{R}^d$, such that if $\tilde{\pi} = \Xi_{\#}\pi$ denotes the push-forward of π into $\mathcal{K}(\mathbb{R}^d)$, and $I \subseteq \mathcal{K}(\mathbb{R}^d)$ is the image of Γ by Ξ, then the following holds:*

(1) There exists a disintegration of π along the map Ξ such that

$$\pi(S) = \int_I \pi_C(S)d\tilde{\pi}(C) \quad \text{for each Borel set } S \subseteq \mathbb{R}^d \times \mathbb{R}^d, \tag{1.2.4}$$

where for $\tilde{\pi}$-a.e. $C \in I$, π_C is a probability measure supported on $\Gamma_C := \Gamma \cap (C \times \overline{C})$.

(2) If $\pi \in MT(\mu,\nu)$ is optimal for problem (1.1.1), then for every $C \in I$, π_C is optimal in $MT(\mu_C,\nu_C)$, where μ_C, ν_C are marginals of π_C. Furthermore, dual optimal solution is attained for π_C.

(3) If μ_C is absolutely continuous with respect to Lebesgue measure on C, and $c(x,y) = |x-y|$, then for μ_C-almost all x, $\Gamma_x \subseteq \mathrm{Ext}(\overline{\mathrm{conv}}(\Gamma_x))$.

As applications of the above theory, the following theorems are presented; note that the dual attainment assumption in Theorem 1.2.2 is no longer assumed.

Theorem 1.2.5. *Let π be a solution of* (1.1.1) *with* $c(x, y) = |x - y|$. *Suppose* $\mu << \mathcal{L}^d$. *Then for μ a.e. x, the conditional probability π_x is supported on a set of Hausdorff dimension at most d-1.*

Theorem 1.2.6. *If every "irreducible components" in the decomposition theorem has dimension $\geq$ d-1 (in particular, when $d = 2$ this holds automatically), then for μ a.e. x, $\Gamma_x \subseteq \mathrm{Ext}\left(\overline{\mathrm{conv}}(\Gamma_x)\right)$.*

Theorem 1.2.7. *Let $c(x, y) = |x - y|$ and suppose ν is discrete; i.e. ν is supported on a countable set. Then for μ a.e. x, the conditional probability π_x is supported on d+1 points which are vertices of a polytope in $\mathbb{R}^d$, and therefore the optimal solution is unique.*

In the attempt to remove the dimensional assumption in Theorem 1.2.6, we encounter a bizarre object, called Nikodym set [2]. In chapter 2, an optimal martingale on the Nikodym set is constructed, such that if we apply the decomposition theorem 2.2.14, then the disintegrated source measure μ_C is singular even if $\mu << \mathcal{L}^d$, due to the bizarre structure of the Nikodym set. As μ_C is not absolutely continuous, our regularity theory does not apply, and thus we cannot deduce the extreme property from our theory. This suggests that, the martingale transport problem in higher dimensions indeed is connected with one of the deepest problems in analysis, and makes problem 1 very intriguing.

Meanwhile, Theorem 1.2.7 and 1.2.1 suggest that, in certain cases, a finer structure of optimal martingale might be obtained. Inspired by these results, we propose the following problem:

Problem 2 (Minimization:) *Assume that π is a minimizer for the problem* (1.1.1) *with distance cost. Then for μ a.e. x, the conditional probability π_x is supported on $k+1$ points, where $k = k(x) \leq d$, that form the vertices of a k-dimensional polytope, and therefore, the minimizing solution is unique.*

In this regard, we show a simple counterexample that in the maximization problem, neither polytope-type fine structure nor uniqueness of optimal solution can be expected even under radially symmetric marginals μ and ν. Therefore, one of my next research is to understand why the maximization and minimization problem can be fundamentally different, which is also a very interesting analytic / geometric aspect of the optimal martingale transport problem in higher dimensions.

Next, we introduce another problem, called the Skorokhod embedding and its connection with the martingale transport problem.

1.3 Subharmonic martinagale transport problem a.k.a Skorokhod embedding problem

Let B be a Brownian motion with initial law μ. Consider the set of stopping times

$$T = \{\tau : \tau \text{ is a stopping time and } B_\tau \sim \nu\} \tag{1.3.1}$$

We say that $\tau \in T$ solves the Skorokhod embedding problem. Now, we consider

$$\text{Maximize / Minimize} \quad \mathbb{E}_{\mathrm{P}}\, c(B_0, B_\tau) \quad \text{over} \quad \tau \in T \tag{1.3.2}$$

The aim of the study is to characterize such optimal stopping times.

It turns out that this problem is equivalent to the transport problem (1.1.1) over $\mathrm{SMT}(\mu,\nu)$, where $\mathrm{SMT}(\mu,\nu)$ (Subharmonic Martingale Transport plan) is the set of joint probabilities such that its disintegration π_x is in $\mathrm{SH}(x)$ for each x. Now the equivalence stems from the following proposition:

Proposition 1.3.1. *Let $SH(x)$ be the set of probabilities with subharmonic barycenter at x; i.e. every $\psi \in SH(x)$ satisfies the constraint* (1.2.1) *for every subharmonic function φ. Then for any $\psi \in SH(x)$, there exists a stopping time τ such that $\psi = Law(B^x_\tau)$, where B^x is a Brownian motion starting at x.*

We note that, in order for the above proposition to hold, we need to consider the class of randomized stopping times (see [6]). By proposition, we see that the optimal Skorokhod embedding problem and the subharmonic martingale transport problem are equivalent, and it also shows the equivalence of martingale transport problem and Skorokhod embedding problem in one dimension.

While in Skorokhod formulation we focus on the random movement of the Brownian motion and its stopping rule, in mass transport formulation we focus on the distributions $(\pi_x)_x$ satisfying the defining inequality (1.2.1). It turns out that the interplay between the two viewpoints are very enlightening, as can be seen in chapter 3.

Since initiated by Skorokhod in 1961, the Skorokhod embedding problem has become one of the most important subjects in probability, and there is a large literatures on the subject. We see that most works concern the one-dimensional problem, and one of the potential reasons is that the class of subharmonic functions is much wider than that of convex functions in higher dimensions. We start tackling the problem with the cost $\mathbb{E}_{\mathrm{P}}|B_0 - B_\tau|$ and radially symmetric marginals. However, it turns out that even in the radial marginal case, the subtle constraint (1.2.1) with respect to the cone of subharmonic functions makes the problem hard to handle. In chapter 3, we show that by applying the properties of Brownian motion and stopping time, a certain variational calculus is admissible. As a consequence, we show:

Theorem 1.3.2. *Let $c(x, y) = |x - y|$ and let μ and ν be radially symmetric probability measures on $\mathbb{R}^d$ which are in subharmonic order. Then for μ a.e. x, every optimal stopping time for the Skorokhod embedding problem is the first-hitting time of a barrier set; in particular, every optimal stopping time is of natural (i.e. non-randomized) type and furthermore, the optimal time is unique.*

The naturality (non-randomization property) of the optimal stopping time in the theorem can be seen as the Brenier-type result in Skorokhod embedding, and we see that it possesses the extremal property every optimal solution exhibits.

Chapter 2

Structure of optimal martingale transport in higher dimensions

2.1 Introduction

We study the profile of one-step martingales π on $\mathbb{R}^d \times \mathbb{R}^d$ that optimize the expected value of the modulus of their increment, among all martingales with two given marginals μ and ν in convex order. More precisely, we shall investigate the submanifolds Γ_x of $\mathbb{R}^d$ along which a particle x is propagated under such transport plans. These questions originate in mathematical finance and are variations on the original Monge problem, where one considers all couplings of the given marginals and not only those of martingale type [32], [19], [37, 38]. However, unlike solutions of the Monge problem, which are often supported on graphs (such as the well-known Brenier solution [8] for the cost given by the squared distance), the additional martingale constraint forces the transport to split the elements of the initial measure μ. One cannot therefore expect –but in trivial cases– that optimal martingale plans be supported on graphs.

These martingale transport problems are motivated by problems in mathematical finance, which call for no-arbitrage lower (or upper) bounds on the price of a forward starting straddle, given today's vanilla call prices at the two relevant maturities.

Just like in the Monge-Kantorovich theory for optimal transport, these problems have dual counterparts, whose financial interpretation amount to constructing the most (or least) expensive semi-static hedging strategy which sub-replicates the payoff of the forward starting straddle for any realization of the underlying forward price process.

The minimization and maximization problems are quite different, though by now well understood when the marginals are probability measures on the real line, at least in the case of one-step martingales. We refer to Hobson-Neuberger [28], Hobson-Klimmek [27], and Beiglböck-Juillet [5]. For the multi-step case, see Beiglböck-Henry-Labordere-Penkner [3]. The dynamic case have been also studied by Galichon-Henri-Labordere-Touzi [17] and Dolinsky–Soner [12, 13]. The two cases studied are when the cost is either $c(x, y) = |x - y|$, which is the main focus of this thesis, or the case when the cost satisfies the so-called Mirlees condition. Note that the one-dimensional case is closely related to Skorohod embedding problem, since real valued martingales can be realized as adequately stopped Brownian paths. See for example Hobson [26], Beiglböck-Cox-Huesmann [6] and Beiglböck-Henry-Labordere-Touzi [4].

Surprisingly, much less is known in the case where the marginals are supported on higher dimensional Euclidean spaces $\mathbb{R}^d$. In this chapter, we shall tackle the following general optimization problem associated to a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$:

(2.1.1)
$$\text{Maximize / Minimize } \operatorname{cost}[\pi] = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi(x, y) \quad \text{over} \quad \pi \in MT(\mu, \nu),$$

where $MT(\mu, \nu)$ is the set of *Martingale Transport plans*, that is the set of probabilities π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, such that for μ-almost $x \in \mathbb{R}^d$, the components of its disintegration $(\pi_x)_x$ with respect to μ, i.e. $d\pi(x, y) = d\pi_x(y)d\mu(x)$, have their barycenter at x; in other words, for any convex function φ on $\mathbb{R}^d$,

$$\varphi(x) \leq \int_{\mathbb{R}^d} \varphi(y)\, d\pi_x(y).$$

One can also use the probabilistic notation in the following way:

$$\text{Maximize / Minimize} \quad \mathbb{E}_{\mathrm{P}}\, c(X,Y) \tag{2.1.2}$$

over all martingales (X, Y) on a probability space $(\Omega, \mathcal{F}, P)$ into $\mathbb{R}^d \times \mathbb{R}^d$ (i.e. $E[Y|X] = X$) with laws $X \sim \mu$ and $Y \sim \nu$ (i.e., $P(X \in A) = \mu(A)$ and $P(Y \in A) = \nu(A)$ for all Borel set A in $\mathbb{R}^d$). Note that in this case, the disintegration of π can be written as the conditional probability $d\pi_x(y) = \mathbb{P}(Y = y|X = x)$.

A classical theorem of Strassen [36] states that the set $MT(\mu, \nu)$ of martingale transports is nonempty if and only if the marginals μ and ν are in *convex order*, that is if

1. μ and ν have finite mass and finite first moments, and
2. $\int_{\mathbb{R}^d} \varphi \, d\mu \leq \int_{\mathbb{R}^d} \varphi \, d\nu$ for every convex function φ on $\mathbb{R}^d$.

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

We will mostly consider the Euclidean distance cost $c(x, y) = |x - y|$ unless stated otherwise, although some of the results below hold for more general costs. We shall use the term optimization in problem (2.1.1) whenever the result holds for either maximization or minimization. We shall be more specific otherwise, since it will soon become very clear that the two cases can sometimes be fundamentally different. The following theorem summarizes the main structural result when μ and ν are one-dimensional marginals. Hobson-Neuberger [28] were first to deal with the maximization case while Beiglböck-Juillet [5] and D. Hobson and M. Klimmek [27] deal with the context of minimization. For a subset Γ in $\mathbb{R}^d \times \mathbb{R}^d$, we shall denote by Γ_x, the fiber $\Gamma_x := \{y \in \mathbb{R}^d; \, (x, y) \in \Gamma\}$.

Theorem 2.1.1. (Beiglböck-Juillet [5], Hobson-Neuberger [28], Hobson-Klimmek [27]) *Assume that μ and ν are probability measures in convex order on $\mathbb{R}$, and that μ is continuous. There exists then a unique optimal martingale transport plan π for* *the cost function $c(x;y) = |x - y|$. Moreover there is a set $\Gamma \subseteq \mathbb{R} \times \mathbb{R}$ with $\pi(\Gamma) = 1$ such that:*

1. *in the minimization problem, $|\Gamma_x| \leq 3$ for every $x \in \mathbb{R}$. More precisely, π can be decomposed into $\pi_{stay} + \pi_{go}$, where $\pi_{stay} = (Id \circ \times Id)_{\#}(\mu \wedge \nu)$ (this measure is concentrated on the diagonal of $\mathbb{R}^2$) and π_{go} is concentrated on $\mathrm{graph}(T_1) \cup \mathrm{graph}(T_2)$ where T_1, T_2 are two real-valued functions.*

2. *in the maximization problem, $|\Gamma_x| \leq 2$ for every $x \in \mathbb{R}$, and π is concentrated on $\mathrm{graph}(T_1) \cup \mathrm{graph}(T_2)$ where T_1, T_2 are two real-valued functions.*

Our main goal in this thesis is to consider higher dimensional analogues of the above result. In section 2.6, we show that the above theorem extends, in the case of minimization, to the setting where the marginals are radially symmetric on $\mathbb{R}^d$ and $c(x, y) = |x - y|^p$ for $0 < p \leq 1$. The general case is wide open and our goal is to work towards establishing the following:

Conjecture 1: *Consider the cost function $c(x, y) = |x - y|$ and assume that μ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$ ($\mu << \mathcal{L}^d$). If π is a martingale transport that optimizes* (2.1.1). *xThen, there is a set Γ with $\pi(\Gamma) = 1$ such that for μ-almost every x, $\Gamma_x := \{y \,|\, (x, y) \in \Gamma\}$ is contained in the Choquet boundary (i.e., the set of extreme points) of the closed convex hull of Γ_x, i.e.,*

$$\Gamma_x \subseteq \mathrm{Ext}\left(\overline{\mathrm{conv}}(\Gamma_x)\right).$$

Note that for the minimization problem, we can and will assume that $\mu \wedge \nu = 0$ since any minimizing martingale transport for problem (2.1.1) must let the support of $\mu \wedge \nu$ stay put. See [5] or [31] for a proof. One can then easily see that in the one dimensional case, the above conjecture reduces to Theorem 2.1.1 since then the dimension of the linear span of Γ_x is one and the Choquet boundary consists of exactly two points, unless of course Γ_x is a singleton.

We shall be able to prove the above conjecture in many important cases, in particular, when a natural dual optimization problem is attained (Theorem 2.2.4), or when the linear span of Γ_x has full dimension (Theorem 2.2.15). Another case where the answer is affirmative is in dimension $d = 2$ (Theorem 2.2.16) provided the two marginals are well separated, which as noted above, is not very restrictive in the case of minimization.

We actually expect to have a more rigid structure in the case of minimization. Indeed, we show that in this case, assuming $\mu \wedge \nu = 0$, we also have $\#\Gamma_x \leq 2$ for μ-almost all x, whenever the marginals are radially symmetric on $\mathbb{R}^d$ and $c(x,y) = |x-y|^p$ for $0 < p \leq 1$. The general case remains open as we propose the following:

Conjecture 2 (Minimization): *Consider the cost function $c(x,y) = |x-y|$ and assume that μ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, that $\mu \wedge \nu = 0$. If π is a martingale transport that minimizes* (2.1.1). *Then, there is a set Γ with $\pi(\Gamma) = 1$ such that for μ almost every x, the set Γ_x consists of $k+1$ points that form the vertices of a k-dimensional polytope, where $k := k(x)$ is the dimension of the linear span of Γ_x and therefore, the minimizing solution is unique.*

We shall give a partial answer to the above conjecture under the assumption that the target measure ν is discrete. Actually, in this case the result holds true in both the maximization and minimization cases (Theorem 2.2.16). We note however that – unlike the minimization case– one cannot always expect in higher dimensions neither the uniqueness of a maximizer (Example 2.2.19), nor a polytope-type structure for its support of Γ_x (Example 2.2.18), even when the marginals are radially symmetric.

Just like in the Monge-Kantorovich theory, the above optimization problem (2.1.1) has a dual formulation, which will be crucial to our analysis. Indeed, one can show (see for example [3]) that if the cost c is lower semi-continuous, then for the minimization problem,

(2.1.3)

$$\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,d\pi;\, \pi\in MT(\mu,\nu)\right\} = \sup\left\{\int_{\mathbb{R}^d}\beta d\nu - \int_{\mathbb{R}^d}\alpha d\mu;\ (\alpha,\beta)\in D_m(\mu,\nu,c)\right\},$$

where $D_m(\mu,\nu,c)$ is the set of all $(\alpha,\beta) \in L^1(\mu)\times L^1(\nu)$ such that for some $\gamma \in$

$C_b(\mathbb{R}^d, \mathbb{R}^d)$,

$$\beta(y) - \alpha(x) - \gamma(x)(y-x) \leq c(x,y) \text{ for all } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{2.1.4}$$

Similarly, if the cost c is upper semi-continuous, then
(2.1.5)
$$\max\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\, d\pi;\ \pi \in MT(\mu,\nu)\right\} = \inf\left\{\int_{\mathbb{R}^d} \beta d\nu - \int_{\mathbb{R}^d} \alpha d\mu;\ (\alpha,\beta) \in D_M(\mu,\nu,c)\right\},$$

where $D_M(\mu,\nu,c)$ is the set of all $(\alpha,\beta) \in L^1(\mu) \times L^1(\nu)$ such that for some $\gamma \in C_b(\mathbb{R}^d; \mathbb{R}^d)$,

$$\beta(y) - \alpha(x) - \gamma(x)(y-x) \geq c(x,y) \text{ for all } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{2.1.6}$$

Note that if π is an optimal martingale measure and if the corresponding dual problem is attained on a triplet (α,β,γ), then it is easy to see that there exists a Borel subset $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ with full π-measure such that

$$\beta(y) - \alpha(x) - \gamma(x)(y-x) = c(x,y) \text{ if and only if } (x,y) \in \Gamma. \tag{2.1.7}$$

We shall show that characterizing a concentration set Γ of π with such an identity forces it to have a specific extremal structure. However, unlike the Monge-Kantorovich theory for mass transport, solutions to the dual problem need not exist –at least in the maximization problem– even in the one-dimensional case, as shown in [3]. Example 2.3.2 below provides another simple counterexample.

We shall actually show that attainment in the dual problem is a very strong assumption. Namely, in the case $x \mapsto c(x,y)$ is locally Lipschitz (uniformly in y), once there is a triplet of functions (α,β,γ), where the dual is attained, one can then improve their regularity, that is, one can associate extensions $(\overline{\alpha},\overline{\beta},\overline{\gamma})$ such that $\overline{\alpha}$ and an appropriate orthogonal projection of $\overline{\gamma}$ are differentiable almost everywhere. This additional regularity allows us to show that Conjecture 1 then holds in case of dual attainment (Theorem 2.2.4) for the cost $c(x,y) = |x-y|$.

We then proceed to build on this result to deal with more general cases. One of the notable advances here is a decomposition of any Borel set Γ supporting a given optimal martingale transport π, into components $\{\Gamma_C\}_{C \in I}$, where a dual triplet $(\alpha_C, \beta_C, \gamma_C)$ is found for each piece in such a way that (2.1.4) (resp., (2.1.6)) holds for the minimization (resp., maximization) setting, with equality on each piece Γ_C. See Theorem 2.2.11 for the precise statement.

What is remarkable is that this decomposition, as well as the duality statements, can be done in full generality (i.e., for any cost function) and without any reference to a martingale transport problem or even to any reference measure. The decomposition is done through an equivalence relation on the projection X_Γ of Γ on the first coordinate, that is induced by a well chosen *irreducible convex paving*, that is a collection of mutually disjoint convex subsets in $\mathbb{R}^d$ that covers X_Γ. On the other hand, a good analogy for the (pointwise and not measure-theoretic) duality conditions, is the role played by c-cyclically monotone sets and their corresponding c-convex potentials in the Monge-Kantorovich theory associated to a given cost function c.

Back to the martingale problem, we then consider the disintegration $\{\pi_C\}_{C \in I}$ of any martingale measure π along the above described decomposition of its support Γ (Theorem 2.2.14). This then gives a canonical decomposition of the optimal martingale problem into a collection of non-interactive martingale problems where duality is attained for each piece π_C in $MT(\mu_C, \nu_C)$. We note that this result can be seen as a generalization of the decomposition of Beiglböck-Juillet [5] in the one-dimensional case $d = 1$, where the disintegration comes from restricting the measures μ, ν onto open subintervals of $\mathbb{R}$ obtained by examining the potential functions for μ, ν. Like theirs, our decomposition applies to any cost function c and not only to $c(x, y) = |x - y|$. It is however quite different since it depends on the support of the martingale measure π that we start with. More importantly, our decomposition need not be countable (Example 2.5.2) which creates additional and interesting complications for the higher dimensional cases.

We shall use the above decomposition/disintegration results to establish the above stated conjectures under various conditions. For example, Conjecture 1 holds in the case where the dimensions of all components $(C)_{C \in I}$ are larger than $d - 1$, or in dimension $d = 2$ provided the supports of the marginals μ and ν are well separated from each other (see Theorem 2.2.16).

Remarkably, the results discussed so far do not distinguish between the minimization and maximization problems (except that we assume that $\mu \wedge \nu = 0$ in the case of minimization). The previously mentioned decomposition can be used to prove Conjecture 2) in either the minimization and maximization case, provided the target measure ν has a countable support (Theorem 2.2.17). However, as mentioned above, we believe that these two problems are quite different, at least in terms of finding finer structural results for each of the cases.

In the next section, we give the precise statements of our results. In Section 2.3, we show the regularity and structural results in the case where the dual problem is attained. In Section 2.4, we establish the decomposition/disintegration results, as well as the existence of dual triplets on each of the components in the decomposition. Section 2.5 is devoted to proving under various additional conditions, structural results for sets where optimal martingale transports concentrate.

2.2 Main results

To discuss our main results, we first introduce a few definitions. We also borrow some of the notation from [5].

Definition 2.2.1. *For $A \subseteq \mathbb{R}^d$, we shall write $V(A)$ for the lowest-dimensional affine space containing A. Also define $IC(A) := int(conv(A))$ and $CC(A) := cl(conv(A))$, where again the interior or closure is taken in the topology of $V(A)$, where the topology of a set A is with respect to the Euclidean metric topology of $V(A)$ (and not with respect to the whole space $\mathbb{R}^d$).*

For a Borel set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$, we write $X_\Gamma := \mathrm{proj}_X \Gamma$, $Y_\Gamma := \mathrm{proj}_Y \Gamma$, i.e. X_Γ is the projection of Γ on the first coordinate space $\mathbb{R}^d$, and Y_Γ on the second. For each $x \in X_\Gamma$, we consider the fiber

$$\Gamma_x = \{y \in \mathbb{R}^d \mid (x, y) \in \Gamma\}.$$

If $\Gamma_x = \{x\}$, then $IC(\Gamma_x) = \{x\}$ since the interior of a singleton set is itself in the topology of 0-dimensional space.

Note that if π is a martingale transport plan in $MT(\mu, \nu)$, then there exists a Borel set $\Gamma \in \mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(\Gamma) = 1$ in such a way that μ-almost all $x \in X_\Gamma$ belongs to $IC(\Gamma_x)$. The set Γ may not necessarily be the support of π, which is normally defined as the smallest closed set with full π-measure, but can be constructed in the following way: Consider any Borel set $\Lambda \subseteq \mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(\Lambda) = 1$, then the unique disintegration $(\pi_x)_x$ of π with respect to μ yields a well-defined measurable map $x \to \pi_x$ on a Borel set E in $\mathbb{R}^d$ with $\mu(E) = 1$ such that each x in E is the barycenter of π_x and $\pi_x(\Lambda_x) = 1$. It is clear that $x \in CC(\Lambda_x)$. However, it is not necessarily in $IC(\Lambda_x)$. Note however that for any Borel set B in $\mathbb{R}^d$, the map $x \mapsto \pi_x(B)$ is Borel measurable, hence for each $r > 0$, the set $B_r := \{(x, y) \,|\, x \in E,\, \pi_x(B_r(y)) > 0\}$ is Borel (Here, $B_r(y)$ is the open ball with center y and radius r in $\mathbb{R}^d$) and consequently the set $\Theta := \{(x, y) \,|\, x \in E,\, y \in spt(\pi_x)\} = \bigcap_{n=1}^{\infty} B_{1/n}$ is also Borel. Letting $\Gamma := \Lambda \cap \Theta$, it is clear that $\pi(\Gamma) = 1$ and $\pi_x(\Gamma_x) = 1$ for all $x \in E$. Finally note that the probability measure π_x has its barycenter at x and that $\Gamma_x \subseteq spt(\pi_x)$, hence $x \in IC(\Gamma_x)$ for $x \in E$.

This leads us to the following purely set-theoretic definition.

Definition 2.2.2. Say that a Borel set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ is a *martingale supporting set,* if

$$\text{for every } x \in X_\Gamma,\ x \in IC(\Gamma_x). \tag{2.2.1}$$

We let S_{MT} denote the class of all martingale supporting sets, i.e.

$$S_{MT} := \{\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d : \Gamma \text{ is a Borel set such that for every } x \in X_\Gamma,\ x \in IC(\Gamma_x)\}.$$

And just like martingale supporting sets can be defined independently of transport problems, duality can also be defined without the presence of any prescribed marginals.

Definition 2.2.3. *Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function. We shall say that a subset G of $\mathbb{R}^d \times \mathbb{R}^d$ admits a c-dual, or that G is c-dualizable, if there exists a triplet $\{\alpha, \beta, \gamma\}$, $\alpha : X_G \to \mathbb{R}$, $\beta : Y_G \to \mathbb{R}$, $\gamma : X_G \to \mathbb{R}^d$, such that the following duality relation (for the maximization problem) holds:*

$$\beta(y) - c(x, y) \geq \gamma(x) \cdot (y - x) + \alpha(x) \ \forall x \in X_G, y \in Y_G, \tag{2.2.2}$$

$$\beta(y) - c(x, y) = \gamma(x) \cdot (y - x) + \alpha(x) \ \forall (x, y) \in G. \tag{2.2.3}$$

In the minimization problem, the inequality in (2.2.2) *is reversed.*

We will call the triplet $\{\alpha, \beta, \gamma\}$ a c-dual for G. A function $\beta : Y_G \to \mathbb{R}$ is called a c-dual for G if there exist α and γ such that $\{\alpha, \beta, \gamma\}$ is a c-dual for G.

Our first main result show that c-dualizable martingale supporting sets enjoy strong structural results, which will be proved in sections 2.3.2 and 2.3.3. Recall that in the minimization problem, we can and shall assume without loss of generality, that $\mu \wedge \nu = 0$.

Theorem 2.2.4. *Suppose $x \mapsto c(x, y)$ is locally Lipschitz, where the Lipschitz constants are uniformly bounded in y, (e.g. $c(x, y) = |x - y|$). Then, the following holds:*

1. [**Regularity of dual solutions**] *Let Γ be a Borel set in S_{MT} that admits a c-dual in the sense of definition 2.2.3 above and suppose that $X_\Gamma \subseteq \Omega := IC(Y_\Gamma)$ with Ω being an open set in $\mathbb{R}^d$. Then, there exist $\overline{\alpha} : \Omega \to \mathbb{R}$, $\overline{\beta} : \mathbb{R}^d \to \mathbb{R}$, $\overline{\gamma} : \Omega \to \mathbb{R}^d$, such that the triplet $(\overline{\alpha}, \overline{\beta}, \overline{\gamma})$ is also a c-dual for Γ, with $\overline{\alpha}$ being locally Lipschitz (thus, differentiable a.e) while $\overline{\gamma}$ is locally bounded in Ω.*

   If we further assume that $y \mapsto c(x, y)$ is locally Lipschitz with Lipschitz constants uniformly bounded in x, then $\overline{\beta}$ is also locally Lipschitz.

   In the case of maximization, $\overline{\beta}$ can be taken to be a convex function on $\mathbb{R}^d$ whenever the function $y \mapsto c(x, y)$ is assumed to be convex.

2. [**Structure of an optimal martingale transport when duality is attained**] *Assume now that $c(x, y) = |x - y|$ and that μ is absolutely continuous*

with respect to Lebesgue measure. If $\pi \in MT(\mu,\nu)$ *is a solution of* (2.1.1) *for either the minimization or maximization problem and if the dual problem is attained, then for* μ*-almost all* x, Γ_x *is contained in the Choquet boundary of its closed convex hull i.e.,* $\Gamma_x \subseteq \mathrm{Ext}\left(\overline{\mathrm{conv}}(\Gamma_x)\right)$.

Since duality is not attained in general ([3] or Example 2.3.2 below), we investigate the possibility of decomposing an optimal martingale transport into "irreducible components" on which duality is attained for the "restricted" problem. First, we introduce the concept of a *convex paving.*

Definition 2.2.5. *Let* Φ *be a family of mutually disjoint open convex sets in* $\mathbb{R}^d$ *(Recall that here, the openness of a set* C *is with respect to* $V(C)$*). Given a set* $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$*, we shall say that* Φ *is a convex paving for* Γ *provided*

1. $X_\Gamma \subseteq \bigcup_{C \in \Phi} C$.
2. *each* $C \in \Phi$ *contains at least one element* x *in* X_Γ *(*C *is then denoted* $C(x)$*).*
3. *For any* $z, x \in X_\Gamma$*, we have:* $IC(\Gamma_z) \cap C(x) \neq \emptyset \Rightarrow IC(\Gamma_z) \subseteq C(x)$.

Note that such a paving clearly defines an equivalent relation on X_Γ by simply defining $x \sim_\Phi x'$ if and only if $C(x) = C(x')$. The corresponding equivalent classes are then $[x] = C(x) \cap X_\Gamma$.
There can be many convex pavings of a set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$; take for example $\Phi := \{\mathbb{R}^d\}$ which however doesn't give much information about Γ. We therefore introduce the following concept.

Definition 2.2.6. *For a fixed set* $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$*, we shall say that* Φ *is an irreducible convex paving for* Γ *if for any other convex paving* Ψ *for* Γ*, we have the following property: If* $C \in \Phi$*,* $D \in \Psi$ *are such that* $C \cap D \neq \emptyset$*, then necessarily* $C \subseteq D$.

Note that an irreducible convex paving for a set Γ is necessarily unique. As to their existence, we shall show in section 4 the following result.

Theorem 2.2.7. For every martingale supporting set $\Gamma \in S_{MT}$, there exists a unique irreducible convex paving for Γ.

Now, a key property of an optimal martingale transport $\pi \in MT(\mu,\nu)$ is due to Beiglböck and Juillet [5]. It says that there exists a set Γ of full π-measure in $\mathbb{R}^d \times \mathbb{R}^d$ such that each one of its finite subsets has a c-dual. This is one of the consequences of the variational lemma in [5], where duality on finite sets is obtained via linear programming (see [5] and [27]). We state it here as a proposition.

Definition 2.2.8. *Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function. (We do not assume anything on the cost function c.) We say that a set $G \subseteq \mathbb{R}^d \times \mathbb{R}^d$ is finitely c-dualizable, if every finite subset $H \subseteq G$ admits a c-dual in the sense of definition 2.2.3.*

The following proposition shows that an optimal martingale transport concentrates naturally on a martingale supporting Borel set in S_{MT} that is also finitely c-dualizable.

Proposition 2.2.9. *Let $\pi \in MT(\mu,\nu)$ be an optimal martingale transport for Problem* (2.1.1)*. Assuming the cost c continuous, then there exists a Borel set $\Gamma \in S_{MT}$ such that $\pi(\Gamma) = 1$ and Γ is finitely c-dualizable.*

Indeed, it is shown in [39] that there exists a Borel set Λ in $\mathbb{R}^d \times \mathbb{R}^d$ with $\pi(\Lambda) = 1$, that satisfies a certain monotonicity property, which is in fact the martingale counterpart of the c-cyclic monotonicity that is inherent to the Monge-Kantorovich theory. As mentioned above, by the duality theorem of linear programming, this property is equivalent to saying that Λ is finitely c-dualizable. Starting with such a Λ, use now the remark described before Definition 2.2.2 to find a martingale supporting set $\Gamma \subseteq \Lambda$ such that $\pi(\Gamma) = 1$. Besides being in S_{MT}, it is clear that Γ is also finitely c-dualizable.

Remark 2.2.10. We note that in the above proposition, the continuity assumption on the cost can be relaxed, in particular when one has the "no-duality gap" (2.1.3) or (2.1.5). See [39] for details.

Since duality is not attained in general, an optimal martingale transport measure is not necessarily concentrated on a c-dualizable set $\Gamma \in S_{MT}$. On the other hand, in view of the above result, we can and will assume that it is concentrated on a set

$\Gamma \in S_{MT}$ that is finitely c-dualizable. This leads to the question of finding "maximal" components of Γ that are c-dualizable. It turns out that this is indeed the case as we show that if a set $G \in S_{MT}$ is such that all of its finite subsets have c-duals, then the same hold for $G \cap (C \times \mathbb{R}^d)$, where C is any component of the irreducible convex paving Φ of G.

In view of Theorem 2.2.7, we can now state the following.

Theorem 2.2.11. *Let $\Gamma \in S_{MT}$ be finitely c-dualizable. Then, there exists an irreducible convex paving Φ for Γ such that for each $x \in X_\Gamma$, the set $\Gamma \cap (C(x) \times \mathbb{R}^d)$ has a c-dual in the sense of definition 2.2.3, where the sets $C(x)$ are the convex components in Φ.*

Remark 2.2.12. Theorem 2.2.11 can be seen as a martingale counterpart to a celebrated result of Rockafellar [33] in the Monge-Kantorovich theory for mass transport, which essentially says that the property of c-cyclical monotonicity that characterizes the support of optimal transport plans are c-dualizable via a pair of functions, one being c-convex and the other being its c-Legendre transform. Here, the finitely c-dualizable property replaces cyclic monotonicity, while duality for martingale supporting sets require three functions, instead of the two in Rockafellar's result. However, in the martingale case, "dualizability" does not always hold on the whole support but requires it to be decomposed into irreducible components.

Theorems 2.2.4 and 2.2.11 yield several structural results in general dimensions such as the following. Note that the attainability of the dual problem is not assumed here.

Corollary 2.2.13 (dimensional result). *Let π be a solution of the optimization problem* (2.1.1) *with $c(x, y) = |x - y|$ and suppose μ is absolutely continuous with respect to Lebesgue measure. Then, there is a set Γ with $\pi(\Gamma) = 1$ such that for μ-almost every x in $\mathbb{R}^d$,*

(1) the Hausdorff dimension of Γ_x is at most d-1, and

(2) whenever $dim V(\Gamma_x) = d$, then $\Gamma_x \subseteq \mathrm{Ext}\left(\overline{\mathrm{conv}}(\Gamma_x)\right)$.

Proof. Let Γ be as in Proposition 2.2.9, and first consider those points x with $\dim V(\Gamma_x) = d$. In this case, the disjoint sets $C(x)$ in Theorem 2.2.11 are open sets in $\mathbb{R}^d$ and so, the restriction of μ to each of these dualizable components is again absolutely continuous. Theorem 2.2.4 can then be applied. Note now that the Choquet boundary has dimension at most $d-1$. This shows that for μ-a.e. x in the open set $\bigcup_{\dim V(C)=d} C$, we have that $\dim V(\Gamma_x) \leq d-1$. The property also obviously holds outside that set, which means that item (1) is also verified. $\square$

We now perform a disintegration of π with respect to the decomposition $\{\Gamma_C\}_{C\in\Phi}$. For $\Gamma \in S_{MT}$, consider the irreducible components $\{C; C \in \Phi\}$ as given in Theorem 2.2.7. We shall assume in the sequel that the map $\Xi : \Gamma \to \mathcal{K}(\mathbb{R}^d)$ defined by $(x, y) \mapsto \overline{C(x)}$ is measurable, where $\mathcal{K}(\mathbb{R}^d)$ is the space of convex closed subsets of $\mathbb{R}^d$, which is a separable complete metric space for the Hausdorff metric. We can then show that a martingale transport plan π can be canonically disintegrated into its components given by $(\Gamma \cap (C(x) \times \mathbb{R}^d))_{x\in X_\Gamma}$. See Section 2.4.4 for the proof.

Theorem 2.2.14 (Disintegration of martingale plans). *Let (μ, ν) be probability measures on $\mathbb{R}^d$ in convex order and let $\pi \in MT(\mu, \nu)$. In the case of minimization problem, we assume further that $\mu \wedge \nu = 0$. Then, there exists a set $\Gamma \in S_{MT}$ with $\pi(\Gamma) = 1$ such that if $\tilde{\pi} = \Xi_{\#}\pi$ denotes the push-forward of π into $\mathcal{K}(\mathbb{R}^d)$, and $I \subseteq \mathcal{K}(\mathbb{R}^d)$ is the image of Γ by Ξ, then the following holds:*

(1) There exists a disintegration of π along the map Ξ such that

$$\pi(S) = \int_I \pi_C(S) d\tilde{\pi}(C) \quad \text{for each Borel set } S \subseteq \mathbb{R}^d \times \mathbb{R}^d, \tag{2.2.4}$$

where for $\tilde{\pi}$-a.e. C, π_C is a probability measure supported on $\Gamma_C := \Gamma \cap (C \times \mathbb{R}^d)$.

(2) For $\tilde{\pi}$-a.e. $C \in I$, there exist probability measures μ_C, ν_C such that the couple (μ_C, ν_C) is in convex order, μ_C is supported on $X_C := X_\Gamma \cap C$, ν_C on Y_{Γ_C}, and $\pi_C \in MT(\mu_C, \nu_C)$.

(3) If π is optimal for problem (2.1.1) in $MT(\mu, \nu)$, then for $\tilde{\pi}$-a.e. $C \in I$, π_C is

optimal for the same problem on $MT(\mu_C, \nu_C)$. Furthermore, Γ_C admits a c-dual in the sense of definition 2.2.3. In particular, duality is attained for π_C.

(4) If μ_C is absolutely continuous with respect to Lebesgue measure on $V(C)$, and $c(x,y) = |x-y|$, then for μ_C-almost all x, $\Gamma_x \subseteq \mathrm{Ext}(\overline{\mathrm{conv}}(\Gamma_x))$.

The decomposition of Γ into components $\{C\}_{C\in I}$ given by Theorem 2.2.14 was motivated by a similar one proposed by Beiglböck-Juillet [5] in the one dimensional case ($d=1$). It is however quite different since in their case the decomposition depends only on the marginals μ and ν. Theirs is also a countable partition, which makes the restricted problems much more amenable to analysis. Actually, the intervals in their decomposition are simply the connected components of the set where the potentials of μ and ν are different on the real line.

In fact, our decomposition coincides with the one performed by Beiglböck-Juillet in the one-dimensional case. But, in the higher dimensional cases, our decomposition may also depend on the set Γ in S_{MT}, where the martingale transport concentrates. Also, our decomposition can be uncountable, and the induced probability measures μ_C's can be Dirac measures (see Example 2.5.2). The relationship between the two decompositions will be investigated in a forthcoming paper [21].

The above decomposition yields certain structural results for the original optimal martingale transport, especially in the case $c(x,y) = |x-y|$. For the next result, define the function $\delta(x)$ as the radius of the largest open ball contained in $C(x)$, i.e.

$$\delta(x) = \sup\{r \geq 0 : B(x,r) \subseteq C(x)\} \text{ for } x \in X_\Gamma, \quad \text{and define the set}$$
$$E = \{x \in X_\Gamma : \Gamma_x \not\subseteq \mathrm{Ext}\left(\overline{\mathrm{conv}}(\Gamma_x)\right)\}.$$

Theorem 2.2.15. *Let $c(x,y) = |x-y|$ and suppose $\mu << \mathcal{L}^d$. Let $\pi \in MT(\mu,\nu)$ be a solution of (2.1.1) and let $\Gamma \in S_{MT}$ be a finitely c-dualizable Borel set such that $\pi(\Gamma) = 1$. Assume the function δ and the set E are measurable, and suppose*

$$\dim(V(C(x))) \geq d-1 \text{ for } \mu \text{ a.e. } x. \tag{2.2.5}$$

Then, for μ almost every $x \in \mathbb{R}^d$ we have $\Gamma_x \subseteq \mathrm{Ext}(\overline{\mathrm{conv}}(\Gamma_x))$.

This will be proved in Section 2.5.2, as well as the following corollary.

Corollary 2.2.16. *Let $d = 2$, $c(x, y) = |x - y|$ and suppose $\mu << \mathcal{L}^d$. Let $\pi \in MT(\mu, \nu)$ be a solution of (2.1.1) and let $\Gamma \in S_{MT}$ be a finitely c-dualizable Borel set such that $\pi(\Gamma) = 1$. Assume E is measurable and suppose*

(2.2.6) *μ is concentrated in an open set U in $\mathbb{R}^d$, while ν is supported on U^c.*

Then, for μ almost every $x \in \mathbb{R}^2$, we have $\Gamma_x \subseteq \mathrm{Ext}(\overline{\mathrm{conv}}(\Gamma_x))$.

Note that the separation assumption (2.2.6) is very mild in the minimization problem, since as mentioned above, we can then assume $\mu \wedge \nu = 0$. Also note that in this 2D case, we do not need the measurability of δ once (2.2.6) is assumed.

We shall also give a partial answer to Conjecture 2 in Section 2.5.3. Namely, assuming that the target measure is discrete, and in this case, it holds true in both the maximization and minimization cases:

Theorem 2.2.17. *Let $c(x, y) = |x - y|$ (or more generally for $c(x, y) = |x - y|^p$ with $p \neq 2$), suppose μ is absolutely continuous with respect to the Lebesgue measure, and that ν is discrete; i.e. ν is supported on a countable set. Let $\pi \in MT(\mu, \nu)$ be a solution of (2.1.1) and let $\Gamma \in S_{MT}$ be a finitely c-dualizable Borel set such that $\pi(\Gamma) = 1$. Then for μ a.e. x, the fiber Γ_x consists of exactly $d + 1$ points which are vertices of a d-dimensional polytope in $\mathbb{R}^d$, and therefore the optimal solution is unique.*

Now we give a couple of examples, which illustrate that the above stated conjectures could be the best structural results we can hope for.

Example 2.2.18. The polytope-like structure of the support required in conjecture 2 does not hold in general for the corresponding maximization problem. Indeed, since $\frac{1}{2}(|x - y| - 1)^2 \geq 0$, we have

$$\frac{1}{2}|y|^2 - \frac{1}{2}|x|^2 + 1 - x \cdot (y - x) \geq |x - y| \quad \text{on } \mathbb{R}^d \times \mathbb{R}^d, \tag{2.2.7}$$

with equality on the set $\{(x,y); |x-y|=1\}$. The functions $\alpha(x)=\frac{1}{2}|x|^2-1$, $\beta(y)=\frac{1}{2}|y|^2$ and $\gamma(x)=x$ then form a dual triplet for the maximization problem with cost $|x-y|$. This means that every martingale (X,Y) with $|X-Y|=1$ a.s. is optimal for the maximization problem corresponding to its own marginals $X\sim\mu$ and $Y\sim\nu$. Hence, Γ_x is not in general a discrete set.

Finally, we consider the uniqueness question in conjecture 2, and whether it could hold for the maximization problem. In [5] it is shown that when $d=1$ the solution of the martingale transport problem (2.1.1) is unique for both max/min problem under the assumption that μ is absolutely continuous. Also, it is reported in [31] that in the minimization problem with radially symmetric marginals (μ,ν), the minimizer is again unique in any dimension. We note however that, unlike the minimization case, one cannot expect the uniqueness of a maximizing martingale measure in higher dimensions, even in the radially symmetric case, as the following example indicates.

Example 2.2.19. Let μ be a radially symmetric probability measure on $\mathbb{R}^2\simeq\mathbb{C}$ such that $\mu(\{0\})=0$. Let $z_1=\cos\frac{\pi}{4}+i\sin\frac{\pi}{4}$, $z_2=\cos\frac{\pi}{4}-i\sin\frac{\pi}{4}$, $z_3=-z_1$ and $z_4=-z_2$, and define the probability measures π_1 and π_2 on $\mathbb{C}\times\mathbb{C}$, whose disintegrations π_x^1 and π_x^2 for each $x\in\mathbb{C}, x\neq 0$, are given by,

$$\pi_x^1=\frac{1}{4}\delta_{x+\frac{x}{|x|}z_1}+\frac{1}{4}\delta_{x+\frac{x}{|x|}z_2}+\frac{1}{4}\delta_{x+\frac{x}{|x|}z_3}+\frac{1}{4}\delta_{x+\frac{x}{|x|}z_4}$$
$$\pi_x^2=\frac{1}{8}\delta_{x+\frac{x}{|x|}z_1}+\frac{3}{8}\delta_{x+\frac{x}{|x|}z_2}+\frac{1}{8}\delta_{x+\frac{x}{|x|}z_3}+\frac{3}{8}\delta_{x+\frac{x}{|x|}z_4}.$$

Then, by the discussion around Section (2.5.1), one can see that both π_1 and π_2 are optimal for the maximization problem corresponding to μ and $\nu:=\nu_1=\nu_2$, where $d\nu_i(y)=\int_{\mathbb{C}}\pi_x^i(y)\,d\mu(x)$, $i=1,2$, hence, the maximizer is not unique.

2.3 Profile of an optimal martingale transport when the dual is attained

The goal of this section is to prove Theorem 2.2.4 which shows that dual attainment of the optimization problem (2.1.1) yields good structural results thanks to regularity properties of the dual functions.

2.3.1 Duality for optimal martingale transports

As mentioned in the introduction, there is a dual formulation for problem (3.1.2), just like in the Monge-Kantorovich theory for (non-martingale) mass transport.

Lemma 2.3.1. *(see e.g. [3]) Let μ and ν be two probability measures on $\mathbb{R}^d$ in convex order, and let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function that is upper semi-continuous, then*

(2.3.1)
$$\max\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,d\pi;\ \pi \in MT(\mu,\nu)\right\} = \inf\left\{\int_{\mathbb{R}^d}\beta d\nu - \int_{\mathbb{R}^d}\alpha d\mu;\ (\alpha,\beta)\in D_M(\mu,\nu,c)\right\},$$

where $D_M(\mu,\nu,c)$ is the set of all $(\alpha,\beta) \in L^1(\mu)\times L^1(\nu)$ such that for some $\gamma \in C_b(\mathbb{R}^d)$,

$$\beta(y) - \alpha(x) - \gamma(x)(y-x) \geq c(x,y) \text{ for all } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{2.3.2}$$

A similar result holds for the cost minimization problem, provided c is lower semi-continuous, and the inequality in (2.3.2) is reversed.

Note that if the dual problem is attained at functions α, β and γ in $D_M(\mu,\nu,c)$, then since (in the maximization problem),

$$\beta(y) - \alpha(x) - \gamma(x)(y-x) \geq c(x,y) \quad \text{for all } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d, \tag{2.3.3}$$

and μ and ν are the marginals of some optimal π in $\mathrm{MT}(\mu,\nu)$, then since $\int \gamma(x)\cdot(y-x)\,d\pi(x,y) = 0$ (due to martingale condition), we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\{\beta(y)-\alpha(x)-\gamma(x)(y-x)\}d\pi(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)d\pi(x,y).$$

It follows that

$$\beta(y) - \alpha(x) - \gamma(x)(y-x) = c(x,y) \quad \text{for } \pi \text{ a.e } \ (x,y) \in \mathbb{R}^d \times \mathbb{R}^d,$$

hence the equality holds on a concentration set Γ of π.

Conversely, if $G \subseteq \mathbb{R}^d \times \mathbb{R}^d$ and $\pi^*(G) = 1$ for some $\pi^* \in \mathrm{MT}(\mu, \nu)$ and if there exists a triplet (α, β, γ) in $D_M(\mu, \nu, c)$ with equality (2.3.3) holding on G, then π^* is an optimal solution of (2.1.1). Indeed, let $\pi \in \mathrm{MT}(\mu, \nu)$ and let H be such that $\pi(H) = 1$. As $\mu(X_G) = 1$ and $\nu(Y_G) = 1$, we can then assume that $X_H \subseteq X_G$ and $Y_H \subseteq Y_G$, hence by integrating (2.3.3) with π, we get

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi(x, y) \leq \int_{\mathbb{R}^d} \beta(y)\, d\nu(y) - \int_{\mathbb{R}^d} \alpha(x)\, d\mu(x).$$

(Again $\int \gamma(x) \cdot (y - x)\, d\pi(x, y) = 0$ since $\pi \in \mathrm{MT}(\mu, \nu)$). However, by integrating (2.3.3) with π^* and since we have equality on G, we get

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi^* = \int_{\mathbb{R}^d \times \mathbb{R}^d} \{\beta(y) - \alpha(x) - \gamma(x)(y - x)\} d\pi^* = \int_{\mathbb{R}^d} \beta(y)\, d\nu - \int_{\mathbb{R}^d} \alpha(x)\, d\mu.$$

This shows that π^* is optimal.

This suggests that dual attainability is actually a property of the support of the optimal martingale transport and not of the measure itself, which is why we introduce in Definition 2.2.3 the concept of pointwise c-duality for a set. Note that the triplet (α, β, γ) defining c-duality in Definition 2.2.3 are much more general than the elements in $D_M(\mu, \nu, c)$, and our results indicate that dual attainment in (2.3.1) does not occur in general even in the most relaxed way, where the functions (α, β, γ) are allowed to be highly non-regular.

An obvious but important remark is that if G is c-dualizable in the sense of Definition 2.2.3, then any subset H of G is also c-dualizable, since any c-dual triplet for G would work for H as well. We have also seen above that if G admits a c-dual and if it supports a martingale transport π^* in $\mathrm{MT}(\mu, \nu)$, then π^* is an optimal solution of (2.1.1) for the cost c. Hence, only Borel sets where optimal martingale measure concentrate can admit a c-dual, but not conversely, as we remarked in the introduction that there exists an optimal martingale measure whose support can never have a c-dual (see [3]). We give here a very simple example.

Example 2.3.2. Let $\mu = \nu$ be two identical probability measures on the interval $[0, 1]$, then the only martingale (say π) from μ to itself is the identity transport, hence it is obviously the solution of the maximization problem with respect to the distance cost, and its support is $\Gamma = \{(x,x) : x \in [0,1]\}$. If now $\{\alpha, \beta, \gamma\}$ is a solution to the dual problem, then

$$\beta(y) \geq |x-y| + \gamma(x)\cdot(y-x) + \alpha(x) \ \forall x \in [0,1], \forall y \in [0,1];$$
$$\beta(y) = |x-y| + \gamma(x)\cdot(y-x) + \alpha(x) \ \forall (x,y) \in \Gamma.$$

The above relations easily yield that for any $0 < a < b < 1$, we have $\gamma(a)+2 \leq \gamma(b)$, which means that it is impossible to define a suitable real-valued function γ for a.e. x in $[0,1]$.

The fact that a concentration set of an optimal martingale transport need not admit a c-dual is in sharp contrast with the usual optimal transport problem and makes the martingale transport problem difficult to analyze. Nevertheless, we shall see in section 3 that whenever a concentration set admits a dual, then it exhibits a special extremal configuration.

2.3.2 Regularity of the dual functions

In Definition 2.2.3, the triplets (α, β, γ) are only assumed to be real-valued functions, which means that Lemma 2.3.5 is not directly applicable. However, we will now show that dual triplets have regular extensions in such a way that Theorem 2.2.4 item (1) would apply.

Lemma 2.3.3 (Theorem 2.2.4 item (1)). *Assume that Γ is a Borel martingale supporting set in S_{MT} with a c-dual triplet $\{\alpha, \beta, \gamma\}$. Set $\Omega := IC(Y_\Gamma)$ and suppose that $X_\Gamma \subseteq \Omega$ and Ω is open in $\mathbb{R}^d$. Assume that $x \mapsto c(x,y)$ is locally Lipschitz, where the local Lipschitz constants are uniformly bounded in y (e.g. $c(x,y) = |x-y|$). Then, there exists a locally Lipschitz $\overline{\alpha} : \Omega \to \mathbb{R}$, a locally bounded $\overline{\gamma} : \Omega \to \mathbb{R}^d$, and $\overline{\beta} : \mathbb{R}^d \to \mathbb{R}$ such that $\{\overline{\alpha}, \overline{\beta}, \overline{\gamma}\}$ coincides with $\{\alpha, \beta, \gamma\}$ on $X_\Gamma, Y_\Gamma, X_\Gamma$, respectively (thus, is also a c-dual triplet for Γ). In particular, $\overline{\alpha}$ is differentiable a.e. in Ω.*

Proof. Recall the duality relations for the maximization problem:

$$\beta(y) - c(x,y) \geq \gamma(x) \cdot (y-x) + \alpha(x) \ \forall x \in X_\Gamma, y \in Y_\Gamma \tag{2.3.4}$$

$$\beta(y) - c(x,y) = \gamma(x) \cdot (y-x) + \alpha(x) \ \forall (x,y) \in \Gamma. \tag{2.3.5}$$

Let $B_x(y) = \beta(y) - c(x,y)$. First, we extend the function α on Ω in the following way:

$$\overline{\alpha}(x) := \sup\{c \in \mathbb{R} : \exists a \in \mathbb{R}^d \text{ such that } B_x(y) \geq a \cdot (y-x) + c \ \forall\, y \in Y_\Gamma\} \text{ for } x \in \Omega.$$

Thus $\overline{\alpha} \geq \alpha$ on X_Γ. Also note that (2.3.5) and the fact $x \in IC(\Gamma_x)$ imply that $\overline{\alpha}(x) = \alpha(x)$ for all $x \in X_\Gamma$, thus it is indeed an extension that is finite on X_Γ. We claim that it is also finite at every $x \in \Omega$. Indeed, for $x \in \Omega = IC(Y_\Gamma)$, we can choose $\{y_1, ..., y_s\} \subseteq Y_\Gamma$ such that

$$x \in U := IC(\{y_1, ..., y_s\}) \text{ and } U \text{ is open in } \mathbb{R}^d. \tag{2.3.6}$$

It is clear that $\overline{\alpha}(z) \leq M := \max_{y_i} B_z(y_i)$ for all $z \in U$, and since $x \mapsto c(x,y)$ is locally Lipschitz, we see that $\overline{\alpha}(z)$ is bounded above for all z near x. In other words, $\overline{\alpha}$ is locally upper bounded. The proof below will show that it is also locally lower bounded.

Now we extend γ to Ω in the following way:

$$\begin{aligned}\overline{\gamma}(x) := \{a \in \mathbb{R}^d : \beta(y) - c(x,y) &\geq a \cdot (y-x) + \overline{\alpha}(x) \ \forall y \in Y_\Gamma, \\ \beta(y) - c(x,y) &= a \cdot (y-x) + \overline{\alpha}(x) \ \forall y \in \Gamma_x\}\end{aligned}$$

Note that $\overline{\gamma}(x)$ is a multivalued function. For $x \in X_\Gamma$, $\overline{\alpha}(x) = \alpha(x)$, and so $\gamma(x) \in \overline{\gamma}(x)$. But we must show that $\overline{\gamma}(x)$ is nonempty for any $x \in \Omega$. Choose an approximating sequence $\{c_n\}$ of $\overline{\alpha}(x)$ and corresponding $\{a_n\}$, i.e. $B_x(y) \geq a_n \cdot (y-x) + c_n$ and $c_n \nearrow \overline{\alpha}(x)$. Now by (2.3.6) and $B_x(y_i) \leq M$, it is clear that $\{a_n\}$ must be bounded, therefore $\overline{\gamma}(x)$ is nonempty and compact. Below, we abuse notation by letting $\overline{\gamma}(x)$ denote an arbitrary selection of a from $\overline{\gamma}(x)$.

We now prove that $\overline{\alpha}$ is locally Lipschitz. First, fix $R > 0$ and let $x \in X_\Gamma, x' \in \Omega, y \in Y_\Gamma$ with $|x|, |x'| < R$. By the local Lipschitz property of c in x, i.e.

$$c(x, y) - c(x', y) \leq C|x - x'|$$

for some $C = C(R) > 0$ and for all $|x|, |x'| < R$, we see

$$\alpha(x) + \gamma(x) \cdot (y - x) \leq B_x(y) \leq B_{x'}(y) + C|x' - x|. \tag{2.3.7}$$

Thus,

$$\alpha(x) - C|x' - x| + \gamma(x) \cdot (x' - x) + \gamma(x) \cdot (y - x') \leq B_{x'}(y).$$

The definition of $\overline{\alpha}$ gives

$$\alpha(x) - C|x' - x| + \gamma(x) \cdot (x' - x) \leq \overline{\alpha}(x'). \tag{2.3.8}$$

In particular, $\overline{\alpha}$ is locally lower bounded. Knowing that $\overline{\alpha}$ is finite everywhere in Ω, the above argument can be repeated, giving (2.3.8) for any $x, x' \in \Omega$;

$$\overline{\alpha}(x) - C|x' - x| + \overline{\gamma}(x) \cdot (x' - x) \leq \overline{\alpha}(x').$$

By interchanging x and x', we get

$$|\overline{\alpha}(x) - \overline{\alpha}(x')| \leq \left((|\overline{\gamma}(x)| \vee |\overline{\gamma}(x')|) + C \right) |x' - x|.$$

Therefore, the local boundedness of $\overline{\gamma}$ will imply that $\overline{\alpha}$ is locally Lipschitz. Now, to show this property for $\overline{\gamma}$, use (2.3.6) and let V be a small neighbourhood of x whose closure is in U. Since $\overline{\alpha}$ is bounded on V, there exists $C > 0$ such that

$$\overline{\gamma}(z) \cdot (y_i - z) \leq C \ \forall z \in V, \ i = 1, 2, ..., s, \tag{2.3.9}$$

which says that $\overline{\gamma}$ is bounded on V. Hence, $\overline{\alpha}$ is locally Lipschitz.

Finally, we can extend β to the whole of $\mathbb{R}^d$ via the formula,

$$\overline{\beta}(y) := \sup_{x\in\Omega}\{c(x,y) + \overline{\gamma}(x)\cdot(y-x) + \overline{\alpha}(x)\}. \tag{2.3.10}$$

Clearly $\beta = \overline{\beta}$ on Y_Γ.

If now $y \mapsto c(x,y)$ is convex, the above definition of $\overline{\beta}$ immediately implies the convexity and finiteness on Y_Γ, hence also on $conv(Y_\Gamma)$. Therefore $\overline{\beta}$ is continuous in Ω and differentiable a.e. in Ω. Moreover, in the case $c(x,y) = |x-y|$ we observe that (2.3.5) immediately implies that

$$x \notin \Gamma_x \text{ for all } x \in X_\Gamma \text{ where } \overline{\beta} \text{ is differentiable.}$$

For the minimization problem, the duality is

$$\beta(y) - c(x,y) \leq \gamma(x)\cdot(y-x) + \alpha(x)\ \forall x \in X_\Gamma, y \in Y_\Gamma, \tag{2.3.11}$$

$$\beta(y) - c(x,y) = \gamma(x)\cdot(y-x) + \alpha(x)\ \forall (x,y) \in \Gamma. \tag{2.3.12}$$

Thus, we define the extensions as

$$\overline{\alpha}(x) := \inf\{c \in \mathbb{R} : \exists a \in \mathbb{R}^d \text{ such that } B_x(y) \leq a\cdot(y-x) + c\ \forall\, y \in Y_\Gamma\}$$

$$\begin{aligned}\overline{\gamma}(x) := \{a \in \mathbb{R}^d : \beta(y) - c(x,y) &\leq a\cdot(y-x) + \overline{\alpha}(x)\ \forall y \in Y_\Gamma \\ \beta(y) - c(x,y) &= a\cdot(y-x) + \overline{\alpha}(x)\ \forall y \in \Gamma_x\}\end{aligned}$$

$$\overline{\beta}(y) := \inf_{x\in\Omega}\{c(x,y) + \overline{\gamma}(x)\cdot(y-x) + \overline{\alpha}(x)\} \tag{2.3.13}$$

The same reasoning shows that $\overline{\alpha}$ is locally Lipschitz. But, note that in this case $\overline{\beta}$ may not be a convex function even if $y \mapsto c(x,y)$ is convex.

Finally note that if $y \mapsto c(x,y)$ is locally Lipschitz and the local Lipschitz constants are uniformly bounded in $x \in \Omega$, then in both max/min cases, the function

$\overline{\beta}$ given in (2.3.10), (2.3.13), respectively, are locally Lipschitz in y, as long as $\overline{\gamma}(x)$ is uniformly bounded on Ω, since then it is given by the sup or inf of Lipschitz functions. $\square$

The next lemma shows that essentially γ is differentiable whenever α is. This property will be crucial in the proof of Theorem 2.2.4 item (2).

Lemma 2.3.4. *Use the same conditions as in Lemma 2.3.3, and suppose $x \mapsto c(x,y)$ is differentiable at x whenever $x \neq y$ (e.g. $c(x,y) = |x-y|$). Let $\alpha : \Omega \to \mathbb{R}$, $\beta : \mathbb{R}^d \to \mathbb{R}$, and $\gamma : \Omega \to \mathbb{R}^d$ be a c-dual triplet for Γ, and assume that for the maximization problem, (for minimization the inequality is reversed) there is $s \in \mathbb{R}^d$, such that*

$$\alpha(x') \geq s \cdot (x'-x) + \alpha(x) + o(|x'-x|) \quad \text{as} \quad x' \to x. \tag{2.3.14}$$

Fix $x \in X_\Gamma$, and let V be the vector subspace of $\mathbb{R}^d$ corresponding to the affine space $V(\Gamma_x)$, and assume $\dim V \geq 1$. Let $\mathrm{proj}_V \gamma$ be the orthogonal projection of the value of γ on V. Then α and $\mathrm{proj}_V \gamma$ have a directional derivative at x in every direction $u \in V$.

Proof. By the duality assumption for the maximization problem, for all $x' \in \Omega$ and all $(x,y) \in \Gamma$,

$$c(x',y) + \gamma(x') \cdot (y-x') + \alpha(x') \leq c(x,y) + \gamma(x) \cdot (y-x) + \alpha(x). \tag{2.3.15}$$

Choose a unit vector $u \in V$ and let $x' = x + tu$. Then (2.3.15) is rewritten as

(2.3.16)
$$\frac{\alpha(x+tu)-\alpha(x)}{t} \leq \frac{\gamma(x+tu)-\gamma(x)}{t} \cdot (x+tu-y) + \gamma(x) \cdot u - \frac{c(x+tu,y)-c(x,y)}{t} \text{ if } t > 0$$

(2.3.17)
$$\frac{\alpha(x+tu)-\alpha(x)}{t} \geq \frac{\gamma(x+tu)-\gamma(x)}{t} \cdot (x+tu-y) + \gamma(x) \cdot u - \frac{c(x+tu,y)-c(x,y)}{t} \text{ if } t < 0$$

Let us use the notation $D_{t,u}f(x) = \frac{f(x+tu)-f(x)}{t}$. Now assumption (2.3.14) says that

$$\limsup_{t\uparrow 0} D_{t,u}\alpha(x) \le \liminf_{t\downarrow 0} D_{t,u}\alpha(x). \tag{2.3.18}$$

Since $x \in int(conv(\Gamma_x))$, there exists $y_1, ..., y_k \in \Gamma_x \setminus \{x\}$, $p_1, ..., p_k \ge 0$, $q_1, ..., q_k \ge 0$, $\Sigma p_i = 1$, $\Sigma q_i = 1$, $t_+ > 0$, $t_- < 0$, such that

$$x + t_+ u = \Sigma p_i y_i$$
$$x + t_- u = \Sigma q_i y_i.$$

Note that the first term on the right side of (2.3.16) and (2.3.17) is linear in y, so by summing up the y_i's with the weights p_i's or q_i's, we get (and we write $\gamma_1(x) := \gamma(x) \cdot u$)

$$D_{t,u}\alpha(x) \le D_{t,u}\gamma_1(x)(t - t_\pm) + C_\pm(t) \text{ if } t > 0 \tag{2.3.19}$$
$$D_{t,u}\alpha(x) \ge D_{t,u}\gamma_1(x)(t - t_\pm) + C_\pm(t) \text{ if } t < 0 \tag{2.3.20}$$

Here $C_+(t), C_-(t)$ are functions of $t \ne 0$, but have limits as $t \to 0$ by the differentiability assumption on the cost. Write $C_\pm = \lim_{t\to 0} C_\pm(t)$, respectively.

By taking $\liminf_{t\downarrow 0}$ in (2.3.19) and $\limsup_{t\uparrow 0}$ in (2.3.20) and by recalling that $t_+ > 0, t_- < 0$, we have

$$\liminf_{t\downarrow 0} D_{t,u}\alpha(x) \le (-t_+) \limsup_{t\downarrow 0} D_{t,u}\gamma_1(x) + C_+$$
$$\liminf_{t\downarrow 0} D_{t,u}\alpha(x) \le (-t_-) \liminf_{t\downarrow 0} D_{t,u}\gamma_1(x) + C_-$$
$$\limsup_{t\uparrow 0} D_{t,u}\alpha(x) \ge (-t_+) \liminf_{t\uparrow 0} D_{t,u}\gamma_1(x) + C_+$$
$$\limsup_{t\uparrow 0} D_{t,u}\alpha(x) \ge (-t_-) \limsup_{t\uparrow 0} D_{t,u}\gamma_1(x) + C_-.$$

This and (2.3.18) combine to give

$$\begin{aligned}\liminf_{t\uparrow 0} D_{t,u}\gamma_1(x) \le \limsup_{t\uparrow 0} D_{t,u}\gamma_1(x) \le \liminf_{t\downarrow 0} D_{t,u}\gamma_1(x) \\ \le \limsup_{t\downarrow 0} D_{t,u}\gamma_1(x) \le \liminf_{t\uparrow 0} D_{t,u}\gamma_1(x),\end{aligned}$$

that is $\gamma_1 = \gamma \cdot u$ is differentiable at x in the direction u. Knowing this, we then take $\limsup_{t\downarrow 0}$ in (2.3.19) and $\liminf_{t\uparrow 0}$ on (2.3.20) to get

$$\limsup_{t\downarrow 0} D_{t,u}\alpha(x) \le (-t_+)\nabla_u\gamma_1(x) + C_+$$
$$\liminf_{t\uparrow 0} D_{t,u}\alpha(x) \ge (-t_+)\nabla_u\gamma_1(x) + C_+.$$

Combining this with (2.3.18), we get the differentiability of α at x in the direction u.

Next, choose any unit vector $v \in V$ orthogonal to u and let $\gamma_2(x) := \gamma(x) \cdot v$. We want to show that $\nabla_u\gamma_2(x)$ exists. We proceed just as before; for some $k \in \mathbb{N}$, there exists $y_1, ..., y_k \in \Gamma_x \setminus \{x\}$, $p_1, ..., p_k \ge 0$, $q_1, ..., q_k \ge 0$, $\Sigma p_i = 1$, $\Sigma q_i = 1$, $t_+ > 0$, $t_- < 0$, such that

$$x + t_+ v = \Sigma p_i y_i$$
$$x + t_- v = \Sigma q_i y_i.$$

By summing up the y_i's and the weights p_i's or q_i's as before, we get this time

(2.3.21) $$D_{t,u}\alpha(x) \le tD_{t,u}\gamma_1(x) - t_\pm D_{t,u}\gamma_2(x) + C_\pm(t) \text{ if } t > 0$$

(2.3.22) $$D_{t,u}\alpha(x) \ge tD_{t,u}\gamma_1(x) - t_\pm D_{t,u}\gamma_2(x) + C_\pm(t) \text{ if } t < 0$$

Taking $\liminf_{t\downarrow 0}$ in (2.3.21) and $\limsup_{t\uparrow 0}$ in (2.3.22) and recalling $t_+ > 0, t_- < 0$,

$$\nabla_u \alpha(x) \leq (-t_+) \limsup_{t\downarrow 0} D_{t,u}\gamma_2(x) + C_+$$
$$\nabla_u \alpha(x) \leq (-t_-) \liminf_{t\downarrow 0} D_{t,u}\gamma_2(x) + C_-$$
$$\nabla_u \alpha(x) \geq (-t_+) \liminf_{t\uparrow 0} D_{t,u}\gamma_2(x) + C_+$$
$$\nabla_u \alpha(x) \geq (-t_-) \limsup_{t\uparrow 0} D_{t,u}\gamma_2(x) + C_-,$$

which implies differentiability of $\gamma_2 = \gamma \cdot v$ at x in the direction u. Now choose an orthonormal basis $\{u, v_1, ..., v_m\}$ of V and write $\mathrm{proj}_V \gamma = (\gamma \cdot u)u + \Sigma(\gamma \cdot v_i)v_i$. We observed that each component of $\mathrm{proj}_V \gamma$ is directionally-differentiable.

For the minimization problem, we need to reverse the inequalities in (2.3.14) and (2.3.15), and then proceed in the same way. Hence, the lemma is proved. □

2.3.3 Structure of optimal martingale supporting sets when the dual is attained

In this subsection, we shall consider $c(x, y) = |x - y|$, and a martingale measure $\pi \in MT(\mu, \nu)$ that is a solution of (2.1.1) that is concentrated on a Borel set Γ. Now we prove Theorem 2.2.4 item (2), which shows that Conjecture 1 holds *for both max/min problem* whenever the optimal martingale is concentrated on a Borel set that admits a c-dual. We shall first indicate the role of the regularity of the "potential functions" (α, β, γ).

Lemma 2.3.5. *Let $\Gamma \in S_{MT}$, Ω an open set in $\mathbb{R}^d$ containing X_Γ, $\alpha : \Omega \to \mathbb{R}$ and $\gamma : \Omega \to \mathbb{R}^d$ be two functions. In the maximization problem, define $\beta : \mathbb{R}^d \to \mathbb{R}$ as*

$$\beta(y) = \sup_{x\in\Omega}\{|x - y| + \gamma(x) \cdot (y - x) + \alpha(x)\}; \tag{2.3.23}$$

For the minimization problem, set

$$\beta(y) = \inf_{x\in\Omega}\{|x-y| + \gamma(x)\cdot(y-x) + \alpha(x)\}. \tag{2.3.24}$$

Assume that Γ *satisfies*

$$\beta(y) = |x-y| + \gamma(x)\cdot(y-x) + \alpha(x) \ \text{ for all } \ (x,y)\in\Gamma. \tag{2.3.25}$$

Then if α *and* γ *are differentiable at* $x \in X_\Gamma$*, then* Γ_x *is supported on the Choquet boundary of the closed convex hull of* Γ_x*.*

Proof. Define the "tilted cone"

$$\zeta(x,y) = \zeta_x(y) = \zeta_y(x) := |x-y| + \gamma(x)\cdot(y-x) + \alpha(x).$$

The duality condition (2.3.25) with (2.3.23) tells us the following: if $(x,y)\in\Gamma$, then for all $x'\in\Omega$,

$$\zeta_{x'}(y) \le \zeta_x(y) \tag{2.3.26}$$

Or (2.3.25) with (2.3.24) we get the reverse inequality.

Note that since $\zeta_x(y)$ is continuous, the same inequality holds for all $y\in\overline{\Gamma_x}$. This obviously implies that, if $y\in\overline{\Gamma_x}$ and $x\neq y$, then the gradient with respect to x vanishes:

$$\nabla\zeta_y(x) = 0 \tag{2.3.27}$$

and in fact (2.3.26) also implies that if $y\in\overline{\Gamma_x}$, then necessarily $x\neq y$. (If $x=y$, then the function $\zeta_y(x)$ strictly increases as x moves along the direction $\nabla_x[\gamma(x)\cdot(y-x)+\alpha(x)]$.) We may call this as non-staying property or unstability, for the maximization problem. For the minimization problem, without loss of generality we already assumed that $x\notin\Gamma_x$, but in fact $x\notin\overline{\Gamma_x}$ as well, by (2.3.28) below.

Now suppose the lemma is false. Then we can find $\{y, y_0, ..., y_s\}\subseteq\overline{\Gamma_x}$ for some

$s \geq 1$ with $y = \Sigma_{i=0}^{s} p_i y_i$, $\Sigma_{i=0}^{s} p_i = 1$. Choose a minimum s such that all $p_i > 0$. Now taking directional derivative in the direction $u = \frac{x-y}{|x-y|}$ gives

$$\nabla_u \zeta_y(x) = \nabla_u \zeta_{y_i}(x) = 0 \ \forall i = 0, 1, ..., s.$$

We compute

$$\nabla_u \zeta_{y_i}(x) = \frac{x - y_i}{|x - y_i|} \cdot u + \nabla_u \gamma(x) \cdot (y_i - x) - \gamma(x) \cdot u + \nabla_u \alpha(x).$$

Then, by linearity of $y \mapsto \nabla_u \gamma(x) \cdot y$, the equation $\nabla \zeta_y(x) = 0$ simply becomes

$$1 = \sum_{i=0}^{s} p_i \frac{x - y_i}{|x - y_i|} \cdot \frac{x - y}{|x - y|}.$$

As $\frac{x-y}{|x-y|}$ is a unit vector and all $p_i > 0$, this can hold only if all y_i lie on the ray emanated from x. The minimality of s then implies that $s = 1$, hence $\{y, y_0, y_1\} \subseteq \overline{\Gamma_x}$ would lie on a ray emanating from x, which is a contradiction, once we prove the following claim:

(2.3.28) $\quad \Gamma_x$ is supported on the topological boundary of the convex hull of Γ_x,

Recall that here the topology is not the topology in $\mathbb{R}^d$ but the topology in $V := V(\Gamma_x)$. If our claim is false and assuming first that $\dim(V) \geq 2$, we can find $y \in \Gamma_x \cap IC(\Gamma_x)$ as a barycenter of a triangle joining 3 points y_0, y_1, y_2 in Γ_x. But the above argument implies that y_0, y_1, y_2 have to be aligned, which is a contradiction. If $\dim(V) = 1$, then as $x \in IC(\Gamma_x)$, we can find $\{y, y_0, y_1\} \subseteq \Gamma_x$ such that x and y are in the interior of the line segment $\overline{y_0 y_1}$. But then again by above, $\{y, y_0, y_1\}$ must lie on the ray (i.e. half-line) emanated from x, a contradiction. Finally, we cannot have $\dim(V) = 0$ since this simply means that $\Gamma_x = \{x\}$, but as we already showed above that $x \notin \Gamma_x$ in the case of maximization, while we already assumed without loss of generality that $x \notin \Gamma_x$ in the case of minimization. □

Proof of Theorem 2.2.4, item (2): First, we claim that there exists a set Γ' such

that $\Gamma' \subseteq \Gamma$, $\pi(\Gamma') = 1$ and $X_{\Gamma'} \subseteq IC(Y_{\Gamma'})$. Let X' be the set of Lebesgue points of X_Γ. Then as $\mu << \mathcal{L}^d$, we have $\mu(X') = 1$. Let $\Gamma' := \Gamma \cap (X' \times \mathbb{R}^d)$. Then, $\Gamma' \in S_{MT}$, $\pi(\Gamma') = 1$ and $X' \subseteq IC(X') \subseteq IC(Y_{\Gamma'})$, as claimed.
Since a c-dual for Γ is also a dual for Γ', Lemma 2.3.3 applies to Γ' and allows to assume that α is differentiable $\mathcal{L}^d$ - a.e. x in $\Omega := IC(Y_{\Gamma'})$. Lemma 2.3.4 gives that whenever α is differentiable at x, then $\text{proj}_V \gamma$ is directionally-differentiable in $V = V(\Gamma_x)$ at x. This enables us to replicate the proof of lemma 2.3.5 and we get that Γ_x is supported on the Choquet boundary of the closed convex hull of Γ_x for $\mathcal{L}^d$ - a.e. x in Ω, and hence for μ a.e. x, since $\mu << \mathcal{L}^d$ and $\mu(\Omega) = 1$. □

2.4 A canonical decomposition of martingale transports

We have shown in the last section that Conjecture 1 holds whenever the dual problem is attained. In this section, we shall decompose an optimal martingale transport π into components on which an induced martingale transport problem is defined in such a way that its dual problem is attained. For that, we shall associate an irreducible convex paving Φ to any Borel set $\Gamma \in S_{MT}$ in such a way that dual attainment holds on each $\Gamma \cap (C(x) \times \mathbb{R}^d)$ where $C(x)$ is any component in the convex paving Φ. In the last part, we will use this convex paving to perform the above mentioned decomposition of π.

2.4.1 Irreducible convex pavings associated to martingale supporting sets

Let Γ be a Borel set in S_{MT}. We start by defining an equivalence relation on X_Γ. For each $x \in X := X_\Gamma$, we define inductively an increasing sequence of convex open sets $(C_n(x))_n$ in the following way:

Start with the trivial equivalence relation $x \sim_0 x'$ iff $x = x'$. Let $C_0(x) := IC(\Gamma_x)$ and recall that if $\Gamma_x = \{x\}$, then $C_0(x) = \{x\}$. Now define the following equivalence relation on X: $x \sim_1 x'$ if there exist finitely many $x_1, ...x_k$ in X such that the following chain condition holds:

$$C_0(x) \cap C_0(x_1) \neq \emptyset,$$
$$C_0(x_i) \cap C_0(x_{i+1}) \neq \emptyset \ \forall i = 1, 2, ...k-1,$$
$$C_0(x_k) \cap C_0(x') \neq \emptyset.$$

We then consider the open convex hull:

$$C_1(x) := IC[\bigcup_{x' \sim_1 x} C_0(x')].$$

Note that $x \sim_1 x'$ implies $C_1(x) = C_1(x')$. Unfortunately, the convex sets $C_1(x)$ do not determine the equivalence classes. In particular, they may not be mutually disjoint for elements that are not equivalent for $\sim_1$. So, we proceed to define $\sim_2$ in a similar way: $x \sim_2 x'$ if there exists finitely many $x_1, ...x_k$ in X such that the following chain condition holds:

$$C_1(x) \cap C_1(x_1) \neq \emptyset,$$
$$C_1(x_i) \cap C_1(x_{i+1}) \neq \emptyset \ \forall i = 1, 2, ...k-1,$$
$$C_1(x_k) \cap C_1(x') \neq \emptyset;$$

and we set

$$C_2(x) := IC[\bigcup_{x' \sim_2 x} C_1(x')].$$

Again, $\sim_2$ is an equivalence relation and one can easily see that

- $x \sim_1 x' \Rightarrow x \sim_2 x'$
- $x \sim_2 x' \Rightarrow C_2(x) = C_2(x')$
- $C_1(x) \subseteq C_2(x)$.

But still, the sets $C_2(x)$ may not be mutually disjoint for non-equivalent $x's$. We continue inductively in a similar fashion by defining equivalence relations $\sim_n$ for $n = 1, 2, ...$ and their corresponding classes

$$C_n(x) := IC[\bigcup_{x' \sim_n x} C_{n-1}(x')].$$

It is easy to check that we have the following properties for each n,

$$\begin{aligned} x \sim_n x' &\Rightarrow x \sim_{n+1} x' \\ x \sim_n x' &\Rightarrow C_n(x) = C_n(x') \\ C_n(x) &\subseteq C_{n+1}(x). \end{aligned}$$

Finally, define the equivalence relation

$$x \sim x' \text{ if } x \sim_n x' \text{ for some } n,$$

and its corresponding convex sets

$$C(x) := \lim_{n\to\infty} C_n(x) = \cup_{n=0}^{\infty} C_n(x). \tag{2.4.1}$$

Now, we show that $\Psi = \{C(x)\}_{x\in X}$ is an irreducible convex paving for Γ.

Theorem 2.4.1. *The canonical relation $\sim$ on X_Γ and the components $(C(x))_{x\in X_\Gamma}$ satisfy the following:*

1. $x \sim x' \Rightarrow C(x) = C(x')$, *and* $x \nsim x' \Rightarrow C(x) \cap C(x') = \emptyset$.

2. $C(x)$ *are mutually disjoint, that is either* $C(x) = C(x')$ *or* $C(x) \cap C(x') = \emptyset$.

3. $x' \in X \cap C(x)$ *if and only if* $x' \sim x$.

4. $\Phi = \{C(x)\}_{x\in X}$ *is an irreducible convex paving for* Γ.

5. $C_n(x) = IC[\bigcup_{x' \sim_n x} \Gamma_{x'}]$ *for* $n \geq 0$ *and* $C(x) = IC[\bigcup_{x' \sim x} \Gamma_{x'}]$.

Proof. The fact that $x \sim_n x' \Rightarrow C_n(x) = C_n(x')$ gives the first part of (1). If there exists a $z \in C(x) \cap C(x')$, then there is N such that $z \in C_N(x) \cap C_N(x')$, implying $x \sim_{N+1} x'$ and verifying the second part of (1) of which (2) and (3) are obvious consequences.

To prove (4), let Ψ be any convex paving of Γ and let $z, x \in X_\Gamma$, $D \in \Psi$ be such that $C(z) \cap D(x) \neq \emptyset$. We must show that $C(z) \subseteq D(x)$. We claim that, for any $n \geq 0$,

$$(*) \qquad C_n(z) \cap D(x) \neq \emptyset \Rightarrow C_n(z) \subseteq D(x), \text{ for every } z, x \in X_\Gamma.$$

Indeed, it is true for $n = 0$ by definition. Assume that $(*)$ is true for some n, and suppose $C_{n+1}(z) \cap D(x) \neq \emptyset$. Note that $C_n(z) \subseteq D(z)$, and so if $w \sim_{n+1} z$, by $(*)$ we have that $C_n(w) \subseteq D(z)$. As $C_{n+1}(z) = IC[\bigcup_{w \sim_{n+1} z} C_n(w)]$, this readily implies that $C_{n+1}(z) \subseteq D(z)$, but then $D(z) \cap D(x) \neq \emptyset$ and hence $D(z) = D(x)$. This proves $(*)$ for every $n \geq 0$. Now if $C(z) \cap D(x) \neq \emptyset$, then for all large n $C_n(z) \cap D(x) \neq \emptyset$, hence by $(*)$ we get that $C_n(z) \subseteq D(x)$. Therefore $C(z) \subseteq D(x)$ which proves the irreducibility of Φ.

For (5), let $(A_i)_{i \in I}$ be any family of sets in $\mathbb{R}^d$, where I is an index set. Then it is easy to see that

$$IC(A_i) = IC(CC(A_i)), \text{ and}$$
$$CC(\bigcup_{i \in I} CC(A_i)) = CC(\bigcup_{i \in I} A_i) = CC(\bigcup_{i \in I} IC(A_i)).$$

But note that $A \subseteq B$ does not imply $IC(A) \subseteq IC(B)$ in general. The above implies in particular

$$IC(\bigcup_{i \in I} A_i) = IC(\bigcup_{i \in I} IC(A_i)).$$

In addition, a simple induction shows that for every $n \geq 0$, we have

$$C_n(x) = IC[\bigcup_{x' \sim_n x} \Gamma_{x'}].$$

Indeed, it is true for $n = 0$ by definition. Suppose $C_n(x) = IC[\bigcup_{x' \sim_n x} \Gamma_{x'}]$. Now by definition,

$$C_{n+1}(x) = IC[\bigcup_{x' \sim_{n+1} x} C_n(x')] = IC[\bigcup_{x' \sim_{n+1} x} IC(\bigcup_{x'' \sim_n x'} \Gamma_{x''})] = IC[\bigcup_{x' \sim_{n+1} x} (\bigcup_{x'' \sim_n x'} \Gamma_{x''})].$$

But $\bigcup_{x' \sim_{n+1} x} \bigcup_{x'' \sim_n x'} \Gamma_{x''} = \bigcup_{x' \sim_{n+1} x} \Gamma_{x'}$, hence, $C_{n+1}(x) = IC(\bigcup_{x' \sim_{n+1} x} \Gamma_{x'})$, completing the induction.

Finally, we proceed as follows:

$$\begin{aligned} C(x) = IC[C(x)] = IC[\bigcup_{n \geq 0} C_n(x)] &= IC[\bigcup_{n \geq 0} IC(\bigcup_{x' \sim_n x} \Gamma_{x'})] \\ = IC[\bigcup_{n \geq 0} \bigcup_{x' \sim_n x} \Gamma_{x'}] = IC[\bigcup_{x' \sim x} \Gamma_{x'}], \end{aligned}$$

which completes the proof of (5) and the theorem. □

2.4.2 Duality attainment on each irreducible component

Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function on which we make no assumption. Our aim is to prove Theorem 2.2.11, which will be obtained once we prove the following lemma.

Lemma 2.4.2. *Let $\Gamma \in S_{MT}$ and denote $X := X_\Gamma$. Fix $x_0 \in X$ and set $G := \Gamma \cap (C(x_0) \times \mathbb{R}^d)$, where $C(x_0)$ is the convex set corresponding to the equivalence class of x_0 given in Section 2.4.1. Then, for each $y \in Y_G$, there exists a compact interval $K_y \subseteq \mathbb{R}$ such that any finite subset $H \subseteq G$ has a c-dual triplet (α, β, γ) with $\beta(y) \in K_y$ for all $y \in Y_H$.*

Note that the above lemma states that there is a uniform estimate on the way finite subsets of G have c-duals. This control on the β component of the c-dual triplets will allow us to use Tychonoff's compactness theorem to deduce that G has a c-dual.

To prove Lemma 2.4.2, we first give an idea about the degrees of freedom we have in choosing β. First, note that if β is a c-dual for G and $L : \mathbb{R}^d \to \mathbb{R}$ is an affine function, then $\beta - L$ is also a c-dual for G. Letting $m = \dim(V(Y_G))$, we can find $\{y_0, ..., y_m\} \subseteq Y_G$ such that $V(\{y_0, ..., y_m\}) = V(Y_G)$, i.e. $\{y_0, ..., y_m\}$ constitute vertices of an m-dimensional polytope in $V(Y_G)$. Now for a given c-dual function β for G, let $L : V(Y_G) \to \mathbb{R}$ be an affine function determined by $L(y_i) = \beta(y_i)$ for $i = 0, 1, ..., m$. The function $\beta' := \beta - L$ then satisfies $\beta'(y_i) = 0$ for all $i = 0, 1, ..., m$, which means that we have $m+1$ freedom of choice on the value of β. In other words, if we set $K_{y_i} = \{0\}$ for $i = 0, 1, ..., m$, then we can find β' such that $\beta'(y_i) \in K_{y_i}$ for each y_i. Now, we want to observe how the initial value of β can control its values at other points y. We shall see that thanks to the duality relations (2.2.2) and (2.2.3), the control of the value of β propagates well along a given chain inside the equivalent class $C(x_0)$.

The proof of Lemma 2.4.2 is involved, and requires several key steps. To clarify the idea, we consider first the special case $c = 0$ where we can establish a complete control on the dual functions.

Lemma 2.4.3. *Let $G \in S_{MT}$ and assume that (α, β, γ) is a 0-dual triplet associated to the set G. In other words,*

$$\beta(y) \geq L_x(y) \quad \forall x \in X_G, y \in Y_G \tag{2.4.2}$$

$$\beta(y) = L_x(y) \quad \forall (x, y) \in G, \tag{2.4.3}$$

where for each x, L_x is the affine function

$$L_x(y) := \gamma(x) \cdot (y - x) + \alpha(x).$$

Then, $L_x = L_{x'}$ on $V(C(x))$ whenever $x \sim x'$.

Note that (2.4.3) says that if we have control on L_x, then we have control on β for all $y \in G_x$. In particular, Lemma 2.4.3 implies that if $L_x = 0$ (we can choose such L_x without loss of generality) then, $L_{x'} = 0$ on $V(C(x))$ for all $x' \in C(x)$, thus $\alpha(x') = 0$ for all $x' \in C(x)$ and $\beta(y) = 0$ at each $y \in G_{x'}$.

The above lemma is a consequence of the following proposition:

Proposition 2.4.4. *Let L_1, L_2 be two affine functions on $\mathbb{R}^d$, and let S_1, S_2 be sets in $\mathbb{R}^d$. Suppose that $L_1 \leq L_2$ on S_1, and $L_2 \leq L_1$ on S_2, and that $IC(S_1) \cap IC(S_2) \neq \emptyset$.*

Then, $L_1 = L_2$ on $V(S_1 \cup S_2)$, the latter is the minimal affine space containing the sets S_1 and S_2.

Proof. This follows from two facts:

1. For affine functions, $L \leq L'$ on a set S implies $L_1 \leq L_2$ on $\mathrm{conv}(S)$.
2. If two affine functions L, L' satisfy $L \leq L'$ on a set S and if moreover, $L(z) = L'(z)$ at some interior point of $\mathrm{conv}(S)$, then $L = L'$ on $\mathrm{conv}(S)$, thus on $V(S)$.

Indeed, apply (1) to the case $L = L_1$, $L' = L_2$, and $S = S_1$, and also to the case $L = L_2$, $L' = L_1$, and $S = S_2$. We get $L_1 = L_2$ on $\mathrm{conv}(S_1) \cap \mathrm{conv}(S_2)$. Now, from the assumption, $IC(S_1) \cap IC(S_2) \neq \emptyset$, and also obviously $IC(S_1) \cap IC(S_2) \subseteq IC(S_i)$, $i = 1, 2$. Using (2) we then get that $L_1 = L_2$ on both $\mathrm{conv}(S_i)$, $i = 1, 2$. From this the assertion follows. □

Proof of Lemma 2.4.3. First note that for each $x, x' \in X = X_G$, conditions (2.4.2) and (2.4.3) yield that $L_{x'} \leq L_x$ on G_x, and $L_x \leq L_{x'}$ on $G_{x'}$.

Now to prove the lemma, it suffices to show that for each $n \in \{0, 1, 2, ..., \}$,

$$\text{if } x \sim_n x' \text{ then } L_x = L_{x'} \text{ on } V(C_n(x)). \tag{2.4.4}$$

Here, by $x \sim_0 x'$ we mean $x = x'$. We do this inductively. Our induction hypothesis is (2.4.4) together with

$$L_z \leq L_x \quad \text{on } C_n(x), \text{ for each } z \in X \text{ and } x \in X. \tag{2.4.5}$$

For $n = 0$, (2.4.4) is trivially satisfied and (2.4.5) follows from (2.4.2) and (2.4.3). Now, assume that (2.4.4) and (2.4.5) hold for all $n \leq k$. For $n = k + 1$, if $x \sim_{k+1} x'$ then there are $x = x_0, x_1, ..., x_m = x'$ for some m, such that $C_k(x_i) \cap C_k(x_{i+1}) \neq \emptyset$ for each $0 \leq i \leq m - 1$. From this, and using (2.4.4) and (2.4.5), we can apply Proposition 2.4.4 with the choice $L_1 = L_{x_i}, L_2 = L_{x_{i+1}}$, $S_1 = C_k(x_i)$, $S_2 = C_k(x_{i+1})$, and see $L_{x_i} = L_{x_{i+1}}$ on $V(C_k(x_i) \cup C_k(x_{i+1}))$, for $i = 0, ..., m-1$. Similarly, repeated application of Proposition 2.4.4 eventually yields that $L_x = L_{x_i} = L_{x'}$ on each $C_k(x_i)$, for $i = 1, .., m$. Therefore, $L_x = L_{x'}$ on $\bigcup_i C_k(x_i)$, thus, on $V(\bigcup_{i=0}^m C_k(x_i))$. This holds for any $x \sim_{k+1} x'$, thus by applying the result to all $z \sim_{k+1} x \sim_{k+1} x'$, we also see

$$L_x = L_{x'} \text{ on } V(\bigcup_{z \sim_{k+1} x} C_k(z)) = V(C_{k+1}(x)),$$

verifying (2.4.4) for $n = k + 1$. For (2.4.5), for each $z \in X$, from the assumption (2.4.5) for $n \leq k$ and applying (2.4.4), we have $L_z \leq L_{x'} = L_x$ on $C_k(x')$ for all $x' \sim_{k+1} x$. For the affine functions, this implies $L_z \leq L_x$ on $C_{k+1}(x)$. This completes the induction argument, so the proof. □

We now consider the case of a non-trivial cost c. We first establish a more quantitative version of Proposition 2.4.4.

Proposition 2.4.5. *Let L_1, L_2 be two affine functions on $\mathbb{R}^d$, and let S_1, S_2 be sets in $\mathbb{R}^d$. Suppose that*

- *$L_1 \leq L_2 + \delta_1$ on S_1, and $L_2 \leq L_1 + \delta_2$ on S_2 for some constants $\delta_1, \delta_2 > 0$;*
- *there is a point z in $IC(S_1) \cap IC(S_2)$.*

Then, $|L_1 - L_2| \leq C$ on $\text{conv}(S_1 \cup S_2)$. Here, $C = C(z, S_1, S_2, \delta_1, \delta_2) < \infty$ as long as z stays in the interior $IC(S_1) \cap IC(S_2)$, though as z gets close to the boundaries $\partial(\text{conv}(S_i))$, $i = 1$ or 2, the constant C may go to $+\infty$.

Proof. First, convexity and linearity imply that for each $\delta, \delta' > 0$ we have the following:

1. For affine functions, $L \leq L' + \delta$ on a set S implies $L \leq L' + \delta$ on $\text{conv}(S)$.
2. If two affine functions L, L' satisfy $L \leq L' + \delta$ on a set S and if moreover, $L(z) \geq L'(z) - \delta'$ at some interior point z of $\text{conv}(S)$, then $|L - L'| \leq C = C(z, S, \delta, \delta')$ on $\text{conv}(S)$. Here, the constant $C < \infty$ depends only on δ, δ' and the ratio between the minimum distance from z to $\partial(\text{conv}(S))$ and the maximum distance to $\partial(\text{conv}(S))$, though, as z gets close to $\partial(\text{conv}(S))$, the constant C can go to $+\infty$.

Now, apply (1) to the case $L = L_1$, $L' = L_2$, and $S = S_1$, and also to the case $L = L_2$, $L' = L_1$, and $S = S_2$. Thus, we get $|L_1 - L_2| \leq \max(\delta_1, \delta_2)$ at the point z of $IC(S_1) \cap IC(S_2)$. Now, apply (2), to get $|L_1 - L_2| \leq C$ on both $\mathrm{conv}(S_i)$, $i = 1, 2$, where $C = C(S_1, S_2, \delta_1, \delta_2) < \infty$. Applying (1) again, we have $|L_1 - L_2| \leq C$ on $\mathrm{conv}(S_1 \cup S_2)$, completing the proof. □

We now introduce the following notation.

Definition 2.4.6. *Let $G \in S_{MT}$ and let $H \subseteq G$. Given any triplet $\{\alpha, \beta, \gamma\}$ that forms a c-dual for H, we define for each $x \in X_H$, the affine function*

$$L_x^H(y) = \gamma(x) \cdot (y - x) + \alpha(x).$$

The superscript H indicates that L_x^H arises from a c-dual of H. The c-duality conditions satisfied by α, β, γ on H can be written as:

$$\beta(y) - c(x, y) \geq L_x^H(y), \; \forall x \in X_H, y \in Y_H, \tag{2.4.6}$$

$$\beta(y) - c(x, y) = L_x^H(y), \; \forall (x, y) \in H. \tag{2.4.7}$$

For an affine space V, we write for $x, x' \in X_H$,

$$L_x^H \approx L_{x'}^H \quad \text{on } V$$

if there is a bounded set S with $V = V(S)$ and a constant $M = M(c, H, S)$ depending only on H, the cost function c and the set S, such that for every choice of c-dual $\{\alpha, \beta, \gamma\}$ for H, we have

$$|L_x^H - L_{x'}^H| \leq M \quad \text{on the set } S. \tag{2.4.8}$$

We say $L_x^H \approx L_{x'}^H$ at z, if we have (2.4.8) for $S = \{z\}$.

An immediate observation is that for $x, x', x'' \in X_H$,

$$\text{whenever } L_x^H \approx L_{x'}^H \text{ and } L_{x'}^H \approx L_{x''}^H \text{ on } V, \text{ then } L_x^H \approx L_{x''}^H \text{ on } V.$$

Also, note that if $H' \subseteq H$, any c-dual for H is also a c-dual for H', hence for any $x, x' \in X_{H'}$, we necessarily have

$$L_x^{H'} \approx L_{x'}^{H'} \quad \text{on } V \Rightarrow L_x^H \approx L_{x'}^H \quad \text{on } V. \tag{2.4.9}$$

We shall now prove the analogue of Lemma 2.4.3 in the case of a general cost. We shall again use Proposition 2.4.5, to establish a propagation of control on the affine functions L_x^H's, along an ordered chain of intersecting convex open sets. But, since c is not trivial anymore, the control on L_x^H can be done only in finite steps, since the errors (the constant C in Proposition 2.4.5) can accumulate.

Lemma 2.4.7. *Set $G := \Gamma \cap (C(x_0) \times \mathbb{R}^d)$ and suppose $x, x' \in X_G$ (i.e. $x \sim x'$). Then there exists a finite set $H \subseteq G$ such that $x, x' \in X_H$ and $L_x^H \approx L_{x'}^H$ on $V(C(x))$.*

Proof. First observe that it suffices to prove

Claim 2.4.8. *Suppose $x \sim x'$ and $z \in G_{x'}$. Then there exists a finite set $H \subseteq G$ such that $x, x' \in X_H$, and $L_x^H \approx L_{x'}^H$ at z.*

The lemma follows when we apply this claim to a set of finitely many (u_i, v_i)'s, $i = 1, ..., m$, in G (so $x \sim x' \sim u_i$) with $V(C(x)) = V(\{v_i\}_{i=1}^m)$, and use (2.4.9).

We show this claim using induction on $n = 0, 1, 2, 3, ...$ Our induction hypothesis is if $x \sim_n x'$ and $z \in G_{x'}$, then there exists a finite set $H \subseteq G$ such that

(i1) $x, x' \in X_H$, $z \in Y_H$ and $Y_H \subseteq \overline{C_n(x)}$;

(i2) $L_x^H \approx L_{x'}^H$ on $V(Y_H)$;

(i3) for each finite set $F \subseteq G$ with $H \subseteq F$, and for $w \in X_F$, there is a constant $C = C(H, w)$ depending only on H and w such that

$$L_w^F \leq L_x^F + C \quad \text{on } Y_H.$$

Notice that the claim follows if we verify (i1) – (i3) for each n, since in particular, the values of L_x^H and $L_{x'}^H$ at z is estimated from (i2).

Let us start the induction. For the case $n = 0$, assume $x \sim_0 x'$ (i.e. $x = x'$) and $z \in G_{x'}$. Choose $H = \{(x, z)\}$. Then, (i1) and (i2) are trivially satisfied. Moreover, (i3) holds from (2.4.6) and (2.4.7), where the constant C is estimated by the value $c(w, z) - c(x, z)$. This completes the case $n = 0$.

Suppose the induction hypothesis holds for n. Assume $x \sim_{n+1} x'$ and $z \in G_{x'}$. Then, there is a finite chain of $C_n(x_k)$, $k = 0, 1, ..., m$, $x_0 = x$, $x_m = x'$, such that $C_n(x_k) \cap C_n(x_{k+1}) \neq \emptyset$ for each k. Recall $C_n(x) = IC[\bigcup_{x' \sim_n x} G_{x'}]$. Thus for each $C_n(x_k)$, it is possible to find a finite set $J_k := \{(u_k^i, v_k^i)_{i=1}^{m_k}\} \subseteq G$ such that $x_k \sim_n u_k^i$ for all i and $IC(Y_{J_k})$ is a good approximation of $C_n(x_k)$, i.e. $Y_{J_k} \subseteq \overline{C_n(x_k)} \subseteq V(Y_{J_k})$ and $IC(Y_{J_k}) \cap IC(Y_{J_{k+1}}) \neq \emptyset$. Also, we can let $z \in Y_{J_m}$.

Now, apply the induction hypothesis for n to each x_k, u_k^i, $x_k \sim_n u_k^i$ and find a finite set $H_k^i \subseteq G$ that satisfies (i1)–(i3) for $x = x_k$, $x' = u_k^i$, $z = v_k^i$, and $H = H_k^i$. Let

$$H_k := \bigcup_i H_k^i$$

Then, from (i3) for H_k^i's, we also have (i3) for $H = H_k$ and $x = x_k$. Here, the point of considering H_k is $Y_{J_k} \subseteq Y_{H_k} \subseteq \overline{C_n(x_k)}$, so $V(Y_{J_k}) = V(Y_{H_k})$, hence $IC(Y_{H_k}) \cap IC(Y_{H_{k+1}}) \neq \emptyset$ as well.

Now, to verify the induction hypothesis for $n+1$-th step, let

$$\bar{H} := \bigcup_k H_k$$

We will show properties (i1)–(i3) for this set $\bar{H}$. From the construction, $x, x' \in X_{\bar{H}}$, $z \in Y_{\bar{H}}$, and since $C_n(x_k) \subseteq C_{n+1}(x_k) = C_{n+1}(x)$, (i1) readily follows. For (i2), apply the induction hypothesis (i3) for H_k's to Proposition 2.4.5 iteratively for the pairs Y_{H_1} and Y_{H_2}, $Y_{H_1} \cup Y_{H_2}$ and Y_{H_3}, ... , $Y_{H_1} \cup .... \cup Y_{H_k}$ and $Y_{H_{k+1}}$, so on. Then we see the estimate (2.4.8) holds for $S = Y_{\bar{H}}$, thus,

$$L_x^{\bar{H}} \approx L_{x_1}^{\bar{H}} \approx ... \approx L_{x_{m-1}}^{\bar{H}} \approx L_{x'}^{\bar{H}} \quad \text{on } V(Y_{\bar{H}}), \tag{2.4.10}$$

verifying (i2).

For (i3), let F be a finite set containing $\bar{H}$ and let $w \in X_F$. Then (i3) for each H_k gives that $L_w^F \le L_{x_k}^F + C_k$ on Y_{H_k}, $k = 0, 1, ..., m$. Now applying (2.4.10) and recalling (2.4.9), we conclude that there is a constant $C = C(\bar{H}, w)$ such that

$$L_w^F \le L_x^F + C \quad \text{on} \quad Y_{\bar{H}}.$$

This completes the induction, and the proof. □

Armed with Lemma 2.4.7, we shall establish Lemma 2.4.2.

Proof of Lemma 2.4.2. Recall that we fix $x_0 \in X$ and let $G := \Gamma \cap (C(x_0) \times \mathbb{R}^d)$ and $V := V(Y_G)$.

Let $m = dim(V)$. Then we can find

$$J := \{(u_i, v_i)\}_{i=0}^m \subseteq G \ \text{ such that } \ V(\{v_i\}_{i=0}^m) = V.$$

Define the initial choices $K_{v_i} = \{0\}, i = 0, 1, ..., m$. We want to define the K_y's to be compatible with these initial choices. For $y \in Y_G$, choose $x(y) \in X_G$ such that $(x(y), y) \in G$. By lemma 2.4.7 (especially see Claim 2.4.8), for $y \in Y_G$, we can choose a finite set $H(y)$ such that $J \cup \{(x(y), y)\} \subseteq H(y)$ and

$$L_{x(y)}^{H(y)} \approx L_{u_i}^{H(y)} \text{ at } v_i, \forall i = 0, ..., m.$$

In particular, there exists a constant M, depending only on y and $H(y)$ but not on the choice of dual functions of $H(y)$, such that

$$|L_{x(y)}^{H(y)}(v_i) - L_{u_i}^{H(y)}(v_i)| \le M, \ \forall\, i = 0, ..., m. \tag{2.4.11}$$

Let (α, β, γ) be a c-dual triplet for the set $H(y)$; note that it exists because Γ, thus G is finitely c-dualizable. By subtracting an appropriate affine function from β, we can let $\beta(v_i) = 0$. Duality then gives

$$\beta(y) = c(x(y), y) + L_{x(y)}^{H(y)}(y).$$

Since $L_x^{H(y)}$ is affine and $V(\{v_i\}_{i=0}^m) = V$, the value $L_{x(y)}^{H(y)}(y)$ can be computed from the values $L_{x(y)}^{H(y)}(v_i)$. Hence by (2.4.11), the values $L_{u_i}^{H(y)}(v_i)$ give an estimate of $\beta(y)$. Notice that from duality we have

$$L_{u_i}^{H(y)}(v_i) = \beta(v_i) - c(u_i, v_i) = -c(u_i, v_i).$$

Thus, there exists a constant $N = N(y)$ such that if β is a c-dual for $H(y)$ and if $\beta(v_i) = 0$ for all i, then $-N \leq \beta(y) \leq N$. We set $K_y = [-N, N]$.

To get the claim in Lemma 2.4.2, we let H be any finite set and denote $Y_H = \{y_1, ..., y_s\}$. Let

$$H^* = H \cup H(y_1) \cup ... \cup H(y_s)$$

Now choose β to be a c-dual for H^* with $\beta(v_i) = 0$ for all i. Since β is also a c-dual for $H(y_j)$, we have $\beta(y_j) \in K_{y_j}$ for all $j = 1, ..., s$. Finally, note that β is also a c-dual for H, concluding the proof. □

2.4.3 Proof of Theorem 2.2.11

As before, let $G = \Gamma \cap (C(x_0) \times \mathbb{R}^d)$ and let $V := V(Y_G)$ be the ambient space. We first find the desired function $\beta : Y_G \to \mathbb{R}$ from the compactness argument already used in [5]. Indeed, define $K := \Pi_{y \in Y_G} K_y$, where the K_y's were obtained in Lemma 2.4.2. This is a subset of the space of all functions from Y_G to $\mathbb{R}$. In the topology of pointwise convergence, K is compact by Tychonoff's theorem. Now we claim that, for any finite $H \subseteq G$, the set

$$\Psi_H := \{\beta \in K : \beta \text{ is a } c\text{-dual for } H\}$$

is a non-empty closed subset of K. Indeed, that Ψ_H is non-empty follows from Lemma 2.4.2 since Γ and hence G is finitely c-dualizable: if necessary, one can extend the β found in Lemma 2.4.2 –and originally defined on Y_H– to Y_G, by simply letting $\beta(y) = 0$ for $y \notin Y_H$.

To show that Ψ_H is closed, let $\{\beta_n\}$ be a sequence of c-dual functions for H, and suppose $\beta_n \to \beta$ pointwise on Y_G. We need to show that β is also a c-dual for H. But for each n, we have functions (α_n, γ_n) such that the duality relation holds:

$$\beta_n(y) - c(x,y) \geq \gamma_n(x) \cdot (y - x) + \alpha_n(x)\ \forall x \in X_H, y \in Y_H \tag{2.4.12}$$

$$\beta_n(y) - c(x,y) = \gamma_n(x) \cdot (y - x) + \alpha_n(x)\ \forall (x,y) \in H. \tag{2.4.13}$$

Here, without loss of generality, we can assume that each vector $\gamma_n(x)$ is parallel to $V(Y_H)$. Now since $(\beta_n(y) - c(x,y))_{x \in X_H, y \in Y_H}$ is uniformly bounded in n, we can choose $(\alpha_n(x), \gamma_n(x))$ in such a way that $(\alpha_n(x), \gamma_n(x))_n$ is also uniformly bounded in n. Since X_H is finite, we can find a subsequence of (α_n, γ_n) which converges to (α, γ) at every $x \in X_H$. Then (α, β, γ) is clearly a c-dual triplet for H, establishing the claim on Ψ_H.

Now for any $T \subseteq S$, any c-dual for S is also a c-dual for T. We therefore have the finite intersection property for $\{\Psi_H\}$, i.e. $\emptyset \neq \Psi_{H_1 \cup \ldots \cup H_s} \subseteq \bigcap_{j=1,\ldots,s} \Psi_{H_j}$. By the compactness of K and the closeness of Ψ_H's, we deduce that the set $\Psi_G := \bigcap_{H \subseteq G, |H| < \infty} \Psi_H$ is nonempty.

We claim that any $\beta \in \Psi_G$ is a c-dual for G. Fix $x \in X_G$ and $\beta \in \Psi_G$. We must show that there exists an affine function L_x on $V = V(Y_G)$ such that the duality holds:

$$\beta(y) - c(x,y) \geq L_x(y),\ \forall y \in Y_G \tag{2.4.14}$$

$$\beta(y) - c(x,y) = L_x(y),\ \forall y \in G_x. \tag{2.4.15}$$

Choose a finite set $H_x \subseteq G_x$ such that $V(H_x) = V(G_x)$. Observe that for any finite set F containing $H := \{x\} \times H_x$,

$$L_x^F(y) = \beta(y) - c(x,y) = L_x^H(y)\ \forall y \in H_x, \text{ hence } L_x^F(y) = L_x^H(y)\ \forall y \in V(G_x).$$

In particular, $L_x^F(x) = L_x^H(x)$ since $x \in IC(G_x) \subseteq V(G_x)$. Let us define $\alpha(x) = L_x^H(x)$.

Now we need to construct the last piece which is $\gamma(x)$. For this, in addition to H, we also choose a finite set $\{(v_i, w_i)\}_{i=1}^m \subseteq G$ such that $x \in IC(\{w_i\}_{i=1}^m)$ and $V(\{w_i\}_{i=1}^m) = V$, and define

$$\bar{H} := H \cup \{(v_i, w_i)\}_{i=1}^m.$$

For any finite set $F \subseteq G$ with $\bar{H} \subseteq F$, according to the duality, define the set

$$\gamma_F(x) := \{v \in V : \beta(y) - c(x,y) - \alpha(x) \geq v \cdot (y - x),\ \forall y \in Y_F \tag{2.4.16}$$

$$\beta(y) - c(x,y) - \alpha(x) = v \cdot (y - x),\ \forall y \in F_x\} \tag{2.4.17}$$

The set $\gamma_F(x)$ is nonempty because of finite dualizable property of G. Now, since $x \in IC(\{w_i\}_{i=1}^m)$ and $V(\{w_i\}_{i=1}^m) = V$, we deduce from (2.4.16) that $\gamma_F(x)$ is a closed and bounded set in V, hence compact. Again since every c-dual for S is also a c-dual for T, whenever $T \subseteq S$, we have that $\{\gamma_F(x) : F \supseteq \bar{H}\}$ has the finite intersection property. Hence, we can choose a

$$\gamma(x) \in \bigcap_{F \supseteq \bar{H}, |F| < \infty} \gamma_F(x).$$

Finally, we show that (2.4.14) and (2.4.15) hold for this choice of $(\alpha(x), \gamma(x))$. Indeed, let $(x', y') \in G$, and let $F = \bar{H} \cup \{(x', y')\}$. By (2.4.16), we have $\beta(y') - c(x, y') \geq \alpha(x) + \gamma(x) \cdot (y' - x)$, so (2.4.14) holds. Let $y \in G_x$. Let $F = \bar{H} \cup \{(x, y)\}$. By (2.4.17), we have $\beta(y) - c(x,y) = \alpha(x) + \gamma(x) \cdot (y - x)$, so (2.4.15) holds. This completes the proof of Theorem 2.2.11. □

2.4.4 The disintegration of a martingale transport plan

For a closed convex set $U \subseteq \mathbb{R}^d$, let $\mathcal{K}(U)$ be the space of all closed convex subsets in $\mathbb{R}^d$, equipped with the Hausdorff metric in such a way that it becomes a separable complete metric space (Polish space): This allows for the disintegration of a measure π on X via a measurable map $T : X \to \mathcal{K}(U)$ (see e.g. [7, Corollary 2.4]) in such a way that each piece of the disintegrated measure, say π_C, is a probability measure on $T^{-1}(C)$. In particular, $\pi_C(T^{-1}(C)) = 1$ for $T_\#\pi$-a.e. $C \in \mathcal{K}(U)$, ultimately yielding conditional probabilities.

Proof of Theorem 2.2.14. The above discussion and the measurability hypothesis of the map $\Xi : \Gamma \to \mathcal{K}(\mathbb{R}^d)$ defined by $(x, y) \mapsto \overline{C(x)}$, yield the disintegration of π into $\pi_C d\tilde{\pi}(C)$ in (2.2.4), with π_C supported on Γ_C. The measures μ_C, ν_C are obtained by taking marginals of π_C. The martingale and optimality properties of π_C for $\tilde{\pi}$-a.e. C, follow from those properties of π and the disintegration (2.2.4). For the optimal martingale transport π, the set Γ_C admits a c-dual by Theorem 2.2.11 applied to the finitely c-dualizable Γ. This deals with items 1), 2), and 3) of the theorem. Finally, 4) follows immediately from Theorem 2.2.4. □

2.5 Structural results for general optimal martingale transport plans

One would like to know when we can disintegrate μ into absolutely continuous pieces μ_C, so as to apply theorem 2.2.4 on each partition. We start by a counterexample showing that this is not possible in general, at least in dimension $d \geq 3$. We then proceed in Section 2.5.2 to show Theorem 2.2.15 and 2.2.16, which give instances where the reasoning could apply.

2.5.1 Nikodym sets and martingale transports

Ambrosio, Kirchheim, and Pratelli [2] constructed a Nikodym set in $\mathbb{R}^3$ having full measure in the unit cube, and intersecting each element of a family of pairwise disjoint open lines only in one point. More precisely they showed the following:

Theorem 2.5.1. (Ambrosio, Kirchheim, and Pratelli [2]) *There exist a Borel set* $M_N \subseteq [-1,1]^3$ *with* $|[-1,1]^3 - M_N| = 0$ *and a Borel map* $f = (f_1, f_2) : M_N \to [-2,2]^2 \times [-2,2]^2$ *such that the following holds. If we define for* $x \in M_N$ *the open segment* l_x *connecting* $(f_1(x), -2)$ *to* $(f_2(x), 2)$*, then*

- $\{x\} = l_x \cap M_N$ *for all* $x \in M_N$*,*
- $l_x \cap l_y = \emptyset$ *for all* $x \neq y \in M_N$*.*

Example 2.5.2. One can use the above construction to construct an optimal martingale transport, whose equivalence classes are singletons, hence the disintegration of the first marginal along the partitions $C(x)$ is the Dirac mass δ_x, which is obviously not absolutely continuous w.r.t. $\mathcal{L}^1$.

Consider the obvious inequality $\frac{1}{2\varepsilon}(|x-y|-\varepsilon)^2 \geq 0$, and its equivalent form

$$\frac{1}{2\varepsilon}|y|^2 \geq |x-y| + \frac{1}{\varepsilon}x \cdot (y-x) + \frac{1}{2\varepsilon}|x|^2 - \varepsilon. \tag{2.5.1}$$

Thus by letting $\alpha_\varepsilon(x) = \frac{1}{2\varepsilon}|x|^2 - \varepsilon$, $\beta_\varepsilon(y) = \frac{1}{2\varepsilon}|y|^2$ and $\gamma_\varepsilon(x) = \frac{1}{\varepsilon}x$, (2.5.1) is a dual relation for the maximization problem, and (2.5.1) shows that every martingale $\pi_\varepsilon := (X, Y)$ with $|X - Y| = \varepsilon$ *a.s.* is optimal with its own marginals $X \sim \mu$ and $Y \sim \nu$.

Now fix $\varepsilon > 0$ small and let X be a random variable whose distribution μ has uniform density on $[-1,1]^3$. We define Y conditionally on X by evenly distributing the mass along the lines l_x considered in Theorem 2.5.1 and distance ε, that is Y splits equally in two pieces from $x \in X$ along l_x with distance ε. Then the martingale (X, Y) is optimal for the maximization problem. But note that in this case, each equivalence class $[x]$ is the singleton $\{x\}$, so the disintegration of μ along the partitions $C(x)$ is the Dirac mass δ_x, which is obviously not absolutely continuous w.r.t. $\mathcal{L}^1$. Hence, the decomposition is not useful in this case. One also notices that the convex sets associated to the irreducible paving of the martingale (X, Y) have codimension 2. We leave it as an open problem whether one can do without assumption (2.2.5) in Theorem 2.2.15.

Remark 2.5.3. By letting $\varepsilon \to 0$, the above problem approaches the one considered in Example 2.3.2, that is the case when the marginals $\mu = \nu$ are equal, the only maximal martingale transport is the identity, and the value of the maximal cost is zero. On the other hand, note that $\int \beta_\varepsilon(y)d\nu(y) - \int \alpha_\varepsilon(x)d\mu(x) = \varepsilon$, which means that $(\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon)$ is a minimizing sequence for the dual problem. But neither of the sequences $\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon$ converge (neither pointwise nor in L^1). This is another manifestation of the non-existence of a dual in example 2.3.2. This said, for the minimization problem, we have no example where duality is not attained.

2.5.2 The case where the convex paving has at most 1-codimensional components

We now establish Theorems 2.2.15 and 2.2.16. The following lemma describes a "flattening map" and shows that it is bi-Lipschitz.

Lemma 2.5.4. *Let $\mathbb{R}^d = V \times W$, where $V = \mathbb{R}^{d-1}$ and $W = \mathbb{R}$. Let $\delta > 0$ and let A be a subset of W. Suppose that for each $h \in A$, there is a set D_h which is contained in a hyperplane H_h with $H_h \cap W = \{0, ..., 0, h\}$. Suppose further that $\{D_h\}_{h \in A}$ are mutually disjoint and the projection of every $\{D_h\}$ on V contains the ball B_R with center 0 and radius R in V. Finally, suppose that the angle between H_h and W is bounded; there is $\eta < \pi/2$ such that the normal direction of H_h and the direction of W has angle less than η for every $h \in A$.*

Now define the flattening map $F : \cup_h D_h \to \mathbb{R}^d$ as follows: for $x = (v, w) \in D_h$, $F(v, w) = (v, h)$. Then F is bi-Lipschitz on the set $N := (\cup_h D_h) \cap (B_r \times W)$, where $r < R$.

Proof. First, note that by the disjointness of $\{D_h\}$ the map F is bijective, so F^{-1} is well-defined. The lemma is intuitively clear; the map F cannot move two nearby points too far away, because the hyperdiscs $\{D_h\}$ are disjoint.

First of all, from the bounded angle assumption, F is clearly bi-Lipschitz on each $F(D_h)$ with the same Lipschitz constant for all $h \in A$. Hence, for $x_1 = (v_1, w_1)$, $x_2 = (v_2, w_2)$, we will assume that x_1, x_2 are contained in D_{h_1}, D_{h_2} respectively, and $h_1 \neq h_2$.

We consider the case $v_1 = v_2 \in V$ and $|v_1| = |v_2| \leq r$. Let L be the 1-dimensional subspace of V containing 0 and v_1. Regarding D_{h_1}, D_{h_2} as affine functions on V, since their graphs on $L \cap B_R$ are disjoint and linear and $r < R$, it is clear that $|w_1 - w_2| \approx |h_1 - h_2|$; i.e.

$$C_1|h_1 - h_2| \leq |w_1 - w_2| \leq C_2|h_1 - h_2| \text{ for some } C_1, C_2 > 0. \tag{2.5.2}$$

Next we consider the case $v_1 \neq v_2$. We want to show $|x_1 - x_2| \approx |F(x_1) - F(x_2)|$, or equivalently,

$$|w_1 - w_2| \approx |h_1 - h_2|.$$

Let L be the 1-dimensional subspace of V containing v_1 and v_2. Regarding D_{h_1}, D_{h_2} as affine functions on V, since their graphs on $L \cap B_R$ are disjoint and linear, it is clear that

$$|w_1 - w_2| = |D_{h_1}(v_1) - D_{h_2}(v_2)| \leq max\big(|D_{h_1}(v_1) - D_{h_2}(v_1)|, |D_{h_1}(v_2) - D_{h_2}(v_2)|\big).$$

But by (2.5.2), we have

$$\max\big(|D_{h_1}(v_1) - D_{h_2}(v_1)|, |D_{h_1}(v_2) - D_{h_2}(v_2)|\big) \leq C_2|h_1 - h_2|,$$

which shows that F^{-1} is Lipschitz on $F(N)$. On the other hand, by (2.5.2), we have

$$|h_1 - h_2| \leq (1/C_1) \min\big(|D_{h_1}(v_1) - D_{h_2}(v_1)|, |D_{h_1}(v_2) - D_{h_2}(v_2)|\big)$$

and, again regarding D_{h_1}, D_{h_2} as disjoint linear graphs on $L \cap B_R$, we have

$$\min\big(|D_{h_1}(v_1) - D_{h_2}(v_1)|, |D_{h_1}(v_2) - D_{h_2}(v_2)|\big) \leq |D_{h_1}(v_1) - D_{h_2}(v_2)| = |w_1 - w_2|$$

which shows that F is Lipschitz on N. Therefore, the lemma follows. $\square$

Proof of Theorem 2.2.15. Write $X = X_\Gamma = X_0 \cup X_1$, where

$$X_0 = \{x \in X : \dim V(C(x)) = d\}, \; X_1 = \{x \in X : \dim V(C(x)) = d-1\}.$$

Note that $X_0 = \cup_{x \in X_0}(X \cap C(x)) = X \cap (\cup_{x\in X_0} C(x))$. Since $\cup_{x\in X_0} C(x)$ is open, X_0 is measurable and so is $X_1 = X - X_0$. By considering a Lebesgue point x of X_0, we get that $\mathcal{L}^d(X_0 \cap C(x)) > 0$. But since $\Gamma \cap (C(x) \times \mathbb{R}^d)$ admits a dual, then Theorem 2.2.4 yields that $\Gamma_x \subseteq \text{Ext}\left(\overline{\text{conv}}(\Gamma_x)\right)$ for a.e. x in $X_0 \cap C(x)$. Since X_0 can be approximated by compact sets from the inside and $\{C(x)\}_{x \in X_0}$ is an open cover of X_0, we conclude that $\Gamma_x \subseteq \text{Ext}\left(\overline{\text{conv}}(\Gamma_x)\right)$ for a.e. x in X_0.

Let $X_\delta := \{x \in X_1 : \delta(x) > \delta\}$ for some $\delta > 0$, let x_0 be a Lebesgue point of X_δ, and consider W to be the 1-dimensional affine space containing x_0 and perpendicular to $C(x_0)$. Choose $\varepsilon > 0$ much smaller than δ and let $X_{\delta,\varepsilon} := X_\delta \cap B(x_0, \varepsilon)$. Then $\left\{C(x) : x \in X_{\delta,\varepsilon}\right\}$ is a disjoint, $d-1$ dimensional open convex cover of $X_{\delta,\varepsilon}$ and $C(x) \cap W \neq \emptyset$. Let $F : \bigcup_{x \in X_{\delta,\varepsilon}} C(x) \to \bigcup_{x\in X_{\delta,\varepsilon}} F(C(x))$ be the flattening map with respect to W as in the lemma 2.5.4. Since F is bi-Lipschitz on the appropriate set containing $X_{\delta,\varepsilon}$, $\mathcal{L}^d(F(X_{\delta,\varepsilon})) > 0$.

Now let $E = \{x \in X : \Gamma_x \nsubseteq \text{Ext}\left(\overline{\text{conv}}(\Gamma_x)\right)\}33$. Then by theorem 2.2.4, $\Gamma_x \subseteq \text{Ext}\left(\overline{\text{conv}}(\Gamma_x)\right)$ for $\mathcal{L}^{d-1}$ almost all x in $X_{\delta,\varepsilon} \cap C(x)$, hence $\mathcal{L}^{d-1}(E \cap X_{\delta,\varepsilon} \cap C(x)) = 0$ for each $x \in X_{\delta,\varepsilon}$.

Now $\left\{F(C(x)) : x \in X_{\delta,\varepsilon}\right\}$ is a parallel cover of $F(X_{\delta,\varepsilon})$, so by Fubini's theorem with bi-Lipschitz map F, we conclude that $\mathcal{L}^d(E \cap X_{\delta,\varepsilon}) = 0$. Hence as in the previous case, we conclude that $\mathcal{L}^d(E \cap X_\delta) = 0$. Finally, letting $\delta \to 0$ we get $\mathcal{L}^d(E \cap X_1) = 0$. □

Proof of Corollary 2.2.16. For $x \in U$, consider the problem (2.1.1) with the source μ restricted to a small ball B with center x. Now by the separation assumption (2.2.6), we have $\delta(x) > c$ on $X_1 \cap B$ for some $c > 0$, hence the same proof as for Theorem 2.2.15 yields the result. □

2.5.3 The case of a discrete target

We shall conclude this section by considering the case of a discrete target. Note that this result is another affirmative sign towards Conjecture 2.

Theorem 2.5.5. *Let $c(x, y) = |x - y|$, suppose $\mu << \mathcal{L}^d$ and that ν is discrete, i.e. ν is supported on a countable set. Then for μ a.e. x, Γ_x consists of $d + 1$ points which are vertices of a polytope in $\mathbb{R}^d$, and therefore the optimal solution is unique.*

Proof. Since the result holds true (for more general target measures) when $d = 1$, we shall assume that $d \geq 2$. Let S be the countable support of ν and let $J := \{E \subseteq S : |E| < \infty \quad \& \quad \dim V(E) \leq d - 1\}$, where $|E|$ is the cardinality of the set E. Consider $V_J := \cup_{E \in J} V(E)$. Since $\dim V(E) \leq d - 1$ and J is countable, it follows that $\mathcal{L}^d(V_J) = 0$. Let $X := X_\Gamma \setminus V(J)$ so that $\mu(X) = 1$. Now notice that if $x \in X$, then Γ_x must contain vertices of a polytope which has x in its interior. Now let

$$K := \{E \subseteq S : \ |E| = d + 2 \text{ and } E \text{ contains vertices of a } d\text{-dimensional polytope.}\}$$

Fix $F = \{y_0, y_1, ..., y_d, y\}$ in K, where $y_0, y_1, ..., y_d$ are vertices of a d-dimensional polytope and consider the set $A := \{x \in X : F \subseteq \Gamma_x\}$. In other words, $A = \Gamma^{y_0} \cap ... \cap \Gamma^{y_d} \cap \Gamma^y$, where $\Gamma^y := \{x : (x, y) \in \Gamma\}$. We shall prove that $\mu(A) = 0$.

Indeed, suppose otherwise, that is $\mu(A) > 0$ and let x_0 be a Lebesgue point of A. Let $B = A \cap C(x_0)$ and note that $\mathcal{L}^d(B) > 0$ since $C(x_0)$ is open in $\mathbb{R}^d$. Since the set $\Gamma \cap (C(x_0) \times \mathbb{R}^d)$ has a c-dual, there exist constants $\lambda_0, \lambda_1, ..., \lambda_d, \lambda$ such that for all $x \in B$, we have

$$|x - y_i| + \gamma(x) \cdot (y_i - x) + \alpha(x) = \lambda_i, \ \ i = 0, 1, ..., d$$
$$|x - y| + \gamma(x) \cdot (y - x) + \alpha(x) = \lambda.$$

Also note that $\{y_0, y_1, ..., y_d, y\} \subseteq \text{Ext}\left(\overline{conv}(\Gamma_x)\right)$ for almost all $x \in B$. Let p_i be determined by $y = \Sigma_{i=0}^d p_i y_i$, and $\Sigma_{i=0}^d p_i = 1$, and note that some p_i may be negative.

Then, by the above, we get that the function

$$g(x) := \Sigma_{i=0}^{d} p_i|x - y_i| - |x - y|$$

is constant on B, which has positive measure.

We explain why this leads to a contradiction. First, notice that because g is real analytic in $\Omega := \mathbb{R}^d \setminus \{y_0, ..., y_d, y\}$, it is not constant in any open subset, since otherwise it is constant everywhere, which is not the case. Second, without loss of generality, assume $x_0 = 0$ and $g(0) = 0$, and notice that from the real analyticity of g, one can write $g(x) = P_k(x) + Q(x)$ for some $k \in \mathbb{N}$, where $P_k(x)$ is the first nonzero k-th degree homogeneous polynomial, and $Q(x)$ is a power series of terms with degree greater than k, in particular, $Q(x) = O(|x|^{k+1})$. Now, consider the set

$$\mathcal{S} := \{u \in S^{d-1} \mid \text{ there exists } 0 \neq x_n \to 0,\ x_n/|x_n| \to u, \text{ with } 0 = g(x_n)\}.$$

Then, for each $u \in \mathcal{S}$, $0 = \frac{g(x_n)}{|x_n|^k} = P_k(x_n/|x_n|) + \frac{Q(x_n)}{|x_n|^k}$, showing $P_k(u) = \lim_{n\to\infty} P_k(x_n/|x_n|) = 0$. Thus, $\mathcal{S}$ is a subset of the zero set $\{u;\ P_k(u) = 0\}$.

Now if g is zero on the set B where x_0 is a Lebsegue point, then $\mathcal{S} = S^{d-1}$, hence $P_k = 0$, a contradiction. Hence, $\mu(A) = 0$. The countability of K now implies the theorem.

For the uniqueness, we use the usual argument, namely that the average of two optimal plans is also optimal, which contradicts the polytope-type of their respective supports. □

Remark 2.5.6. As we see from the above proof, Theorem 2.5.5 holds true for a much more general cost $c(x, y)$ than $|x-y|$. Indeed, it is enough (but not necessary) $c(x, y)$ to be analytic in $\{x \neq y\}$, and the function $g(x) = \sum_{i=0}^{d} p_i c(x, y_i) - c(x, y)$ to be non-constant. In particular, we can choose $c(x, y) = |x - y|^p$, with $p \neq 2$.

2.6 Minimization problem under radially symmetric marginals

So far, we have studied the optimal martingale transport problem in higher dimensions, and we conjectured the following extremal property of minimizers.

Conjecture: *Consider the cost function $c(x, y) = |x - y|$ and assume that μ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, and that $\mu \wedge \nu = 0$. If π is a martingale transport that minimizes* (2.1.1)*, then there is a set Γ with $\pi(\Gamma) = 1$ such that for μ almost every x, the fiber $\Gamma_x := \{y : (x, y) \in \Gamma\}$ consists of $k+1$ points that form the vertices of a k-dimensional polytope, where $k := k(x)$ is the dimension of the linear span of Γ_x. Finally, the minimizing solution is unique.*

We have seen that the conjecture is true when the source μ is absolutely continuous and the target ν is discrete. In this section, we show that the above conjecture is true in the minimization case when the marginals μ and ν are radially symmetric on $\mathbb{R}^d$, and determine the explicit structure of the minimizer. Many important probability distributions, e.g. the Gaussians, satisfy the radial symmetry assumption.

Theorem 2.6.1. *Suppose that μ, ν are radially symmetric probability measures on $\mathbb{R}^d$ which are in convex order and $\mu\{0\} = 0$. Assume that either μ is absolutely continuous, or that there exists an open ball $B_r = \{x \,|\, |x| < r\}$ such that $\mu(B_r) = 1$ and $\nu(B_r) = 0$. Then there is a unique minimizer π for the problem* (2.1.1) *with respect to the cost $c(x, y) = |x-y|^p$ where $0 < p \leq 1$, and for each x, the disintegration π_x is concentrated on the one-dimensional subspace $L_x = \{ax : a \in \mathbb{R}\}$. Furthermore, if μ is absolutely continuous with respect to Lebesgue measure and $\mu \wedge \nu = 0$, then π_x is supported at two points, hence verifying the conjecture.*

The organization of this section is as follows. In section 2.6.1, we introduce the monotonicity principle [5] and establish the stability of the common marginal $\mu \wedge \nu$ under every minimizer of (2.1.1). In section 2.6.2, we further apply the monotonicity to determine the structure of the minimizer in one dimension. Finally, in section

2.6.3, we establish the variational lemma and the main theorem.

2.6.1 Monotonicity principle and stability of $\mu \wedge \nu$ under every minimizer

An important basic tool in optimal transport is the notion of *c-cyclical monotonicity.* A parallel statement was given in [5].

Definition 2.6.2. *Let φ be a finite measure supported on a finite set $H \subseteq \mathbb{R}^d \times \mathbb{R}^d$. Let X_H be the orthogonal projection of H onto the first coordinate space $\mathbb{R}^d$. Then we say that ψ is a* competitor *of φ if ψ has the same marginals as φ and for each $x \in X_H$, $\int_{\mathbb{R}^d} y\, d\varphi(x,y) = \int_{\mathbb{R}^d} y\, d\psi(x,y)$.*

Lemma 2.6.3 (Monotonicity principle [5],[39])**.** *Assume that μ, ν are probability measures in convex order and that $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a Borel measurable cost function. Assume that $\pi \in MT(\mu,\nu)$ is an optimal martingale transport plan which leads to finite cost. Then there exists a Borel set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ with $\pi(\Gamma) = 1$ such that the following monotonicity principle holds:*

If φ is a finite measure on a finite set $H \subseteq \Gamma$, then for every competitor ψ of φ,
$$\int c\, d\varphi \le \int c\, d\psi.$$

The meaning of the monotonicity principle is clear; $spt(\varphi) \subseteq \Gamma$ means that φ is a "subplan" of the full transport plan π, and the definition of competitor means that if we change the subplan φ to ψ, then the martingale structure of π is not disrupted. Now if we have $\int c\, d\varphi > \int c\, d\psi$, then we may modify π to have ψ as its subplan, achieving less cost, therefore the current plan π is not a minimizer.

We recall the notations, which are introduced in [5]:

For a set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$, we write $X_\Gamma := \mathrm{proj}_X \Gamma$, $Y_\Gamma := \mathrm{proj}_Y \Gamma$, i.e. X_Γ is the projection of Γ on the first coordinate space $\mathbb{R}^d$, and Y_Γ on the second. For each $x \in \mathbb{R}^d$, we let $\Gamma_x = \{y \in \mathbb{R}^d \mid (x,y) \in \Gamma\}$.

Now as an application of the monotonicity principle, we prove the stability of $\mu \wedge \nu$ under every optimal martingale transport. [5] discusses the following theorem in one-dimensional setup with $p = 1$. We prove it here in general dimension with every $0 < p \leq 1$. Note that radial symmetry of μ, ν is not assumed.

Theorem 2.6.4. *Let π be any minimizer of the problem* (2.1.1) *with cost $c(x,y) = |x-y|^p$, $0 < p \leq 1$. Then under π the common mass $\mu \wedge \nu$ stays put, in the sense that if we define $D : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ by $D(x) = (x,x)$, then the push-forward measure of $\mu \wedge \nu$ by the map D is dominated by π, i.e. $D_\#(\mu \wedge \nu) \leq \pi$.*

Proof. Suppose that the theorem is false. Then there exists a minimizer π such that $D_\#(\mu \wedge \nu) \not\leq \pi$. Let Γ be a monotone concentration set of π as in Lemma 2.6.3 and denote $(\pi_x)_{x \in \mathbb{R}^d}$ as its integration. Since $D_\#(\mu \wedge \nu) \not\leq \pi$, we can find a point $x \in spt(\mu \wedge \nu)$ such that π_x is not a Dirac mass δ_x and $(z,x) \in \Gamma$ for some $z \neq x$. Then as $\pi_x(\Gamma_x) = 1$, we can find a probability measure ψ_x such that $\psi_x \neq \delta_x$, ψ_x is supported on a finite subset of Γ_x, and ψ_x has its barycenter at x.

Now for every $0 < p \leq 1$ and $z \in \mathbb{R}^d$, we have that $|x-y|^p + |z-x|^p \geq |y-z|^p$, hence

$$\int |x-y|^p d\psi_x(y) + |z-x|^p \geq \int |y-z|^p d\psi_x(y). \tag{2.6.1}$$

But the inequality is strict since $\psi_x \neq \delta_x$, a contradiction to the fact that Γ is monotone. Therefore, the theorem holds and every minimizer π makes the common mass $\mu \wedge \nu$ stay put. □

Remark 2.6.5. By the theorem, we can reduce the problem between disjoint marginals $\bar{\mu} := \mu - \mu \wedge \nu$ and $\bar{\nu} := \nu - \mu \wedge \nu$. Thus, from now on we will always assume that $\mu \wedge \nu = 0$, and therefore, for any minimizer $\pi \in MT(\mu, \nu)$, we have a monotone set Γ such that $\pi(\Gamma) = 1$ and $\Gamma \cap \Delta = \emptyset$, where $\Delta := \{(x,x) \,|\, x \in \mathbb{R}^d\}$.

2.6.2 Structure of optimal martingale transport in one dimension

In this section, we study the minimization problem (2.1.1) in one dimension, i.e. the marginals μ, ν are defined on the real line $\mathbb{R}$. We will consider the cost function $c(x, y) = |x - y|^p$ with $0 < p \leq 1$ and will determine the structure of optimal coupling. In this section, we do not assume the symmetry of marginals μ, ν with respect to the origin. Recall that we can assume $\mu \wedge \nu = 0$. Finally, we will say that μ is continuous if μ does not assign positive measure at any point; $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}$.

The following theorem for the 1-dimensional case was shown in [5] when $p = 1$, and we extend it for every $0 < p \leq 1$.

Theorem 2.6.6. *Assume that $\mu \wedge \nu = 0$, μ is continuous and $c(x, y) = |x - y|^p$ for some $0 < p \leq 1$. Let π be a minimizer for the problem* (2.1.1) *with $d = 1$. Then, there exists a monotone set Γ such that $\pi(\Gamma) = 1$ and for every $x \in X_\Gamma$, we have $|\Gamma_x| = 2$. Hence if we define two functions $S : X_\Gamma \to \mathbb{R}$ and $T : X_\Gamma \to \mathbb{R}$ by $\Gamma_x = \{S(x), T(x)\}$ and $S(x) < x < T(x)$, then π is concentrated on graph$(S) \cup$ graph(T). Therefore, the minimizer is unique.*

Proof. Let Γ be any monotone concentration set of π and suppose $(x, y^-), (x, y^+), (x', y') \in \Gamma$, with $y^- < y' < y^+$. Then we claim that neither $y^- < x' < x \leq y'$ nor $y' \leq x < x' < y^+$ is possible. To prove the claim, suppose $y^- < x' < x \leq y'$ and let $0 < t < 1$ be such that $ty^- + (1 - t)y^+ = y'$. Now consider the function

$$G(z) = t|z - y^-|^p + (1 - t)|z - y^+|^p - |z - y'|^p.$$

If $y^- < z < y'$,

$$G(z) = t(z - y^-)^p + (1 - t)(y^+ - z)^p - (y' - z)^p.$$

Taking derivative, we get

$$G'(z) = p[t(z-y^-)^{p-1} - (1-t)(y^+ - z)^{p-1} + (y'-z)^{p-1}].$$

We observe

$$\text{If } 0 < p < 1, \quad (y'-z)^{p-1} > (y^+ - z)^{p-1} \quad \text{hence} \quad G'(z) > 0.$$
$$\text{If} \qquad p = 1, \quad G'(z) = t - (1-t) + 1 = 2t > 0.$$

Hence for $y^- < x' < x \le y'$, we have $G(x') < G(x)$. In other words,

$$t(x'-y^-)^p + (1-t)(y^+ - x')^p + |y'-x|^p < t(x-y^-)^p + (1-t)(y^+ - x)^p + |y'-x'|^p.$$

This means that if we define a measure φ by $\varphi = t\delta_{(x,y^-)} + (1-t)\delta_{(x,y^+)} + \delta_{(x',y')}$, then we have a cost-efficient competitor ψ by $\psi = t\delta_{(x',y^-)} + (1-t)\delta_{(x',y^+)} + \delta_{(x,y')}$. Note that ψ satisfies the assumption to be a competitor of φ. Hence by Lemma 2.6.3, $(x,y^-),(x,y^+),(x',y') \in \Gamma$ with $y^- < y' < y^+$ and $y^- < x' < x \le y'$ cannot occur. The case $y' \le x < x' < y^+$ cannot occur as well.

Now we follow the proof of [5]; Suppose the set $A := \{x \in \mathbb{R} : |\Gamma_x| \ge 3\}$ is uncountable. Then we will have $(x,y^-),(x,y^+),(x,y) \in \Gamma$, with $y^- < x < y < y^+$ or $y^- < y < x < y^+$ (Recall that $\Gamma \cap \Delta = \emptyset$, where $\Delta := \{(x,x) \,|\, x \in \mathbb{R}^d\}$, since $\mu \wedge \nu = 0$). Assume the first case. Then the Lemma 3.2 in [5] shows that any given $\varepsilon > 0$, we have $(x',y') \in \Gamma$ with $x - \varepsilon < x' < x$ and $|y'-y| < \varepsilon$ by the uncountability of A. Then for small ε we have the first forbidden case, and similarly if $y^- < y < x < y^+$ then we have $(x',y') \in \Gamma$ with $x < x' < x + \varepsilon$ and $|y'-y| < \varepsilon$, the second forbidden case. Hence, A must be countable, therefore by continuity of μ, A is negligible.

Uniqueness follows by usual argument, namely, if π_1 and π_2 are optimal solutions realized by (S_1,T_1) and (S_2,T_2) respectively, then the average $\frac{\pi_1+\pi_2}{2}$ is also optimal and hence it must also be realized by two functions (S,T). This implies that $S_1(x) = S_2(x)$ and $T_1(x) = T_2(x)$ for μ a.e. x, yielding uniqueness. $\square$

In fact, we can say more on the structure of optimal martingale. Note that in the following lemma, the continuity of μ is not assumed.

Lemma 2.6.7. *Let $I := (a,b)$ be an open interval and suppose $\nu(I) = 0$. Let $c(x,y) = |x-y|^p$ for some $0 < p \leq 1$ and let π be a minimizer for the problem* (2.1.1)*. Denote $(\pi_x)_x$ be its disintegration and if $x \in I$, then denote π_x^+ as the restriction of π_x on $[b,\infty)$ and π_x^- as the restriction of π_x on $(-\infty,a]$.*

Then if $x, x' \in I$ and $x < x'$, then $\sup(spt(\pi_{x'}^+)) \leq \inf(spt(\pi_x^+))$ and $\sup(spt(\pi_{x'}^-)) \leq \inf(spt(\pi_x^-))$. In other words, the set valued functions $spt(\pi_x^+)$ and $spt(\pi_x^-)$ decrease on I.

Proof. Let Γ be a monotone concentration set of π with $Y_\Gamma \cap I = \emptyset$. If $x, x' \in I \cap X_\Gamma$ and $x < x'$, then we claim that $\sup(\Gamma_{x'}^+) \leq \inf(\Gamma_x^+)$, where $\Gamma_x^+ := \Gamma_x \cap [b,\infty)$. If not, then we can find $y' > y \geq b$ such that $(x,y), (x',y') \in \Gamma$. As π is a martingale, we can also find $y'' \leq a$ with $(x', y'') \in \Gamma$. Then the configuration $(x,y), (x',y'), (x',y'') \in \Gamma$ is forbidden by the proof of Theorem 2.6.6, a contradiction. As $\pi_x(\Gamma_x) = 1$ for every $x \in X_\Gamma$, π_x^+ has its full mass on Γ_x^+, hence $\sup(spt(\pi_{x'}^+)) \leq \inf(spt(\pi_x^+))$. The other case $\sup(spt(\pi_{x'}^-)) \leq \inf(spt(\pi_x^-))$ can be proved similarly. $\square$

We may say the above result as "local decreasing property", as the function $x \mapsto spt(\pi_x^+)$ and $x \mapsto spt(\pi_x^-)$ decrease locally, i.e. on an open interval I whenever $\nu(I) = 0$. Thus if we make the following assumption, we will have the global decreasing property for any optimal martingale transport.

Separation Assumption: There is an open interval I so that $\mu(I) = 1$ and $\nu(I) = 0$.

For example, two Gaussian measures μ, ν in convex order will satisfy this assumption, after $\mu \wedge \nu$ is subtracted from each marginal. Now we observe that the global decreasing property also yields the uniqueness of optimal solution, without assuming the continuity of μ.

Theorem 2.6.8. *Under the separation assumption, a solution for the problem* (2.1.1) *with $c(x,y) = |x-y|^p$ for some $0 < p \leq 1$ is unique. Moreover, the optimal solution is identical for all $0 < p \leq 1$.*

Proof. Let $0 < p \leq 1$ be fixed and let π be a minimizer. Then the separation assumption yields that $(\pi_x)_{x\in I}$ decreases on $I = (a, b)$, as shown in Lemma 2.6.7. By the decreasing property, π must take the mass of μ from the left and transport it to fill out the ν^+ (ν restricted on $[b, \infty)$) and ν^- (ν restricted on $(-\infty, a]$) from the right in a martingale way. When μ, ν are continuous, this is described by the following equations with functions $S(x) \in (-\infty, a]$ and $T(x) \in [b, \infty)$:

$$\mu\left((a, x]\right) = \nu^-\left((S(x), a]\right) + +\nu^+\left((T(x), \infty)\right),$$
$$\int_a^x t\, d\mu(t) = \int_{S(x)}^a t\, d\nu^-(t) + \int_{T(x)}^\infty t\, d\nu^+(t).$$

The first equation says the preservation of mass, and the second says the preservation of barycenter. In the case μ, ν are general, with functions $0 \leq \lambda^-(x), \lambda^+(x) \leq 1$ it may be written as

$$\mu\left((a, x]\right) = \lambda^-(x)\, \nu^-(S(x)) + \nu^-\left((S(x), a]\right) + \lambda^+(x)\, \nu^+(T(x)) + \nu^+\left((T(x), \infty)\right),$$
$$\int_{(a,x]} t\, d\mu(t) = S(x)\, \lambda^-(x)\, \nu^-(S(x)) + \int_{(S(x),a]} t\, d\nu^-(t)$$
$$+ T(x)\, \lambda^+(x)\, \nu^+(T(x)) + \int_{(T(x),\infty)} t\, d\nu^+(t).$$

Note that $S(x), T(x), \lambda^-(x), \lambda^+(x)$ are uniquely determined, if $S(x), T(x)$ are chosen as the largest numbers which satisfy the above equations. Furthermore, it is clear that these equations uniquely determine the martingale coupling π. Finally, the above equations are derived only from the decreasing property of π and do not depend on p, therefore the theorem follows. $\square$

In particular if we assume symmetry of μ, ν, then we have:

Corollary 2.6.9. *If μ, ν are symmetric with respect to the origin, then the unique coupling in Theorem 2.6.6 or 2.6.8 is also symmetric.*

Proof. We can prove it directly, or we let σ be an optimal coupling in Theorem 2.6.6 or 2.6.8, and let $\tau = \frac{1}{2}(\sigma + \sigma')$ be a symmetrization of σ, where σ' is the reflection of σ with respect to the origin. Then τ is also optimal, so by uniqueness, $\sigma = \tau$. $\square$

2.6.3 Structure of optimal martingale transport in higher dimensions

We have studied the structure of the martingale transport in one dimension which minimizes $\mathbb{E}_{\mathrm{P}}\,|X - Y|^p$ where $0 < p \leq 1$, and in particular have shown its uniqueness either when μ is continuous or when the separation assumption holds. In this section, we will introduce the notion of symmetrization of a transport plan, and then will present a variational calculus which will lead the higher dimensional problem under radially symmetric marginals into the one-dimensional situation.

L- and R- symmetrizations

We first introduce the notion of L- symmetrization.

Definition 2.6.10. *Let L be a one-dimensional subspace of $\mathbb{R}^d$ and let ψ be a probability measure on $\mathbb{R}^d$. We say that ψ is L-symmetric if it is symmetric with respect to L, i.e. for any Borel set B and any orthogonal matrix M which fixes L, we have $\psi(B) = \psi(M(B))$.*

We say that the probability measures φ and ψ are L-equivalent if $\varphi(B) = \psi(B)$ for every $B \subseteq \mathbb{R}^d$ which is symmetric with respect to L, i.e. $x \in B$ implies $z \in B$ for every z with $dist(z, L) = dist(x, L)$ and $x - z \perp L$. Then we write $\varphi \cong_L \psi$. Finally, we define $L[\psi]$ to be the unique L-symmetric measure that is L-equivalent to ψ.

Now we turn to the notion of R-symmetrization. Let $S_r = \{x \,|\, |x| = r\}$ be the unit sphere in $\mathbb{R}^d$ with radius r, let ζ be a probability measure on S_r, and let $(\pi_x)_{x \in S_r}$ be a set of probability measures on $\mathbb{R}^d$ attached on each $x \in S_r$. We can view this as a mass transport plan π with initial mass ζ, by seeing π_x as its disintegration. We may denote this as $\pi = (\zeta, \pi_x)$. Fix a vector $w \in S_r$, and let $(M_x)_{x \in S_r}$ be a choice of orthogonal matrices with the property $M_x(x) = w$. For $u \in \mathbb{R}^d$, let $T_u(z) = z + u$ be translations. Finally, let L_x be the one-dimensional subspace spanned by $x \neq 0$.

Now consider the following probability measure

$$\sigma(\cdot) := \int_{x \in S_r} (M_x \circ T_{-x} \circ L_x(\pi_x))\,(\cdot)\, d\zeta(x)$$

Note that σ does not depend on the choice of $(M_x)_{x \in S_r}$, due to the presence of the operator L_x in the definition. Now we define the R- symmetrization operator acting on the transport plan $\pi = (\zeta, \pi_x)$.

Definition 2.6.11. *The R-symmetrization operator is defined by*

$$R[\pi] = R[(\zeta, \pi_x)] = (U, \sigma_x)$$

where $\sigma_x = (M_x \circ T_{-x})^{-1}(\sigma)$ *and* U *is the uniform probability measure on* S_r.

Thus, σ is an average of appropriately translated and rotated π_x's with weight ζ, and R-symmetrization operator uniformly pushes σ back on S. Now for any transport plan $\pi = (\mu, \pi_x)$ with general initial distribution μ, one can similarly apply the R-symmetrization, by applying the above R-symmetrization on each disintegration of μ along the spherical layers S_r.

Finally, we introduce the notion of R- equivalence on the space of probability measures on $\mathbb{R}^d$.

Definition 2.6.12. *Probability measures* φ *and* ψ *are called R-equivalent if they contain the same mass on any annulus, i.e. for any* $B \subseteq \mathbb{R}_+$ *and any* $A_B := \{x \in \mathbb{R}^d \,|\, |x| \in B\}$, *we have* $\varphi(A_B) = \psi(A_B)$. *We write* $\varphi \cong_R \psi$.

Next, we will apply the symmetrization ideas to study the structure of optimal martingale transport in higher dimensions, when the marginals are radially symmetric.

Variational Lemma and Main Theorem

In this section, we will present a variational lemma which will allow martingale transport problem under radial marginals to be reduced to the problem on the one-dimensional subspaces, where we can apply the results in the previous section.

Lemma 2.6.13 (Variational Lemma)**.** *Consider the cost function of the form* $c(x,y) = h(|x-y|)$ *and let* L_x *be the one-dimensional subspace spanned by* x, $x \neq 0$. *Let* φ *be a probability measure on* $\mathbb{R}^d$ *with barycenter at* x *and assume that* φ *is not supported on* L_x. *Suppose that* $r \mapsto h'(r)/r$ *is strictly decreasing for* $r > 0$. *Then, there exists a probability measure* ψ *with barycenter at* x, *supported on* L_x *and is R- equivalent to* φ, *such that*

$$\int_{\mathbb{R}^d} h(|x-y|)\, d\varphi(y) > \int_{\mathbb{R}^d} h(|x-y|)\, d\psi(y). \tag{2.6.2}$$

For example, $h(r) = r^p$, $0 < p < 2$, or $h(r) = -r^p$, $p > 2$, satisfies the assumption of the lemma.

Proof. We can assume that φ is L_x-symmetric, as it does not change the cost $\int_{\mathbb{R}^d} h(|x-y|)\, d\varphi(y)$. Thus, it is sufficient to consider the two-dimensional case $d = 2$, as we will see. Now we will explain how to deform φ to obtain ψ in the lemma.

For this, let us consider the family of probability measures $\psi(t)$ supported on the four points $z_n^1(t), z_n^2(t), z_s^1(t), z_s^2(t)$, where $0 \leq t \leq 1$ is a parameter. We will observe that the cost of $\psi(t)$ strictly decreases as t increases, which will be the desired deformation process.

To begin, without loss of generality let the barycenter x be a point in $\mathbb{R}^2$, $x = (0,b), b \neq 0$. Let $z^1, z^2 \in \mathbb{R}^2$, $|z^1| = |z^2| = r > 0$ and let $z^1 = (a,z)$, $z^2 = (-a,z)$. Now for $0 \leq t \leq 1$, let $z_n(t) = z + t(r-z)$, $z_s(t) = z - t(r+z)$, and let

$$z_n^1(t) = \big(\sqrt{r^2 - (z_n(t))^2}, z_n(t)\big), \quad z_n^2(t) = \big(-\sqrt{r^2 - (z_n(t))^2}, z_n(t)\big),$$
$$z_s^1(t) = \big(\sqrt{r^2 - (z_s(t))^2}, z_s(t)\big), \quad z_s^2(t) = \big(-\sqrt{r^2 - (z_s(t))^2}, z_s(t)\big).$$

Thus, the four points $z_n^1(t), z_n^2(t), z_s^1(t), z_s^2(t)$ are on the circle of center 0 and radius r, and they are symmetrically located with respect to the vertical axis. Now define $\psi(t)$ and its transportation cost

$$\psi(t) = \frac{r+z}{4r}\delta_{z_n^1(t)} + \frac{r+z}{4r}\delta_{z_n^2(t)} + \frac{r-z}{4r}\delta_{z_s^1(t)} + \frac{r-z}{4r}\delta_{z_s^2(t)},$$
$$C(t) = \frac{r+z}{2r}h(||z_n^1(t)-x||) + \frac{r-z}{2r}h(||z_s^1(t)-x||).$$

Thus, $C(t)$ is the cost of transporting the point mass δ_x to $\psi(t)$. Note that $\psi(0) = \frac{1}{2}\delta_{(-a,z)} + \frac{1}{2}\delta_{(a,z)}$ and $\psi(1) = \frac{r+z}{2r}\delta_{(0,r)} + \frac{r-z}{2r}\delta_{(0,-r)}$, so $\psi(t)$ is a continuous deformation from $\psi(0)$ to $\psi(1)$ along the circle of radius r. Now, observe that for all $0 \le t \le 1$, the barycenter of $\psi(t)$ is fixed at $(0,z)$ and they are obviously R-equivalent. We will show $C'(t) < 0$ if $h'(r)/r$ is strictly decreasing for $r > 0$; we compute

$$\begin{aligned} C'(t) = {} & \left(\frac{r+z}{2r}\right)\frac{h'(||z_n^1(t)-x||)}{||z_n^1(t)-x||}\Big\langle z_n^1(t)-x, \frac{d}{dt}\big(z_n^1(t)\big)\Big\rangle \\ & + \left(\frac{r-z}{2r}\right)\frac{h'(||z_s^1(t)-x||)}{||z_s^1(t)-x||}\Big\langle z_s^1(t)-x, \frac{d}{dt}\big(z_s^1(t)\big)\Big\rangle \end{aligned}$$

where $\langle , \rangle$ is the inner product. Note that $\big\langle z_n^1(t), \frac{d}{dt}\big(z_n^1(t)\big)\big\rangle = \big\langle z_s^1(t), \frac{d}{dt}\big(z_s^1(t)\big)\big\rangle = 0$, and

$$\Big\langle x, \frac{d}{dt}\big(z_n^1(t)\big)\Big\rangle = b(r-z), \quad \Big\langle x, \frac{d}{dt}\big(z_s^1(t)\big)\Big\rangle = -b(r+z), \quad \text{hence}$$
$$C'(t) = \left(\frac{b(r+z)(r-z)}{2r}\right)\left[\frac{h'(||z_s^1(t)-x||)}{||z_s^1(t)-x||} - \frac{h'(||z_n^1(t)-x||)}{||z_n^1(t)-x||}\right].$$

Now we compute

$$\begin{aligned} ||z_n^1(t)-x||^2 &= r^2+b^2-2b\,z_n(t) = r^2+b^2-2b(z+t(r-z)) \\ ||z_s^1(t)-x||^2 &= r^2+b^2-2b\,z_s(t) = r^2+b^2-2b(z-t(r+z)) \\ ||z_n^1(t)-x||^2 - ||z_s^1(t)-x||^2 &= -4brt \end{aligned}$$

Hence, we see that

$$||z_n^1(t) - x|| < ||z_s^1(t) - x|| \quad \text{if} \quad b > 0$$
$$||z_n^1(t) - x|| > ||z_s^1(t) - x|| \quad \text{if} \quad b < 0$$

Thus in any case $C'(t) < 0$. Hence, $C(0) > C(1)$ and note that $\psi(1)$ is supported on the line L_x.

Now we return to the general case and assume that φ is L_x-symmetric (if not, consider $L_x(\varphi)$) but not supported on L_x. Consider a pair of points z^1, z^2 in $\mathbb{R}^d$ that are symmetric with respect to L_x, i.e. $|z^1| = |z^2| = r$, $\langle x, z^1 - z^2 \rangle = 0$ and $\frac{z^1+z^2}{2} \in L_x$. Now by L_x-symmetry of φ, $d\varphi(z^1) = d\varphi(z^2)$, so we can perform the continuous deformation along a great circle of radius r, and by the above computation the cost is strictly decreasing. Hence, after performing the deformation to all such pair of points z^1, z^2 in the support of φ, we obtain ψ which is supported on the ray L_x, and we have shown that (2.6.2) holds. $\square$

Finally, we apply the symmetrization arguments in conjunction with the variational lemma, to conclude the following theorem. Recall that without loss of generality we can assume $\mu \wedge \nu = 0$, by Remark 2.6.5.

Theorem 2.6.14. *Suppose that $\mu \wedge \nu = 0$, $\mu(\{0\}) = 0$ and μ, ν are radially symmetric probability measures on $\mathbb{R}^d$ which are in convex order. Assume that either μ is absolutely continuous, or that there exists an open ball $B_r = \{x \,|\, |x| < r\}$ such that $\mu(B_r) = 1$ and $\nu(B_r) = 0$. Then there is a unique minimizer π for the problem* (2.1.1) *with respect to the cost $c(x, y) = |x - y|^p$, where $0 < p \leq 1$. Furthermore, for each x, the disintegration π_x is concentrated on the one-dimensional subspace L_x.*

Proof. Let π be a minimizer. First of all, we claim that for μ a.e. x, the disintegration π_x is concentrated on the line L_x. If not, then we apply the variational lemma 2.6.13 for each π_x and get a competitor ψ_x, so that we get another martingale $\psi = (\mu, \psi_x)$ having ψ_x as its disintegrations. Then $\text{cost}(\pi) > \text{cost}(\psi)$, but ψ is not necessarily be in $\text{MT}(\mu, \nu)$. However, since π_x and ψ_x are R-equivalent, if we apply the R-symmetrization to the martingale ψ, then $R[\psi]$ is in $\text{MT}(\mu, \nu)$, by radial symmetry of μ and ν. Now $\text{cost}(\pi) > \text{cost}(\psi) = \text{cost}(R[\psi])$, a contradiction. Hence the claim is true for every minimizer $\pi \in \text{MT}(\mu, \nu)$. This implies that the problem (2.1.1) is decomposed to the problem on each one-dimensional subspaces, which respect to radially disintegrated marginals along each subspaces. Then the corollary 2.6.9 says that the solution for each reduced problem is unique, yielding uniqueness for the whole problem. $\square$

Remark 2.6.15. When μ and ν are defined by density functions f and g on $\mathbb{R}^d$, then the radially disintegrated marginals along one-dimensional subspace have corresponding density functions $f_0(r) = f(r)|r|^{d-1}$ and $g_0(r) = g(r)|r|^{d-1}$, $-\infty < r < \infty$, by radial symmetry of μ and ν. Let us check that f_0 and g_0 are in convex order.

By radial symmetry of f and g, f_0 and g_0 have the same mass and the same barycenter 0. Thus, we only need to check that

$$\int_{\mathbb{R}} (r-k)_+ f_0(r)dr \leq \int_{\mathbb{R}} (r-k)_+ g_0(r)dr$$

for all real k. Let $s_k(r) = (r-k)_+$ and $h_k(r) = \frac{1}{2}(s_k(r) + s_k(-r))$. Now let $H(x) = h_k(|x|)$ for $x \in \mathbb{R}^d$. Then H is a radially symmetric convex function, hence

$$\int_{\mathbb{R}^d} H(x)f(x)dx \leq \int_{\mathbb{R}^d} H(x)g(x)dx$$

and it is clear that

$$\int_{\mathbb{R}^d} H(x)f(x)dx = C\int_{\mathbb{R}} h_k(r)f(r)|r|^{d-1}dr = C\int_{\mathbb{R}} (r-k)_+ f_0(r)\, dr$$

where C is the surface area of unit sphere in $\mathbb{R}^d$.

Chapter 3

Structure of optimal subharmonic martingale transport in higher dimensions

3.1 Formulation of the problem

3.1.1 Skorokhod embedding problem

Let μ and ν be probability measures on $\mathbb{R}^d$, $d \geq 2$, and let B denote a Brownian motion with initial law μ, i.e. $B_0 \sim \mu$.

Consider the set of stopping times

$$T = \{\tau : \tau \text{ is a stopping time and } B_0 \sim \mu, B_\tau \sim \nu.\}$$

Thus, a stopping time $\tau \in T$ solves Skorokhod embedding problem with terminal law ν. We note that T contains all randomized stopping times; definition will be given later.

Next, we introduce the cost function

$$c(x, y) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}.$$

Let $f : \mathbb{R}_+ \to \mathbb{R}$, $f(0) = 0$, be a continuous function. In this chapter, we are interested in the following form:

$$c(x, y) = f(|x - y|).$$

A particular example is when $f(r) = r^\alpha, 0 < \alpha \neq 2$,

$$c(x, y) = |x - y|^\alpha. \tag{3.1.1}$$

Now we consider the following optimization problem

$$\text{Maximize / Minimize} \quad \mathbb{E}f(|B_0 - B_\tau|) \quad \text{over} \quad \tau \in T \tag{3.1.2}$$

The purpose of this chapter is to characterize such optimal stopping times, especially when the initial and terminal laws (a.k.a. marginals) μ and ν are radially symmetric.

3.1.2 Subharmonic martingale transport problem

[21] considered the following problem

$$\text{Maximize / Minimize} \quad \text{Cost}[\pi] = \iint c(x, y) d\pi(x, y) \quad \text{over} \quad \pi \in MT(\mu, \nu)$$

where $\mathrm{MT}(\mu, \nu)$ (Martingale Transport plan) is the set of joint probabilities of μ and ν on $\mathbb{R}^d \times \mathbb{R}^d$, such that for each $x \in \mathbb{R}^d$ and $\pi \in \mathrm{MT}(\mu, \nu)$, the disintegration π_x has its barycenter at x; in other words, for any convex function f on $\mathbb{R}^d$,

$$f(x) \leq \int f(y)\, d\pi_x(y). \tag{3.1.3}$$

Now we define $\mathrm{SMT}(\mu, \nu)$ (Subharmonic Martingale Transport plan) as above, but the constraint (3.1.3) must hold for every subharmonic function f. Since every convex function is subharmonic, SMT problem is more restricted; $\mathrm{SMT}(\mu, \nu) \subseteq \mathrm{MT}(\mu, \nu)$.

Example 3.1.1 ($\mathrm{SMT}(\mu, \nu) \neq \mathrm{MT}(\mu, \nu)$). Let μ be the uniform measure on the unit sphere, and let ν have $\frac{1}{2}$ mass at the origin and $\frac{1}{2}$ mass uniformly on the sphere of radius 2. Then $\mathrm{MT}(\mu,\nu) \neq \emptyset$; for each x on the unit sphere, define the disintegration π_x by sending $\frac{1}{2}$ mass to the origin and $\frac{1}{2}$ mass to the point $2x$. Then the resulting π is an MT plan and clearly satisfies the marginal condition. However, if we consider a subharmonic function $f(z) = log\,|z|$, we have

$$\int f(z)\, d\mu\,(z) = 0 > -\infty = \int f(z)\, d\nu\,(z) \tag{3.1.4}$$

Hence, $\mathrm{SMT}(\mu,\nu) = \emptyset$.

Definition 3.1.2. *We call a probability measure ψ a* SH-*measure with barycenter x if* (3.1.3) *holds for every subharmonic function f, and* $\mathrm{SH}(x)$ *be the set of all* SH-*measures at x.*

We denote B^x a Brownian motion starting at x. Let τ be a stopping time. Observe that, for any subharmonic function f, $(f(B^x_t), t \geq 0)$ is a submartingale, hence the measure $\mathrm{Law}(B^x_\tau)$ is in $\mathrm{SH}(x)$. In fact, the coverse is also true, and we give a proof of the following proposition in the appendix.

Proposition 3.1.3. *Let $\psi \in \mathrm{SH}(x)$ and suppose* $\mathrm{supp}(\psi)$ *is compact. Then there exists a stopping time τ such that $\psi = \mathrm{Law}(B^x_\tau)$.*

For the proposition to hold, we need to consider the randomized stopping times; see [6]. Now by proposition, we see that the optimal Skorokhod embedding problem and the optimal subharmonic martingale transport problem are equivalent.

Remark 3.1.4. In $d = 2$, SH-measures are also called Jensen measures. However, in $\mathbb{R}^{2d} \cong \mathbb{C}^d$ with $d \geq 2$, the two notions are not equivalent, so in this chapter we will use the notation SH-measure.

Remark 3.1.5. For simplicity, we will assume throughout this chapter

Marginals μ and ν are supported in a compact ball $K \subseteq \mathbb{R}^d$.

Therefore every measure appearing in the sequel are also supported in K, and for every stopping time τ,

$\tau = \tau \wedge \kappa$, where κ is the first time for Brownian motion to hit the boundary ∂K.

In other words, no one can escape the "universe" K. However, the assumption is only to avoid some convergence issues, and we think that it is sufficient to assume appropriate moment conditions on the marginals μ, ν according to α in (3.1.1).

3.2 Symmetrization of transport plans

We recall the notion of symmetrization of transport plans, which will be crucial for our analysis of the optimal transport plans under radially symmetric marginals. This section is the same as the first part of the section 2.6.3 for reader's convenience.

Definition 3.2.1. *Let L be a one-dimensional subspace of $\mathbb{R}^d$ and let ψ be a probability measure on $\mathbb{R}^d$. We say that ψ is L-symmetric if it is symmetric with respect to L, i.e. for any Borel set B and any orthogonal matrix M which fixes L, we have $\psi(B) = \psi(M(B))$.*

We say that the probability measures φ and ψ are L-equivalent if $\varphi(B) = \psi(B)$ for every $B \subseteq \mathbb{R}^d$ which is symmetric with respect to L, i.e. $x \in B$ implies $z \in B$ for every z with $dist(z, L) = dist(x, L)$ and $x - z \perp L$. Then we write $\varphi \cong_L \psi$. Finally, we define $L[\psi]$ to be the unique L-symmetric measure that is L-equivalent to ψ.

Now we turn to the notion of R-symmetrization. Let $S_r = \{x \,|\, |x| = r\}$ be the sphere in $\mathbb{R}^d$ with radius r, let ζ be a probability measure on S_r, and let $(\pi_x)_{x \in S_r}$ be a set of probability measures on $\mathbb{R}^d$ attached on each $x \in S_r$. We can view this as a mass transport plan π with initial mass ζ, by seeing π_x as its disintegration. We may denote this as $\pi = (\zeta, \pi_x)$. Fix a vector $w \in S_r$, and let $(M_x)_{x \in S_r}$ be a choice of orthogonal matrices with the property $M_x(x) = w$. For $u \in \mathbb{R}^d$, let $T_u(z) = z + u$ be translations. Finally, let L_x be the one-dimensional subspace spanned by $x \neq 0$.

Now consider the following probability measure

$$\sigma(\cdot) := \int_{x \in S_r} (M_x \circ T_{-x} \circ L_x[\pi_x])(\cdot)\, d\zeta(x)$$

Note that σ does not depend on the choice of $(M_x)_{x\in S_r}$, due to the presence of the operator L_x in the definition. Now we define the R- symmetrization operator acting on the transport plan $\pi = (\zeta, \pi_x)$.

Definition 3.2.2. *The R-symmetrization operator is defined by*

$$R[\pi] = R[(\zeta, \pi_x)] = (U, \sigma_x)$$

where $\sigma_x = (M_x \circ T_{-x})^{-1}(\sigma)$ and U is the uniform probability measure on S_r.

Thus, σ is an average of appropriately translated and rotated π_x's with weight ζ, and R-symmetrization operator uniformly pushes σ back on S. Now for any transport plan $\pi = (\mu, \pi_x)$ with general initial distribution μ, one can similarly apply the R-symmetrization, by applying the above R-symmetrization on each disintegration of μ along the spherical layers S_r.

Finally, we introduce the notion of R-equivalence on the space of probability measures on $\mathbb{R}^d$.

Definition 3.2.3. *Probability measures φ and ψ are called R-equivalent if they contain the same mass on any annulus, i.e. for any $B \subseteq \mathbb{R}_+$ and any $A_B := \{x \in \mathbb{R}^d \,|\, |x| \in B\}$, we have $\varphi(A_B) = \psi(A_B)$. We write $\varphi \cong_R \psi$.*

3.3 Randomized stopping times

Definition 3.3.1. *Let $C(\mathbb{R}_+) = \{\omega : \mathbb{R}_+ \to \mathbb{R}^d \mid \omega(0) = 0,\ \omega \text{ is continuous.}\}$ be the continuous path space starting at 0. Let $\Omega := C(\mathbb{R}_+) \times \mathbb{R}^d$ be the probability space, and we denote for $(\omega, x) \in \Omega$ that $(\omega, x)(t) = \omega^x(t) := x + \omega(t)$, thus the second component of Ω represents the starting point of the path. Finally, let S be the set of* *all* stopped paths

$$S = \{(f^x, s) \mid f^x : [0, s] \to \mathbb{R}^d \text{ is continuous and } f^x(0) = x\}. \tag{3.3.1}$$

Definition 3.3.2 (Randomized stopping time)**.** *(See [6, Definition 4.17]) Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be the (filtered) probability space over the path space* Ω*, equipped with the probability measure* $\mathbb{P} := \mathbb{W} \otimes \mu$*, where* $\mathbb{W}$ *is the Wiener measure and* μ *is a probability measure on* $\mathbb{R}^d$*. Consider its extension*

$$(\Omega \times [0, 1], \mathcal{F} \otimes \mathcal{B}([0, 1]), \mathbb{P} \otimes \mathcal{L}^1)$$

where, $\mathcal{B}$ *is the Borel* σ*-algebra, and* $\mathcal{L}$ *is the Lebesgue measure. A probablity measure* ξ *on* $\Omega \times \mathbb{R}_+$ *is called a* randomized stopping time *if for each* $u \in \mathbb{R}_+$ *and* $\omega^x \in \Omega$*, the random time*

$$\rho_u(\omega^x) := \inf\{t \geq 0 : \xi_{\omega^x}([0, t]) \geq u\}$$

gives a stopping time adapted to $\mathcal{F} \otimes \mathcal{B}([0, 1])$*. Here,* ξ_{ω^x} *is a disintegration of* ξ *along the path* ω^x *according to* $\mathbb{P}$*;* $\xi(d\omega, dx, dt) = \xi_{\omega^x}(dt)\mu(dx)\mathbb{W}(d\omega)$*.*

Throughout this chapter, the terminology "stopping time" means the randomized stopping time, and we will say that a stopping time is *non-randomized* if the disintegration ξ_{ω^x} is the Dirac measure on $\mathbb{R}_+$ for $\mathbb{P}$ a.e. ω^x; thus, a non-randomized stopping time is the stopping time usually described as a random variable on Ω.

Next, we introduce the *conditional randomized stopping time given* $(f^x, s) \in S$, that is, the normalized stopping measure given that we followed the path f^x up to time s. For $(f^x, s) \in S$ and $\omega \in C(\mathbb{R}_+)$, define the concatenate path $f^x \oplus \omega$ by $(f^x \oplus \omega)(t) = f^x(t)$ if $t \leq s$, and $(f^x \oplus \omega)(t) = f^x(s) + \omega(t - s)$ if $t > s$.

Definition 3.3.3 (Conditional stopping time [6])**.** *The conditional randomized stop-*

ping time of ξ, *given* $(f^x, s) \in S$, *denoted by* $\xi^{(f^x,s)}$, *is defined to be*

$$\xi_\omega^{(f^x,s)}([0,t]) := \frac{1}{1-\xi_{f^x\oplus\omega}([0,s])}\left(\xi_{f^x\oplus\omega}([0,t+s]) - \xi_{f^x\oplus\omega}([0,s])\right) \ \textit{if} \ \xi_{f^x\oplus\omega}([0,s]) < 1,$$
$$\xi_\omega^{(f^x,s)}(\{0\}) := 1 \ \textit{if} \ \xi_{f^x\oplus\omega}([0,s]) = 1.$$

According to [6], this is the normalized stopping measure of the "bush" which follows the "stub" (f^x, s). Note that $\xi_{f^x\oplus\omega}([0,s])$ does not depend on the bush ω.

3.4 Monotonicity principle and stability of $\mu \wedge \nu$

Imagine that a particle of unit mass at x starts a Brownian motion and suppose that it reaches y at time t, and continues to travel until stopping time τ; i.e. $B_t^x(\omega) = y$ and $t < \tau(\omega^x)$. Then by the strong Markov property, the particle at y will diffuse for the remaining time $\tau - t$ and thus will create a SH-measure $\psi_y = \text{Law}(B^y_{\tau-t})$. Also imagine that another unit mass particle starting at x' stops at y' under τ, i.e. $B_\tau^{x'}(\omega') = y'$. In this situation, we will use the notation

$$(x, \psi_y : x', \delta_{y'}) \in \tau \qquad (3.4.1)$$

and we define its cost

$$C(x, \psi_y : x', \delta_{y'}) = \int c(x,z)\, d\psi_y(z) + c(x', y').$$

More precise definition is the following; let $\Gamma \subseteq S$ be a concentration set of τ, i.e. $\tau(\Gamma) = 1$. Then we say that

$$(x, \psi_y : x', \delta_{y'}) \in \tau \ \text{ with respect to } \Gamma$$

if there exist (f^x, t), $(g^{x'}, t') \in \Gamma$ and $s < t$, such that $y = f^x(s)$, $y' = g^{x'}(t')$, and $\psi_y = \text{Law}(B^y(\tau^{(f^x,s)}))$.

In [6], the following theorem is proved for more general path-dependent cost.

Theorem 3.4.1 (Monotonicity principle [6]). *Suppose the cost is of the form* $c(x, y)$ *and that* τ *is an optimal stopping time for the minimization problem* (3.1.2)*. Then there exists a Borel set* $\Gamma \subseteq S$ *such that* $\tau(\Gamma) = 1$*, and* Γ *is c-monotone, in the following sense: If* $(x, \psi_y : x', \delta_y) \in \tau$ *with respect to* Γ*, then*

$$C(x, \psi_y : x', \delta_y) \leq C(x, \delta_y : x', \psi_y). \tag{3.4.2}$$

Now we introduce a variant of the theorem, which is for the case of radially symmetric marginals μ and ν.

Theorem 3.4.2 (Radial monotonicity principle). *Suppose the cost is of the form* $c(x, y) = f(|x - y|)$ *and that* τ *is an optimal stopping time for the minimization problem* (3.1.2) *where the marginals* μ, ν *are radially symmetric. Then there exists a Borel set* $\Gamma \subseteq S$ *such that* $\tau(\Gamma) = 1$*, and* Γ *is radially c-monotone, in the following sense: If* $(x, \psi_y : x', \delta_{y'}) \in \tau$ *with respect to* Γ *and* $|y| = |y'|$*, then*

$$C(x, \psi_y : x', \delta_{y'}) \leq C(x, \delta_y : x', \varphi_{y'}) \tag{3.4.3}$$

for any $\varphi_{y'} \in \mathrm{SH}(y')$ *which is R-equivalent to* ψ_y*.*

Remark 3.4.3. The radial monotonicity principle presented here is an adapted version of the principle given in [6], for our radially symmetric marginal case. Its intuitive meaning is clear: if the current optimal stopping rule τ allows a particle starting at x to diffuse when it reaches y so that it becomes the probability measure ψ_y, but takes another particle at x' to stop at y', and if we have the opposite inequality

$$C(x, \psi_y : x', \delta_{y'}) > C(x, \delta_y : x', \varphi_{y'}) \tag{3.4.4}$$

for some $\varphi_{y'} \in \mathrm{SH}(y')$, then we can "modify" the stopping time τ to τ' in such a way that the particle at x now stops at y by τ', but instead the particle at x' starts diffusing at y' until it becomes $\varphi_{y'}$. Then (3.4.4) means that the cost for τ' is smaller than that of τ, but the modified time τ' may not satisfy the terminal marginal condition ν. However, as $\psi_y \cong_R \varphi_{y'}$ and $|y| = |y'|$, τ' is also R-equivalent to τ. Hence, regarding a stopping time as a SMT plan, we can apply the symmetrization operator to τ' to get $R[\tau']$. Now since the terminal marginal ν is radially symmetric, $R[\tau']$ also solves the Skorokhod embedding problem with less cost, a contradiction (note that the symmetrization operator does not alter the cost as the cost is of the form $f(|x-y|)$). The monotonicity principle asserts the existence of a set Γ of stopped paths, which supports the optimal stopping time and resists any such modification.

Now we give a simple application of the monotonicity principle.

Theorem 3.4.4. *Let τ be a minimizer of the problem* (3.1.2) *and assume $c(x, y) = |x-y|^\alpha$ with $0 < \alpha \leq 1$. Then under τ, the common mass $\mu \wedge \nu$ stays put; i.e. on $\mu \wedge \nu$, $\tau = 0$.*

Proof. Suppose the theorem is false. Then heuristically, there is a particle at x in $\mu \wedge \nu$ which diffuses to some $\psi_x \in \mathrm{SH}(x) \setminus \{\delta_x\}$, and so another particle at y, $y \neq x$, has to flow into x and make up for the loss. Mathematically, for $\Gamma \subseteq S$ with $\tau(\Gamma) = 1$, there should exist $x, y \in \mathbb{R}^d$, $x \neq y$ and $\psi_x \in \mathrm{SH}(x) \setminus \{\delta_x\}$, such that

$$(x, \psi_x : y, \delta_x) \in \tau \text{ with respect to } \Gamma. \tag{3.4.5}$$

Now for $0 < \alpha \leq 1$ and $z \in \mathbb{R}^d$, $|x-z|^\alpha + |y-x|^\alpha \geq |y-z|^\alpha$, so

$$\int |x-z|^\alpha d\psi_x(z) + |y-x|^\alpha \geq \int |y-z|^\alpha d\psi_x(z). \tag{3.4.6}$$

In other words,

$$C(x, \psi_x : y, \delta_x) \geq C(x, \delta_x : y, \psi_x). \tag{3.4.7}$$

But the inequality is strict because $\psi_x \neq \delta_x$, hence by theorem 3.4.1, τ is not a minimizer. □

Note that, for the above result, we do not need to assume the radial symmetry of marginals μ and ν. And by the theorem, when we consider the minimization problem with $0 < \alpha \leq 1$, we can assume $\mu \wedge \nu = 0$ without loss of generality.

3.5 Variational lemma and its consequences

3.5.1 Heuristic observations

Let $S_{y,r}$ be the uniform probability measure on the sphere of center y and radius r, the most simple element in $\mathrm{SH}(y)$. We choose $c(x,y) = |x-y|$ for simplicity and observe the following "gain" function:

$$G(x) = G(x,y,r) = \int |x-z|\, dS_{y,r}(z) - |x-y|$$

and its gradient $\nabla_x G$. G is the increase of cost when the dirac mass at y diffuses uniformly onto the sphere.

1. If $|x-y| < |x'-y'|$ and r is fixed, then $G(x,y,r) > G(x',y',r)$.
   Roughly speaking, for the same diffusion, the increase of cost is greater when the distance $|x-y|$ is small. Hence, for cost minimization, we expect that it is better to stop particles near the source x, and let the particles diffuse which are far from x, as far as the given marginal condition holds.

2. $\nabla G(x)$ is toward the direction $y-x$, thus the directional derivative $\nabla_u G(x)$ is

   $$\nabla_u G(x) < 0 \quad \text{if} \quad \langle u, y-x\rangle < 0.$$

   Hence, the gain function decreases when x moves away from the "center of diffusion" y, and we may link this with monotonicity principle to tell something about optimality. This is the key idea of this chapter.

3.5.2 Variational lemma and the main theorem

From now on, $c(x,y) = |x-y|^\alpha$, $0 < \alpha \neq 2$. We define the gain function

$$G(x,\psi_y) := \int |x-z|^\alpha d\psi_y(z) - |x-y|^\alpha \quad \text{for} \quad \psi_y \in \mathrm{SH}(y), \quad \text{and note that}$$

$$G(x,\psi_y) - G(x',\varphi_{y'}) = C(x,\psi_y : x',\delta_{y'}) - C(x,\delta_y : x',\varphi_{y'}).$$

Lemma 3.5.1. *Let x, y be nonzero vectors in $\mathbb{R}^d$ with $r = |x|$. Let u be a unit tangent vector to the sphere S_r at x, such that $\langle u, y\rangle < 0$. Let $\psi_y \in \mathrm{SH}(y) \setminus \{\delta_y\}$. Then, there exists a $\varphi_y \in \mathrm{SH}(y)$ which is R-equivalent to ψ_y, such that*

$$\begin{aligned} G(x,\varphi_y) &= G(x,\psi_y) \\ \nabla_u G(x,\varphi_y) &< 0 \quad \textit{if} \quad 0<\alpha<2 \\ \nabla_u G(x,\varphi_y) &> 0 \quad \textit{if} \qquad\quad \alpha>2 \end{aligned}$$

where the directional derivative is applied to the x variable.

For proof, we first present some calculations. Let $c(x,z) = |x-z|^\alpha$. Take partial derivative $\frac{\partial}{\partial x_d}$ and put $x = 0$, then we get

$$h(z) := \nabla_{e_d} c(x,z)\big|_{x=0} = -\alpha\,|z|^{\alpha-2}\,z_d \tag{3.5.1}$$

Then we take Laplacian in z,

$$\Delta\,h(z) = -\alpha(\alpha-2)(\alpha+d-2)\,|z|^{\alpha-4}\,z_d \tag{3.5.2}$$

We see that the function h is

$$\begin{aligned} &\text{strictly superharmonic in the lower half-space } \{z_d<0\} \quad \text{if} \quad 0<\alpha<2 \\ &\text{strictly subharmonic in the lower half-space } \{z_d<0\} \quad \text{if} \qquad\quad \alpha>2. \end{aligned}$$

Proof of Lemma 3.5.1: Without loss of generality, assume $x = x_1\,e_1 = (x_1, 0, ..., 0)$ and $u = e_d$. Then $\langle u, y\rangle < 0$ means that y is in the lower half-space. Let $H = \{z \in \mathbb{R}^d : z_d = 0\}$ and $0 < \alpha < 2$.

Choose a stopping time τ for Brownian motion B^y such that $\mathrm{Law}(B^y_\tau) = \psi_y$, and let η be the first time for B^y to hit H. Let $\sigma_y = \mathrm{Law}(B^y_{\tau\wedge\eta})$. Then σ_y is supported in the lower half-space, and it is nontrivial, hence by the strict subharmonicity (3.5.2),

$$\nabla_u G(x,\sigma_y) < 0.$$

Now we modify τ to τ' in the following way; if $\tau \leq \eta$, then we let $\tau' = \tau$. But if $\tau > \eta$, in other words if a particle at y has landed on H but not completely stopped by τ, then we symmetrize the remaining time of τ (i.e. the conditional stopping time) with respect to H and get τ'. More precisely, let τ_H be the reflection of the conditional stopping time of τ w.r.t. H; if τ stops a path emanating from H, then τ_H stops the reflected path at the same time. Now define $\tau' := \frac{\tau+\tau_H}{2}$ to be a randomization; before re-starting Brownian motion on H, we flip a coin and choose either τ or τ_H.

Now, define $\varphi_y = \text{Law}(B^y_{\tau'})$ and observe

1. $\varphi_y \cong_R \psi_y$ by the symmetry with respect to H in the definition of τ'.

2. $\nabla_u G(x, \varphi_y) = \nabla_u G(x, \sigma_y)$, since the function $z \mapsto \nabla_{e_d} c(x, z)$ is odd in z_d (see (3.5.2)) hence the symmetrization in the definition of τ' does not change $\nabla_u G$.

Hence, the lemma follows. □

Next, we prove the following consequence of the variational lemma.

Theorem 3.5.2. *Let x_0, x_1, y be nonzero vectors in $\mathbb{R}^d$ and let $r = |x_0| = |x_1|$. Assume $|x_0 - y| < |x_1 - y|$. Fix a measure $\varphi_y \in \text{SH}(y) \setminus \{\delta_y\}$ and define*

$$\underline{\Re}(x, \varphi_y) := \{\psi_y \in \text{SH}(y) : \psi_y \cong_R \varphi_y \text{ and } \psi_y \text{ is a solution of } Min \int |x - z|^\alpha d\psi_y(z)\},$$

$$\overline{\Re}(x, \varphi_y) := \{\psi_y \in \text{SH}(y) : \psi_y \cong_R \varphi_y \text{ and } \psi_y \text{ is a solution of } Max \int |x - z|^\alpha d\psi_y(z)\},$$

$$\underline{G}(x) := G(x, \psi_y) = \int |x - z|^\alpha d\psi_y(z) - |x - y|^\alpha \quad \text{for any} \quad \psi_y \in \underline{\Re}(x, \varphi_y),$$

$$\overline{G}(x) := G(x, \psi_y) = \int |x - z|^\alpha d\psi_y(z) - |x - y|^\alpha \quad \text{for any} \quad \psi_y \in \overline{\Re}(x, \varphi_y).$$

Then we have

$$\underline{G}(x_0) > \underline{G}(x_1) \quad \text{and} \quad \overline{G}(x_0) < \overline{G}(x_1) \quad \text{if} \quad 0 < \alpha < 2$$

$$\underline{G}(x_0) < \underline{G}(x_1) \quad \text{and} \quad \overline{G}(x_0) > \overline{G}(x_1) \quad \text{if} \quad \alpha > 2.$$

Proof. First, we claim that

$$\underline{G}(x) \text{ and } \overline{G}(x) \text{ are continuous.}$$

The proof of this plausible fact is given in the appendix. Now let $0 < \alpha < 2$, and chooose any differentiable curve $x(t) : [0,1] \to S_r$ with $x(0) = x_0$, $x(1) = x_1$ and $|\,x(t) - y\,|$ is strictly increasing. In other words, we choose $x(t)$ in such a way that $\langle \frac{d}{dt}x(t), y\rangle < 0$ for all t. Now for a fixed $t \in [0,1]$, choose any $\psi_y \in \underline{\Re}(x(t), \varphi_y)$. Then Lemma 3.5.1 gives $\sigma_y \in \underline{\Re}(x(t), \varphi_y)$ with $\frac{d}{dt}G(x(t), \sigma_y) < 0$. By definition, $\underline{G}(x(s)) \leq G(x(s), \sigma_y)$ for any s, and $\underline{G}(x(t)) = G(x(t), \sigma_y)$. Hence

$$\begin{aligned}\limsup_{\varepsilon\downarrow 0} \frac{\underline{G}(x(t+\varepsilon)) - \underline{G}(x(t))}{\varepsilon} &\leq \limsup_{\varepsilon\downarrow 0} \frac{G(x(t+\varepsilon), \sigma_y) - G(x(t), \sigma_y)}{\varepsilon} \\ &= \frac{d}{dt}G(x(t), \sigma_y) < 0.\end{aligned}$$

We showed that the function $\underline{G}(x(t))$ is continuous and has the above strict inequality for each $t \in [0,1]$, hence it must be strictly decreasing. Therefore, the theorem follows. The other three cases can be shown in the same way. □

Finally, armed with theorem 3.5.2, we establish the main theorem.

Theorem 3.5.3. *Suppose μ, ν are radially symmetric probability measures on $\mathbb{R}^d$, $d \geq 2$, and $\mu(\{0\}) = 0$. Let $c(x,y) = |x-y|^\alpha$, where $0 < \alpha \neq 2$. Then the solution for the optimization problem* (3.1.2) *under the marginals μ, ν is unique and it is given by a non-randomized stopping time.*

Proof. Let τ be a minimizer for (3.1.2). The proof consists of two parts. The first part claims that there exists a $\Gamma \subseteq S$ on which τ is concentrated, such that Γ does not admit a "forbidden pair", an element in $\Gamma \times \Gamma$. Then the second part claims that if Γ does not allow such forbidden pair, then every stopping time concentrated on Γ is necessarily of non-randomized type. This will yield uniqueness immediately.

We begin the first part. For $x, y \neq 0$ in $\mathbb{R}^d$, define the following barrier set:

$$U_x^y = \{z \in \mathbb{R}^d : |z| = |y| \text{ and } \langle x, y\rangle < \langle x, z\rangle\}.$$

Now fix $0 < \alpha < 2$. We then describe the forbidden pairs:

$$\mathrm{FP} = \{((f,s),(g,t)) \in S \times S : f(0) = g(0) \neq 0 \text{ and } \exists s' < s \text{ such that } f(s') \in U_{g(0)}^{g(t)}\}.$$

Now choose a radially c-monotone Γ for τ as in theorem 3.4.2. Notice that theorem 3.5.2 and 3.4.2 imply that there must be no forbidden pair in Γ, i.e. $\mathrm{FP} \cap (\Gamma \times \Gamma) = \emptyset$.

Now the second step relies on the oscillation property of Brownian motion. Let Γ be such that $\mathrm{FP} \cap (\Gamma \times \Gamma) = \emptyset$ and ξ be a stopping time with $\xi(\Gamma) = 1$. Then we claim that ξ must be of non-randomized type.

Suppose not. Then there exists an element $(f^x, s) \in \Gamma$ such that the conditional stopping time $\xi^{(f^x,s)}$ is nonzero. This means that the Brownian motion which has followed the path f^x up to time s will continue its motion at $y := f^x(s)$. Now consider the barrier $U := U_x^y$ and note that, Brownian motion starting at y under any non-zero stopping time will go through the surface U before its complete stop, since y is on the boundary circle of U. This implies that there is a stopped path $(g^y, t) \in S$ such that $(f^x \oplus g^y, s+t) \in \Gamma$ and for some $s' < s+t$, $(f^x \oplus g^y)(s') \in U$. Now notice that the pair $((f^x \oplus g^y, s+t), (f^x, s)) \in \Gamma \times \Gamma$ is forbidden, a contradiction.

Therefore, every minimizer is of non-randomized type, and uniqueness follows in the usual way; let τ and τ' be two minimizers. If their disintegrations do not agree, i.e. $\tau_{\omega^x} \neq \tau'_{\omega^x}$ for all $\omega^x \in B$ with $\mathbb{P}(B) > 0$, then the stopping time $\frac{\tau+\tau'}{2}$ must be of randomized type but it is obviously a minimizer, a contradiction. □

Remark 3.5.4. Let Γ be the radial c-monotone set on which the optimizer in Theorem 3.5.3 is concentrated. The proof of the above theorem in fact tells us that, for minimization problem with $0 < \alpha < 2$ or maximization problem with $\alpha > 2$, the optimal stopping time applied on the Brownian motion B^x is given by the first hitting time of the following union of barriers

$$\mathcal{U}_x := \cup_y U_x^y, \text{ where } y = f^x(s) \text{ for some } (f^x, s) \in \Gamma$$

and by uniqueness and the radial symmetry of μ and ν, the $\mathcal{U}_x$'s are congruent under rotation; if M is an orthogonal matrix and $M(x) = x'$, then $M(\mathcal{U}_x) = \mathcal{U}_{x'}$.

For minimization problem with $\alpha > 2$ or maximization problem with $0 < \alpha < 2$, we have the same type of result, but the barrier will be reversed: it will be given by

$$\mathcal{V}_x := \cup_y V_x^y, \text{ where } y = f^x(s) \text{ for some } (f^x, s) \in \Gamma$$

where V_x^y is the reversed barrier: $V_x^y = \{z \in \mathbb{R}^d : |z| = |y| \text{ and } \langle x, y \rangle > \langle x, z \rangle\}$.

This is due to the interchange of the superharmonic and subharmonic region of the derivative of the cost function (3.5.1), according to α. Also note that when dealing with maximization problem, the inequalities (3.4.2) and (3.4.3) in the monotonicity principles must be reversed.

Chapter 4

Conclusion

In this chapter, we summarize the contributions of this thesis and discuss directions for future work.

4.1 Contributions

In chapter 2, we study the martingale optimal transport problem (2.1.1), especially when the marginals μ, ν lie on the higher dimensional spaces $\mathbb{R}^d, d \geq 2$. First, we notice that when the dual problem attains an optimal solution, we could study the dual solution and conclude that the conjecture 1 holds true. But in that proof, we need regularity of the dual solution for the variational argument to work. Thus, our first major contribution is to verify such regularity improvement whenever a triple (α, β, γ) solves the dual problem.

Then the question is whether the dual problem always attains a solution or not. Unlike the Monge-Kantorovich mass transport theory, it is shown that the martingale transport problem does not always attain a solution for the dual, revealing the deep aspect of the problem. In regards to this matter, our second major contribution is to develop a general theory of dual attainment, which enables us to decompose a given optimal martingale transport and to study each piece.

However, this partial dual attainment result still does not fully resolve the conjecture 1, due to the presence of certain factors such as, the Nikodym set. Therefore, the conjecture 1 still remains open, especially in dimension $d \geq 3$, and this reveals the fact that the martingale transport problem in higher dimensions is connected to one of the deepest topics in analysis.

Towards conjecture 2, even under the assumption of full dual attainment, we do not see a general argument which leads to its resolution, making the conjecture even deeper. We find an example which shows that the conjecture 2 does not hold for maximization problem, even under radial symmetric marginals. But with the additional assumption of discreteness of the target ν, we could resolve the conjecture for both maximization / minimization problem. Another case where we could show that conjecture 2 is affirmative is when the marginals μ, ν are radially symmetric, only in the case of minimization. Here the sharp distinction between the solutions for maximization and minimization problem emerges, and this is the reason why we guess that the conjecture 2 would hold only for the minimization problem. We should mention that even in dimension 2, the conjecture has not been fully solved.

In this thesis, we also define the Subharmonic martingale transport problem as well as we notice that the problem is in fact equivalent to the Skorokhod embedding problem which is one of the most important problems in probability theory. Inspired by the previous result in martingale transport, we tackle the Skorokhod problem in higher dimensions and in case of marginals being assumed to be radially symmetric. Instead of appealing to the dual problem, we rather develop a direct method which involves fine properties of Brownian motion and nicely exploits the radial symmetry of the marginals. It eventually leads to the conclusion saying that every optimal embedding time must exhibit a certain extremal property which can be seen as the Brenier-type result in the case of Skorokhod embedding. Now we see that the central ideas in two important areas in mathematics, namely the mass transport theory in analysis and the Skorokhod embedding theory in probability, culminate in meeting at their peaks, and this harmony has not been fully studied yet.

4.2 Directions for future work

My further research interest lies on the multi-step martingale transport problem; even in one-dimensional setting, not much is known for the following problem; see, e.g., [11], [23], [24], [29]. Continued from the Conjecture 1 and 2, we propose the following problems:

Problem 3 (Multi-marginal case) *If a martingale* $(X_1, X_2, ..., X_n)$ *optimizes* $\mathbb{E}_{\mathbb{P}}\, c(X_1, X_2, ..., X_n)$ *where* $c(x_1, ..., x_n) : \mathbb{R}^d \times ... \times \mathbb{R}^d \to \mathbb{R}$ *is an appropriate path-dependent cost function and the marginals* $X_i \sim \mu_i$ *are known, then what can one say on the structure of the martingale* $(X_1, X_2, ..., X_n)$*?*

Even further, I am interested in the case where a continuum of marginals is imposed. We note that recently the classical optimal transport problem with the multiple or continuum marginals has been intensively studied by Gangbo, Swiech, Carlier, Ekeland, Agueh, Pass, Ghoussoub, Kim and etc, but the corresponding martingale transport problem is wide open:

Problem 4 (Continuum-marginal case) *If a continuum of probabilities* $(\mu_t)_{t\in[0,1]}$ *(having the right convex order) is given, then what can one say on the structure of the martingale* $(X_t)_{t\in[0,1]}$*, in case in which it solves the optimization problem with respect to an appropriate path-dependent cost function, and has* $(\mu_t)_{t\in[0,1]}$ *as its marginal?*

Meanwhile, many important results are kept hidden in the Skorokhod embedding problem, especially in higher dimensions. Inspired by the result in Chapter 3, we are working on the following more ambitious problem:

Problem 5: *Without assuming any symmetry on the marginals* μ, ν *in* $\mathbb{R}^d$*, every optimal stopping time is of non-randomized type (i.e. the stopping rule can be solely determined by the Brownian path itself). Furthermore, the optimal time is unique.*

Even further, while working on the above projects, we see that a general unifying theory can be built and each extremal property in the above problems can be seen as special cases. One of the most important cases should be the *Plurisubharmonic Martingale Transport Problem*, where the defining constraint (1.2.1) holds with respect to the cone of plurisubharmonic functions on $\mathbb{C}^d$, which plays a key role in the theory of several complex variables. Then, we ask:

Problem 6: *If π is an optimal plurisubharmonic martingale transport with respect to an appropriate cost functional, then what kind of extremal property must it exhibit?*

Problem 7: *What is an equivalent formulation of the plurisubharmonic martingale transport in terms of the Brownian motion, and what novelty would it tell us on the theory of complex variables?*

Final remark is that these optimal martingale problems do not need to be considered only on $\mathbb{R}^d$, but on other interesting spaces as well; for example, on the Riemannian manifolds and even on the infinite-dimensional spaces. Uncountable interesting results should be awaiting for someone to discover them.

Bibliography

[1] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. *Optimal transportation and applications. Springer Berlin Heidelberg.*, Volume 1813 of the series Lecture Notes in Mathematics (2003) 123-160.

[2] L. Ambrosio, B. Kirchheim, and A. Pratelli. Existence of optimal transport maps for crystalline norms. *Duke Mathematical Journal.*,Volume 125, Number 2 (2004) 207-241.

[3] M. Beiglböck, P. Henry-Labordere, and F. Penkner. Model-independent bounds for option prices: A linear programming approach. *Finance and Stochastics.*, 17 (2013) 477-501.

[4] M. Beiglböck, P. Henry-Labordere, and N. Touzi. Monotone martingale transport plans and Skorohod Embedding. Preprint (2015).

[5] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. http://arxiv.org/abs/1208.1509 (2014).

[6] M. Beiglböck, A.M.G. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. http://arxiv.org/abs/1307.3656, 2014.

[7] S. Bianchini and F. Cavalletti. The Monge Problem for Distance Cost in Geodesic Spaces. *Comm. Math. Phys.*, Volume 318, Issue 3 (2013) 615-673.

[8] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, Volume 44, Issue 4 (1991) 375-417.

[9] S. Bu and W. Schachermayer. Approximation of Jensen measures by image measures under holomorphic functions and applications. *Trans. Amer. Math. Soc.*, 331 (1992), 585-608.

[10] G. Choquet. Forme abstraite du théorème de capacitabilité. *Ann. Inst. Fourier. Grenoble*, 9 (1959) 83–89.

[11] A.M.G. Cox, J. Obłój, and N. Touzi. The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. http://arxiv.org/abs/1505.03169, (2015).

[12] Y. Dolinsky and H.M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields.*, 160 (2014) 391-427.

[13] Y. Dolinsky and H.M. Soner. Martingale optimal transport in the Skorokhod space. *Stochastic Processes and their Applications.*, 125(10) (2015) 3893-3931.

[14] G.A. Edgar. Complex martingale convergence. *Banach Spaces, Volume 1166 of the series Lecture Notes in Mathematics.*, 38-59.

[15] G.A. Edgar. Analytic martingale convergence. *Journal of Functional Analysis.*, Volume 69, Issue 2, November 1986, 268-280.

[16] P. Fischer and J. A. R. Holbrook. Balayage defined by the nonnegative convex functions. *Proc. Amer. Math. Soc.*, 79(3) (1980) 445-448.

[17] A. Galichon, P. Henry-Labordere, and N. Touzi. A Stochastic Control Approach to No-Arbitrage Bounds Given Marginals, with an Application to Lookback Options. *Annals of Applied Probability.*, Volume 24, Number 1 (2014) 312-336.

[18] T. W. Gamelin. *Uniform Algebras And Jensen Measures*, Cambridge University Press, 1979.

[19] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, Volume 177, Issue 2 (1996) 113-161.

[20] N. Ghoussoub, Y-H. Kim, and T. Lim. Optimal skorokhod embedding between radially symmetric marginals in general dimensions. Preprint (2015).

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

[22] N. Ghoussoub and B. Maurey. Plurisubharmonic martingales and barriers in complex quasi-Banach spaces. *Annales de l'institut Fourier.*, Volume 39, Issue 4 (1989)1007-1060.

[23] G. Guo, X. Tan, and N. Touzi. Optimal Skorokhod embedding under finitely-many marginal constraints. http://arxiv.org/abs/1506.04063, (2015).

[24] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series.* Springer, Milan (2011).

[25] D. Hobson. Robust hedging of the lookback option. *Finance and Stochastics.*, 2 (1998) 329-347.

[26] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, Springer, Berlin (2011) 267-318.

[27] D. Hobson and M. Klimmek. , Robust price bounds for the forward starting straddle. *Finance and Stochastics.*, Volume 19, Issue 1 (2014) 189-214.

[28] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Mathematical Finance.*, Volume 22, Issue 1 (2012) 31-56..

[29] S. Kallblad, X. Tan, and N. Touzi. Optimal Skorokhod embedding given full marginals and Azema-Yor peacocks. http://arxiv.org/abs/1503.00500, (2015).

[30] H. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4) (1984) 399-432.

[31] T. Lim. Optimal martingale transport between radially symmetric marginals in general dimensions. Preprint, 2015.

[32] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*, 1781.

[33] R.T. Rockafellar Characterization of the subdifferentials of convex functions. *Pacific J. Math*, 17 (1966) 497-510.

[34] L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32(1) (1990) 48-54.

[35] L. Rüschendorf and L. Uckelmann. Numerical and analytical results for the transportation problem of Monge-Kantorovich. *Metrika*, 51(3) (electronic) (2000) 245-258.

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

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

[38] C. Villani. *Optimal Transport. Old and New*, Vol. 338, *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

[39] D. Zaev. On the Monge-Kantorovich problem with additional linear constraints. *http://arxiv.org/abs/1404.4962v1*, 2014.

Appendix A

A proof of Proposition 3.1.3

Here, we prove the following proposition.

Proposition A.0.1. *Let* $\psi \in \mathrm{SH}(x)$ *and suppose* $\mathrm{supp}(\psi)$ *is compact. Then there exists a stopping time* τ *such that* $\psi = \mathrm{Law}(B^x_\tau)$.

We introduce the notion of spherical martingale.

Definition A.0.2. *Let* U *be the uniform probability measure on the unit sphere in* $\mathbb{R}^d$. *Let* $(X_i)_{i=1}^\infty$ *be i.i.d random variables on some probability space* $(\Omega, \mathbb{P})$, *whose distribution is* U. *We define the spherical martingales as follows:*

$$F_0 = x \in \mathbb{R}^d \tag{A.0.1}$$

$$F_n(X_1, ..., X_n) - F_{n-1}(X_1, ..., X_{n-1}) = r_n(X_1, ..., X_{n-1}) \cdot X_n. \tag{A.0.2}$$

Thus, at each time n, a particle splits uniformly to the sphere of radius r_n. (Compare this definition with the analytic martingales in [9].) A simple but important observation is that, the push-forwarded measure of δ_x by spherical martingale F_n can be exactly obtained by $\mathrm{Law}(B^x_\tau)$, where τ is defined as follows: τ_1 is the first hitting time of B^x on the sphere $S(x, r_1)$, the sphere of center x and radius r_1. Now $B_{\tau_1}(\omega) = x_1 \in S(x, r_1)$, and define τ_2 as the first hitting time of B^{x_1} on the sphere $S(x_1, r_2)$. We inductively define $\tau_1 \leq \tau_2 \leq ...$ and we see that $F_n(\mathbb{P}) = \mathrm{Law}(B^x_{\tau_n})$. Now we reproduce the proof of proposition 2.1 in [9] for our spherical martingale.

Proposition A.0.3. *Let O be an open set in $\mathbb{R}^d$ and $f : O \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function. Define $f_0 = f$ and for $n \geq 1$*

$$f_n(x) = inf \left\{ \int f_{n-1}(x + ry)\, dU(y) \right\}$$

where the infimum is taken over all $r \geq 0$ such that $\{x + r\overline{B}\} \subseteq O$. Then $(f_n)_{n=0}^{\infty}$ decreases pointwise to the largest subharmonic function $\hat{f}$ on O dominated by f.

Proof. The proof here follows the same way as in [9] but we provide it here for completeness. $(f_n)_{n=0}^{\infty}$ obviously decreases. We show upper-semicontinuity of $\hat{f}$. Suppose f_{n-1} is upper semicontinuous and let $(x_k)_{k=0}^{\infty}$ in O be such that $lim_{k\to\infty} x_k = x_0$. If $r \geq 0$ is such that $\{x_0 + r\overline{B}\} \subseteq O$ where $\overline{B}$ is the closed unit ball, then there is k_0 such that $\{x_k + r\overline{B}\} \subseteq O$ for $k \geq k_0$. The upper semicontinuous function f_{n-1} is bounded above on the relatively compact set $\cup_{k=k_0}^{\infty} \{x_k + r\overline{B}\}$ and, for every $z \in \overline{B}$, $f_{n-1}(x_0 + rz) \geq \limsup_{k\to\infty} f_{n-1}(x_k + rz)$. Hence by Fatou's lemma,

$$\int f_{n-1}(x_0 + ry)\, dU(y) \geq \limsup_{k\to\infty} \int f_{n-1}(x_k + ry)\, dU(y) \geq \limsup_{k\to\infty} f_n(x_k).$$

Thus $f_n(x_0) \geq \limsup_{k\to\infty}(x_k)$, showing upper-semicontinuity of f_n and hence of f. Let g be subharmonic on O with $g \leq f$. Suppose $g \leq f_{n-1}$. Then for $\{x_0 + r\overline{B}\} \subseteq O$,

$$\int f_{n-1}(x_0 + ry)\, dU(y) \geq \int g(x_0 + ry)\, dU(y) \ \geq \ g(x_0)$$

so $f_n(x_0) \geq g(x_0)$, hence $\hat{f} \geq g$. Finally, for $\{x_0 + r\overline{B}\} \subseteq O$, by monotone convergence

$$\hat{f}(x_0) = \lim_{n\to\infty} f_n(x_0) \leq \lim_{n\to\infty} \int f_{n-1}(x_0 + ry)\, dU(y) = \int \hat{f}(x_0 + ry)\, dU(y).$$

This shows that $\hat{f}$ is subharmonic, and the proof is complete. □

By following G.A. Edgar [14], one can write the definition of f_n as the following:

(A.0.3)
$$f_n(x) = \inf\{\mathbb{E}[f(F_n)] : (F_i)_{i=0}^n \text{ is a spherical martingale valued in } O \text{ with } F_0 = x\}.$$

Now the following theorem and its proof is a close analogue of theorem (A) in [9].

Theorem A.0.4. *Any* SH-*measure in* $\mathbb{R}^d$ *can be approximated by the push-forward of spherical martingales.*

Proof. Let $\psi \in \mathrm{SH}(x)$, and define the set of push-forwarded measures

$$\Phi = \{F_n(\mathbb{P}) : (F_i)_{i=0}^n \text{ is a spherical martingale with } F_0 = x\}.$$

Then Φ is convex: let $(F_i')_{i=0}^n$ and $(F_i'')_{i=0}^m$ be two spherical martingales, and we may assume $n = m$. Define a spherical martingale $(F_i)_{i=0}^{n+1}$ by letting $F_0 = F_1 = x$ and for $1 \leq i \leq n$,

$$F_i(X_1, X_2, ..., X_{i+1}) = \begin{cases} F_i'(X_2, ..., X_{i+1}) & \text{if } X_1 \text{ is in the upper hemisphere} \\ F_i''(X_2, ..., X_{i+1}) & \text{if } X_1 \text{ is in the lower hemisphere.} \end{cases}$$

Then clearly $F_{n+1}(\mathbb{P}) = \{F_n'(\mathbb{P}) + F_n''(\mathbb{P})\}/2$, hence Φ is convex.

Now if the theorem is false, then by Hahn-Banach theorem we could find a continuous and bounded function f on O and real numbers $a < b$ such that

$$\int f\, d\psi \leq a, \text{ while } \int f \circ F_n\, d\mathbb{P} \geq b \text{ for every } F_n(\mathbb{P}) \in \Phi.$$

Hence by (A.0.3) we have $\hat{f}(x) \geq b$, but then $a \geq \int f\, d\psi \geq \int \hat{f}\, d\psi \geq \hat{f}(x) \geq b$, a contradiction. □

Proof of Proposition A.0.1: For $\psi \in \mathrm{SH}(x)$, we have a sequence $\{\psi_n\} \subseteq \Phi$ such that $\psi_n \to \psi$. We know that $\psi_n = \mathrm{Law}(B^x_{\tau_n})$ for a sequence of stopping times τ_n. Since $\mathrm{supp}(\psi)$ is compact, $\mathrm{Law}(B^x_{\tau_n})$ are also compactly supported, hence in particular $\mathbb{E}\tau_n = \mathbb{E}|B^x_{\tau_n}|^2 \leq V$ for some constant V. Then by following the proof of theorem 4.28 in [6], we see that the sequence (τ_n) is tight, hence by Prokhorov's theorem we have a subsequence (τ_k) which converges to a (randomized) stopping time τ. Now let f be a continuous and bounded function. Then the function $(\omega, t) \to (f \circ B^x)(\omega, t)$ is continuous and bounded on $C(\mathbb{R}_+) \times \mathbb{R}_+$, hence $\mathbb{E}f(B^x_{\tau_k}) \to \mathbb{E}f(B^x_\tau)$, which means that $\mathrm{Law}(B^x_{\tau_k}) \to \mathrm{Law}(B^x_\tau)$. Therefore, $\mathrm{Law}(B^x_\tau) = \psi$. $\square$

Appendix B

Continuity of the functions in Theorem 3.5.2

Here, we show the continuity of the functions in Theorem 3.5.2.

Lemma B.0.1. *The functions $\underline{G}(x)$ and $\overline{G}(x)$ in Theorem 3.5.2 are continuous.*

Proof. We will use the following:
$(*)$ Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers. Then $\{a_n\}$ converges to a if and only if each subsequence $\{a_k\}$ of $\{a_n\}$ has a further subsequence $\{a_l\}$ that converges to a.

Now let $x_n \to x$ in $\mathbb{R}^d$, and define

$$\underline{C}(x) := \underline{G}(x) + |x-y|^\alpha = \int |x-z|^\alpha d\psi_y(z) \quad \text{for any} \quad \psi_y \in \underline{\Re}(x, \varphi_y).$$

Let $a_n = \underline{C}(x_n)$, $a = \underline{C}(x)$. We need to show that $a_n \to a$ as $n \to \infty$.

Choose any subsequence $\{a_k\}$ of $\{a_n\}$, and choose a corresponding sequence of measures $\psi_k \in \underline{\Re}(x_k, \varphi_y)$. By compactness, $\{\psi_k\}$ has a subsequence $\{\psi_l\}$ which converges to, say ψ. Note that $\psi \in \mathrm{SH}(y)$ and $\psi \cong_R \varphi_y$ by weak convergence. Now

$$\begin{aligned} &\int |x_l - z|^\alpha d\psi_l(z) - \int |x-z|^\alpha d\psi(z) \\ &= \left[\int \left(|x_l - z|^\alpha - |x-z|^\alpha\right) d\psi_l(z)\right] + \left[\int |x-z|^\alpha d\psi_l(z) - \int |x-z|^\alpha d\psi(z)\right]. \end{aligned}$$

The first bracket goes to zero as $l \to \infty$ since $|x_l - z|^\alpha - |x - z|^\alpha \to 0$ uniformly on every compact set in $\mathbb{R}^d$, and the second bracket goes to zero since $\psi_l \to \psi$.

Now we claim that

$$\int |x - z|^\alpha d\psi(z) = \underline{C}(x) = a, \quad \text{i.e.} \quad \psi \in \underline{\Re}(x, \varphi_y). \tag{B.0.1}$$

If not, then there exists a $\rho \in \underline{\Re}(x, \varphi_y)$ such that

$$\int |x - z|^\alpha d\psi(z) > \int |x - z|^\alpha d\rho(z), \text{ hence}$$
$$\int |x_l - z|^\alpha d\psi_l(z) > \int |x_l - z|^\alpha d\rho(z) \text{ for all large } l,$$

a contradiction since $\psi_l \in \underline{\Re}(x_l, \varphi_y)$. Therefore, $a_n \to a$ and the lemma follows. $\square$