Università degli studi di Pavia
Joint Ph.D. program in Mathematics Milano Bicocca - Pavia - INdAM

Doctoral Thesis

Efficient methods for Discrete Optimal Transport

Author:
Gennaro Auricchio

Supervisor:
Prof. Marco Veneroni
Prof. Stefano Gualandi

Contents

Chapter 1

Introduction

In 1781 Gaspard Monge published the work "*Mémoire sur la théorie des déblais et remblais*" [61], in which he studied the following problem: how to find the optimal way to move a starting configuration to a final one. As the title of the work suggests, the work was inspired by the construction of structures with material extracted from the soil.
The distribution of materials and the final location in which we need to move them can be represented from two probability measures μ and ν in a suitable space $\mathbb{R}^n$. In this framework, the Monge problem is to find an assignation rule $T : \mathbb{R}^n \to \mathbb{R}^n$ which has to satisfy the following transportation constraint

$$\mu(T^{-1}(A)) = \nu(A)$$

for each Borel set $A \subset \mathbb{R}^n$. Even when there exists at least a function T satisfying this constraint, it is not generally unique. For this reason, Monge introduced a cost function that helps to understand which rules are cheaper or more convenient than others. In his work, this cost was equal to the mass of the moved object multiplied by the length of the path along which it moved. With a slight generalization, we can pick any measurable function $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and express the total cost as

$$\int_{\mathbb{R}^n} c(x, T(x)) d\mu.$$

As we will see, in this form, the problem is quite hard to handle and can present many issues, as, for example, the lack of a minimizer.

In the twentieth century, through an independent work, Kantorovich proposed and studied in [41, 39] the following minimum problem: given two probability measures on $\mathbb{R}^n$, namely μ and ν, and a cost function $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty]$, we want to minimize the pay-off functional

$$\int_{\mathbb{R}^n} c(x, y) d\pi(x, y)$$

over all the measures π over $\mathbb{R}^n \times \mathbb{R}^n$ whose first marginal is equal to μ and the second one to ν, which means that, for any $A, B \subset \mathbb{R}^n$ it holds true

$$\pi(A \times \mathbb{R}^n) = \mu(A) \qquad \text{and} \qquad \pi(\mathbb{R}^n \times B) = \nu(B).$$

A measure satisfying these relations is also called a transportation plan between μ and ν. Describing the movements of mass through probability measures rather than with functions allows us to generalize the problem previously proposed by Monge (as Kantorovich itself notices in [40]) and to set the problem in a simpler framework, where we can guarantee the existence of a minimizer. Moreover, when the cost function c is a distance, the above minimal problem induces a distance over the space of probability measures over $\mathbb{R}^n$.

For this reason, throughout the second half of 20^{-th} century, the optimal transportation problem has been a topic of major interest for probabilists and statisticians. In this period the theory has only been applied to the study of systems of particles, by Dobrushin [31] and to the Boltzmann equation, by Tanaka in [86, 85, 63].

At the end of the eighties, however, this branch of interest started thriving further.

John Mather, through the study of Lagrangian dynamical systems in [56, 57], was able to provide a new point of view on the Kantorovich-Monge problem. He formulated a variational problem on the space of measures, which generalizes the classical action minimizing curves problem. This study gave a fruitful insight into the structure of the solution of Monge problem.

In the field of fluidodynamics, Yann Brenier studied the set of mass-preserving maps and defined a projection over this set by choosing the map that minimizes the transportation functional, i.e. the minimal map of the Monge transportation problem [19]. This was the starting point of research that brought to light an important link between optimal transport and fluid mechanics and, in particular, the theory of Monge-Ampere equations. The growing taste for this subject started spreading also outside mathematics: Mike Cullen [33], who was a meteorologist, was able to successfully apply it to explain a famous change of unknowns discovered by Brian Hoskins [37].

All those results pointed out that optimal transport could be applied to differential equations across different applied fields and, in particular, that by a qualitative description of the optimal transport it is possible to gain insightful information on many open problems. This attracted various mathematicians, willing of giving a better description of the underlying structure of the optimal transportation problem. Among the many, we recall Felix Otto, who formulated an appealing formalism who lead to a different point of view, that opened the way to a more geometric description of the space of probability measures [38].

In recent years, optimal transport has been used to compute the similarity (or dissimilarity) between pairs of objects, such as images. This is a crucial subproblem in several applications in Computer Vision [69, 72, 74], Computational Statistic [50], Probability [9, 7], and Machine Learning [5, 83, 34, 28]. The optimization problem that yields the Kantorovich-Wasserstein distance can be solved with different methods. Nowadays, the most popular methods are based on (i) the Sinkhorn's algorithm [36, 3, 27, 82], which solves a regularized version of the basic optimal transport problem, and (ii) Linear Programming-based algorithms [32, 51], which exactly solve the basic optimal transport problem by formulating and solving an equivalent uncapacitated minimum cost flow problem.

Our work aims to provide new and efficient ways to compute the Kantorovich-Wasserstein distances. We achieve that by exploiting the geometric structure induced by the chosen cost function. We focus our investigation on two different classes of cost functions and two different frameworks.

In the first one we specialize to compare two measures whose support is a regular grid (as is the case for images) when the cost function is separable. A cost function is said to be separable if it can be expressed as the sum of two independent pieces. On $\mathbb{R}^2$, this means

$$\begin{aligned} c : \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R} \\ c(\mathbf{x}, \mathbf{y}) &= c_1(x_1, y_1) + c_2(x_2, y_2), \end{aligned}$$

where $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$. The most famous ground distances meeting this condition are the cost functions induced by l^p-norms, which are

$$c_p(\mathbf{x}, \mathbf{y}) = (|x_1 - y_1|)^p + (|x_2 - y_2|)^p. \tag{1.1}$$

Our main result shows that, for this class of functions, it is possible to compute the Wasserstein distance through another and more concise minimal problem. We will move the attention from the transportation plans to the cardinal flows. Given two measures μ and ν over $\mathbb{R}^2$, a couple $(f^{(1)}, f^{(2)})$ of probability measures over $\mathbb{R}^3$ is said to be a cardinal flow between μ and ν if

$$\int_{\mathbb{R}} f^{(1)}(x_1, x_2, y_1)dy_1 = \mu(x_1, x_2), \qquad \int_{\mathbb{R}} f^{(2)}(x_2, y_1, y_2)dy_2 = \nu(y_1, y_2),$$
$$\int_{\mathbb{R}} f^{(1)}(x_1, x_2, y_1)dx_1 = \int_{\mathbb{R}} f^{(2)}(x_2, y_1, y_2)dy_2.$$

Roughly speaking, one can think of the cardinal flow as a decomposition in two steps of a transportation plan: the measure $f^{(1)}$ describes the horizontal movements, while $f^{(2)}$ describes the vertical one. It is worthy of notice that according to this description, the order in which we have to commit those operations cannot be swapped: the first cardinal flow has to act before the second cardinal flow $f^{(2)}$. According to this interpretation, the total effort we need to invest in order to perform this couple of transformations is given by

$$\mathbb{CT}_c(f^{(1)}, f^{(2)}) = \int_{\mathbb{R}^3} c_1 df^{(1)} + \int_{\mathbb{R}^3} c_2 df^{(2)}.$$

We can then prove the key result of this argument, which states that the search for the cardinal flow that minimizes the functional $\mathbb{CT}_c$ is equivalent to the formulation by Kantorovich, i.e.

$$W_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \mathbb{T}_c(\pi) = \min_{(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)} \mathbb{CT}_c(f^{(1)}, f^{(2)}).$$

Delving deeper into the properties of the cardinal flows, we are able to showcase a remarkable structure of the optimal ones. Let $(f^{(1)}, f^{(2)})$ be an optimal cardinal flow between μ and ν for the generic separable cost function $c = c_1 + c_2$. If we denote with ζ the common marginal between $f^{(1)}$ and $f^{(2)}$, i.e.

$$\zeta(y_1, x_2) = \int_{\mathbb{R}} f^{(1)}(x_1, x_2, y_1)dx_1 = \int_{\mathbb{R}} f^{(2)}(x_2, y_1, y_2)dy_2,$$

the restriction of $f^{(1)}$ to horizontal lines gives birth to a transportation plan between the restriction of μ and ζ on the same line. Similarly, the restriction of $f^{(2)}$ on any vertical line is an optimal transportation plan between the restriction of ζ and ν. This property can be written through the conditional laws of μ and ζ and allows us to rewrite the transportation functional as

$$W_c(\mu, \nu) = \int_{\mathbb{R}} W_{c_1}(\mu_{|x_2}, \zeta_{|x_2})d\mu_2 + \int_{\mathbb{R}} W_{c_2}(\zeta_{|y_1}, \nu_{|y_1})d\nu_1. \tag{1.2}$$

Formula (1.2) is particular insightful when we take as a cost function the cost c_p defined in (1.1). Since each restriction of $f^{(1)}$ and $f^{(2)}$ gives birth to a one dimensional problem, and since we know how the optimal transportation plan between one dimensional measures looks like for this class of cost functions, the only real unknown is the measure on which both $f^{(1)}$ and $f^{(2)}$ glue, which is ζ. In particular, we can redefine once again the minimum problem that defines the Kantorovich-Wasserstein distance induced by the cost c_p. In this case we need to minimize the functional

$$\zeta \to \int_{\mathbb{R}} W_p^p(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2 + \int_{\mathbb{R}} W_p^p(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1 \tag{1.3}$$

over all ζ such that

$$\int_{\mathbb{R}} \zeta(y_1, x_2) dy_1 = \mu_2(x_2) \qquad \int_{\mathbb{R}} \zeta(y_1, x_2) dx_2 = \nu_1(y_1). \tag{1.4}$$

We denote the set of all the measures ζ as above with $\mathcal{I}(\mu, \nu)$. It is possible to show that, given any cardinal flow $(f^{(1)}, f^{(2)})$ between μ and ν, their common marginal satisfies the properties (1.4). Moreover, it is also possible to show that, for any measure $\zeta \in \mathcal{I}(\mu, \nu)$, we can find at least a cardinal flow that glues on ζ, which makes the two problems equivalent.
Given any measure $\zeta \in \mathcal{I}(\mu, \nu)$, there is a plethora of possible cardinal flows that glue on ζ, however, the monotone structure of the optimal transportation plan between one-dimensional measures, suggests a particular cardinal flow, which we call *co-monotone* cardinal flow associated to ζ. Given ζ, we can define the cardinal flow $(f^{(1)}, f^{(2)})$ as

$$f^{(1)}(x_1, x_2, y_1) = \pi^{(x_2)}(x_1, y_1) \otimes \mu_2(x_2)$$

and

$$f^{(2)}(y_1, x_2, y_2) = \pi^{(y_1)}(x_2, y_2) \otimes \nu_1(y_1)$$

where $\pi^{(x_2)}(x_1, y_1)$ is the monotone transportation plan between $\mu_{|x_2}$ and $\zeta_{|x_2}$ and $\pi^{(y_1)}(x_2, y_2)$ is the monotone transportation plan between $\zeta_{|y_1}$ and $\nu_{|y_1}$. Given a intermediate configuration $\zeta \in \mathcal{I}(\mu, \nu)$, we can recreate the co-monotone cardinal flow and hence a transportation plan *pi* that passes from ζ, i.e.

$$\int_{\mathbb{R}^2} \pi(\mathbf{x}, \mathbf{y}) dx_1 dy_2 = \zeta(y_1, x_2). \tag{1.5}$$

This plan minimizes the transportation cost restricted to the plans that satisfy relation (1.5) and also offers significative insight on the nature of the

Knothe-Rosenblatt rearrangement [44], which we will prove to be a plan induced by a particular intermediate configuration $\zeta_{KR} \in \mathcal{I}(\mu, \nu)$.

Last but not least, the cardinal flow formulation has a major impact when translated into an uncapacitated minimum flow problem. Given two discrete measures μ and ν defined over a $d-$dimensional grid, the classical transportation problem has always been described through a bipartite graph (in which each point in the support of μ is connected to each point of the support of ν); the cardinal flows, instead, are described through a new graph, the $(d+1)-$partite one. This new graph connects every point only to points with a common coordinate. This models the fact that, in the bidimensional case, the first flow can move the mass only horizontally (i.e. between points with the same second coordinate) and the second one only vertically (i.e. between points with the same first coordinate). Solving the problem on this graph is significatively faster than solving the problem on the bi-partite one. This is due to the fact that the $(d+1)-$partite graph requires considerably less memory and has fewer unknowns: if the starting measures are supported on a $d-$dimensional cube and each of its sides contains N points, the bipartite graph requires $N^d \times N^d = N^{2d}$ unknowns, in the $(d+1)-$partite graph, only dN^{d+1} are needed, which is significatively less, especially as d grows. This intuition will be supported by several tests that will show that this method is now the state of the art when it comes to computing Wasserstein distances between measures with structured support.

The other case we study in this work concerns truncated cost functions. The study of these problems has been approached for the first time in [69], where the authors showed that, given a threshold t, it was possible to simplify the bipartite graph and speed up the computation of the Wasserstein distance related to the truncated cost. Given two discrete sets X and Y, a cost function $c : X \times Y \to \mathbb{R}$ and $t \in \mathbb{R}$, the $t-$truncated cost function is defined as

$$c^{(t)}(x, y) := \min\{c(x, y), t\}.$$

In this case, we will keep the result rather general and will not require any other property nor on structure on the cost function or the support of the measures. Starting from this cost, we can define an auxiliary one

$$s^{(t)}(x, y) := (t - c(x, y))_+$$

where $(\circ)_+$ is the positive part function. We denote as nearby points (and we indicate them with $N_c^{(t)}$) the couples (x, y) on which $s^{(t)}$ is positive. For suitable values of t this set is considerably smaller than the whole $X \times Y$ and, for this reason, it may be convenient to consider only the points in $N^{(t)}$. We can do that by defining a new item: the nearby flow. Given two probability measures μ and ν over X and Y, respectively, a measure η is a nearby flow if for any $x \in X$

$$\sum_{y \in Y \ \ s.t. \ \ c(x,y)<t} \eta(x,y) \leq \mu(x)$$

and, for any $y \in Y$

$$\sum_{x \in X \ \ s.t. \ \ c(x,y)<t} \eta(x,y) \leq \nu(y).$$

The functional to maximize is then defined as

$$\eta \to \sum_{(x,y) \in N^{(t)}} s^{(t)}(x,y)\eta(x,y),$$

and it is related to the classical transportation functional through the relation

$$\min_{\pi} \sum_{(x,y) \in X \times Y} c(x,y)\pi(x,y) = t - \max_{\eta} \sum_{(x,y) \in N^{(t)}} s^{(t)}(x,y)\eta(x,y).$$

Unlike what happens in the classical transportation plan, the measure η that maximizes the functional in our problem has not unitary mass. Our main contribution states that this lack of mass can be used to bound from above both the absolute and the relative error. In particular we have

$$W_c(\mu,\nu) - W_{c^{(t)}}(\mu,\nu) \leq \sum_{(x,y) \in X \times Y} s^{(t)}(x,y) \frac{\tilde{\mu}(x)\tilde{\nu}(y)}{1 - \sum_{(x,y) \in N_c^{(t)}} \eta(x,y)},$$

and

$$\frac{|W_c(\mu,\nu) - W_{c^{(t)}}(\mu,\nu)|}{|W_c(\mu,\nu)|} \leq \frac{\sum_{x \in X, y \in Y} s^{(t)}(x,y)\tilde{\mu}(x)\tilde{\nu}(y)}{\left(1 - \sum_{(x,y) \in N_c^{(t)}} \eta(x,y)\right)\left(\sum_{(x,y) \in N_c^{(t)}} s^{(t)}(x,y)\eta(x,y)\right)},$$

where η is the maximal nearby flow and $\tilde{\mu}$ and $\tilde{\nu}$ are the slack measures defined as

$$\tilde{\mu}(x) = \mu(x) - \sum \eta(x,y) \qquad \text{and} \qquad \tilde{\nu}(y) = \nu(y) - \sum \eta(x,y).$$

We use these bounds to define several stopping criteria for an iterative algorithm, that we call *climbing algorithm*. It consists in an iterative increase of the threshold t at which we truncate the cost function. At each step, we solve the truncated problem, which can be done quickly since we use a small amount of arcs in our graph. We then check the upper bounds on the errors and decide if the approximation is accurate enough.

We apply these algorithms to two problems.

The first one is the computation of the infinity Wasserstein cost, defined as

$$W_c^{(\infty)}(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \max\big\{c(x,y) \text{ s.t. } \pi(x,y)>0\big\}.$$

This cost can be obtained as the limit of $p\to\infty$ of $W_p(\mu,\nu)$ and it measures the minimal maximal movement we have to make to transform μ into ν. This distance appears naturally in many modeling issues, for example, the first appearance of this distance in a scientific paper is in [59], where it was used to describe the shape of rotating stars.

As far as we know, ours is the first algorithm which is able to compute this quantity.

The other application we consider concerns the computation of the projection of a measure over a specific set of measures. Given a positive function $f : X \to \mathbb{R}$ such that

$$1 \le \sum_{x\in X} f(x) < +\infty,$$

we define the set K_f as

$$K_f := \Big\{\mu \in \mathcal{P}(X) \quad \text{s.t.} \quad \mu(x) \le f(x)\Big\}.$$

Since K_f is both closed and convex, the projection operator

$$P_{K_f} : \mu \to \operatorname*{arg\,min}_{\rho\in K_f} W_c(\mu,\rho)$$

is well defined. We will prove that, due to its peculiar definition, it is possible to compute any projection over sets of this kind.

The outline of the work is the following:

In Chapter 2 we recall the basic notions of measure theory, allowing us to fix the notation that will be used throughout the whole work. We will then introduce the optimal transportation problem and state the classical results concerning the existence, uniqueness, and stability of the solution. These results will allow a better understanding of the upsides and problems that arise when dealing with discrete measures. To conclude the chapter, we review some well-known methods used to solve (or approximate) the linear problem that defines the Wasserstein distance. Since an exhaustive and complete guide to this method is out of the goals of this work, we choose to showcase only the methods that can offer a significative parallelism with ours, or the most established methods nowadays to have a significative confrontation on the efficiency.

In Chapter 3, we study the optimal transportation problem for a separable cost function and we introduce the concept of cardinal flow between measures. After proving the main result in its general form, we show how it is possible to retrieve a transportation plan from any cardinal flow. This will allow us to make some consideration about the uniqueness of the solutions. We continue our study by introducing the pivot measures. Given any couple of measures μ and ν, the pivot measure is an intermediate configuration that, for suitable cost functions, fully characterizes the optimal transportation plan between the given pair. More in general, given any feasible measure ζ, we can build a cardinal flow that reshapes μ into ν and "passes" by ζ in a monotone way. As we will see, this object is unique for any given measure ζ and we will call it the co-monotone cardinal flow. To conclude the Chapter, we showcase the numerical framework in which we can set this problem. We will generalize the discussion to any dimension and we will define a flow problem over a $(d+1)-$partite graph. Through this formulation, we will be able to compute the exact value of any p-th power of the Wasserstein distance between measure supported over regular grids. Through several tests and comparison with the established state of the art, we will showcase how our method outclasses them, both in accuracy and in runtime.

Chapter 4 is devoted to truncated cost functions and their related transportation problems. Given a positive threshold t, the $t-$truncated cost function is obtained by choosing the minimum between t and the actual value of the cost function. This cost function is always smaller than the original one and hence induces a distance between measures that is smaller than the original one. This class of simplified problems was proposed by Ofir and Pele in [69], we take their work to a step further and propose an alternative formulation that lives on a smaller network. Our formulation allows us to give an estimation of both the absolute and the relative error we commit by approximating the real Wasserstein distance with the one induced by the cost function truncated at the threshold t. Remarkably, we can compute this estimation explicitly by only knowing the starting measures, the threshold t, and the optimal solution for the approximated problem. We use these estimations to define a stopping criterium for an iterative algorithm that allows us to compute the Wasserstein distance through a sequence of approximations from below. Unlike the method proposed in Chapter 2, this one will work for any cost function and without making assumptions on the support of the measures. As an interesting side note, these methods compute both the infinity Wasserstein distance and the projections of measures in an efficient way. We also provide an estimation of the $W^{(\infty)}$ distance between discrete measures, extending the result proved in [18] to the discrete setting. To conclude, we report the results of our tests.

Chapter 2

Preliminary Results

This chapter aims to introduce the Optimal Transport problem and to showcase the techniques used to explicitly solve it. To do so, we divided the chapter into three main sections.
The first one is devoted to introduce all the fundamental concepts of measure theory and fix the notation that we will use throughout the whole work. For a more detailed description of this topic, we recommend any good probability book, in particular [15, 25]. In the second one, we briefly introduce the Optimal Transport problem and recall the main results about the existence, uniqueness, and stability. All the results, in a more generic form, can be found in [76, 88, 4]. In the last one, we will review a small portion of the computation methods that are used nowadays. We preferred to highlight those that are either renowned for their efficiency or that can offer significative parallelism with the methods we will present in the following Chapters.

2.1 Basic Notions of Measure Theory

Given a non empty set X, we indicate with $P(X)$ the set of all parts, i.e. the set of its subsets. We indicate with $\emptyset$ the empty set. For any $A \in P(X)$, we indicate with A^c its complement, and, for any couple $A, B \in P(X)$, we denote the set A minus B as $A \backslash B := A \cap B^c$.

Definition 2.1 (σ-algebra)**.** *A non empty set $\mathcal{A} \subset P(X)$ is an algebra over X if*

- $\emptyset, X \in \mathcal{A}$*;*
- *if $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$;*

- *if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.*

Moreover, we say that $\mathcal{A}$ is a $\sigma-$algebra if, for any sequence A_n in $\mathcal{A}$, we have

$$\bigcup_{i=1}^{\infty} A_n \in \mathcal{A}.$$

Given $\mathcal{F} \subseteq P(X)$, the σ-algebra generated by $\mathcal{F}$ is defined as the smallest σ-algebra that contains $\mathcal{F}$.

Given a non empty set X, we can endow it with a lot of possible $\sigma-$algebras. However, when X is a polish space, i.e. a complete and separable metric space, we have a standard one: the Borel $\sigma-$algebra.

Definition 2.2 (Borel $\sigma-$algebra)**.** *Given a separable metric space (X, d), the Borel Algebra $\mathcal{B}(X)$ is the smallest $\sigma-$algebra generated by the open balls of d.*

Remark 2.1. *Notice that each σ-algebra on X is also a topology on the same set, hence $\mathcal{B}(X)$ is also characterizable as the smallest topology that makes the distance function d continuous over $X \times X$.*

From now on, we will deal with Polish spaces, i.e. metric spaces that are both complete and separable, which we will endow with the topology induced by the Borel $\sigma-$algebra. This is a classical framework that allows us to avoid a lot of technicalities while leaving the discussion pretty general.

Definition 2.3 (Borel Measure)**.** *Given a polish space X, we say that a function $\mu : \mathcal{B}(X) \to [0, +\infty]$ is a Borel measure over X if it is $\sigma-$additive, i.e. for any given collection of disjoint sets $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{B}(X)$, we have*

$$\mu\bigg(\bigcup_{n \in \mathbb{N}} A_n\bigg) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Moreover, we say that μ is a Borel probability measure if

$$\mu(X) = 1.$$

Given a polish space X we indicate with $\mathcal{M}(X)$ the set of all the Borel measures over X and, with $\mathcal{P}(X)$ the set of all the Borel probability measures over X.

Among those measures, a special role will be covered by the discrete ones. Roughly speaking, a discrete measure is a measure supported on sparse points. Those entities have proven to be a valuable tool to model both structured and sparse data.

Definition 2.4 (Negligible sets)**.** *Let $\mu \in \mathcal{M}(X)$. A set $B \in \mathcal{B}(X)$ is $\mu-$negligible if*

$$\mu(B) = 0.$$

Definition 2.5 (Support of a measure)**.** *Let $\mu \in \mathcal{M}(X)$. The support of μ, namely $spt(\mu)$, is the smallest closed set whose complementary is $\mu-$negligible, i.e.,*

$$spt(\mu) := \bigcap \Big\{ A : A \quad \textit{is closed and} \quad \mu(X \backslash A) = 0 \Big\}.$$

Definition 2.6 (Discrete measures)**.** *Let $\mu \in \mathcal{M}(X)$. The measure μ is discrete if it is concentrated on the union of a countable set of points $\big\{x_n\big\}_{n\in\mathbb{N}}$ in X, i.e.*

$$\mu\Bigg(\Big(\bigcup_{n\in\mathbb{N}} \{x_n\} \Big)^c \Bigg) = 0.$$

Example 2.1 (Dirac's Delta)**.** *Given $x_0 \in X$ we define the Dirac's Delta centered in x_0 as the probability measure defined as*

$$\delta_{x_0}(A) := \begin{cases} 1 & \textit{if} \quad x_0 \in A, \\ 0 & \textit{otherwise.} \end{cases}$$

Any discrete measure can be described as a countable combination of Dirac's deltas, in fact, if μ is concentrated on $\bigcup_{i=1}^{\infty}\{x_n\}$, we can write

$$\mu = \sum_{i=1}^{\infty} \mu(x_n)\delta_{x_n}.$$

2.1.1 Integrals with respect to Radon measures

The transportation problem aims to minimize an integral over a set of measures. To define a well-posed problem, we need to make assumptions on the functions involved in the formulation. For this reason, we briefly recall the basic notions and notation of integration theory. For a complete discussion on these topics, we refer to [15, 4, 88].

Definition 2.7 (Measurable Function)**.** *Let X and Y be two polish spaces and $\phi : X \to Y$ a function. We say that ϕ is (Borel-)measurable if*

$$\phi^{-1}(A) \in \mathcal{B}(X) \qquad\qquad \forall A \in \mathcal{B}(Y),$$

where $\phi^{-1}(A)$ is the pre-image of A through ϕ.

Definition 2.8 (L^p norms)**.** *Let X be a polish space. For any $1 \leq p < +\infty$, we denote by L^p_μ the set of all the measurable functions $f : X \to \mathbb{R}$ with finite L^p_μ norm*

$$||f||_p^p := \int_X |f|^p d\mu < +\infty.$$

We denote with L^∞_μ the set (classes of $\mu-$a.e. identical) of functions with finite $\infty-$norm

$$||f||_\infty := \inf \Big\{ \alpha \in \mathbb{R}, |f(x)| \leq \alpha, \quad \mu - \textit{almost everywhere} \Big\}.$$

We briefly recall here some classical inequalities that will be useful in the following.

Theorem 2.1 (Jensen's Inequality)**.** *Let X be a polish space, $\mu \in \mathcal{P}(X)$, and $a : \mathbb{R} \to \mathbb{R}$ be a convex measurable functions. Then, given any integrable function $f : X \to \mathbb{R}$, it holds true that*

$$a\bigg(\int_X f d\mu \bigg) \leq \int_X (a \circ f) d\mu.$$

Definition 2.9. *The values $p, q \in (1, \infty)$ are conjugate indices if*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Theorem 2.2 (Holder's inequality)**.** *Let $p, q \in (1, \infty)$ be conjugate indices. Then, for all measurable function f and g, we have*

$$\int_X |fg| d\mu \leq ||f||_p ||g||_q,$$

for any $\mu \in \mathcal{P}(X)$.

Theorem 2.3 (Minkowski's inequality)**.** *For any $p \in [1, \infty)$ and any measurable functions f and g, we have*

$$||f + g||_p \leq ||f||_p + ||g||_p.$$

Given $\mu \in \mathcal{P}(X)$, we can deduce key information on its shape through the integration of specific functions. In particular, given $x_0 \in X$, we consider

$$x \to d(x, x_0)$$

from which we can deduce partial information on the behaviour of the tails of μ.

Definition 2.10 (p-th moments). *Let us take $\mu \in \mathcal{P}(X)$ and $x_0 \in X$. For any $p \geq 1$, we define the $p-th$ moment of μ as*

$$M_p(\mu) := \int_X d(x, x_0)^p d\mu,$$

whenever the integral is finite. We say that the measure μ has finite $p-th$ moment if $M_p(\mu) < +\infty$ and denote with $P_p(X)$ the set of measures with finite $p-th$ moment.

Remark 2.2. *If $\mu \in \mathcal{P}(X)$ has finite $q-th$ moment, then it has finite $p-th$ moment for each $p \leq q$. In fact, since the map*

$$x \to x^\alpha$$

is convex for each $\alpha \geq 1$, from Jensen's inequality we have

$$\left(\int_X d^p(x, x_0) d\mu\right)^{\frac{q}{p}} \leq \int_X \left(d^p(x, x_0)\right)^{\frac{q}{p}} d\mu = \int_X d^q(x, x_0) d\mu < +\infty.$$

Weak convergence in $\mathcal{P}(X)$

As we mentioned above, images are described through discrete measures, then, if we can define a distance between measures, it can be used to quantify how much two images are different. Since not all the distances are equivalent, when it comes to practical applications, we have to choose the one that better fits the aim of the application. For example, given X a Polish space, we can define a distance over the space $\mathcal{P}(X)$ by setting

$$\delta_{TV}(\mu, \nu) := \sup_{A \in \mathcal{B}(X)} \big|\mu(A) - \nu(A)\big|. \tag{2.1}$$

The value introduced above is also called total variation distance between μ and ν and it is indeed a distance.
The total variation between two discrete measures is easy to compute (it can be done through a sum), however, it induces a notion of convergence that is too strict, which makes it less appealing for practical applications.

Example 2.2. *Let us take $X = \mathbb{R}$ and consider the sequence*

$$\mu_n = \delta_{\frac{1}{n}}.$$

According to the human perception of shapes, one would say that this sequence converges toward the measure $\mu = \delta_0$, however, through a simple computation we can see that

$$\delta_{TV}(\mu, \nu) := 1$$

for any $n \in \mathbb{N}$, hence, according to the total variation distance, this sequence does not converge.

We can endow the space $\mathcal{P}(X)$ with another topology, the one induced by the duality of the measures with the space of continuous and bounded functions, the weak convergence.

Definition 2.11 (Weak Convergence)**.** *Given $\{\mu_n\}_{n\in\mathbb{N}}$ a sequence of measures on X, we say that μ_n weakly converges to $\mu \in \mathcal{M}(X)$ if*

$$\int_X f d\mu_n \to \int_X f d\mu$$

for any $f \in C_b(X)$. The topology induced by this notion of convergence is called weak topology.

Example 2.3. *Although the sequence in Example 2.2 does not converge according to the δ_{TV} metric, it does converge toward the Dirac's delta centered in 0 according the weak convergence. In fact*

$$\int_{\mathbb{R}} f d\mu_n = f\left(\frac{1}{n}\right)$$

which, since f is a generic continuous function, does converge to $f(0)$, which can be written as

$$f(0) = \int_{\mathbb{R}} f d\delta_0$$

and, hence, $\delta_{\frac{1}{n}} \to \delta_0$ weakly.

Remark 2.3. *It is worth of notice that bounded sets of $P_p(X)$ introduced in Definition 2.10 are closed with respect to the weak topology. Given μ_n a sequence in $P_p(X)$ that weakly converges toward $\mu \in \mathcal{P}(X)$, if there exists a constant C such that*

$$\int_X d(x, x_0)^p d\mu_n \leq C \tag{2.2}$$

for any $n \in \mathbb{N}$, then $\mu \in P_p(X)$. This follows from the fact that, by definition of the Borel $\sigma-$algebra the function $d(\circ, x_0)$ is continuous for any $x_0 \in X$, and so is its $p-$th power, hence

$$\lim_{n\to\infty} \int_X d(x, x_0)^p d\mu_n = \int_X d(x, x_0)^p d\mu,$$

which, along with relation (2.2) *allow us to conclude* $\mu \in P_p(X)$.

Thanks to a classical result from Prokhorov, the compact sets in this topology have a neat characterization, he showed that tightness is a sufficient condition.

Definition 2.12 (Tightness)**.** *A subset A of $\mathcal{P}(X)$ is tight if, for every $\epsilon > 0$, there exists a compact set $K_\epsilon \subset X$ such that*

$$\mu(X \backslash K_\epsilon) \leq \epsilon \qquad\qquad \forall \mu \in A.$$

In particular, a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ is tight if $\bigcup_{n\in\mathbb{N}}\{\mu_n\} \subset \mathcal{P}(X)$ is tight.

Theorem 2.4. *Let (X, d) be a polish space, then any finite Borel measure is also tight. In particular each probability over the Borel algebra is tight.*

Theorem 2.5 (Prokhorov)**.** *Let μ_n be a tight sequence in $\mathcal{P}(X)$, then there exists a measure $\mu \in \mathcal{P}(X)$ and sub-sequence μ_{n_k} such that*

$$\mu_{n_k} \to \mu.$$

As we will see in the next section, the weak topology is also the one induced by the Wasserstein distance. This relation between topology and distance has been of key importance for its success in the application in image retrieval, machine learning, and many other applied fields.

Pushforward of Measures

The concept of pushforward has been the starting point for developing the optimal transportation cost. Roughly speaking, a pushforward of a measure is a deterministic rearrangement of a measure into another one. This can be done through the use of functions that send a polish space into another one.

Definition 2.13 (Push-forward of measures)**.** *Let X and Y be two polish spaces, $T : X \to Y$ a measurable function, and $\mu \in \mathcal{M}(X)$. The push-forward measure of μ through T is defined as*

$$T_\#\mu(B) := \mu(T^{-1}(B))$$

for each $B \in \mathcal{B}(Y)$.

Remark 2.4. *Given a measurable $T : X \to Y$ and $\mu \in \mathcal{P}(X)$, there is an integral way to characterize the push-forward measure $T_\#\mu$. Given any measurable function ϕ on Y the push-forward measure of μ through T is the only probability measure ν satisfying the identity*

$$\int_X \phi \circ T d\mu = \int_Y \phi d\nu. \tag{2.3}$$

It is also worth of notice that the push-forward operator preserves the total mass of the measure, since $T^{-1}(Y) = X$ and, hence

$$T_\# \mu(Y) = \mu(X).$$

In particular, any measurable map $T : X \to Y$ can be lifted to a map $\Phi : \mathcal{P}(X) \to \mathcal{P}(Y)$.
Since the composition of two measurable functions is still measurable, we can define a chain rule for push-forwards.

Lemma 2.1 (Chain Rule for Push-forwards). *Let X, Y and Z be three polish spaces and let $T_1 : X \to Y$, $T_2 : Y \to Z$ two measurable functions. Given any $\mu \in \mathcal{P}(X)$, it holds true the following chain formula*

$$(T_2 \circ T_1)_\# \mu = (T_2)_\# (T_1)_\# \mu.$$

Lemma 2.2 (Linearity of Push-forwards). *Given $T : X \to Y$ a measurable function between two polish spaces and $\mu, \mu' \in \mathcal{P}(X)$, then*

$$T_\#(\mu + \mu') = T_\# \mu + T_\# \mu'.$$

Let us assume that the polish space X is the direct product of two polish spaces X_1 and X_2. We can then define the projections over X_1 and X_2 as the functions

$$(\mathfrak{p}_{X_1})(\mathbf{x}) := x_1 \quad \text{and} \quad (\mathfrak{p}_{X_2})(\mathbf{x}) := x_2$$

respectively, where $\mathbf{x} = (x_1, x_2)$ is a generic point of X. Those functions are always continuous (and hence measurable) and allow us to define the marginals of a measure.

Definition 2.14 (Marginal Probabilities). *Let $X = X_1 \times X_2$ be a polish space and $\mu \in \mathcal{P}(X)$, the first marginal of μ is $\mu_1 \in \mathcal{P}(X_1)$ defined as*

$$\mu_1 := (\mathfrak{p}_{X_1})_\# \mu.$$

Similarly, the second marginal of μ is defined as

$$\mu_2 := (\mathfrak{p}_{X_2})_\# \mu.$$

Given a measure $\mu \in \mathcal{P}(X)$, where $X = X_1 \times X_2$, its marginals are always uniquely determined, but the vice versa is not true unless one of the two marginal is a Dirac's delta. In general, if we are given the marginal on X_1 and X_2 of μ there is a plethora of measures over the space $X_1 \times X_2$ whose marginals are μ_1 and μ_2. Among those many probability measures, a classical one is the direct product of the marginal.

Definition 2.15 (Direct Product of Measures)**.** *Let us take two measures with the same mass, i.e.*

$$\mu(X) = \nu(Y) > 0.$$

The direct product measure of μ and ν is the measure $\mu \otimes \nu \in \mathcal{M}(X \times Y)$ defined as

$$\mu \otimes \nu(A \times B) := \frac{\mu(A)\nu(B)}{\mu(X)} = \frac{\mu(A)\nu(B)}{\nu(Y)}$$

for each Borel set $A \subset X$ and $B \subset Y$. When $\mu(X) = \nu(Y) = 0$, we define the direct product $\mu \otimes \nu$ as the null measure.

Remark 2.5. *The direct product of two measures has the same mass of the single measures that compose the product, in particular, if μ and ν are probability measures, so is their product.*

Lemma 2.3 (Gluing Lemma)**.** *For $i = 1, 2, 3$, let X_i be polish spaces and $\mu_i \in \mathcal{P}(X_i)$. If $\pi_1 \in \Pi(\mu_1, \mu_2)$ and $\pi_2 \in \Pi(\mu_2, \mu_3)$ then there exists $\pi \in \mathcal{P}(X_1 \times X_2 \times X_3)$ such that*

$$(\mathfrak{p}_{X_1 \times X_2})_\# \pi = \pi_1$$

and

$$(\mathfrak{p}_{X_2 \times X_3})_\# \pi = \pi_2.$$

Another classical notion related to push-forwards is the disintegration of a measure. Being able to disintegrate a measure allows one to decompose a complex integral into two smaller and, presumably, simpler ones.

Definition 2.16 (Disintegration of a Measure)**.** *Let $f : X \to Y$ be a measurable function and $\mu \in \mathcal{P}(X)$. We say that a family $\{\mu_y\}_{y \in Y}$ is a disintegration of μ according to f if every μ_y is a probability measure concentrated on $f^{-1}(\{y\})$ such that*

$$y \to \int_X \phi d\mu_y$$

is a measurable function and

$$\int_X \phi d\mu = \int_Y \left(\int_X \phi d\mu_y \right) d(f_\# \mu) \tag{2.4}$$

for every $\phi \in C(X)$. With a slight abuse of notation, the disintegration of μ is also often written as

$$\mu = \mu_y \otimes \nu \tag{2.5}$$

where $\nu = f_\# \mu$.

Theorem 2.6 (Existence of Disintegration). *Let λ be a σ-finite Radon measure on a metric space X and let T be a measurable map from X into Y. If there exists a σ-finite measure on $B(Y)$ that dominates the image measure $T_{\#}\lambda$ then λ has a $T-$disintegration. The λ_y measures are uniquely determined up to an almost sure equivalence.*

Also in this case, a special class of disintegration is induced by the projection functions. If $X = X_1 \times X_2$ and we choose $f = (\mathfrak{p}_{X_1})$, formula (2.5) reads as

$$\mu = \mu_{|x_1} \otimes \mu_1.$$

The measure $\mu_{|x_1}$ is called the conditional law of μ given x_1. We can think of this measure as the normalized restriction of μ over the set $\{x_1\} \times X_2 \subset X$.

Measures over the Euclidean spaces

We will now specialize the discussion further, by choosing the $d-$dimensional Euclidean space $\mathbb{R}^d$ as the polish space. In this framework, there is a canonical measure, the Lebesgue one.

Definition 2.17 (Lebesgue Measure). *The Lebesgue measure $\mathcal{L}$ is the only $\sigma-$measure on $\mathbb{R}$ such that, for any closed set $I = [a, b]$, it holds that*

$$\mathcal{L}([a, b]) = b - a.$$

The $d-$dimensional Lebesgue measure is then obtained by producing d times the one-dimensional measure.

Through this measure, we can define the class of absolutely continuous measures, which can be expressed as integral of positive functions with respect to the Lebesgue measure.

Definition 2.18 (Absolutely Continuous Measures). *Let us take $f : \mathbb{R}^d \to [0, +\infty)$ an integrable function such that*

$$\int_{\mathbb{R}^d} f(x)dx = 1.$$

We define $\mu_f \in \mathcal{P}(\mathbb{R}^d)$ as it follows

$$\mu_f(A) := \int_A f(x)dx \tag{2.6}$$

for each $A \in \mathcal{B}(\mathbb{R}^d)$. This class of probability measures is called absolutely continuous probability measures and the function f that induces the probability measure is called density of the probability.

Remark 2.6. *Owing to the $\sigma-$additivity of the Lebesgue measure, it is possible to prove that any measure μ_f defined as in* (2.6) *is $\sigma-$additive as well. Hence, any μ_f as above is a Borel measure. Moreover, the density of the measure completely characterize the measure itself, hence, when it does not create confusion, we indicate with the same symbol both the measure μ and its density $\mu(x_1, x_2)$.*

Given an absolutely continuous measure μ, we can write its marginals as an integral and, moreover, the regularity of μ is inherited by its marginal and conditional laws.

Theorem 2.7. *Let $\mu \in \mathcal{P}(\mathbb{R}^2)$ be an absolutely continuous measure of density $\mu(x_1, x_2)$. Then its marginals are absolutely continuous and their density is given by*

$$\mu_1(x_1) := \int_{\mathbb{R}} \mu(x_1, t)dt$$

and

$$\mu_2(x_2) := \int_{\mathbb{R}} \mu(s, x_2)ds.$$

Similarly, also the conditional laws are absolutely continuous and they can be expressed through the following formula

$$\mu_{|x_1}(x_2) := \begin{cases} \dfrac{\mu(x_1, x_2)}{\mu_1(x_1)} & \textit{if} \quad \mu_1(x_1) > 0 \\ 0 & \textit{otherwise}, \end{cases}$$

and

$$\mu_{|x_2}(x_1) := \begin{cases} \dfrac{\mu(x_1, x_2)}{\mu_2(x_2)} & \textit{if} \quad \mu_2(x_2) > 0 \\ 0 & \textit{otherwise}. \end{cases}$$

An important feature of multidimensional Euclidean spaces is that they are totally ordered on each of their components. This is because $\mathbb{R}^d$ is a direct product of $\mathbb{R}$, which is totally ordered. In particular, given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we can determine which point has the biggest $j-$th component. Given a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, we can then define its cumulative function F_μ.

Definition 2.19 (Cumulative Function)**.** *Given $\mu \in \mathcal{P}(\mathbb{R}^d)$, we define the cumulative function (or CDF) $F_\mu : \mathbb{R}^d \to [0, 1]$ as*

$$F_\mu(t_1, \ldots, t_d) := \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} d\mu.$$

This classic function has remarkable properties that we briefly recall here

- F_μ is always monotone non decreasing and right-continuous with respect to each coordinates. Moreover, it holds true

$$\lim_{t_1,\dots,t_d \to -\infty} F_\mu(t_1,\dots,t_d) = 0$$

and

$$\lim_{t_1,\dots,t_d \to +\infty} F_\mu(t_1,\dots,t_d) = 1.$$

- It holds true that $\mu \neq \mu'$ iff $F_\mu \neq F_{\mu'}$, which means that the cumulative function uniquely identifies the measure and vice versa.

- If μ_j is the $j-$th marginal of μ, it holds true that

$$\lim_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_d \to +\infty} F_\mu(t_1,\dots,t_{j-1},\alpha,t_{j+1},\dots,t_d) = F_{\mu_j}(\alpha),$$

for any $\alpha \in \mathbb{R}$.

When $d = 1$ and then the measure μ is supported on $\mathbb{R}$, we can also introduce the generalized inverse of the CDF. The pseudo-inverse covers an important role in the resolution of the transport problem in the mono-dimensional case.

Definition 2.20 (Pseudo-Inverse Function)**.** *Let $F : \mathbb{R} \to [0, 1]$ be a non-increasing right-continuous function, its pseudo-inverse $F^{[-1]} : [0, 1] \to \mathbb{R}$ is defined as*

$$F^{[-1]}(t) := \inf_{x\in\mathbb{R}} \big\{F(x) \geq t\big\}.$$

Remark 2.7. *We notice that, given $\mu \in \mathcal{P}(\mathbb{R})$, both its cumulative function and its pseudo inverse are monotone non decreasing. In particular given two measures μ and ν, both the compositions $F_\mu \circ F_\nu^{[-1]}$ and $F_\nu \circ F_\mu^{[-1]}$ are monotone non decreasing as well.*

Given a measure $\mu \in \mathcal{P}(\mathbb{R})$, we can define the CDF by integrating it on a suitable set, we can then recover the pseudo-inverse by inverting it. The next result states that we can recover the probability measure from a push-forward of the pseudo-inverse function.

Lemma 2.4. *Given $\mu \in \mathcal{P}(\mathbb{R})$ let $F_\mu^{[-1]}$ be the pseudo-inverse of the cumulative function of μ. Then*

$$(F_\mu^{[-1]})_\#(\mathcal{L}_{|[0,1]}) = \mu$$

where $\mathcal{L}_{|[0,1]}$ is the Lebesgue measure restricted on $[0, 1]$.

Proof. It follows from this straightforward considerations

$$\begin{aligned}\mathcal{L}_{|[0,1]}(\{x \in [0,1] : F_\mu^{[-1]}(x) \leq a\}) &= \mathcal{L}_{|[0,1]}(\{x \in [0,1] : F_\mu(a) \leq x\}) \\ &= F_\mu(a)\end{aligned}$$

which allows us to conclude thanks to the characterization of the measure through their cumulative function. □

2.2 The Optimal Transport Problem

The first formulation of the transportation problem is the one due to Monge. In modern language this formulation reads as it follows: given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a cost function $c : X \times Y \to [0, \infty]$, we have to minimize the functional

$$T \to \int_X c(x, T(x))d\mu$$

over all the functions $T : X \to Y$ such that

$$T_\# \mu = \nu.$$

Let us assume $X = Y = \mathbb{R}^d$ and μ, ν be two absolutely continuous probabilities. Let us denote with f and g the densities of μ and ν respectively. In this setting, we can use the change of variable formula to rewrite the relation $T_\# \mu = \nu$ and find the Monge-Ampere equation, which reads as it follow

$$g(T(x))|det(DT(x))| = f(x) \tag{2.7}$$

if we suppose T to be injective and regular enough. Even under those assumptions, equation (2.7) is highly non linear. This feature of the equation makes hard the study of the existence of a solution, which is usually done by taking a minimizing succession T_n that it is known to converge, respect a certain weak topology, to a function T. By using the lower semi-continuity of the functional it is possible to infer that T is a minimal point. However, we should also be able to show that the limit T still satisfies equation (2.7), which, since it is nonlinear, prevents us from concluding.
If we also drop the conditions over the regularity of μ and ν, we could also incur other problems, as the next example shows.

Example 2.4. *Let us take $\mu = \delta_x$ and $\nu = \frac{1}{2}\Big(\delta_{y_1} + \delta_{y_2}\Big)$ where x and y_1, y_2 are fixed points in two polish spaces X and Y. Given any measurable function $T : X \to Y$, we have*

$$T_\# \mu = \delta_{T(x)},$$

hence it is not possible to find a function T *as above that satisfies the equation* $T_{\#}\mu = \nu$.
It is also worth of notice that, if we swap the roles of μ *and* ν*, i.e. we search for a* $S : Y \to X$ *such that* $S_{\#}\nu = \mu$*. By the same argument used above and from the linear property of pushforwards (Lemma 2.2), we have*

$$S_{\#}\nu = \frac{1}{2}\Big(\delta_{S(y_1)} + \delta_{S(y_2)}\Big)$$

hence, whenever $S(y_1) = S(y_2) = x$*, the function* S *satisfies the relation* $S_{\#}\nu = \mu$*. In particular, the function* $S(y) = x$ *for each* $y \in Y$*, it is a feasible transportation map from* ν *to* μ.
This showcases two of the many problems of the Monge formulation: the set on which we want to minimize could be empty and the problem is not symmetric, which means that moving μ *into* ν *it is different from moving* ν *into* μ.

To solve those issues of the Monge formulation, Kantorovich proposed an alternative formulation. He modelized the transshipment of mass through probability measure over the product space $X \times Y$ and not through maps.

Definition 2.21 (Transportation Plan)**.** *Let* μ *and* ν *be two measures over two polish spaces* X *and* Y*. The probability measure* $\pi \in \mathcal{P}(X \times Y)$ *is a transportation plan between* μ *and* ν *if*

$$(\mathfrak{p}_X)_{\#}\pi = \mu$$

and

$$(\mathfrak{p}_Y)_{\#}\pi = \nu.$$

We denote with $\Pi(\mu, \nu)$ *the set of all the transportation plans between* μ *and* ν.

Unlike the push-forwards, transportation plans allow the mass in a certain point to split, more in general, given $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(Y)$, the quantity $\pi(A \times B)$ is the amount of mass that from the set A travels to the set B.

Remark 2.8. *Given two measures* μ *and* ν*, the set* $\Pi(\mu, \nu)$ *is nonempty. In fact we can define the independent transportation plan between* μ *and* ν*, defined as*

$$\pi = \mu \otimes \nu,$$

is in $\Pi(\mu, \nu)$.

According to this plan, the amount of mass that travels from a set $A \subset X$ to a set $B \subset Y$ depends only on the amount of mass that μ and ν assign to the respective sets. Although it is easy to define, this plan can never be the optimal one, except in some trivial cases.

Remark 2.9 (Deterministic Plans)**.** *Let $T : X \to Y$ be a measurable map such that*

$$T_{\#}\mu = \nu,$$

then the measure $\pi_T := \big(Id_X, T\big)_{\#}\mu$ is a transportation plan between μ and ν. In fact, from the composition of the push forwards we have

$$(\mathfrak{p}_X)_{\#}\pi_T = \Big((\mathfrak{p}_X)_{\#} \circ \big(Id_X, T\big)\Big)_{\#}\mu = \big(Id_X\big)_{\#}\mu = \mu$$

and, similarly, we prove $(\mathfrak{p}_Y)_{\#}\pi_T = \nu$. According to this plan, all the mass stored in x is sent into $T(x)$, for this reason the plan is also called deterministic plan. From the definition of push-forward, we have

$$\int_{X\times Y} c(x,y)d\pi_T = \int_X c(x,T(x))d\mu,$$

hence the cost of π_T according to the Kantorovich functional is equal to the cost of the map T according to the Monge functional. More in general, it is possible to show that the Kantorovich formulation is the relaxation of the Monge one.

Theorem 2.8. *Given two tight sets $A \subset \mathcal{P}(X)$ and $B \subset \mathcal{P}(Y)$, the set C defined as*

$$C := \Big\{\pi \in \mathcal{P}(X \times Y) : \exists \mu \in A, \nu \in B \quad \textit{such that} \quad \pi \in \Pi(\mu,\nu)\Big\}$$

is tight in $\mathcal{P}(X \times Y)$. In particular, the set $\Pi(\mu,\nu)$ is tight in $\mathcal{P}(X \times Y)$.

Given any couple of measures μ and ν, there are, in general, a plethora of transportation plans between them. We need to introduce a way to evaluate and compare them. This can be done by introducing a cost function that measures how much costs moving a unity of mass from any point of X to any point of Y.

Definition 2.22 (Cost Function)**.** *Let us take two polish spaces X and Y. A function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is a feasible cost function if it is l.s.c. and symmetric.*

Definition 2.23 (Transportation Functional)**.** *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a cost function such that there exist two u.s.c. functions $a \in L^1_\mu$ and $b \in L^1_\nu$ such that*

$$c(x, y) \geq a(x) + b(x) \tag{2.8}$$

for each $(x, y) \in X \times Y$. The transportation functional $\mathbb{T}_c : \Pi(\mu, \nu) \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\mathbb{T}_c(\pi) := \int_{X \times Y} c d\pi. \tag{2.9}$$

Remark 2.10. *The conditions asked to the cost function in Definition 2.22 and Definition 2.23 are the minimal ones for which it makes sense defining the integral in* (2.9)*. However, to grant the well posedness of the minimization problem, we have to ask for some further regularity.*

Definition 2.24 (Minimal Transportation Cost)**.** *Let us take a feasible cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. The minimal transportation cost functional $C : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ is defined as*

$$(\mu, \nu) \to C(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \mathbb{T}_c(\pi). \tag{2.10}$$

Remark 2.11. *Minimal transportation cost has many abstract applications and interpretations. For instance, it can be seen as a reallocation of a good. In this case, the measure μ modelizes the distribution of a certain facility (the supply) while ν is the distribution of the need among customers (the demand). We can then think of the transportation plan as a policy adopted by a fictionary company in charge of reallocating the goods accordingly to the demands of customers. Since moving the facilities is not free, the company has to decide wisely how to move the goods. In particular, if we can associate a cost to any possible association, the company will be incentivized to minimize their expenses, which can be expressed through the transportation functional defined in* (2.10)*.*

By making some further assumption, we can prove that the infimum in (2.10) is actually a minimum, i.e. there exist a minimizer. This is done through a classical variational method. Hence we have to ensure that the functional $\mathbb{T}_c$ is both convex and lower semi-continuous for a suitable topology.
From the linearity of the integral, we can already prove that $\mathbb{T}_c$ is convex. In fact, let us take $\pi, \pi' \in \Pi(\mu, \nu)$ and $\lambda \in [0, 1]$, then, by the convexity of $\Pi(\mu, \nu)$, we deduce

$$\lambda\pi + (1 - \lambda)\pi' \in \Pi(\mu, \nu),$$

and, by the linearity of the integral

$$\begin{aligned}\mathbb{T}_c(\lambda\pi + (1-\lambda)\pi') &= \int_{X\times Y} cd\Big(\lambda\pi + (1-\lambda)\pi'\Big) \\ &= \lambda\int_{X\times Y} cd\pi + (1-\lambda)\int_{X\times Y} cd\pi' \\ &= \lambda\mathbb{T}_c(\pi) + (1-\lambda)\mathbb{T}_c(\pi').\end{aligned}$$

Hence $\mathbb{T}_c$ is convex, although not strictly. Regarding the lower semi-continuity can be recovered by making minimal assumption on the cost functions.

Lemma 2.5 (Lower semi continuity of $\mathbb{T}_c$, Villani [88], Chapter 4, Lemma 4.3)**.** *If the cost function c is bounded from below by an upper semi continuous function $h : X \times Y \to \mathbb{R}$, the functional $\mathbb{T}_c$ is weakly l.s.c., i.e. if $\pi_n \in \mathcal{P}(X\times Y)$ is a sequence that weakly converges to $\pi \in \mathcal{P}(X \times Y)$, then it holds true that*

$$\mathbb{T}_c(\pi) \leq \liminf_{n\to\infty} \mathbb{T}_c(\pi_n).$$

We are now ready to state the classical result concerning the existence of a minimizer for the minimal transportation problem induced by $\mathbb{T}_c$.

Theorem 2.9 (Existence, Villani [88], Chapter 4, Theorem 4.1)**.** *Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be two probabilities. Given a cost function $c : X \times Y \to \mathbb{R}\cup\{+\infty\}$, if there exist two u.s.c. functions $a \in L^1_\mu$ and $b \in L^1_\nu$ for which*

$$c(x, y) \geq a(x) + b(x) \tag{2.11}$$

for each $(x, y) \in X \times Y$, then $\exists\pi \in \Pi(\mu, \nu)$ which minimizes $\mathbb{T}_c$. Furthermore, if there exist two functions $c_X \in L^1_\mu$ and $c_Y \in L^1_\nu$ for which

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

then

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) < \infty.$$

Remark 2.12. *Since $a \in L^1_\mu$ and $b \in L^1_\nu$ are u.s.c., their sum is u.s.c. over $X \times Y$, hence also the conditions of Theorem 2.9 assure the lower semi-continuity of $\mathbb{T}_c$.*
As a plus, the requirement (2.11)*, bounds the infimum value from below, since*

$$\int_{X\times Y} cd\pi \;\; \geq \;\; \int_{X\times Y} ad\pi + \int_{X\times Y} bd\pi$$

$$= \int_X a d\mu + \int_Y b d\nu > -\infty,$$

for any $\pi \in \Pi(\mu, \nu)$. However, during our work we will deal with positive cost functions (which corresponds to $a = b = 0$), hence the functional $\mathbb{T}_c$ will be l.s.c. and will admit a minimizer.

As we saw above, the functional $\mathbb{T}_c$ is not strictly convex, hence, even if we can guarantee the existence of a solution for the minimization problem it induces, the solution will not be unique. Unfortunately, there is no hope to recover the uniqueness by making some mild assumption as has been done for the lower semi-continuity and we need to make assumptions on both the cost function and the starting measures μ and ν.

Theorem 2.10 (Uniqueness, Gangbo and McCann [35], Theorem 6.3)**.** *Let be given μ and ν two probability measures on a compact domain $\Omega \subset \mathbb{R}^d$ and a cost function $c(x, y) = h(|x - y|)$ where $h : \mathbb{R} \to [0, +\infty]$ is a strictly convex function such that $h(0) = 0$. Then the optimal plan π associated to this problem is unique and, if μ is absolutely continuous and the boundary of Ω is negligible, that plan is induced by a push-forward, i.e.*

$$\pi = (Id, T)_{\#}\mu.$$

Unfortunately, there is no hope of obtaining uniqueness without making assumptions on the measures μ and ν. In the next example, we show that the mere convexity of the cost function is not enough.

Example 2.5. *Let us take $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ defined as*

$$\mu = \frac{1}{2}\Big[\delta_{(0,0)} + \delta_{(1,1)}\Big] \qquad \text{and} \qquad \nu = \frac{1}{2}\Big[\delta_{(1,0)} + \delta_{(0,1)}\Big].$$

For any $p > 1$, we consider the $p-$th power cost function

$$c_p(\boldsymbol{x}, \boldsymbol{y}) := (|x_1 - y_1|)^p + (|x_2 - y_2|)^p,$$

which is indeed strictly convex. Since each source is equally distant from each sink, it is easy to see that the function $\mathbb{T}_{c_p}$ is constant over the set $\Pi(\mu, \nu)$. In particular, each element of $\Pi(\mu, \nu)$ is minimal and hence the convexity of the cost is not sufficient to assure the uniqueness of the solution.

Corollary 2.1. *Let us take $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. If c and c' are two cost functions such that*

$$c'(x, y) \leq c(x, y) \qquad \forall (x, y) \in X \times Y,$$

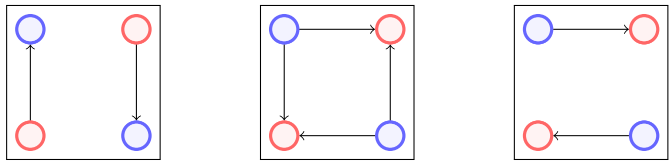


Figure 2.1: Three examples of optimal transportation plans between the measures μ and ν defined in Example 2.5. The plans described in the first and third images are the extremal points of $\Pi(\mu, \nu)$, so that each $\gamma \in \Pi(\mu, \nu)$ can be described as a convex combination of those two.

then

$$\min_{\pi \in \Pi(\mu,\nu)} \mathbb{T}_{c'}(\pi) \leq \inf_{\pi \in \Pi(\mu,\nu)} \mathbb{T}_c(\pi).$$

To conclude this brief introduction to Optimal Transport, we showcase some properties the functional $\mathcal{C}$ enjoys and that will come handy later.

c-cyclical monotonicity

Given two measures, the optimal transportation plan between them depends on the cost function. Roughly speaking, this is because each cost function induces a different geometrical structure over $X \times Y$, which makes some association better than other ones. Remarkably, this allows us to check the optimality of a transportation plan by looking at its support.

Definition 2.25 (c-cyclical monotonicity)**.** *Let X and Y be two polish spaces and let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a cost function. A subset $\Gamma \subset X \times Y$ is $c-$cyclically monotone if, $\forall n \in \mathbb{N}$ and for any $(x_i, y_i) \in \Gamma$, with $i = 1, \dots, n$,*

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

with the convention $y_{n+1} = y_1$. A transportation plan $\pi \in \Pi(\mu, \nu)$ is $c-$cyclically monotone if its support is $c-$cyclically monotone.

Remark 2.13. *It is easy to see that, if $\Gamma \subset X \times Y$ is a $c-$cyclically monotone set, any sub-set $\Gamma' \subset \Gamma$ is so.*

As we noticed before, a couple (x, y) belongs to the support of a transportation plan π if, according to π, there is mass moving from x to y. Let us now take a collection of pairs (x_i, y_i) (with $i = 1, \dots, N$) that belongs to the support of a transportation plan, if this collection is $c-$cyclically monotone, it means

that we cannot improve the overall cost by rearranging the starting points and the arrival points.
This property translates to the whole plan if its support is $c-$cyclically monotone.

Theorem 2.11 (Santambrogio [76], Chapter 1, Theorem 1.38 and Theorem 1.42)**.** *Let c be a continuous cost function. If $\pi \in \Pi(\mu, \nu)$ is an optimal transportation plan for the cost function c, then spt(π) is a $c-$cyclically monotone subset of $X \times Y$. Moreover, if $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is uniformly continuous and $\rho \in \mathcal{P}(X \times Y)$ such that spt(ρ) is $c-$cyclically monotone then ρ is the optimal transportation plan between its marginals $\mu = (\mathfrak{p}_X)_{\#}\rho$ and $\nu = (\mathfrak{p}_Y)_{\#}\rho$ for the cost c.*

Remark 2.13 and Theorem 2.11 allow us to deduce that any restriction of an optimal transportation plan is optimal.

Theorem 2.12 (Restriction Property, Villani [88], Chapter 4, Theorem 4.6)**.** *Let X and Y be two polish spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, $a \in L^1_\mu$, $b \in L^1_\nu$, and let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be an uniformly continuous cost function such that*

$$c(x, y) \geq a(x) + b(y)$$

for all $x \in X$ and $y \in Y$. Let C be the optimal transportation functional defined in (2.10)*. Moreover, let π be one of the optimal transportation plans between μ and ν and assume that $C(\mu, \nu) = \mathbb{T}_c(\pi) < +\infty$. Then, if π' is a positive measure such that $\pi' \leq \pi$ and $\pi'(X \times Y) > 0$, the probability measure*

$$\tilde{\pi} = \frac{\pi'}{\pi'(X \times Y)}$$

is an optimal transportation plan between its marginals.

Stability and Convexity

Theorem 2.13 (Stability of the Optimal Plan, Villani [88], Chapter 5, Theorem 5.20)**.** *Let X and Y be two polish spaces and $c : X \times Y \to \mathbb{R}$ a continuous cost function satisfying* (2.11)*. Let c_k be a sequence of continuous cost functions converging uniformly to c on $X \times Y$. Let μ_k and ν_k be two sequences of probability measures on X and Y respectively. Assume that μ_k converges to μ (respectively ν_k converges to ν) weakly. For each k, let π_k be an optimal plan between μ_k and ν_k. If, for each $k \in \mathbb{N}$*

$$\int_{X \times Y} c_k d\pi_k < +\infty,$$

then, up to a sub-sequence, π_k weakly converges to some $c-$ciclically monotone transference plan $\pi \in \Pi(\mu, \nu)$. If moreover

$$\liminf_{k \in \mathbb{N}} \int_{X \times Y} c_k d\pi_k < +\infty$$

then π is an optimal transportation plan between μ and ν.

Theorem 2.13 states that the optimal transportation plan is stable under perturbations, i.e. similar couples of measures induce similar optimal transportation plans. In some sense, also the measurability is preserved: if we have two measurable collections of measures, also the collection of optimal transportation plans will be measurable.

Lemma 2.6 (Measurable Selection of Plans, Villani [88], Chapter 5, Corollary 5.22)**.** *Let X and Y be two polish spaces and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ a continuous cost function such that $\inf c > -\infty$. Given Ω a polish space and $\lambda \in \mathcal{P}(\Omega)$, consider a measurable map*

$$\omega \to (\mu_\omega, \nu_\omega)$$

that goes from Ω to $\mathcal{P}(X) \times \mathcal{P}(Y)$. Then there is a measurable choice

$$\omega \to \pi_\omega$$

where for each ω, π_ω is the optimal transportation plan between μ_ω and ν_ω.

Through the use of this result, it is possible to show that the transportation cost functional is convex in its arguments.

Theorem 2.14 (Convexity Property, Villani [88], Chapter 4, Theorem 4.8)**.** *Let X and Y be two polish spaces, let c be a feasible cost function between X and Y and let C be the optimal transportation functional associated to c. Let then (Ω, λ) be a probability space and let μ_ω and ν_ω be two measurable functions defined in $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ respectively. Then*

$$C\left(\int_\Omega \mu_\omega d\lambda, \int_\Omega \nu_\omega d\lambda \right) \leq \int_\Omega C(\mu_\omega, \nu_\omega) d\lambda.$$

The Kantorovich problem as relaxation of the Monge one

In the first formulation of the Optimal Transportation Problem, Monge utilized the notation of push-forwards to define the transport between measures. This formulation is difficult to study in many ways. Later, Kantorovich was able to give an alternative formulation of the same problem which included the older one as special types of transportation plans, the deterministic ones. Those formulations, however, are closely related. Under mild assumption over the starting measure μ, it is possible to show that the infimum of the Monge problem is equal to the minimum of the Kantorovich one.

Theorem 2.15. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and let us assume that μ is absolutely continuous. Then, there exists at least one measurable function $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_\# \mu = \nu$.*

Assuming μ absolutely continuous already assures us that the set of maps on which we have to minimize is not empty, as we highlighted in Example 2.4, making the Monge problem well posed. By adding some requirements on the supports of μ and ν, we can characterize the Kantorovich problem as a relaxation of the Monge one.

Theorem 2.16. *Let $\Omega \subset \mathbb{R}^d$ be a compact set. If $\mu, \nu \in \mathcal{P}(\Omega)$ and μ is absolutely continuous, then the set of all deterministic plans*

$$D(\mu, \nu) := \Big\{ \pi = \big(Id_X, T\big)_\# \mu, \quad \text{such that} \quad T_\# \mu = \nu \Big\}$$

is dense (with respect to the weak topology) in $\Pi(\mu, \nu)$. In particular, we have that

$$\inf_{\pi \in \Pi(\mu,\nu)} J(\pi) = \min_{\pi \in \Pi(\mu,\nu)} K(\pi),$$

where J is defined as

$$J(\pi) = \begin{cases} \mathbb{T}_c(\pi) & \exists T : X \to Y \quad \text{such that} \quad \pi = \pi_T, \\ +\infty & \quad \text{otherwise} \end{cases}$$

where π_T is the transportation plan induced by T defined in Remark 2.9.

2.2.1 Wasserstein Distance

Given a Polish space (X, d), we can use the optimal transportation problem to define a distance over the space $\mathcal{P}(X)$. In particular, we can lift the distance d from X to the space $\mathcal{P}(X)$, by choosing d as a cost function in (2.9).

Definition 2.26 (Wasserstein Distance)**.** *Let (X, d) be a polish space and $p \in [1, \infty)$. The $p-$order Wasserstein distance between the probability measures μ and ν on X is defined as*

$$W^p_{d^p}(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \mathbb{T}_{d^p}(\pi).$$

When $p = 1$, the $1-$Wasserstein distance is also known as Kantorovich-Rubistein distance.

Remark 2.14. *When $X \subset \mathbb{R}^n$ and $d = |\circ|$, we denote the $p-$order Wasserstein distance with W^p, without specifying the distance.*

The Wasserstein distance $W^p_{d^p}$ is well defined when used to compare measures whose $p-$th moments are finite.

Definition 2.27 (Wasserstein Space)**.** *Within the same context of Definition 2.26, the $p-$Wasserstein space is defined as*

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

where x_0 is an arbitrary point of X.

Remark 2.15. *The space P_p does not depend on the choice of x_0. In fact, given another point x_1, if we know that μ is such that*

$$\int_X d^p(x, x_0) d\mu < +\infty$$

then, since d is a distance

$$\begin{aligned} \int_X d^p(x, x_1) d\mu &\leq \int_X (d(x, x_0) + d(x_0, x_1))^p d\mu \\ &\leq 2^p \int_X d^p(x, x_0) d\mu + 2^p \int_X d^p(x_0, x_1) d\mu \\ &\leq 2^p \int_X d^p(x, x_0) d\mu + 2^p d^p(x_0, x_1) < +\infty. \end{aligned}$$

Theorem 2.17. *The W_{d^p} distance is a finite distance over P_p.*

Proof. Let us take $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(X)$ and π_1,π_2 two optimal transportation plans between μ_1,μ_2 and μ_2,μ_3 respectively.
Since π_1 and π_2 share a common marginal, μ_2, we can find a probability measure $\tilde{\pi} \in \mathcal{P}(X \times X \times X)$ such that

$$\pi_1 = (\mathfrak{p}_{X_1 \times X_2})_\# \tilde{\pi}$$

and

$$\pi_2 = (\mathfrak{p}_{X_2 \times X_3})_\# \tilde{\pi}.$$

We can then define

$$\pi = (\mathfrak{p}_{X_1 \times X_3})_\# \tilde{\pi}.$$

It is easy to check that $\pi \in \Pi(\mu_1, \mu_3)$ so that

$$\begin{aligned}
W_{d^p}(\mu_1, \mu_3) &\leq \left(\int_{X \times X} d^p(\mathbf{x}_1, \mathbf{x}_3) d\pi \right)^{\frac{1}{p}} \\
&= \left(\int_{X \times X \times X} d^p(\mathbf{x}_1, \mathbf{x}_3) d\tilde{\pi} \right)^{\frac{1}{p}} \\
&\leq \left(\int_{X \times X \times X} (d(\mathbf{x}_1, \mathbf{x}_2) + d(\mathbf{x}_2, \mathbf{x}_3))^p d\tilde{\pi} \right)^{\frac{1}{p}} \qquad (2.12) \\
&\leq \left(\int_{X \times X \times X} d^p(\mathbf{x}_1, \mathbf{x}_2) d\tilde{\pi} \right)^{\frac{1}{p}} + \left(\int_{X \times X \times X} d^p(\mathbf{x}_2, \mathbf{x}_3) d\tilde{\pi} \right)^{\frac{1}{p}} \\
&= \left(\int_{X \times X} d^p(\mathbf{x}_1, \mathbf{x}_2) d\pi_1 \right)^{\frac{1}{p}} + \left(\int_{X \times X} d^p(\mathbf{x}_2, \mathbf{x}_3) d\pi \right)^{\frac{1}{p}} \\
&= W_{d^p}(\mu_1, \mu_2) + W_{d^p}(\mu_2, \mu_3),
\end{aligned}$$

where, in (2.12) we used the Minkowsky's inequality. □

Whenever d is continuous, it is easy to see that the Wasserstein distance is not equivalent to the total variation one. In fact, let us take the sequence introduced in Example 2.2, for any $n \in \mathbb{N}$, the only element of $\Pi(\mu_n, \mu)$ is $\delta_{\frac{1}{n}, 0}$, hence

$$W_d(\delta_{\frac{1}{n}}, \delta_0) = d\Big(\frac{1}{n}, 0\Big) = 0$$

and, since d is continuous, we get

$$\lim_{n \to +\infty} W_d(\delta_{\frac{1}{n}}, \delta_0) = \lim_{n \to +\infty} d\Big(\frac{1}{n}, 0\Big) = 0$$

while, according to the total variation, there was no convergence. Those distances induce a different notion of convergence and, hence, a different topology. In particular, we have the following result.

Theorem 2.18. *When the set X is bounded, the W_{d^p} distance induces the weak topology on the $p-$Wasserstein space $P_p(X)$. Moreover, $(P_p(X), W_{d^p})$ is a polish space.*

2.2.2 The one-dimensional case

When the two measures μ and ν are supported on the same line and the cost function is convex, the minimal transportation plan can be computed through a closed formula. This comes from the fact that, unlike the other euclidean spaces, $\mathbb{R}$ is totally ordered.

Definition 2.28. *Given $\mu, \nu \in \mathcal{P}(\mathbb{R})$, the co-monotone transportation plan γ_{mon} between μ and ν is defined as*

$$\gamma_{mon} := (F_\mu^{[-1]}, F_\nu^{[-1]})_{\#}\mathcal{L}_{|[0,1]}.$$

Remark 2.16. *From the chain rule for push-forwards and Lemma 2.4, we have $\gamma_{mon} \in \Pi(\mu, \nu)$ for any $\mu, \nu \in \mathcal{P}(\mathbb{R})$.*

Roughly speaking, the monotone transportation plan moves the mass of μ preserving the order, i.e. if two points in the support of μ have a certain order, so will be the order of their arrival points.

Lemma 2.7 (Santambrogio [76], Chapter 2, Lemma 2.8)**.** *Let us take $\mu, \nu \in \mathcal{P}(\mathbb{R})$. If $\gamma \in \Pi(\mu, \nu)$ is monotone, i.e. if $(x, y), (x', y') \in spt(\gamma)$, then*

$$x < x' \to y < y'$$

then $\gamma = \gamma_{mon}$. Furthermore, if μ is absolutely continuous the optimal plan π is induced by the pushforward through the monotone non decreasing map

$$T_{mon} := F_\nu^{[-1]} \circ F_\mu. \tag{2.13}$$

When the cost function c is a positive and increasing convex function of the Euclidean distance, this property coincides with the $c-$cyclically monotonicity, and so, from Theorem 2.11, the monotone transportation plan is optimal.

Theorem 2.19 (Optimality of the co-monotone plan, Santambrogio [76], Chapter 2, Theorem 2.9)**.** *Let $h : \mathbb{R} \to \mathbb{R}_+$ be a strictly convex function such that $h(0) = 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{R})$. Consider the cost*

$$c(x, y) = h(|x - y|)$$

and suppose that this cost is feasible for the transportation problem. Then the problem has a unique solution which is γ_{mon}. Furthermore, if μ is absolutely continuous the plan is inducted by the function T_{mon} defined in (2.13)*.*

Remark 2.17. *If we drop the assumption of strict convexity of h, γ_{mon} is still an optimal transportation plan, but the uniqueness is not guaranteed anymore.*

Since the optimal transportation plan can be written as the push-forward of a measure, also the Wasserstein distance can be expressed through a closed integral formula.

Corollary 2.2 (Santambrogio [76], Chapter 2, Proposition 2.17)**.** *If $c(x, y) = (|x - y|)^p$, with $p \geq 1$, then*

$$W_p^p(\mu, \nu) = \int_{[0,1]} (|F_\mu^{[-1]} - F_\nu^{[-1]}|)^p d\mathcal{L}.$$

Moreover, for $p = 1$,

$$W_1(\mu, \nu) = \int_{\mathbb{R}} |F_\mu(t) - F_\nu(t)| dt.$$

Unfortunately, for higher-dimensional problems, this formula does not have an equivalent, however, due to its simple computation, monotone transportation maps have been used to build transportation plans and give estimations. We recall the Knothe-Rosenblatt rearrangement [44, 71], which uses monotone plans between marginals and conditional laws to define a transportation plan between two generic measures. Although this plan has nothing to deal with the optimal one, it has been used to prove the isoperimetric inequality.

Remark 2.18 (The Knothe-Rosenblatt Rearrangement)**.** *Let μ and ν be two absolutely continuous measures over $\mathbb{R}^2$. Let μ_1 and ν_1 be the marginal of μ and ν, respectively. Since μ_1 and ν_1 are absolutely continuous, from Theorem 2.19, we can recover the monotone map that sends μ in ν, which we denote T_1.*
Let us now consider the conditional laws with respect to the first variable, $\mu_{|x_1}$ and $\nu_{|y_1}$, of μ and ν, respectively. For any $x_1 \in \mathbb{R}$, we can find an optimal transportation map T_{x_1} between $\mu_{|x_1}$ and $\nu_{|y_1}$, where y_1 is defined as $T_1(x_1)$. We can then consider the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$T(\boldsymbol{x}) := (T_1(x_1), T_{x_1}(x_2))$$

and define the Knothe-Rosenblatt rearrangement as

$$\pi_{KR} := (Id_{\boldsymbol{x}}, T(\boldsymbol{x}))_{\#}\mu. \tag{2.14}$$

2.3 Wasserstein Distance in Image Classification

Due to its wide applicability in practical problems, the optimal transportation problem has become a topic of major interest in many fields: in statistic and probability it has been used in assessment of goodness of fit between distributions [84, 62], in Machine Learning as a loss function in Generative Adversarial Networks (GAN) [5], in unsupervised domain adaptation [26] and as a loss function for learning probability distributions [34], and also in medical fields as a way to compare cytometry diagrams [13, 67] or in radiation therapy [42].
An image can be represented as a discrete probability over a bi-dimensional regular grid $G_N := \Big\{0, \dots, N\Big\}^2$. In this setting, each element of G_N will be a couple of natural numbers (i, j), which represents a bin (or a pixel) of the image. To each pixel, it is associated a positive, usually ranging between 0 and 255, value that represents the intensity of the grey in the original image. The measure obtained in this way is then normalized, so that the transportation problem between any pair of images is well posed.
Given a distance between points of G_N, the dichotomy between discrete measures and images allow to use the Wasserstein distance as a comparison criterion. The problem that defines this distance, also called Hitchcock-Koopmans transportation problem [32], is an hystorical problem in mathematical programming [81] and played a fundamental role in the development of the Network Simplex Algorithm and all the related network flow problem [36, 2, 81].

The Bi-partite Graph

Given two measures μ and ν over two discrete sets X and Y, we can solve the transportation problem between those measures by formulating it as a minimum cost flow problem on a directed graph.

Definition 2.29 (Directed Graph)**.** *A directed graph is a couple $G := (V, E)$, where V is a set of points, vertex of the graph, and $E \subset V \times V$ is the set of edges, or arcs.*

Remark 2.19. *Given a directed graph (V, E), each element $e \in E$ has the form*

$$e := (x, y),$$

where x is the starting point and y is the ending point of the arc. Roughly speaking, each element of E can be thought of as a directed path that goes *from x to y. In this setting, the cost function tells us how expensive it is to use a certain path.*

Definition 2.30 ($b-$flows)**.** *Let us take a graph $G := (V, E)$ and a cost function c. Given any function $b : V \to \mathbb{R}$ such that*

$$\sum_{v \in V} b_v = 0, \tag{2.15}$$

we define a $b-$flow as a function $f : E \to [0, +\infty)$ such that

$$\sum_{v \text{ s.t. } (v,u) \in E} f_{v,u} - \sum_{v \text{ s.t. } (u,v) \in E} f_{u,v} = b_u. \tag{2.16}$$

We denote with $\mathcal{F}(G, b)$ the set of all the $b-$flows.

Remark 2.20. *A $b-$flow describes the movement of mass along all the arcs, $f_{u,v}$ is the amount of mass that goes from u to v. We can then think of the equation* (2.16) *as a continuity one:*

- $\sum_{v \text{ s.t. } (v,u) \in E} f_{v,u}$ *is the amount of mass entering in the node u;*
- $\sum_{v \text{ s.t. } (u,v) \in E} f_{u,v}$ *is the amount of mass leaving the node u;*

hence (2.16) *tells us that b_u is the change of mass in the node u. Condition* (2.15) *tells us that there is no dissipation nor creation of mass during the process.*

Problem 2.1. *Let us take a graph $G = (V, E)$, a flow $b : V \to \mathbb{R}$ and $c : E \to \mathbb{R}$ a cost on the arcs. The uncapacitated minimum cost flow problem consists in finding the $b-$flow that minimizes the flow cost*

$$\sum_{e \in E} c_e f_e.$$

The minimum of this problem will be denoted with $F_{G,c}(b)$.

Definition 2.31 (Sub-Graph)**.** *Given a graph $G' := (V', E')$, we say that $G = (V, E)$ is a sub-graph of G' if*

$$V \subset V' \quad \textit{and} \quad E \subset E'.$$

Let G' be a graph and G a sub-graph of G'. By restricting them, we can associate any function on G' to a function over G. In particular, any uncapacitated $b-$flow problem over G' defines a $b-$flow problem over any of its sub-graphs and, for those problems it holds

$$F_{G,c}(b) \geq F_{G',c}(b).$$

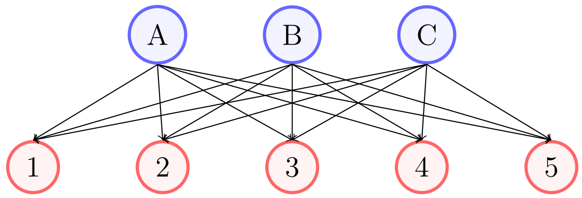


Figure 2.2: An example of bi-partite graph. Each point of the set $X := \Big\{A, B, C\Big\}$ is connected to each point of $Y := \Big\{1, 2, 3, 4, 5\Big\}$.

Definition 2.32 (Bi-partite Graph)**.** *Given two discrete sets X and Y, we define the bi-partite graph $G_{X\to Y}$ as the couple $(X \cup Y, E_{X\to Y})$, where*

$$E_{X\to Y} := \Big\{(x, y) : x \in X, y \in Y\Big\}.$$

Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, we can define $b : X \cup Y \to \mathbb{R}$ as

$$b_u := \begin{cases} \mu_x & \text{if} \quad u = x \in X, \\ -\nu_y & \text{if} \quad u = y \in Y. \end{cases}$$

In this setting, any $b-$flow can be seen as a transportation plan between μ and ν. Given $f \in \mathcal{F}(G_{X\to Y}, b)$, if $u = x \in X$, we have

$$\mu_x = b_u = \sum_{v \text{ s.t. } (u,v)\in E_{X\to Y}} f_{u,v} - \sum_{v \text{ s.t. } (v,u)\in E_{X\to Y}} f_{v,u} = \sum_{v \text{ s.t. } (u,v)\in E_{X\to Y}} f_{u,v} \tag{2.17}$$

while, if $u = y \in Y$, we have

$$\nu_y = b_u = \sum_{v \text{ s.t. } (u,v)\in E_{X\to Y}} f_{u,v} - \sum_{v \text{ s.t. } (v,u)\in E_{X\to Y}} f_{v,u} = - \sum_{v \text{ s.t. } (v,u)\in E_{X\to Y}} f_{v,u}, \tag{2.18}$$

hence, if we define

$$\pi_{x,y} = f_e = f_{(x,y)},$$

by (2.17) and (2.18), we conclude $\pi \in \Pi(\mu, \nu)$. Vice versa, if $\pi \in \Pi(\mu, \nu)$, we can define a $b-$flow by setting

$$f_e := \pi_{x,y}$$

where $e = (x, y)$. Hence $\mathcal{F}(G_{X\to Y}, b) = \Pi(\mu, \nu)$, and, as a consequence,

$$F_{G_{X\to Y},c}(b) = W_c(\mu, \nu)$$

for any cost function c.
Although the bi-partite graph allows us to compute exactly the Wasserstein distance between two measures, often it is too clunky to be practical. This issue comes from the fact that, when $X = Y = G_N$ the set of arcs (and, consequentially, of unknowns) is $N^2 \times N^2$, since each point of X has to be connected to each point of Y. The situation gets worse as the dimension of the grid grows.
Since each grid contains $(N+1)^2$ points the solution of the problem can be tough to compute, making the whole process unappealing, especially in fields where the time is an issue and having a fast answer (even if not completely correct) is a necessity. For this reason, through the recent years, many alternatives were proposed to this classical method.

The Fast EMD

Ling and Okada proposed in their work [51] an alternative way to compute the W_1 distance between measures. This reformulation shows a lot of improvements with respect to the classical bi-partite graph method presented above. First of all, it only needs to compute an $O(N)$ number of unknowns, which is way less then $O(N^2)$. As other considerable upsides, it has a half of the usual number of constraints and it does not require to store any cost matrix.
Ling and Okada considered the case in which both measures are supported on two-dimensional regular grids (or histograms, as they call them). In this setting, they exploit the structure of the l^1 norm on those grids. The key feature used is the linearity of this function once it is fixed a direction. As Figure 2.3 shows, this feature allows us to describe any movement as a succession of smaller movements.
Thanks to this feature, the authors were able to focus only on the movements between adjacent points, i.e. the points whose l^1-distance is unitary.

Definition 2.33. *Let us take G_N a regular $N \times N$ grid. We denote with $\mathcal{I}_1$ the set of all the arcs in $G_N \times G_N$ whose l^1-cost is unitary, i.e.*

$$\mathcal{I}_1 := \Big\{(i,j;k,l) \in G_N \times G_N \quad s.t. \quad |i-k| + |j-l| = 1\Big\} \tag{2.19}$$

As we anticipated above, the l^1 norm well behaves with respect to this set of arcs. In fact, given any arc connecting two points $(i,j),(k,l) \in G_N$, thanks to the property of module, we can always decompose it as a sequence of arcs

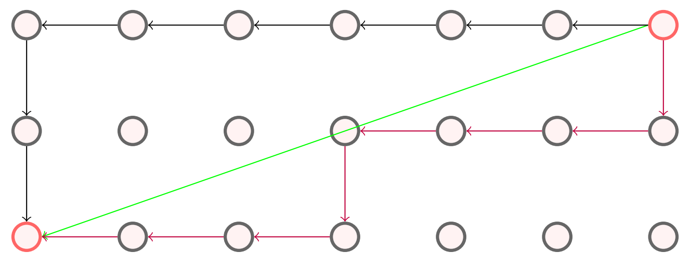


Figure 2.3: The green path connecting the two red points can be decomposed in a plethora of subsequential smaller paths. In the picture, we present two of them (in black and magenta). It is important to notice that the total cost of those sequences of paths is equal to the cost of the green path.

in $\mathcal{I}_1$ without changing the total cost, in fact

$$|i-k|+|j-l| := \sum_{\alpha=i}^{k-1} |\alpha-(\alpha+1)| + \sum_{\beta=j}^{l-1} |\beta-(\beta+1)|,$$

where, without loss of generality, we supposed $i \leq j$ and $k \leq l$.

Definition 2.34 (l^1 Flows)**.** *Let us take two measures μ and ν on the regular grid $G = I_1 \times I_2$. A collection of real values $\{g_{i,j,k,l}\}$ is a feasible flow between μ and ν if, given any $(i,j) \in G_N$,*

$$\sum_{k,l \text{ s.t. } (k,l;i,j)\in\mathcal{I}_1} g_{k,l;i,j} - \sum_{k,l \text{ s.t. } (i,j;k,l)\in\mathcal{I}_1} g_{i,j;k,l} = b_{i,j}$$

where $b_{i,j} := \nu_{i,j} - \mu_{i,j}$. We denote with $G(\mu,\nu)$ the set of all the feasible flows between μ and ν.

Definition 2.35 (l^1 Flow Functional)**.** *The l^1 Flow Functional $\mathcal{G} : G(\mu,\nu) \to [0,+\infty)$ is defined as*

$$\mathcal{G}(g) := \sum_{(i,j;k,l)\in\mathcal{I}_1} g_{i,j;k,l}.$$

In their main result, Ling and Okada proved that minimizing the functional $\mathcal{G}$ is equivalent to minimizing the transportation functional associated to the cost function

$$c(\mathbf{x},\mathbf{y}) := |x_1-y_1| + |x_2-y_2|.$$

Theorem 2.20. *Given μ and ν two measures, we have*

$$\min_{\pi\in\Pi(\mu,\nu)} \sum_{(i,j;k,l)\in G\times G} \Big(|i-j|+|k-l|\Big)\pi_{i,j;k,l} = \min_{g\in G(\mu,\nu)} \mathcal{G}(g).$$

In their proof, it is showed how from a transportation plan it is always possible to retrieve a l^1 flow and vice versa. Although there is no closed formula to retrieve a transportation plan from an l^1-flow, this framework gives a more concise description of how the mass moves since it needs only $4N^2$ variables to compute. Moreover, we do not need to store any cost function, since each flow travels along an arc that has a unitary cost.

Reduced Graphs

Bassetti, Gualandi, and Veneroni in [8] generalized the idea behind the l^1-flows to compute the Wasserstein distance associated with each l^p-norms. To this intent, they proposed a family of sub-graphs of the bi-partite one depending on a parameter L, that ranges from 0 to N, and determines what directions are allowed.

Definition 2.36. *Given $L \in \mathbb{N}$, we define the set of co-prime directions as*

$$V_0 := \big\{(1,0),(0,1),(-1,0),(0,-1)\big\}$$

for $L = 0$ and

$$V_L := V_0 \cup \big\{(i_1,i_2) \in \mathbb{Z}_0 \times \mathbb{Z}_0 \text{ s.t. } |i_1| \leq L, |i_2| \leq L, \text{ s.t } |i_1|,|i_2| \text{ are co-prime.}\big\}.$$

Fixed L as above, the set of arcs E consists in each couple of $(x,y) \in G_N \times G_N$, whose difference $x - y$ lies in V_L. When $L = 0$, we find the set I_1 defined in 2.19.

Definition 2.37 ($L-$reduced Graphs)**.** *Let G_N be a $N \times N$ grid. For any $L \in \mathbb{N}$, we define the $L-$reduced graph as the couple*

$$G_L := \big(G_N \cup G_N, E_L\big)$$

where

$$E_L := \bigcup_{(i_1,i_2)\in G_N} \Big\{\Big((i_1,i_2),(i_1,i_2)+V_L\Big)\Big\}$$

This graphs are suitable for approximating the transportation costs induced from a $1-$homogeneous costs, as the ones induced from l^p-norms

$$c_p(x,y) := \sqrt[p]{(|x_1-y_1|)^p + (|x_2-y_2|)^p},$$

Similarly to what happened for the l^1-norm, the homogeneity allows us to decompose any movement along a line as a sequence of smaller ones,

but, unlike what happened for the l^1-norm, we have to pick every possible independent direction to be sure of computing the exact Wasserstein distance. Picking every direction corresponds to $L = N$, in general, when $L < N$, we obtain an upper bound of the desired distance, which becomes more precise as L grows.
Through this generalization, Bassetti et al. were able to prove a result similar to the one proved by Ling and Okada for the $l^\infty-$norm

$$c_\infty(x, y) := \max\big\{|x_1 - y_1|, |x_2 - y_2|\big\}.$$

Similarly to what happens for the l_1 one, it is possible the exact computation of the Wasserstein distance by taking $L = 1$.
Even if for the other l_p-norms it is not possible to retrieve the exact value of the Wasserstein distance if $L < N$, it is possible to obtain an a priori estimation on the error we commit. The authors focused on the case $p = 2$, for which proved that

$$\frac{|W_{l_2}(\mu,\nu) - F_{G_L,c_2}(\mu-\nu)|}{|W_{l_2}(\mu,\nu)|} \leq c(L)\left(1 - \frac{L}{\sqrt{1+L^2}}\right)$$

where

$$c(L) = \left(2 + \sqrt{2 + \frac{2L}{\sqrt{1+L^2}}}\right)^{-1}.$$

The Fast and robust EMD

Pele and Werman proposed in [69] a new family of transportation problems in which the ground distance is truncated at a given threshold t. This allows the simplification of another family of distances introduces by the same authors, the $\widehat{EMD}_\alpha$.

Definition 2.38 ($\widehat{EMD}_\alpha$ Distance)**.** *Let us take μ and ν two measures over discrete spaces X and Y and c a cost function between those spaces. Given $\alpha \geq 0$, the $\widehat{EMD}_\alpha$ distance is defined as*

$$\widehat{EMD}_\alpha(\mu,\nu) := \min_{\pi\in\mathfrak{J}(\mu,\nu)} \Big\{ \sum_{x\in X, y\in Y} c_{x,y}\pi_{x,y}\Big\} + \alpha C\Big|\sum_{x\in X}\mu_x - \sum_{y\in Y}\nu_y\Big|$$

where $C := \max_{(x,y)\in X\times Y} c_{x,y}$ *and*

$$\mathfrak{J}(\mu,\nu) := \left\{\pi \in \mathcal{M}(X\times Y) \;\middle|\; \sum_{y\in Y}\pi_{x,y} \leq \mu_x, \qquad \sum_{x\in X}\pi_{x,y} \leq \nu_y,\right.$$

$$\sum_{(x,y)\in X\times Y} \pi_{x,y} \quad = \quad \max\left\{ \sum_{x\in X} \mu_x, \sum_{y\in Y} \nu_y \right\}\Bigg\}.$$

The authors proved that, for any $\alpha > 0$ this is indeed a distance, but, unlike the classical Wasserstein distance, it can compare measures with different mass. The $\widehat{EMD}_\alpha$ can be be computed through a bi-partite graph at which it was added an auxiliary node, the sink. This sink is connected to each node of μ, has a cost of αD and takes care of all the mass that cannot be moved from μ to ν. It is worth of notice that, when $\alpha = 0$ or μ and ν have the same mass, the $\widehat{EMD}_\alpha$ is equal to the classical Wasserstein distance associated to the cost function c.
To reduce the number of arcs of this formulation, the authors introduced a threshold t and considered the $t-$truncation of the cost function.

Definition 2.39 (t-truncated Cost Function)**.** *Let c be a cost function on $X \times Y$. Given a positive threshold $t \in \mathbb{R}^+$, we define the* t-truncated *cost function $c^{(t)}$ as*

$$c^{(t)}_{x,y} := \min\ \left\{c_{x,y}, t\right\} \qquad \forall x \in X, \quad \forall y \in Y.$$

This allowed the authors to reduce the number of arcs of the graph by removing every arc whose cost was higher than t and introducing another auxiliary node, which is connected to all the points of both grids. Roughly speaking, the arcs connecting to this new auxiliary point will take care of all the mass that would have traveled on the removed arcs, for this reason the cost of reaching this node is equal to the threshold t, and the cost of leaving it is null.

Remark 2.21. *The arcs connecting the two auxiliary nodes added to the reduced bi-partite graph have all cost equal to t. If we know how much mass travels through the other arcs, we could deduce how much does it cost to travel the remaining one by multiplying this amount for a suitable constant. Since it is not possible know that a priori, this models needs those arcs to implement the constraints on the flow.*
In Chapter 4, we will show how, by changing approach to the problem, it is indeed possible to get rid of those arcs.

Unlike the previous methods, this one simplifies the formulation by approximating the cost function. Since the truncated functions have a simpler structure, the whole computation benefits as well. Moreover, since $c^{[t]} \leq c$, this approximation gives us a lower bound on the Wasserstein distance.

The Sinkhorn Distance

The Sinkhorn distance is the most used approximation of the Wasserstein distance. It is obtained through an entropic regularization of the Optimal Transportation problem, which makes the problem strictly convex.
The convexity benefits the resolution by granting the uniqueness of the solution and by allowing the use of techniques proper of convex optimization. In particular, the author used the Sinkhorn-Knopp's algorithm, after which the distance has been named.

Definition 2.40 (Logarithmic Entropy)**.** *Given a discrete measure μ_x over a finite set X, we define the entropy of μ as*

$$h(\mu) := -\sum_{x\in X} \mu_x \log(\mu_x).$$

Remark 2.22. *From Entropy Theory we know that if $\pi \in \Pi(\mu, \nu)$, this classical bound holds*

$$h(\pi) \leq h(\mu) + h(\nu) \tag{2.20}$$

which is sharp, since the equality holds for $\pi = \mu \otimes \nu$.

Through the logarithmic entropy, it is possible to define the Kullback-Leiber divergence, a classical function that measures the relative entropy between measures.

Definition 2.41 (Kullback-Leiber divergence)**.** *Given a discrete polish space X, we define the Kullback-Leiber divergence $KL : \mathcal{P}(X) \times \mathcal{P}(X) \to [0, +\infty]$ as*

$$KL(\pi||\rho) := \sum_{x\in X} \pi_x \log\left(\frac{\pi_x}{\rho_x}\right)$$

for any pair $\pi, \rho \in \Pi(\mu, \nu)$.

The author used this divergence to introduce a further constraint on the set of transportation plans between two measures. This will smooth the set on which we have to minimize the functional. Since we will deal with discrete images, from now on we will assume $X = G_N$ (and $X = G_N \times G_N$) so that X will be a couple of integers (or a couple of couples for the transportation plans).

Definition 2.42 (Transportation Plans with Entropic constraints)**.** *Given two probability measures μ, ν and $\alpha \geq 0$, we define*

$$U_\alpha(\mu, \nu) := \Big\{\pi \in \Pi(\mu, \nu) \ s.t. \ KL(\pi||\mu \otimes \nu) \leq \alpha\Big\}. \tag{2.21}$$

Thank to Remark 2.22, we have that $U_\alpha(\mu, \nu)$ is not empty.

Remark 2.23. *It is possible to characterize the sets in* (2.21) *through the logarithmic entropy*

$$U_\alpha(\mu,\nu) := \Big\{\pi \in \Pi(\mu,\nu) \ s.t.\ h(\pi) \geq h(\mu) + h(\nu) - \alpha\Big\}.$$

Lemma 2.8 (Gluing Lemma with entropic contrastraints)**.** *Given* $\alpha \geq 0$*, let* μ*,* ν *and* ζ *three probability measures over a discrete set* X*. If* $\pi \in U_\alpha(\mu,\nu)$ *and* $\rho \in U_\alpha(\nu,\zeta)$*, then the measure defined as*

$$\gamma_{i,k} := \sum_j \frac{\pi_{i,j}\rho_{j,k}}{\nu_j}$$

is an element of $U_\alpha(\mu,\zeta)$*.*

Definition 2.43 (Sinkhorn distance)**.** *Let* c *be a cost function and* α *a positive value. Given two probability measures* μ *and* ν*, the Sinkhorn distance is defined as*

$$d_{c,\alpha}(\mu,\nu) := \min_{\pi \in U_\alpha(\mu,\nu)} \sum_{x,y} c_{x,y}\pi_{x,y}.$$

Remark 2.24. *When* α *is large enough, the bound* (2.20) *ensures us that*

$$U_\alpha(\mu,\nu) = \Pi(\mu,\nu),$$

hence the Sinkhorn distance coincides with the usual transportation cost. On the other hand, when $\alpha = 0$*, we have that* $U_\alpha(\mu,\nu) := \big\{\mu \otimes \nu\big\}$*, hence, in this case*

$$d_{c,\alpha}(\mu,\nu) := \sum_{x,y} c_{x,y}\mu_x\nu_y.$$

By using the Lemma 2.8, we can prove that, for any given $\alpha \geq 0$ the function $d_{c,\alpha}$ satisfy the triangular inequality. Moreover, it is easy to see that the Sinkhorn distance is symmetric. The coincidence axiom, however, does not hold true, so that, in general $d_{c,\alpha}(\mu,\mu) \neq 0$. For instance, take $\alpha = 0$ and any probability measure μ that is not a Dirac's delta. This issue can be fixed by multiplying the Sinkhorn distance by a suitable indicator function, as the following Theorem states.

Theorem 2.21. *For all* $\alpha \geq 0$*, the function*

$$(\mu,\nu) \to d_{c,\alpha}(\mu,\nu)\chi_{\mu\neq\nu}$$

it is a distance on $\mathcal{P}(X)$*.*

From duality theory, it is possible to show that given any $\alpha \geq 0$ there exists a $\lambda \geq 0$ for which holds true

$$d_{c,\alpha}(\mu,\nu) = d_c^\lambda(\mu,\nu) := \sum_{x,y} c_{x,y}\pi^\lambda_{x,y}$$

where

$$\pi^\lambda := \underset{\pi\in\Pi(\mu,\nu)}{\arg\min} \sum_{x,y} c_{x,y}\pi_{x,y} - \frac{1}{\lambda}h(\pi). \tag{2.22}$$

The problem defining π^λ is an entropic regularization of the classical transportation problem, which is now strictly convex thanks to the convexity of the entropy functional h. This makes the computation of d_c^λ much easier since it is now possible to use all the tools from convex optimization. From the Sinkhorn and Knopp's theorem we can deduce that the solution π^λ is the only transportation plan $\pi \in \Pi(\mu,\nu)$ of the form

$$\pi^\lambda = diag(u)e^{-\lambda M}diag(v) \tag{2.23}$$

where

- u and v are vectors whose entries are all non negative;
- M is the matrix whose (i,j) component is the cost between the i^{th} and the j^{th} point and $e^{-\lambda M}$ is matrix whose entries are $e^{-\lambda c(x_i,y_j)}$;
- $diag(u)$ and $diag(v)$ are the matrix whose diagonal contains the elements of u and v respectively.

Thank to the Sinkhorn and Knopp's algorithm it is possible to compute the solution quickly, since the whole process can be parallelized. To go back from λ to α is harder, and has to be done iteratively until we find a π^λ whose entropy is close to $h(\mu)+h(\nu)-\alpha$. However, it is possible to see from (2.22) that the entropy of the solution π^λ decreases monotonically as λ increases, hence it is possible to apply a plethora of method to find the desired λ.
When the cost function is the square of the distance d, the objective function in (2.22) reads as

$$\sum_{i,j} d^2_{i,j}\pi_{i,j} + \gamma\sum_{i,j}\pi_{i,j}\log(\pi_{i,j}) \tag{2.24}$$

hence, by defining

$$\mathcal{K}^\gamma_{i,j} := e^{-\frac{d^2_{i,j}}{\gamma}}, \tag{2.25}$$

we can rewrite (2.24) as

$$\begin{aligned} -\gamma \sum_{i,j} \log(\mathcal{K}^{\gamma}_{i,j})\pi_{i,j} + \gamma \sum_{i,j} \pi_{i,j} \log(\pi_{i,j}) &= -\gamma \sum_{i,j} \pi_{i,j} \log\left(\frac{\pi_{i,j}}{\mathcal{K}^{\gamma}_{i,j}}\right) \\ &= \gamma \Big[1 + KL(\pi || \mathcal{K}^{\gamma}) \Big]. \end{aligned}$$

The Sinkhorn distance is then equal to

$$d^{\gamma}_{c}(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} \gamma \Big[1 + KL(\pi || \mathcal{K}^{\gamma}) \Big],$$

which is indeed strictly convex.
When the supports of the two measures coincide and has a regular structure (for the sake of simplicity assume them supported on G_N), there is an interesting link between the kernel $\mathcal{K}^{\gamma}$ (defined in (2.25)) and the discretized heat kernel.
We recall that the heat kernel is a time dependent family of function that describe the diffusion between points of a given set. In particular, it holds true that

$$f_t(x) := \int_{\Omega} f_0(y) H_t(x,y) dy, \tag{2.26}$$

where f_t is the solution of the heat equation on a regular set $\Omega \subset \mathbb{R}^2$, i.e.

$$\partial_t f_t = \Delta f_t \qquad \text{on} \quad \Omega \times [0,T]$$

with initial condition f_0 and homogeneous Dirichlet boundary condition.
If we take $\Omega = [0,N]^2$, we can discretize this set and the equation on it as it follow

$$\begin{aligned} f_t(x) &\approx f^{(t)}_{i,j}, \\ f_0(x) &\approx f^{(0)}_{k,l}, \\ H_t(x,y) &\approx H^{(t)}_{i,j,k,l} \end{aligned}$$

hence, equation (2.26) becomes

$$f^{(t)}_{i,j} := \sum_{k,l} f^{0}_{k,l} H^{(t)}_{i,j,k,l}. \tag{2.27}$$

From the heat kernel it is possible to retrieve the distance over the space through the Varadhan's formula, which states

$$d^2(x,y) := \lim_{t \to 0^+} \Big[-2t \log(H_t(x,y)) \Big],$$

which, in the discrete setting becomes

$$d^2_{i,j,k,l} = \lim_{t \to 0^+} \Big[-2t \log(H^{(t)}_{i,j,k,l}) \Big].$$

By definition of $\mathcal{K}^\gamma$, we have

$$d^2_{i,j,k,l} = -\gamma \log(\mathcal{K}^\gamma_{i,j,k,l}),$$

hence, for t small enough we can conclude

$$-\gamma \log(\mathcal{K}^\gamma_{i,j,k,l}) \approx -2t \log(H^{(t)}_{i,j,k,l}) \tag{2.28}$$

so, if we set $\gamma = 2t$, we can simplify relation (2.28) further and obtain

$$\mathcal{K}^\gamma_{i,j,k,l}. \approx H^{\frac{\gamma}{2}}_{i,j,k,l}.$$

The relation between the discretized heat kernel and $\mathcal{K}^\gamma$ can be fruitfully be used to reduce even more the complexity of the problem. In fact, given a vector v, it holds true

$$w_{i,j} = \sum_{k,l} \mathcal{K}^\gamma_{i,j,k,l} v_{k,l} \iff w_{i,j} = \sum_{k,l} H^{\frac{\gamma}{2}}_{i,j,k,l} v_{k,l}.$$

Hence, since $H^{\frac{\gamma}{2}}_{i,j,k,l}$ is the discretized heat kernel, the vector w is the discretization of the solution of the diffusion equation after $\frac{\gamma}{2}$ and with initial data v, which means

$$w_{i,j} = \sum_{k,l} \mathcal{K}^\gamma_{i,j,k,l} v_{k,l} \iff v = \Big(Id + \frac{\gamma}{2} L \Big) w$$

where L is the cotangent Laplacian. The matrix $D + \frac{\gamma}{2} L$ can be pre-factorized before computing the distance, hence, the heat kernel convolution becomes equivalent to a near-linear time back-substitution, which is a huge boost in execution speed since the algorithm apply the convolution repeatedly.

Chapter 3

Separable Cost Functions

As we saw in the previous Chapter, the Wasserstein distance has proved to be a reliable tool in many applied fields, ranging from cytometry flows to image retrieval. This huge range of applications came with a huge request for fast computation methods or, at least, for numerical ways to achieve insightful information about the distance.
Our research aims to improve the answers to this challenging task.
In this section, we will introduce and study the transportation problem related to particular cost functions, the separable ones. For this class of functions, it is possible to reformulate the transportation problem (2.10) in a new fashion. Through this reformulation, we will be able to highlight the significative properties of the optimal transportation plan, allowing us to introduce two new items, the cardinal flow, and the pivot measure.
Moreover, the reformulation has an intuitive interpretation that can be easily translated into a simple network. As we previously pointed out, due to their efficiency and sharp use of the memory, the newest methods are overshadowing more and more methods based on the Network Simplex Method. However, we will show that our model can fruitfully utilize the possibilities offered by the Network Simplex to compute W_p distances, outperforming every existing method, and becoming the new state of the art for computing the Wasserstein distance between measures supported on regular grids of any dimension.

3.1 The Cardinal Flows Formulation

In this section, we introduce the key notion of our study i.e., *separability*. Roughly speaking, a cost function is separable if we can decompose the cost of each movement into two smaller and independent contributions.

This peculiar structure will allow us to introduce a reformulation that mimics the structure of the cost function.

Definition 3.1 (Separable Cost Function). *Let $X = X_1 \times X_2$ and $Y = Y_1 \times Y_2$ be two polish spaces. We say that a cost function $c : X \times Y \to [0, \infty)$ is* semi-separable *if there exist two functions $c_1 : X \times Y_1 \to [0, \infty)$ and $c_2 : X_2 \times Y \to [0, \infty)$ such that*

$$c(\boldsymbol{x}, \boldsymbol{y}) := c_1(\boldsymbol{x}, y_1) + c_2(x_2, \boldsymbol{y}).$$

We say that c is totally separable *(or just* separable*) if the function c_1 does not depend on x_2 and the function c_2 does not depend on y_1, i.e.*

$$c(\boldsymbol{x}, \boldsymbol{y}) := c_1(x_1, y_1) + c_2(x_2, y_2).$$

Remark 3.1. *Semi-separable cost functions have a very undesirable flaw, they are not symmetric. So that, in general, $\mathcal{C}(\mu, \nu) \neq \mathcal{C}(\nu, \mu)$, which is a very undesirable feature. In order to keep the results as generic as possible, for the rest of this section, we will talk of semi-separable cost functions. Notice that if c is totally separable, then it is semi-separable as well, which means that every result stated here will also stand for separable cost functions.*

Remark 3.2. *From now on, we assume that X and Y are of the form $X = X_1 \times X_2$ and $Y = Y_1 \times Y_2$ where X_i and Y_i are polish spaces.*

Definition 3.2 (Cardinal Flows). *Let us take $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We say that the couple of measures $(f^{(1)}, f^{(2)}) \in \mathcal{P}(X \times Y_1) \times \mathcal{P}(X_2 \times Y)$ is a* cardinal flow *between μ and ν if it satisfies the following conditions*

- *The marginal on X of $f^{(1)}$ is equal to the starting measure μ, i.e.*

$$\mu = (\mathfrak{p}_X)_\# f^{(1)}. \tag{3.1}$$

- *The marginal on Y of $f^{(2)}$ is equal to the arriving measure ν, i.e.*

$$\nu = (\mathfrak{p}_Y)_\# f^{(2)}. \tag{3.2}$$

- *$f^{(1)}$ and $f^{(2)}$ have the same marginal on $Y_1 \times X_2$, i.e.*

$$(\mathfrak{p}_{Y_1 \times X_2})_\# f^{(1)} = (\mathfrak{p}_{Y_1 \times X_2})_\# f^{(2)}. \tag{3.3}$$

We call the measures $f^{(1)}$ and $f^{(2)}$ first and second cardinal flow, respectively. We denote with $\mathcal{F}(\mu, \nu)$ the set of all cardinal flows between μ and ν.

Remark 3.3. *For any couple of probability measures μ and ν, the set $\mathcal{F}(\mu,\nu)$ is nonempty. In fact, the couple $(f^{(1)}, f^{(2)})$, defined as*

$$f^{(1)} = \mu \otimes \nu_1$$

and

$$f^{(2)} = \mu_2 \otimes \nu,$$

is an element of $\mathcal{F}(\mu,\nu)$.

Remark 3.4. *The sets $\mathcal{F}(\mu,\nu)$ and $\mathcal{F}(\nu,\mu)$ can be different, even if we take $X = Y$. For instance, just take $X = Y = \mathbb{R}^2$, $\mu = \delta_{(0,0)}$ and $\nu = \delta_{(1,1)}$. In this case we have that $\mathcal{F}(\mu,\nu) = \{\delta_{(0,0);1}\}$ and $\mathcal{F}(\nu,\mu) = \{\delta_{(1,1);0}\}$.*
Since the cost function does not change in any way the definition of cardinal flows, this problem is not related to the missing symmetry of semi-separable cost functions, although, when the cost function is totally separable, this difference has a neat interpretation.

Definition 3.3 (Cardinal Flows Functional)**.** *Given two probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a semi-separable cost $c = c_1 + c_2$ function over $X \times Y$. We define the first and second cardinal transportation functionals as*

$$\mathbb{CT}_c^{(1)}(F) = \int_{X \times Y_1} c_1 df^{(1)}$$

and

$$\mathbb{CT}_c^{(2)}(F) = \int_{X_2 \times Y} c_2 df^{(2)}$$

where $F = (f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$.
The total cardinal flow functional is then defined as the sum of those two, i.e.

$$\mathbb{CT}_c(F) = \mathbb{CT}_c^{(1)}(F) + \mathbb{CT}_c^{(2)}(F).$$

We are now ready to prove the first main result of the Chapter, the equivalence between the minimal cardinal flow problem and the minimal transportation plan one. We will prove that by showing that from each transportation plan π we can recover a cardinal flow $(f^{(1)}, f^{(2)})$ such that

$$\mathbb{T}_c(\pi) = \mathbb{CT}_c(f^{(1)}, f^{(2)})$$

and vice versa. As an interesting consequence, we will see that there is only one cardinal flow induced by a transportation plan, while the opposite is not always true. This will give us some insight into the lack of uniqueness highlighted in Example 2.5.

Theorem 3.1. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and let $c : X \times Y \to [0, \infty)$ be a separable cost function, then*

$$\inf_{\pi \in \Pi(\mu,\nu)} \Big\{ \mathbb{T}_c(\pi) \Big\} = \inf_{(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)} \Big\{ \mathbb{CT}_c(f^{(1)}, f^{(2)}) \Big\}.$$

Proof. Let us consider the function $L : \Pi(\mu, \nu) \to \mathcal{F}(\mu, \nu)$ defined as

$$L(\pi) = ((\mathfrak{p}_{X\times Y_1})_\#(\pi), (\mathfrak{p}_{X_2\times Y})_\#(\pi)) =: (f^{(1)}, f^{(2)}). \tag{3.4}$$

For any given $\pi \in \Pi(\mu, \nu)$ we have that $L(\pi) \in \mathcal{F}(\mu, \nu)$, in fact from the chain rule for pushforwards (Lemma 2.1) we get

$$\mu = (\mathfrak{p}_X)_\#(\pi) = ((\mathfrak{p}_X)_\#((\mathfrak{p}_{X\times Y_1})_\#))(\pi) = ((\mathfrak{p}_X)_\#)f^{(1)} \tag{3.5}$$

and, similarly

$$\nu = ((\mathfrak{p}_Y)_\#)f^{(2)}.$$

To prove that $L(\pi) \in \mathcal{F}(\mu, \nu)$ we need to prove the gluing condition. This follows from the identity

$$((\mathfrak{p}_{X_2\times Y_1}) \circ (\mathfrak{p}_{X\times Y_1}))_\# = (\mathfrak{p}_{X_2\times Y_1})_\#(\mathfrak{p}_{X_2\times Y})_\#$$

and, using once again the chain rule for pushforwards, we get

$$\begin{aligned}
(\mathfrak{p}_{Y_1\times X_2})_\#(f^{(1)}) &= (\mathfrak{p}_{Y_1\times X_2})_\#(\mathfrak{p}_{X\times Y_1})_\#(\pi) \\
&= (\mathfrak{p}_{Y_1\times X_2})_\#(\mathfrak{p}_{X_2\times Y})_\#(\pi) \\
&= (\mathfrak{p}_{Y_1\times X_2})_\#(f^{(2)}),
\end{aligned}$$

which proves that $L(\pi) \in \mathcal{F}(\mu, \nu)$.
By definition of $L(\pi)$ we have that

$$\begin{aligned}
\mathbb{CT}_c(L(\pi)) &= \int_{X\times Y_1} c_1 df^{(1)} + \int_{X_2\times Y} c_2 df^{(2)} \\
&= \int_{X\times Y_1} c_1 d(\mathfrak{p}_{X\times Y_1})_\#(\pi) + \int_{X_2\times Y} c_2 d(\mathfrak{p}_{X_2\times Y})_\#(\pi) \\
&= \int_{X\times Y} c_1(\mathfrak{p}_{X\times Y_1}) d\pi + \int_{X\times Y} c_2(\mathfrak{p}_{X_2\times Y}) d\pi \\
&= \int_{X\times Y} c_1 d\pi + \int_{X\times Y} c_2 d\pi \qquad (3.6) \\
&= \mathbb{T}_c(\pi)
\end{aligned}$$

where (3.6) comes from the fact that c_1 and c_2 does not depend on y_2 and x_1 respectively.
This far, we only proved that

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) = \inf_{(f^{(1)},f^{(2)})\in L(\Pi(\mu,\nu))} \mathbb{CT}_c(f^{(1)}, f^{(2)}). \tag{3.7}$$

To conclude the proof we show that $L(\Pi(\mu,\nu)) = \mathcal{F}(\mu,\nu)$, i.e. that L is surjective. Let us then take $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$. Thanks to condition (3.3) we can apply the gluing lemma and find a measure $\pi \in \mathcal{P}(X \times Y)$ such that

$$(\mathfrak{p}_{X\times Y_1})_\#(\pi) = f^{(1)} \tag{3.8}$$

and

$$(\mathfrak{p}_{X_2\times Y})_\#(\pi) = f^{(2)}. \tag{3.9}$$

Using once again the chain rule for pushforwards and the identity

$$(\mathfrak{p}_X) = (\mathfrak{p}_X) \circ (\mathfrak{p}_{X\times Y_1})$$

we find

$$(\mathfrak{p}_X)_\#(\pi) = (\mathfrak{p}_X)_\#((\mathfrak{p}_{X\times Y_1})_\#(\pi)) = (\mathfrak{p}_X)_\#(f^{(1)}) = \mu,$$

where the last equality follows from (3.1). Similarly we get $(\mathfrak{p}_Y)_\#(\pi) = \nu$. Finally, by definition of the function L and from identities (3.8) and (3.9), we get $L(\pi) = (f^{(1)}, f^{(2)})$, hence

$$L(\Pi(\mu,\nu)) = \mathcal{F}(\mu,\nu),$$

which allows us to conclude the proof. □

The functional $L : \Pi(\mu,\nu) \to \mathcal{F}(\mu,\nu)$ defined in (3.4), allows us to relate the functionals $\mathbb{T}_c$ and $\mathbb{CT}_c$, as it follows

$$\mathbb{T}_c(\pi) = \mathbb{CT}_c(L(\pi)) \qquad \forall \pi \in \Pi(\mu,\nu),$$

which, in conjunction with the identity $L(\Pi(\mu,\nu)) = \mathcal{F}(\mu,\nu)$, allows us to conclude that the infimum of $\mathbb{CT}_c$ is actually a minimum and the set of minimizers of $\mathbb{CT}_c$ coincides with the image of $\Gamma_o(\mu,\nu)$ through L. In particular, we have that the cardinal flow problem inherits the uniqueness of the solution from the optimal transportation problem.

Corollary 3.1. *If the optimal transportation plan is unique, then the optimal cardinal flow is also unique.*

Since the operator L is only surjective and not also injective, the reverse implication is not true, i.e., given an optimal cardinal flow, we can recover many optimal transportation plans. We can still use optimal cardinal flows to retrieve optimal transportation plans with an explicit formula.

Corollary 3.2. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let c be a separable cost function on $X \times Y$. If $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$ is an optimal couple of cardinal flows, then any $\pi \in \mathcal{P}(X \times Y)$ such that*

$$f^{(1)} = (\mathfrak{p}_{X\times Y_1})_{\#}(\pi)$$

and

$$f^{(2)} = (\mathfrak{p}_{X_2\times Y})_{\#}(\pi)$$

is an optimal transportation plan between μ and ν. In particular, the plan π defined as

$$\begin{aligned} \pi(A\times B) &= \frac{f^{(1)}((\mathfrak{p}_{X\times Y_1})(A\times B))f^{(2)}((\mathfrak{p}_{X_2\times Y})(A\times B))}{(\mathfrak{p}_{Y_1\times X_2})_{\#}f^{(1)}(A\times B)} \\ &= \frac{f^{(1)}((\mathfrak{p}_{X\times Y_1})(A\times B))f^{(2)}((\mathfrak{p}_{X_2\times Y})(A\times B))}{(\mathfrak{p}_{Y_1\times X_2})_{\#}f^{(2)}(A\times B)} \end{aligned} \tag{3.10}$$

for all $A \times B$ such that

$$(\mathfrak{p}_{Y_1\times X_2})_{\#}f^{(1)}(A\times B) = (\mathfrak{p}_{Y_1\times X_2})_{\#}f^{(2)}(A\times B) \neq 0$$

is an optimal transportation plan.

Proof. Let us take $A = A_1 \times A_2 \subset X$ and $B_1 \subset Y_1$. By the definition of pushforward and formula (3.10) we have

$$\begin{aligned} & (\mathfrak{p}_{X\times Y_1})_{\#}\pi(A\times B_1) \\ = \;& \pi(A\times(B_1\times Y_2)) \\ = \;& \frac{f^{(1)}((\mathfrak{p}_{X\times Y_1})(A\times(B_1\times Y_2)))f^{(2)}((\mathfrak{p}_{X_2\times Y})(A\times(B_1\times Y_2))}{(\mathfrak{p}_{Y_1\times X_2})_{\#}f^{(2)}((\mathfrak{p}_{X_2\times Y_1})(A\times(B_1\times Y_2)))} \\ = \;& \frac{f^{(1)}((\mathfrak{p}_{X\times Y_1})(A\times B_1))f^{(2)}(A_2\times(B_1\times Y_2))}{(\mathfrak{p}_{Y_1\times X_2})_{\#}f^{(2)}(A_2\times B_1)}. \end{aligned} \tag{3.11}$$

Using again the definition of pushforward, we get

$$\zeta(A_2\times B_1) = f^{(2)}(A_2\times(B_1\times Y_2))$$

for all $A_2 \subset X_2$, $B_1 \subset Y_1$, which allows us to simplify the relation (3.11) and finally find

$$(\mathfrak{p}_{X\times Y_1})_\# \pi(A \times B_1) = f^{(1)}((\mathfrak{p}_{X\times Y_1})(A \times B_1))$$

for every $A \subset X$ and $B_1 \subset Y_1$, such that $(\mathfrak{p}_{Y_1\times X_2})_\# f^{(2)}(A \times (B_1 \times Y_2)) \neq 0$, hence $(\mathfrak{p}_{X\times Y_1})_\# \pi = f^{(1)}$. Similarly it is possible to show that $(\mathfrak{p}_{X_2\times Y})_\# \pi = f^{(2)}$, hence $L(\pi) = (f^{(1)}, f^{(2)})$. □

Given $\pi \in \Pi(\mu, \nu)$, since the marginals of π are unique, we have only one cardinal flow induced by π. However, given a cardinal flow, there might be more than one transportation plan that induces it. The latter corollary tells us how to retrieve one. From this property, it is already possible to infer an important characteristic of the cardinal flows: their description of the transportation of μ into ν is more concise than the one made through the transportation plans. This feature will allow us to define a new and faster way to compute the optimal transportation cost between measures.

3.1.1 Pivot Measures

Given a cardinal flow $(f^{(1)}, f^{(2)})$, by definition we have that $f^{(1)}$ and $f^{(2)}$ glue on a common marginal. Studying the measures on which optimal cardinal flows glue is of key importance for understanding the structure of the optimal transportation plan.
As we will see, by assuming more structure on the cost function c and on the spaces X and Y, we can move our attention from the cardinal flows to the common marginal.

Definition 3.4 (Intermedium Measures)**.** *Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We define the set of intermedium measures between μ and ν as*

$$\mathcal{I}(\mu,\nu) := \left\{ \lambda \in \mathcal{P}(Y_1 \times X_2) \ \textit{s.t.}\ (\mathfrak{p}_{X_2})_\#(\lambda) = \mu_2 \ \textit{and}\ (\mathfrak{p}_{Y_1})_\#(\lambda) = \nu_1 \right\}.$$

Remark 3.5. *Notice that, given a cardinal flow $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$, the measure*

$$\zeta = (\mathfrak{p}_{Y_1\times X_2})_\# f^{(1)} = (\mathfrak{p}_{Y_1\times X_2})_\# f^{(2)}$$

belongs to $\mathcal{I}(\mu,\nu)$. In fact from a direct computation we find

$$(\mathfrak{p}_{X_2})_\# \zeta = (\mathfrak{p}_{X_2})_\# (\mathfrak{p}_{Y_1\times X_2})_\# f^{(1)} = (\mathfrak{p}_{X_2})_\# \mu = \mu_2.$$

and, through similar argument, $(\mathfrak{p}_{Y_1})_\# \zeta = \nu_1$.

Lemma 3.1. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $\lambda \in \mathcal{I}(\mu,\nu)$. Then, there exists $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$ such that*

$$(\mathfrak{p}_{X_2 \times Y_1})_\# f^{(1)} = \lambda = (\mathfrak{p}_{X_2 \times Y_1})_\# f^{(2)}.$$

Proof. Let $\lambda \in \mathcal{I}(\mu,\nu)$. By the disintegration problem we can write

$$\lambda = \lambda_{|x_2} \otimes \lambda_2 = \lambda_{|x_2} \otimes \mu_2$$

and

$$\mu = \mu_{|x_2} \otimes \mu_2.$$

We can then define $f^{(1)} \in \mathcal{P}(X \times Y_1)$ as

$$f^{(1)} = \left(\lambda_{|x_2} \otimes \mu_{|x_2}\right) \otimes \mu_2.$$

It is easy to see that $(\mathfrak{p}_X)_\# f^{(1)} = \mu$ and $(\mathfrak{p}_{X \times Y_1})_\# f^{(1)} = \lambda$. Similarly, we can define $f^{(2)}$ as

$$f^{(2)} = \left(\lambda_{|y_1} \otimes \nu_{|y_1}\right) \otimes \nu_1,$$

so that

$$\nu = (\mathfrak{p}_Y)_\# f^{(2)}$$

and

$$\lambda = (\mathfrak{p}_{Y_1 \times X_2})_\# f^{(2)}$$

hence $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$. □

Definition 3.5. *Given $\lambda \in \mathcal{I}(\mu,\nu)$, we say that the cardinal flow $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$ glues on λ if*

$$(\mathfrak{p}_{Y_1 \times X_2})_\# f^{(1)} = (\mathfrak{p}_{Y_1 \times X_2})_\# f^{(2)} = \lambda.$$

For any $\lambda \in \mathcal{I}(\mu,\nu)$ we can then recover a cardinal flow that glues on it. Although this cardinal flow is not unique in general, we can think of λ as of a hybrid configuration that stands in the middle of the measures μ and ν, an intermediate step of a transformation that will change μ into ν.
By introducing a cost function on $X \times Y$, we are able to discriminate which cardinal flows are better than others and, by consequence, which intermediate configurations are better.

Definition 3.6 (Pivot Measure)**.** *Let us take $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and c a separable cost function. We say that $\zeta \in \mathcal{P}(X_2 \times Y_1)$ is a* pivot measure *between μ and ν if there exists an optimal cardinal flow $(f^{(1)}, f^{(2)})$ that glues on it.*

Remark 3.6. *From Remark 3.5 it trivially follows that all the pivot measures are also intermediate measures.*

Remarkably, if the supports of the measures μ and ν are "perpendicular", the uniqueness can be obtained without any other request on the cost function rather then the separability.

Lemma 3.2. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. If there exist $\bar{x}_2 \in X_2$ for which*

$$\mu_2(\{\bar{x}_2\}) = \mu(X_1 \times \{\bar{x}_2\}) = 1$$

then the pivot measure ζ has to be $\delta_{\bar{x}_2} \otimes \nu_1$. Similarly, if there exists $\bar{y}_1$ such that

$$\nu_1(\{\bar{y}_1\}) = \nu(\{\bar{y}_1\} \times Y_2) = 1,$$

then the pivot measure $\zeta = \delta_{\bar{y}_1} \otimes \mu_2$. Moreover, if both $\bar{x}_2$ and $\bar{y}_1$ exists the Dirac's delta centered in $(\bar{y}_1, \bar{x}_2)$ is the only intermediate measure and the only cardinal flow (and, therefore, the optimal one) is

$$f^{(1)} = \mu \otimes \delta_{\bar{y}_1}$$

and

$$f^{(2)} = \delta_{\bar{x}_2} \otimes \nu.$$

In this case each $\pi \in \Pi(\mu, \nu)$ is also optimal.

Proof. The first two points follows follow from the fact that $\mathcal{I}(\mu, \nu)$ contains only one measure, which is $\nu_1 \otimes \delta_{\bar{x}_2}$, $\delta_{\bar{y}_1} \otimes \mu_2$ and $\delta_{(\bar{y}_1, \bar{x}_2)}$ respectively. In the latter case, also $\mathcal{F}(\mu, \nu)$ contains only one element, therefore $L(\pi) = (f^{(1)}, f^{(2)})$, which means that each $\pi \in \Pi(\mu, \nu)$ is optimal. $\square$

3.2 The reverse problem

Up to now we always supposed that the cost functions were only semi-separable. From now on, we will focus on the separable functions.

There are two main reasons for this choice. The first one is that dealing with totally separable cost functions is simpler and allows us to give more information on the structure of the optimal transportation plan itself. The second one is that, when it comes to applications, the greater part of the cost functions are separable. For instance, each l^p norm can be used to induce cost function c_p on euclidean spaces

$$c_p(\mathbf{x}, \mathbf{y}) := ||\mathbf{x} - \mathbf{y}||_p^p = \sum_{i=1}^{N} |x_i - y_i|^p$$

which is indeed separable.

When the cost function is (totally) separable we can define a "reverse" version of the cardinal flow. Due to the semi-separable structure of the cost function, the flows had to move the mass first from X_1 to Y_1 and, afterward, from X_2 to Y_2. Once this structure ceases to exist we are allowed to define a problem in which this order is reversed.

In this case, the set of cardinal flows will be then defined as

$$\mathcal{G}(\mu, \nu) \subset \mathcal{P}(X \times Y_2) \times \mathcal{P}(X_1 \times Y)$$

containing all the couples namely $(g^{(1)}, g^{(2)})$ satisfying the following conditions

$$\mu = (\mathfrak{p}_X)_\# g^{(1)}, \qquad \nu = (\mathfrak{p}_Y)_\# g^{(2)}$$

and

$$(\mathfrak{p}_{X_1 \times Y_2})_\# g^{(1)} = (\mathfrak{p}_{X_1 \times Y_2})_\# g^{(2)}.$$

The flow functional associated with this reformulation is

$$(g^{(1)}, g^{(2)}) \to \int_{X \times Y_2} c_1 dg^{(2)} + \int_{X_1 \times Y} c_2 dg^{(1)},$$

and the pivot measures will still be the one on which the two flows glue.

Remark 3.7. *All the results of Section 3.1 can be recovered in this setting by applying very minimal changes to make them fit this framework.*

In the following two propositions, we aim to highlight some of the relations that bound those two formulations. The first one tells us that $\mathcal{F}(\mu, \nu)$ contains all the reverse transformations of $\mathcal{G}(\nu, \mu)$. The second one will show us that we can associate with each element of $\mathcal{F}(\mu, \nu)$ an element of $\mathcal{G}(\mu, \nu)$ that has the same cost along both the directions.

Proposition 3.1. *For each $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, it holds true*

$$\mathcal{F}(\mu, \nu) = \mathcal{G}(\nu, \mu).$$

Proof. We show that each $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$ uniquely defines an element of $\mathcal{G}(\nu, \mu)$. By swapping the role of $\mathcal{F}(\mu, \nu)$ and $\mathcal{G}(\nu, \mu)$ and repeating the following argument we can prove also the reverse implication.
Let us then take $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$. We define

$$g^{(1)} = f^{(2)}$$

and

$$g^{(2)} = f^{(1)}$$

then, by definition

$$\begin{aligned}(\mathfrak{p}_Y)_\# g^{(1)} &= (\mathfrak{p}_Y)_\# f^{(2)} = \nu,\\ (\mathfrak{p}_X)_\# g^{(2)} &= (\mathfrak{p}_X)_\# f^{(1)} = \mu\end{aligned}$$

and

$$(\mathfrak{p}_{Y_1\times X_2})_\# g^{(1)} = (\mathfrak{p}_{Y_1\times X_2})_\# f^{(2)} = (\mathfrak{p}_{Y_1\times X_2})_\# f^{(1)} = (\mathfrak{p}_{Y_1\times X_2})_\# g^{(2)}$$

hence $(g^{(1)}, g^{(2)}) \in \mathcal{G}(\nu, \mu)$. □

Proposition 3.2. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and c a separable cost function. For any $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$ there exists a $(g^{(1)}, g^{(2)}) \in \mathcal{G}(\mu, \nu)$ such that*

$$\int_{X\times Y_1} c_1 df^{(1)} = \int_{X_1\times Y} c_1 dg^{(2)}$$

and

$$\int_{X_2\times Y} c_2 df^{(2)} = \int_{X\times Y_2} c_2 dg^{(1)}.$$

Proof. From Corollary 3.2 we know that, for a given $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$, there exists a $\pi \in \Pi(\mu, \nu)$ such that

$$L(\pi) = (f^{(1)}, f^{(2)}).$$

We can then define $g^{(1)}$ and $g^{(2)}$ as

$$g^{(1)} = (\mathfrak{p}_{X\times Y_2})_\# \pi$$

and

$$g^{(2)} = (\mathfrak{p}_{X_1\times Y})_\# \pi.$$

It is easy to check that $(g^{(1)}, g^{(2)}) \in \mathcal{G}(\mu, \nu)$. The thesis follows from considering that, for totally separable cost functions, holds true that

$$c_1 \circ (\mathfrak{p}_{X\times Y_1}) = c_1 \circ (\mathfrak{p}_{X_1\times Y})$$

hence

$$\int_{X\times Y_1} c_1 df^{(1)} = \int_{X\times Y_1} c_1 d(\mathfrak{p}_{X\times Y_1})_\# \pi$$

$$\begin{aligned}
&= \int_{X\times Y} c_1 \circ (\mathfrak{p}_{X\times Y_1}) d\pi \\
&= \int_{X\times Y} c_1 \circ (\mathfrak{p}_{X_1\times Y}) d\pi \qquad (3.12) \\
&= \int_{X_1\times Y} c_1 d(\mathfrak{p}_{X_1\times Y})_{\#}\pi \\
&= \int_{X_1\times Y} c_1 dg^{(2)}.
\end{aligned}$$

Similarly, we can prove the other equality. □

Given two probability measures μ and ν, we can define two different pivot measures: the one obtained considering first the flows that move the mass along the first direction and then the flow that moves the mass along the second one. We denote with ζ the first one and η the other one.

Definition 3.7. *Let μ and ν be two probability measures and π be an optimal transportation plan between them according to the separable cost function c. We say that the pair (ζ, η), defined as*

$$\zeta := (\mathfrak{p}_{Y_1\times X_2})_{\#}\pi$$

and

$$\eta := (\mathfrak{p}_{X_1\times Y_2})_{\#}\pi,$$

is a couple of pivot measure for μ and ν.

When $X = Y$, we can think of the measures μ, ν, ζ, and η as the vertex of a rectangle in the space $\mathcal{P}(X)$. For example, let us take $\mu = \delta_{\mathbf{x}}$ and $\nu = \delta_{\mathbf{y}}$ where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, then the two pivot measures ζ and η are also Dirac's deltas centered in the (y_1, x_2) and (x_1, y_2) respectively. Those points, along with $\mathbf{x}$ and $\mathbf{y}$ identify a unique rectangle in the space.
For more complex measures can be tougher understanding how the rectangle is made and, in general, is not even uniquely defined, i.e. there might be more than one couple of pivot measures.

Theorem 3.2. *Given two measures $\mu, \nu \in \mathcal{P}(X)$, we have that*

$$W_c(\mu, \nu) = W_c(\zeta, \eta)$$

where (ζ, η) is a couple of pivot measures. In particular, we have that μ and ν are a couple of pivot measures for each of their own couple of pivot measures.

Proof. Let us denote π the optimal transportation plan between μ and ν for which holds true the identity

$$\zeta := (\mathfrak{p}_{X_2 \times Y_1})_{\#}\pi$$

and

$$\eta := (\mathfrak{p}_{X_1 \times Y_2})_{\#}\pi,$$

this means that π is also a transportation plan between ζ and η. According to Theorem 2.11, π must be $c-$ciclically monotone.
Since c_1 and c_2 are both symmetric, it holds true that

$$\begin{aligned} c((y_1, x_2), (x_1, y_2)) &= c_1(x_1, y_1) + c_2(y_2, x_2) \\ &= c_1(x_1, y_1) + c_2(x_2, y_2) \\ &= c(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Which means that π is ciclically monotone also respect the ground distance

$$\hat{c} : (Y_1 \times X_2) \times (X_1 \times Y_2) \to \mathbb{R},$$

defined as

$$\hat{c}((y_1, x_2), (x_1, y_2)) := c_1(x_1, y_1) + c_2(y_2, x_2)$$

From Theorem 3.2, we can conclude that π is optimal, hence the thesis. The last point follows from the fact that π is a transportation plan between μ and ν. $\square$

3.2.1 Separable Distances

As we saw in Section 2.2.1, whenever $X = Y$ and the cost function c is a distance d over X, the minimal transportation problem induces a distance over the space $\mathcal{P}(X)$, the Wasserstein-Kantorovich distance W_d. Roughly speaking, the distance structure of d is lifted from the space X to space $\mathcal{P}(X)$.
In this section, we will show that if the distance d is separable, so is W_d. Moreover, we will show how the pivot measure covers a major role in the separability of the Wasserstein-Kantorovich distance.

Remark 3.8. *Since in the case $X = Y$, we need to slightly change the notations in order to avoid confusion. We denote the generic point of $X \times X$ with*

$$(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \big((x_1^{(1)}, x_2^{(1)}), (x_1^{(2)}, x_2^{(2)})\big)$$

so that, in general, with $\boldsymbol{x}^{(i)}$ *we denote the* $i-th$ *component in the space* $X \times X$ *and, with* $x_j^{(i)}$, *we denote the* $j-th$ *component of* $\boldsymbol{x}^{(i)}$. *With this notations we can then define, for* $i = 1, 2$, *the projections* $(\mathfrak{p})^{(i)} : X \times X \to X$ *defined as*

$$(\mathfrak{p})^{(i)}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := \boldsymbol{x}^{(i)} \tag{3.13}$$

and

$$p_j^{(i)}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := x_j^{(i)} \tag{3.14}$$

for $i = 1, 2$ *and* $j = 1, 2$, *so that*

$$(\mathfrak{p})^{(i)} = (p_1^{(i)}, p_2^{(i)}).$$

Theorem 3.3. *Let us take* $\mu, \nu \in \mathcal{P}(X)$ *(where* $X = X_1 \times X_2$*) and let* d *be a separable distance over* X, *i.e.*

$$d(\boldsymbol{x}, \boldsymbol{y}) = d_1(x_1, y_1) + d_2(x_2, y_2) \qquad\qquad \forall \boldsymbol{x}, \boldsymbol{y} \in X.$$

Then, for any pivot measure $\zeta \in \mathcal{I}(\mu, \nu)$, *we have*

$$W_d(\mu, \nu) = W_d(\mu, \zeta) + W_d(\zeta, \nu). \tag{3.15}$$

Proof. From Theorem 2.17 we know that W_d is a distance over $\mathcal{P}(X)$, then, by the triangular inequality, we get that

$$W_d(\mu, \nu) \leq W_d(\mu, \zeta) + W_d(\zeta, \nu), \tag{3.16}$$

for any $\zeta \in \mathcal{P}(X)$, and, in particular, for each $\zeta \in \mathcal{I}(\mu, \nu)$.
To prove the other inequality, let us take $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$ two optimal cardinal flows. We can then define

$$\pi^{(1)} = \big((\mathfrak{p})^{(2)}, (p_1^{(2)}, p_2^{(1)})\big)_{\#} f^{(1)}$$

and

$$\pi^{(2)} = \big((p_1^{(2)}, p_2^{(1)}), (\mathfrak{p})^{(2)}\big)_{\#} f^{(2)}.$$

Remembering the functions defined above, it holds true that

$$(\mathfrak{p})^{(1)} \circ \big((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)})\big) = (p_1, p_2) \tag{3.17}$$

and

$$(\mathfrak{p})^{(2)} \circ \big((p_1^{(2)}, p_2^{(1)}), (\mathfrak{p})^{(2)}\big) = (p_3, p_2). \tag{3.18}$$

Then, through the relations (3.17), (3.18) and the chain rule for the push-forwards (Lemma 2.1), we get that

$$
\begin{aligned}
(\mathfrak{p})^{(1)}_{\#}\pi^{(1)} &= (p_1^{(1)}, p_2^{(1)})_{\#}\Big(\big((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)})\big)_{\#} f^{(1)}\Big) \\
&= (p_1^{(1)}, p_2^{(1)})_{\#} f^{(1)} \\
&= \mu
\end{aligned}
$$

and

$$
\begin{aligned}
(\mathfrak{p})^{(2)}_{\#}\pi^{(1)} &= (p_1^{(2)}, p_2^{(1)})_{\#}\Big(\big((p_1^{(2)}, p_2^{(1)}), (\mathfrak{p})^{(2)}\big)_{\#} f^{(1)}\Big) \\
&= (p_1^{(2)}, p_2^{(1)})_{\#} f^{(1)} \\
&= \zeta,
\end{aligned}
$$

hence $\pi^{(1)} \in \Pi(\mu, \zeta)$. Similarly, we can show that $\pi^{(2)} \in \Pi(\zeta, \nu)$.
Finally, by Theorem 3.1 we have that

$$
\begin{aligned}
W_d(\mu,\nu) &= \int_{X\times X_1} d_1(x_1,y_1)df^{(1)}(\mathbf{x},y_1) + \int_{X_2\times X} d_2(x_2,y_2)df^{(2)}(x_2,\mathbf{y}) \\
&= \int_{X\times X_1} d((x_1,x_2),(y_1,x_2))df^{(1)}(\mathbf{x},y_1) \\
&\quad + \int_{X_2\times X} d((y_1,x_2),(y_1,y_2))df^{(2)}(x_2,\mathbf{y}) \\
&= \int_{X\times X} d(\mathbf{x},\mathbf{y})d\pi^{(1)}(\mathbf{x},\mathbf{y}) + \int_{X\times X} d(\mathbf{x},\mathbf{y})d\pi^{(2)}(\mathbf{x},\mathbf{y}) \\
&\geq W_d(\mu,\zeta) + W_d(\zeta,\nu).
\end{aligned}
$$

The latter inequality, in conjunction with (3.16), allows us to conclude

$$W_d(\mu,\nu) = W_d(\mu,\zeta) + W_d(\zeta,\nu)$$

for any pivot measure ζ. □

Remark 3.9. *The previous Theorem states that any pivot measure minimizes the functional*

$$\Theta : \lambda \to W_d(\mu,\lambda) + W_d(\lambda,\nu).$$

Unfortunately, the reverse implication is not true, i.e. there might be measures that minimize Θ that are not pivot measure. Consider, for instance, $X = \mathbb{R}^2$, $\mu = \frac{1}{2}\big[\delta_{(0,0)} + \delta_{(7,1)}\big]$ and $\nu = \frac{1}{2}\big[\delta_{(1,1)} + \delta_{(8,0)}\big]$. It is easy to see that, in this *case, that the only pivot measure is* $\zeta = \frac{1}{2}\left[\delta_{(1,0)} + \delta_{(8,1)}\right]$. *Anyway, since* $\mu_2 = \nu_2$, *we have that* $\nu \in \mathcal{I}(\mu,\nu)$, *so that*

$$\begin{aligned} W_d(\mu,\nu) &= \inf_{\lambda\in\mathcal{I}(\mu,\nu)} W_d(\mu,\lambda) + W_d(\lambda,\nu) \\ &\leq W_d(\mu,\nu) + W_d(\nu,\nu) \\ &= W_d(\mu,\nu) \end{aligned}$$

which means that ν *minimizes* Ω, *although it is not a pivot measure.*

More in general, if the cost function is not a distance, we cannot recover this property, i.e. it is not true that, if $c = c_1 + c_2$ then

$$W_c(\mu,\nu) = \inf_{\sigma\in\mathcal{I}(\mu,\nu)} W_c(\mu,\sigma) + W_c(\sigma,\nu)$$

as we will show in the following example.

Example 3.1. *Let us take* $X = Y = \mathbb{R}^2$ *and* $\mu,\nu \in \mathcal{P}(\mathbb{R}^2)$ *defined as*

$$\mu = \frac{1}{2}\delta_{(1,1)} + \frac{1}{2}\delta_{(0,2)}$$

and

$$\nu = \frac{1}{2}\delta_{(3,0)} + \frac{1}{2}\delta_{(5,3)}$$

where $\delta_{(a,b)}$ *is the Dirac delta centered in* (a,b).
As cost function on $\mathbb{R}^2$, *we will take the square Euclidean one, i.e.*

$$c(\boldsymbol{x},\boldsymbol{y}) = (x_1 - y_1)^2 + (x_2 - y_2)^2$$

which is indeed separable.
From a direct argument, it is possible to see that the optimal transportation plan is

$$\pi = \frac{1}{2}\delta_{((1,1);(3,1))} + \frac{1}{2}\delta_{((0,2);(5,2))},$$

hence, $W_2^2(\mu,\nu) = \frac{31}{2}$.
By applying L *to* π *we find the optimal cardinal flow*

$$f^{(1)} = \frac{1}{2}\delta_{((1,1);3)} + \frac{1}{2}\delta_{((0,2);5)}$$

and

$$f^{(2)} = \frac{1}{2}\delta_{(2;(5,3))} + \frac{1}{2}\delta_{(1;(3,0))}$$

hence the related pivot measure is

$$\zeta = \frac{1}{2}\delta_{(3,1)} + \frac{1}{2}\delta_{(5,2)}.$$

Let us now compute $W_2^2(\mu,\zeta) + W_2^2(\zeta,\nu)$, we will show that it is less than $W_2^2(\mu,\nu)$.
Again from a direct method, we will find that the optimal plan $\tilde{\pi}_1$ between μ and ζ is

$$\tilde{\pi}_1 = \frac{1}{2}\delta_{((0,2);(3,1))} + \frac{1}{2}\delta_{((1,1);(5,2))}$$

while the optimal one between ζ and ν is

$$\tilde{\pi}_2 = \frac{1}{2}\delta_{((3,1);(3,2))} + \frac{1}{2}\delta_{((5,2);(5,3))}$$

hence

$$W_2^2(\mu,\zeta) = \frac{10}{2} + \frac{17}{2} = \frac{27}{2},$$

and

$$W_2^2(\zeta,\nu) = \frac{1}{2} + \frac{1}{2} = 1,$$

so that $W_2^2(\mu,\zeta) + W_2^2(\zeta,\nu) = \dfrac{29}{2}$, hence

$$W_2^2(\mu,\nu) > W_2^2(\mu,\zeta) + W_2^2(\zeta,\nu).$$

More in general, for the squared Euclidean cost function holds true

$$W_2^2(\mu,\nu) \geq \inf_{\zeta\in\mathcal{I}(\mu,\nu)} \Big\{ W_2^2(\mu,\zeta) + W_2^2(\zeta,\nu) \Big\}.$$

3.3 Singular Costs

As we pointed out in the previous section, the separability is lifted from X to $\mathcal{P}(X)$ whenever the cost function is a distance over X. For generic cost functions, we can still recover similar results by introducing the singular cost functions. Roughly speaking, those cost functions force the mass to travel horizontally and vertically. Through the use of those functions, we will also be able to find a more precise characterization of the pivot measures.

Definition 3.8 (Singular Costs)**.** *Let c be a cost function on $X \times X$. The singular cost functions inducted by c are defined as it follows*

$$c_1^{(\infty)}(\boldsymbol{x}, \boldsymbol{y}) := \begin{cases} c(\boldsymbol{x}, \boldsymbol{y}) & \text{if } x_2 = y_2 \\ +\infty & \text{otherwise} \end{cases}$$

and

$$c_2^{(\infty)}(\boldsymbol{x}, \boldsymbol{y}) := \begin{cases} c(\boldsymbol{x}, \boldsymbol{y}) & \text{if } x_1 = y_1 \\ +\infty & \text{otherwise} \end{cases}.$$

Remark 3.10. *As long as the sets $\{x_1^{(1)} = x_1^{(2)}\}$ and $\{x_2^{(1)} = x_2^{(2)}\}$ are weakly close in $X \times X$ (which is the case when $X = \mathbb{R}^2$), the functions $c_1^{(\infty)}$ and $c_2^{(\infty)}$ are both l.s.c. if so is c. Moreover, since*

$$c \leq c_1^{(\infty)} \qquad \text{and} \qquad c \leq c_2^{(\infty)},$$

if c satisfies the conditions of Theorem 2.9, so does $c_1^{(\infty)}$ and $c_2^{(\infty)}$.

Definition 3.9. *The singular transport cost problem associated to c is defined as*

$$\mathcal{C}^{(\infty)}(\mu, \nu) := \inf_{\sigma \in \mathcal{I}(\mu,\nu)} \mathcal{C}_1^{(\infty)}(\mu, \sigma) + \mathcal{C}_2^{(\infty)}(\sigma, \nu), \tag{3.19}$$

where $\mathcal{C}_1^{(\infty)}$ and $\mathcal{C}_2^{(\infty)}$ are defined as it follows

$$\mathcal{C}_i^{(\infty)}(\alpha, \beta) := \inf_{\pi \in \Pi(\alpha,\beta)} \int_{X \times X} c_i^{(\infty)}(\boldsymbol{x}, \boldsymbol{y}) d\pi.$$

In what follows, we show that the functional $\mathcal{C}^{(\infty)}(\mu, \nu)$ is equal to $\mathcal{C}(\mu, \nu)$ whenever the cost function is separable. Furthermore, all the $\sigma \in \mathcal{I}(\mu, \nu)$ that minimize (3.19) are actually pivot measures. To do so, we need to show two following technical lemma, which assures us the well-posedness of $\mathcal{C}^{(\infty)}$. Notice that Remark 3.10 assures us that both $\mathcal{C}_1^{(\infty)}$ and $\mathcal{C}_2^{(\infty)}$ are both well defined.

Lemma 3.3. *The functional $\mathcal{C}^{(\infty)}$ is proper,* i.e.

$$\mathcal{C}^{(\infty)}(\mu, \nu) < +\infty$$

for each couple of $\mu, \nu \in \mathcal{P}(X)$.

Proof. Given μ and ν, let us consider

$$\tilde{\sigma} = \nu_1 \otimes \mu_2$$

where ν_1 and μ_2 are the first marginal of ν and the second marginal of μ respectively. By definition $\tilde{\sigma} \in \mathcal{I}(\mu, \nu)$, hence

$$\mathcal{C}^{(\infty)}(\mu, \nu) \leq \mathcal{C}_1^{(\infty)}(\mu, \tilde{\sigma}) + \mathcal{C}_2^{(\infty)}(\tilde{\sigma}, \nu).$$

Let us now consider the measure $\mu \otimes \nu_1$, is then well defined the measure $\gamma^{(1)} \in \mathcal{P}(X \times X)$ as

$$\gamma^{(1)} := ((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)}))_{\#}(\mu \otimes \nu_1).$$

Using the chain rule (2.1), we have

$$(p_1^{(1)}, p_2^{(1)})_{\#}\gamma^{(1)} = (p_1^{(1)}, p_2^{(1)})_{\#}(\mu \otimes \nu_1) = \mu$$

and

$$(p_1^{(2)}, p_1^{(2)})_{\#}\gamma^{(1)} = (p_1^{(2)}, p_2^{(1)})_{\#}(\mu \otimes \nu_1) = \nu_1 \otimes \mu_2,$$

so that $\gamma^{(1)} \in \Pi(\mu, \tilde{\sigma})$.
Hence we have

$$\begin{aligned}
\mathcal{C}_1^{(\infty)}(\mu, \tilde{\sigma}) &\leq \int_{X \times X} c_1^{(\infty)}(\mathbf{x}, \mathbf{y}) d\gamma^{(1)}(\mathbf{x}, \mathbf{y}) \\
&= \int_{X \times X} c_1^{(\infty)}(\mathbf{x}, \mathbf{y}) d((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)}))_{\#}(\mu \otimes \nu_1)(\mathbf{x}, y_1) \\
&= \int_{X \times X_1} c_1^{(\infty)}((x_1, x_2), (y_1, x_2)) d(\mu \otimes \nu_1)(\mathbf{x}, y_1) \\
&= \int_{X \times X_1} c(\mathbf{x}, \mathbf{y}) d(\mu \otimes \nu_1)(\mathbf{x}, y_1) \\
&< +\infty
\end{aligned}$$

which is finite thanks to the conditions on c.
Similarly, we can show that $\mathcal{C}_2^{(\infty)}(\mu, \nu) < +\infty$ and, by summing up the two estimations, conclude the proof. □

Lemma 3.4. *Given $\mu, \nu \in \mathcal{P}(X)$, the functional*

$$\theta \to \mathcal{C}_1^{(\infty)}(\mu, \theta) + \mathcal{C}_2^{(\infty)}(\theta, \nu) \tag{3.20}$$

is l.s.c. (with respect the weak topology) and convex over its domain.
In particular, the infimum defined in (3.19) *is actually a minimum and, given any couple $\mu, \nu \in \mathcal{P}(X)$, there exists $\zeta \in \mathcal{I}(\mu, \nu)$ such that*

$$\mathcal{C}^{(\infty)}(\mu, \nu) = \mathcal{C}_1^{(\infty)}(\mu, \zeta) + \mathcal{C}_2^{(\infty)}(\zeta, \nu). \tag{3.21}$$

Definition 3.10. *We denote with $\Sigma_o(\mu,\nu)$ the non empty set of optimizer for the problem defined in* (3.19)*, i.e. the set of $\zeta \in \mathcal{I}(\mu,\nu)$ for which holds true formula* (3.21)*.*

Up to now we never asked the separability on the cost function c. This further assumption is indeed unnecessary to define the minimum problem in (3.19), but it is needed to obtain the equivalence with the optimal transportation problem. In the following, we will show this equivalence and also that the set $\Sigma_o(\mu,\nu)$ coincides with the set of pivot measures between μ and ν.

Theorem 3.4. *Given $\mu,\nu \in \mathcal{P}(X)$, it holds true that*

$$W_c(\mu,\nu) = \mathcal{C}^{(\infty)}(\mu,\nu), \tag{3.22}$$

whenever the cost function c is separable. In particular, each $\theta \in \mathcal{I}(\mu,\nu)$ for which holds true

$$W_c(\mu,\nu) = \mathcal{C}_1^{(\infty)}(\mu,\theta) + \mathcal{C}_2^{(\infty)}(\theta,\nu)$$

is a pivot measure, and vice versa.

Proof. First of all, we notice that, since it is defined as the infimum of the sum of two l.s.c. functional, $\mathcal{C}^{(\infty)}$ is l.s.c. as well. We now prove that $\mathcal{C}^{(\infty)}$ is convex too.
Let us take $\mu_i,\nu_i \in \mathcal{P}(X)$ and $\alpha_i \in [0,1]$, where $i = 1,2$, such that

$$\alpha_1 + \alpha_2 = 1.$$

From Lemma 3.4, we can find two measures $\theta_i \in \Sigma_o(\mu_i,\nu_i)$ for which holds true that

$$\mathcal{C}^{(\infty)}(\mu_i,\nu_i) = \mathcal{C}_1^{(\infty)}(\mu_i,\theta_i) + \mathcal{C}_2^{(\infty)}(\theta_i,\nu_i)$$

for $i = 1,2$.
We can then define the measures

$$\bar{\mu} = \alpha_1\mu_1 + \alpha_2\mu_2,$$

$$\bar{\nu} = \alpha_1\nu_1 + \alpha_2\nu_2$$

and

$$\bar{\theta} = \alpha_1\theta_1 + \alpha_2\theta_2.$$

Since $\theta_i \in \mathcal{I}(\mu_i,\nu_i)$, we have that

$$\bar{\theta} \in \mathcal{I}(\bar{\mu},\bar{\nu})$$

and hence

$$\begin{aligned}
\mathcal{C}^{(\infty)}(\bar{\mu},\bar{\nu}) &\leq \mathcal{C}_1^{(\infty)}(\bar{\mu},\bar{\theta}) + \mathcal{C}_2^{(\infty)}(\bar{\theta},\bar{\nu}) \\
&\leq \alpha_1\big(\mathcal{C}_1^{(\infty)}(\mu_1,\theta_1) + \mathcal{C}_2^{(\infty)}(\theta_1,\nu_1)\big) \\
&\quad +\alpha_2\big(\mathcal{C}_1^{(\infty)}(\mu_2,\theta_2) + \mathcal{C}_2^{(\infty)}(\theta_2,\nu_2)\big) \\
&= \alpha_1\mathcal{C}^{(\infty)}(\mu_1,\nu_1) + \alpha_2\mathcal{C}^{(\infty)}(\mu_2,\nu_2),
\end{aligned}$$

hence $\mathcal{C}^{(\infty)}$ is convex.
Let us now take two points $\mathbf{x},\mathbf{y} \in X$ and $\delta_{\mathbf{x}},\delta_{\mathbf{y}}$ the relative Dirac's delta associated with them, it holds true that

$$\begin{aligned}
\mathcal{C}^{(\infty)}(\delta_{\mathbf{x}},\delta_{\mathbf{y}}) &\leq \mathcal{C}_1^{(\infty)}(\delta_{\mathbf{x}},\delta_{(y_1,x_2)}) + \mathcal{C}_2^{(\infty)}(\delta_{(y_1,x_2)},\delta_{\mathbf{y}}) \\
&= c_1(x_1,y_1) + c_2(x_2,y_2) \\
&= W_c(\delta_{\mathbf{x}},\delta_{\mathbf{y}}).
\end{aligned}$$

Given then any two discrete measures, we can describe them as a convex combination of Dirac's deltas

$$\mu = \sum_{i\in I}\alpha_i\delta_{\mathbf{x}_i}$$

and

$$\nu = \sum_{j\in J}\beta_j\delta_{\mathbf{y}_j}$$

where $\{\mathbf{x}_i\}_{i\in I}$ and $\{\mathbf{y}_j\}_{j\in J}$ and I,J are finite sets.
Let $\gamma = \sum_{i,j}\gamma_{i,j}\delta_{(\mathbf{x}_i,\mathbf{y}_j)}$ be an optimal transportation plan between μ and ν. Since $\mathcal{C}^{(\infty)}$ is convex, we get that

$$\begin{aligned}
\mathcal{C}^{(\infty)}(\mu,\nu) &\leq \sum_{i\in I,j\in J}\gamma_{i,j}\mathcal{C}^{(\infty)}(\delta_{\mathbf{x}_i},\delta_{\mathbf{y}_j}) \\
&= \sum_{i\in I,j\in J}\gamma_{i,j}c(\mathbf{x}_i,\mathbf{y}_j) \\
&= \mathbb{T}_c(\gamma),
\end{aligned}$$

hence, since, γ was optimal $\mathbb{T}_c(\pi) = W_c(\mu,\nu)$, hence

$$\mathcal{C}^{(\infty)}(\mu,\nu) \leq W_c(\mu,\nu)$$

for each couple of discrete measures μ and ν.

Let us now take two generic measures μ and ν. By approximation we can find two sequences of discrete measures, $\mu^{(k)}$ and $\nu^{(k)}$, which weakly converge to μ and ν, respectively, as $k \to \infty$. Since $\mathcal{C}^{(\infty)}$ is l.s.c., we can conclude that

$$\mathcal{C}^{(\infty)}(\mu, \nu) \leq \liminf_{k\to\infty} \mathcal{C}^{(\infty)}(\mu^{(k)}, \nu^{(k)}) \leq \liminf_{k\to\infty} W_c(\mu^{(k)}, \nu^{(k)}) = W_c(\mu, \nu) \tag{3.23}$$

for each couple of measures μ and ν.
To conclude, we will prove the other inequality.
Let us take $\mu, \nu \in \mathcal{P}(X)$, $(f^{(1)}, f^{(2)})$ an optimal cardinal flow and ζ the pivot measure on which the flow glues.
Let us define the function $T : X \times X_1 \to X \times X$ as

$$T(\mathbf{x}, y_1) = \Big(\mathfrak{p}^{(1)}, (p_1^{(2)}, p_2^{(1)})\Big)(\mathbf{x}, y_1).$$

For that function T holds true the following identities

$$\mathfrak{p}^{(1)} \circ T(\mathbf{x}, y_1) = (x_1, x_2) \tag{3.24}$$

and

$$\mathfrak{p}^{(2)} \circ T(\mathbf{x}, y_1) = (y_1, x_2), \tag{3.25}$$

where $\mathfrak{p}^{(1)}$ and $\mathfrak{p}^{(2)}$ are defined above. We can then define $\gamma^{(1)} \in \mathcal{P}(X \times X)$ as

$$\gamma^{(1)} = T_{\#} f^{(1)}$$

and $\gamma^{(1)} \in \Pi(\mu, \zeta)$, in fact from identity (3.24) we get

$$(\mathfrak{p})^{(1)}\gamma = (\mathfrak{p})^{(1)} T_{\#} f^{(1)} = (\mathfrak{p})^{(1)} f^{(1)} = \mu.$$

Similarly, from identity (3.25)

$$(\mathfrak{p})^{(2)}\gamma^{(1)} = (p_1^{(2)}, p_2^{(1)})_{\#} f^{(1)} = \zeta.$$

Moreover, we have that

$$\begin{aligned}
\int_{X\times X} c_1^{(\infty)}(\mathbf{x}, \mathbf{y}) d\gamma^{(1)} &= \int_{X\times X_1} c_1^{(\infty)}(T(\mathbf{x}, y_1)) df^{(1)} \\
&= \int_{X\times X_1} c_1^{(\infty)}((x_1, x_2), (y_1, x_2)) df^{(1)} \\
&= \int_{X\times X_1} c_1(x_1, y_1) df^{(1)}.
\end{aligned}$$

Similarly, we can define $\gamma^{(2)} \in \Pi(\zeta, \nu)$ from $f^{(2)}$ such that

$$\int_{X \times X} c_2^{(\infty)}(\mathbf{x}, \mathbf{y}) d\gamma^{(2)} = \int_{X_2 \times X} c_2(x_2, y_2) df^{(2)}.$$

Finally, we can conclude, since

$$\begin{aligned} \mathcal{C}^{(\infty)}(\mu, \nu) &= \inf_{\sigma \in \mathcal{I}(\mu,\nu)} \mathcal{C}_1^{(\infty)}(\mu, \sigma) + \mathcal{C}_2^{(\infty)}(\sigma, \nu) \\ &\leq \mathcal{C}_1^{(\infty)}(\mu, \zeta) + \mathcal{C}_2^{(\infty)}(\zeta, \nu) \\ &\leq \int_{X \times X} c_1^{(\infty)}(\mathbf{x}, \mathbf{y}) d\gamma^{(1)} + \int_{X \times X} c_1^{(\infty)}(\mathbf{x}, \mathbf{y}) d\gamma^{(2)} \\ &= \int_{X \times X_1} c_1(x_1, y_1) df^{(1)} + \int_{X_2 \times X} c_2(x_2, y_2) df^{(2)}, \end{aligned} \tag{3.26}$$

and, since $(f^{(1)}, f^{(2)})$ are optimal,

$$\mathcal{C}^{(\infty)}(\mu, \nu) \leq W_c(\mu, \nu).$$

Taking into account the identity (3.23) we get, from (3.26), that

$$\mathcal{C}^{(\infty)}(\mu, \nu) = \mathcal{C}_1^{(\infty)}(\mu, \zeta) + \mathcal{C}_2^{(\infty)}(\zeta, \nu)$$

where ζ is a pivot measure, proving that each minimizer of (3.20) is also a pivot measure and vice versa. □

3.3.1 Multiseparability

Up to now, we have assumed the sets X and Y to be the direct product of two smaller polish spaces. This subsection aims to show how we can extend this framework to the case when both X and Y are the product of more sets. Intuitively, to deal with those richer structures we just need to decompose the total transportation plan into more cardinal flows, one for each subspace.

Definition 3.11. *Let $X = \times_{i=1}^N X_i$ and $Y = \times_{i=1}^N Y_i$ be two polish spaces. A function $c : X \times Y \to \mathbb{R}$ is separable if*

$$c(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^N c_i(x_i, y_i)$$

where $c_i : X_i \times Y_i \to \mathbb{R}$.

Definition 3.12. *Let us take $X = \times_{i=1}^N X_i$ and $Y = \times_{i=1}^N Y_i$ as above. We define the $i-$th interpolation set as*

$$Z_i = Y_1 \times \cdots \times Y_{i-1} \times X_i \times \cdots \times X_N$$

with the convention that $Z_{N+1} = Y$ and $Z_1 = X$. The $i-$th transportation set is defined as

$$T_i = Y_1 \times \cdots \times Y_i \times X_i \times \cdots \times X_N$$

for $i = 1, \dots, N$.

In this framework, the cardinal flow between two measures is defined as follows.

Definition 3.13. *Let us take X and Y as above. Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, we say that the family $F = (f^{(1)}, \dots, f^{(N)})$, where $f^{(i)} \in \mathcal{P}(T_i)$ for $i = 1, \dots, N$, is an n dimensional cardinal flow if*

$$\mu = (\mathfrak{p}_X)_\# f^{(1)},$$

$$\nu = (\mathfrak{p}_Y)_\# f^{(N)}$$

and, for each $i = 1, \dots, N-1$ holds true that

$$(\mathfrak{p}_{Z_i})_\# f^{(i)} = (\mathfrak{p}_{Z_i})_\# f^{(i+1)}.$$

We indicate that by $F \in \mathcal{F}(\mu, \nu)$.

Finally, we define the total cardinal flow cost as

$$\mathbb{CT}_c^{(N)}(F) := \sum_{i=1}^N \int_{Z_i \times Y_i} c_i df^{(i)}$$

where $c = \sum_{i=1}^N c_i$.

Theorem 3.5. *Let us take $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R}$ a separated cost function. Then, it holds true*

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathbb{T}_c(\pi) = \inf_{(f^{(1)}, \dots, f^{(N)}) \in \mathcal{F}(\mu,\nu)} \mathbb{CT}_c^{(N)}(F).$$

In this framework, given μ and ν two measures, we do not have a singular intermediate configuration that interpolates between them, but rather a plethora, one for each space Z_i. Those configurations will become more similar to ν as i grows.

Definition 3.14. *We call a pivoting chain between μ and ν a $(N+1)$ collection of measures*

$$\zeta := (\zeta_0, \dots, \zeta_N)$$

where $\zeta_0 = \mu$, $\zeta_N = \nu$ and $\zeta_i \in \mathcal{P}(Z_{i+1})$ are such that

$$\zeta_i = (\mathfrak{p}_{Z_i})_\# f^{(i)}$$

where $f^{(i)}$ are the optimal cardinal flows between μ and ν.

In this setting, we can still recover the results of the Theorem 3.3 and 3.4. In this case, we will have more pieces composing the various functionals we want to minimize, one for each measure in the pivoting chain.

Theorem 3.6. *Let $\mu, \nu \in \mathcal{P}(X)$ and c be a separable cost function. Then, we have*

$$W_c(\mu, \nu) = \inf_{\zeta = (\zeta_0, \dots, \zeta_d)} \sum_i^N \mathcal{C}^{(i)}(\zeta_i, \zeta_{i+1})$$

where $\mathcal{C}^{(i)}$ are the transportation functionals associated to the singular costs $c_i^{(\infty)}$.

Theorem 3.7. *Let d be a $n-$separable distance, then, if $\zeta = (\zeta_0, \dots, \zeta_N)$ is a pivoting chain, holds true the relation*

$$W_d(\mu, \nu) = \sum_{i=1}^N W_{d_i}(\zeta_{i-1}, \zeta_i).$$

Furthermore, every pivoting chain is a minimum of the functional

$$F(\zeta) = F((\zeta_0, \dots, \zeta_N)) := \sum_{i=1}^N W_d(\zeta_{i-1}, \zeta_i).$$

3.4 Separable costs on the Euclidean Plane

By requiring additional structure on the polish spaces X and Y, we can give a more specific description of what the cardinal flow is. In particular, we can delve further into the Formula (3.22) and describe the transportation cost as the integral of a "fiber by fiber" transportation between the elements of a pivot chain. We will also able to highlight an interesting connection with the Sliced Wasserstein distances, which recently have received a lot of attention from the applied fields. Furthermore, in the Euclidean space, we can give an

integral version of the conditions that define the cardinal flows between two given measures μ and ν.
For the sake of simplicity, in what follows we will then take $X = Y = \mathbb{R}^2$, before closing the section we will show how to generalize to higher dimension the results we will show.

Remark 3.11. *In this framework we can give an integral description of the cardinal flow. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$, a cardinal flow is a couple of measures $F = (f^{(1)}, f^{(2)})$ on $\mathbb{R}^3$ such that*

(A) $\int_{\mathbb{R}} f^{(1)}(x_1, x_2, dy_1) = \mu,$

(B) $\int_{\mathbb{R}} f^{(2)}(dx_2, y_1, y_2) = \nu,$

(C) $\int_{\mathbb{R}} f^{(1)}(dx_1, x_2, y_1) = \int_{\mathbb{R}} f^{(2)}(x_2, y_1, dy_2).$

With those notations the total cardinal flow cost is

$$\mathbb{CT}_c(F) = \sum_{i=1}^{2} \int_{\mathbb{R}^2 \times \mathbb{R}} c_i df^{(i)},$$

where $c = c_1 + c_2$.

Recalling what said in Chapter 1 about the interpretation of what a transportation plan is, we can give a similar meaning to the cardinal flow. Let us take $f^{(1)}$ and $f^{(2)}$ a cardinal flow and let us assume, for the sake of simplicity that both the measures are absolutely continuous respect the Lebesgue measure over $\mathbb{R}^3$ and let us denote with $f^{(1)}(x_1, x_2, y_1)$ and $f^{(2)}(x_2, y_1, y_2)$ their respective densities. The value $f^{(1)}(x_1, x_2, y_1)$ will then represent the mass moving from the point x_1 to the point y_1 on the line whose second coordinate is constantly x_2. Similarly, $f^{(2)}(x_2, y_1, y_2)$ will be the mass moving from the point x_2 to y_2 on the line whose first coordinate is constantly equal to y_1.
In this mindset, also the conditions we need to impose on the cardinal flow have a natural interpretation.

- The integral $\int_{\mathbb{R}} f^{(1)}(x_1, x_2, y_1) dy_1$ represent the total amount of mass that is spread from the point (x_1, x_2) along the horizontal line that passes from it. Condition A assures us that this amount of mass is equal to $\mu(x_1, x_2)$ i.e. to the source in (x_1, x_2).

- Similarly, the integral $\int_{\mathbb{R}} f^{(2)}(x_2, y_1, y_2) dx_2$ is the total amount of mass arriving in (y_1, y_2) through the vertical flow $f^{(2)}$. Conditions B is telling us that this amount is equal to $\nu(y_1, y_2)$, i.e. the sink.

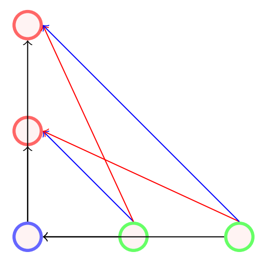


Figure 3.1: The lack of uniqueness showcased in example 3.2: every possible transportation plan induces the same pivot measure, hence, by knowing the pivot measure, we cannot retrieve the optimal transportation plan that generated it.

- Condition C is a continuity equation. Since $\int_{\mathbb{R}} f^{(1)}(x_1, x_2, y_1)dx_1$ is the total mass moved to (y_1, x_2) and $\int_{\mathbb{R}} f^{(2)}(x_2, y_1, y_2)dy_2$ is the total mass departing from (y_1, x_2), condition C is telling us that no mass is left untraveled during the "change of direction" that the mass accomplishes.

Remark 3.12. *As we exposed in Remark 2.11, we can think of π as the reallocating strategy of a company in charge of a transportation. This classical economic interpretation can be translated in the language of the cardinal flows. One can think of the cardinal flows as at two action made by two different companies that are in charge of moving the configuration μ into the configuration ν. The first company can move only horizontally, while the second one only vertically and they have to cooperate in order to move the mass in the cheaper way.*
In this setting, it is also possible to give an interpretation to the pivot measures. They represent the configuration which is cheapest for both the companies for "passing the baton" of traveling the mass. Obviously once the mass is settled in the shape of the pivot measure the company that moves it vertically has no clue on where the mass that he has to pick up and reshape came from. Roughly speaking, this lack of knowledge is the reason for which we can, given an optimal cardinal flow, rebuild a multitude of optimal transportation plans. As the following example shows.

Example 3.2. *Let us take μ and ν measures on $\mathbb{R}^2$ defined as*

$$\mu := \frac{1}{2}\left(\delta_{(1,0)} + \delta_{(2,0)}\right)$$

and

$$\nu := \frac{1}{2}\left(\delta_{(0,1)} + \delta_{(0,2)}\right).$$

The Corollary 3.2 tells us that the pivot measure has to be

$$\zeta = \delta_{(0,0)},$$

which means that the only possible cardinal flow (and therefore the optimal one), is

$$f^{(1)} := \frac{1}{2}\Big(\delta_{(1,0;0)} + \delta_{(2,0;0)}\Big)$$

and

$$f^{(2)} := \frac{1}{2}\Big(\delta_{(0,0;1)} + \delta_{(0,0;2)}\Big).$$

From formula (3.10) *in the Corollary 3.2, we find that the measure* π *defined as*

$$\pi := \frac{1}{4}\Big(\delta_{(1,0),(0,1)} + \delta_{(1,0),(0,2)} + \delta_{(2,0),(0,1)} + \delta_{(2,0),(0,2)}\Big)$$

is an optimal transportation plan between μ *and* ν*.*
Since they induce the same cardinal flows, also the plans

$$\pi_1 := \frac{1}{2}\Big(\delta_{(1,0),(0,1)} + \delta_{(2,0),(0,2)}\Big)$$

and

$$\pi_2 := \frac{1}{2}\Big(\delta_{(1,0),(0,2)} + \delta_{(2,0),(0,1)}\Big)$$

are optimal.
This lack of uniqueness is due to a natural "lack of memory" that is possible to see thanks to the cardinal flow formulation. Roughly speaking, once the mass is allocated into $(0,0)$ *by the first cardinal flow* $f^{(1)}$*, it merges in one point and loses its identity, which means that, once that the two masses are merged, we are no more able to tell from which point it came. When the second cardinal flows* $f^{(2)}$ *"picks up" the mass in* $(0,0)$ *and moves it vertically to complete the transportation we will not be able to remember how much of the mass that ended in* $(0,1)$*, came from the point* $(1,0)$ *or* $(2,0)$*.*
The plans π*,* π_1 *and* π_2 *are different for this reason: for* π *just half of the mass in* $(1,0)$ *goes to* $(0,1)$*, for* π_1 *all the mass in* $(1,0)$ *goes to* $(0,1)$ *and, for* π_2*, none of the mass in* $(1,0)$ *goes to* $(0,1)$*.*

Theorem 3.8. *Let* $(f^{(1)}, f^{(2)})$ *be a cardinal flow between* μ *and* ν *two given measures. The transportation plan that generates* $(f^{(1)}, f^{(2)})$ *is unique if and only if for almost every* $(y_1, x_2) \in \mathbb{R}^2$ *there exist a* x_1 *or a* y_2 *for which one of the following conditions holds true*

- $f^{(1)}_{|(y_1,x_2)} := \delta_{x_1}$,
- $f^{(2)}_{|(y_1,x_2)} := \delta_{y_2}$,

where $f^{(1)}_{|(y_1,x_2)}$ *and* $f^{(2)}_{|(y_1,x_2)}$ *are the conditional laws of* $f^{(1)}$ *and* $f^{(2)}$ *given* (y_1, x_2).
Furthermore, the unique transportation plan is given by the formula (3.10). *i.e.*

$$\pi := \begin{cases} f^{(1)} \otimes \delta_{y_2} & \textit{if} \quad f^{(2)}_{|(y_1,x_2)} := \delta_{x_1}, \\ f^{(2)} \otimes \delta_{x_1} & \textit{if} \quad f^{(1)}_{|(y_1,x_2)} := \delta_{x_1}, \\ 0 & \textit{otherwise.} \end{cases}$$

Remark 3.13. *The condition required in the Theorem 3.8 are the natural solutions of the problems highlighted in the Example 3.2: if we require that the flows are somehow deterministic, we have no problems discerning where the mass comes from.*

Theorem 3.9. *Let us take* $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ *and* c *a separable cost function. For any pivot measure,* ζ *holds true the formula*

$$W_c(\mu, \nu) = \int_{\mathbb{R}} W_{c_1}(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2 + \int_{\mathbb{R}} W_{c_2}(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1. \tag{3.27}$$

Proof. Let ζ be a pivot measure between μ and ν, then, from Remark 3.6 we know that $\zeta \in \mathcal{I}(\mu, \nu)$. The Disintegration Theorem 2.16 allow us to write

$$\zeta = \sigma^{(1)}_{|x_2} \otimes \mu_2 \tag{3.28}$$

and

$$\zeta = \sigma^{(2)}_{|y_1} \otimes \nu_1. \tag{3.29}$$

Similarly we can decompose μ and ν as

$$\mu = \mu_{|x_2} \otimes \mu_2 \tag{3.30}$$

and

$$\nu = \nu_{|y_1} \otimes \nu_1 \tag{3.31}$$

respectively. For μ_2-almost every $x_2 \in \mathbb{R}$ is then well defined the problem

$$W_{c_1}(\mu_{|x_2}, \sigma^{(1)}_{|x_2}) = \inf_{\pi_{|x_2} \in \Pi(\mu_{|x_2}, \sigma^{(1)}_{|x_2})} \int_{\mathbb{R}^2} c_1 d\pi_{|x_2}.$$

Theorem 2.16 assures us that the selections $x_2 \to \mu_{|x_2}$ and $x_2 \to \sigma^{(1)}_{|x_2}$ are both measurable, hence, according to Lemma 2.6, we know that there exists a measurable selection of optimal plans $\pi_{|x_2}$ for which holds true

$$W_{c_1}(\mu_{|x_2}, \sigma^{(1)}_{|x_2}) = \int_{\mathbb{R}^2} c_1 d\pi_{|x_2} \tag{3.32}$$

μ_2-almost everywhere $x_2 \in \mathbb{R}$. Similarly, there exists a measurable selection $\pi_{|y_1}$ for which, for ν_1-almost every $y_1 \in \mathbb{R}$, holds true that

$$W_{c_2}(\sigma^{(2)}_{|y_1}, \nu_{|y_1}) = \int_{\mathbb{R}^2} c_2 d\pi_{|y_1}. \tag{3.33}$$

Let us now consider the measures $f^{(1)} \in \mathcal{P}(\mathbb{R}^2 \times \mathbb{R})$ and $f^{(2)} \in \mathcal{P}(\mathbb{R} \times \mathbb{R}^2)$, defined as it follows

$$f^{(1)} = \pi_{|x_2} \otimes \mu_2 \tag{3.34}$$

and

$$f^{(2)} = \pi_{|y_1} \otimes \nu_1. \tag{3.35}$$

The couple $(f^{(1)}, f^{(2)})$ is a cardinal flow between μ and ν, in fact

$$\begin{aligned}
\int_{\mathbb{R}} f^{(1)} dx_1 &= \int_{\mathbb{R}} \pi_{|x_2}(x_1, y_1) \otimes \mu_2(x_2) dy_1 \\
&= \mu_2(x_2) \otimes \int_{\mathbb{R}} \pi_{|x_2}(x_1, y_1) dy_1 \\
&= \mu_{|x_2}(x_1) \otimes \mu_2(x_2) \\
&= \mu(x_1, x_2)
\end{aligned}$$

and

$$\begin{aligned}
\int_{\mathbb{R}} f^{(1)} dx_1 &= \int_{\mathbb{R}} \pi_{|x_2}(x_1, y_1) \otimes \mu_2(x_2) dx_1 \\
&= \mu_2(x_2) \otimes \int_{\mathbb{R}} \pi_{|x_2}(x_1, y_1) dx_1 \\
&= \sigma_{|x_2}(y_1) \otimes \mu_2(x_2) \\
&= \zeta(y_1, x_2).
\end{aligned}$$

Similarly, it is possible to show that

$$\int_{\mathbb{R}} f^{(2)} dy_2 = \zeta(y_1, x_2)$$

and

$$\int_{\mathbb{R}} f^{(2)} dx_2 = \nu(y_1, y_2),$$

hence $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$. From the identities (3.32) and (3.33), we have

$$\begin{aligned}
\int_{\mathbb{R}} W_{c_1}(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2 + \int_{\mathbb{R}} W_{c_2}(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1 \quad &= \quad \int_{\mathbb{R}} \int_{\mathbb{R}^2} c_1(x_1, y_1) d\pi_{|x_2} d\mu_2 \\
&\quad + \int_{\mathbb{R}} \int_{\mathbb{R}^2} c_2(x_1, y_1) d\pi_{|y_1} d\nu_1 \\
&= \quad \int_{\mathbb{R}^3} c_1(x_1, y_1) df^{(1)} \\
&\quad + \int_{\mathbb{R}^3} c_2(x_2, y_2) df^{(2)},
\end{aligned}$$

so that

$$\begin{aligned}
\int_{\mathbb{R}} W_{c_1}(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2 + \int_{\mathbb{R}} & W_{c_2}(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1 \\
& \geq \min_{(f^{(1)}, f^{(2)})} \mathbb{CT}_c(f^{(1)}, f^{(2)}). \qquad (3.36)
\end{aligned}$$

To prove the other inequality, let us now take $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$. By definition, we have that

$$\int_{\mathbb{R}\times\mathbb{R}} f^{(1)} dx_1 dy_1 = \int_{\mathbb{R}} \mu(x_1, x_2) dx_1 = \mu_2(x_2)$$

and

$$\int_{\mathbb{R}\times\mathbb{R}} f^{(2)} dx_2 dy_2 = \int_{\mathbb{R}} \nu(y_1, y_2) dy_2 = \nu_1(y_1)$$

which allow us to disintegrate $f^{(1)}$ and $f^{(2)}$ as it follows

$$f^{(1)} = \psi_{|x_2}(x_1, y_1) \otimes \mu_2(x_2)$$

and

$$f^{(2)} = \phi_{|y_1}(x_2, y_2) \otimes \nu_1(y_1).$$

If we denote with ζ the measure on which $f^{(1)}$ and $f^{(2)}$ glue, we have that $\psi_{|x_2} \in \Pi(\mu_{|x_2}, \zeta_{|x_2})$ μ_2-a.e. and $\phi_{|y_1} \in \Pi(\zeta_{|y_1}, \nu_{|y_1})$ ν_1-a.e.. In fact we have

$$\mu \quad = \quad \int_{\mathbb{R}} f^{(1)} dy_1 = \int_{\mathbb{R}} \psi_{|x_2} \otimes \mu_2 dy_1$$

$$= \left(\int_{\mathbb{R}} \psi_{|x_2} dy_1 \right) \otimes \mu_2$$

so that, by uniqueness of the conditional law, we have

$$\int_{\mathbb{R}} \psi_{|x_2} dy_1 = \mu_{|x_2},$$

similarly, we can prove

$$\int_{\mathbb{R}} \psi_{|x_2} dx_1 = \zeta_{|x_2}$$

so that $\psi_{|x_2} \in \Pi(\mu_{|x_2}, \zeta_{|x_2})$. Through a similar argument, we can prove that $\phi_{|y_1} \in \Pi(\zeta_{|y_1}, \nu_{|y_1})$.
We can then estimate each term of $\mathbb{CT}_c$ as

$$\begin{aligned} \int_{\mathbb{R}^2 \times \mathbb{R}} c^{(1)}(x_1, y_1) df^{(1)} &= \int_{\mathbb{R}} \int_{\mathbb{R} \times \mathbb{R}} c^{(1)}(x_1, y_1) d\psi_{|x_2} d\mu_2 \\ &\geq \int_{\mathbb{R}} W_{c_1}(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2 \end{aligned} \tag{3.37}$$

and

$$\begin{aligned} \int_{\mathbb{R}^2 \times \mathbb{R}} c^{(2)}(x_2, y_2) df^{(1)} &= \int_{\mathbb{R}} \int_{\mathbb{R} \times \mathbb{R}} c^{(2)}(x_2, y_2) d\phi_{|y_1} d\nu_1 \\ &\geq \int_{\mathbb{R}} W_{c_2}(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1. \end{aligned} \tag{3.38}$$

By summing up the relations (3.37) and (3.38) we can then conclude

$$W_c(\mu, \nu) \geq \int_{\mathbb{R}} W_{c_1}(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2 + \int_{\mathbb{R}} W_{c_2}(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1,$$

that, along with the relation (3.36), concludes the proof. □

In the previous theorem, we saw that, by choosing a suitable $\zeta \in \mathcal{I}(\mu, \nu)$, we can express the Wasserstein distance as a sum of two contributions. This suggests the definition of a function that evaluates how those contribution does change as the intermediate measure changes.

Definition 3.15 (Pivoting functional)**.** *Given two probability measures μ, ν and $c = c_1 + c_2$ a separable cost function, we define the pivoting functional $\mathbb{Z} : \mathcal{I}(\mu, \nu) \to \mathbb{R}$ as*

$$\mathbb{Z} : \zeta \to \int_{\mathbb{R}} W_{c_1}(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2 + \int_{\mathbb{R}} W_{c_2}(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1. \tag{3.39}$$

Theorem 3.10. *Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ and c a separable cost function, it holds true that*

$$W_c(\mu,\nu) = \inf_{\zeta \in \mathcal{I}(\mu,\nu)} \mathbb{Z}(\zeta).$$

Proof. Since any pivot measures ζ is an element of $\mathcal{I}(\mu,\nu)$, Theorem 3.9, assure us

$$W_c(\mu,\nu) \geq \inf_{\zeta \in \mathcal{I}(\mu,\nu)} \mathbb{Z}(\zeta).$$

To conclude we just need to prove the other inequality.
Let us fix $\zeta \in \mathcal{I}(\mu,\nu)$. Following the steps of the proof of Theorem 3.9, we can disintegrate μ, ζ and ν (see (3.28)-(3.31)), find the optimal transportation plans between the conditional measures and, finally, define the cardinal flow as done in (3.34) and (3.35). Since the couple $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$, we have that

$$\mathbb{Z}(\zeta) = \mathbb{CT}_c((f^{(1)}, f^{(2)})) \geq \inf_{(f^{(1)},f^{(2)}) \in \mathcal{F}(\mu,\nu)} \mathbb{CT}_c((f^{(1)}, f^{(2)})),$$

hence

$$\inf_{\sigma \in \mathcal{I}(\mu,\nu)} \mathbb{Z}(\zeta) \geq \inf_{(f^{(1)},f^{(2)}) \in \mathcal{F}(\mu,\nu)} \mathbb{CT}_c((f^{(1)}, f^{(2)})) = W_c(\mu,\nu).$$

□

If we can compute the Wasserstein distance between two one-dimensional probability measures, the last Theorem tells us that all the information needed to retrieve an optimal transportation plan between two bi-dimensional configurations is enclosed in the pivot measure.
This is the case when the cost function c_i are of the form

$$c_i(x,y) = h_i(|x-y|) \tag{3.40}$$

where $h_i : \mathbb{R} \to [0,\infty)$ is a convex function. In this scenario, Theorem 2.19 allows us to express the one-dimensional Wasserstein distance between two measures in terms of their cumulative functions.
We can then improve the previous result for separable cost functions whose components are of the form (3.40) and obtain the following.

Theorem 3.11. *Let us take $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ and c a separable cost function. Assume that $\zeta \in \mathcal{I}(\mu,\nu)$ is a pivot measure between μ and ν, then it holds true*

$$W_c(\mu,\nu) \quad = \quad \int_{\mathbb{R}} \int_{[0,1]} h(|F^{[-1]}_{\mu_{|x_2}}(s) - F^{[-1]}_{\zeta_{|x_2}}(s)|) ds d\mu_2$$

$$+\int_{\mathbb{R}}\int_{[0,1]} h(|F^{[-1]}_{\zeta_{|y_1}}(t) - F^{[-1]}_{\nu_{|y_1}}(t)|)dtd\nu_1,$$

where $F^{[-1]}_{\mu_{|x_2}}, F^{[-1]}_{\zeta_{|x_2}}, F^{[-1]}_{\zeta_{|y_1}}$ and $F^{[-1]}_{\nu_{|y_1}}$ are the pseudo-inverse functions of the cumulative function of $\mu_{|x_2}, \zeta_{|x_2}, \zeta_{|y_1}$ and $\nu_{|y_1}$, respectively.
In particular, if $h_1 = h_2 = |\circ|$, we have

$$W_1(\mu,\nu) = \int_{\mathbb{R}}\int_{\mathbb{R}} |F_{\mu_{|x_2}}(s) - F_{\zeta_{|x_2}}(s)|dsd\mu_2 + \int_{\mathbb{R}}\int_{\mathbb{R}} |F_{\zeta_{|y_1}}(t) - F_{\nu_{|y_1}}(t)|dtd\nu_1.$$

3.5 The co-monotone cardinal flow

In the previous section we showed that solving the cardinal flow problem can be done by considering two sub-problems: finding the pivot measure and computing the optimal transportation plan between a pair of one dimensional measures. In general, finding the pivot measure is much harder than solving the optimal transport problem between a pair of one dimensional measures. Since the decomposition done in Theorem 3.9 can be done for any intermediate measure. Given $\lambda \in \mathcal{I}(\mu,\nu)$, it is possible to define a class of one dimensional problems between the conditional laws, in the same way we did for the pivot measure. By doing so, we will find the best cardinal flow that glues on λ.

Definition 3.16 (co-monotone cardinal flow)**.** *Let $\mu,\nu \in \mathcal{P}(\mathbb{R}^2)$ and $\zeta \in \mathcal{I}(\mu,\nu)$. The co-monotone cardinal flow associated to ζ is the couple $(f^{(1)}, f^{(2)})$ in $\mathcal{F}(\mu,\nu)$ satisfying the following*

- *Given $x_2 \in \mathbb{R}$, we have that the conditional law of $f^{(1)}$ given x_2 is the monotone transportation plan between $\mu_{|x_2}$ and $\zeta_{|x_2}$.*
- *Given $y_1 \in \mathbb{R}$, we have that the conditional law of $f^{(2)}$ given y_1 is the monotone transportation plan between $\zeta_{|y_1}$ and $\nu_{|y_1}$.*

Remark 3.14. *Since Theorem 2.19 assures us that the monotone transport is unique, the definition above is well posed. Moreover, given $\mu,\nu \in \mathcal{P}(\mathbb{R}^2)$ and a $\zeta \in \mathcal{I}(\mu,\nu)$, we can explicitly write the co-monotone transportation plan associated to ζ. From Theorem 2.19, given $x_2 \in \mathbb{R}^2$, the optimal plan between $\mu_{|x_2}$ and $\zeta_{|x_2}$ is given by*

$$(F^{[-1]}_{\mu_{|x_2}}, F^{[-1]}_{\zeta_{|x_2}})_{\#}\mathcal{L}_{[0,1]}.$$

Similarly, given y_1, the monotone transportation plan between $\zeta_{|y_1}$ and $\nu_{|y_1}$ is

$$(F^{[-1]}_{\zeta_{|y_1}}, F^{[-1]}_{\nu_{|y_1}})_{\#}\mathcal{L}_{[0,1]}.$$

We can then define $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$ *as*

$$f^{(1)} = (F^{[-1]}_{\mu_{|x_2}}, F^{[-1]}_{\zeta_{|x_2}})_{\#}\mathcal{L}_{[0,1]} \otimes \mu_2$$

and

$$f^{(2)} = \nu_1 \otimes (F^{[-1]}_{\zeta_{|y_1}}, F^{[-1]}_{\nu_{|y_1}})_{\#}\mathcal{L}_{[0,1]}.$$

Given μ, ν and ζ as in Definition 3.16 we can find a co-monotone flow and then, from formula (3.10), we can recover a transportation plan whose marginals on the first three and on the last three components are $f^{(1)}$ and $f^{(2)}$ respectively. We will denote with $\mu \times_\zeta \nu$ this plan. Given μ, ν we can define a functional $\mathcal{R} : \mathcal{I}(\mu, \nu) \to \Pi(\mu, \nu)$ as

$$\mathcal{R}(\zeta) := \mu \otimes_\zeta \nu,$$

and, in this notations, it holds true that

$$\mathbb{Z}(\zeta) = \mathbb{B}(L(\mathcal{R}(\zeta))) := \mathbb{T}_c(\mathcal{R}(\zeta)). \tag{3.41}$$

We can think of the co-monotone cardinal flow as the optimal plan that passes through the intermedium measure ζ.

Theorem 3.12. *Let us take* $\mu, \nu \in \mathcal{I}(\mu, \nu)$ *and* c *a separable cost function of the form* (3.40). *Given any* $\zeta \in \mathcal{I}(\mu, \nu)$, *let* $(f^{(1)}, f^{(2)})$ *be the co-monotone cardinal flow associated to* ζ *then*

$$\inf_{\pi \in \Pi_\zeta(\mu,\nu)} \mathbb{T}_c(\pi) = \mathcal{B}\big((f^{(1)}, f^{(2)})\big)$$

where

$$\Pi_\zeta(\mu, \nu) := \left\{ \pi \in \Pi(\mu, \nu) s.t. \int_{\mathbb{R}^2} \pi dx_1 dy_2 = \zeta \right\}.$$

Given μ and ν two probability measures, we can arbitrarily fix a measure $\zeta \in \mathcal{I}(\mu, \nu)$ and always compute the co-monotone cardinal flow associated to it, consequentially, we can also find a whole transportation plan.

In what follows, we present two remarkable examples of those constructions: the first one is found by taking an intermediate measure that is completely independent on its components, this will allow us to define an improved version of the completely independent transportation plan. The second one will highlight an interesting relationship with a classical transportation plan: the Knothe-Rosenblatt rearrangement.

The co-monotone independent transportation plan

Given two probability measures μ and ν, we can always define the independent transportation plan between them as

$$\pi = \mu \otimes \nu.$$

Although it is easy to define, this plan is well known to be very unefficient whenever we take the L^p norm as a cost function. Part of this inefficiency is due to the lack of monotonicity "fiber by fiber". In other words, the cardinal flows induced by this plan are not co-monotone. We can then improve the optimality of the independent plan by taking the measure

$$\zeta = \nu_1 \otimes \mu_2 = \int_{\mathbb{R}^2} \pi dx_1 dy_2$$

and defining the co-monotone cardinal flow associated to this pivot measure. This allows us to define a new transportation plan between measures, the co-monotone independent plan. This is defined

$$\pi^{(o,v)} := (F^{[-1]}_{\mu_{|x_2}}, F^{[-1]}_{\mu_2}, F^{[-1]}_{\nu_1}, F^{[-1]}_{\nu_{|y_1}})_{\#}\mathcal{L}_{|[0,1]} \tag{3.42}$$

where $\mathcal{L}_{|[0,1]}$ is the Lebesgue measure restricted to $[0,1]$. If we consider the reverse problem we can define the vertical-orizontal co-monotone transportation plan as

$$\pi^{(v,o)} := (F^{[-1]}_{\mu_1}, F^{[-1]}_{\mu_{|x_1}}, F^{[-1]}_{\nu_{|y_2}}, F^{[-1]}_{\nu_2})_{\#}\mathcal{L}_{|[0,1]}.$$

Theorem 3.13 (Closed Formula for Independent Measures)**.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ be independent, i.e., if we call μ_i, ν_i their $i^{th}-$marginals, we have*

$$\mu = \mu_1 \otimes \mu_2 \qquad \textit{and} \qquad \nu = \nu_1 \otimes \nu_2.$$

Then, the plan $\pi^{(o,v)}$ defined in (3.42)*, is optimal for any power-like separable cost function.*

Proof. From hypotesis, we have $\mu = \mu_1 \otimes \mu_2$ and $\zeta = \nu_1 \otimes \mu_2$, so that

$$W_c(\mu_{|x_2}, \zeta_{|x_2}) = W_c(\mu_1, \nu_1)$$

for μ_2-almost each $x_2 \in \mathbb{R}$. Then, the map

$$x_2 \to \pi^{(x_2)}$$

where $\pi^{(x_2)}$ is the optimal transportation plan between $\mu_{|x_2}$ and $\zeta_{|x_2}$ is constant. By denoting with $\pi^{(1)}$ the constant value, we have that the first cardinal flow is

$$f^{(1)} = \pi^{(1)} \otimes \mu_2.$$

Similarly, also the map that to each y_1 associate the optimal transportation plan between $\zeta_{|y_1}$ and $\nu_{|y_1}$

$$y_1 \to \pi^{(y_1)}$$

is constant. Denoted with $\pi^{(2)}$ the constant value, we have that the second cardinal flow is $f^{(2)} = \pi^{(2)} \otimes \nu_1$.
By simplifying the formula (3.10), we have that a plan that induces this cardinal flow is

$$\pi = \pi^{(1)} \otimes \pi^{(2)}.$$

Notice that $(\mathbf{x}, \mathbf{y}) \in spt(\pi)$ if and only if $(x_1, y_1) \in spt(\pi^{(1)})$ and $(x_2, y_2) \in spt(\pi^{(2)})$.
If we show that the support of π is $c-$ciclically monotone, we will conclude the optimality of π from Theorem 2.11.
By absurd, let us take $((x_1, x_2), (y_1, y_2)), ((x_1', x_2'), (y_1', y_2')) \in spt(\pi)$ such that

$$c((x_1, x_2), (y_1', y_2')) + c((x_1', x_2'), (y_1, y_2)) < c((x_1, x_2), (y_1, y_2)) + c((x_1', x_2'), (y_1', y_2')).$$

From separability of c, we can rewrite the relation above as

$$\begin{aligned} c_1(x_1, y_1') + c_1(x_1', y_1) + c_2(x_2, y_2') + c_2(x_2', y_2) \quad < \quad & c_1(x_1, y_1) + c_1(x_1', y_1') \\ & + c_2(x_2, y_2) + c_2(x_2', y_2'). \end{aligned}$$

Hence, either

$$c_1(x_1, y_1') + c_1(x_1', y_1) < c_1(x_1, y_1) + c_1(x_1, y_1)$$

or

$$c_2(x_2, y_2') + c_2(x_2', y_2) < c_2(x_2', y_2') + c_2(x_2', y_2')$$

must hold true, but both are impossible since the supports of $\pi^{(1)}$ and $\pi^{(2)}$ are monotone. Since this argument can be applied to any given set of points $(x_i, y_i) \in spt(\pi)$, we can conclude the thesis. $\square$

The Knothe-Rosenblatt rearrangement

This paragraph is devoted to the study of an other intermediate measure, the Knothe-Rosenblatt pivoting measure. As the name suggests, this measure is closely related to the Knothe-Rosenblatt rearrangement.

Definition 3.17 (Knothe-Rosenblatt pivoting measure)**.** *Let us take $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ and assume that $\mu_1 = (p_1^{(1)})_\#\mu$ is absolutely continuous. The Knothe-Rosenblatt pivoting measure between them is defined as it follow*

$$\zeta_{KR}^{(\mu,\nu)} := \big(T_1, p_2^{(1)}\big)_\#\mu$$

where T_1 is the monotone rearrangement that sends μ_1 into ν_1.

Proposition 3.3. *Given a pair $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$, we have $\zeta_{KR}^{(\mu,\nu)} \in \mathcal{I}(\mu, \nu)$.*

Proof. Let us fix $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$. For the sake of semplicity, we denote with ζ the Knothe-Rosenblatt pivoting measure.
From a direct computation, we get

$$\begin{aligned}
(p_2^{(1)})_\#\zeta &= (p_2^{(1)})_\#(T_1, p_2^{(1)})_\#\mu \\
&= (p_2^{(1)})_\#\mu \\
&= \mu_2.
\end{aligned}$$

To conclude the proof we just need to show $\nu_1 = (p_1^{(1)})_\#\zeta$, once again, by a direct computation we find

$$\begin{aligned}
(p_1^{(1)})_\#\zeta &= (p_1^{(1)})_\#(T_1, p_2^{(1)})_\#\mu \\
&= (T_1)_\#\mu \\
&= \nu_1.
\end{aligned}$$

Where the last equality comes from the relation

$$T_1 = T_1 \circ p_1^{(1)}$$

and by applying the chain rule for pushforwards. □

Now that we proved $\zeta_{KR}^{(\mu,\nu)} \in \mathcal{I}(\mu, \nu)$, we can study the co-monotone cardinal flow associated to this measure. For the sake of simplicity, we make a further assumption on the measure ν and require that also its first marginal is absolutely continuous.
In this case, we can define the inverse function of T_1, namely S_1. This hypothesis allow us to express the conditional laws of $\zeta_{KR}^{(\mu,\nu)}$ through the conditional laws of μ and ν.

Lemma 3.5. *Let us take $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$, such that both μ_1 and ν_1 are absolutely continuous. Denoted with T_1 the monotone rearrangement that sends μ_1 into ν_1, we have*

$$(\zeta_{KR}^{(\mu,\nu)})_{|y_1} = \mu_{|x_1}$$

if $T_1(x_1) = y_1$ and

$$(\zeta_{KR}^{(\mu,\nu)})_{|x_2} = (T_1)_{\#}\mu_{|x_2}.$$

Proof. Let us fix $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ as above. For the sake of simplicity, let us denote with ζ the Knothe-Rosenblatt pivoting measure $\zeta_{KR}^{(\mu,\nu)}$. Let us then take $\phi \in C(\mathbb{R}^2)$, by definition of ζ and from the disintegration formula (2.4), we have

$$\begin{aligned} \int_{\mathbb{R}^2} \phi(y_1, x_2) d\zeta &= \int_{\mathbb{R}^2} \phi(T_1(x_1), x_2) d\mu \\ &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(T_1(x_1), x_2) d\mu_{|x_2} \Big) d\mu_2. \end{aligned}$$

Recalling Remark 3.5, we have that $\mu_2 = \zeta_2$, then we can write

$$\begin{aligned} \int_{\mathbb{R}^2} \phi(y_1, x_2) d\zeta &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(T_1(x_1), x_2) d\mu_{|x_2} \Big) d\zeta_2 \\ &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(y_1, x_2) d(T_1)_{\#}\mu_{|x_2} \Big) d\zeta_2 \end{aligned}$$

hence, from the uniqueness of the conditional law, we find $\zeta_{|x_2} = (T_1)_{\#}\mu_{|x_2}$. Similarly, if we decompose μ respect the other variable we find

$$\int_{\mathbb{R}^2} \phi(y_1, x_2) d\zeta = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(T_1(x_1), x_2) d\mu_{|x_1} \Big) d\mu_1.$$

Since both μ_1 and ν_1 are both absolutely continuous we can invert the optimal transportation map between μ_1 and ν_1. We denote by S_1 the inverse function. Both T_1 and S_1 are measurable and then, by consequence, so is the function

$$\Psi : y_1 \to \int_{\mathbb{R}} \phi(y_1, x_2) d\mu_{|S_1(y_1)}.$$

Notice that, in these notations, we have

$$\begin{aligned} \int_{\mathbb{R}^2} \phi(y_1, x_2) d\zeta &= \int_{\mathbb{R}} \big(\Psi \circ T_1\big) d\mu_1 \\ &= \int_{\mathbb{R}} \Psi d(T_1)_{\#}\mu_1 \end{aligned}$$

$$= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(y_1, x_2) d\mu_{|S_1(y_1)} \Big) d\nu_1.$$

Again from Remark 3.5, we have $\nu_1 = \zeta_1$, so that

$$\int_{\mathbb{R}^2} \phi(y_1, x_2) d\zeta = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(y_1, x_2) d\mu_{|S_1(y_1)} \Big) d\zeta_1$$

and, therefore, we have $\zeta_{|y_1} = \mu_{|S_1(y_1)}$ or, equivalentely, $\mu_{|x_1} = \zeta_{|T_1(x_1)}$. □

As we said above, it is possible to relate the measure defined in Definition 3.17 to the Knothe-Rosenblatt rearrangement π_{KR}. In the next result, we highlihgt this relationship by proving that the co-monotone cardinal flow associated to ζ_{KR} are the same flow induced by the plan π_{KR}.

Theorem 3.14. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$. If μ_1 and ν_1 are absolutely continuous, it holds true that*

$$T_c(\pi_{KR}) := \mathbb{Z}(\zeta_{KR}).$$

where π_{KR} is the Knothe-Rosenblatt transportation plan (defined in Remark 2.18). Moreover, the cardinal flow $L(\pi_{KR})$ is the co-monotone cardinal flow associated to ζ_{KR}.

Proof. We will show that

$$\int (x_1 - y_1)^2 d\pi_{KR} = \int W_2^2(\mu_{|x_2}, \zeta_{|x_2}) d\mu_2$$

and

$$\int (x_2 - y_2)^2 d\pi_{KR} = \int W_2^2(\zeta_{|y_1}, \nu_{|y_1}) d\nu_1.$$

We remind that the Knothe-Rosenblatt rearrangement plan π_{KR} is given by

$$\pi_{KR} := (Id; T_1, T_2)_{\#}\mu$$

where T_1 is the optimal map that sends μ_1 into ν_1, while T_2 is such that, fixed x_1 the map

$$T_2^{(x_1)} : x_2 \to T_2(x_1, x_2)$$

is the optimal transportation map between $\mu_{|x_1}$ and $\nu_{|T_1(x_1)}$, in particular, it holds true that

$$W_2^2(\mu_{|x_1}, \nu_{|T_1(x_1)}) = \int_{\mathbb{R}} (x_2 - T_2(x_1, x_2))^2 d\mu_{|x_1} \tag{3.43}$$

for μ_1-almost every $x_1 \in \mathbb{R}$.

By integrating both sides of relation (3.43) respect μ_1 we get

$$\begin{aligned}\int_{\mathbb{R}} W_2^2(\mu_{|x_1}, \nu_{|T_1(x_1)})d\mu_1 &= \int_{\mathbb{R}}\int_{\mathbb{R}}(x_2 - T_2(x_1,x_2))^2 d\mu_{|x_1} d\mu_1 \\ &= \int_{\mathbb{R}^2}(x_2 - T_2(x_1,x_2))^2 d\mu \\ &= \int_{\mathbb{R}^2\times\mathbb{R}^2}(x_2 - y_2)^2 d\pi_{KR}.\end{aligned}$$

From the characterization of the push-forward measure (2.3), we have that

$$\int_{\mathbb{R}} W_2^2(\zeta_{|T(x_1)}, \nu_{|T_1(x_1)})d\mu_1 = \int_{\mathbb{R}} W_2^2(\zeta_{|y_1}, \nu_{|y_1})d\nu_1$$

and then

$$\int_{\mathbb{R}} W_2^2(\zeta_{|y_1}, \nu_{|y_1})d\nu_1 = \int_{\mathbb{R}^2\times\mathbb{R}^2}(x_2 - y_2)^2 d\pi_{KR}.$$

If we show that

$$\int_{\mathbb{R}} W_2^2(\mu_{|x_2}, \zeta_{|x_2})d\nu_1 = \int_{\mathbb{R}^2\times\mathbb{R}^2}(x_1 - y_1)^2 d\pi_{KR}$$

we conclude the proof. Using again (2.3) we have that

$$(\zeta_{KR})_{|x_2} = (T_1)_{\#}\mu_{|x_2}.$$

Since T_1 is a monotone map, from Theorem 2.19, it is also the optimal map between $\mu_{|x_2}$ and $(T_1)_{\#}\mu_{|x_2}$, i.e.

$$W_2^2(\mu_{|x_2}, (\zeta_{KR})_{|x_2}) = \int_{\mathbb{R}}(x_1 - T_1(x_1))^2 d\mu_{|x_2}$$

for μ_2-almost each x_2. If we integrate both sides of the previous relationship with respect μ_2

$$\begin{aligned}\int_{\mathbb{R}} W_2^2(\mu_{|x_2}, (\zeta_{KR})_{|x_2})d\mu_2 &= \int_{\mathbb{R}}\int_{\mathbb{R}}(x_1 - T_1(x_1))^2 d\mu_{|x_2} d\mu_2 \\ &= \int_{\mathbb{R}^2}(x_1 - T_1(x_1))^2 d\mu \\ &= \int_{\mathbb{R}^2\times\mathbb{R}^2}(x_2 - y_2)^2 d\pi_{KR},\end{aligned}$$

hence

$$\mathbb{Z}(\zeta_{KR}^{(\mu,\nu)}) = \mathbb{T}_c(\pi_{KR}).$$

To prove that $L(\pi_{KR})$ is the co-monotone cardinal flow associated to $\zeta_{KR}^{(\mu,\nu)}$ it suffice to notice that

$$\left(p_1^{(2)}, p_2^{(1)}\right)_{\#}\pi_{KR} = \zeta_{KR}^{(\mu,\nu)}.$$

$\square$

3.5.1 Cardinal flows on $\mathbb{R}^d$

Once again, everything said up to now can be extended to higher dimensional euclidean spaces through minimal changes. We denote with $(\mathfrak{p}_{<i})$ the projection on the first $(i-1)-$components of $\mathbf{x} \in \mathbb{R}^d$, i.e.

$$(\mathfrak{p}_{<i})(\mathbf{x}) = \mathbf{x}_{<i} := (x_1, \dots, x_{i-1}).$$

while, with $(\mathfrak{p}_{>i})$ we denote the projection on the last $(d-(i+1))-$components, i.e.

$$(\mathfrak{p}_{>i})(\mathbf{x}) = \mathbf{x}_{>i} := (x_{i+1}, \dots, x_d).$$

Accordingly to those notations, given a measure $\zeta \in \mathcal{P}(\mathbb{R}^d)$, let us denote with $\zeta_{<i,>i}^{(i)}$ is the marginal of $\zeta^{(i)}$ of the variables $(\mathbf{y}_{<i}, \mathbf{x}_{>i})$, i.e.

$$\zeta_{<i,>i} := (((\mathfrak{p}_{<i}), (\mathfrak{p}_{>i}))_{\#}\zeta$$

and with $\zeta_{|(\mathbf{y}_{<i},\mathbf{x}_{>i})}$ the conditional law of ζ given the first $(i-1)-$th and the last $(d-(i-1))-$th components so that it holds true that

$$\zeta = \zeta_{|(\mathbf{y}_{<i},\mathbf{x}_{>i})} \otimes \zeta_{<i,>i}.$$

With this notations we can extend Theorem 3.9 as it follows.

Theorem 3.15. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a separable cost function, i.e. of the form*

$$c(\boldsymbol{x}, \boldsymbol{y}) := \sum_{i=1}^d c_i(x_i, y_i).$$

Using the notations above, it holds true

$$W_c(\mu,\nu) := \sum_{i=1}^d \int_{\mathbb{R}^{d-1}} W_{c_i}\left(\zeta^{(i-1)}_{|(\boldsymbol{y}_{<i},\boldsymbol{x}_{>i})}, \zeta^{(i)}_{|(\boldsymbol{y}_{<i},\boldsymbol{x}_{>i})}\right) d\zeta^{(i)}_{<i,>i}$$

where $\zeta := (\zeta_0, \dots, \zeta_d)$ is a chain of pivot measures between μ and ν.

3.6 The Radiant Formula

On vectorial spaces, the separability of the cost function can depend on the choice of the base we describe the space with. In this section, we point out how this freedom affects the cardinal flow formulation. This allows us to introduce an extension of the previous result to a wider class of cost functions. Moreover, we study the squared Euclidean cost function. By exploiting its invariance under isometries, we introduce the Radiant Formula, which we use to highlight an interesting connection between the cardinal flow formulation and the sliced Wasserstein Distance.

3.6.1 Separability

Up to now we assumed to have separable metrics on $\mathbb{R}^d$. To be more precise, we considered metrics separable with respect the canonical base on $\mathbb{R}^d$. In this subsection, we will point out how we can extend those reformulation to a slightly more general class of cost functions, the $\mathcal{B}-$*separable* one.

Definition 3.18 ($\mathcal{B}-$separable cost functions)**.** *Given a base $\mathcal{B}$ of $\mathbb{R}^d$, we say that a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ is $\mathcal{B}-$separable if it holds true*

$$c(\boldsymbol{x}_{\mathcal{B}}, \boldsymbol{y}_{\mathcal{B}}) := \sum_{i=1}^{d} c_i((x_{\mathcal{B}})_i, (y_{\mathcal{B}})_i)$$

where $\boldsymbol{x}_{\mathcal{B}} = ((x_{\mathcal{B}})_1, \dots, (x_{\mathcal{B}})_d)$ and $\boldsymbol{y}_{\mathcal{B}} = ((y_{\mathcal{B}})_1, \dots, (y_{\mathcal{B}})_d)$ are the coordinates with respect to the base $\mathcal{B}$. Equivalently, we say that the base $\mathcal{B}$ separates the cost function c.

Given any cost function, there might not exist a basis that separates it. However, when we can separate the cost function, we can also define an *ad hoc* cardinal flow problem that inherits all the good properties of the classical cardinal flow formulation.

Definition 3.19. *Given two measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and a base $\mathcal{B} := \{v_1, \dots, v_d\}$, we define the $\mathcal{B}-$cardinal flow between μ and ν as the collection*

$$\mathcal{F}_{\mathcal{B}} := ((f_{\mathcal{B}}^{(1)}), \dots, (f_{\mathcal{B}}^{(d)}))$$

where the $f_{\mathcal{B}}^{(i)} \in \mathcal{P}(\mathbb{R}^{d+1})$ are such that

- $\int_{\mathbb{R}} df_{\mathcal{B}}^{(1)}((y_{\mathcal{B}})_1) = \mu,$

- $\int_{\mathbb{R}} df_{\mathcal{B}}^{(d)}((x_{\mathcal{B}})_d) = \nu$,
- $\int_{\mathbb{R}} df_{\mathcal{B}}^{(i)}((x_{\mathcal{B}})_i) = \int_{\mathbb{R}} df_{\mathcal{B}}^{(i+1)}((y_{\mathcal{B}})_{i+1})$ *for each* $i = 1, \dots, d-1$.

We denote with $\mathcal{F}_{\mathcal{B}}(\mu, \nu)$ *the set of all* $\mathcal{B}-$*cardinal flows between* μ *and* ν.

Roughly speaking, the main difference between the cardinal flows and the $\mathcal{B}-$flows are the directions on which the mass travels.

Definition 3.20. *Let* $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ *and* $\mathcal{B}$ *a base on* $\mathbb{R}^d$*.Given* c *a* $\mathcal{B}-$*separable cost function, the* $\mathcal{B}-$*cardinal flow functional is defined as*

$$\mathbb{CT}_c^{(\mathcal{B})}(\mathcal{F}_{\mathcal{B}}) := \sum_{i=1}^{d} \int_{\mathbb{R}^{d+1}} c_i((x_{\mathcal{B}})_i, (y_{\mathcal{B}})_i) df_{\mathcal{B}}^{(i)}.$$

Example 3.3. *Let us take the cost function*

$$c(\boldsymbol{x}, \boldsymbol{y}) := (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{y})$$

where $\boldsymbol{A}$ *is a symmetric definite positive matrix. We know that there exists an orthogonal matrix* $\mathbf{O}$ *such that*

$$\boldsymbol{A} := \mathbf{O}^T \mathbf{D} \mathbf{O}$$

where $\mathbf{D}$ *is a diagonal matrix with positive diagonal entries, namely* λ_i*, for* $i = 1, \dots, d$*. It is well known that the columns of the matrix* $\mathbf{O}$ *are an (orthogonal) base of* $\mathbb{R}^d$*, that we call* $\mathcal{B}_O$*, and* O^T *is the matrix that describes the change of coordinate with from the base* $\mathcal{B}_O$ *to the canonical one.*
We can then define the cost function c_O *as it follows*

$$c_{\mathbf{O}}(\boldsymbol{x}', \boldsymbol{y}') := c(\mathbf{O}^T \boldsymbol{x}', \mathbf{O}^T \boldsymbol{y}') = (\boldsymbol{x}' - \boldsymbol{y}')^T \mathbf{D} (\boldsymbol{x}' - \boldsymbol{y}') = \sum_{i=1}^{d} \lambda_i (\boldsymbol{x}'_i - \boldsymbol{y}'_i)^2$$

which is totally separate, i.e. the base $\mathcal{B}_{\mathbf{O}}$ *separates the cost function* c.
Notice that, following the same argument, every cost function $c_{\boldsymbol{A}}$ *of the form*

$$c_{\boldsymbol{A}}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{y})$$

with $\boldsymbol{A}$ *diagonalizable matrix is a* $\mathcal{B}_{\mathbf{O}}-$*separable cost function for an orthogonal matrix* $\mathbf{O}$.

3.6.2 The Radiant Formula

A classical and omnipresent cost function in literature is the one induced by the squared Euclidean distance. In this section, we focus on this cost function and study its invariance under a specific class of isometries: the rotations.
Any rotation sends the canonical base into a new and orthogonal base of the space, using then the $\mathcal{B}-$flows introduced above, we can define a class of equivalent problems that will lead us to the Radiant Formula.
For the sake of clarity, we restrict to the bi-dimensional problem.

Definition 3.21. *Given $\theta \in [0, 2\pi]$, we define the $\theta-$rotated canonical base as $V_\theta := \{v_\theta, v_{\theta^\perp}\}$, where*

$$v_\theta := (cos(\theta), sen(\theta))$$

and

$$v_{\theta^\perp} := (-sen(\theta), cos(\theta)).$$

With $\boldsymbol{x} = (x_1, x_2)$ we denote the coordinates of a point in $\mathbb{R}^2$ with respect the canonical base, while with $\boldsymbol{x}^\theta = (x_1^\theta, x_2^\theta)$ we denote its coordinates with respect the base V_θ. With $(\mathfrak{p}_\theta)$ we denote the projection on span$\{v_\theta\}$,

$$c(\boldsymbol{x}^\theta) = x_1^\theta$$

while, with $(\mathfrak{p}_{\theta^\perp})$, we denote the projection on span$\{v_{\theta^\perp}\}$,

$$(\mathfrak{p}_{\theta^\perp})(\boldsymbol{x}^\theta) = x_2^\theta.$$

Accordingly to the previous notations, given $\mu \in \mathcal{P}(\mathbb{R}^2)$, we define as μ_θ the marginal of μ on span$\{v_\theta\}$,

$$\mu_\theta := (\mathfrak{p}_\theta)_\# \mu$$

and with $(\mu)_{|\theta}$ the conditional of μ given x_1^θ, so that holds true

$$\mu = (\mu)_{|\theta} \otimes \mu_\theta$$

for any $\theta \in (0, 2\pi]$.

Since the cost function

$$c(\mathbf{x}, \mathbf{y}) := (x_1 - y_1)^2 + (x_2 - y_2)^2$$

is invariant under rotations, we have that

$$(x_1 - y_1)^2 + (x_2 - y_2)^2 = (x_1^\theta - y_1^\theta)^2 + (x_2^\theta - y_2^\theta)^2 \tag{3.44}$$

for any given $\theta \in [0, 2\pi)$. In particular, for any V_θ, we can define the $V_\theta-$flow and find the associated pivot measure $\zeta^{(\theta)}$. If μ and ν are such that the optimal transportation plan between them is unique, so is the pivot measure $\zeta^{(\theta)}$, and in particular, the function

$$\theta \to \zeta^{(\theta)}$$

is well defined.
As the next example highlights, we can think of each $\zeta^{(\theta)}$ as of the same pivot measure seen under a different point of view.

Example 3.4. *Let us take $\mu = \delta_{(0,0)}$ and $\nu = \delta_{(1,1)}$, from a simple computation we find that $W_2^2(\mu, \nu) = 2$. Given $\theta \in (0, 2\pi]$, the pivot measure $\zeta^{(\theta)}$ associated to the base V_θ is*

$$\zeta^{(\theta)} = \delta_{\big(cos(\theta)+sen(\theta)\big)(cos(\theta),sen(\theta))}$$

hence the first cardinal flow is given by

$$f^{(\theta)} = \delta_{\big((0,0);cos(\theta)+sen(\theta)\big)}$$

and the cost of moving the mass along the first direction v_θ is

$$\int_{\mathbb{R}^3} \big(x_1^{(\theta)} - y_1^{(\theta)}\big)^2 df^{(\theta)} = \big(cos(\theta) - sen(\theta)\big)^2 = 1 - 2cos(\theta)sen(\theta).$$

Again, we remark that this formula holds true for each $\theta \in (0, 2\pi]$, so that, by taking the integral media over $(0, 2\pi]$, we find

$$\frac{1}{2\pi}\int_0^{2\pi} \big(1 - 2cos(\theta)sen(\theta)\big)d\theta = 1 = \frac{1}{2}W_2^2(\mu, \nu). \tag{3.45}$$

The latter example shows us that, if we know all the pivot measures $\zeta^{(\theta)}$, we are able to compute the Wasserstein distance. Moreover, this can be done without knowing both the second cardinal flow and the measure ν.
We can then think of the Radiant Formula as a generalization of (3.45).

Theorem 3.16 (The Radiant Formula)**.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ be absolutely continuous, so that the optimal transportation plan between μ and ν is unique. Then, using the notation introduced above*

$$W_2^2(\mu, \nu) = \frac{1}{2\pi}\int_0^{2\pi} \bigg(\int_{\mathbb{R}} W_2^2(\mu_{|\theta}, \zeta_{|\theta}^{(\theta)})d\mu_\theta\bigg)d\theta, \tag{3.46}$$

where $\zeta_{|\theta}^{(\theta)}$ is the conditional law of ζ given x_2^θ.

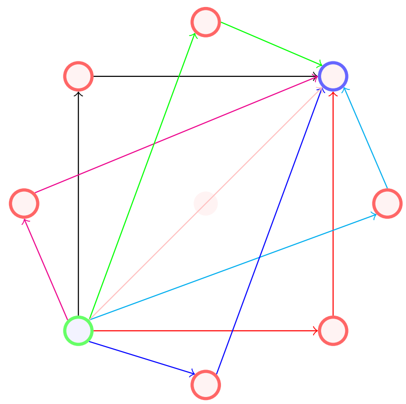


Figure 3.2: Some of the pivot measures described in Example 3.4. All those measures lies on a circumference centered in $\left(\frac{1}{2}, \frac{1}{2}\right)$

Proof. For any $\theta \in (0, 2\pi]$ let us denote with $\zeta^{(\theta)}$ the pivot measure for the base V_θ. Since the optimal transportation is unique, so that $\zeta^{(\theta)}$ is uniquely defined. For any of such $\zeta^{(\theta)}$ holds then true formula (3.27)

$$W_2^2(\mu, \nu) = \int_{\mathbb{R}} W_2^2(\mu_{|\theta}, \zeta_{|\theta}^{(\theta)}) d\mu_\theta + \int_{\mathbb{R}} W_2^2(\zeta_{|\theta^\perp}^{(\theta)}, \nu_{|\theta^\perp}) d\nu_{\theta^\perp}. \tag{3.47}$$

From Proposition 3.2 we know that

$$\int_{\mathbb{R}} W_2^2(\zeta_{|\theta^\perp}^{(\theta)}, \nu_{|\theta^\perp}) d\nu_{\theta^\perp} = \int_{\mathbb{R}} W_2^2(\mu_{|\theta}, \zeta_{|\theta}^{(\theta+\frac{\pi}{2})}) d\mu_{(\theta+\frac{\pi}{2})} \tag{3.48}$$

for each $\theta \in (0, 2\pi]$.
By substituting (3.48) in (3.47) we find

$$W_2^2(\mu, \nu) = \int_{\mathbb{R}} W_2^2(\mu_{|\theta}, \zeta_{|\theta}^{(\theta)}) d\mu_\theta + \int_{\mathbb{R}} W_2^2(\mu_{|\theta}, \zeta_{|\theta}^{(\theta+\frac{\pi}{2})}) d\mu_{(\theta+\frac{\pi}{2})}. \tag{3.49}$$

Finally, since (3.49) holds true for any θ, we can take the integral media and find the radiant formula

$$\begin{aligned} W_2^2(\mu, \nu) &= \frac{1}{2\pi} \int_{[0,2\pi]} W_2^2(\mu, \nu) d\theta \\ &= \frac{1}{2\pi} \int_{[0,2\pi]} \int_{\mathbb{R}} W_2^2(\mu_{|\theta}, \zeta_{|\theta}^{(\theta)}) d\mu_\theta d\theta \end{aligned}$$

$$+\frac{1}{2\pi}\int_{[0,2\pi]}\int_{\mathbb{R}} W_2^2(\mu_{|\theta},\zeta_{|\theta}^{(\theta+\frac{\pi}{2})})d\mu_{(\theta+\frac{\pi}{2})}d\theta \tag{3.50}$$
$$= 2\frac{1}{2\pi}\int_{[0,2\pi]}\int_{\mathbb{R}} W_2^2(\mu_{|\theta},\zeta_{|\theta}^{(\theta)})d\mu_\theta d\theta.$$

□

Remark 3.15. *We talked about the case in which we have measures over a bi-dimensional space. When we are dealing with measures over a higher dimensional case we can recover a similar formula*

$$W_2^2(\mu,\nu) = \frac{d}{|S^{(d-1)}|}\int_{S^{(d-1)}}\int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|\theta},\zeta_{|\theta}^{(\theta)})d\mu_\theta d\theta$$

where d is the dimension on which we are working. To obtain it is sufficient to repeat the passages made in before. The coefficient d comes from the fact that we need to use d different relations.

Relation with the Sliced Wasserstein Distance

The Sliced Wasserstein Distance is an alternative way to compare measures which shares some properties with the regular Wasserstein distance but is much simpler to compute. For this reason, this distance has recently proven to be a reliable tool in the computation of barycenters [70, 16], in generative modeling [47, 65, 46, 89], and in many others applied fields [48, 21]. Due to this success, during the last year have been developed variants of this distance [10, 45, 68].
The idea behind it is the following: given two measures μ and ν, the Sliced Wasserstein Distance takes all the marginals over all the possible one-dimensional spaces and compares them through the classical W_2 distance. The Sliced Wasserstein Distance is then computed by taking the average of the computed distances over all those directions (for more details, see [70]). Since all the marginals are one dimensional, the computation of those Wasserstein distances is easy and fast, making the computation of the whole Sliced Wasserestain Distance fast and efficient.

Definition 3.22. *Let $\mu,\nu \in \mathcal{P}(\mathbb{R}^2)$, then the sliced Wasserstein distance between μ and ν is defined as it follows*

$$SW_p^p(\mu,\nu) := \frac{1}{2\pi}\int_{[0,2\pi]} W_p^p(\mu_\theta,\nu_\theta)d\theta,$$

where the W_p^p is the p− Wasserstein Distance between 1−dimensional measures and μ_θ and ν_θ are the marginals of μ and ν with respect the first coordinate of the base V_θ.

Remark 3.16. *It is possible to show that SW_p is indeed a distance over $\mathcal{P}_p(\mathbb{R}^2)$ and that it is equivalent to the classical W_p distance. This can be done by using the Fourier Analysis for measures, as it has been done in [17].*

Proposition 3.4. *Given two measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, it holds true that*

$$W_2^2(\mu,\nu) \geq dSW_2^2(\mu,\nu).$$

Proof. It follow from the convexity of the W_2^2 functional stated in Theorem 2.14. In fact, from the radiant formula (3.46) we know that

$$W_2^2(\mu,\nu) = \frac{d}{|S^{(d-1)}|}\int_{S^{(d-1)}}\left(\int_{\mathbb{R}} W_2^2(\mu_{|\theta},\zeta_{|\theta}^{(\theta)})d\mu_\theta\right)d\theta$$

then, by taking $\lambda = \mu_\theta$, $\mu = \mu_{|\theta}^{(\theta)}$ and $\nu = \zeta_{|\theta}^{(\theta)}$, we can conclude

$$\begin{aligned} W_2^2(\mu,\nu) &= \frac{d}{|S^{(d-1)}|}\int_{S^{(d-1)}}\int_{\mathbb{R}^{d-1}} W_2^2(\mu_{|\theta},\zeta_{|\theta}^{(\theta)})d\mu_\theta d\theta \\ &\geq \frac{d}{|S^{(d-1)}|}\int_{S^{(d-1)}} W_2^2\left(\int_{\mathbb{R}^{d-1}}\mu_{|\theta}^{(\theta)}d\mu_\theta, \int_{\mathbb{R}^{d-1}}\zeta_{|\theta}^{(\theta)}d\mu_\theta\right)d\theta \\ &= \frac{d}{|S^{(d-1)}|}\int_{S^{(d-1)}} W_2^2(\mu_\theta,\nu_\theta)d\theta \qquad (3.51) \\ &= dSW_2^2(\mu,\nu) \end{aligned}$$

where the equality (3.51) comes from the fact that each $\zeta^{(\theta)}$ is an intermediate measure and so

$$\int_{\mathbb{R}^{d-1}}\zeta_{|\theta}^{(\theta)}d\mu_\theta = (\mathfrak{p}_\theta)_\#\zeta = (\mathfrak{p}_\theta)_\#\nu = \nu_\theta.$$

□

3.7 Formulation on the (d+1)-partite graph

As we saw in Section 2.3, we can describe grey-scales images as a probability measure on the regular bi-dimensional grids. In this setting, we can describe the measure as a stochastical matrix and the optimal transportation problem

can be formulated as a minimal uncapacitated flow problem on a bipartite graph. This setting can be easily generalized to higher dimensions, even if the resolution of the flow problem grows significantly in complexity as the dimension d grows.
Similarly, we can translate the cardinal flow problem as a flow problem on a suitable structure, the $(d+1)-$partite graph.

Definition 3.23 ($(d+1)-$partite graph)**.** *Let $G_N = (i_1, \dots, i_d), i_j = 1, \dots, N$ be the d dimensional regular grid. We define the $(d+1)-$partite graph as the couple (G, E), where*

$$G = G_0 \times G_1 \times \cdots \times G_N$$

and

$$E = \cup_{i=1}^{N} E_i$$

with

$$E_i := \{(x, y), x \in G_{i-1}, y \in G_i \quad \text{such that} \quad x_j = y_j, \quad \forall i \neq j\}.$$

Given μ and ν probability measures on G_N, we can then define the uncapacitated flow problem on the $(d+1)-$partite graph as

$$\mathcal{U} := \bigcup_{i=1}^{N} \mathcal{U}_i$$

and $\mathcal{U}_i = \bigcup_{i=1}^{N} \{u_{e_i}\}_{e_i \in E_i}$, satisfying the following conditions

$$\mu_x = \sum_{e_1 \in E_1, x = (e_1)_1} u_{e_1},$$

$$\nu_y = \sum_{e_d \in E_d, y = (e_N)_2} u_{e_N}$$

and

$$\sum_{e_i \in E_i, \alpha = (e_i)_2} u_{e_i} = \sum_{e_{i+1} \in E_{i+1}, \alpha = (e_{i+1})_1} u_{e_{i+1}}.$$

For the sake of clarity, but without loss of generality, we present first our construction considering 2-dimensional histograms and the ℓ_2 Euclidean ground distance.

Let us consider the following flow problem: let $\mu = \{\mu_{i,j}\}_{i,j}$ and $\nu = \{\nu_{i,j}\}_{i,j}$ be two probability measures over a $N \times N$ regular grid denoted by

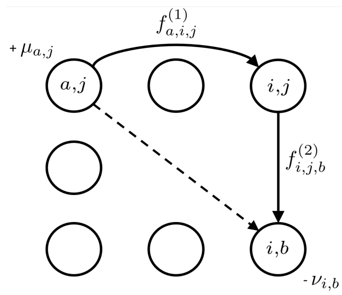


Figure 3.3: In order to send a unit of flow from point (a, j) to point (i, b), we either send a unit of flow directly along arc $((a, j), (i, b))$ of cost $c((a, j), (i, b)) = (a - i)^2 + (j - b)^2$, or, we first send a unit of flow from (a, j) to (i, j), and then from (i, j) to (i, b), having total cost $c((a, j), (i, j)) + c((i, j), (i, b)) = (a - i)^2 + (j - j)^2 + (i - i)^2 + (j - b)^2 = (a - i)^2 + (j - b)^2 = c((a, j), (i, b))$. Indeed, the cost of the two different path is exactly the same.

G. In the following paragraphs, we use the notation sketched in Figure 3.3. In addition, we define the set $U := \{1, \ldots, N\}$.

Since we are considering the ℓ_2 norm as ground distance, we minimize the functional

$$R : (F_1, F_2) \to \sum_{i,j=1}^{N} \left[\sum_{a=1}^{N} (a - i)^2 f^{(1)}_{a,i,j} + \sum_{b=1}^{N} (j - b)^2 f^{(2)}_{i,j,b} \right] \tag{3.52}$$

among all $F_i = \{f^{(i)}_{a,b,c}\}$, with $a, b, c \in \{1, ..., N\}$ real numbers (i.e., flow variables) satisfying the following constraints

$$\sum_{i=1}^{N} f^{(1)}_{a,i,j} = \mu_{a,j}, \qquad \forall a, j \in U \times U \tag{3.53}$$

$$\sum_{j=1}^{N} f^{(2)}_{i,j,b} = \nu_{i,b}, \qquad \forall i, b \in U \times U \tag{3.54}$$

$$\sum_{a} f^{(1)}_{a,i,j} = \sum_{b} f^{(2)}_{i,j,b}, \qquad \forall i, j \in U \times U, a \in U, b \in U. \tag{3.55}$$

Constraints (3.53) impose that the mass $\mu_{a,j}$ at the point (a, j) is moved to the points $(k, j)_{k=1,...,N}$. Constraints (3.54) force the point (i, b) to receive from the points $(i, l)_{l=1,...,N}$ a total mass of $\nu_{i,b}$. Constraints (3.55) require that all the mass that goes from the points $(a, j)_{a=1,\dots,N}$ to the point (i, j) is moved to the points $(i, b)_{b=1,\dots,N}$. We call a pair (F_1, F_2) satisfying the constraints (3.53)–(3.55) a *feasible flow* between μ and ν. We denote by $\mathcal{F}(\mu, \nu)$ the set of all feasible flows between μ and ν.

Indeed, we can formulate the minimization problem defined by (3.52)–(3.55) as an uncapacitated minimum cost flow problem on a tripartite graph $G = (V, A)$. The set of nodes of G is $V := G^{(1)} \cup G^{(2)} \cup G^{(3)}$, where $G^{(1)}, G^{(2)}$ and $G^{(3)}$ are the nodes corresponding to three $N \times N$ regular grids. We denote by $(i, j)^{(l)}$ the node of coordinates (i, j) in the grid $G^{(l)}$. We define the two disjoint set of arcs between the successive pairs of node partitions as

$$\begin{aligned} A^{(1)} &:= \{((a, j)^{(1)}, (i, j)^{(2)}) \mid i, a, j \in U\}, && (3.56) \\ A^{(2)} &:= \{((i, j)^{(2)}, (i, b)^{(3)}) \mid i, b, j \in U\}, && (3.57) \end{aligned}$$

and, hence, the arcs of G are $A := A^{(1)} \cup A^{(2)}$. Note that in this case the graph G has $3N^2$ nodes and $2N^3$ arcs. Whenever (F_1, F_2) is a feasible flow between μ and ν, we can think of the values $f^{(1)}_{a,i,j}$ as the quantity of mass that travels from (a, j) to (i, j) or, equivalently, that moves along the arc $((a, j), (i, j))$ of the tripartite graph, while the values $f^{(2)}_{i,j,b}$ are the mass moving along the arc $((i, j), (i, b))$ (e.g., see Figures 3.4– and 3.3).

Now we can give an idea of the roles of the grids $G^{(1)}$, $G^{(2)}$ and $G^{(3)}$: $G^{(1)}$ is the grid where is drawn the initial distribution μ, while on $G^{(3)}$ it is drawn the final configuration of the mass ν. The grid $G^{(2)}$ is an auxiliary grid that hosts a intermediate configuration between μ and ν. From a strict mathematical point of view, this configuration drawn on $G^{(2)}$ is a measure that lies on a geodetic connecting the measures μ and ν.

We are now ready to state our main contribution.

Theorem 3.17. *For each measure π on $G \times G$ that transports μ into ν, we can find a feasible flow (F_1, F_2) such that*

$$R(F_1, F_2) = \sum_{((a,j),(i,b))} ((a - i)^2 + (b - j)^2)\pi((a, j), (i, b)). \tag{3.58}$$

Indeed, we can compute the Kantorovich-Wasserstein distance of order 2 between a pair of discrete measures μ, ν, by solving an uncapacitated minimum cost flow problem on the given tripartite graph $G := (G^{(1)} \cup G^{(2)} \cup G^{(3)}, A^{(1)} \cup A^{(2)})$.

We remark that our approach is very general and it can be directly extended to deal with the following generalizations:

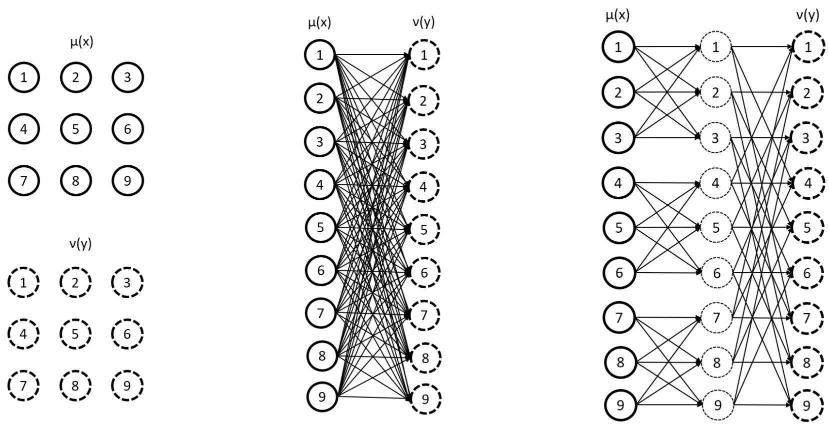


Figure 3.4: Two given 2-dimensional histograms of size $N \times N$, with $N = 3$ (left). The complete bipartite graph with N^4 arcs (center). The 3-partite graph with $2N^3$ arcs (right).

Higher dimensional grids. Our approach can handle discrete measures in spaces of any dimension d, that is, for instance, any d-dimensional histogram. In dimension $d = 2$, we get a tripartite graph because we decomposed the transport along the two main directions. If we have a problem in dimension d, we need a $(d+1)$-plet of grids connected by arcs oriented as the d fundamental directions, yielding a $(d + 1)$-partite graph. As the dimension d grows, our approach gets faster and more memory efficient than the standard formulation given on a bipartite graph.

3.8 Computational Results

In this section, we report the results obtained on two different set of instances. The goal of our experiments is to show how our approach scales with the size of the histogram N and with the dimension of the histogram d. As cost distance $c(x, y)$, with $x, y \in \mathbb{R}^d$, we use the squared ℓ_2 norm. As problem instances, we use the gray scale images (i.e., 2-dimensional histograms) proposed by the DOTMark benchmark [80], and a set of d-dimensional histograms obtained by biomedical data measured by flow cytometer [13].

Implementation details. We run our experiments using the Network Simplex as implemented in the Lemon C++ graph library since it provides the fastest implementation of the Network Simplex algorithm to solve uncapaci-

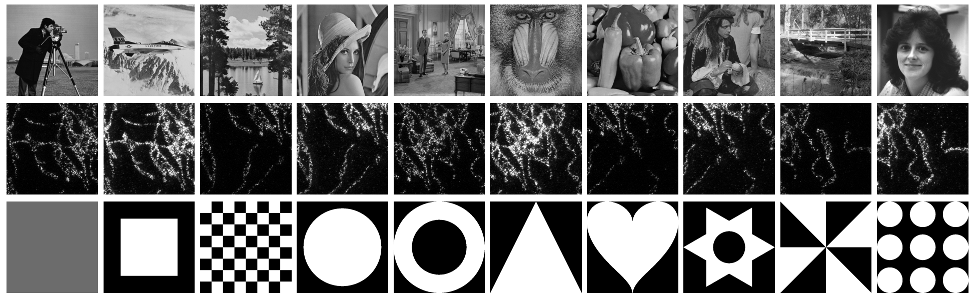


Figure 3.5: DOTmark benchmark: Classic, Microscopy, and Shapes images.

tated minimum cost flow problems [49]. The tests are executed on a gaming laptop with Windows 10 (64 bit), equipped with an Intel i7-6700HQ CPU and 16 GB of Ram. The code was compiled with MS Visual Studio 2017, using the ANSI standard C++17. The code execution is single threaded. The Matlab implementation of the Sinkhorn's algorithm [27] runs in parallel on the CPU cores, but we do not use any GPU in our test. Our C++ code is freely available at

Results for the DOTmark benchmark. The DOTmark benchmark contains 10 classes of gray scale images related to randomly generated images, classical images, and real data from microscopy images of mitochondria [80]. In each class there are 10 different images. Every image is given in the data set at the following pixel resolutions: 32×32, 64×64, 128×128, 256×256, and 512×512. The images in Figure 3.5 are respectively the *ClassicImages*, *Microscopy*, and *Shapes* images (one class for each row), shown at highest resolution, that are the three classes for which we report our results. For the lack of space, we do not report here the results for the other seven classes included in the DOTmark benchmark, but the results are essentially the same.

In our test, we first compared four approaches to compute the Kantorovich-Wasserstein distances on images of size 32×32:

1. **EMD**: The implementation of Transportation Simplex provided by [73], known in the literature as EMD code, that is an exact general method to solve optimal transport problem.

2. **Sinkhorn**: The Matlab implementation of the Sinkhorn's algorithm[1] [27], that is an approximate approach whose performance in terms of

[1] `http://marcocuturi.net/SI.html` (last visited on May, 18th, 2018)

speed and numerical accuracy depends on a parameter λ: for smaller values of λ, the algorithm is faster but the solution value has a large gap with respect to the optimal value of the transportation problem; for larger values of λ, the algorithm is more accurate (i.e., smaller gap), but it becomes slower. Unfortunately, for very large value of λ the method becomes numerically unstable. The best value of λ is problem dependent, and for our test we used $\lambda = 1$ and $\lambda = 1.5$. The second one, is the largest value we found for which the algorithm computes the distances for all the instances considered without facing numerical issues.

3. **Bipartite**: The bipartite formulation presented in Figure 3.4, which is the same as [73], but it is solved with the Network Simplex implemented in the Lemon Graph library [49].

4. **3-partite**: The 3-partite formulation proposed in this thesis, which for 2-dimensional histograms is represented in Figure 3.4. Again, we use the Network Simplex of the Lemon Graph Library to solve the corresponding uncapacitated minimum cost flow problem.

Table 3.1 reports the averages of our results over the three classes of images shown in Table 3.5. Each class contains 10 instances, and we compute the distance between every possible pair of images within the same class: this corresponds to have 45 instances for each class. We report the means and the standard deviations (between brackets) of the runtime, measured in seconds. Table 3.1 shows in the second column the runtime for EMD [73]. The third and fourth columns gives the runtime and the optimality gap for the Sinkhorn's algorithm with $\lambda = 1$; the 6-*th* and 7-*th* columns for $\lambda = 1.5$. The percentage gap is computed as Gap $= \frac{UB - opt}{opt} \cdot 100$, where UB is the upper bound computed by the Sinkhorn's algorithm, and *opt* is the optimal value computed by EMD. The last two columns report the runtime for the bipartite and 3-partite approaches.

As Table 3.1 shows, the 3-partite approach is clearly faster than any of the three alternatives considered here, despite being an exact method. In addition, we remark that, even on the bipartite formulation, the Network Simplex implementation of the Lemon Graph library is order of magnitude faster than EMD, and hence it should be the best choice in this particular type of instances.

Table 3.2 reports the results for the bipartite and 3-partite approaches for increasing size of the 2-dimensional histograms. The table report for each of the two approaches, the number of vertices $|V|$ and of arcs $|A|$, and the

Table 3.1: Comparison of different approaches on 32×32 images. The runtime (in seconds) is given as "Mean (StdDev)". The gap to the optimal value *opt* is computed as $\frac{UB-opt}{opt} \cdot 100$, where UB is the upper bound computed by Sinkhorn's algorithm. Each row reports the averages over 45 instances.

	EMD[73]	Sinkhorn [27]				3-partite
		$\lambda = 1$		$\lambda = 1.5$		
Image	Runtime	Runtime	Gap	Runtime	Gap	Runtime
Classic	24.0 (3.3)	6.0 (0.5)	17.3%	8.9 (0.7)	9.1%	0.07 (0.01)
Micr.	35.0 (3.3)	3.5 (1.0)	2.4%	5.3 (1.4)	1.2%	0.08 (0.01)
Shapes	25.2 (5.3)	1.6 (1.1)	5.6%	2.5 (1.6)	3.0%	0.05 (0.01)

Table 3.2: Comparison of the bipartite and the 3-partite approaches on 2-dimensional histograms.

		Bipartite			3-partite		
Size	Image	$\|V\|$	$\|A\|$	Runtime	$\|V\|$	$\|A\|$	Runtime
64×64	Classic	8 193	16 777 216	16.3 (3.6)	12 288	524 288	2.2 (0.2)
	Micr.			11.7 (1.4)			1.0 (0.2)
	Shape			13.0 (3.9)			1.1 (0.3)
128×128	Classic	32 768	268 435 456	1 368 (545)	49 152	4 194 304	36.2 (5.4)
	Micr.			959 (181)			23.0 (4.8)
	Shape			983 (230)			17.8 (5.2)

means and standard deviations of the runtime. As before, each row gives the averages over 45 instances. Table 3.2 shows that the 3-partite approach is clearly better in terms of memory, the 3-partite graph has a fraction of the number of arcs (which has to be stored in memory), and runtime, since it is at least an order of magnitude faster. Indeed, the 3-partite formulation is better essentially because it exploits the structure of the ground distance $c(x, y)$ used, that is, the squared ℓ_2 norm.

Flow Cytometry biomedical data. Flow cytometry is a laser-based biophysical technology used to study human health disorders. Flow cytometry experiments produce huge set of data, which are very hard to analyze with standard statistics methods and algorithms [13]. Currently, such data is used two study the correlations of only two factors (e.g., biomarkers) at the time, by visualizing 2-dimensional histograms and by measuring the (dis-)similarity between pairs of histograms [67]. However, during a flow cytometry experiment up to hundreds of factors (biomarkers) are measured and stored in digital format. Hence, we can use such data to build d-dimensional histograms that consider up to d biomarkers at the time, and then comparing the similarity among different individuals by measuring the distance between the corresponding histograms. In this work, we have used the flow cytometry data related to *Acute Myeloid Leukemia (AML)*, available at `http://flowrepository.org/id/FR-FCM-ZZYA`, which contains cytometry data for 359 patients, classified as "normal" or affected by AML. This dataset has been used by the bioinformatics community to run clustering algorithms, which should predict whether a new patient is affected by AML [1].

Table 3.3 reports the results of computing the distance between pairs of d-dimensional histograms, with d ranging in the set $\{2, 3, 4\}$, obtained using the AML biomedical data. For simplicity, we considered regular histograms of size $n = N^d$ (i.e., n is the total number of bins), using $N = 16$ and $N = 32$. Table 3.3 compares the results obtained by the bipartite and $(d+1)$-partite approach, in terms of graph size and runtime. Again, the $(d+1)$-partite approach, by exploiting the structure of the ground distance, outperforms the standard formulation of the optimal transport problem. We remark that for $N = 32$ and $d = 3$, we pass from going out-of-memory with the bipartite formulation, to compute the distance in around 5 seconds with the 4-partite formulation.

Table 3.3: Comparison between the bipartite and the $(d+1)$-partite approaches on Flow Cytometry data.

		Bipartite Graph			$(d+1)$-partite Graph		
N	d	$\lvert V\rvert$	$\lvert A\rvert$	Runtime	$\lvert V\rvert$	$\lvert A\rvert$	Runtime
16	2	512	65 536	0.024 (0.01)	768	8 192	0.003 (0.00)
	3	8 192	16 777 216	38.2 (14.0)	16 384	196 608	0.12 (0.02)
	4	*out-of-memory*			327 680	4 194 304	4.8 (0.84)
32	2	2 048	1 048 756	0.71 (0.14)	3072	65 536	0.04 (0.01)
	3	*out-of-memory*			131 072	3 145 728	5.23 (0.69)

Chapter 4

Truncated Distances

As we stressed through all the last chapter, exploiting the structure of the cost function can result in a simplification of the classical bipartite graph to solve. In this chapter, we study how to use the saturation of a cost function.
Roughly speaking, a cost function is saturated if it assumes its maximum value on a wide set. We show that it is possible to ignore all the arcs that connect couples in the saturated zone, thus reducing the dimension of the problem and, as a consequence, also its complexity.
We focus our attention on a particular subclass of cost functions: the truncated ones. Given a cost function c and a suitable parameter t, the $t-$truncated cost function $c^{(t)}$ is obtained by choosing, for every pair $(x, y) \in X \times Y$, the lower value between $c_{x,y}$ and t. The smaller t is, the wider the saturated zone is.
This problem has been introduced by Ofir and Pele in their work [69]. We improve their results by proposing a new and equivalent formulation. Through this formulation, we are be able to give an upper bound on both the absolute and the relative error we commit by taking the truncated cost function over the original one.
We use these bounds as stopping criteria in an iterative algorithm that allow us to compute the Wasserstein distance between measures. The key idea behind the algorithm is to iteratively increase the threshold of t until a certain accuracy is reached. The bounds allow us to estimate this accuracy.
We also study what happens if the cost function we truncate is separable. By introducing a new truncation method that preserves the separability, we define another problem. The number of arcs of this problem is smaller of the number of arcs required by the classical truncated problem.
In this chapter we assume that μ and ν are supported over a finite set

of points X and Y, respectively. According to Example 2.6, we identify the measure with the collection of values that defines it, so that we write equivalently μ or μ_x (where $x \in X$). The same holds true for any generic cost function $c : X \times Y \to \mathbb{R}$, that is uniquely identified by the collection of values $c_{x,y} := c(x, y)$. A partial and preliminary version of this chapter appeared in [6].

4.1 The maximum Nearby Flow problem

In this section, we introduce and study the maximization of the nearby flow functional. This problem can be thought of as a variant of the classical transportation one that takes care only of movements between points that are considered near. This problem has a natural interpretation in the study of the transportation cost with truncated cost functions.

Definition 4.1 (Nearby Point sets). *Given a cost function c and a fixed parameter $t > 0$, we define the set of near points*

$$N_c^{(t)} := \big\{(x, y) \in X \times Y : c_{x,y} < t\big\}.$$

Similarly, we define the set of points near to x and y, respectively, as

$$\begin{aligned} O_c^{(t)}(x) &:= \big\{y \in Y : c_{x,y} < t\big\}, \\ I_c^{(t)}(y) &:= \big\{x \in X : c_{x,y} < t\big\}. \end{aligned}$$

In our formulation, the pairs of points in $N_c^{(t)}$ are the extremes of the arcs that compose our bipartite graph. If $t < \max_{(x,y)\in X\times Y} c_{x,y}$ we have $N_c^{(t)} \subsetneq X \times Y$, therefore a feasible transportation plan between two given measures might not exist on this graph. For this reason, we introduce a new mathematical tool, the nearby flow.

Definition 4.2 (Nearby Flow). *Let us take $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We say that a subprobability measure $\eta \in \mathcal{M}(N_c^{(t)})$ is a* ***nearby flow*** *if*

$$\sum_{x\in I_c^{(t)}(y)} \eta_{x,y} \leq \nu_y, \quad \forall y \in Y, \tag{4.1}$$

$$\sum_{y\in O_c^{(t)}(x)} \eta_{x,y} \leq \mu_x, \quad \forall x \in X. \tag{4.2}$$

We denote by $\mathcal{N}^{(t)}(\mu, \nu)$ the set of all nearby flows between μ and ν.

Remark 4.1. *Unlike what happened for the transportation plans, the definition of nearby flow does not require μ and ν to have the same mass. We can then define the set $\mathcal{N}^{(t)}(\mu, \nu)$ for any pair of positive measures μ and ν. This will come in handy in subsection 4.4.1, where we will show the relation between this problem and the computation of projections. However, for now, we assume that μ and ν have unitary mass.*

Definition 4.3 (Nearby Flow Functional)**.** *The* nearby flow transport functional $\mathbb{B}_c^{(t)}$ *is defined as*

$$\mathbb{B}_c^{(t)}(\eta) := \sum_{(x,y)\in N_c^{(t)}} s_{x,y}\eta_{x,y}$$

where $s_{x,y} := t - c_{x,y}$.

Remark 4.2. *The nearby flow describes the movements between points whose distance is less than the threshold t. Since moving the mass through more expensive arcs increases the overall transportation cost, it is reasonable to suppose that the optimal transportation plan should encourage the movements between nearby points.*
The next theorem states that searching for the minimal transportation plan for the truncated cost is equivalent to searching for the maximal nearby flow.

Theorem 4.1. *Given a threshold value $t > 0$, a cost function c on $X \times Y$, two discrete probability measures μ and ν defined over X and Y, respectively, the following relation holds:*

$$W_{c^{(t)}}(\mu, \nu) := \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi) = t - \max_{\eta\in\mathcal{N}^{(t)}(\mu,\nu)} \mathbb{B}_c^{(t)}(\eta), \tag{4.3}$$

where $c_{x,y}^{(t)} := \min\{c_{x,y}, t\}$ is the truncated cost.

Remark 4.3. *Given any couple of measures μ, ν, and a positive threshold t, the set $\mathcal{N}^{(t)}(\mu, \nu)$ is not empty since the null flow, defined as $\eta_{x,y} = 0$ for all $(x, y) \in N_c^{(t)}$, is in this set. Moreover, it is easy to check that $N_c^{(t)}$ is also convex and, since we are dealing with discrete measures, closed, therefore compact. Hence the minimum and maximum defined in* (4.3) *are well defined.*

Proof. We first show how, given a flow π, we can get a feasible nearby flow η, and, later, we show the opposite, that is, how to get a feasible flow π given a nearby flow η.
Let us start with a given $\pi \in \Pi(\mu, \nu)$. Then, we can define

$$\eta_{x,y} := \pi_{x,y} \qquad \forall (x, y) \in N_c^{(t)},$$

for which we can easily check that

$$\begin{aligned}\sum_{x\in I_c^{(t)}(y)} \eta_{x,y} &= \sum_{x\in I_c^{(t)}(y)} \pi_{x,y}\\ &\leq \sum_{x\in X} \pi_{x,y} = \nu_y, \quad \forall y \in Y,\end{aligned}$$

and

$$\begin{aligned}\sum_{y\in O_c^{(t)}(x)} \eta_{x,y} &= \sum_{y\in O_c^{(t)}(x)} \pi_{x,y}\\ &\leq \sum_{y\in Y} \pi_{x,y} = \mu_x, \quad \forall x \in X.\end{aligned}$$

Hence, we have that $\eta \in \mathcal{N}^{(t)}(\mu,\nu)$.
By a simple computation, we get

$$\begin{aligned}&\sum_{(x,y)\in X\times Y} c_{x,y}^{(t)}\pi_{x,y}\\ &\quad= \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\pi_{x,y} + t \sum_{(x,y)\in (X\times Y)\setminus N_c^{(t)}} \pi_{x,y}\\ &\quad= \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\pi_{x,y} + t \left(\sum_{(x,y)\in X\times Y} \pi_{x,y} - \sum_{(x,y)\in N_c^{(t)}} \pi_{x,y} \right)\\ &\quad= \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\pi_{x,y} + t \left(1 - \sum_{(x,y)\in N_c^{(t)}} \pi_{x,y} \right) \qquad (4.4)\\ &\quad= t - \sum_{(x,y)\in N_c^{(t)}} \left(t - c_{x,y}^{(t)} \right) \pi_{x,y}\\ &\quad= t - \mathbb{B}_c^{(t)}(\eta).\end{aligned}$$

In (4.4), since π is a transportation plan, we have

$$\sum_{(x,y)\in X\times Y} \pi_{x,y} = 1.$$

We have shown that for every $\pi \in \Pi(\mu,\nu)$ we can define a nearby flow $\eta \in \mathcal{N}^{(t)}(\mu,\nu)$ such that

$$\mathbb{T}_{c^{(t)}}(\pi) = t - \mathbb{B}_c^{(t)}(\eta).$$

However, it could exist a nearby flow η with a larger value of $\mathbb{B}_c^{(t)}(\eta)$, and, hence, so far we have only proved that

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi) \geq t - \max_{\eta\in\mathcal{N}^{(t)}(\mu,\nu)} \mathbb{B}_c^{(t)}(\eta). \tag{4.5}$$

Now we want to show that the previous relation holds with equality. We need to show that, given a nearby flow $\eta \in \mathcal{N}^{(t)}(\mu,\nu)$, we can get a feasible π. We start by introducing the slack variables corresponding to constraints (4.1) and (4.2):

$$\tilde{\mu}_x := \mu_x - \sum_{y\in O_c^{(t)}(x)} \eta_{x,y}, \quad \forall x \in X, \tag{4.6}$$

$$\tilde{\nu}_y := \nu_y - \sum_{x\in I_c^{(t)}(y)} \eta_{x,y}, \quad \forall y \in Y. \tag{4.7}$$

We have $\tilde{\mu}_x \geq 0$ and $\tilde{\nu}_y \geq 0$ for each $x \in X$ and $y \in Y$. Since μ and ν are probability measures, we have that

$$\begin{aligned}
\sum_{x\in X} \tilde{\mu}_x &= \sum_{x\in X} \left(\mu_x - \sum_{y\in O_c^{(t)}(x)} \eta_{x,y} \right) \\
&= 1 - \sum_{x\in X} \sum_{y\in O_c^{(t)}(x)} \eta_{x,y} \\
&= \sum_{y\in Y} \nu_y - \sum_{y\in Y} \sum_{x\in I_c^{(t)}(y)} \eta_{x,y} \\
&= \sum_{y\in Y} \tilde{\nu}_y.
\end{aligned}$$

We introduce the value $M = \sum_{y\in Y} \tilde{\nu}_y = \sum_{x\in X} \tilde{\mu}_x$, which is used to define

$$\pi_{x,y} := \begin{cases} \eta_{x,y} + \dfrac{\tilde{\mu}_x\tilde{\nu}_y}{M} & \text{if } (x,y) \in N_c^{(t)}, \\ \dfrac{\tilde{\mu}_x\tilde{\nu}_y}{M} & \text{otherwise.} \end{cases} \tag{4.8}$$

We now show that $\pi \in \Pi(\mu,\nu)$. For all $y \in Y$ we have

$$\sum_{x\in X} \pi_{x,y} = \sum_{x\in X\setminus I_c^{(t)}(y)} \pi_{x,y} + \sum_{x\in I_c^{(t)}(y)} \pi_{x,y}$$

$$
\begin{aligned}
&= \sum_{x\in X\setminus I_c^{(t)}(y)} \frac{\tilde{\mu}_x\tilde{\nu}_y}{M} + \sum_{x\in I_c^{(t)}(y)} \eta_{x,y} + \sum_{x\in I_c^{(t)}(y)} \frac{\tilde{\mu}_x\tilde{\nu}_y}{M} \\
&= \sum_{x\in X} \frac{\tilde{\mu}_x\tilde{\nu}_y}{M} + \sum_{x\in I_c^{(t)}(y)} \eta_{x,y} \\
&= \tilde{\nu}_y + \sum_{x\in I_c^{(t)}(y)} \eta_{x,y} \\
&= \nu_y.
\end{aligned}
$$

Similarly, we can show that

$$\sum_{y\in Y} \pi_{x,y} = \mu_x,$$

and, hence, the measure π defined in (4.8) belongs to $\Pi(\mu,\nu)$. Regarding its cost, we have

$$
\begin{aligned}
\sum_{(x,y)\in X\times Y} c_{x,y}^{(t)}\pi_{x,y} &= \sum_{(x,y)\in (X\times Y)\setminus N_c^{(t)}} c_{x,y}^{(t)}\pi_{x,y} + \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\pi_{x,y} \\
&= t \sum_{(x,y)\in (X\times Y)\setminus N_c^{(t)}} \frac{\tilde{\mu}_x\tilde{\nu}_y}{M} + \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\eta_{x,y} \\
&\quad + \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\frac{\tilde{\mu}_x\tilde{\nu}_y}{M} \\
&\leq t \sum_{(x,y)\in X\times Y} \frac{\tilde{\mu}_x\tilde{\nu}_y}{M} + \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\eta_{x,y} \\
&= tM + \sum_{(x,y)\in N_c^{(t)}} c_{x,y}^{(t)}\eta_{x,y}.
\end{aligned}
$$

By definition, the constant $M = \sum_{y\in Y} \tilde{\nu}_y$ can be rewritten as

$$
\begin{aligned}
M &= \sum_{y\in Y} \left(\nu_y - \sum_{x\in I_c^{(t)}(y)} \eta_{x,y} \right) \\
&= \sum_{y\in Y} \nu_y - \sum_{(x,y)\in N_c^{(t)}} \eta_{x,y} \\
&= 1 - \sum_{(x,y)\in N_c^{(t)}} \eta_{x,y}.
\end{aligned}
$$

Since $s_{x,y} = t - c_{x,y}$, we can write

$$\begin{aligned}\sum_{(x,y)\in X\times Y} c^{(t)}_{x,y}\pi_{x,y} &\leq tM + t\sum_{(x,y)\in N^{(t)}_c}\eta_{x,y} - \sum_{(x,y)\in N^{(t)}_c} s_{x,y}\eta_{x,y}\\ &= tM + t(1-M) - \sum_{(x,y)\in N^{(t)}_c} s_{x,y}\eta_{x,y}\\ &= t - \sum_{(x,y)\in N^{(t)}_c} s_{x,y}\eta_{x,y}.\end{aligned}$$

Thus, we showed that for each nearby flow $\eta \in \mathcal{N}^{(t)}(\mu,\nu)$ there exists a $\pi \in \Pi(\mu,\nu)$ such that

$$\mathbb{T}_{c^{(t)}}(\pi) \leq t - \mathbb{B}^{(t)}_c(\eta),$$

and hence

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi) \leq t - \max_{\eta\in\mathcal{N}^{(t)}(\mu,\nu)} \mathbb{B}^{(t)}_c(\eta),$$

which together with the inequality (4.5) completes the proof. □

Remark 4.4. *In the proof of the previous Theorem we showed that, given two measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and any $\pi \in \Pi(\mu,\nu)$, the restriction of π on $N^{(t)}_c$ is a nearby flow between μ and ν. The reverse implication is not always true: given a nearby flow η, there might not exist a transportation plan whose restriction is equal to η.*
For instance, consider the case $X = Y$ and $\mu = \nu = \delta_x$ where $x \in X$ is a given point. The set $\Pi(\mu,\nu)$ is then composed of only one element, $\pi = \delta_{x,x}$. As we noticed in Remark 4.3, the null flow is always a nearby flow between any pair of measures, but, in this case, it is not the restriction of π and, therefore, of any element of $\Pi(\mu,\nu)$.
However, as we did in the proof of the previous Theorem, given a flow η, it is always possible to retrieve a transportation plan $\pi \in \Pi(\mu,\nu)$ such that

$$\eta_{x,y} \leq \pi_{x,y}$$

for any $(x,y) \in N^{(t)}_c$, whatever the value of t is.

Definition 4.4 (Maximal Nearby Flows)**.** *Let us take a cost function c and $t \in \mathbb{R}$. Given any couple of measures μ and ν, we denote with $\Theta^{(t)}(\mu,\nu)$ the set of all the maximal flows between them, according to the functional $\mathbb{B}^{(t)}_c$, i.e. $\eta^* \in \Theta^{(t)}(\mu,\nu)$ if*

$$\mathbb{B}^{(t)}_c(\eta^*) \geq \mathbb{B}^{(t)}_c(\eta)$$

for every $\eta \in \mathcal{N}^{(t)}(\mu,\nu)$.

Corollary 4.1. *Let us take $\eta^* \in \Theta^{(t)}(\mu,\nu)$. Then, for any pair $(\bar{x},\bar{y}) \in N_c^{(t)}$ at least one of the following conditions holds*

$$\sum_{x \in I_c^{(t)}(\bar{y})} \eta^*_{x,\bar{y}} = \nu_{\bar{y}} \quad \textit{or} \quad \sum_{y \in O_c^{(t)}(\bar{x})} \eta^*_{\bar{x},y} = \mu_{\bar{x}}. \tag{4.9}$$

Proof. We argue by contradiction: let η^* be a maximal nearby flow and let $(\bar{x},\bar{y}) \in N_c^{(t)}$ be a pair of points for which both relations (4.9) are false, i.e.

$$\sum_{x \in I_c^{(t)}(\bar{y})} \eta^*_{x,\bar{y}} < \nu_{\bar{y}}$$

and

$$\sum_{y \in O_c^{(t)}(\bar{x})} \eta^*_{\bar{x},y} < \mu_{\bar{x}}.$$

Then we can find $\epsilon > 0$ such that

$$\sum_{x \in I_c^{(t)}(\bar{y})} \eta^*_{x,\bar{y}} + \epsilon < \nu_{\bar{y}}$$

and

$$\sum_{y \in O_c^{(t)}(\bar{x})} \eta^*_{\bar{x},y} + \epsilon < \mu_{\bar{x}}.$$

If we define $\tilde{\eta}$ as

$$\tilde{\eta}_{x,y} := \begin{cases} \eta^*_{x,y} + \epsilon & \text{if } (x,y) = (\bar{x},\bar{y}), \\ \eta^*_{x,y} & \text{otherwise,} \end{cases}$$

we have that $\tilde{\eta} \in \mathcal{N}^{(t)}(\mu,\nu)$. In fact, if $x \neq \bar{x}$, we have that

$$\sum_{y \in O_c^{(t)}(\bar{x})} \eta_{x,y} = \sum_{y \in O_c^{(t)}(\bar{x})} \eta^*_{x,y} \leq \mu_x,$$

while, if $x = \bar{x}$, we have

$$\sum_{y \in O_c^{(t)}(\bar{x})} \tilde{\eta}_{\bar{x},y} = \sum_{y \in O_c^{(t)}(\bar{x})} \eta^*_{\bar{x},y} + \epsilon \leq \mu_{\bar{x}}.$$

Similarly we can show that also

$$\sum_{x \in I_c^{(t)}(y)} \eta_{x,y} \leq \nu_y$$

for any $y \in Y$. To conclude notice that

$$\begin{aligned}\mathbb{B}_c^{(t)}(\eta) &= \sum_{(x,y)\in N_c^{(t)}} (t-c)\eta_{x,y} \\ &= \sum_{(x,y)\in N_c^{(t)}} (t-c)\eta^*_{x,y} + \epsilon(t - c_{\bar{x},\bar{y}}) \\ &> \sum_{(x,y)\in N_c^{(t)}} (t-c)\eta^*_{x,y},\end{aligned}$$

which contradicts the maximality of η^*. □

As we saw in the proof of Theorem 4.1, by restricting π to $N_c^{(t)}$, we find a nearby flow. Unfortunately, it is not true that the restriction of an optimal plan is still optimal for the nearby flow problem, as we can see from the following example.

Example 4.1. *Let $\mu, \nu \in \mathcal{P}([0,\ldots,10])$ be defined as*

$$\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_5$$

$$\nu = \frac{1}{2}\delta_6 + \frac{1}{2}\delta_{10}.$$

and let us set $c_{x,y} = |x-y|^2$. By Theorem 2.19, the optimal transportation plan π^* coincides with the monotone transportation plan, i.e.

$$\pi^* = \frac{1}{2}\delta_0 \otimes \delta_6 + \frac{1}{2}\delta_5 \otimes \delta_{10}.$$

However, if we fix the threshold $t = 2$, each $i \in [0,\ldots,10]$ is connected only to $i-1$ and $i+1$, hence the maximal nearby flow is

$$\eta_{x,y} = \begin{cases} \frac{1}{2} & \textit{if} \quad x = 5,\ y = 6 \\ \\ 0 & \textit{otherwise} \end{cases}$$

which is not the restriction of π^* to $N_c^{(t)}$.

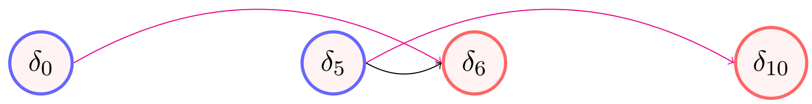


Figure 4.1: The transportation problem showcased in Example 4.1. The green line is the maximal nearby flow, the purple ones are the optimal transportation plan. We notice that their arcs are not in common, hence, we cannot see the maximal nearby flow as a restriction of an optimal transportation plan.

4.2 The Faraway Flow problem

As Corollary 2.1 states, since $c^{(t)} \leq c$ for any threshold t, the transportation cost related to $c^{(t)}$ gives a lower bound on the original transportation cost. The aim of this section is to use an auxiliary problem to deduce an upper bound on both the absolute error $|W_c - W_{c^{(t)}}|$ and the relative error $\dfrac{|W_c - W_{c^{(t)}}|}{|W_c|}$. In Theorem 4.2, we estimate these errors through a maximal nearby flow and the starting measures μ and ν.

Definition 4.5 (Faraway Cost Function)**.** *Given a positive parameter t, we define the faraway cost function $c^{[t]}$ as*

$$c^{[t]}_{x,y} = \max\left\{c_{x,y} - \ t, 0\right\},$$

for all $x \in X, y \in Y$.

The faraway cost function can be thought as the residual of the cost c after approximating it with $c^{(t)}$. In fact,

$$c_{x,y} = c^{(t)}_{x,y} + c^{[t]}_{x,y} \tag{4.10}$$

for all $x \in X$ and $y \in Y$. As for the nearby flow we take into account only a subset of all the possible arcs. In this case, we focus on the ones whose distance is greater than the given threshold t.

Definition 4.6 (Faraway Points)**.** *We define the set of faraway points as*

$$F^{[t]}_c := \Big\{(x, y) \in X \times Y : c_{x,y} \geq t\Big\}.$$

Similarly, we define the set of faraway points from x and y, respectively, as

$$\begin{aligned} U^{[t]}_c(x) &:= \left\{y \in Y : c_{x,y} \geq t\right\}, \\ E^{[t]}_c(y) &:= \left\{x \in X : c_{x,y} \geq t\right\}. \end{aligned}$$

Definition 4.7 (Faraway Flows). *Let us fix $t > 0$. Let μ and ν be two probability measures and $\eta^* \in \Theta^{(t)}(\mu, \nu)$. A sub-probability $\zeta \in \mathcal{M}(X \times Y)$ is a* faraway flow *from μ to ν if it satisfies the following conditions*

$$\sum_{x \in X} \zeta_{x,y} = \tilde{\nu}_y \quad = \quad \nu_y - \sum_{x \in I_c^{(t)}(y)} \eta^*_{x,y}, \tag{4.11}$$

$$\sum_{y \in Y} \zeta_{x,y} = \tilde{\mu}_x \quad = \quad \mu_x - \sum_{y \in O_c^{(t)}(x)} \eta^*_{x,y}. \tag{4.12}$$

Given $\eta^* \in \Theta^{(t)}(\mu, \nu)$, we denote by $\mathcal{Z}^{[t]}(\mu, \nu, \eta^*)$ the set of all faraway flows from μ to ν.

Definition 4.8 (Faraway Transport Functional). *We define the faraway transport functional $\mathbb{S}_c^{[t]} : \mathcal{Z}^{[t]}(\mu, \nu, \eta^*) \to \mathbb{R}$ as*

$$\mathbb{S}_c^{[t]}(\zeta) := \sum_{(x,y) \in X \times Y} c^{[t]}_{x,y}\, \zeta_{x,y}. \tag{4.13}$$

As it happened for the truncated cost function, the faraway flow takes care of the residual mass that was not moved through the nearby flow. This transportation allows us to define a transportation plan that extends the fixed maximal nearby flow η^*. In the following Lemma, we see that the support of each faraway flow is a subset of $F_c^{[t]}$.

Lemma 4.1. *Each faraway flow $\zeta \in \mathcal{Z}^{[t]}(\mu, \nu, \eta^*)$ satisfies*

$$\zeta_{x,y} = 0, \quad \forall (x, y) \in N_c^{(t)}. \tag{4.14}$$

Proof. Let $\eta^* \in \Theta^{(t)}(\mu, \nu)$ and $(\bar{x}, \bar{y}) \in N_c^{(t)}$ be fixed. By Corollary 4.1, we have

$$\mu_x - \sum_{y \in O_c^{(t)}(x)} \eta^*_{\bar{x},y} = 0$$

or

$$\nu_y - \sum_{x \in I_c^{(t)}(y)} \eta^*_{x,\bar{y}} = 0.$$

We can suppose, without loss of generality, that the first equality is satisfied, in this case

$$\sum_{y \in Y} \zeta_{\bar{x},y} = 0.$$

Since $\zeta_{x,y} \geq 0$ for each $x \in X, y \in Y$, we can deduce that $\zeta_{\bar{x},y} = 0$ for each $y \in Y$. Since Corollary 4.1 holds true for any pair $(x, y) \in N_c^{(t)}$, we deduce $\zeta = 0$ over $N_c^{(t)}$. $\square$

In Remark 4.4 we saw that not all the nearby flows can be seen as the restriction to $N_c^{(t)}$ of a $\pi \in \Pi(\mu, \nu)$. However, this is not the case if η^* is maximal for a certain t. Thanks to Lemma 4.1, we are indeed able to use the faraway flows to complete a maximal nearby flow.
Given $\eta^* \in \Theta^{(t)}(\mu, \nu)$ and a faraway flow ζ, we can define

$$\pi_{x,y} := \begin{cases} \eta^*_{x,y} & \text{if } (x,y) \in N_c^{(t)}, \\ \zeta_{x,y} & \text{otherwise.} \end{cases} \tag{4.15}$$

This π is a transportation plan between μ and ν. In fact, using (4.14), we have

$$\begin{aligned} \sum_{x \in X} \pi_{x,y} &= \sum_{x \in I_c^{(t)}(y)} \eta^*_{x,y} + \sum_{x \in E_c^{[t]}(y)} \zeta_{x,y} \\ &= \sum_{x \in I_c^{(t)}(y)} \eta^*_{x,y} + \sum_{x \in X} \zeta_{x,y} \\ &= \sum_{x \in I_c^{(t)}(y)} \eta^*_{x,y} + \nu_y - \sum_{x \in I_c^{(t)}(y)} \eta^*_{x,y} \\ &= \nu_y. \end{aligned}$$

Similarly, we can show that

$$\sum_{y \in Y} \pi_{x,y} = \mu_x. \tag{4.16}$$

Starting from relation (4.16), we can sum over x on both sides to obtain

$$\begin{aligned} \sum_{(x,y) \in X \times Y} \pi_{x,y} &= 1 \\ \sum_{(x,y) \in F_c^{[t]}} \zeta_{x,y} + \sum_{(x,y) \in N_c^{(t)}} \eta^*_{x,y} &= 1 \\ \sum_{(x,y) \in F_c^{[t]}} \zeta_{x,y} &= 1 - \sum_{(x,y) \in N_c^{(t)}} \eta^*_{x,y}. \end{aligned}$$

Let us now take an optimal transportation plan $\pi^*_{x,y}$ between μ and ν for the cost function c and let π be the transportation plan defined in (4.15). We have

$$\mathbb{T}_c(\pi^*) = \sum_{(x,y) \in X \times Y} c_{x,y} \pi^*_{x,y}$$

$$
\begin{aligned}
&\leq \sum_{(x,y)\in X\times Y} c_{x,y}\pi_{x,y} \\
&= \sum_{(x,y)\in X\times Y} \Big(\max\{c_{x,y}-t,0\} + \min\{c_{x,y},t\} \Big)\pi_{x,y} \\
&= \sum_{(x,y)\in F_c^{[t]}} c^{[t]}_{x,y}\zeta_{x,y} + \sum_{(x,y)\in X\times Y} \min\{c_{x,y},t\}\pi_{x,y} \\
&= \sum_{(x,y)\in F_c^{[t]}} c^{[t]}_{x,y}\zeta_{x,y} + t\sum_{(x,y)\in F_c^{[t]}} \zeta_{x,y} + \sum_{(x,y)\in N_c^{(t)}} c_{x,y}\eta^*_{x,y} \\
&= \sum_{(x,y)\in F_c^{[t]}} c^{[t]}_{x,y}\zeta_{x,y} + t\Big(1 - \sum_{(x,y)\in N_c^{(t)}} \eta^*_{x,y}\Big) \\
&\quad + \sum_{(x,y)\in N_c^{(t)}} c_{x,y}\eta^*_{x,y} \\
&= \sum_{(x,y)\in F_c^{[t]}} c^{[t]}_{x,y}\zeta_{x,y} + t - \sum_{(x,y)\in N_c^{(t)}} s_{x,y}\eta^*_{x,y},
\end{aligned}
$$

so that

$$
W_c(\mu,\nu) = \mathbb{T}_c(\pi^*) \leq \sum_{(x,y)\in F_c^{[t]}} c^{[t]}_{x,y}\zeta_{x,y} + t - \sum_{(x,y)\in N_c^{(t)}} s_{x,y}\eta^*_{x,y}.
$$

Since $\eta^* \in \Theta^{(t)}(\mu,\nu)$, for Theorem 4.1, we can write

$$
\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) \leq \sum_{(x,y)\in F_c^{[t]}} c^{[t]}_{x,y}\zeta_{x,y} + \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi),
$$

which leads to

$$
\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) - \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi) \leq \min_{\zeta\in\mathcal{Z}^{[t]}(\mu,\nu,\eta^*)} \mathbb{S}_c^{[t]}(\zeta). \tag{4.17}
$$

In particular, since

$$
\tilde{\zeta}_{x,y} := \frac{\Big(\mu_x - \sum_{y\in O_c^{(t)}(x)} \eta^*_{x,y}\Big)\Big(\nu_y - \sum_{x\in I_c^{(t)}(y)} \eta^*_{x,y}\Big)}{1 - \sum_{(x,y)\in X\times Y} \eta^*_{x,y}}. \tag{4.18}
$$

is a feasible faraway flow, we get the following estimation.

Theorem 4.2. *Let us take a threshold $t > 0$ and a cost function c. Given any two probability measures, μ and ν, let us take $\eta^* \in \Theta^{(t)}(\mu, \nu)$. Then, the following relation holds true*

$$\sum_{(x,y)\in X\times Y} c^{[t]}_{x,y}\tilde{\zeta}_{x,y} \geq \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) - \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi), \tag{4.19}$$

where $\tilde{\zeta}_{x,y}$ is the faraway flow defined in (4.18).

Moreover, since the the transportation cost related to the original cost function is always greater than the truncated transportation cost, we can also give an estimation of the relative error.

Corollary 4.2. *Given any $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $t > 0$, it holds*

$$\frac{|W_c(\mu,\nu) - W_{c^{(t)}}(\mu,\nu)|}{|W_c(\mu,\nu)|} \leq \frac{\sum_{x\in X, y\in Y} \tilde{\mu}_x\tilde{\nu}_y(t - c_{x,y})_+}{\left(1 - \sum_{(x,y)\in N^{(t)}_c} \eta_{x,y}\right) W_{c^{(t)}}(\mu,\nu)} \tag{4.20}$$

where $\tilde{\mu}$ and $\tilde{\nu}$ are the slack variables defined in (4.6) *and* (4.7).

Since the slack variables depend only on the maximal nearby flow and the known measure μ and ν, Corollary 4.2 tells us that from the nearby flow it is also possible to deduce the relative precision with which $W^{(t)}_c$ estimates W_c. Moreover, if the nearby flow is a probability measure, the threshold t does not alter the solution of the transportation problem.

Corollary 4.3. *Let be given a positive threshold $t > 0$ and two probability measures μ and ν. If $\eta^* \in \Theta^{(t)}_{\mu,\nu}$ satisfies the equality*

$$\sum_{(x,y)\in N^{(t)}_c} \eta^*_{x,y} = 1, \tag{4.21}$$

then

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) = \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi).$$

Proof. Since $c_{x,y} \geq c^{(t)}_{x,y}$, we have

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) - \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi) \geq 0.$$

From relations (4.16) and (4.17) we have that

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) - \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi) \quad \leq \quad \sum_{(x,y)\in F^{[t]}_c} c^{[t]}_{x,y}\zeta_{x,y}$$

$$\begin{aligned} &\le (C-t)_+ \sum_{(x,y)\in F_c^{[t]}} \zeta_{x,y} &(4.22)\\ &= (C-t)_+\Big(1-\sum_{(x,y)\in N_c^{(t)}} \eta_{x,y}\Big) \end{aligned}$$

where $C = \max\{c_{x,y} : (x,y) \in X \times Y\}$. Since η satisfies (4.21), we conclude

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) - \min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_{c^{(t)}}(\pi) = 0.$$

□

4.3 The Scattered Flow problem

Up to now, our results do not require any specific structure on the cost function c, nor on the spaces X and Y. In this section, we specialize our discussion on the separable cost functions over two regular grids.
Unfortunately, the $t-$truncated cost function of a separable cost function is not separable, i.e. the separability is not preserved through truncation. For this reason, we introduce a new way to truncate cost functions that gives birth to the bi-truncated cost functions, which are indeed separable. For this class of functions is it possible to retrieve a formulation that mimics both the cardinal flow and the max nearby flow.

Definition 4.9 (Bitruncated Cost Function)**.** *Let $c = c^{(1)}+c^{(2)}$ be a separable cost function on $G \times G := (I_1 \times I_2) \times (I_1 \times I_2)$ as in Definition 3.1. Given a couple of positive thresholds (t_1, t_2), we define the (t_1, t_2)-*bitruncated *cost $c^{(t_1,t_2)}$ as*

$$c^{(t_1,t_2)}_{a,b} := \min\{c^{(1)}_{a_1,b_1}, t_1\} + \min\{c^{(2)}_{a_2,b_2}, t_2\},$$

where $a = (a_1, a_2)$ and $b = (b_1, b_2)$. When $t_1 = t_2 = t$, we say that the bi-truncate is $t-$univariate.

Remark 4.5. *The bi-truncate $t-$univariate cost function is a better approximation of c than the $t-$truncated one. In fact, since both the sub-costs $c^{(1)}$ and $c^{(2)}$ are positive, we have that*

$$\min\Big\{c^{(1)} + c^{(2)}, t\Big\} \le \min\Big\{c^{(1)}, t\Big\} + \min\Big\{c^{(2)}, t\Big\}$$

and, by definition we deduce

$$c^{(t)} \le c^{(t,t)}.$$

From Corollary 2.1, this bound can be lifted to the transportation functionals, hence, given any pair of probability measure, μ and ν, we deduce

$$W_{c^{(t)}}(\mu,\nu) \leq W_{c^{(t,t)}}(\mu,\nu) \leq W_c(\mu,\nu).$$

Definition 4.10. *Given a separable cost function c and a couple of positive thresholds (t_1, t_2), we define the set of near points along the* i-th *direction as*

$$N_{t_i}^{(i)} := \Big\{(a_i, b_i) \in I_i \times I_i \text{ such that } c^{(i)}(a_i, b_i) < t_i\Big\}$$

and the set of points near to a_i and b_i along the i-th *direction, respectively, as*

$$O_{t_i}^{(i)}(a_i) := \Big\{b_i \in I_i \text{ such that } c^{(i)}(a_i, b_i) < t_i\Big\},$$

$$I_{t_i}^{(i)}(b_i) := \Big\{a_i \in I_i \text{ such that } c^{(i)}(a_i, b_i) < t_i\Big\}$$

As for the nearby flow problem, we do not need to consider the pairs of point whose distance is greater than the fixed threshold. Roughly speaking, in this case we have to do the same manipulation for each cardinal direction. As a consequence, the number of required arcs drops, while the elements on which we have to maximize have a more complex definition.

Definition 4.11 (Scattered Flow)**.** *Let us take a separable cost function c and a couple of positive values (t_1, t_2). Given two probability measures $\mu := \{\mu_{a_1,a_2}\}$ and $\nu := \{\nu_{b_1,b_2}\}$, we say that the couple $F := (F^{(1)}, F^{(2)})$, with*

$$F^{(1)} := \Big\{f^{(1)}_{a_1,a_2,b_1} \geq 0 \text{ such that } (a_1, b_1) \in N^{(1)}_{t_1}, a_2 \in I_2\Big\}$$

and

$$F^{(2)} := \Big\{f^{(2)}_{b_1,a_2,b_2} \geq 0 \text{ such that } (a_2, b_2) \in N^{(2)}_{t_2}, b_1 \in I_1\Big\}$$

is a scattered flow between μ and ν if it satisfies the following conditions:

M1 (first marginal condition)

$$\sum_{b_1 \in O^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} \leq \mu_{a_1,a_2}, \tag{4.23}$$

M2 (second marginal condition)

$$\sum_{a_2 \in I^{(2)}_{t_2}(b_2)} f^{(2)}_{b_1,a_2,b_2} \leq \nu_{b_1,b_2}, \tag{4.24}$$

P1 (first patching condition)

$$\sum_{a_2\in I_2} \max\left\{ \sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1},\ \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{b_1,a_2,b_2}\right\} \leq \sum_{b_2\in I_2} \nu_{b_1,b_2} \quad (4.25)$$

P2 (second patching condition)

$$\sum_{b_1\in I_1} \max\left\{ \sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1},\ \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{b_1,a_2,b_2}\right\} \leq \sum_{a_1\in I_1} \mu_{a_1,a_2}. \quad (4.26)$$

We call the collections $F^{(1)}$ and $F^{(2)}$ first and second scattered flow, respectively. We denote by $\mathcal{SF}(\mu,\nu)$ the set of all scattered flows between μ and ν.

The conditions $M1 - M2$ are similar to the ones that define the nearby flow and have the same interpretation. The patching conditions $P1 - P2$ are the truncated version of the gluing condition for cardinal flows. To be more specific, they give a bound on the total amount of mass that arrives horizontally and leaves vertically at a certain point.
In Remark 4.7 we show how, under further assumptions, the patching conditions are equivalent to the gluing one.

Definition 4.12 (Scattered Flow Functional)**.** *Let us take a separable cost function $c = c^{(1)} + c^{(2)}$ and a couple of positive values (t_1, t_2). We define the first and second scattered flow functionals as*

$$\begin{aligned}
\mathbb{B}^{(t_1)}(F^{(1)}) &:= \sum_{(a_1,a_2)\in G} \sum_{b_1\in O^{(1)}_{t_1}(a_1)} s^{(1)}_{a_1,b_1} f^{(1)}_{a_1,a_2,b_1}\\
\mathbb{B}^{(t_2)}(F^{(2)}) &:= \sum_{(b_1,b_2)\in G} \sum_{a_2\in I^{(2)}_{t_2}(b_2)} s^{(2)}_{a_2,b_2} f^{(2)}_{b_1,a_2,b_2},
\end{aligned}$$

where $s^{(1)}_{a_1,b_1} := (t_1 - c^{(1)}_{a_1,b_1})$ and $s^{(2)}_{a_2,b_2} := (t_2 - c^{(2)}_{a_2,b_2})$.
The total scattered flow functional $\mathbb{B}^{(t_1,t_2)}$ is then defined as

$$\mathbb{B}^{(t_1,t_2)}(F) := \mathbb{B}^{(t_1)}(F^{(1)}) + \mathbb{B}^{(t_2)}(F^{(2)}). \quad (4.27)$$

Remark 4.6. *Let μ and ν be two probability measures, t be a positive threshold, and $\eta \in \mathcal{N}^{(t)}(\mu,\nu)$. We can build a scattered flow from a nearby flow η*

between μ and ν by setting

$$f^{(1)}_{a_1,a_2,b_1} := \sum_{\substack{b_2 \ s.t. \\ (a_1,a_2,b_1,b_2)\in N_c^{(t)}}} \eta_{a_1,a_2,b_1,b_2}$$

and

$$f^{(2)}_{a_2,b_1,b_2} := \sum_{\substack{a_1 \ s.t. \\ (a_1,a_2,b_1,b_2)\in N_c^{(t)}}} \eta_{a_1,a_2,b_1,b_2}.$$

If we extend the couple defined above on all the missing indices, we can conclude that it is indeed a scattered flow. In fact

$$\sum_{b_1\in O^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} = \sum_{\substack{b_1,b_2 \ s.t. \\ (a_1,a_2,b_1,b_2)\in N_c^{(t)}}} \eta_{a_1,a_2,b_1,b_2} \leq \mu_{a_1,a_2}$$

and

$$\sum_{a_2\in I^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2} = \sum_{\substack{a_1,a_2 \ s.t. \\ (a_1,a_2,b_1,b_2)\in N_c^{(t)}}} \eta_{a_1,a_2,b_1,b_2} \leq \nu_{b_1,b_2}.$$

To prove the patching conditions notice that

$$\sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1} = \sum_{\substack{a_1,b_2 \ \boldsymbol{s.t.} \\ (a_1,a_2,b_1,b_2)\in N_c^{(t)}}} \eta_{a_1,a_2,b_1,b_2} = \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{a_2,b_1,b_2}$$

hence

$$\sum_{a_2\in I_2} \max\left\{ \sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{a_2,b_1,b_2} \right\} = \sum_{\substack{a_1,a_2,b_2 \ \boldsymbol{s.t.} \\ (a_1,a_2,b_1,b_2)\in N_c^{(t)}}} \eta_{a_1,a_2,b_1,b_2} \leq \sum_{b_2\in I_2} \nu_{b_1,b_2}$$

and, similarly, we can prove

$$\sum_{b_1\in I_1} \max\left\{ \sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{a_2,b_1,b_2} \right\} \leq \sum_{a_1\in I_1} \mu_{a_1,a_2}.$$

Moreover, notice that for this couple it holds true

$$\sum_{(x,y)\in N_c^{(t)}} s_{a_1,a_2,b_1,b_2}\eta_{a_1,a_2,b_1,b_2} = \sum_{(a_1,a_2)\in G}\sum_{b_1\in O_{t_1}^{(1)}(a_1)} s^{(1)}_{a_1,b_1} f^{(1)}_{a_1,a_2,b_1} + \sum_{(b_1,b_2)\in G}\sum_{a_2\in I_{t_2}^{(2)}(b_2)} s^{(2)}_{a_2,b_2} f^{(2)}_{a_2,b_1,b_2}.$$

The goal of this section is to show that the search for a maximal scattered flow, i.e. a scattered flow that maximizes functional (4.27), is related to the optimal transport problem associated to the cost $c^{(t_1,t_2)}$. In particular, we show that

$$\min_{\pi\in\Pi(\mu,\nu)}\left[\mathbb{T}_{c^{(t_1,t_2)}}(\pi)\right] = t_1+t_2-\max_{(F^{(1)},F^{(2)})\in\mathcal{SF}(\mu,\nu)}\left[\mathbb{B}^{(t_1)}(F^{(1)})+\mathbb{B}^{(t_2)}(F^{(2)})\right].$$

We show this identity by constructing a scattered flow from a transportation plan and *vice versa*. Finding a nearby flow from a transportation plan is easy: it suffices to restrict the plan over $N_c^{(t)}$. On the other hand, the reverse construction is much harder.

To simplify the study, we introduce a preliminary result.

Lemma 4.2 (Patching Lemma)**.** *Let α_{a_1,a_2,b_1} and β_{b_1,a_2,b_2} be two sub-probability measures on $G\times I_1$ and on $G\times I_2$. Furthermore, assume that there exist two probability measures on G, $\mu := \{\mu_{a_1,a_2}\}$ and $\nu := \{\nu_{b_1,b_2}\}$ such that*

$$\sum_{b_1\in I_1}\alpha_{a_1,a_2,b_1}\le\mu_{a_1,a_2}, \qquad \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\le\nu_{b_1,b_2}, \tag{4.28}$$

$$\sum_{a_2\in I_2}\max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1},\sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\}\le\sum_{b_2\in I_2}\nu_{b_1,b_2} \tag{4.29}$$

and

$$\sum_{b_1\in I_1}\max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1},\sum_{a_2\in I_2}\beta_{b_1,a_2,b_2}\right\}\le\sum_{a_1\in I_1}\mu_{a_1,a_2}. \tag{4.30}$$

We can then find a probability measure $\pi := \pi_{a_1,a_2,b_1,b_2}$ on $G\times G$ such that

$$\sum_{(a_1,a_2)\in G}\pi_{a_1,a_2,b_1,b_2}=\nu_{b_1,b_2}, \qquad \sum_{(b_1,b_2)\in G}\pi_{a_1,a_2,b_1,b_2}=\mu_{a_1,a_2}, \tag{4.31}$$

and

$$\sum_{a_1\in I_1}\pi_{a_1,a_2,b_1,b_2}\ge\beta_{b_1,a_2,b_2}, \qquad \sum_{b_2\in I_2}\pi_{a_1,a_2,b_1,b_2}\ge\alpha_{a_1,a_2,b_1}. \tag{4.32}$$

Proof. Given any couple $(a_2, b_1) \in I_2 \times I_1$, we can find two values $k^{(1)}_{a_2,b_1}, k^{(2)}_{a_2,b_1} \in [0,1]$ such that

$$k^{(1)}_{a_2,b_1} \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} = k^{(2)}_{a_2,b_1} \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} \tag{4.33}$$

and

$$\max\left\{k^{(1)}_{a_2,b_1}, k^{(2)}_{a_2,b_1}\right\} = 1. \tag{4.34}$$

It holds true that

$$\begin{aligned} \max\left\{\sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1}, \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}\right\} &= \left(1 - k^{(1)}_{a_2,b_1}\right) \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} \\ &\quad + \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}. \end{aligned} \tag{4.35}$$

In fact, if $k^{(1)}_{a_2,b_1} = 1$, we have that

$$\sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} = k^{(2)}_{a_2,b_1} \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} \leq \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}$$

so that

$$\max\left\{\sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1}, \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}\right\} = \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}. \tag{4.36}$$

If we substitute $k^{(1)}_{a_2,b_1} = 1$ in the second member of (4.35), we get

$$(1 - k^{(1)}_{a_2,b_1}) \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} + \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} = \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}$$

so that (4.35) holds true. If $k^{(1)}_{a_2,b_1} < 1$, by condition (4.34) we get $k^{(2)}_{a_2,b_1} = 1$, hence

$$\sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} = k^{(1)}_{a_2,b_1} \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} \leq \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} \tag{4.37}$$

and then

$$\max\left\{\sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1}, \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}\right\} = \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1}. \tag{4.38}$$

Again, by using (4.33), we find

$$\left(1 - k^{(1)}_{a_2,b_1}\right) \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} + \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}$$

$$= \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} + \left(1 - k^{(2)}_{a_2,b_1}\right) \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2}$$

and, by substituting $k^{(2)}_{a_2,b_1} = 1$, we get (4.35).
The identity (4.35) allows us to rewrite the condition (4.29) as

$$\sum_{a_2 \in I_2} \left[(1 - k^{(1)}_{a_2,b_1}) \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1} + \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} \right] \leq \sum_{b_2 \in I_2} \nu_{b_1,b_2} \tag{4.39}$$

hence

$$\sum_{(a_1,a_2) \in G} \left(1 - k^{(1)}_{a_2,b_1}\right) \alpha_{a_1,a_2,b_1} \leq \sum_{b_2 \in I_2} \left[\nu_{b_1,b_2} - \sum_{a_2 \in I_2} \beta_{b_1,a_2,b_2} \right]. \tag{4.40}$$

Similarly, we can show that

$$\begin{aligned} \max\left\{ \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1}, \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} \right\} \quad = \quad & (1 - k^{(2)}_{a_2,b_1}) \sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} \\ & + \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1}, \end{aligned}$$

and rewrite condition (4.30) as

$$\sum_{(b_2,b_1) \in G} \left(1 - k^{(2)}_{a_2,b_1}\right) \beta_{b_1,a_2,b_2} \leq \sum_{a_1 \in I_1} \left[\mu_{a_1,a_2} - \sum_{b_1 \in I_1} \alpha_{a_1,a_2,b_1} \right]. \tag{4.41}$$

We are now ready to build the measure π.
Let us fix a couple of indexes $(a_2, b_1) \in I_2 \times I_1$. The relation (4.33) allows us to use the gluing lemma and find a measure $\rho^{a_2,b_1}_{a_1,b_2}$ such that

$$\sum_{a_1 \in I_1} \rho^{a_2,b_1}_{a_1,b_2} = k^{(2)}_{a_2,b_1} \beta_{b_1,a_2,b_2} \tag{4.42}$$

and

$$\sum_{b_2 \in I_2} \rho^{a_2,b_1}_{a_1,b_2} = k^{(1)}_{a_2,b_1} \alpha_{a_1,a_2,b_1} \tag{4.43}$$

for any $(a_2, b_1) \in I_2 \times I_1$.

Let us now fix $b_1 \in I_1$. By relation (4.40), we can find a constant $0 \leq s^{(2)}_{b_1} \leq 1$ such that

$$\sum_{(a_1,a_2)\in G} \big(1 - k^{(1)}_{a_2,b_1}\big)\alpha_{a_1,a_2,b_1} = s^{(2)}_{b_1} \sum_{b_2\in I_2} \left[\nu_{b_1,b_2} - \sum_{a_2\in I_2} \beta_{b_1,a_2,b_2}\right] \tag{4.44}$$

and, again by using the gluing lemma, we are able to find a measure $\psi^{b_1}_{a_1,a_2,b_2}$ such that

$$\sum_{b_2\in I_2} \psi^{b_1}_{a_1,a_2,b_2} = \big(1 - k^{(1)}_{a_2,b_1}\big)\alpha_{a_1,a_2,b_1} \tag{4.45}$$

and

$$\sum_{(a_1,a_2)\in G} \psi^{b_1}_{a_1,a_2,b_2} = s^{(2)}_{b_1} \left[\nu_{b_1,b_2} - \sum_{a_2\in I_2} \beta_{b_1,a_2,b_2}\right] \tag{4.46}$$

for any $b_1 \in I_1$.
Similarly, fixed $a_2 \in I_2$. We can find a constant $0 \leq s^{(1)}_{a_2} \leq 1$ such that

$$\sum_{(b_1,b_2)\in G} (1 - k^{(2)}_{a_2,b_1})\beta_{b_1,a_2,b_2} = s^{(1)}_{a_2} \sum_{a_1\in I_1} \left[\mu_{a_1,a_2} - \sum_{b_1\in I_1} \alpha_{a_1,a_2,b_1}\right] \tag{4.47}$$

and a measure $\phi^{a_2}_{a_1,b_1,b_2}$ such that

$$\sum_{(b_1,b_2)\in G} \phi^{a_2}_{a_1,b_1,b_2} = s^{(1)}_{a_2} \left[\mu_{a_1,a_2} - \sum_{b_1\in I_1} \alpha_{a_1,a_2,b_1}\right], \tag{4.48}$$

and

$$\sum_{a_1\in I_1} \phi^{a_2}_{a_1,b_1,b_2} = (1 - k^{(2)}_{a_2,b_1})\beta_{b_1,a_2,b_2} \tag{4.49}$$

for any $a_2 \in I_2$.
Let us now define

$$\tilde{\pi}_{a_1,a_2,b_1,b_2} := \rho^{a_2,b_1}_{a_1,b_2} + \psi^{b_1}_{a_1,a_2,b_2} + \phi^{a_2}_{a_1,b_1,b_2}.$$

By a simple computation we get

$$\sum_{(a_1,a_2)\in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} = \sum_{(a_1,a_2)\in G} \rho^{a_2,b_1}_{a_1,b_2} + \sum_{(a_1,a_2)\in G} \psi^{b_1}_{a_1,a_2,b_2} + \sum_{(a_1,a_2)\in G} \phi^{a_2}_{a_1,b_1,b_2}$$

$$
\begin{aligned}
&= \sum_{a_2 \in I_2} k^{(2)}_{a_2,b_1} \beta_{b_1,a_2,b_2} + s^{(2)}_{b_1 \in I_1} \left[\nu_{b_1,b_2} - \sum_{a_2 \in I_2} \beta_{b_1,a_2,b_2} \right] \\
&+ \sum_{a_2 \in I_2} \big(1 - k^{(2)}_{a_2,b_1}\big) \beta_{b_1,a_2,b_2} \qquad (4.50) \\
&= \sum_{a_2 \in I_2} \beta_{b_1,a_2,b_2} + s^{(2)}_{b_1 \in I_1} \left[\nu_{b_1,b_2} - \sum_{a_2 \in I_2} \beta_{b_1,a_2,b_2} \right]
\end{aligned}
$$

By condition (4.28), the quantity $(\nu_{b_1,b_2} - \sum_{a_2 \in I_2} \beta_{b_1,a_2,b_2})$ is positive and, by definition, $s^{(2)}_{b_1} \leq 1$, hence, from (4.50), we are able to conclude

$$
\sum_{(a_1,a_2) \in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} \leq \sum_{a_2 \in I_2} \beta_{a_1,b_1,b_2} + \left[\nu_{b_1,b_2} - \sum_{a_2 \in I_2} \beta_{b_1,a_2,b_2} \right] = \nu_{b_1,b_2}. \qquad (4.51)
$$

Similarly, we can show

$$
\sum_{(b_1,b_2) \in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} \leq \mu_{a_1,a_2}. \qquad (4.52)
$$

We can then define the slack measures

$$
\tilde{\mu}_{a_1,a_2} = \mu_{a_1,a_2} - \sum_{(b_1,b_2) \in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} \qquad (4.53)
$$

and

$$
\tilde{\nu}_{b_1,b_2} = \nu_{b_1,b_2} - \sum_{(a_1,a_2) \in G} \tilde{\pi}_{a_1,a_2,b_1,b_2}. \qquad (4.54)
$$

From a direct computation, we find

$$
M = \sum_{a \in G} \tilde{\mu}_{a_1,a_2} = \sum_{b \in G} \tilde{\nu}_{b_1,b_2}
$$

and, then, it is possible to take the direct product of those measure $\tilde{\mu} \otimes \tilde{\nu}$ and, finally define

$$
\pi_{a_1,a_2,b_1,b_2} = \tilde{\pi}_{a_1,a_2,b_1,b_2} + \tilde{\mu}_{a_1,a_2} \otimes \tilde{\nu}_{b_1,b_2}. \qquad (4.55)
$$

The measure π defined above is indeed a transportation plan between μ and ν. In fact,

$$
\sum_{(a_1,a_2) \in G} \pi_{a_1,a_2,b_1,b_2} \quad = \quad \sum_{(a_1,a_2) \in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} + \sum_{(a_1,a_2) \in G} \tilde{\mu}_{a_1,a_2} \times \tilde{\nu}_{b_1,b_2}
$$

$$\begin{aligned} &= \sum_{(a_1,a_2)\in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} + \tilde{\nu}_{b_1,b_2} \\ &= \nu_{b_1,b_2} \end{aligned}$$

and

$$\begin{aligned} \sum_{(b_1,b_2)\in G} \pi_{a_1,a_2,b_1,b_2} &= \sum_{(b_1,b_2)\in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} + \sum_{(b_1,b_2)\in G} \tilde{\mu}_{a_1,a_2} \times \tilde{\nu}_{b_1,b_2} \\ &= \sum_{(b_1,b_2)\in G} \tilde{\pi}_{a_1,a_2,b_1,b_2} + \tilde{\mu}_{a_1,a_2} \\ &= \mu_{a_1,a_2}, \end{aligned}$$

so that $\pi \in \Pi(\mu,\nu)$.
To conclude we have to show the relations in (4.32). Again, by a direct computation

$$\begin{aligned} \sum_{a_1\in I_1} \pi_{a_1,a_2,b_1,b_2} &\geq \sum_{a_1\in I_1} \tilde{\pi}_{a_1,a_2,b_1,b_2} \\ &= \sum_{a_1\in I_1} \rho^{a_2,b_1}_{a_1,b_2} + \sum_{a_1\in I_1} \psi^{b_1}_{a_1,a_2,b_2} + \sum_{a_1\in I_1} \phi^{a_2}_{a_1,b_1,b_2} \\ &\geq \sum_{a_1\in I_1} \rho^{a_2,b_1}_{a_1,b_2} + \sum_{a_1\in I_1} \phi^{a_2}_{a_1,b_1,b_2} \end{aligned}$$

where the last inequality follow from $\psi^{b_1}_{a_1,a_2,b_2} \geq 0$. Since

$$\begin{aligned} \sum_{a_1\in I_1} \rho^{a_2,b_1}_{a_1,b_2} &= k^{(2)}_{a_2,b_1} \beta_{b_1,a_2,b_2} \\ \sum_{a_1\in I_1} \phi^{a_2}_{a_1,b_1,b_2} &= (1-k^{(2)}_{a_2,b_1})\beta_{b_1,a_2,b_2} \end{aligned}$$

we can conclude

$$\sum_{a_1\in I_1} \pi_{a_1,a_2,b_1,b_2} \geq k^{(2)}_{a_2,b_1}\beta_{b_1,a_2,b_2} + (1-k^{(2)}_{a_2,b_1})\beta_{b_1,a_2,b_2} = \beta_{b_1,a_2,b_2}.$$

Similarly we can show that

$$\sum_{b_2\in I_2} \pi_{a_1,a_2,b_1,b_2} \geq \alpha_{a_1,a_2,b_1}.$$

This proves that π is indeed the probability measure we were looking for. $\square$

Remark 4.7. *Let us take α and β as in Lemma 4.2, such that the conditions (4.28), (4.29) and (4.30) are equalities,* i.e.

$$\sum_{b_1\in I_1} \alpha_{a_1,a_2,b_1} = \mu_{a_1,a_2} \qquad \forall (a_1,a_2)\in G, \tag{4.56}$$

$$\sum_{a_2\in I_2} \beta_{b_1,a_2,b_2} = \nu_{b_1,b_2} \qquad \forall (b_1,b_2)\in G, \tag{4.57}$$

and, for all $(b_2,a_1)\in I_2\times I_1$,

$$\sum_{a_2\in I_2} \max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\} = \sum_{b_2\in I_2}\nu_{b_1,b_2}, \tag{4.58}$$

$$\sum_{b_1\in I_1} \max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\} = \sum_{a_1\in I_1}\mu_{a_1,a_2}. \tag{4.59}$$

We can then recover the hypothesis of the gluing lemma.
In fact, by taking the relations (4.58) and (4.56), we find

$$\sum_{a_2\in I_2} \max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\} = \sum_{(a_2,b_2)\in I_2\times I_2}\beta_{b_1,a_2,b_2}, \tag{4.60}$$

which can be rewritten as

$$\sum_{b_2\in I_2}\left[\max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\} - \sum_{a_2\in I_2}\beta_{b_1,a_2,b_2}\right] = 0. \tag{4.61}$$

Since

$$\max_{(b_2,a_1)\in I_2\times I_1}\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\} - \sum_{a_2\in I_2}\beta_{b_1,a_2,b_2} \geq 0$$

we can deduce from (4.61) that

$$\max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\} = \sum_{a_2\in I_2}\beta_{b_1,a_2,b_2} \tag{4.62}$$

for all $(b_2,a_1)\in I_2\times I_1$. Similarly, if we consider the relations (4.57) and (4.59) we can prove that

$$\max\left\{\sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \sum_{b_2\in I_2}\beta_{b_1,a_2,b_2}\right\} = \sum_{a_1\in I_1}\alpha_{a_1,a_2,b_1}, \tag{4.63}$$

for all $(b_2, a_1) \in I_2 \times I_1$. *Putting together* (4.62) *and* (4.63), *we find*

$$\sum_{b_2 \in I_2} \beta_{b_1,a_2,b_2} = \sum_{a_1 \in I_1} \alpha_{a_1,a_2,b_1}$$

i.e. *the condition of the gluing lemma.*
In particular, we notice that, in this case, any scattered flow is actually a cardinal flow.

Definition 4.13 (Trivial Extension). *Let* $F := \left(F^{(1)}, F^{(2)}\right)$ *be a scattered flow between two measures* μ *and* ν. *The trivial extension of* F *is defined as the couple* $\left(\tilde{F}^{(1)}, \tilde{F}^{(2)}\right)$, *where* $\tilde{F}^{(i)}$ *is the collection of positive value given by the formula*

$$\tilde{f}^{(1)}_{a_1,a_2,b_1} := \begin{cases} f^{(1)}_{a_1,a_2,b_1} & \text{if } (a_1, b_1) \in N^{(1)}_{t_1} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\tilde{f}^{(2)}_{a_2,b_1,b_2} := \begin{cases} f^{(2)}_{a_2,b_1,b_2} & \text{if } (a_2, b_2) \in N^{(2)}_{t_2} \\ 0 & \text{otherwise.} \end{cases}$$

Remark 4.8. *Let us take* μ *and* ν *two probability measures and* t *a positive value. If a* $t-$*univariate scattered flow* $(f^{(1)}, f^{(2)})$ *has unitary mass, i.e. such that*

$$\sum_{(a_1,a_2)\in G, b_1 \in O^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} = \sum_{(b_1,b_2)\in G, a_2 \in I^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2} = 1 \tag{4.64}$$

then its trivial extension $(\tilde{f}^{(1)}, \tilde{f}^{(2)})$ *are cardinal flow. In fact, since*

$$\sum_{x \in X} \mu_{x_1,x_2} = \sum_{y \in Y} \nu_{y_1,y_2} = 1$$

all the inequalities in (4.28), (4.29) and (4.30), must hold as equivalences. Then, by Remark 4.7, we get that $\tilde{f}^{(1)}$ *and* $\tilde{f}^{(2)}$ *glue on a common marginal, i.e.*

$$\sum_{a_1 \in I_1} \tilde{f}^{(1)}_{a_1,a_2,b_1} = \sum_{b_2 \in I_2} \tilde{f}^{(2)}_{a_2,b_1,b_2}$$

hence the couple $(\tilde{f}^{(1)}, \tilde{f}^{(2)})$ *is a cardinal flow.*

We are now ready to prove the equivalence between the maximization of the scattered flow functional and the minimization of the transportation one for bi-truncated cost functions. As always, we show that for any given transportation plan we can recover a scattered flow and vice versa.

Theorem 4.3. *Given two probability measures μ, ν and a couple of positive parameters (t_1, t_2) then the following identity holds true*

$$\min_{\Pi(\mu,\nu)} \Big[\mathbb{T}_{c^{(t_1,t_2)}}(\pi)\Big] = t_1 + t_2 - \max_{\mathcal{F}(\mu,\nu)} \Big[\mathbb{B}^{(t_1,t_2)}(F^{(1)}, F^{(2)})\Big].$$

Proof. The proof is divided in two different parts. In the first part we show that

$$\min_{\Pi(\mu,\nu)} \Big[\mathbb{T}_{c^{(t_1,t_2)}}(\pi)\Big] \leq t_1 + t_2 - \max_{\mathcal{F}(\mu,\nu)} \Big[\mathbb{B}^{(t_1,t_2)}(F^{(1)}, F^{(2)})\Big].$$

In the second part we show the other inequality. Let us take $\pi \in \Pi(\mu, \nu)$. We can then define

$$f^{(1)}_{a_1,a_2,b_1} := \sum_{b_2 \in I_2} \pi_{a,b}$$

for each $(a_1, a_2, b_1) \in G \times I_1$ and

$$f^{(2)}_{b_1,a_2,b_2} := \sum_{a_1 \in I_1} \pi_{a,b}$$

for each $(b_1, a_2, b_2) \in G \times I_2$.

The couple $\Big(\{f^{(1)}_{a_1,a_2,b_1}\}, \{f^{(2)}_{b_1,a_2,b_2}\}\Big)$ is a cardinal flow between μ and ν, hence it holds true

$$\sum_{b_1 \in I_1} f^{(1)}_{a_1,a_2,b_1} = \sum_{(b_1,b_2) \in G} \pi_{a_1,a_2,b_1,b_2} = \mu_{a_1,a_2}$$

for each $(a_1, a_2) \in G$ and

$$\sum_{a_2 \in I_2} f^{(2)}_{b_1,a_2,b_2} = \sum_{(a_1,a_2) \in G} \pi_{a_1,a_2,b_1,b_2} = \nu_{b_1,b_2}$$

for each $(b_1, b_2) \in G$.

Furthermore, they have the same marginal on $I_2 \times I_1$, i.e.

$$\sum_{a_1 \in I_1} f^{(1)}_{a_1,a_2,b_1} = \sum_{b_2 \in I_2} f^{(2)}_{b_1,a_2,b_2}$$

hence

$$\max\Bigg\{\sum_{a_1 \in I_1} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2 \in I_2} f^{(2)}_{b_1,a_2,b_2}\Bigg\} = \sum_{a_1 \in I_1} f^{(1)}_{a_1,a_2,b_1} = \sum_{b_2 \in I_2} f^{(2)}_{b_1,a_2,b_2}. \quad (4.65)$$

By summing upon all the $a_2 \in I_2$ and upon all the $b_1 \in I_1$ respectively, from (4.65) we get

$$\begin{aligned}\sum_{a_2\in I_2} \max\Big\{\sum_{a_1\in I_1} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in I_2} f^{(2)}_{b_1,a_2,b_2}\Big\} &= \sum_{(a_1,a_2)\in G} f^{(1)}_{a_1,a_2,b_1} \\ &= \sum_{b_2\in I_2} \nu_{b_1,b_2}\end{aligned}$$

for all $b_1 \in I_1$, and

$$\begin{aligned}\sum_{b_1\in I_1} \max\Big\{\sum_{a_1\in I_1} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in I_2} f^{(2)}_{b_1,a_2,b_2}\Big\} &= \sum_{(b_1,b_2)\in G} f^{(2)}_{b_1,a_2,b_2} \\ &= \sum_{a_1\in I_1} \mu_{a_1,a_2}\end{aligned}$$

for each $a_2 \in I_2$.
Finally we can define a scattered flow by taking the restriction of the above cardinal flow

$$F^{(1)} := \big\{f^{(1)}_{a_1,a_2,b_1}\big\}_{(a_1,a_2)\in G, b_1\in O^{(1)}_{t_1}(a_1)},$$

and

$$F^{(2)} := \big\{f^{(2)}_{b_1,a_2,b_2}\big\}_{(b_1,b_2)\in G, a_2\in I^{(2)}_{t_2}(b_2)}.$$

The couple $F := (F^{(1)}, F^{(2)})$ is a scattered flow between μ and ν since

$$\sum_{b_1\in O^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} \le \sum_{b_1\in I_1} f^{(1)}_{a_1,a_2,b_1} = \mu_{a_1,a_2}$$

for all $(a_1, a_2) \in G$ and

$$\sum_{a_2\in I^{(2)}_{t_2}(b_2)} f^{(2)}_{b_1,a_2,b_2} \le \sum_{b_2\in I_2} f^{(2)}_{b_1,a_2,b_2} = \nu_{b_1,b_2}$$

for all $(b_1, b_2) \in G$, so that relations (4.23) and (4.24) are satisfied.
Similarly, since it holds true

$$\begin{aligned}\sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1} &\le \sum_{a_1\in I_1} f^{(1)}_{a_1,a_2,b_1}, &\qquad \forall (a_2,b_1) \in I_2\times I_1 \\ \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{b_1,a_2,b_2} &\le \sum_{b_2\in I_2} f^{(2)}_{b_1,a_2,b_2}, &\qquad \forall (a_2,b_1) \in I_2\times I_1\end{aligned}$$

we can conclude

$$\max\Big\{\sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{b_1,a_2,b_2}\Big\} \le \max\Big\{\sum_{a_1\in I_1} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in I_2} f^{(2)}_{b_1,a_2,b_2}\Big\} \tag{4.66}$$

for each couple $(a_2, b_1) \in I_2 \times I_1$.
By summing (4.66) upon $a_2 \in I_2$ and remembering the relation (4.65), we get

$$\max\Big\{\sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{b_1,a_2,b_2}\Big\} \le \sum_{b_2\in I_2} \nu_{b_1,b_2}.$$

Similarly, by summing upon all the $b_1 \in I_1$, we get

$$\sum_{b_1\in I_1}\max\Big\{\sum_{a_1\in I^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in O^{(2)}_{t_2}(a_2)} f^{(2)}_{b_1,a_2,b_2}\Big\} \le \sum_{a_1\in I_1} \mu_{a_1,a_2}.$$

So that the couple $(F^{(1)}, F^{(2)})$ is a scattered flow between μ and ν.
To conclude this first part we just need to show that

$$\sum_{(a,b)\in G\times G} c^{(t_1,t_2)}_{a,b}\pi_{a,b} = t_1 + t_2 - \mathbb{B}(F).$$

This follows from the definition of the transportation cost functional related to the ground distance $c^{(t_1,t_2)}$, in fact

$$\begin{aligned}
\sum_{(a,b)\in G\times G} c^{(t_1,t_2)}_{a,b}\pi_{a,b} &= \sum_{(a,b)\in G\times G} c^{(t_1)}_{a_1,b_1}\pi_{a,b} + \sum_{(a,b)\in G\times G} c^{(t_2)}_{a_2,b_2}\pi_{a,b}\\
&= \sum_{\substack{(a_1,a_2)\in G,\\ b_1\in I_1(a_1)}} c^{(t_1)}_{a_1,b_1}\sum_{b_2\in I_2}\pi_{a,b} + \sum_{\substack{(b_1,b_2)\in G,\\ a_2\in I_2(b_2)}} c^{(t_2)}_{a_2,b_2}\sum_{a_1\in I_1}\pi_{a,b}\\
&= t_1 \sum_{\substack{(a_1,a_2)\in G, b_2\in I_2,\\ b_1\notin O^{(1)}_{t_1}(a_1)}} \pi_{a,b} + \sum_{\substack{(a_1,a_2)\in G,\\ b_1\in O^{(1)}_{t_1}(a_1)}} c^{(t_1)}_{a_1,b_1}\sum_{b_2\in I_2}\pi_{a,b}\\
&\quad + t_2 \sum_{\substack{a_1\in I_1(b_1),(b_1,b_2)\in G,\\ a_2\notin I^{(2)}_{t_2}(b_2)}} \pi_{a,b} + \sum_{\substack{(b_1,b_2)\in G,\\ a_2\in I^{(2)}_{t_2}(b_2)}} c^{(t_2)}_{a_2,b_2}\sum_{a_1\in I_1}\pi_{a,b}\\
&= t_1 \sum_{(a,b)\in G\times G}\pi_{a,b} + \sum_{\substack{(a_1,a_2)\in G,\\ b_1\in O^{(1)}_{t_1}(a_1)}} (c^{(t_1)}_{a_1,b_1} - t_1)\sum_{b_2\in I_2}\pi_{a,b}
\end{aligned}$$

$$
\begin{aligned}
& +t_2 \sum_{(a,b)\in G\times G} \pi_{a,b} + \sum_{\substack{(b_1,b_2)\in G,\\ a_2\in I_{t_2}^{(2)}(b_2)}} (c^{(t_2)}_{a_2,b_2} - t_2) \sum_{a_1\in I_1} \pi_{a,b} \\
= \quad & t_1 - \sum_{\substack{(a_1,a_2)\in G,\\ b_1\in O_{t_1}^{(1)}(a_1)}} s^{(t_1)}_{a_1,b_1} f^{(1)}_{a_1,a_2,b_1} \\
& +t_2 - \sum_{\substack{(b_1,b_2)\in G,\\ a_2\in I_{t_2}^{(2)}(b_2)}} s^{(t_2)}_{a_2,b_2} f^{(2)}_{b_1,a_2,b_2} \\
= \quad & t_1 + t_2 - \mathbb{B}(F).
\end{aligned}
$$

Hence

$$\min_{\Pi(\mu,\nu)} \mathbb{T}_{c^{(t_1,t_2)}}(\pi) \geq t_1 + t_2 - \max_{\mathcal{F}(\mu,\nu)} \mathbb{B}(F).$$

To complete the proof, we need to show that, for each $F \in \mathcal{F}(\mu,\nu)$, there exists a $\pi \in \Pi(\mu,\nu)$ for which holds true

$$\sum_{(a,b)\in G\times G} c^{(t_1,t_2)}_{a,b} \pi_{a,b} \leq t_1 + t_2 - \mathbb{B}(F).$$

Let us then take $F = (\{f^{(1)}_{a_1,a_2,b_1}\}, \{f^{(2)}_{b_1,a_2,b_2}\})$ a scattered flow between μ and ν, we define with $(\tilde{F}^{(1)}, \tilde{F}^{(2)})$ their trivial extension. Since both $F^{(1)}$ and $F^{(2)}$ are sub-probability measures, so are $\tilde{F}^{(1)}$ and $\tilde{F}^{(1)}$. Moreover, those extension verify the condition required by the patching Lemma 4.2, hence we can find a $\pi \in \Pi(\mu,\nu)$ such that

$$\sum_{b_2\in I_2} \pi_{a,b} \geq \tilde{f}^{(1)}_{a_1,a_2,b_1} \tag{4.67}$$

and

$$\sum_{a_1\in I_1} \pi_{a,b} \geq \tilde{f}^{(2)}_{b_1,a_2,b_2}. \tag{4.68}$$

Again, by the definition of $\mathbb{T}_{c^{(t_1,t_2)}}$, we deduce

$$
\begin{aligned}
\sum_{(a,b)\in G\times G} c^{(t_1,t_2)}_{a,b} \pi_{a,b} \quad = \quad & t_1 - \sum_{\substack{(a_1,a_2)\in G,\\ b_1\in O_{t_1}^{(1)}(a_1)}} s^{(1)}_{a_1,b_1} \sum_{b_2\in I_2} \pi_{a,b} \\
& +t_2 - \sum_{\substack{(b_1,b_2)\in G,\\ a_2\in I_{t_2}^{(2)}(b_2)}} s^{(2)}_{a_2,b_2} \sum_{a_1\in I_1} \pi_{a,b}
\end{aligned}
$$

which, by relations (4.67) and (4.68), becomes

$$\begin{aligned}\sum_{(a,b)\in G\times G} c_{a,b}^{(t_1,t_2)}\pi_{a,b} \quad \leq \quad & t_1 - \sum_{\substack{(a_1,a_2)\in G,\\ b_1\in O_{t_1}^{(1)}(a_1)}} s_{a_1,b_1}^{(1)}\tilde{f}_{a_1,a_2,b_1}^{(1)} \\ & +t_2 - \sum_{\substack{(b_1,b_2)\in G,\\ a_2\in I_{t_2}^{(2)}(b_2)}} s_{a_2,b_2}^{(2)}\tilde{f}_{b_1,a_2,b_2}^{(2)}.\end{aligned}$$

Since $\tilde{f}_{a_1,a_2,b_1}^{(1)} = f_{a_1,a_2,b_1}^{(1)}$ when $b_1 \in O_{t_1}^{(1)}(a_1)$, and $\tilde{f}_{b_1,a_2,b_2}^{(2)} = f_{b_1,a_2,b_2}^{(2)}$ when $a_2 \in I_{t_2}^{(2)}(b_2)$, we can conclude

$$\sum_{(a,b)\in G\times G} c_{a,b}^{(t_1,t_2)}\pi_{a,b} \leq t_1 + t_2 - \mathbb{B}(F)$$

and, therefore the thesis. □

As for the Nearby Flow problem, we are able to use the maximal scattered flow to bound the absolute error and the relative error.

Theorem 4.4 (Bound on the Absolute Error)**.** *Let us take a separable cost function $c = c^{(1)} + c^{(2)}$ and two discrete measures μ, ν supported on the regular grid $G = I_1 \times I_2$. Given a couple of positive thresholds (t_1, t_2) and an optimal scattered flow $(f^{(1)}, f^{(2)})$ between μ and ν, the following estimate on the absolute and relative error hold*

$$\begin{aligned}|W_c(\mu,\nu) - W_{c^{(t_1,t_2)}}(\mu,\nu)| \quad \leq \quad & \sum_{\substack{a_2\in I_2,\\ (a_1,b_1)\notin N_{t_1}^{(1)}}} (c_{a_1,b_1}^{(1)} - t_1)\gamma_{a_1,a_2,b_1}^{(1)} \\ & + \sum_{\substack{b_1\in I_1,\\ (a_2,b_2)\notin N_{t_2}^{(2)}}} (c_{a_2,b_2}^{(2)} - t_2)\gamma_{a_2,b_1,b_2}^{(2)} \quad (4.69)\end{aligned}$$

where

$$\gamma_{a_1,a_2,b_1}^{(1)} := \frac{\left(\zeta_{b_1,a_2} - \sum_{a_1\in N_{t_1}^{(1)}(b_1)} f_{a_1,a_2,b_1}^{(1)}\right)\left(\mu_{a_1,a_2} - \sum_{b_1\in N_{t_1}^{(1)}(a_1)} f_{a_1,a_2,b_1}^{(1)}\right)}{\sum_{a_1\in I_1}\left(\mu_{a_1,a_2} - \sum_{b_1\in N_{t_1}^{(1)}(a_1)} f_{a_1,a_2,b_1}^{(1)}\right)},$$

$$\gamma^{(2)}_{a_2,b_1,b_2} := \frac{\left(\zeta_{b_1,a_2} - \sum_{b_2\in N^{(2)}_{t_2}(a_2)} f^{(2)}_{a_2,b_1,b_2}\right)\left(\nu_{b_1,b_2} - \sum_{a_2\in N^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2}\right)}{\sum_{b_2\in I_2}\left(\nu_{b_1,b_2} - \sum_{a_2\in N^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2}\right)}$$

and $\zeta \in \mathcal{I}(\mu,\nu)$ *depends only on* $f^{(1)}, f^{(2)}, \mu$ *and* ν.

Proof. Let us define

$$\tilde{\zeta}_{b_1,a_2} = \max\left\{\sum_{a_1\in N^{(1)}_{t_1}} f^{(1)}_{a_1,a_2,b_1}, \sum_{b_2\in N^{(2)}_{t_2}} f^{(2)}_{a_2,b_1,b_2}\right\}, \tag{4.70}$$

since $(f^{(1)}, f^{(2)})$ is a scattered flow between μ and ν, we have

$$\sum_{b_1\in I_1} \tilde{\zeta}_{b_1,a_2} \leq \sum_{a_1\in I_1} \mu_{a_1,a_2}, \tag{4.71}$$

$$\sum_{a_2\in I_2} \tilde{\zeta}_{b_1,a_2} \leq \sum_{b_2\in I_2} \nu_{b_1,b_2} \tag{4.72}$$

hence we can define the slack variables

$$\tilde{\mu}^{(2)}_{a_2} := \sum_{a_1\in I_1} \mu_{a_1,a_2} - \sum_{b_1\in I_1} \tilde{\zeta}_{b_1,a_2}$$

and

$$\tilde{\nu}^{(1)}_{b_1} := \sum_{b_2\in I_2} \nu_{b_1,b_2} - \sum_{a_2\in I_2} \tilde{\zeta}_{b_1,a_2}.$$

The measures $\tilde{\mu}$ and $\tilde{\nu}$ are positive and, moreover, have the same mass, hence the independent product between them is well defined and so is the following probability measure

$$\zeta := \tilde{\zeta} + \tilde{\nu} \otimes \tilde{\mu}.$$

From the definition of ζ we find

$$\sum_{b_1\in I_1} \zeta_{b_1,a_2} = \sum_{a_1\in I_1} \mu_{a_1,a_2}$$

and

$$\sum_{a_2\in I_2} \zeta_{b_1,a_2} = \sum_{b_2\in I_2} \nu_{b_1,b_2},$$

hence ζ belongs to the set of intermediate measures $\mathcal{I}(\mu,\nu)$ (Definition 3.4). We show that there exists a cardinal flow $(F^{(1)}, F^{(2)})$ that glues on ζ and such that

$$F^{(1)}_{a_1,a_2,b_1} = f^{(1)}_{a_1,a_2,b_1} \qquad \forall (a_1,a_2,b_1) \in N^{(1)}_{t_1}, \tag{4.73}$$

$$F^{(2)}_{a_2,b_1,b_2} = f^{(2)}_{a_2,b_1,b_2} \qquad \forall (a_2,b_1,b_2) \in N^{(2)}_{t_2}. \tag{4.74}$$

Let us define the slack variable

$$\alpha_{b_1,a_2} = \zeta_{b_1,a_2} - \sum_{a_1 \in N^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1}.$$

Since $\zeta \in \mathfrak{I}(\mu,\nu)$, we have

$$M_{a_2} := \sum_{a_1 \in I_1} \left(\mu_{a_1,a_2} - \sum_{b_1 \in N^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} \right) = \sum_{b_1 \in I_1} \left(\zeta_{b_1,a_2} - \sum_{a_1 \in N^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1} \right)$$

for any $a_2 \in I_2$. For any given $a_2 \in I_2$, we can then define the direct product of those measures and obtain the collection of measures

$$\gamma^{(1)}_{a_1,a_2,b_1} := \frac{\alpha_{b_1,a_2} \left(\mu_{a_1,a_2} - \sum_{b_1 \in N^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} \right)}{M_{a_2}}. \tag{4.75}$$

We can then define $F^{(1)}$ as

$$F^{(1)}_{a_1,a_2,b_1} := f^{(1)}_{a_1,a_2,b_1} + \gamma^{(1)}_{a_1,a_2,b_1}$$

which is the first cardinal flow we were looking for since

$$\sum_{b_1 \in I_1} F^{(1)}_{a_1,a_2,b_1} = \sum_{b_1 \in N^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} + \sum_{b_1 \in I_1} \gamma^{(1)}_{a_1,a_2,b_1} = \mu_{a_1,a_2}$$

and

$$\sum_{a_1 \in I_1} F^{(1)}_{a_1,a_2,b_1} = \sum_{a_1 \in N^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1} + \sum_{a_1 \in I_1} \gamma^{(1)}_{a_1,a_2,b_1} = \zeta_{b_1,a_2}.$$

Similarly we can define the second cardinal flow that glues on ζ by defining the slack variable

$$\beta_{b_1,a_2} = \zeta_{b_1,a_2} - \sum_{b_2 \in N^{(2)}_{t_2}(a_2)} f^{(2)}_{a_2,b_1,b_2},$$

and the relative correction term

$$\gamma^{(2)}_{a_2,b_1,b_2} = \frac{\beta_{a_2,b_1,b_2} \left(\nu_{b_1,b_2} - \sum_{a_2 \in N^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2} \right)}{M_{b_1}}, \tag{4.76}$$

where

$$M_{b_1} = \sum_{b_2 \in I_2} \left(\nu_{b_1,b_2} - \sum_{a_2 \in N^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2} \right).$$

Conditions (4.73) and (4.74) follow from the optimality of $f^{(1)}$ and $f^{(2)}$. Indeed, owing to the definition of $F^{(1)}$ and $F^{(2)}$, we can find two scattered flows $\tilde{f}^{(1)}$ and $\tilde{f}^{(2)}$ such that

$$\begin{aligned} \tilde{f}^{(1)}_{a_1,a_2,b_1} &\geq f^{(1)}_{a_1,a_2,b_1} && (4.77)\\ \tilde{f}^{(2)}_{a_2,b_1,b_2} &\geq f^{(2)}_{a_2,b_1,b_2}, && (4.78) \end{aligned}$$

since both $\gamma^{(1)}$ and $\gamma^{(2)}$ are positive measures. Without loss of generality, assume that there exists (a_1, a_2, b_1) such that $\tilde{f}^{(1)}_{a_1,a_2,b_1} > f^{(1)}_{a_1,a_2,b_1}$, hence

$$\sum_{a_2 \in I_2} \sum_{(a_1,b_1) \in I_1} (t_1 - c^{(1)}_{a_1,b_1}) \tilde{f}^{(1)}_{a_1,a_2,b_1} > \sum_{a_2 \in I_2} \sum_{(a_1,b_1) \in I_1} (t_1 - c^{(1)}_{a_1,b_1}) f^{(1)}_{a_1,a_2,b_1}$$

which contradicts the optimality of the couple $(f^{(1)}, f^{(2)})$. To conclude the proof, notice that

$$\begin{aligned} 0 \;&\leq\; W_c(\mu,\nu) - W_{c^{(t_1,t_2)}}(\mu,\nu) \\ &\leq\; \sum_{(a_1,a_2)\in G} \sum_{b_1 \in I_1} c^{(1)}_{a_1,b_1} F^{(1)}_{a_1,a_2,b_1} + \sum_{(b_1,b_2)\in G} \sum_{x_2 \in I_2} c^{(2)}_{a_2,b_2} F^{(2)}_{a_2,b_1,b_2} \\ &\qquad + t_1 - \sum_{a_2 \in I_2} \sum_{(a_1,b_1) \in N^{(1)}_{t_1}} (t_1 - c^{(1)}_{a_1,b_1}) f^{(1)}_{a_1,a_2,b_1} \\ &\qquad + t_2 - \sum_{b_1 \in I_1} \sum_{(a_2,b_2) \in N^{(2)}_{t_2}} (t_2 - c^{(2)}_{a_2,b_2}) f^{(2)}_{a_2,b_1,b_2} \\ &=\; \sum_{a_2 \in I_2} \left(\sum_{a_1 \in I_1, b_1 \in I_1} c^{(1)}_{a_1,b_1} F^{(1)}_{a_1,a_2,b_1} - \sum_{(a_1,b_1) \in N^{(1)}_{t_1}} c^{(1)}_{a_1,b_1} f^{(1)}_{a_1,a_2,b_1} \right) \\ &\qquad + \sum_{b_1 \in I_1} \left(\sum_{a_2 \in I_2, b_2 \in I_2} c^{(2)}_{a_2,b_2} F^{(2)}_{a_2,b_1,b_2} - \sum_{(a_2,b_2) \in N^{(2)}_{t_2}} c^{(2)}_{a_2,b_2} f^{(2)}_{a_2,b_1,b_2} \right) \\ &\qquad + t_1 \left(1 - \sum_{a_2 \in I_2} \sum_{(a_1,b_1) \in N^{(1)}_{t_1}} f^{(1)}_{a_1,a_2,b_1} \right) + t_2 \left(1 - \sum_{b_1 \in I_1} \sum_{(a_2,b_2) \in N^{(2)}_{t_2}} f^{(2)}_{a_2,b_1,b_2} \right) \\ &=\; \sum_{a_2 \in I_2} \left(\sum_{(a_1,b_1) \notin N^{(1)}_{t_1}} c^{(1)}_{a_1,b_1} F^{(1)}_{a_1,a_2,b_1} \right) + \sum_{b_1 \in I_1} \left(\sum_{(a_2,b_2) \notin N^{(2)}_{t_2}} c^{(2)}_{a_2,b_2} F^{(2)}_{a_2,b_1,b_2} \right) \end{aligned}$$

$$
\begin{aligned}
&+t_1\Big(\sum_{a_2\in I_2}\sum_{(a_1,b_1)\notin N^{(1)}_{t_1}} F^{(1)}_{a_1,a_2,b_1}\Big)+t_2\Big(\sum_{b_1\in I_1}\sum_{(a_2,b_2)\notin N^{(2)}_{t_2}} F^{(2)}_{a_2,b_1,b_2}\Big)\\
= &\sum_{a_2\in I_2}\Big(\sum_{(a_1,b_1)\notin N^{(1)}_{t_1}}(c^{(1)}_{a_1,b_1}-t_1)F^{(1)}_{a_1,a_2,b_1}\Big)\\
&+\sum_{b_1\in I_1}\Big(\sum_{(a_2,b_2)\notin N^{(2)}_{t_2}}(c^{(2)}_{a_2,b_2}-t_2)F^{(2)}_{a_2,b_1,b_2}\Big).
\end{aligned}
$$

By definition, we have

$$F^{(1)}_{a_1,a_2,b_1}=\gamma^{(1)}_{a_1,a_2,b_1}$$

for all $(a_1,b_1)\notin N^{(1)}_{t_1}$ and all $a_2\in I_2$ and, similarly,

$$F^{(2)}_{a_2,b_1,b_2}=\gamma^{(2)}_{a_2,b_1,b_2}$$

for all $(a_2,b_2)\notin N^{(1)}_{t_1}$ and all $b_1\in I_1$. Hence we can conclude

$$
\begin{aligned}
|W_c(\mu,\nu)-W_{c^{(t_1,t_2)}}(\mu,\nu)| \;\le\; &\sum_{a_2\in I_2}\sum_{(a_1,b_1)\notin N^{(1)}_{t_1}}(c^{(1)}_{a_1,b_1}-t_1)\gamma^{(1)}_{a_1,a_2,b_1}\\
&+\sum_{b_1\in I_1}\sum_{(a_2,b_2)\notin N^{(2)}_{t_2}}(c^{(2)}_{a_2,b_2}-t_2)\gamma^{(2)}_{a_2,b_1,b_2}
\end{aligned}
$$

and, remembering the relations (4.75) and (4.76) we find (4.69). □

Owing to Theorem 4.4 we can deduce two useful corollaries. The first gives us a bound on the relative error and it is obtained by simply remembering that $W_{c^{(t_1,t_2)}}(\mu,\nu)\le W_c(\mu,\nu)$ for any couple $\mu,\nu\in\mathcal{P}(G)$. The second establishes a criterion to understand when the optimal scattered flow is also an optimal cardinal flow.

Corollary 4.4 (Bound on the Relative Error)**.** *Given any couple of positive thresholds* (t_1,t_2)*, it holds true that*

$$\frac{|W_c(\mu,\nu)-W_{c^{(t_1,t_2)}}(\mu,\nu)|}{|W_c(\mu,\nu)|}\le\frac{\mathcal{E}_{c^{(t_1,t_2)}}(\mu,\nu)}{|W_{c^{(t_1,t_2)}}(\mu,\nu)|},\tag{4.79}$$

where $\mathcal{E}_{c^{(t_1,t_2)}}(\mu,\nu)$ *is the right-hand side of equation* (4.69).

Remark 4.9. *The bound granted by Theorem 4.4 is unpractical to compute. Fortunately, it is possible to simplify it by dropping a bit of accuracy. In fact, since*

$$\frac{\mu_{a_1,a_2} - \sum_{b_1 \in N^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1}}{\sum_{a_1 \in I_1} \left(\mu_{a_1,a_2} - \sum_{b_1 \in N^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1}\right)} \leq 1$$

and

$$\frac{\nu_{b_1,b_2} - \sum_{a_2 \in N^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2}}{\sum_{b_2 \in I_2} \left(\nu_{b_1,b_2} - \sum_{a_2 \in N^{(2)}_{t_2}(b_2)} f^{(2)}_{a_2,b_1,b_2}\right)} \leq 1$$

we can simplify the bound in (4.69) *and obtain*

$$\begin{aligned}
&|W_c(\mu,\nu) - W_{c^{(t_1,t_2)}}(\mu,\nu)| \\
&\quad \leq \sum_{\substack{a_2 \in I_2, \\ (a_1,b_1) \notin N^{(1)}_{t_1}}} (c^{(1)}_{a_1,b_1} - t_1)\left(\zeta_{b_1,a_2} - \sum_{a_1 \in N^{(1)}_{t_1}(b_1)} f^{(1)}_{a_1,a_2,b_1}\right) \\
&\quad + \sum_{\substack{b_1 \in I_1, \\ (a_2,b_2) \notin N^{(2)}_{t_2}}} (c^{(2)}_{a_2,b_2} - t_2)\left(\zeta_{b_1,a_2} - \sum_{b_2 \in N^{(2)}_{t_2}(a_2)} f^{(2)}_{a_2,b_1,b_2}\right) \\
&\quad = \sum_{b_1 \in I_1} (c^{(1)}_{a_1,b_1} - t_1)\left(\nu^{(1)}_{b_1} - \sum_{\substack{a_2 \in I_2, \\ a_1 \in N^{(1)}_{t_1}(b_1)}} f^{(1)}_{a_1,a_2,b_1}\right) \\
&\quad + \sum_{a_2 \in I_2} (c^{(2)}_{a_2,b_2} - t_2)\left(\mu^{(2)}_{a_2} - \sum_{\substack{b_1 \in I_1, \\ b_2 \in N^{(2)}_{t_2}(a_2)}} f^{(2)}_{a_2,b_1,b_2}\right)
\end{aligned}$$

which depends only on the initial data and on the computed solutions. We can thus find the following result.

Corollary 4.5. *Let $c = c^{(1)} + c^{(2)}$ be a separable cost function and (t_1, t_2) a couple of positive parameters. Given two probability measures μ and ν, let $(F^{(1)}, F^{(2)}) := \left(\{f^{(1)}_{a_1,a_2,b_1}\}, \{f^{(2)}_{b_1,a_2,b_2}\}\right)$ be the optimal scattered flow between them. If*

$$\sum_{(a_1,a_2) \in G, b_1 \in O^{(1)}_{t_1}(a_1)} f^{(1)}_{a_1,a_2,b_1} = \sum_{(b_1,b_2) \in G, a_2 \in I^{(2)}_{t_2}(b_2)} f^{(2)}_{b_1,a_2,b_2} = 1, \tag{4.80}$$

then

$$W_c(\mu,\nu) = t_1 + t_2 - \mathbb{B}\big((F^{(1)}, F^{(2)})\big).$$

4.4 Minimal Saturation Threshold

Since the transportation problem is linear, a classical result of convex optimization tells us that the minimal solution of the problem lies in an extremal point of the set of feasible flows. In the discrete setting, this means that the matrix describing the optimal transportation plan has a huge amount of null entries, i.e. many arcs of the bipartite graph are left unused.
As we saw in the previous section, it is reasonable to assume that, when possible, the more expensive is the arc, the lesser is the likelihood that it is used. This phenomenon can be related to the truncated cost function by searching the lowest threshold at which the two transportation problems, the truncated one and the original one, have the same value. It is then interesting to study what is the minimal threshold for which this happens or, as we call it, the minimal saturation threshold.

Definition 4.14 (Minimal Saturation Threshold)**.** *Given two measures μ, ν on X and Y, respectively and a cost function c, we define the* minimal saturation threshold *as*

$$\tau_c(\mu,\nu) := \inf\left\{t > 0 \ s.t. \ \exists \eta \in \Theta^{(t)}(\mu,\nu) \ s.t. \sum_{(x,y)\in N_c^{(t)}} \eta_{x,y} = 1\right\}. \tag{4.81}$$

We notice that the set

$$T(\mu,\nu) := \left\{t > 0 \text{ s.t. } \exists \eta \in \Theta^{(t)}(\mu,\nu) \text{ s.t.} \sum_{(x,y)\in N_c^{(t)}} \eta_{x,y} = 1\right\}$$

is not empty since $C = \max_{(x,y)\in X\times Y}\{c_{x,y}\}$ belongs to it, so the inf in definition (4.81) is always well defined. The following result shows us that the value $\tau_c(\mu,\nu)$ can also be characterized through the truncated Wasserstein distances themselves.

Theorem 4.5. *Let c be a cost function and t a threshold. Given any two probability measures μ and ν. It holds true that*

$$\tau_c(\mu,\nu) = \inf\big\{t > 0 \ s.t. \ W_{c^{(t)}}(\mu,\nu) = W_c(\mu,\nu)\big\}. \tag{4.82}$$

In particular

$$\tau_c(\mu,\nu) \le \mathbb{T}_c^{(\infty)}(\pi^*)$$

for any $\pi^* \in \Pi(\mu,\nu)$ that minimizes $\mathbb{T}_c$.

Remark 4.10. *We notice that the infimum in (4.82) is finite since the set is not empty, again, the value* $C = \max_{(x,y)\in X\times Y}\{c_{x,y}\}$ *belongs to it. We can then take a sequence of values* t_n *decreasing to the infimum of the set, that we call* $\bar{t}$ *for simplicity. It is straightforward to see that the sequence of functions* $\{c^{(t_n)}\}$ *uniformly converge to* $c^{(\bar{t})}$ *on* $X \times Y$*. In particular, if* $\pi \in \Pi(\mu,\nu)$ *it holds true that*

$$\sum_{(x,y)\in X\times Y} c^{(t_n)}_{x,y}\pi_{x,y} \to \sum_{(x,y)\in X\times Y} c^{(\bar{t})}_{x,y}\pi_{x,y} \tag{4.83}$$

for $n \to \infty$.
If we choose $\tilde{\pi} \in \Pi(\mu,\nu)$ *optimal for the Wasserstein problem with ground distance* $c^{(\bar{t})}$*, the relation* (4.83) *reads as*

$$\sum_{(x,y)\in X\times Y} c^{(t_n)}_{x,y}\tilde{\pi}_{x,y} \to \sum_{(x,y)\in X\times Y} c^{(\bar{t})}_{x,y}\tilde{\pi}_{x,y} = W_{c^{(\bar{t})}}(\mu,\nu).$$

Since for each $n \in \mathbb{N}$*, we have that*

$$\sum_{(x,y)\in X\times Y} c^{(t_n)}_{x,y}\tilde{\pi}_{x,y} \geq W_{c^{(t_n)}}(\mu,\nu) = W_c(\mu,\nu)$$

we deduce

$$W_{c^{(\bar{t})}}(\mu,\nu) \geq W_c(\mu,\nu).$$

To conclude, notice that $c \geq c^{(\bar{t})}$ *so that* $W_c(\mu,\nu) \geq W_{c^{(\bar{t})}}(\mu,\nu)$*, hence*

$$W_c(\mu,\nu) = W_{c^{(\bar{t})}}(\mu,\nu).$$

In particular, the infimum in (4.82) *can be replaced with a minimum.*

Proof. (Theorem 4.5) The proof is divided in two steps. In the first one, we prove that, for each of those t, it holds true that

$$W_c(\mu,\nu) = W_{c^{(t)}}(\mu,\nu).$$

This allow us to conclude that

$$W_c(\mu,\nu) \leq \mathbb{T}_c(\pi) = \mathbb{T}_{c^{(t)}}(\pi) = W_{c^{(t)}}(\mu,\nu). \tag{4.84}$$

Then we conclude the proof by showing that the inequality (4.84) cannot be strict.
Given two probability measures μ and ν, let us take $t > 0$ such that there exists a maximal nearby flow whose mass is unitary. The first part of the proof follows easily from Corollary 4.3, since η is maximal and $\sum_{(x,y)\in N_c^{(t)}} \eta = 1$.

Regarding the second part, let us indicate with T the minimum in (4.82) and, by absurd, let us assume that $T \neq \tau_c(\mu, \nu)$, we can then find $s \in \mathbb{R}$ for which holds true

$$T < s < \tau_c(\mu, \nu).$$

From Corollary 4.3, we know that the optimal $\pi \in \Pi(\mu, \nu)$ according to c is also optimal for the cost function $c^{(t)}$. This means that, whenever $c_{x,y} \geq T$, $\pi_{x,y} = 0$ must hold true. Since $s > T$, we can conclude

$$c_{x,y} \geq s \quad \rightarrow \quad \pi_{x,y} = 0. \tag{4.85}$$

Finally, let us define $\eta_{x,y} = \pi_{x,y}$ for any $(x, y) \in N_c^{(s)}$. Since η is a restriction of a transportation plan, we can conclude $\eta \in \mathcal{N}^{(s)}(\mu, \nu)$ and, since $\pi_{x,y} = 0$ for any (x, y) such that $c_{x,y} \geq s$, we also have $\sum_{(x,y)\in N_c^{(s)}} \eta_{x,y} = 1$. Moreover, η is maximal since

$$\begin{aligned} s - \sum_{(x,y)\in N_c^{(s)}} (s - c_{x,y})\eta_{x,y} &= s - s \sum_{(x,y)\in N_c^{(s)}} \eta + \sum_{(x,y)\in N_c^{(s)}} c_{x,y}\eta_{x,y} \\ &= \sum_{(x,y)\in N_c^{(s)}} c_{x,y}\pi_{x,y} \\ &= \sum_{(x,y)\in X\times Y} c^{(s)}_{x,y}\pi_{x,y} \\ &= \mathbb{W}_{c^{(s)}}(\mu, \nu) \end{aligned} \tag{4.86}$$

so that, according to Theorem 4.1, $\eta \in \Theta^{(t)}(\mu, \nu)$, which is absurd since $s < \tau_c$. □

Proposition 4.1. *Let c be a cost function. Given two probability measures μ and ν, if π minimizes the functional $\mathbb{T}_c$ over $\Pi(\mu, \nu)$, then it minimizes also $\mathbb{T}_{c^{(\tau_c(\mu,\nu))}}$ over the same set.*

Proof. Let π be a minimizer for $\mathbb{T}_c$. Then it holds true that

$$W_c(\mu, \nu) = \mathbb{T}_c(\pi) \geq \mathbb{T}_{c^{(\tau_c(\mu,\nu))}}(\pi) \geq W_{c^{(\tau_c(\mu,\nu))}}(\mu, \nu) = W_c(\mu, \nu)$$

so that all the inequalities have to be equalities. □

Up to now, we saw that the nearby flow problem can be fruitfully used to compute in an alternative way the minimal transportation cost. However, this new approach to this problem opens us the possibility of approaching the computation of other mathematical entities: the projections in Wasserstein spaces and the infinite Wasserstein distance.

4.4.1 Computing Projections

A classical problem of metric spaces is the one concerning the projections. Given a subset $K \subset \mathcal{P}(X)$ and a measure $\mu \in \mathcal{P}(X)$, the projection of μ over K is the element of K that is closer to μ according to a certain distance.
When we endow $\mathcal{P}(X)$ with the Wasserstein distance, the projection operator $P_K : \mathcal{P}(X) \to K$ is defined through the following minimum problem

$$P_K[\mu] := \arg\min_{\rho \in K} W_c(\mu, \rho). \tag{4.87}$$

This operator has been recently used in a splitting scheme that aims to solve density constrained equations, like the ones that describe crowded motions [55, 58]. This scheme, whose convergence has been proven in [30, 58, 64, 76], consists of two steps.
In the first one, the equation evolves freely, without considering the density constraint, for a small amount of time. To perform this step there is a plethora of methods [12, 11, 43, 60] which have solid results and theoretical backgrounds.
In the second one, the solution previously found is projected into the space of admissible ones (i.e. the ones satisfy the density constraint). The projection performed during the second step is the one induced by problem (4.87) and, unlike the first step, there is not a wide range of possible and efficient tools. One possibility is to exploit Brownian random walks, which was proposed in [30, 58]. Even if the convergence of the method is not proven, the results presented in [30, 14] are convincing.
For a more complete review of the whole topic we refer to [77].

Remark 4.11. *If we do not make any request on the set K we might have issues with the minimization problem* (4.87)*. In particular, a solution might not exist or might not be unique. Classical results [75] guarantee us that if K is both convex and closed (according to the distance we adopt) the projection exists and is unique.*
In what follows, we focus on sets of the form

$$K_f := \left\{ \mu \in \mathcal{P}(X) : \mu_x \leq f_x, \quad \forall x \in X \right\} \tag{4.88}$$

where f is such that

$$\sum_{x \in X} f_x \geq 1,$$

otherwise the projection does not exist.

As we pointed out in Remark 4.1, the maximum nearby flow problem is well defined even if μ and ν do not have the same mass, i.e. they are not balanced. Assuming that μ and ν have the same mass is necessary to relate the maximal nearby flow to the minimal transportation one, however, we stress that this assumption is not necessary in general.
In what follows we consider two unbalanced measures μ and f such that

$$\sum_{x\in X}\mu_x \leq \sum_{x\in X} f_x.$$

Without loss of generality, we suppose that $\mu \in \mathcal{P}(X)$.

Definition 4.15. *Let $\mu \in \mathcal{P}(X)$, $f \in \mathcal{M}(X)$ and $t > 0$ be a threshold. We say that $\eta \in \mathcal{N}^{(t)}(\mu, f)$ is a $t-$saturated flow between μ and f if*

$$\sum_{(x,y)\in N_c^{(t)}} \eta_{x,y} = \min\left\{\sum_{x\in X}\mu_x, \sum_{y\in X} f_y\right\} = 1.$$

Remark 4.12. *If we extend a $t-$saturated flow between μ and f by setting it to zero outside $N_c^{(t)}$, we find a probability measure over $X \times X$, whose marginals are*

$$\sum_{y\in X}\eta_{x,y} = \mu_x$$

and

$$\sum_{x\in X}\eta_{x,y} = \zeta_y \leq f_y.$$

Theorem 4.6. *Let $t > 0$ be a threshold, let $\mu \in \mathcal{P}(X)$ and $f \in \mathcal{M}(X)$. If η is a $t-$saturated flow between μ and f which is maximal for the nearby flow problem, then the second marginal of η,*

$$\zeta_y = \sum_{x\in I_c^{(t)}(y)} \eta_{x,y}$$

is the projection of μ on K_f.

Proof. For the sake of simplicity, we denote by η the trivial extension of the saturated maximal solution related to the parameter t introduced in Remark 4.12.
We notice that $\zeta \in \mathcal{P}(X)$ since $\zeta_y \geq 0$ and

$$\sum_{y\in X}\zeta_y = \sum_{y\in X}\sum_{x\in X}\eta_{x,y} = \sum_{x\in X}\mu_x = 1.$$

Since η is maximal, we have that η is optimal between μ and ζ, hence, from Corollary 4.3, we have that

$$W_{c^{(t)}}(\mu,\zeta) = W_c(\mu,\zeta). \tag{4.89}$$

By absurd let us assume that $P_{K_f}[\mu] = \rho \neq \zeta$. Then, by definition of projection we have

$$W_c(\mu,\zeta) > W_c(\mu,\rho) = W_{c^{(t)}}(\mu,\rho)$$

and, remembering relation (4.89), we find

$$W_{c^{(t)}}(\mu,\zeta) > W_{c^{(t)}}(\mu,\rho).$$

From Theorem 4.1 we can then find a nearby flow η^* between μ and ρ such that

$$W_{c^{(t)}}(\mu,\rho) = t - \mathbb{B}_c^{(t)}(\eta^*).$$

Since $\rho \in K_f$, η^* is a nearby flow between μ and f, which contradicts the maximality of η. $\square$

4.4.2 The Infinity Wasserstein

Given a $p > 0$ and a distance d over X, by considering d^p as a cost function, we can define the W_p distance over $\mathcal{P}(X)$. If we take the limit as $p \to +\infty$, we can define a new distance between measures, the infinity Wasserstein distance, $W^{(\infty)}$. Roughly speaking, the $W^{(\infty)}$ distance is obtained by searching for the transportation plan that has the smallest movements. This item has been widely studied in [87] and has been fruitfully applied in several fields, as branching models [54, 52], Vlasov equation [23, 22], travelling policies [20] and also astronomy [59]. Roughly speaking, the $W^{(\infty)}$ distance is obtained by searching for the transportation plan that has the smallest movements. In this section, we showcase the relationship between this distance and the minimal saturation threshold introduced in Section 4.4.

Definition 4.16 (Infinity Wasserstein Distance)**.** *Given a cost function c, the $W_c^{(\infty)}$ distance between two measures μ and ν is defined as*

$$W_c^{(\infty)}(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left[\mathbb{T}_c^{\infty}(\pi)\right]$$

where $\mathbb{T}_c^{\infty} : \Pi(\mu,\nu) \to [0,\infty)$ is defined as

$$\mathbb{T}_c^{\infty}(\pi) := \max\left\{c_{x,y} \;\; s.t. \;\; \pi_{x,y} > 0\right\}.$$

Corollary 4.6. *Given two measures μ and ν, we have that*

$$\min_{\pi \in \Pi_o(\mu,\nu)} \mathbb{T}_c^{\infty}(\pi) = \tau_c$$

where $\Pi_o(\mu, \nu)$ is the set of minimizers for the functional $\mathbb{T}_c$. By particular, we have

$$W_c^{(\infty)}(\mu, \nu) \leq \tau_c.$$

Proof. Given μ,$\nu \in \mathcal{P}(X)$, let π be a minimizer for $\mathbb{T}_c$ and let τ be the minimal saturation threshold $\tau_c(\mu, \nu)$. By Proposition 4.1, we have that

$$\sum_{(x,y)\in X\times X} c_{x,y}\pi_{x,y} = \sum_{(x,y)\in X\times X} c_{x,y}^{(\tau)}\pi_{x,y}$$

which can be rewritten as

$$\sum_{(x,y)\in X\times Y} \left(c_{x,y} - c_{x,y}^{(\tau)}\right)\pi_{x,y} = 0.$$

Since $c_{x,y} - c_{x,y}^{(\tau)} \geq 0$ and $\pi_{x,y} \geq 0$ $\forall x, y \in X$, we have that $\pi_{x,y} = 0$ for all $x, y \in X$ such that $c_{x,y} > \tau$. By definition of $\mathbb{T}_c^{\infty}$, we have

$$\mathbb{T}_c^{\infty}(\pi) \leq \tau_c.$$

By contradiction, let us suppose that

$$T := \min_{\pi \in \Pi_o(\mu,\nu)} \mathbb{T}_c^{\infty}(\pi) < \tau_c. \tag{4.90}$$

We can then find a $t \in \mathbb{R}$ such that $T < t < \tau_c$. Let us denote with π^* the transportation plan that minimizes (4.90). By definition of $\mathbb{T}_c^{\infty}$, we get that the support of π^* is a subset of $N_c^{(t)}$, hence, if we set

$$\eta_{x,y} = \pi_{x,y} \qquad \forall (x, y) \in N_c^{(t)},$$

we find an optimal nearby flow whose mass is equal to 1, which is absurd. The latter statement of the Theorem flows from the definition of $W_c^{(\infty)}$ itself. $\square$

The minimal saturation threshold is then only an upper bound of $\mathbb{W}_c^{\infty}(\mu, \nu)$. By changing the objective function to maximize in the algorithm, we can compute its actual value.

Definition 4.17 ($t-$truncated mass functional)**.** *Given a cost function c and a threshold $t > 0$, we define the $t-$truncated mass function $\mathbb{B}_\infty^{(t)} : \mathcal{N}^{(t)}(\mu, \nu) \to [0, \infty)$ as*

$$\mathbb{B}_\infty^{(t)}(\eta) := \sum_{(x,y) \in N_c^{(t)}} \eta_{x,y}. \tag{4.91}$$

Theorem 4.7. *Let us take a cost function c. Given two probability measures μ and ν on X and Y respectively, we have that*

$$W_c^\infty(\mu, \nu) = \inf \Big\{ t > 0 \ s.t. \ \exists \eta \in \mathcal{N}^{(t)}(\mu, \nu) \ s.t. \ \mathbb{B}_\infty^{(t)}(\eta) = 1 \Big\}.$$

Proof. Let $t > 0$ be such that there exists $\eta \in \mathcal{N}^{(t)}(\mu, \nu)$ for which $\mathbb{B}_\infty^{(t)}(\eta) = 1$. Then, by extending η trivially on $X \times Y$, we find $\pi \in \Pi(\mu, \nu)$ such that $\mathbb{T}_c^\infty(\pi) \leq t$. Since this is true for any t as above, we can conclude

$$\mathbb{W}_c^\infty(\mu, \nu) \leq \inf \Big\{ t > 0 \text{ s.t. } \exists \eta \in \mathcal{N}^{(t)}(\mu, \nu) \text{ s.t. } \mathbb{B}_\infty^{(t)}(\eta) = 1 \Big\}.$$

In order to conclude we must prove the inverse inequality. If we set $T = \mathbb{W}_c^\infty(\mu, \nu)$, then, for each $\epsilon > 0$, we can find a transportation plan $\pi \in \Pi(\mu, \nu)$ such that

$$\pi_{x,y} = 0 \qquad \text{if} \qquad c_{x,y} > T_\epsilon := T + \epsilon.$$

If we define

$$\eta_{x,y} = \pi_{x,y} \qquad\qquad (x, y) \in N_c^{T_\epsilon}$$

we have that

$$\mathbb{B}_\infty^{(T_\epsilon)}(\eta) = \sum_{(x,y) \in N_c^{(T_\epsilon)}} \eta_{x,y} = \sum_{(x,y) \in X \times X} \pi_{x,y} = 1$$

so that

$$\inf \Big\{ t > 0 \text{ s.t. } \exists \eta \in \mathcal{N}^{(t)}(\mu, \nu) \text{ s.t. } \mathbb{B}_\infty^{(t)}(\eta) = 1 \Big\} \leq T_\epsilon \qquad \forall \epsilon > 0,$$

which concludes the proof. □

4.4.3 An Upper Bound for the Infinity Wasserstein Distance in the Discrete Setting

Let μ and ν be two probability measures on a Lipschitz regular and bounded subset $\Omega \subset \mathbb{R}^n$. If μ is absolutely continuous with respect to the Lebesgue measure, and we endow Ω with a cost function of the form

$$c_p(\mathbf{x}, \mathbf{y}) := \left(\sqrt{\sum_{i=1}^n |x_i - y_i|^2} \right)^p, \qquad p > 1,$$

Theorem 2.10 states that the optimal transportation plan π between μ and ν is unique and it is induced by a transportation map T_p, i.e.

$$\pi = (Id, T_p)_{\#}\mu.$$

In [18], Bouchitté et al. established an L^∞ bound on the displacement map $Id - T_p$, which only depends on the shape of Ω, p and the density of μ. This estimate allowed the authors to give an upper bound on the $W^{(\infty)}$ distance between μ and ν, which reads as it follows.

Theorem 4.8 (Theorem 1.2, [18])**.** *Let Ω be a bounded connected open subset of $\mathbb{R}^n$ with Lipschitz boundary and denote by $\mathcal{P}_{ac}(\Omega)$ the set of Borel absolutely continuous probability measures on $\overline{\Omega}$. Then, for every $p > 1$, and every pair $(\mu, \nu) \in \mathcal{P}_{ac}(\Omega) \times \mathcal{P}(\overline{\Omega})$ there holds*

$$(W^{(\infty)}(\mu, \nu))^{p+n} \leq C_{p,d}(\Omega)||f^{-1}||_{L^\infty(\Omega)} W_p^p(\mu, \nu), \tag{4.92}$$

where f is the density of μ and $C_{p,n}(\Omega)$ is a positive constant depending only on p, n and Ω.

The proof of this result heavily relies on the regularity of μ, hence, when μ and ν are both discrete, this result does not apply. In particular, we are no longer able to find a constant depending only on μ and the geometry of the support of μ, as the following example shows.

Example 4.2. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ be defined as*

$$\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \qquad \nu = \frac{1-\epsilon}{2}\delta_0 + \frac{1+\epsilon}{2}\delta_1,$$

for $\epsilon \in (0, 1)$, and let us endow $\mathbb{R}$ with the cost function $c_2(x, y) = |x - y|^2$. By a simple computation we have that

$$W_2^{(\infty)}(\mu, \nu) = 1, \qquad W_2^2(\mu, \nu) = \frac{\epsilon}{2}.$$

When $\epsilon \leq \frac{1}{2}$, we have

$$\frac{1}{4} \leq \nu_y, \mu_x \leq \frac{3}{4}$$

hence the estimate (4.92) *does not hold true.*

In what follows, we prove the existence of an optimal transportation plan between two discrete measures that is induced by the action of two push-forward functions, one going from X to Y and one going from Y to X. This allows us to establish a bound on $W^{(\infty)}(\mu, \nu)$ similar to the one proved in [18].

Definition 4.18 (Trim solution)**.** *Let $\mu, \nu \in \mathcal{P}(X)$ be two measures on a discrete set X and let $c : X \times X \to \mathbb{R}$ be a cost function. A minimal solution π^* of the transportation problem* (2.10) *is said to be trim if*

$$\#spt(\pi^*) \leq \#spt(\pi)$$

for each optimal solution π.

Lemma 4.3. *Let $\pi \in \Pi(\mu, \nu)$ be a trim solution. Then each restriction of π is a trim solution for its marginals. In particular, if $\pi^{(1)}$ and $\pi^{(2)}$ are such that*

$$\pi = \pi^{(1)} + \pi^{(2)}$$

and $spt(\pi^{(1)}) \cap spt(\pi^{(2)}) = \emptyset$, then $\pi^{(1)}$ and $\pi^{(2)}$ are trim solutions for their marginals.

Proof. Let π^* be a restriction of π. By Theorem 2.12, we know that π^* is optimal between its marginals, hence we only need to prove that its support has minimal cardinality.
Arguing by contradiction, let us assume that π^* is not trim, hence there exists another optimal plan η between the marginals of π^* such that

$$\#spt(\eta) < \#spt(\pi^*).$$

We can define the measure $\hat{\pi}$ as

$$\hat{\pi} = \pi - \pi^* + \eta,$$

since $\pi \geq \pi^*$ and $\eta \geq 0$, we have $\hat{\pi} \geq 0$. Moreover, since π^* and η have the same marginals, $\hat{\pi}$ has the same marginals of π, therefore $\hat{\pi} \in \Pi(\mu, \nu)$. Moreover, since π^* and η are optimal between their marginals, we have

$$\sum_{(x,y)\in X\times X} c_{x,y}\pi^*_{x,y} = \sum_{(x,y)\in X\times X} c_{x,y}\eta_{x,y},$$

thus

$$\begin{aligned}
\sum_{(x,y)\in X\times X} c_{x,y}\hat{\pi}_{x,y} &= \sum_{(x,y)\in X\times X} c_{x,y}\pi_{x,y} - \sum_{(x,y)\in X\times X} c_{x,y}\pi^*_{x,y} \\
&\quad + \sum_{(x,y)\in X\times X} c_{x,y}\eta_{x,y} \\
&= \sum_{(x,y)\in X\times X} c_{x,y}\pi_{x,y}.
\end{aligned}$$

In particular, π and $\hat{\pi}$ have the same cost, therefore $\hat{\pi}$ is an optimal transportation plan between μ and ν.
To conclude, we notice that, since π^* is a restriction of π, we have

$$\#spt(\pi) = \#spt(\pi - \pi^*) + \#spt(\pi^*) > \#spt(\pi - \pi^*) + \#spt(\eta) \geq \#spt(\hat{\pi}),$$

which concludes the contradiction, since π is trim by hypothesis. □

Theorem 4.9. *Let X be a discrete polish space and let μ and ν be two positive measures over the set X such that*

$$\mu_a > 0 \qquad \forall a \in X,$$

$$\nu_b > 0 \qquad \forall b \in X,$$

and

$$\sum_{a \in X} \mu_a = \sum_{b \in X} \nu_b.$$

Given a cost function $c : X \times X \to \mathbb{R}$, let π be a trim solution of the transportation problem. We can then find two couples of measures $(\mu^{(d)}, \mu^{(c)})$ and $(\nu^{(d)}, \nu^{(c)})$ and a couple of functions $h^{(1)}$ and $h^{(2)}$ such that

- $\mu = \mu^{(d)} + \mu^{(c)}$ *and* $\nu = \nu^{(d)} + \nu^{(c)}$,
- $\pi = (Id, h^{(1)})_\# \mu^{(d)} + (h^{(2)}, Id)_\# \nu^{(d)}$.

Proof. We proceed by induction on the cardinality of X.
If $\#X = 1$, the thesis follows trivially.
Let us now assume that the statement holds for each couple of measures whose support has cardinality $(n-1)$ and let μ and ν be two measures supported on a set with cardinality n, namely X_n.
Given a trim solution π, it is well known (see Chapter 7 of [29]) that

$$\#spt(\pi) \leq 2n - 1.$$

Since μ and ν have n points in their support, we can find $\bar{a} \in X$ such that there exists a unique $\bar{b} \in spt(\nu)$ for which

$$\pi_{\bar{a},\bar{b}} > 0,$$

hence $\mu_{\bar{a}} = \pi_{\bar{a},\bar{b}} \leq \nu_{\bar{b}}$. Similarly, we can find $\underline{b} \in X$ such that there exists a unique $\underline{a} \in spt(\mu)$ for which

$$\pi_{\underline{a},\underline{b}} > 0,$$

so that $\nu_{\underline{b}} = \pi_{\underline{a},\underline{b}} \le \mu_{\underline{a}}$.
If $\mu_{\bar{a}} = \pi_{\bar{a},\bar{b}} = \nu_{\bar{b}}$, we can restrict the plan π to the set $spt(\pi)\backslash\{(\bar{a},\bar{b})\}$. We denote this restriction with π_*. By definition, the marginals of π_* are

$$\mu_* = \mu - \mu_{\bar{a}}\delta_{\bar{a}}$$

and

$$\nu_* = \nu - \nu_{\bar{b}}\delta_{\bar{b}}.$$

In particular, both the supports of μ_* and ν_* contain $(n-1)$ points. By induction we can find $(\mu_*^{(d)}, \mu_*^{(c)})$, $(\nu_*^{(d)}, \nu_*^{(c)})$ and $(h_*^{(1)}, h_*^{(2)})$ such that

$$\mu_* = \mu_*^{(d)} + \mu_*^{(c)},$$

$$\nu_* = \nu_*^{(d)} + \nu_*^{(c)}$$

and

$$\pi_* = (Id, h_*^{(1)})_\# \mu_*^{(d)} + +(h_*^{(2)}, Id)_\# \nu_*^{(d)}.$$

We can then define

$$\mu^{(d)} = \mu_*^{(d)} + \mu_{\bar{a}}\delta_{\bar{a}}, \qquad \mu^{(c)} = \mu_*^{(c)},$$

$$\nu^{(d)} = \nu_*^{(d)}, \qquad \nu^{(c)} = \nu_*^{(c)} + \nu_{\bar{b}}\delta_{\bar{b}},$$

and

$$h^{(1)}(a) = \begin{cases} h_*^{(1)}(a) & \text{if } a \neq \bar{a}, \\ \bar{b} & \text{otherwise.} \end{cases}, \qquad h^{(2)}(b) = h_*^{(2)}(b).$$

It easy to see that

$$\mu = \mu^{(d)} + \mu^{(c)}, \qquad \nu = \nu^{(d)} + \nu^{(c)}$$

and, since $h_\#^{(1)}\delta_{\bar{a}} = \delta_{\bar{b}}$, we have

$$\pi = (Id, h^{(1)})_\# \mu^{(d)} + (h^{(2)}, Id)_\# \nu^{(d)}, \tag{4.93}$$

which concludes the proof in the case $\mu_{\bar{a}} = \pi_{\bar{a},\bar{b}} = \nu_{\bar{b}}$. We proceed similarly if $\nu_{\underline{b}} = \pi_{\underline{a},\underline{b}} = \mu_{\underline{a}}$.
To conclude, consider the case in which $\mu_{\bar{a}} = \pi_{\bar{a},\bar{b}} < \nu_{\bar{b}}$ and $\nu_{\underline{b}} = \pi_{\underline{a},\underline{b}} < \mu_{\underline{a}}$. In this case, we restrict π to the set $spt(\pi)\backslash\{(\bar{a},\bar{b}), (\underline{a},\underline{b})\}$. Let us denote again with π_* the restriction and with μ_* and ν_* its marginals. Since both μ_*

and ν_* have $(n-1)$ points in their supports, we can again decompose them as

$$\mu_* = \mu_*^{(d)} + \mu_*^{(c)}, \qquad \nu_* = \nu_*^{(d)} + \nu_*^{(c)}$$

and find a couple of functions $h_*^{(1)}, h_*^{(2)}$ for which

$$\pi_* = (Id, h_*^{(1)})_\# \mu_*^{(d)} + (h_*^{(2)}, Id)_\# \nu_*^{(d)}.$$

We can then define

$$\mu^{(d)} = \mu_*^{(d)} + \mu_{\bar{a}}\delta_{\bar{a}}, \qquad \mu^{(c)} = \mu_*^{(c)} + \mu_{\underline{a}}\delta_{\underline{a}},$$

$$\nu^{(d)} = \nu_*^{(d)}(c) + \nu_{\underline{b}}\delta_{\underline{b}}, \qquad \nu^{(c)} = \nu_*^{(c)} + \nu_{\bar{b}}\delta_{\bar{b}},$$

and

$$h^{(1)}(a) = \begin{cases} h_*^{(1)}(a) & \text{if } a \neq \bar{a}, \\ \bar{b} & \text{otherwise.} \end{cases} \qquad h^{(2)}(b) = \begin{cases} h_*^{(2)}(b) & \text{if } b \neq \underline{b}, \\ \underline{a} & \text{otherwise.} \end{cases}$$

which concludes the thesis. □

Remark 4.13. *Theorem 2.10 states that, whenever μ is an absolutely continuous measure supported over a compact set $\Omega \subset \mathbb{R}^n$ and the cost function c is a strictly convex function of the euclidean distance, the optimal transportation plan is induced by a transportation map, regardless of the regularity of ν. When both μ and ν are discrete this result is generally false, however in Theorem 4.9, we proved that there exists at least one optimal transportation plan between two measures that can be recreated as the action of two functions, one acting from a subset $\tilde{X} \subset X$ to Y and one acting from a subset $\tilde{Y} \subset Y$ to X.*

Definition 4.19 (Diffusive Model)**.** *The decomposition ensured by Theorem 4.9 will be named a diffusive model associated with the given (trim) solution π. We call $\mu^{(d)}$ and $\nu^{(d)}$ the diffusive part of μ and ν, respectively. Similarly, we denote with $\mu^{(c)}$ and $\nu^{(c)}$ the concentrating part of μ and ν, respectively. Finally, we call $h^{(1)}$ the diffusive scheme of μ and $h^{(2)}$ the diffusive scheme of ν.*

Remark 4.14. *Given two measures as in the hypothesis of Theorem 4.9, let $\mu^{(d)}$ and $\nu^{(d)}$ be their diffusive part. Since the support of spt$(\mu^{(d)}) \subset$ spt(μ) and, similarly, spt$(\nu^{(d)}) \subset$ spt(ν), the support of the transportation plan defined by the formula* (4.93) *will have, at most, $2n$ points in its support. Thus the trim condition on the optimal transportation plan is necessary, as we are going to show in the next example.*

Example 4.3. *Let us take*

$$\mu = \frac{1}{4}\Big(\delta_{(0,0,0)} + \delta_{(1,1,0)} + \delta_{(1,0,1)} + \delta_{(0,1,1)}\Big)$$

and

$$\nu = \frac{1}{4}\Big(\delta_{(1,1,1)} + \delta_{(0,0,1)} + \delta_{(0,1,0)} + \delta_{(1,0,0)}\Big),$$

and, as a cost function, let us take the regular Euclidean distance in $\mathbb{R}^3$*, i.e.*

$$|\boldsymbol{x} - \boldsymbol{y}| := \sqrt{\sum_{i=1}^{3}(x_i - y_i)^2}.$$

It is easy to see that the plan

$$\begin{aligned}\pi \quad := \quad & \frac{1}{12}\delta_{(0,0,0)} \otimes \Big(\delta_{(1,0,0)} + \delta_{(0,1,0)} + \delta_{(0,0,1)}\Big) \\ & + \frac{1}{12}\delta_{(1,1,0)} \otimes \Big(\delta_{(0,1,0)} + \delta_{(1,0,0)} + \delta_{(1,1,1)}\Big) \\ & + \frac{1}{12}\delta_{(1,0,1)} \otimes \Big(\delta_{(1,0,0)} + \delta_{(0,0,1)} + \delta_{(1,1,1)}\Big) \\ & + \frac{1}{12}\delta_{(0,1,1)} \otimes \Big(\delta_{(0,1,0)} + \delta_{(0,0,1)} + \delta_{(1,1,1)}\Big)\end{aligned}$$

is optimal. However, according to Remark 4.14, it cannot be decomposed as in formula (4.93)*, since*

$$\#spt(\pi) = 12 > 2\#spt(\mu) = 8.$$

Remark 4.15 (Lack of Uniqueness)**.** *Given a trim solution, there might be more than one diffusive model associated with it. For example, let us take*

$$\mu = \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(1,1)}, \qquad \nu = \frac{1}{4}\delta_{(-1,1)} + \frac{3}{4}\delta_{(1,0)},$$

two discrete measures over $\mathbb{R}^2$*. As a cost function, let us take the euclidean distance*

$$c(\boldsymbol{x}, \boldsymbol{y}) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2},$$

then, the probability measure

$$\pi = \frac{1}{4}\delta_{(0,0)} \otimes \delta_{(-1,1)} + \frac{1}{4}\delta_{(0,0)} \otimes \delta_{(1,0)} + \frac{1}{2}\delta_{(1,1)} \otimes \delta_{(1,0)}$$

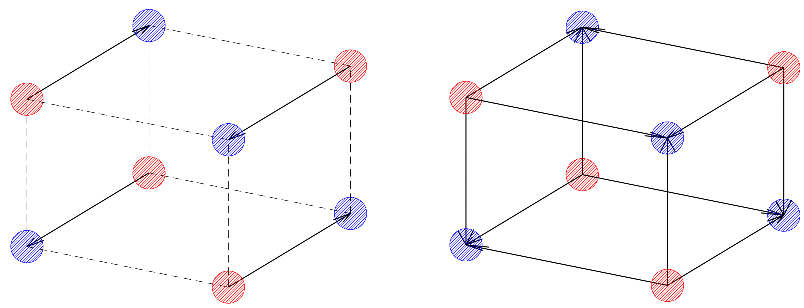


Figure 4.2: Visual comparison between an optimal trim plan (left) and an optimal plan that is not trim (right). The red points are the support of μ, the blue points the support of ν (from Example 4.3).

is a trim plan between μ and ν.
It easy to check that

$$\mu^{(d)} = \frac{1}{4}\delta_{(0,0)} + \frac{1}{2}\delta_{(1,1)}, \qquad \mu^{(c)} = \frac{1}{4}\delta_{(0,0)},$$
$$\nu^{(c)} = \frac{1}{4}\delta_{(-1,1)} + \frac{1}{2}\delta_{(1,0)}, \qquad \nu^{(d)} = \frac{1}{4}\delta_{(1,0)},$$

and

$$h^{(1)} := \begin{cases} (-1,1) & if\ x = (0,0), \\ (+1,0) & if\ x = (1,1), \\ (0,0) & otherwise, \end{cases} \qquad h^{(2)}(y) = (0,0) \quad \forall y \in \mathbb{R}^2,$$

is a decomposition of the trim plan. However, we can also decompose ν as

$$\tilde{\nu}^{(d)} = \frac{1}{4}\delta_{(-1,1)}, \qquad \tilde{\nu}^{(c)} = \frac{3}{4}\delta_{(1,0)},$$

define the functions as

$$h^{(1)}(\boldsymbol{x}) = (1,0) \quad \forall \boldsymbol{x} \in \mathbb{R}^2, \qquad h^{(2)}(\boldsymbol{y}) = (0,0) \quad \forall \boldsymbol{y} \in \mathbb{R}^2,$$

and still obtain an admissible decomposition of π.

Corollary 4.7. *Let $\mu, \nu \in \mathcal{P}(X)$ be two discrete measures, $c : X \times X \to \mathbb{R}$ a cost function, and π a trim solution of the transportation problem. Given a diffusive model for π, we have*

$$W_c(\mu,\nu) = \sum_{x\in X} c(x, h^{(1)}(x))\mu_x^{(d)} + \sum_{y\in X} c(h^{(2)}(y), y)\nu_y^{(d)}$$

and

$$\mathbb{T}_c^{(\infty)}(\pi) = \max\left\{||c(x,h^{(1)}(x))||_{L^\infty_{\mu^{(d)}}}, ||c(h^{(2)}(y),y)||_{L^\infty_{\nu^{(d)}}}\right\}.$$

In particular, we have

$$W_c(\mu,\nu) \geq \alpha W_c^{(\infty)}(\mu,\nu), \tag{4.94}$$

where

$$\alpha = \min_{a\in spt(\mu^{(d)}),b\in spt(\nu^{(d)})}\{\nu_b^{(d)},\mu_a^{(d)}\}. \tag{4.95}$$

The value α, defined in (4.95), only depends on the particular diffusive model we choose. However, since $W_c(\mu,\nu)$ and $W_c^{(\infty)}$ do not depend on the choice of the diffusive model, if we can give a lower bound on α for a particular diffusive model, we can generalize the estimate (4.94).

Corollary 4.8. *Let $\mu,\nu \in \mathcal{P}(X)$ be two discrete measures and $c : X \times X \to \mathbb{R}_+$ be a cost function. For any trim plan π, there exists a diffusive model for which*

$$\alpha \geq \min_{(A,B)\in K(\mu,\nu)}\left\{\left|\sum_{x\in A}\mu_x - \sum_{y\in B}\nu_y\right|\right\}, \tag{4.96}$$

where α is defined in relation (4.95) and

$$K(\mu,\nu) := \left\{(A,B)\subset X\times X \quad s.t. \quad \left|\sum_{x\in A}\mu_x - \sum_{y\in B}\nu_y\right| > 0\right\}.$$

Proof. Let n be the cardinality of X. Since π is trim between μ and ν, we have $\#spt(\pi) \leq 2n-1$, hence we can find $\bar{x}_1$ such that

$$\exists! \;\; \bar{y}_1 \quad s.t. \quad \pi_{\bar{x}_1,\bar{y}_1} \neq 0$$

and $\underline{y}_1$ such that

$$\exists! \;\; \underline{x}_1 \quad s.t. \quad \pi_{\underline{x}_1,\underline{y}_1} \neq 0.$$

If $\underline{x}_1 = \bar{x}_1$ (and hence $\underline{y}_1 = \bar{y}_1$), we have $\mu_{\bar{x}_1} = \nu_{\bar{y}_1}$ and we define

$$\mu^{(d)}_{\bar{x}_1} = \mu_{\bar{x}_1}, \qquad \nu^{(c)}_{\bar{y}_1} = \mu_{\bar{x}_1},$$

and

$$\mu^{(1)} := \mu - \mu_{\bar{x}_1}\delta_{\bar{x}_1}, \quad \nu^{(1)} := \nu - \nu_{\bar{y}_1}\delta_{\bar{y}_1}, \quad \pi^{(1)} = \pi - \pi_{\bar{x}_1,\bar{y}_1}\delta_{\bar{x}_1,\bar{y}_1}.$$

Otherwise, if $\underline{x}_1 \neq \bar{x}_1$ (and hence $\underline{y}_1 \neq \bar{y}_1$), we set

$$\begin{aligned} \mu^{(d)}_{\bar{x}_1} &= \mu_{\bar{x}_1}, & \mu^{(c)}_{\underline{x}_1} &= \nu_{\underline{y}_1}, \\ \nu^{(d)}_{\underline{y}_1} &= \nu_{\underline{y}_1}, & \nu^{(c)}_{\bar{y}_1} &= \mu_{\bar{x}_1}, \end{aligned}$$

and

$$\begin{aligned} \mu^{(1)} &= \mu - \mu_{\bar{x}_1}\delta_{\bar{x}_1} - \nu_{\underline{y}_1}\delta_{\underline{x}_1}, \\ \nu^{(1)} &= \nu - \nu_{\underline{y}_1}\delta_{\underline{y}_1} - \mu_{\bar{x}_1}\delta_{\bar{y}_1} \\ \pi^{(1)} &= \pi - \pi_{\bar{x}_1,\bar{y}_1}\delta_{\bar{x}_1,\bar{y}_1} - \pi_{\underline{x}_1,\underline{y}_1}\delta_{\underline{x}_1,\underline{y}_1}. \end{aligned}$$

In both cases, we find two measures, $\mu^{(1)}$ and $\nu^{(1)}$, whose support has, at most, $n-1$ points. Since $\pi^{(1)}$ is a restriction of a trim plan, by Lemma 4.3, also $\pi^{(1)}$ is trim between its marginals $\mu^{(1)}$ and $\nu^{(1)}$. Therefore we can repeat the process, finding two points $\bar{x}_2$ and $\underline{y}_2$ for which

$$\exists! \;\; \bar{y}_2 \quad s.t. \quad \pi_{\bar{x}_2,\bar{y}_2} \neq 0$$

and

$$\exists! \;\; \underline{x}_2 \quad s.t. \quad \pi_{\underline{x}_2,\underline{y}_2} \neq 0.$$

We can then extend the definition of the measures $\mu^{(d)}, \mu^{(c)}, \nu^{(d)}$, and $\nu^{(c)}$, define the measures $\mu^{(2)}$, $\nu^{(2)}$, and $\pi^{(2)}$ and start all over again.
At each step, we define two measures $\mu^{(i)}$ and $\nu^{(i)}$ and increase the cardinality of the supports of $\mu^{(d)}, \mu^{(c)}, \nu^{(d)}$, and $\nu^{(c)}$. Given any $x \in spt(\mu^{(d)})$, we can then find $i \in \{0, 1, \ldots, n-1\}$ such that

$$\mu^{(d)}_x = \mu^{(i)}_x, \tag{4.97}$$

and, similarly, for any $y \in spt(\nu^{(d)})$, we can find a $j \in \{0, 1, \ldots, n-1\}$ such that

$$\nu^{(d)}_y = \nu^{(j)}_y,$$

with the convention $\mu^{(0)} = \mu$ and $\nu^{(0)} = \nu$.
The relation between $\mu^{(i)}$ and $\mu^{(i+1)}$ is either

$$\mu^{(i+1)} = \mu^{(i)} - \mu^{(i)}_{\bar{x}_{i+1}}\delta_{\bar{x}_{i+1}}$$

or

$$\mu^{(i+1)} = \mu^{(i)} - \mu^{(i)}_{\bar{x}_{i+1}}\delta_{\bar{x}_{i+1}} - \nu^{(i)}_{\underline{y}_{i+1}}\delta_{\underline{x}_{i+1}}.$$

Similarly, we have
$$\nu^{(i+1)} = \nu^{(i)} - \nu^{(i)}_{\underline{y}_{i+1}} \delta_{\underline{y}_{i+1}}$$
or
$$\nu^{(i+1)} = \nu^{(i)} - \nu^{(i)}_{\underline{y}_{i+1}} \delta_{\underline{y}_{i+1}} - \mu^{(i)}_{\bar{x}_{i+1}} \delta_{\bar{y}_{i+1}}.$$
Similarly, we can write $\mu^{(i)}$ and $\nu^{(i)}$ as a function of $\mu^{(i-1)}$ and $\nu^{(i-1)}$, and then express $\mu^{(i+1)}$ through $\mu^{(i-1)}$ and $\nu^{(i-1)}$ as it follows
$$\mu^{(i+1)}_x = \sum_{a \in \tilde{A}_2} \mu^{(i-1)}_a - \sum_{b \in \tilde{B}_2} \nu^{(i-1)}_b, \tag{4.98}$$
where $\tilde{A}_2$ and $\tilde{B}_2$ are two subsets of X whose cardinality is at most two. By iterating this process, we are able to find
$$\mu^{(i+1)}_x = \sum_{a \in \tilde{A}_{n-(i+1)}} \mu_a - \sum_{b \in \tilde{B}_{n-(i+1)}} \nu_b, \tag{4.99}$$
where $\tilde{A}_{n-(i+1)}$ and $\tilde{B}_{n-(i+1)}$ are subsets of X, whose cardinality is $n-(i+1)$. Since the left side of (4.98) is positive, we can rewrite (4.99) as
$$\mu^{(i+1)}_x = \left| \sum_{a \in \tilde{A}_2} \mu^{(i-1)}_a - \sum_{b \in \tilde{B}_2} \nu^{(i-1)}_b \right|. \tag{4.100}$$
By taking the minimum over $K(\mu, \nu)$ of the right side in (4.100), we find
$$\mu^{(i)}_x \geq \min_{(A,B) \in K(\mu,\nu)} \left\{ \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| \right\},$$
for any $i \in \{0, 1, \ldots, n-1\}$ and each $x \in spt(\mu^{(i)})$, therefore, from relation (4.97), we get
$$\mu^{(d)} \geq \min_{(A,B) \in K(\mu,\nu)} \left\{ \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| \right\}.$$
Similarly, one can prove
$$\nu^{(d)}_y \geq \min_{(A,B) \in K(\mu,\nu)} \left\{ \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| \right\},$$
for each $y \in spt(\nu^{(d)})$, hence relation (4.96) is proven.

$\square$

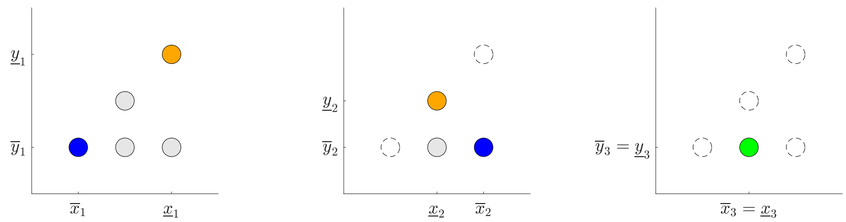


Figure 4.3: Visual description of the decomposition process used in proof of Theorem 4.9. For the sake of simplicity, the measures μ and ν are both one dimensional and have 3 points in their supports.

In Corollary 4.9, we bound $W_c^{(\infty)}$ from above with W_c. However, due to the properties of $W_c^{(\infty)}$, it is possible to relate this distance to the Wasserstein cost induced by any $p-$power of the same cost function.

Lemma 4.4. *Let $\mu, \nu \in \mathcal{P}(X)$ and let $c : X \times X \to \mathbb{R}_+$ be a cost function. Given any $p > 0$, it holds true*

$$W_{c^p}^{(\infty)}(\mu,\nu) = \big(W_c^{(\infty)}(\mu,\nu)\big)^p.$$

Proof. Let $\pi \in \Pi(\mu,\nu)$ be a plan such that

$$T_c(\pi) = W_c^{(\infty)}(\mu,\nu),$$

then

$$W_{c^p}^{(\infty)}(\mu,\nu) \leq T_{c^p}(\pi) = T_c(\pi)^p = \big(W_c^{(\infty)}(\mu,\nu)\big)^p.$$

Similarly, one can prove $\big(W_c^{(\infty)}(\mu,\nu)\big)^p \leq W_{c^p}^{(\infty)}(\mu,\nu)$ and conclude the thesis. $\square$

Thanks to the latter Lemma, we are able to prove the following result.

Theorem 4.10. *Given a cost function $c : X \times Y \to [0,\infty)$, let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ be two discrete measures. For any $p \geq 1$,*

$$W_c^{(\infty)}(\mu,\nu) \leq \frac{W_{c_p}(\mu,\nu)}{(\alpha_p)^{\frac{1}{p}}}, \tag{4.101}$$

where α_p is the constant defined in (4.95).

Proof. Given a $p \geq 1$, let us denote with $\pi^{(p)}$ the trim optimal transportation plan between μ and ν according to the cost function c_p. Given a diffusive model for $\pi^{(p)}$, we denote with α_p the constant defined in (4.95).
From Lemma 4.4 we have

$$W_{c_p}^{(\infty)}(\mu,\nu) = (W_c^{(\infty)}(\mu,\nu))^p,$$

hence, for any p, we have

$$(W_c^{(\infty)}(\mu,\nu))^p = W_{c_p}^{(\infty)}(\mu,\nu) \leq \frac{W_{c_p}^p(\mu,\nu)}{\alpha_p},$$

i.e.,

$$W_c^{(\infty)}(\mu,\nu) \leq \frac{W_{c_p}(\mu,\nu)}{(\alpha_p)^{\frac{1}{p}}}.$$

□

Remark 4.16. *In particular, since the constant α from Corollary 4.8 bounds from below every α_p and does not depend on the cost function but only on the starting measures μ and ν, we have*

$$W_c^{(\infty)}(\mu,\nu) \leq \frac{W_{c_p}(\mu,\nu)}{(\alpha)^{\frac{1}{p}}}$$

for any $p \geq 1$. In particular, if we take

$$c(\boldsymbol{x},\boldsymbol{y}) := \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2},$$

we recover the bound proposed in Theorem 4.8 for discrete measures.

Remark 4.17. *The estimate in* (4.101) *is sharp. To prove it, let us take*

$$\mu = \delta_a \qquad \textit{and} \qquad \nu = \delta_b$$

where $a, b \in \mathbb{R}^n$. By definition (4.95)*, we have $\alpha = 1$. Moreover, it is easy to see that*

$$W^{(\infty)}(\mu,\nu) = |a-b| \qquad \textit{and} \qquad W_p(\mu,\nu) = |a-b|,$$

which proves the sharpness of inequality (4.94).

Algorithm 1 Climbing Algorithm

Input: μ, ν, c

$t_0 \leftarrow 1$
$\Delta m \leftarrow 1$

while $\Delta m \geq \delta$ **do**
  $\eta \leftarrow$ Maximize functional $\mathbb{B}_c^{(t_i)}$
  $\Delta m \leftarrow 1 - \sum \eta$
  $t_{i+1} \leftarrow t_i + \Delta_t^{(i)}$
end while
return $t_{i+1} - \mathbb{B}_c^{(t_i)}(\eta)$

Output: $W_c(\mu, \nu)$

4.5 The Climbing Algorithm

As shown in Subsection 4.4.3, the optimal transportation plan utilizes a very small portion of the arcs of the bipartite graph. This is one of the main issues of the classical formulation showcased in Subsection 2.3, most of the arcs are not used, and hence there is a waste of memory.

By considering the truncated cost function, we wipe away a lot of arcs, making the whole computation faster, since we are solving the problem on a sub-graph of the bipartite one.

In this section, we exploit these simpler structures and the estimation introduced in Section 4.2. We define an iterative algorithm, the Climbing algorithm. The key idea behind this algorithm is to iteratively increase the threshold at which the ground distance is truncated. After solving the nearby flow problem, we check, through the bounds presented in Corollary 4.2, how accurate the solution is. Eventually, the threshold becomes greater than the minimal saturation threshold, hence this method takes a finite number of iterations to reach any given precision.

We will use this algorithm to compute both the infinity Wasserstein distance and the projections over sets of the form (4.88).

In Algorithm 1 , we sketch the pseudocode of the climbing algorithm. where δ is a pre-fixed tolerance on untraveled mass Δm.

We notice that if $t \geq C := \max_{(x,y)\in X\times Y} c_{x,y}$ we have $c^{(t)} = c$, so that $W_{c^{(t)}}(\mu, \nu) = W_c(\mu, \nu)$, hence this algorithm terminates after a finite number of steps, $\frac{C-t_0}{\epsilon}$ at the most, where ϵ is the minimum incrementation step $\Delta_t^{(i)}$.

The number of iterations needed to the algorithm to converge grows as ϵ gets smaller. On the other hand, the smaller ϵ is, the more accurate the returned value of $\tau_c(\mu,\nu)$ is. In the next theorem, we show that there exists a positive value, namely ϵ^*, for which the algorithm has its maximal precision.

Theorem 4.11. *If we set the starting threshold $t_0 = 0$ and we take a discretization step as it follows*

$$0 < \epsilon < \min_{x\in X, y\neq z, y,z\in Y} \big\{ |c(x,y) - c(x,z)| \big\} =: \epsilon^*,$$

then the final threshold t^* returned by the climbing algorithm is such that there exists only one value $k \in Im(c)$ such that $k \in (t^* - \epsilon, t^*]$.

Proof. Let us take ϵ as in the theorem. By absurd, let us assume that there exist two couples $(x,y),(x,z) \in X \times Y$ such that

$$t^* - \epsilon < c(x,y) < c(x,z) < t^* = t^*,$$

we could rewrite the relation as

$$0 < c(x,z) - c(x,y) < t^* - c(x,y) = (t^* - \epsilon - c(x,y)) + \epsilon.$$

Since $t^* - \epsilon - c(x,y) < 0$ we have that

$$0 < c(x,z) - c(x,y) < \epsilon$$

and then

$$|c(x,z) - c(x,y)| < \epsilon$$

which is absurd by definition of ϵ. □

Remark 4.18 (Climbing Algorithm for Separated Cost Function)**.** *We can adapt the climbing algorithm to the scattered flow framework. In this case, we have to check the total mass of both the flows and, if at least one of those is not equal to* 1*, we increase the threshold related to the flow with less mass. It is also possible to consider the $t-$univariate cost, so that we have only one threshold to iteratively increase at each iteration.*

Finally, notice that we can customize the climbing algorithm to make it suitable for the computation of the infinite Wasserstein distance. It suffices to change the object function to maximize in Algorithm 1, see Algorithm 2.

Algorithm 2 Climbing Algorithm for $W^{(\infty)}$

Input: μ, ν, c
$t_0 \leftarrow 1$
$\Delta m \leftarrow 1$

while $\Delta m \geq \delta$ **do**
  $\eta \leftarrow$ Maximize functional (4.91) for t_i
  $\Delta m \leftarrow 1 - \sum \eta$
  if $\Delta m \geq \delta$ **then**
    $t_{i+1} \leftarrow t_i + \Delta_t^{(i)}$
  else
    return t_i
  end if
end while
Output: $W_c^{(\infty)}(\mu, \nu)$

4.6 Computational Results

In this section we introduce the linear programming models for the Nearby Flow Problem and the Scattered Flow Problem. We compare those models and analyze the differences in memory usage, precision with respect to the threshold, and runtime. Finally, we use the climbing algorithm to compute the $W^{(\infty)}$ distance and the projection of measures over sets of the form (4.88).

4.6.1 Linear Model for the Nearby flow problem

The main consequence of Theorem 4.1 is that whenever we are using a truncated ground distance $c_{a,b}^{(t)}$, the transportation problem can be reformulated as following Maximum Nearby Flow Problem (MNF):

$$\textbf{(MNF)} \quad W_{c^{(t)}}(\mu, \nu) := t - \max \sum_{(a,b) \in N_c^{(t)}} s_{a,b} \eta_{a,b}$$
$$\text{s.t.} \quad \sum_{a \in I_c^{(t)}(b)} \eta_{a,b} \leq \nu_b, \qquad \forall b \in G_d,$$
$$\sum_{b \in O_c^{(t)}(b)} \eta_{a,b} \leq \mu_a, \qquad \forall a \in G_d,$$
$$\eta_{a,b} \geq 0 \qquad \forall (a,b) \in N_c^{(t)}.$$

Depending on the type of cost function c and on the value of the threshold parameter t, the number of variables $\eta_{a,b}$ can be reduced to a small fraction of the number of variables required by the bipartite graph. Indeed, when $t = \max_{a,b} c_{a,b}$, then the previous problem gives the optimal value of the original problem; for smaller values of t, it provides a lower bound.

We run several numerical tests with the objective of assessing the impact of the threshold value t on the gap between the value of the lower bound given by problem (MNF) with respect to the optimal solution value of problem (EMD), that is, the ratio $\frac{W_{c^{(t)}}}{W_c}$. In addition, we compute the ratio between the runtime of solving the two corresponding linear problems using the commercial solver Gurobi v8.0.

As problem instances, we use a collection of 10 different gray scale images with 32×32 pixels, which belong to the DOTmark benchmark [80]. We compute the distance between every pair of images, for a total of 45 pairs. For each pair of images, first, we solve once problem (EMD), and then, we solve problem (MNF) with the threshold value t ranging from 1 up to $\lceil 32\sqrt{2} \rceil$, increasing t by 1 each time.

Figure 1.a shows the results as function of the threshold value t over the maximum distance $C_{max} = 32\sqrt{2}$, while Figure 1.b shows the same results, but as a function of the number of the arc variables in (MNF) over (EMD), that is, the ratio $\frac{|N_C^{(t)}|}{|G_d \times G_d|}$. Note that in (EMD) we have a complete bipartite graph, while in (MNF) we have a subset of the complete arc set. Both figures show in the left y-axis the gap value, and on the right y-axis the ratio of the runtime for solving the problems.

Clearly, we can remark two main features for this type of instances. First, a small value of the threshold t permits to obtain already a gap close to zero (see Figure 1.a). However, as expected, the complexity of the problem does not really depend on the value of the threshold t, but on the number of arc variables that a given value of t implies, that is, the cardinality of the set $N_c^{(t)}$ (see Figure 1.b).

Finally, in Figure 4.5 and 4.6, we compare the bounds given in Theorem 4.2 and Corollary 4.2 with the real absolute and relative errors. As a cost function we adopted the l^1 norm. We notice that the estimation of relative error explodes for small thresholds.

4.6.2 Linear Model for the Scattered Flow Problem

The scattered flow problem is a flow problem on a subgraph of the tripartite graph.

We consider cost functions c_α induced by l^α norms, that is

$$c_\alpha(x,y) := l^\alpha(x,y) = |x_1 - y_1|^\alpha + |x_2 - y_2|^\alpha, \tag{4.102}$$

where $\alpha > 0$. Therefore, given two thresholds t_1 and t_2, the functional to be maximized is

$$\begin{aligned} SF_\alpha : (f^{(1)}, f^{(2)}, z) \quad \to \quad & \sum_{j \in I_2} \left(\sum_{(i,k) \in N^{(1)}_{t_1}} (t_1 - |i-k|^\alpha) f^{(1)}_{i,j,k} \right) \\ & + \sum_{k \in I_1} \left(\sum_{(j,l) \in N^{(2)}_{t_2}} (t_2 - |j-l|^\alpha) f^{(2)}_{k,j,l} \right) \end{aligned} \tag{4.103}$$

among all the $(f^{(1)}, f^{(2)}, z)$ such that

$$\begin{aligned}
\sum_{k \in N^{(1)}_{t_i}} f^{(1)}_{i,j,k} &\leq \mu_{i,j}, & \forall (i,j) \in I_1 \times I_2, \quad (4.104)\\
\sum_{j \in N^{(2)}(l)} f^{(2)}_{j,k,l} &\leq \nu_{k,l}, & \forall (k,l) \in I_1 \times I_2, \quad (4.105)\\
\sum_{k \in I_1} z_{k,j} &= \sum_{i \in I_1} \mu_{i,j}, & \forall j \in I_2, \quad (4.106)\\
\sum_{j \in I_2} z_{k,j} &= \sum_{l \in I_2} \nu_{k,l}, & \forall k \in I_1, \quad (4.107)\\
\sum_{i \in N^{(1)}(k)} f^{(1)}_{i,j,k} &\leq z_{k,j}, & \forall (i,j) \in I_1 \times I_2, \quad (4.108)\\
\sum_{l \in N^{(1)}(j)} f^{(2)}_{j,k,l} &\leq z_{k,j}, & \forall (k,l) \in I_1 \times I_2. \quad (4.109)
\end{aligned}$$

Conditions (4.104) and (4.105) are equivalent to the conditions defining the nearby flow and they have the same interpretation. Notice that the constraints 4.25 and 4.26 in Definition 4.11 are not linear, since they involve a maximum operator. To linearize those constraints, we introduced the auxiliary variables, $z_{k,j}$ for $(k,j) \in G_N := I_1 \times I_2$. Conditions (4.106) and (4.107) assure us the auxiliary variable $z_{k,j}$is an intermediate measure between μ and ν. Finally, conditions (4.108) and (4.109) assure that the two scattered flows can glue in a feasible way. In particular, there exists an intermediate measure that dominates the marginals of $f^{(1)}$ and $f^{(2)}$.

Remark 4.19. *The constraints* (4.106) *and* (4.107) *can be relaxed, i.e. they can be replaced with*

$$\sum_{k \in I_1} z_{k,j} \leq \sum_{i \in I_1} \mu_{i,j}, \quad \text{and} \quad \sum_{j \in I_2} z_{k,j} \leq \sum_{l \in I_2} \nu_{k,l}.$$

Since the objective function (4.103) *does not depend on* z*, this change does not affect the maximum. In our implementation, we adopt those conditions over the original ones.*

4.6.3 Comparison between the Scattered Flow and the Nearby Flow

In Section 4.3, we showed that, when we deal with measures on a grid and consider a separable cost function, we can truncate the cost function in two fashions: through the classic truncation and the bi-truncation. In this subsection, we inquire how using a method over the other affects memory usage, accuracy, and runtime.

Complexity and memory usage

As we saw in Section 4.6.1 and 4.6.2, both the problems related to those cost functions can be formulated as a linear problem on a sub-graph of the bi-partite and the tri-partite graph, respectively. As a cost function we choose the l^1 norm, i.e.

$$c_1(x, y) := |x_1 - y_1| + |x_2 - y_2|$$

and we truncate it at $t = [2, 3, 4, 5, 6, 7]$. For each of these threshold, we report the number of arcs and the percentage of arcs given by the two formulations (with respect to the complete bi-partite and tripartite graph) on images of size 32×32, 64×64, and 128×128. We report our results in Figure 4.7.
We observe a quadratic growth for the number of arcs required by the nearby flow, while for the scattered flow model it is linear. On the other hand, the percentage of arcs used by the scattered flow is greater than the percentage required by the nearby flow.
In Figure 4.8, we report the same test for various values of α.

Accuracy and Runtime

We ran our experiments on the class of Shapes in the DOTmark benchmark. For each couple of those images, we solve both the nearby flow and the scattered flow problem related to thresholds ranging from 1 to the size of the side. In Figure 4.9, we report the relative error we commit with the two methods and confront the runtimes for the l^1, l^2 and $l^{\frac{1}{2}}$ norms, when we fix the size at 32.

According to Remark 4.5, the scattered flow has a better accuracy over the nearby flow, for any threshold. Moreover, we observe that the difference in accuracy gets greater as the threshold grows.

For small thresholds, the nearby flow problem is slightly faster to solve than the scattered flow problem. However, for greater thresholds, the scattered flow becomes faster. It is also worth of notice how the run time required by the scattered flow stabilizes for high values of the threshold.

4.6.4 Computing the Infinity Wasserstein Distance and Projections Through the Climbing Algorithm

As shown in Section 4.5, the climbing algorithm can be used to compute the $W_2^{(\infty)}$ and the projections of measures over specific subsets of $\mathcal{P}(X)$. We run several experiments to show the potential of this approach. In this section, we report the results of our tests.
First, we show how the $W_2^{(\infty)}$ distance can be used as an initial threshold for the climbing algorithm. In particular, we show that computing the $W_2^{(\infty)}$ distance among measures requires less time than performing all the iterations needed to reach the threshold $T = W_2^{(\infty)}(\mu, \nu)$.
Second, we use the climbing algorithm to compute the projections over sets of the form (4.88). We approach the problem in the unbounded and in the constrained setting. In the first setting, we expand the boundaries of the image to remove any obstacle. In the second case, we add constraints to make areas inaccessible.
We run our experiments in the following setting.

As cost function we fix the squared Euclidean distance

$$c(x, y) = (x_1 - y_1)^2 + (x_2 - y_2)^2. \tag{4.110}$$

We fix $t_0 = 1$ and, if t_i denotes the threshold used at the i^{th} step, we define the threshold t_{i+1} as

$$t_{i+1} := t_i + 2\sqrt{t_i} + 1.$$

We notice that, according to the cost function chosen, we have $N_c^{t_i} \subsetneq N_c^{t_{i+1}}$ for each $i \in \mathbb{N}$ lesser than $C_{max} := \max_{(x,y)\in X\times X} c_{x,y}$. Therefore, at each step, we are assured to solve a different nearby flow problem. We keep iterating this process until the solution $\bar{\eta}$ satisfy the condition

$$1 - \sum_{(x,y)\in N_c^{(t)}} \bar{\eta}_{x,y} \leq 0.05.$$

The $W_2^{(\infty)}$ Distance

Given two measures $\mu, \nu \in \mathcal{P}(X)$, thanks to Theorem 4.7, we are able to compute the $W_2^{(\infty)}(\mu,\nu)$ as the minimal saturation threshold of a climbing algorithm. Given t_i the threshold obtained at the i^{th}step, the linear problem to solve is

$$\begin{aligned} \max \sum_{(x,y)\in N_c^{(t_i)}} & \eta_{x,y} & \\ \sum_{y\in N_c^{(t_i)}(x)} & \eta_{x,y} \leq \mu_x, & \forall x \in X, \\ \sum_{x\in N_c^{(t_i)}(y)} & \eta_{x,y} \leq \nu_y, & \forall y \in X. \end{aligned}$$

We run our algorithm on three classes of the DOTmark: the ClassicImages, the Microscopy, and the Shapes. Each of those classes contains 10 different images, for a total of 45 comparisons. We repeat the experiment for different resolutions: 32×32, 64×64, and 128×128. For any couple of images, namely μ and ν, we compute $W_2^{(\infty)}(\mu,\nu)$. Afterwards, we run the climbing algorithm until we reach the threshold $W_2^{(\infty)}(\mu,\nu)$ (i.e. we iteratively solve the nearby flow problem for $t_i := 1^2, 2^2, \ldots, W_2^{(\infty)}(\mu,\nu)$). In Table 4.1, we compare the times required by those two processes.

Comparison with the Sinkhorn Algorithm

To conclude this section, we compare the results of the Climbing Algorithm with the results of the stabilized sparse scaling algorithm, proposed in [79].

Table 4.1: Comparison between the computational times (in seconds) of the $W_2^{(\infty)}$ and the Climbing Algorithm (C.A.) up to the threshold $T = W_2^{(\infty)}(\mu, \nu)$.

		$W_2^{(\infty)}$ Runtime		C.A. Runtime	
Size	Image	Mean	(std)	Mean	(std)
32×32	Classic	1.01	(0.70)	4.53	(1.36)
	Microscopy	1.33	(0.67)	15.10	(0.2)
	Shape	0.32	(3.33)	9.56	(6.30)
64×64	Classic	2.73	(2.71)	7.27	(4.24)
	Microscopy	0.23	(0.21)	15.42	(0.2)
	Shape	0.81	(0.56)	19.19	(8.09)

Table 4.2: Comparison between the runtimes required by the Sinkhorn Algorithm and the Climbing Algorithm (C.A.). We also report the gap between the optimal solution and the approximation obtained through the Sinkhorn Algorithm.

		C.A. Runtime		Sinkhorn Runtime		Gap	
Size	Image	Mean	(std)	Mean	(std)	Mean	(std)
32×32	Classic	0.44	(0.22)	0.77	(0.08)	0.02	(0.00)
	Shape	0.36	(0.32)	0.55	(0.10)	0.01	(0.00)
64×64	Classic	7.33	(3.69)	3.64	(0.47)	0.02	(0.01)
	Shape	6.65	(6.09)	2.82	(0.53)	0.03	(0.01)

The applied the stabilized sparse scaling algorithm includes four modifications (partially already presented in [24, 53, 78, 66]) in order to improve its numerical stability, to solve the issues related to the large kernel matrix, and to speed up the convergence of the Sinkhorn Algorithm.
We run our algorithm using the Network Simplex implemented in the Lemon C++ graph library [1], for the stabilized sparse scaling algorithm we used the implementation available on GitHub [2].
We compare them to compute the W_2^2 distance. We run the Climbing Algorithm using the $W^{(\infty)}$ distance as the initial threshold.
In Table 4.2, we report the results of our tests for the Classic and Shapes classes. The average is taken over 45 instances. We observe that for 32×32

[1]http://lemon.cs.elte.hu
[2]https://github.com/bernhard-schmitzer/optimal-transport/tree/master/v0.2.0

images, the time required by our algorithm to find the exact solution is less than the time required by the Sinkhorn Algorithm to find an approximation. Instead, for higher resolutions (even for 64×64 images) the Sinkhorn Algorithm outperforms the Climbing Algorithm. However, we want to stress that the Sinkhorn Algorithm we used is fully parallelized, while our code is not.

Projections of Measures over K_α

In this section, we compute the projection of measures over the set of measures with bounded densities, that is

$$K_\alpha := \big\{\rho \in \mathcal{P}(X), \ \big|\ \rho_x \le \alpha \ \forall x \in X\big\}, \tag{4.111}$$

where α is a positive value satisfying the condition

$$\alpha \# X \ge 1. \tag{4.112}$$

In both the settings, given a discrete $\mu \in \mathcal{P}(X)$, we compute its projection over the sets K_{α_j}, where

$$\alpha_j = p_j \cdot \max_{x \in X} \mu_x \tag{4.113}$$

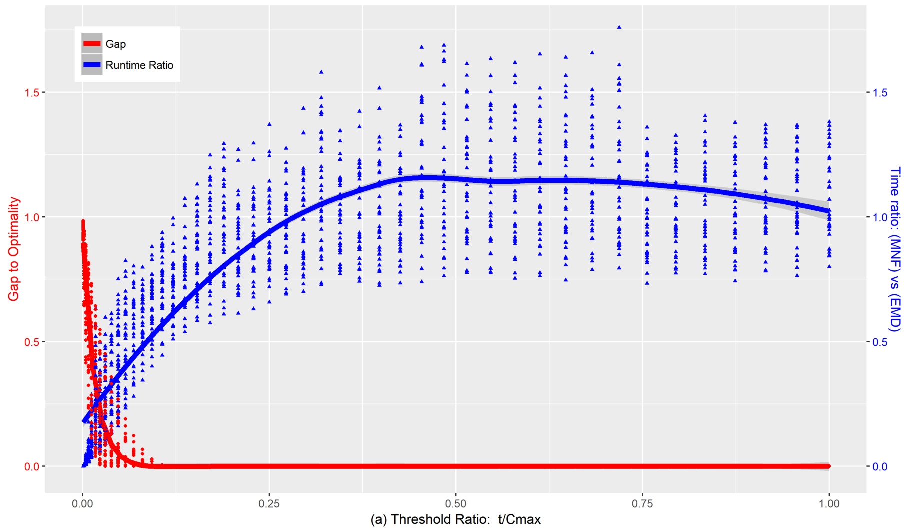
with $p_j = 1 - 0.05 \cdot j$, and $j = 0, \dots, 19$.
In the unbounded setting, we take $X = \tilde{G}$, where $\tilde{G}$ is a grid whose side has a double the length of the side of G, but the same center.
In the constrained setting, we fix $X = G' \subset G$. The set $C = G \backslash G'$ is, therefore, an inaccessible area. We can think of those areas as of obstacles (see Figure 4.10). For all of our tests, we always added the same three rectangular obstacles. A visual sample of those operations is reported in Figure 4.11.
We run our tests on three classes of the DOTmark benchmark: Shapes, ClassicImages, and MicroscopyImages. In Table 4.3, we report the mean runtimes and standard deviations in the unbounded setting. In Table 4.4, we report the mean runtimes and standard deviations for the constrained setting.
Finally, in Figure 4.12, we report the projections of three Shapes with definition 32×32. In Figure 4.13, we report the projections of three Shapes with definition 128×128 in the constrained setting.

Table 4.3: Computational times (in seconds) of the W_2 projections over the sets K_{α_j} for $j = 2, 4, 6$ in the unbounded setting.

		32×32 Runtime		64×64 Runtime	
p_j	Image Class	Mean	(std)	Mean	(std)
0.9	Classic	4.93	(0.27)	67.20	(3.72)
	Microscopy	4.96	(0.32)	79.15	(13.33)
	Shape	8.74	(0.23)	204.59	(7.72)
0.8	Classic	9.90	(0.27)	111.83	(4.54)
	Microscopy	9.62	(0.33)	126.24	(23.12)
	Shape	15.73	(0.40)	331.19	(13.86)
0.7	Classic	19.71	(0.94)	182.38	(5.77)
	Microscopy	18.61	(0.66)	180.22	(34.12)
	Shape	28.72	(0.69)	560.04	(24.21)

Table 4.4: Computational times (in seconds) of the W_2 projections over the sets K_{α_j} for $j = 2, 4, 6$ in the constrained setting.

		32×32 Runtime		64×64 Runtime	
p_j	Image Class	Mean	(std)	Mean	(std)
0.9	Classic	1.65	(0.08)	20.20	(0.61)
	Microscopy	1.56	(0.06)	19.04	(0.91)
	Shape	1.89	(0.80)	31.95	(12.80)
0.8	Classic	3.90	(0.09)	38.64	(1.04)
	Microscopy	3.81	(0.29)	33.21	(1.96)
	Shape	3.83	(1.70)	53.46	(21.27)
0.7	Classic	8.55	(0.13)	70.62	(1.90)
	Microscopy	7.48	(0.81)	53.03	(5.02)
	Shape	7.21	(2.92)	87.23	(36.05)



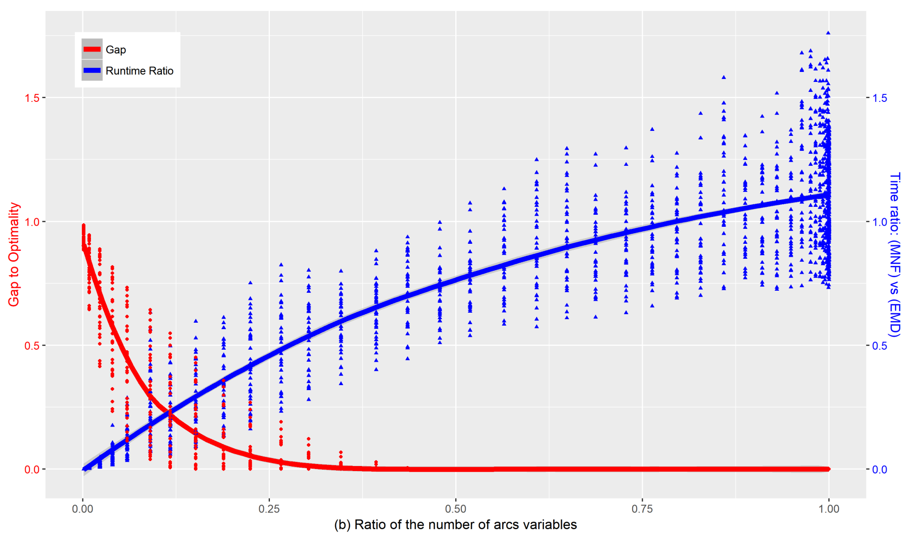


Figure 4.4: Gap to optimality (left y-scale) and time ratio between solving problem (MNF) and (EMD) (right y-scale): in Figure (1.a) as a function of the threshold $\frac{t}{C_{max}}$, where t is the distance threshold and C_{max} is the maximum distance. In Figure (1.b) as a function of the number of arc variables in problem (MNF) against the number of arc variables in problem (EMD), that is, the ratio $\frac{|N_C^{(t)}|}{|G_d \times G_d|}$.

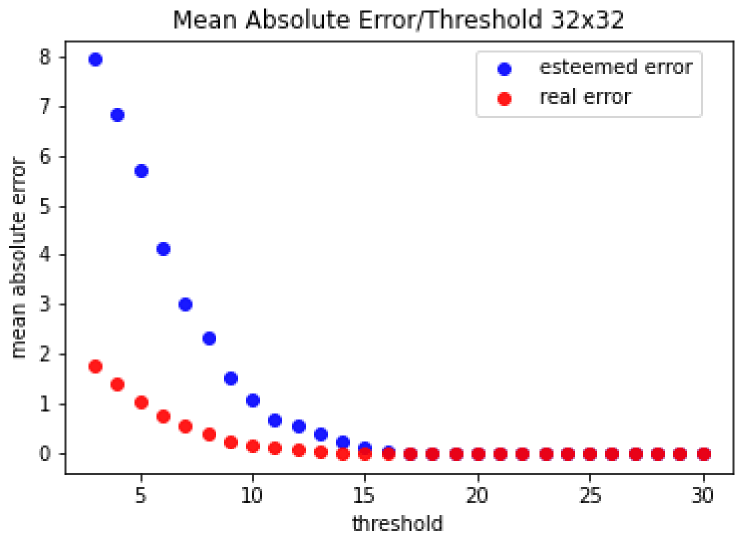


Figure 4.5: Gap between the mean absolute error predicted with formula (4.19) and the real mean absolute error. The test has been run over the Shapes class, therefore the average is taken over 45 instances.

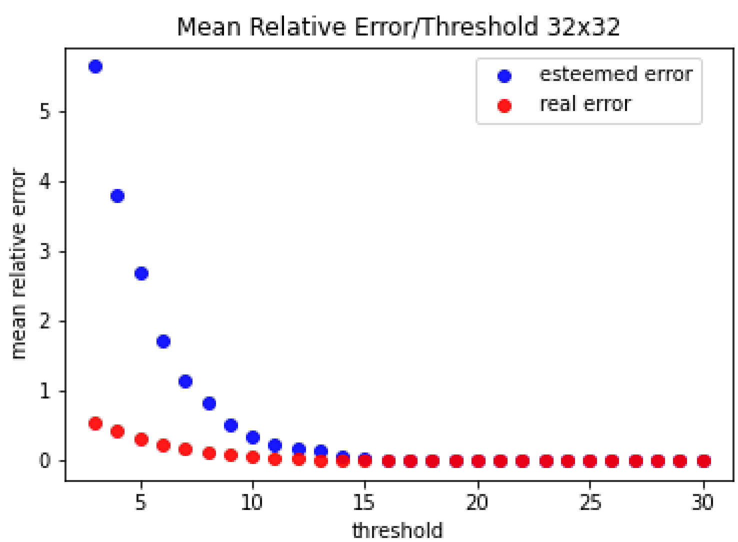


Figure 4.6: Gap between the mean relative error predicted with formula (4.20) and the real mean relative error. The test has been run over the Shapes class, therefore the average is taken over 45 instances.

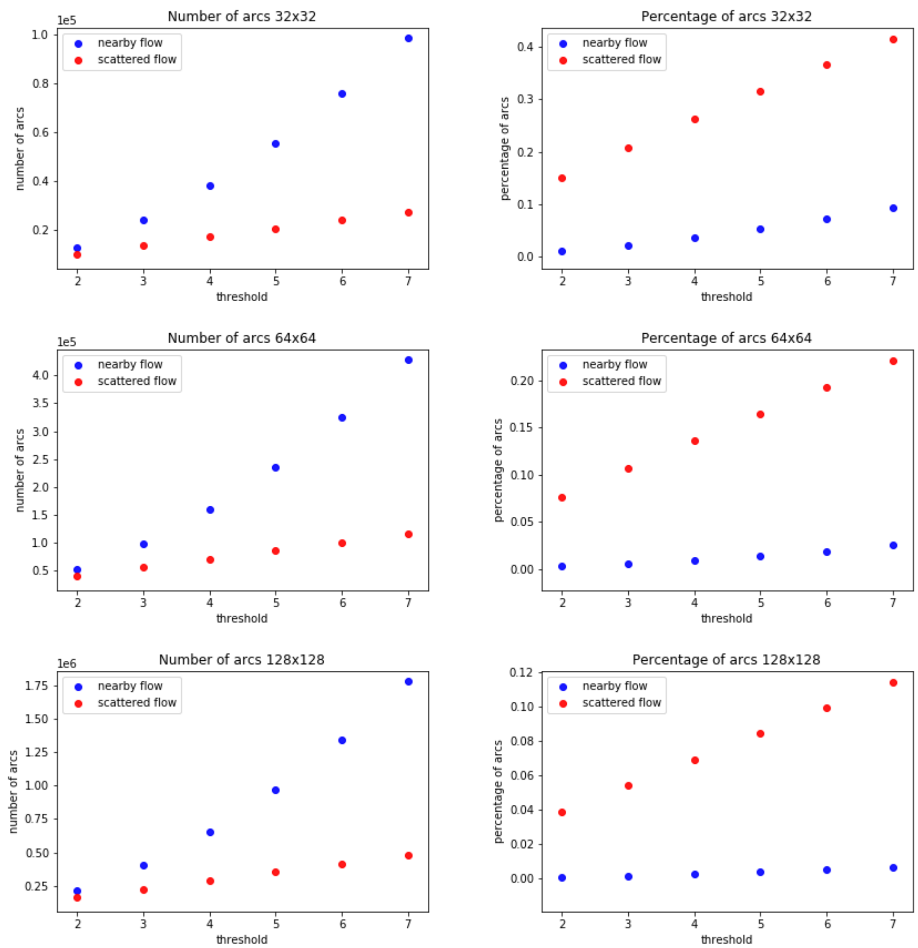


Figure 4.7: Comparison between threshold and number of arcs required by the two methods for various size of the grid. On the left, we report the raw number of arcs. On the right, the percentage.

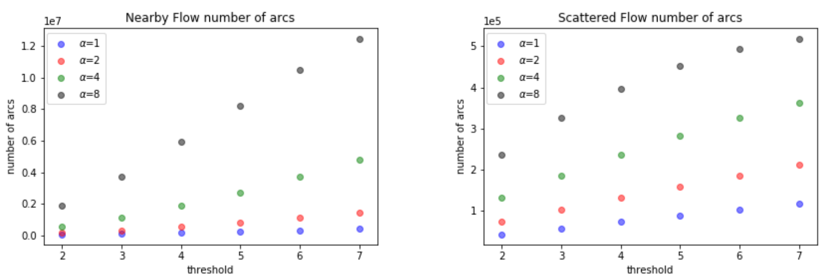


Figure 4.8: Comparison between the threshold used and the number of arcs required by the nearby flow problem and the scattered flow problem for $\alpha = [1, 2, 4, 8]$ in (4.102). For each $t_i = [2, 3, 4, 5, 6, 7]$, we report on the left the raw number of arcs for the thresholds t_i^α. On the right, we report the percentage for the thresholds t_i^α.

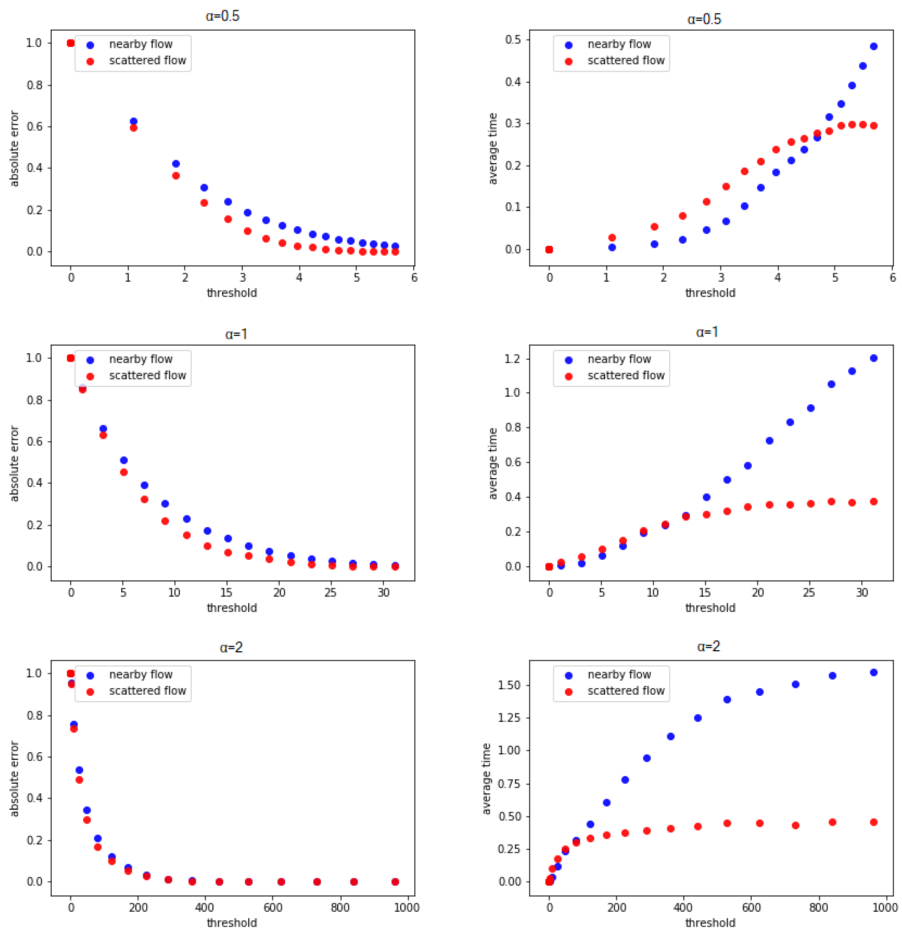


Figure 4.9: On the left we report the mean relative error of the two methods with respect to the used threshold. On the right, we report the mean runtime (in seconds).

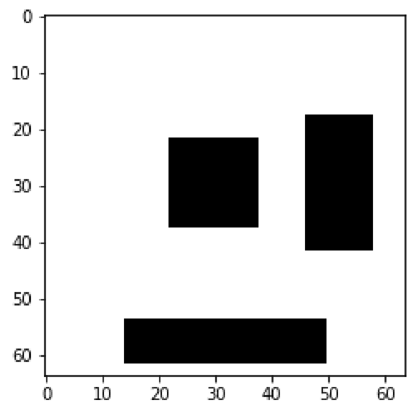


Figure 4.10: Set of obstacles (in black) for resolution 64×64.

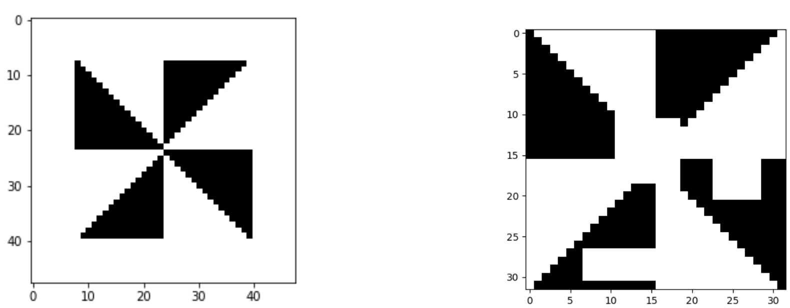


Figure 4.11: Comparison between an extended 32×32 image and the same image with the additional constraints.

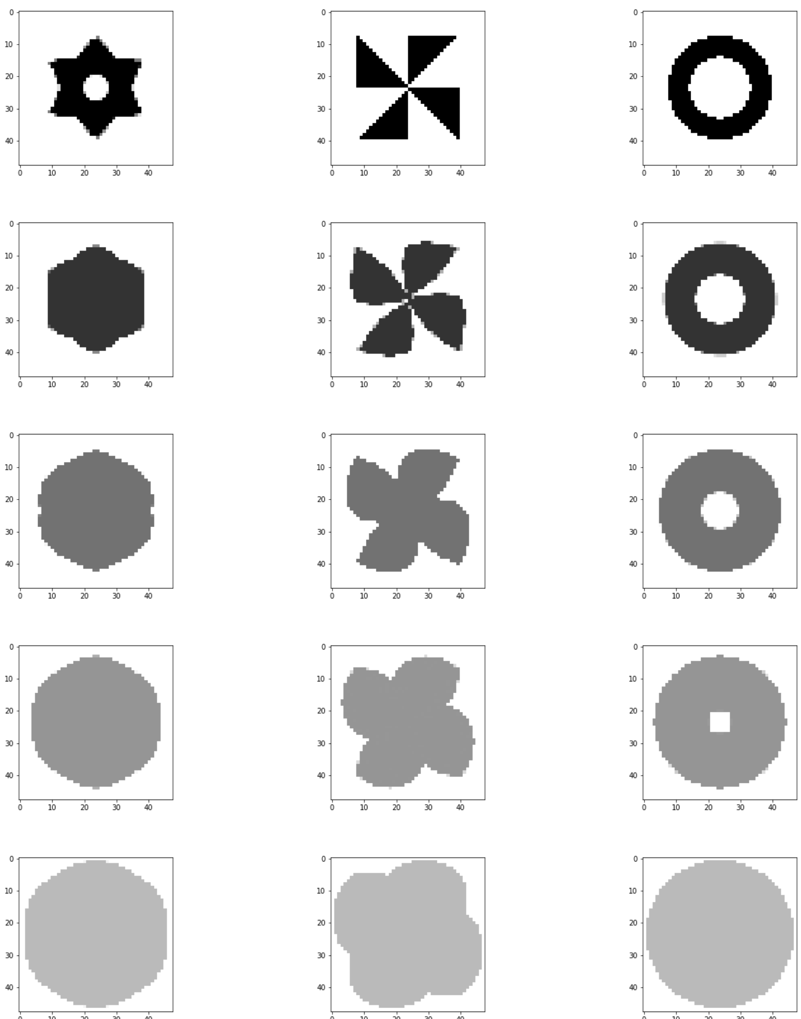


Figure 4.12: Three unbounded projections of Shapes from DOTmark benchmark with definition 32×32. We projected on the sets K_{α_j} for $j = 0, 2, 4, 6, 8$.

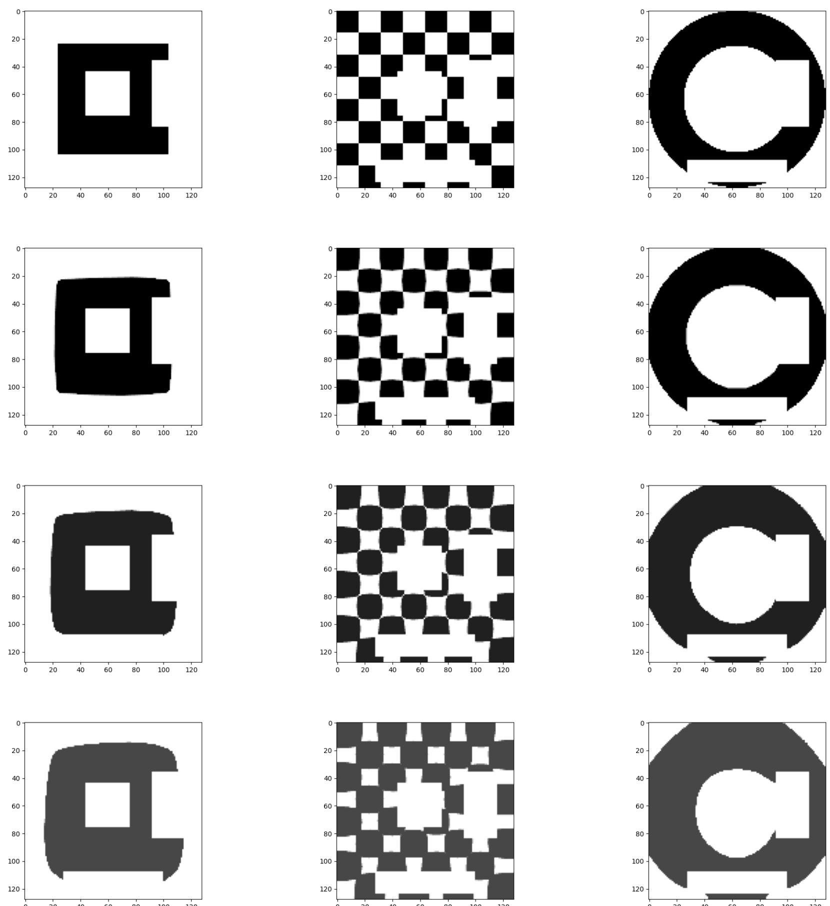


Figure 4.13: Three constrained projections Shapes of the DOTmark benchmark with definition 128×128. We projected on the sets K_{α_j} for $j = 0, 2, 4, 6$.

Bibliography

[1] Nima Aghaeepour, Greg Finak, Holger Hoos, Tim R. Mosmann, Ryan Brinkman, Raphael Gottardo, Richard H. Scheuermann, FlowCAP Consortium, DREAM Consortium, et al. Critical assessment of automated flow cytometry data analysis techniques. *Nature methods*, 10(3):228, 2013.

[2] Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network flows: Theory, Algorithms, and Applications*. Cambridge, Mass.: Alfred P. Sloan School of Management, Massachusetts Institute of Technology, 1988.

[3] Jason Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In *NIPS*, 2017.

[4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[5] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. *Proceedings of Machine Learning Research*, 70:214–223, 06–11 Aug 2017.

[6] Gennaro Auricchio, Stefano Gualandi, and Marco Veneroni. The maximum nearby flow problem. In *Advances in Optimization and Decision Science for Society, Services and Enterprises*, pages 23–33. Springer, 2019.

[7] Federico Bassetti, Antonella Bodini, and Eugenio Regazzini. On minimum Kantorovich distance estimators. *Statistics and Probability Letters*, 76(12):1298–1302, 2006.

[8] Federico Bassetti, Stefano Gualandi, and Marco Veneroni. On the computation of Kantorovich-Wasserstein distances between 2D-histograms by uncapacitated minimum cost flows. *SIAM, Journal on Optimization*, 30(3):2441–2469, 2020.

[9] Federico Bassetti and Eugenio Regazzini. Asymptotic properties and robustness of minimum dissimilarity estimators of location-scale parameters. *Society for Industrial and Applied Mathematics*, 50:312–330, 01 2005.

[10] Erhan Bayraktar and Gaoyue Guo. Strong equivalence between metrics of Wasserstein type, 2019.

[11] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 2015.

[12] Benamou, Jean-David, Carlier, Guillaume, and Laborde, Maxime. An augmented lagrangian approach to Wasserstein gradient flows and applications. *ESAIM: ProcS*, 54:1–17, 2016.

[13] Tytus Bernas, Elikplimi K. Asem, J. Paul Robinson, and Bartek Rajwa. Quadratic form: a robust metric for quantitative comparison of flow cytometric histograms. *Cytometry Part A*, 73(8):715–726, 2008.

[14] Maury Bertrand, Roudneff-Chupin Aude, Filippo Santambrogio, and Juliette Venel. Handling congestion in crowd motion modeling. *Networks and Heterogeneous Media*, 6:485, 2011.

[15] Vladimir I. Bogachev. *Measure theory*, volume 1. Springer Science & Business Media, 2007.

[16] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein Barycenters of Measures. *Journal of Mathematical Imaging and Vision*, 1(51):22–45, 2015.

[17] Nicolas Bonnotte. Unidimensional and evolution methods for optimal transportation. *Ph.D. Thesis, Université Paris-Sud*, 2013.

[18] Guy Bouchitte, Chloé Jimenez, and Rajesh Mahadevan. A new $l\infty$ estimate in optimal mass transport. *Proceedings of the American Mathematical Society*, 135:3525–3535, 11 2007.

[19] Yann Brenier. On the translocation of masses. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[20] Giuseppe Buttazzo, Aldo Pratelli, and Eugene Stepanov. Optimal pricing policies for public transportation networks. *SIAM J. Opt*, 16(3):826–853, 2006.

[21] Mathieu Carrière, Marco Cuturi, and Steve Oudot. Sliced Wasserstein kernel for persistence diagrams. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 664–673, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

[22] José A. Carrillo, Marco Di Francesco, and Corrado Lattanzio. Contractivity and asymptotics in Wasserstein metrics for viscous nonlinear scalar conservation laws. *Journal of Differential Equations*, 10(2):277–292, June 2007.

[23] José A. Carrillo, Maria Gualdani, and Giuseppe Toscani. Finite speed of propagation in porous media by mass transportation methods. *Comptes Rendus Mathematique*, 338, 05 2004.

[24] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced transport problems. *arXiv preprint arXiv:1607.05816*, 2016.

[25] Erhan Çınlar. *Probability and Stochastics*. Graduate Texts in Mathematics. Springer New York, 2011.

[26] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2017.

[27] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.

[28] Marco Cuturi and Arnaud Doucet. Fast computation of Wasserstein barycenters. *Proceedings of Machine Learning Research*, 32(2):685–693, 22–24 Jun 2014.

[29] George B. Dantzig and Mukund N. Thapa. *Linear Programming 1: Introduction*. Springer-Verlag, Berlin, Heidelberg, 1997.

[30] Simone Di Marino, Bertrand Maury, and Fillipo Santambrogio. Measure sweeping processes. *J. Convex Anal.*, 23:567–601, 01 2016.

[31] Roland Dobrushin. Vlasov equations. *Funct. Anal. Appl.*, 13(2):115–123, 1979.

[32] Merrill M. Flood. On the Hitchcock distribution problem. *Pacific Journal of Mathematics*, 3(2):369–386, 1953.

[33] Aimé Fournier. Review of a mathematical theory of large-scale atmosphere/ocean flow. *Bulletin of the American Meteorological Society*, 89:1735–, 11 2008.

[34] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya, and Tomaso A. Poggio. Learning with a Wasserstein loss. In *Advances in Neural Information Processing Systems*, pages 2053–2061, 2015.

[35] Wilfred Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[36] Andrew V. Goldberg, Éva Tardos, and Robert E. Tarjan. *Network Flow Algorithms*. Number No. 216 in CS-TR. Princeton University, Department of Computer Science, 1989.

[37] Brian. J. Hoskins. Atmospheric frontogenesis models: Some solution. *Quarterly Journal Of The Royal Meteorological Society*, 97:139–153., 01 1971.

[38] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[39] Leonid V. Kantorovich. Mathematical methods of organizing and planning production. *Management science*, 6(4):366–422, 1960.

[40] Leonid V. Kantorovich. On a problem of Monge. *Uspekhi Mat. Nauk.*, 3:225–226, 1980.

[41] Leonid V. Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133(4):1381–1382, 2006.

[42] Michael Kazhdan, Patricio Simari, Todd McNutt, Binbin Wu, Robert Jacques, Ming Chuang, and Russell Taylor. A shape relationship descriptor for radiation therapy planning. In *International Conference on*

Medical Image Computing and Computer-Assisted Intervention, pages 100–108. Springer, 2009.

[43] Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a Newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society*, 21, March 2019.

[44] Herbert Knothe. Contributions to the theory of convex bodies. *Michigan Math. J.*, 4(1):39–52, 1957.

[45] Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. Generalized sliced Wasserstein distances. In *NeurIPS*, 02 2019.

[46] Soheil Kolouri, Phillip E. Pope, Charles E. Martin, and Gustavo K. Rohde. Sliced Wasserstein auto-encoders. In *International Conference on Learning Representations*, 2019.

[47] Soheil Kolouri, Gustavo K. Rohde, and Heiko Hoffmann. Sliced Wasserstein distance for learning gaussian mixture models. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3427–3436, 2018.

[48] Soheil Kolouri, Yang Zou, and Gustavo K. Rohde. Sliced Wasserstein kernels for probability distributions. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5258–5267, 2016.

[49] Péter Kovács. Minimum-cost flow algorithms: an experimental evaluation. *Optimization Methods and Software*, 30(1):94–127, 2015.

[50] Elizaveta Levina and Peter Bickel. The Earth Mover's Distance is the Mallows distance: Some insights from statistics. *Proceedings of the IEEE International Conference on Computer Vision*, 2:251 – 256 vol.2, 02 2001.

[51] Haibin Ling and Kazunori Okada. An efficient Earth Mover's Distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):840–853, 2007.

[52] Francesco Maddalena, Sergio Solimini, and Jean-Michel Morel. A variational model of irrigation patterns. *Interfaces and Free Boundaries*, 5:391–416, 01 2003.

[53] Manish Mandad, D. Cohen-Steiner, L. Kobbelt, P. Alliez, and Mathieu Desbrun. Variance-minimizing transport plans for inter-surface mapping. *ACM Transactions on Graphics (TOG)*, 36:1 – 14, 2017.

[54] Vincent Caselles Marc Bernot and Jean-Michel Morel. Are There Infinite Irrigation Trees? *Journal of Mathematical Fluid Mechanics*, 40(8):311–332, 2006.

[55] Simone Di Marino and Alpár Richárd Mészáros. Uniqueness issues for evolution equations with density constraints. *Math. Models Methods Appl. Sci.*, 26 (09):1761–1783., 2016.

[56] John N. Mather. Minimal measures. *Commentarii Mathematici Helvetici*, 64,:375–394, 1989.

[57] John N. Mather. Action minimizing invariant measures for positive definite lagrangian systems. *Mathematische Zeitschrift*, 207(2):169–208, 1991.

[58] Bertrand Maury, Aude Roudneff-Chupin, and Santambrogio Filippo. A macroscopic crowd motion model of gradient flow type. *Mathematical Models and Methods in Applied Sciences*, 20, 02 2010.

[59] Robert J. McCann. Stable rotating binary stars and fluid in a tube. *HOUSTON JOURNAL OF MATHEMATICS*, 32(2):603–631, 2006.

[60] Quentin Mérigot. A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5):1584–1592, August 2011. 18 pages.

[61] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[62] Axel Munk and Claudia Czado. Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):223–241, 1998.

[63] Hiroshi Murata, Hiroshi; Tanaka. An inequality for certain functional of multidimensional probability distributions. *Hiroshima Math*, 4(1):75–81, 1974.

[64] Alpár R. Mészáros and Filippo Santambrogio. Advection-diffusion equations with density constraints. *Analysis and Partial Differential Equations*, 9:615–644, 03 2016.

[65] Kimia Nadjahi, Alain Durmus, Umut Simsekli, and Roland Badeau. Asymptotic guarantees for learning generative models with the sliced-Wasserstein distance. In *NeurIPS*, 06 2019.

[66] Adam M. Oberman and Yuanlong Ruan. An efficient linear programming method for optimal transportation. *arXiv: Numerical Analysis*, 2015.

[67] Darya Y. Orlova, Noah Zimmerman, Stephen Meehan, Connor Meehan, Jeffrey Waters, Eliver E.B. Ghosn, Alexander Filatenkov, Gleb A. Kolyagin, Yael Gernez, Shanel Tsuda, et al. Earth Mover's Distance (EMD): a true metric for comparing biomarker expression levels in cell populations. *PloS one*, 11(3):1–14, 2016.

[68] François-Pierre Paty and Marco Cuturi. Subspace robust Wasserstein distances. *Proceedings of Machine Learning Research*, 97:5072–5081, 09–15 Jun 2019.

[69] Ofir Pele and Michael Werman. Fast and robust Earth Mover's Distances. In *Computer vision, 2009 IEEE 12th international conference on*, pages 460–467. IEEE, 2009.

[70] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. *Scale Space and Variational Methods in Computer Vision*, pages 435–446, 2012.

[71] Murray Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Statist.*, 23(3):470–472, 09 1952.

[72] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. Metric for distributions with applications to image databases. *Proceedings of the IEEE International Conference on Computer Vision*, pages 59–66, 02 1998.

[73] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. A metric for distributions with applications to image databases. In *Computer Vision, 1998. Sixth International Conference on*, pages 59–66. IEEE, 1998.

[74] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The Earth Mover's Distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[75] Walter Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., USA, 1987.

[76] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 2015.

[77] Filippo Santambrogio. Crowd motion and evolution PDEs under density constraints. *ESAIM: ProcS*, 64:137–157, 2018.

[78] Bernhard Schmitzer. A sparse multiscale algorithm for dense optimal transport. *Journal of Mathematical Imaging and Vision*, 56(2):238–259, Oct 2016.

[79] Bernhard Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, 2019.

[80] Jörn Schrieber, Dominic Schuhmacher, and Carsten Gottschlich. DOTmark–a benchmark for discrete optimal transport. *IEEE Access*, 5:271–282, 2017.

[81] Alexander Schrijver. On the history of the transportation and maximum flow problems. *Mathematical Programming*, 91(3):437–445, 2002.

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

[83] Justin Solomon, Raif Rustamov, Leonidas J. Guibas, and Adrian Butscher. Wasserstein propagation for semi-supervised learning. *Proceedings of Machine Learning Research*, 32(1):306–314, 22–24 Jun 2014.

[84] Max Sommerfeld and Axel Munk. Inference for empirical Wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):219–238, 2018.

[85] Hiroshi Tanaka. An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a maxwellian gas. *Z. Wahrscheinlichkeitstheorie verw Gebiete*, 27:47–52, 1973.

[86] Hiroshi Tanaka. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrscheinlichkeitstheorie verw Gebiete*, 46:67–105, 1978.

[87] Petri Juutinen Thierry Champion, Luigi De Pascale. The ∞-Wasserstein Distance: Local Solutions and Existence of Optimal Transport Maps. *SIAM Journal on Mathematical Analysis*, 40(1):1–20, 2008.

[88] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[89] Jiqing Wu, Zhiwu Huang, Dinesh Acharya, Wen Li, Janine Thoma, Danda Pani Paudel, and Luc Van Gool. Sliced Wasserstein generative models. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3713–3722, 2019.