

**OPTIMAL PATHS RELATED TO DISCRETE TRANSPORT
PROBLEMS**

SHARMIN SULTANA

.

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Mathematics and Computer Science
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

© Sharmin Sultana, 2015

OPTIMAL PATHS RELATED TO DISCRETE TRANSPORT PROBLEMS

SHARMIN SULTANA

Date of Defense: August 31, 2015

Dr. Abbas Momemi Supervisor	Professor	PhD
Dr. Nathan Ng Committee Member	Professor	PhD
Dr. Amir Akbary-Majdabadno Committee Member	Professor	PhD
Dr. Hadi Kharaghani Chair, Thesis Examination Com- mittee	Professor	PhD

Dedication

To Dad and Mom

For instilling the importance of hard work and Higher education.

Abstract

In this thesis we describe Xia's results from [13] giving a solution to a general optimal transport problem. The transport problem was first proposed by Monge in the 1780's as an earth-moving problem, where the goal is to move one or more piles of soil to one or more destination points, so as to minimize the cost involved. This cost may depend on factors such as the distances involved, the weight of the piles, the time needed, and so on. A standard example to consider is the case with two source points and one destination or sink point in \mathbb{R}^2 . In this setting, Xia shows that a "Y-shaped" path can be less expensive than a "V-shaped path". More abstractly, Xia has shown that any Radon probability measure can be transported to another Radon probability measure through a general optimal transport path, which is given by a vector measure. Xia also defines a new distance function d_α on the space of probability measures and shows that this function metrizes the weak* topology of measures.

Contents

Contents	v
List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Introduction	1
1.2 Generalization of the Basic Problem	4
1.3 Transport Maps and Transference Plans	6
2 Background	10
2.1 Graph theory	10
2.2 Kirchoff's law	13
2.3 Measure Theory	15
2.4 Notation	21
3 Optimal paths related to transport problems	22
3.1 Introduction	22
3.2 Transport Paths Between Measures	23
3.3 Transportation of General Measures	42
3.4 A New Distance d_α on the Space of Probability Measures	47
Bibliography	57

List of Tables

3.1	α and corresponding $\theta_1 + \theta_2$	30
-----	--	----

List of Figures

1.1	“V” shaped path	2
1.2	“Y” shaped path	3
1.3	“T” shaped path	4
1.4	Another possible path	5
2.1	Graph	11
2.2	Directed Graph	13
2.3	Current Flow example ([8])	14
2.4	Voltage flow example ([8])	15
3.1	Transportation from two initial positions to one final destination	24
3.2	Possible paths from two sources to two sinks	25
3.3	Graph G with one cycle	35
3.4	Graph G' with no cycle	35
3.5	Graph T with cycles	36
3.6	Graph from Figure 3.5 with one cycle removed	36
3.7	Graph from Figure 3.6 with remaining cycle removed	37

Chapter 1

Introduction to Transport Problems

1.1 Introduction

In this chapter we will introduce several basic transportation problems, including the Monge and the Kantorovich optimal transport problems. The aim of the optimal transportation problem is to find an optimal way to transport one or more given objects with some weight or volume, from one or several places to another place or places while preserving the weight or volume of the objects and minimizing some cost function. The optimal transportation problem was first introduced by French mathematician Gaspard Monge in 1781, in his famous paper “Memoire sur la theorie des deblais et des remblais” [6]. In this paper he introduced the theory of “clearings and fillings”, where he tried to find the best possible way to transport a given amount of soil to fill up holes so that the total transportation cost is minimized. According to Monge the volume of the soil to be transported from the first location, must occupy the same volume in the hole after the transfer. He called the process of moving soil from the first location “clearing” and pouring the soil in the hole “filling”.

A simplified model of optimization can be found from a nicely described story about Lego pieces by Zemel in [15]. His model is a structure made of Lego blocks, which a child wants to move from one place to another for storage. The goal then is to move the pieces of the structure from one place to another, with minimal effort. This Lego model shows how we can consider Monge’s problem as an optimization question

in three-dimensional space, and also as a transportation problem.

To model a simple version of these situations from Monge or Zemel, we will start with two sources and one destination. We think of these as points of two or three dimensional space, so that we can draw graphs of them. Suppose we are shipping two items from two nearby cities x_1 and x_2 to a city y , where the distance from x_1 to y is the same as the distance from x_2 to y , the mass or volume to be shipped from each of the cities x_1 and x_2 is the same, and the amount received in the city y is the same total amount as is shipped. To find the best possible route from x_1 and x_2 to y , we can imagine several possible paths to move the source objects to the destination. As a first attempt, we can move the items directly from each of x_1 and x_2 to y , transporting the items separately, creating a V-shaped path as illustrated in Figure 1.1.

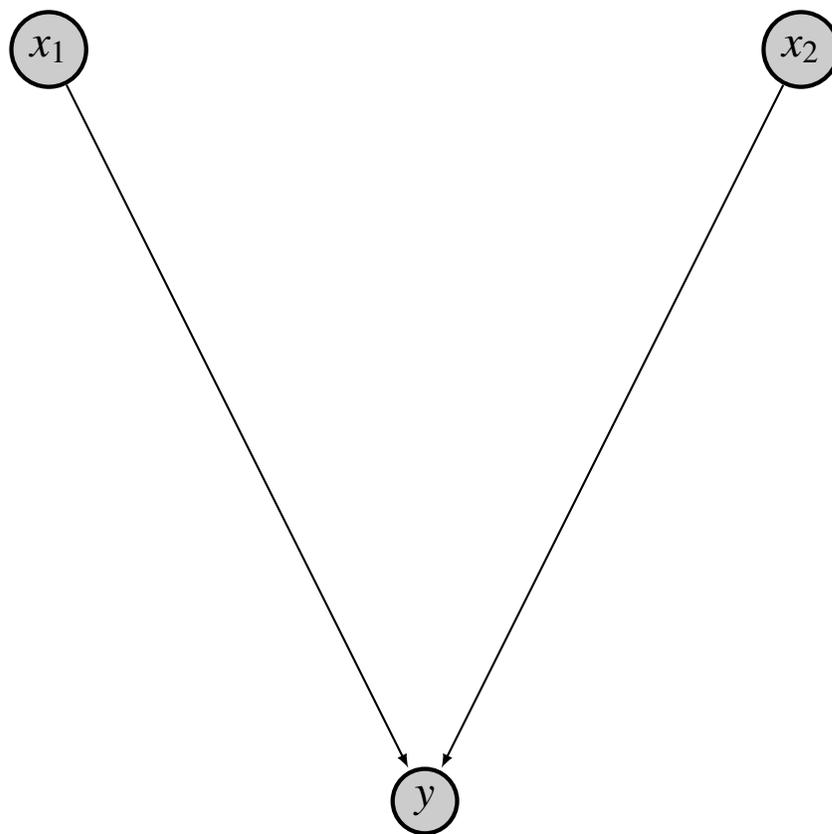


Figure 1.1: “V” shaped path

Another possible path occurs if we bring the items together at a point x_3 equidistant from x_1 and x_2 , and put the items in a truck at location x_3 and then transport them together to the destination y . In this way, it might be cheaper to transport the items in a Y-shaped path as shown in Figure 1.2, rather than a V-shaped path. The T -shaped path shown in Figure 1.3 is a special case of this. Another possible path is shown in Figure 1.4.

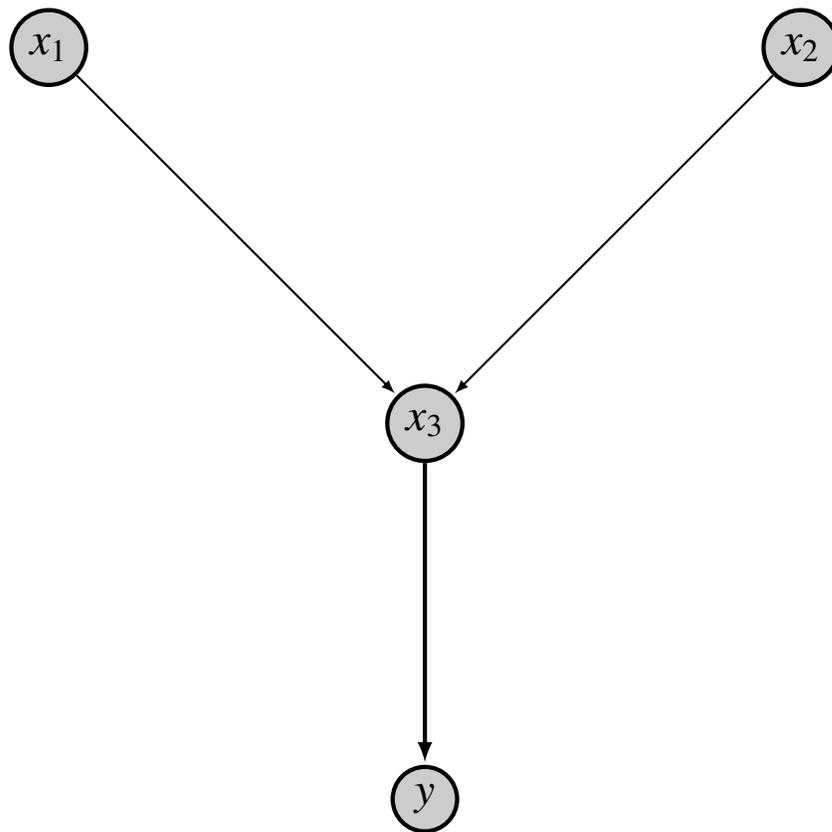


Figure 1.2: “Y” shaped path

Transport problems occur in many different fields, such as probability theory, economics and optimization ([1], [7], [3]). Many phenomena can be modelled in this way, such as circulatory systems in trees or animals, river channel networks and postal delivery systems.

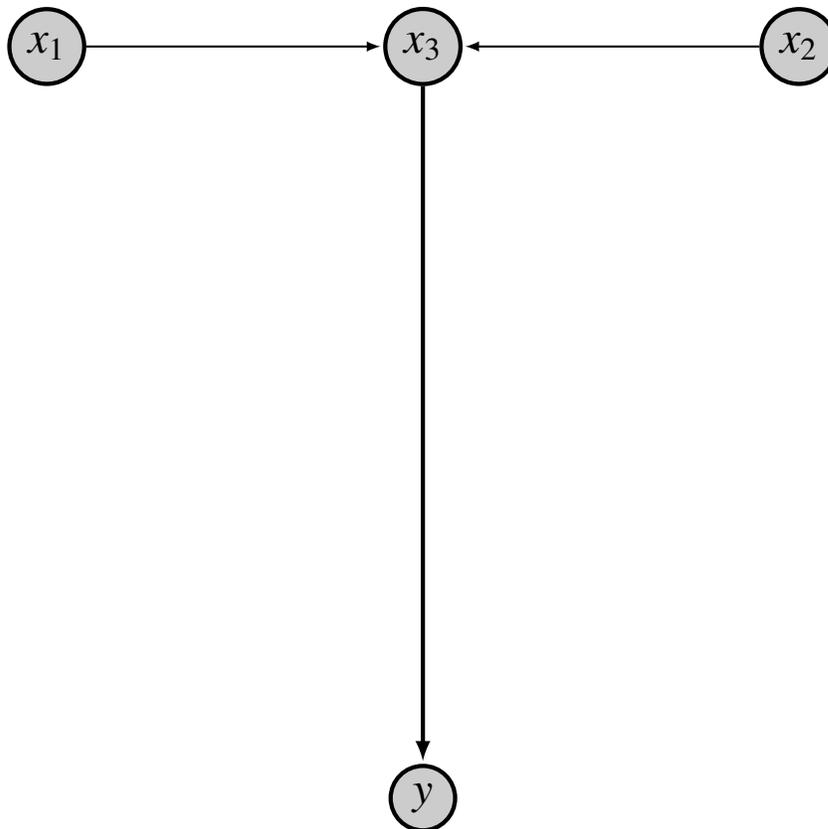


Figure 1.3: “T” shaped path

1.2 Generalization of the Basic Problem

There are several ways the basic problems described in the previous section can be generalized.

As we have seen above in Monge’s Transport problem there can be several ways to move the soil from the initial location(s) to the final destination. Consider two spaces $X \subseteq \mathbb{R}^3$ and $Y \subseteq \mathbb{R}^3$ as the initial and final destinations in the Y-shaped model shown in Figure 1.2. We can think of the model in Figure 1.1 as a directed graph, with two edges e_1 from x_1 to y and e_2 from x_2 to y . Similarly in the graph in Figure 1.3, we can consider the set X consisting of three vertices x_1, x_2 and x_3 , with three edges e_1, e_2 and e_3 connecting the vertices. The objects being moved have some mass or weight or volume, and there is some carrying cost to transport them, which needs to be defined on $X \times Y$. Let μ and ν be functions on the power sets of X and Y respectively, with $\mu(P)$

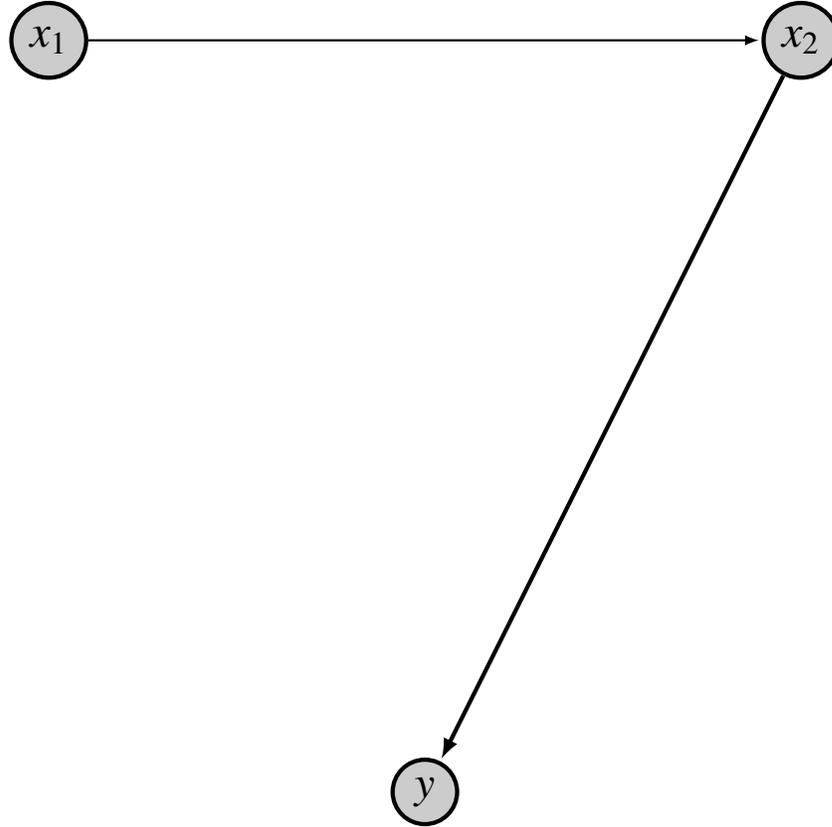


Figure 1.4: Another possible path

and $\mathbf{v}(Q)$ denoting the mass of subsets P of X and Q of Y respectively. We also denote by $c : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$ the cost function to be considered. We assume that c takes non-negative but possibly infinite values. A map that minimizes the total transport cost from the source(s) to the destination is called an optimal map. In [6], Monge considered the Euclidean distance as the cost function; that is he used $c(x,y) = |x - y|$ in \mathbb{R}^3 .

Let us now consider Zemel's Lego model from [15], with the structure X to be moved and the storage container Y as subsets of three dimensional space, so $X \subseteq \mathbb{R}^3$ and $Y \subseteq \mathbb{R}^3$. More generally, let X and Y be metric spaces. A metric space (X, d) is a space X with metric function d from $X \times X$ to \mathbb{R} , such that for any $x, y, z \in X$,

1. $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$.

2. $d(x,y) = d(y,x)$.
3. $d(x,y) \leq d(x,z) + d(z,y)$.

Suppose the structure X is made up on n Lego pieces, modelled by n points in three-space, $x_i(a_i, b_i, c_i)$, for $1 \leq i \leq n$, and a point $Y(a_0, b_0, c_0)$ as the position of the container. The distance from a Lego piece at location (a_i, b_i, c_i) to the container location $Y(a_0, b_0, c_0)$ can be defined as $d = (x_i, Y) = \sqrt{(a_i - a_0)^2 + (b_i - b_0)^2 + (c_i - c_0)^2}$. The structure of the Lego model has some mass or weight and we want to preserve that mass as we move the structure from one location to other.

There is usually a cost associated with moving the given items. This cost can depend on the distance between the points or objects, the weights to be moved, or other factors such as time or method of transportation. Suppose our aim is to move the Lego structure of n pieces to the container in the shortest possible time. In this case our cost function will be measured in units of time. Let us assume that it takes t seconds to move something from X to Y . If we can move the whole structure at once, the move will then take only t seconds. But if we move the items in two separate batches, with a return trip of our transportation vehicle needed in between, the time factor would be $3t$ instead. In general, if we use m batches to transfer all the items, we get a cost function $c : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$ with $c(x,y) = (2m - 1)t$. Of course, it would be more efficient if we can transfer all the items at once.

1.3 Transport Maps and Transference Plans

The work presented in this section is modelled on that of Zemel in [15]. In our model of transport problems, we consider sets X and Y corresponding to the source and destination locations respectively. We will now assume throughout that these sets are compact, convex subsets of a Euclidean space \mathbb{R}^m , with the usual Euclidean distance metric. We want to consider functions $T : X \rightarrow Y$ which represent the transport of objects. In addition to minimizing the cost of the transport, as given by some cost

function $c : X \times Y \longrightarrow \mathbb{R} \cup \{\infty\}$, we need a function which measures the mass of the objects being moved. Let μ be a function from the power set of X to \mathbb{R} , where $\mu(P)$ is the measure or mass assigned to P for any set $P \subseteq X$. Similarly, $\nu : P(Y) \longrightarrow \mathbb{R}$ will assign a mass $\nu(Q)$ to any subset Q of Y . Specific technical properties of these measure functions will be described in Chapter 2. The condition that the transport function T preserves mass can now be expressed by requiring $\mu(T^{-1}(D)) = \nu(D)$ for any subset D of Y . A function $T : X \longrightarrow Y$ which satisfies this condition will be called a transport map from X to Y . Zemel in [15] called ν a push-forward of μ by T , indicating this by $\nu = T\#\mu$.

Setting $D = Y$ in the requirement that $\mu(T^{-1}(D)) = \nu(D)$ gives the special case that $\mu(X) = \nu(Y)$ as a condition that any transport map must meet. We can allow the measure functions μ and ν to take on an infinite value ∞ as well, but we assume that $\mu(X)$ is always finite. The measure functions μ and ν can also be normalized, so that their values are restricted to the unit interval from 0 to 1. In this case, the Monge problem stated in [15] can be formulated as the search for a transport map $T : X \longrightarrow Y$ that minimizes $I(T) = \int_X c(x, T(x)) d\mu(X)$ over all maps T that satisfy $\nu = T\#\mu$.

Consider a transport map $T : X \rightarrow Y$ which maps each $x \in X$ to some $T(x) \in Y$. If $\{x\}$ has a positive mass, then this mass has to be moved entirely to $T(x)$. The Kantorovich problem in [5] is basically a relaxation of the Monge model, in which we allow the mass to be split up. That is, we could move the mass to be moved from one source point in X to several points in Y . Zemel in [15] construct a probability measure $\pi_x \in P(Y)$, for any $x \in X$, which describes how the mass of x is distributed in Y . Formally, we consider a measure π on the product space $X \times Y$ with marginal distributions μ on X and ν on Y , that is

$$\pi(A \times Y) = \mu(A) \quad \forall A \subseteq X \text{ measurable.} \tag{1.1}$$

and

$$\pi(X \times B) = \mathbf{v}(B) \quad \forall B \subseteq Y \text{ measurable.} \quad (1.2)$$

The relation (1.1) means that the mass $\pi(A \times B)$ transported into B equals the mass of A . The relation (1.2) means that the mass transported into B , $\pi(X \times B)$, equals the mass of B . If π satisfies these two requirements, it is called a **transference plan**.

Transference plans are a generalization of transport plans, since every transport plan in fact induces a transference plan. For example, if \mathbf{v} is the push-forward of μ by T , we can define a transference plan $\pi = (\text{id} \times T)\#\mu$ by

$$\pi(A \times B) = \mu(\{x \in A : T(x) \in B\}) = \mu(A \cap T^{-1}(B)). \quad (1.3)$$

This means that π gives measure zero to subsets of $X \times Y$ that are disjoint to the set $\{(x, T(x)) : x \in X\}$. More precisely, the equation (1.3) can be written as

$$d\pi(x, y) = d\mu(x)1\{y = T(x)\},$$

where $1\{\cdot\}$ is the indicator function. Equivalently, for any continuous and bounded function $\phi : X \times Y \rightarrow \mathbb{R}$,

$$\int_{X \times Y} \phi \, d\pi = \int_X \phi(x, T(x)) \, d\mu(x)$$

Therefore a transport map T always induces a transference plan π .

We shall denote the set of transference plans from (X, μ) to (Y, \mathbf{v}) by

$$\Pi(\mu, \mathbf{v}) \stackrel{\text{def}}{=} \{\pi \in P(X \times Y) : \forall A \pi(A \times Y) = \mu(A) \text{ and } \forall B \pi(X \times B) = \mathbf{v}(B)\},$$

where, $A \subseteq X$ and $B \subseteq Y$ run over all measurable sets. The Kantorovich problem is to

find a minimizer for

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\pi(x, y). \quad (1.4)$$

The Kantorovich problem is then a relaxation of the Monge problem as any transference map induces a transference plan that allocates I the same value where $I(T) = \int_X c(x, T(x)) d\mu(X)$ over all maps T that satisfy $\nu = T\#\mu$.

After introducing some preliminary notation and results in Chapter 2, in Chapter 3 we consider a space whose elements will actually be measure functions rather than points in \mathbb{R}^m , with a metric defined on the space based on the distribution distance between measures. Then a transport path between two elements will be a path between two atomic measures and is just a weighted directed graph which follows Kirchoff's law at each interior vertex. In general, a transport path between two arbitrary measures, is a vector measure given by a limit of some weighted directed graphs. The cost on each transport path is a suitably modified weighted mass of the vector measure. Then we describe Xia's ([13]) proof of an Existence Theorem which guarantees the existence of an optimal transport path.

In Chapter 3, we also describe Xia's work from [13] on a new distance function on the space of probability measures on a fixed convex set. Such a distance function is different from any of the Wasserstein distances but still metrizes the weak* topology on the space of probability measures. In the last section of Chapter 3, we describe Xia's proof that the space of probability measures with this new distance function becomes a length space.

Chapter 2

Background

In this chapter we review some preliminaries required to describe Xia's result in Chapter 3. These include graph theory as a way to model our transport paths, weighted directed graphs to model the weights or other costs associated with transport paths, Kirchoff's laws to describe the balance of weights coming into or out of vertices in graphs, and some basic background from measure theory.

2.1 Graph theory

As we saw in Chapter 1, we can draw pictures to represent the paths traced out in moving our source mass to the end location. One way to formalize these drawings is by graph theory. A graph consists of a set of points called vertices, along with a set of ordered pairs called edges that connect vertices. For more background in graph theory, see [9].

Definition 2.1. A graph is a structure $G = (V, E)$, where V is a set of objects called vertices and E is set of unordered pairs of vertices, whose elements are called the edges of the graph. A graph is said to be directed if any pair (u, v) is not considered the same as the pair (v, u) , but is otherwise called undirected.

Example 2.2. The graph $G = (V, E)$ in Figure 2.1 is an undirected graph with five vertices, labelled v_1 through v_5 , and five edges, labelled e_1 to e_5 .

A graph with no edges (i.e. if E is empty) is called an empty graph. A graph

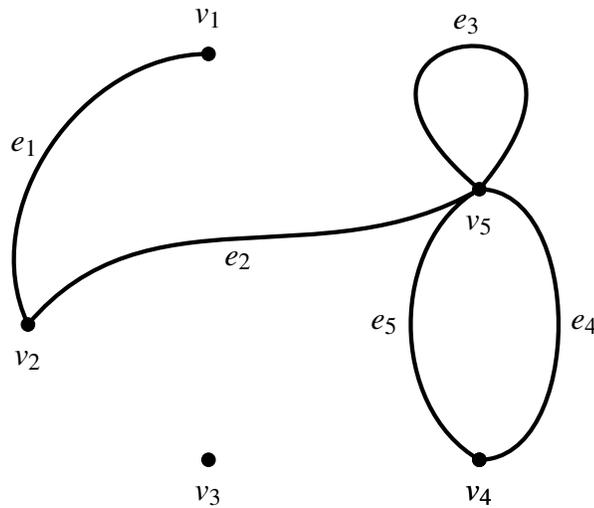


Figure 2.1: Graph

which has no vertices (i.e. if both V and E are empty) is called a null graph. A graph is called trivial if it has only one vertex.

Two edges in a graph are said to be parallel if they have the same vertices. An edge of the form (v,v) is called a loop. A graph with no parallel edges or loops is called a simple graph. Two edges are adjacent if they are of the form (u,v) and (v,w) for some vertices u , v and w .

The degree of a vertex v , written as $d(v)$, is the number of edges with v as an end vertex. A vertex with degree 1 is called a pendant vertex. An edge that has a pendant vertex as an end vertex is a pendant edge. A vertex with degree 0 is called an isolated vertex, for instance v_3 in Figure 2.1. The maximum degree of the vertices in the graph G is denoted by $\Delta(G)$. The minimum degree of the vertices in the graph G is denoted by $\delta(G)$. The graph G shown in Figure 2.1 has maximum degree 4 and minimum degree 0.

Definition 2.3. For a graph $G = (V,E)$, a walk is defined as a sequence of alternating vertices and edges of the form $v_0e_1,v_1e_2,v_2e_3,,e_kv_k$ for some $k \geq 0$, where each edge e_i is defined as $e_i = (v_{i-1},v_i)$. The length of this walk is k .

A closed walk is a walk whose first vertex is the same as the last. An open walk is a walk whose first vertex and last vertex are distinct; that is, it is a walk which ends on a different vertex from the one where it starts.

Definition 2.4. A trail is a walk where there are no repeated edges. A trail between two vertices u and v is called a $u - v$ trail.

Definition 2.5. A path is a trail in which all vertices (except perhaps the first and last ones) are distinct. A path between two vertices u and v is called a $u - v$ path. An open path is a path in which the first and last vertices are distinct. If the first and last vertices are the same, a path is called a **cycle**.

Example 2.6. The walk $v_2, e_7, v_5, e_6, v_4, e_3, v_3$ in Figure 2.1 is a path and the walk $v_2, e_7, v_5, e_6, v_4, e_3, v_e, e_2, v_2$ is a cycle.

A walk starting at u and ending at v is called a $u - v$ walk. Vertices u and v are said to be connected if there is a $u - v$ walk in the graph. If u and v are connected and v and w are connected, then u and w are connected; if there is a $u - v$ walk and there is a $v - w$ walk, then there is also a $u - w$ walk. A graph is connected if all the vertices are connected to each other. A trivial graph is assumed to be connected.

The graph in Figure 2.1 is not connected.

Definition 2.7. A tree is an undirected graph with no cycles in it. In other words, in a tree any two vertices are connected by exactly one path. A tree is called a rooted tree if one vertex has been designated the root, in which case the edges have a natural orientation, towards or away from the root.

Definition 2.8. A directed graph or digraph is a graph $G = (V, E)$ in which each edge has a direction.

As shown in Figure 2.2, edges (v, u) in a digraph are indicated by an arrow showing direction, with an arrowhead at the end vertex v . The direction of an edge (u, v) is

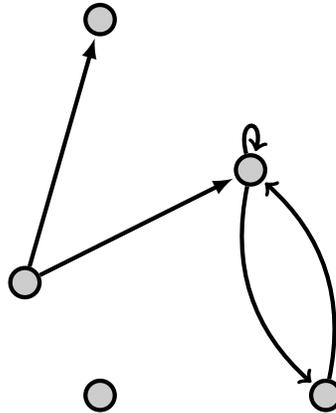


Figure 2.2: Directed Graph

opposite to the direction of the edge (v, u) . In the edge (u, v) , vertex u is called the initial vertex and v the terminal vertex of the edge. We also say that the edge (u, v) is incident out of u and incident into v .

For any two vertices u and v , we say that v is an ancestor of u and u is a descendant of v , if there exists a list of vertices $v_1 = v, v_2, \dots, v_{h-1}, v_h = u$ such that each (v_i, v_{i+1}) is a directed edge in $E(G)$ for $i = 1, \dots, h-1$. If (v, u) is a directed edge in $E(G)$, then v is a parent of u and u is a child of v .

Definition 2.9. A directed graph with a weight attached to each edge is called a weighted directed graph.

We will use weighted directed graphs to model transport paths, with weights on edges corresponding to the mass to be moved along that edge from one place to another.

2.2 Kirchoff's law

In 1845 German physicist Gustav Kirchoff described laws dealing with the conservation of current and energy within electrical circuits. In this section we describe two of Kirchoff's laws, the Current Law and the Voltage Law, and show how they may be used to describe the balancing of weights in a weighted directed graph, to use in

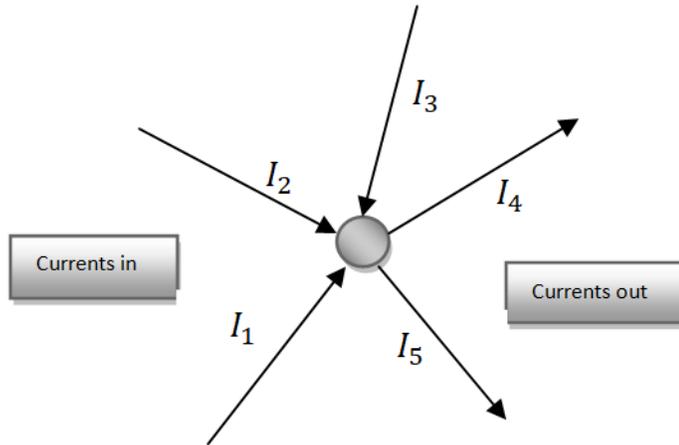


Figure 2.3: Current Flow example ([8])

our transport path models. Material in this section follows the presentation of circuit analysis in Paul's book [8].

Kirchoff's Current Law (KCL)

Kirchoff's Current Law, KCL for short, as described in [8] states that “the total current or charge entering a junction or node is exactly equal to the charge leaving the node as it has no other place to go except to leave, as no charge is lost within the node”. In other words the algebraic sum of all the currents entering and leaving a node must be equal to zero, where incoming currents are seen as positive and outgoing currents as negative. In our context, the currents coming in and out will represent the masses moving, and we can think of KCL as a statement of conservation of mass.

In Figure 2.3, which is described in [8], the three currents I_1, I_2 and I_3 entering the node are balanced by the two currents I_4 and I_5 leaving the node. Then we describe this by the equation $I_1 + I_2 + I_3 - I_4 - I_5 = 0$.

Kirchoff's Voltage Law (KVL)

Kirchoff's Voltage Law, KVL for short, as described in [8] states that “in any closed loop network, the total voltage around the loop is equal to the sum of all the voltage

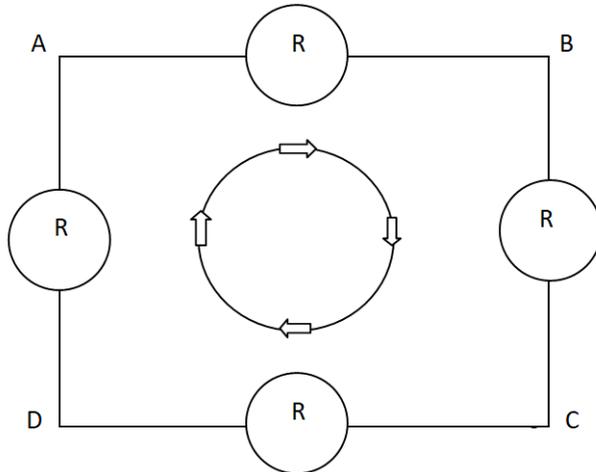


Figure 2.4: Voltage flow example ([8])

drops within the same loop” which is also equal to zero. In other words the algebraic sum of all voltages within the loop must be equal to zero. Then we describe this by the equation $V_{AB} + V_{BC} + V_{CD} + V_{DA} = 0$. This idea by Kirchoff also express a kind of law of conservation of mass.

This law means that if we start at any point in a loop, and continue in the same direction along the loop until we return back to the starting point, and record all the voltage drops, we should get a sum of zero. It is important here to maintain the same direction either clockwise or counter-clockwise. We will use Kirchoff’s Laws to model preservation of mass in our transport paths.

2.3 Measure Theory

In this section we define measure functions, and give some examples, as well as a distance function between measures. Results in this section are taken from [2] and [12]. We begin by introducing some notation. We denote the extended set of real numbers by $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$.

Definition 2.10. ([4]) Let X be a set. Then a σ -algebra Σ is a nonempty collection of subsets of X which satisfies the following properties:

1. X is in Σ .
2. If Y is in Σ , then complement of Y is also in Σ .
3. If $(Y_n)_{n \geq 1}$ is a sequence of elements of Σ , then the union $\bigcup Y_n$ of the Y_n is in Σ .

Definition 2.11. Let X be a set and Σ a σ -algebra over X . A function μ from Σ to the extended real number line is called a measure if it satisfies the following properties:

1. Non-negativity: $\mu(E) \geq 0$, for all E in Σ .
2. Null empty set: $\mu(\emptyset) = 0$.
3. Countable additivity (or σ -additivity): For all countable collections $\{E_i\}_{i \in \mathbb{N}}$ of pairwise disjoint sets in Σ ,

$$\mu\left(\bigcup_{i \in \mathbb{N}} E_i\right) = \sum_{i \in \mathbb{N}} \mu(E_i).$$

If only the second and third conditions of the definition of measure above are met, and μ takes on at most one of the values $\pm \infty$, then μ is called a signed measure. The pair (X, Σ) is called a measurable space, and the members of Σ are called measurable sets.

If (X, Σ_X) and (Y, Σ_Y) are two measurable spaces, then a function $f : X \rightarrow Y$ is called measurable if for every Y -measurable set $B \in \Sigma_Y$, the inverse image is X -measurable i.e. $f^{(-1)}(B) \in \Sigma_X$ for every $B \in \Sigma_Y$.

Definition 2.12. The family of Borel sets on a topological space is the smallest family that contains all the open sets and is closed under the operations of countable union, countable intersection, and relative complement.

Next we define a probability space (Ω, \mathcal{F}, P) consisting of three parts.

Definition 2.13. ([10]) In probability theory, the set of all possible outcomes or results of an experiment or random trial is called a sample space. A sample space is usually denoted by S , Ω , or U .

Definition 2.14. A subset of the sample space of a probability space is called an event. In other words, a set of outcomes of an experiment of a probability space is called an event.

Definition 2.15. A function P is said to be a probability measure on a probability space or the collection S of events if

1. $P(A) \geq 0$, for every event A .
2. $P(S) = 1$.
3. If $\{A_i : i \in I\}$ is a countable pairwise disjoint collection of events then $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$.

Definition 2.16. A probability space consists of three parts:

1. A sample space, Ω , which is the set of all possible outcomes.
2. A set of events \mathcal{F} , where each event is a set containing zero or more outcomes.
3. The assignment of probabilities to the events; that is, a function P from events to probabilities.

Definition 2.17. Let (X, Σ) be a measurable space. Let $x \in X$ and $A \subset X$ be any (measurable) set. Then the Dirac Measure δ_x is the measure function $\delta_x : \Sigma \rightarrow \overline{\mathbb{R}}$ defined by

$$\delta_x(A) = \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{if } x \in A. \end{cases}$$

Note that the Dirac measure is a probability measure. In Chapter 3, we shall use the Dirac measures as building blocks for other measures, and produce transport paths on Dirac measures which we can then extend to other kinds of measures.

Definition 2.18. Given a metric space (X, d) , a metric outer measure is an outer measure μ defined on the subsets of X such that

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

for every pair of positively separated subsets A and B of X .

Definition 2.19. Let (X, d) be a metric space. The diameter of a set $A \subseteq X$, denoted by $\text{diam } A$, is defined by

$$\text{diam } A := \sup\{d(x, y) \mid x, y \in A\}, \quad \text{with } \text{diam } \emptyset := 0.$$

Let U be any subset of X , and $\delta > 0$ a real number. Define

$$H_\delta^n(U) = \inf \left\{ \sum_{i=1}^{\infty} (\text{diam } U_i)^n : \bigcup_{i=1}^{\infty} U_i \supseteq U, \text{diam } U_i < \delta \right\},$$

where the infimum is over all countable covers of U by sets $U_i \subset X$ satisfying $\text{diam } U_i < \delta$.

As δ gets larger, more collection of sets can be accumulated in $H_\delta^n(U)$, which makes the infimum smaller. The values $H_\delta^n(U)$ are monotone decreasing in δ , so the limit $\lim_{\delta \rightarrow 0} H_\delta^n(U)$ can be infinite. Let

$$H^n(U) := \sup_{\delta > 0} H_\delta^n(U) = \lim_{\delta \rightarrow 0} H_\delta^n(U).$$

Then $H^n(U)$ is a metric outer measure, called the n -dimensional Hausdorff measure of U .

Definition 2.20. Let μ be a measure on the σ -algebra of Borel sets of a Hausdorff topological space X . The measure μ is called inner regular or tight if for any Borel set B , $\mu(B)$ is the supremum of $\mu(K)$ over all compact subsets K of B . i.e.

$$\mu(B) = \sup\{\mu(K) : K \subset B, K \text{ compact}\}.$$

The measure μ is called locally finite if every point of X has a neighborhood U for which $\mu(U)$ is finite. (If μ is locally finite, then it follows that μ is finite on compact sets.) The measure μ is called a Radon measure if it is inner regular and locally finite.

Definition 2.21. A Radon measure a on X is atomic if a is a finite sum of Dirac measures with positive multiplicities. That is

$$a = \sum_{i=1}^k a_i \delta_{x_i},$$

for some integer $k \geq 1$ and some points $x_i \in X$, $a_i > 0$ for each $i = 1, \dots, k$. For such a Radon measure a , we denote by $\Lambda = \sum_{i=1}^k a_i$ the total mass of a .

For any $\Lambda > 0$, let

$$A_\Lambda(X) \subset M_\Lambda(X)$$

be the space of all atomic measures on X of equal total mass Λ .

Here we recall the definition of weak convergence for Radon measures.

Definition 2.22. Let μ and $\mu_1, \mu_2, \dots, \mu_k$ be measures on \mathbb{R}^n . We say that the sequence μ_k converges weakly to μ , written as $\mu_k \rightarrow \mu$, if for each compactly supported continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f(x) d\mu_k = \int_{\mathbb{R}^n} f(x) d\mu.$$

Definition 2.23. A Polish space is a separable completely metrizable topological space.

Definition 2.24. ([10]) Consider a real-valued function $f: X \rightarrow \mathbf{R}$ where an arbitrary set X is the domain of f . Then the support of f , denoted by $\text{spt}(f)$, is defined by

$$\text{spt}(f) = \{x \in X : f(x) \neq 0\}.$$

Definition 2.25. ([11]) Consider X be a measurable space. The set of all \mathbf{R}^+ -valued measures on X is a cone $\text{Meas}(X)$, algebraic operations being defined in the usual “pointwise” way (e.g. $(\boldsymbol{\mu} + \boldsymbol{\nu})(U) = \boldsymbol{\mu}(U) + \boldsymbol{\nu}(U)$) and norm given by $\|\boldsymbol{\mu}\|_{\text{Meas}(X)} = \boldsymbol{\mu}(X)$, where \mathbf{R}^+ is the set of non-negative real numbers and $\text{Meas}(X)$ is the set of all bounded measures over the measurable space X .

At the end of Chapter 3, we will discuss metric distances defined between probability measures, and introduce both the standard Wasserstein distance and Xia’s new distance function.

Divergence Theorem

In vector calculus, the Divergence Theorem, also known as Gauss’s Theorem, can be stated as follows. Let R be a simple solid region and S be the boundary surface of R with positive orientation. Let \vec{F} be a vector field whose components have continuous first order partial derivatives. Then the Divergence Theorem states the equality

$$\iint_S \vec{F} \cdot d\vec{S} = \iiint_R \text{div } \vec{F} \, dR.$$

This equation relates the flow or flux of a vector field through a surface to the behavior of the vector field inside the surface.

In his theory of clearings and fillings of soil, Monge ([6]) assumed that the total volume of the soil used to fill the hole must equal the total volume of soil taken out of the source locations. The Divergence Theorem captures this assumption in a more technical setting. We can consider the soil taken out from the source(s) as outward

flux and the soil moved into the hole or destination location as inward flux. The Divergence Theorem then says that the outward flux of a vector field through a closed surface is equal to the volume integral of the divergence over the region inside the surface; that is the sum of all sources minus the sum of all sinks gives the net flow out of a region.

2.4 Notation

Throughout the remainder of this thesis, we will use the following notation, taken from Xia in ([13]).

- X : a compact convex subset of a Euclidean space \mathbb{R}^m .
- μ^+ : a probability Radon measure on X as the initial measure.
- μ^- : a probability Radon measure on X as the target measure.
- $\text{Path}(\mu^+, \mu^-)$: the space of all transport paths from μ^+ to μ^- .
- $\mathcal{M}_\Lambda(X)$: the space of Radon vector measure m -tuples $\mu = (\mu_1, \dots, \mu_m)$ on X .
- $\mathcal{M}^m(X)$: the space of Radon vector measures $\mu = \mu_1, \mu_2, \dots, \mu_m$ on X .
- $\|\mu\|$: the total variational measure of any vector measure $\mu \in \mathcal{M}^m(X)$.
- W : the Wasserstein 1-distance on $\mathcal{M}_\Lambda(X)$.
- $\mathcal{H}^m(X)$: m -dimensional Hausdorff measure on X .

Chapter 3

Optimal paths related to transport problems

In this chapter we will give a detailed exposition of the work of Q. Xia in [13] on optimal mass transport problems. In this paper, Xia has defined a particular cost function for which we can prove that there is an optimal-cost solution for certain kinds of transport paths.

3.1 Introduction

The transport problem described in earlier chapters deals with moving weighted objects from one location to another. We have generalized this to use measure functions on a set X as indicators of the mass to be moved, so that we now think of the transport problem as the problem of how to optimally move or transform one measure into another. Thus in [13] Xia formally stated the transport problem as follows.

Problem 3.1. ([13]) Let X be a compact subset of \mathbb{R}^m for some natural number m . Given two arbitrary probability measure μ^+ and μ^- on X , we want to find an optimal path to transport μ^+ to μ^- . This path will be a weighted directed graph.

Xia identifies the solution to this problem as consisting of two steps, as follows:

1. Identifying a class of transport paths which we can ensure contains some optimal paths.
2. Identifying a reasonable cost function on such paths that allows some paths

to overlap in a cost-efficient way. This overlap efficiency is a way to model the difference between a V-shaped and a Y-shaped path, in the setting of two source points and one sink point.

Now we introduce some notation and terminology to be used in this chapter.

- $\{x_1, x_2, \dots, x_k\}$ is the set of initial points (the sources) in \mathbb{R}^m for the transport.
- $\{y_1, y_2, \dots, y_l\}$ is the set of final points (the sinks or destinations) for the transport.
- a is the atomic measure on X determined by the sources, as defined below.
- b is the atomic measure on X determined by the sinks, as defined below.
- For an edge e in a transport path T , e^+ denotes the end vertex of e .
- For an edge e in a transport path T , e^- denotes the start vertex of e .

3.2 Transport Paths Between Measures

Definition 3.2. ([13]) Suppose we are given the following information for a transport problem: A set $\{x_1, x_2, \dots, x_k\}$ of source points and a set $\{y_1, y_2, \dots, y_l\}$ of sink points in \mathbb{R}^m , along with for each x_i a weight a_i to be moved from it, for $1 \leq i \leq k$, and for each y_j for $1 \leq j \leq l$ a weight b_j to be moved into it. Given the constraints of the transport problem, we will assume that the sums of the weights a_i are equal to the sums of the weights b_j , and call this equal sum Λ .

We can use this given information to construct two atomic measures a and b on X , as follows. Recall that an atomic measure is a weighted sum of Radon measures δ_x for points x . Then we set

$$a = \sum_{i=1}^k a_i \delta_{x_i} \quad \text{and} \quad b = \sum_{j=1}^l b_j \delta_{y_j}.$$

These measures are atomic measures, and can be represented by the vectors $\bar{a} = (a_1, \dots, a_k)$ and $\bar{b} = (b_1, \dots, b_l)$ respectively.

Conversely, suppose we are given two atomic measures $a = \sum_{i=1}^k a_i \delta_{x_i}$ and $b = \sum_{j=1}^l b_j \delta_{y_j}$, with $a_1 + \dots + a_k = b_1 + \dots + b_l = \Lambda$. Then we can use the x_i for $1 \leq i \leq k$ as source points, with each x_i having a total weight a_i coming out of it; and similarly the sink points y_j for $1 \leq j \leq l$ each have total weight b_j coming in. Thus, each given source-sink-weights scenario induces a pair of atomic measures, and vice versa.

Given such information, we now show how to construct a family of transport paths T from the source to the sink points. The vertex set of T will consist of the union of the sink and the source sets, along with one or more new points to be added as internal vertices. How many vertices we add, and where exactly they are added, will determine a unique transport path. For example, Figure 3.1 shows a graph for the

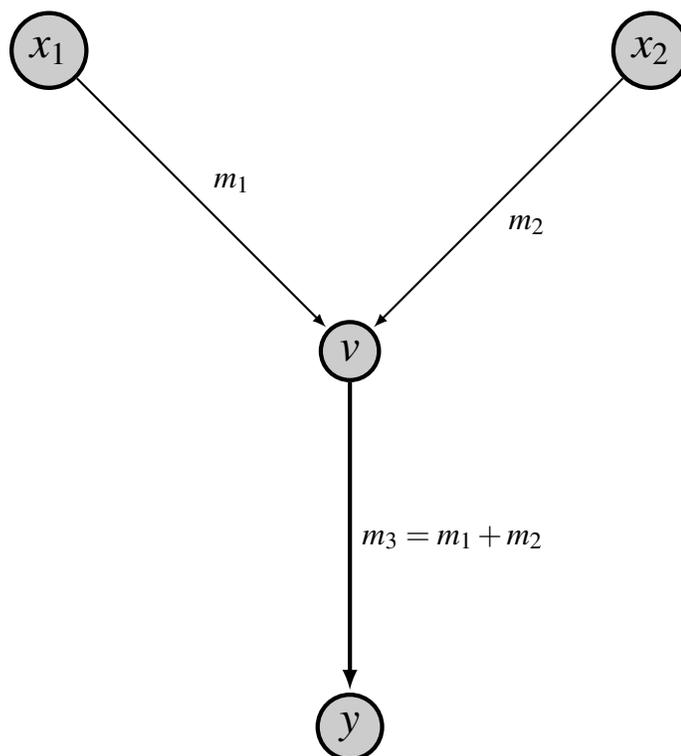


Figure 3.1: Transportation from two initial positions to one final destination

scenario with two source points and one sink point in \mathbb{R}^2 , and one new internal vertex

v , which results in a Y-shaped path; the lengths of the two edges leading into v , and the angle between these edges, will depend on the location of v , and hence so will the cost.

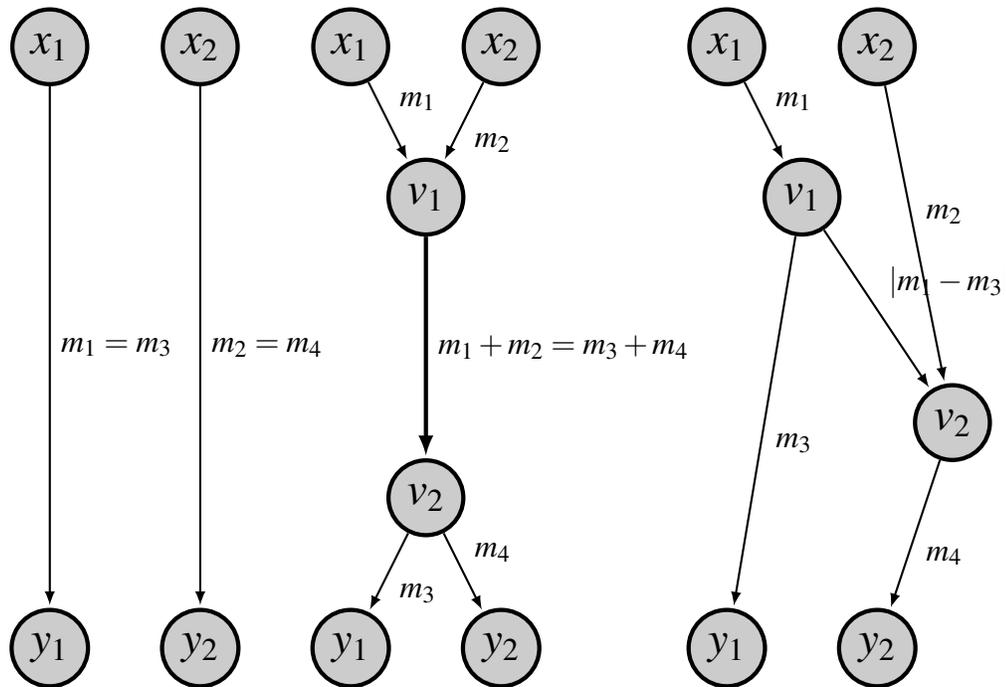


Figure 3.2: Possible paths from two sources to two sinks

Figure 3.2 shows some possible choices of new vertices for the case with two sources and two sinks. In general, the goal is to add some internal vertices so as to minimize the overall cost of the transport. Figure 3.1 also shows weights on the edges, which are determined by the values a_i and b_j as above. In Figure 3.1, each of the source points has exactly one edge coming out of it, and so the entire weight a_i out of vertex x_i is used on the edge from x_i to v ; and similarly for the sink point. In general, we define a weight function $w : E(T) \rightarrow (0, \infty)$, which assigns a weight to each directed edge in the graph. To maintain the weight balance, we must ensure that as we add new internal vertices v , we always have

$$\sum_{\substack{e \in E(T) \\ e^- = v}} w(e) = \sum_{\substack{e \in E(T) \\ e^+ = v}} w(e). \quad (3.1)$$

The space of all possible transport paths between two given measures a and b will be denoted by $\text{Path}(a, b)$. For a given total weight \wedge , the set of all the transport paths between two atomic measures on X of equal total mass \wedge will be denoted by

$$\mathcal{G}_\wedge(X) = \bigcup_{(a,b) \in \mathcal{A}_\wedge(X) \times \mathcal{A}_\wedge(X)} \text{Path}(a, b).$$

Each transport path $T \in \text{Path}(a, b)$, regarded as a weighted directed graph, also determines a vector measure on X , by

$$T = \sum_{e \in E(T)} w(e) [[e]], \quad (3.2)$$

where $[[e]] = \text{length}(e) \bar{e}$, and \bar{e} is the unit directional vector of the edge e considered as a vector in \mathbf{R}^m . Equation 3.2 means that we can think of the path T as both a graph and a collection of edges which are vectors in \mathbf{R}^m , each carrying a weight as well as a length.

Once again, the Divergence Theorem lets us simplify the balance of mass condition on a path T by

$$\text{div}(T) = a - b, \quad (3.3)$$

in the sense of distribution. Therefore

$$\text{Path}(a, b) = \left\{ T = \sum_{e \in E(T)} w(e) [[e]] \in \mathcal{M}^m(X) : \text{div}(T) = a - b \right\}.$$

Next we consider the cost of the transport. Xia's method allows for the overlapping of some parts of the transport path, for instance a Y -shaped path instead of a V -

shape from two sources to one sink, so that total cost can be minimized. He uses a parameter α to represent the magnitude of any such savings in the overall cost. We assume $0 \leq \alpha \leq 1$. The smaller the value of α , the more efficient the transport is. Our cost function will depend on the value of α , the length of the vector edges in the graph T and the weights assigned to the edges, as shown in the next Definition.

Definition 3.3. ([13]) The H^α Cost function on $\mathcal{G}_\Lambda(X)$ is defined by

$$H^\alpha(T) = \sum_{e \in E(T)} w(e)^\alpha \text{length}(e),$$

for any transport path $T = \sum_{e \in E(T)} w(e)[[e]] \in \mathcal{G}_\Lambda(X)$.

An optimal transport path T with respect to this cost function will be one that minimizes $H^\alpha(T)$, over all paths in $\text{Path}(a, b)$. Such a minimizer is called an α -optimal or simply an optimal transport path from a to b .

Note that when $\alpha = 0$, the new cost function $H^\alpha(T)$ simply gives the sum of the length of all the vectors in the graph T . When $\alpha = 1$, we have $H^\alpha(T) = \|T\|(X)$, where $\|T\|$ denotes the total variational measure of T . Moreover, Xia notes that in the case $\alpha = 1$, the edges in the graph T are straight lines.

We shall show later (see Proposition 3.7) that in the case of two source points and one sink points, Xia's method allows the addition of at most one internal vertex. Thus our transport graph will have either a V -shape (no internal vertices) or a Y -shape, with one new internal vertex. In the next example, expanded from Xia, we consider this particular setting with one vertex added.

Example 3.2.1. The following example is based on Xia ([13]). In this example, we consider a situation where we have two source points x_1 and x_2 , and one sink point y , and we want to move items of weights m_1 and m_2 from the sources to the sink. Xia's method appears to be based on the concept of the Fermat point of a triangle, a point inside the triangle which minimizes the sum of the distances from that point

each of the three vertices. In our situation, we want to make a Y -shaped path as shown in Figure 3.1 by adding one new interior vertex v to the graph. Our graph will thus have vertex set $V(T) = \{x_1, x_2, y_1, v\}$, and edges as shown in Figure 3.1. We also assign the weight m_i to the edge from vertex x_i to v , for $i = 1, 2$, and the balance equation for internal vertices then requires that the weight m_3 from v to y must equal $m_1 + m_2$.

The problem then is to find optimal positions for the new interior vertices, specifically to locate the vertex v in our Y -shaped example. An optimal positioning for v depends on the weights m_i and their relative proportions, the length of the straight-line edges and the cost savings factor α . We shall describe the process of positioning v in terms of the angles between the three edges labelled by m_1 , m_2 and m_3 in Figure 3.1. Thinking of these edges as vectors in \mathbb{R}^m , we will use the unit vectors n_1 , n_2 and n_3 respectively.

Here $n_i = \frac{v-x_i}{|v-x_i|}$ is the unit vector from x_i to v , for $i = 1, 2$, and similarly n_3 is the unit vector from v to y_1 . We will use k_1 and k_2 to represent the fractions of total weight from each vertex, so $k_i = \frac{m_i}{m_1+m_2}$ for $i = 1, 2$ and of course $k_1 + k_2 = 1$. We will let θ_i be the angle between n_i and $-n_3$, for $i = 1, 2$ which makes the angle between n_1 and n_2 equal to $\theta_1 + \theta_2$. Then we have the following balance formula (see [13]):

$$m_1^\alpha n_1 + m_2^\alpha n_2 = m_3^\alpha n_3 \tag{3.4}$$

Taking dot products between the equation (3.4) and n_1 , n_2 and n_3 respectively, we have the following equations:

$$m_1^\alpha + m_2^\alpha \cos(\theta_1 + \theta_2) = m_3^\alpha \cos \theta_1, \tag{3.5}$$

$$m_1^\alpha \cos(\theta_1 + \theta_2) + m_2^\alpha = m_3^\alpha \cos \theta_2, \tag{3.6}$$

$$m_1^\alpha \cos(\theta_1) + m_2^\alpha \cos(\theta_2) = m_3^\alpha. \tag{3.7}$$

Multiplying equation (3.5) by m_1^α and equation (3.6) by m_2^α , we get respectively

$$m_1^{2\alpha} + m_1^\alpha m_2^\alpha \cos(\theta_1 + \theta_2) = m_1^\alpha m_3^\alpha \cos \theta_1, \quad (3.8)$$

$$m_1^\alpha m_2^\alpha \cos(\theta_1 + \theta_2) + m_2^{2\alpha} = m_2^\alpha m_3^\alpha \cos \theta_2. \quad (3.9)$$

By adding equation (3.8) and equation (3.9) we get

$$\begin{aligned} m_1^{2\alpha} + m_2^{2\alpha} + 2m_1^\alpha m_2^\alpha \cos(\theta_1 + \theta_2) &= m_3^\alpha (m_1^\alpha \cos \theta_1 + m_2^\alpha \cos \theta_2) \\ &= m_3^\alpha m_3^\alpha \\ &= m_3^{2\alpha}. \end{aligned}$$

Then the above formula implies that the angles satisfy

$$\begin{aligned} \cos(\theta_1 + \theta_2) &= \frac{m_3^{2\alpha} - m_1^{2\alpha} - m_2^{2\alpha}}{2m_1^\alpha m_2^\alpha} \\ &= \frac{\left(\frac{m_3}{m_1+m_2}\right)^{2\alpha} - \left(\frac{m_1}{m_1+m_2}\right)^{2\alpha} - \left(\frac{m_2}{m_1+m_2}\right)^{2\alpha}}{2\left(\frac{m_1}{m_1+m_2}\right)^\alpha \left(\frac{m_2}{m_1+m_2}\right)^\alpha} \\ &= \frac{1 - k_1^{2\alpha} - k_2^{2\alpha}}{2k_1^\alpha k_2^\alpha}. \end{aligned}$$

Substituting the value of $\cos(\theta_1 + \theta_2)$ in equation (3.5) we get

$$\cos \theta_1 = \frac{k_1^{2\alpha} + 1 - k_2^{2\alpha}}{2k_1^\alpha}.$$

Similarly from equation (3.6) we get

$$\cos \theta_2 = \frac{k_2^{2\alpha} + 1 - k_1^{2\alpha}}{2k_2^\alpha}.$$

Now we can use these various equations for the angles in Figure 3.1 to consider some

specific cases for the weights and their proportions. First, the equation above for $\cos(\theta_1 + \theta_2)$ shows exactly how the angles being determined by the position of v depend on the values of α and the proportions k_i . For example, in the case $\alpha = \frac{1}{2}$, we have

$$\begin{aligned} \cos(\theta_1 + \theta_2) &= \frac{1 - k_1^{2\alpha} - k_2^{2\alpha}}{2k_1^\alpha k_2^\alpha} \\ &= \frac{1 - k_1^{2(\frac{1}{2})} - k_2^{2(\frac{1}{2})}}{2k_1^{\frac{1}{2}} k_2^{\frac{1}{2}}} \\ &= \frac{1 - k_1 - k_2}{2k_1^{(\frac{1}{2})} k_2^{(\frac{1}{2})}} \\ &= \frac{1 - k_1 - (1 - k_1)}{2k_1^{\frac{1}{2}} k_2^{\frac{1}{2}}} = 0, \end{aligned}$$

so that $\theta_1 + \theta_2 = 90^\circ$. Similar calculations show the value of θ for other values of α , as shown in Table 3.1.

Table 3.1: α and corresponding $\theta_1 + \theta_2$

α	$\theta_1 + \theta_2$
0	120°
$\frac{1}{2}$	90°
1	0°

We see here that the larger α is, the smaller the centre angle at v is, and the steeper the vectors n_1 and n_2 .

Another special case occurs if the weights m_1 and m_2 to be moved from each of the two vertices are equal. In this situation, clearly $k_1 = k_2$, and we get $\cos(\theta_1 + \theta_2) = 2^{2\alpha-1} - 1$. From this we have that $\theta_1 + \theta_2 = \arccos(2^{2\alpha-1} - 1)$.

Example 3.2.2. ([13]) Figure 3.2 illustrates several graphs possible in the scenario

where we have two source points x_1 and x_2 , and two sink points y_1 and y_2 , with weights satisfying $m_1 + m_2 = m_3 + m_4$ as shown in the Figure. Three optimal transport paths are shown, one with no internal vertices added and two with two internal vertices added. These paths depend on the positions of the source and sink points, the distances involved, and the ratio of the various weights.

We have now shown Xia's method to create a transport path T , a weighted directed graph, between atomic measures a and b . However, the transport path graph created here is not always a directed tree, because there could be some loops or cycles in it. The next Proposition shows that we can however always modify the graph T into a directed tree \tilde{T} from a to b , with cost no greater than H^α .

Proposition 3.4. (*[13]*) *For any $T \in \text{Path}(a, b)$, there exists a $\tilde{T} \in \text{Path}(a, b)$ which contains no cycles and for which $H^\alpha(\tilde{T}) \leq H^\alpha(T)$.*

Proof. Suppose that O is some cycle, that is a list of edges, in the graph T . For each edge e in O , there is a weight on the edge in the graph T , to be denoted by $w(e)$. We will define a new mass on edges by

$$m(e) = \frac{\alpha \text{length}(e)}{w(e)^{1-\alpha}}$$

Without loss of generality, we pick an orientation for the cycle O , and use this orientation to form two sets corresponding to a partition of the set of all edges of O into the set of edges with the same orientation as O and the set of those with opposite orientation. Recall that $[[e]] = \text{length}(e) \bar{e}$ is the vector in m -space determined by the edge e in the graph. We let

$$O_1 = \sum \{ [[e]] : \text{edge } e \text{ of } O \text{ has the same orientation as } O \} \text{ and}$$

$$\mathcal{O}_2 = \sum \{[e] : \text{edge } e \text{ of } \mathcal{O} \text{ has the opposite direction to } \mathcal{O}\}.$$

Note that one of \mathcal{O}_1 or \mathcal{O}_2 could be empty, and as they both each have the same start and end vertices but with opposite orientation, equation (3.3) guarantees that the weights balance and so $\text{div}(\mathcal{O}_1) = \text{div}(\mathcal{O}_2)$. By reversing the orientation on \mathcal{O} if necessary, we may assume that

$$\sum_{e \in \mathcal{O}_1} m(e) \leq \sum_{e \in \mathcal{O}_2} m(e).$$

We use these two sets to construct a new graph T' from T , as follows. Assuming as above that the vector set \mathcal{O}_2 has the larger sum of $m(e)$ values, we take $w_0 = \min\{w(e) : e \in \mathcal{O}_2\}$. Now we go through the edges in the loop, and add w_0 units of weight to each of the edges in \mathcal{O}_1 , and remove w_0 units from each edge in \mathcal{O}_2 . Note that we have maintained the weight balance at each of the internal vertices. After this addition and subtraction of weights, at least one edge in \mathcal{O}_2 will now have weight of zero. We then remove any such zero-weight edges from the graph, resulting in the new graph T' . Xia has denoted this process by $T' = T + w_0(\mathcal{O}_1 - \mathcal{O}_2)$. Note that if $\alpha = 0$, then all the edges $m(e)$ are zero, so either partition set can be used, and we adjust edge weights by the minimum edge weight.

This process will be illustrated below in Examples 3.2.3 and 3.2.4. Note that T' now has one loop removed, but may still have more loops; if so, we repeat this process as necessary. We finish the proof of the Proposition by showing that at any stage in this process, $H^\alpha(T') \leq H^\alpha(T)$.

To see this, we consider the function on $[0, w]$ defined by

$$\begin{aligned} f(\lambda) &:= H^\alpha(T + \lambda(\mathcal{O}_1 - \mathcal{O}_2)) - H^\alpha(T) \\ &= \sum_{e \in \mathcal{O}_1} \text{length}(e)[(w(e) + \lambda)^\alpha - w(e)^\alpha] + \sum_{e \in \mathcal{O}_2} \text{length}(e)[(w(e) - \lambda)^\alpha - w(e)^\alpha]. \end{aligned}$$

Here,

$$\begin{aligned}
 f(0) &= \sum_{e \in O_1} \text{length}(e)[(w(e) + \lambda)^\alpha - w(e)^\alpha] + \sum_{e \in O_2} \text{length}(e)[(w(e) - \lambda)^\alpha - w(e)^\alpha] \\
 &= \sum_{e \in O_1} \text{length}(e)[(w(e) + 0)^\alpha - w(e)^\alpha] + \sum_{e \in O_2} \text{length}(e)[(w(e) - 0)^\alpha - w(e)^\alpha] \\
 &= \sum_{e \in O_1} \text{length}(e)[(w(e))^\alpha - w(e)^\alpha] + \sum_{e \in O_2} \text{length}(e)[(w(e))^\alpha - w(e)^\alpha] \\
 &= \sum_{e \in O_1} \text{length}(e)[0] + \sum_{e \in O_2} \text{length}(e)[0] \\
 &= 0.
 \end{aligned}$$

One can easily compute the derivative of f ,

$$f'(\lambda) = \sum_{e \in O_1} \text{length}(e)[\alpha(w(e) + \lambda)^{\alpha-1}] + \sum_{e \in O_2} \text{length}(e)[\alpha(w(e) - \lambda)^{\alpha-1}].$$

Therefore,

$$\begin{aligned}
 f'(0) &= \sum_{e \in O_1} \text{length}(e)[\alpha(w(e) + 0)^{\alpha-1}] - \sum_{e \in O_2} \text{length}(e)[\alpha(w(e) - 0)^{\alpha-1}] \\
 &= \sum_{e \in O_1} \text{length}(e)[\alpha(w(e))^{\alpha-1}] - \sum_{e \in O_2} \text{length}(e)[\alpha(w(e))^{\alpha-1}] \\
 &= \sum_{e \in O_1} \frac{\text{length}(e)\alpha}{(w(e))^{1-\alpha}} - \sum_{e \in O_2} \frac{\text{length}(e)\alpha}{(w(e))^{1-\alpha}} \\
 &= \sum_{e \in O_1} m(e) - \sum_{e \in O_2} m(e).
 \end{aligned}$$

Since $\sum_{e \in O_1} m(e) \leq \sum_{e \in O_2} m(e)$, we have that

$$f'(0) = \sum_{e \in O_1} m(e) - \sum_{e \in O_2} m(e) \leq 0.$$

We shall now compute the second derivative of f ,

$$\begin{aligned} f''(\lambda) &= \sum_{e \in \mathcal{O}_1} \text{length}(e) \alpha(\alpha - 1) [(w(e) + \lambda)^{\alpha-2}] \\ &\quad + \sum_{e \in \mathcal{L}_2} \text{length}(e) \alpha(\alpha - 1) [(w(e) - \lambda)^{\alpha-1}]. \end{aligned}$$

Then, since $\alpha \leq 1$, we always have

$$f''(\lambda) \leq 0, \quad f'(\lambda) \leq f'(0) \leq 0,$$

from which we obtain $f(\lambda) \leq f(0) = 0$. Thus,

$$\begin{aligned} f(w) &= H^\alpha(T + w(\mathcal{O}_1 - \mathcal{O}_2)) - H^\alpha(T) \\ &= H^\alpha(T') - H^\alpha(T). \end{aligned}$$

Therefore, $H^\alpha(T') \leq H^\alpha(T)$ as $f(w) \leq 0$. □

Example 3.2.3. Let us consider the graph G shown in the Figure 3.3, situated as a vector measure in \mathbb{R}^2 , with vertices $x_1(-4,4)$, $x_2(0,4)$, $x_3(4,4)$, $v_1(-2,2)$, $v_2(2,2)$ and $y(0,0)$. Here, we are shipping items from x_1 , x_2 and x_3 with weights 3, 7 and 4 units respectively to the sink y , and two new interior vertices v_1 and v_2 have been added. All the weights $w(e)$ on each edge are given in Figure 3.3. A cycle $\mathcal{O} = x_2 v_2 y v_1 x_2$ is present in this graph which has two partition sets \mathcal{O}_1 containing the edges $x_2 v_2$ and $v_2 y$, and \mathcal{O}_2 containing the edges $x_2 v_1$ and $v_1 y$.

Let us consider $\alpha = \frac{1}{2}$. The length of each edge in this graph is 2.8284 approximately. By Proposition 3.4 we have $m(x_2 v_1) = 0.8000$, $m(x_2 v_2) = 0.5060$, $m(v_1 y) = 0.5060$ and $m(v_2 y) = 0.3771$. Then we have

$$\sum_{e \in \mathcal{O}_1} m(e) = m(x_2 v_2) + m(v_2 y) = 0.8831$$

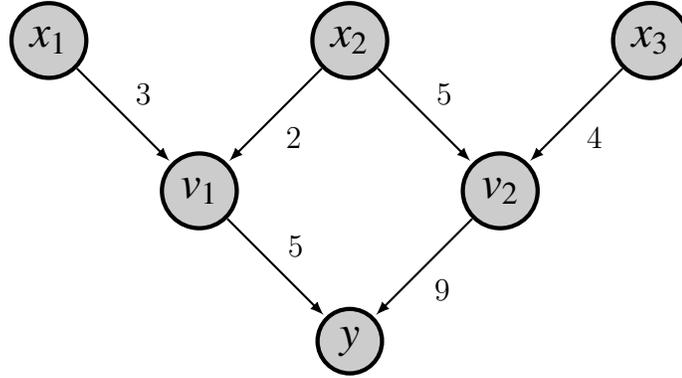


Figure 3.3: Graph G with one cycle

and

$$\sum_{e \in O_2} m(e) = m(x_2v_1) + m(v_1y) = 1.3060.$$

Clearly, $\sum_{e \in O_1} m(e) \leq \sum_{e \in O_2} m(e)$, and $w_0 = \min\{w(e) : e \in O_2\} = \min\{2, 5\} = 2$. To form the new graph G' , we therefore need to add two units of weight to each of the O_1 edges, and remove two units of weight from each of the O_2 edges. As in Figure 3.4 then, the weights in our loop become 7, 10, 0 and 3. Since the edge x_2v_1 now has no mass on it, it can be removed from the graph, resulting in the new graph G' with no cycles.

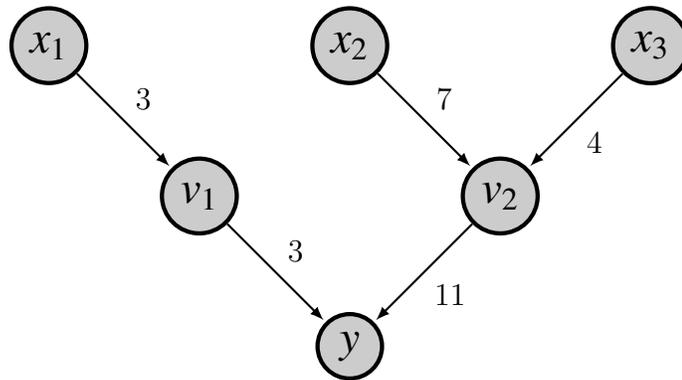


Figure 3.4: Graph G' with no cycle

Example 3.2.4. Consider the graph T shown in Figure 3.5, where we are shipping items from three sources x_1 , x_2 and x_3 to one sink point y , and two interior points v_1

and v_2 have been added, with weights as shown. This graph has three loops (ignoring direction). We look first at the loop $O = x_2x_3yv_2x_2$, and show how to remove one edge from the graph to remove this loop. We partition O into two sets, one with edges x_2x_3 and x_3y , and the other with edges x_2v_2 and v_2y . Let us consider the case $\alpha = 1$, so that each $m(e)$ is simply the length of e as a vector. The minimum weight of an edge in the partition set with larger m sum is then 6, so we increase the weights by 6 units on each of x_2v_2 and v_2y , and decrease by 6 on each of x_2x_3 and x_3y . Removing the edge that is left with weight zero then results in the graph T' shown in Figure 3.6.

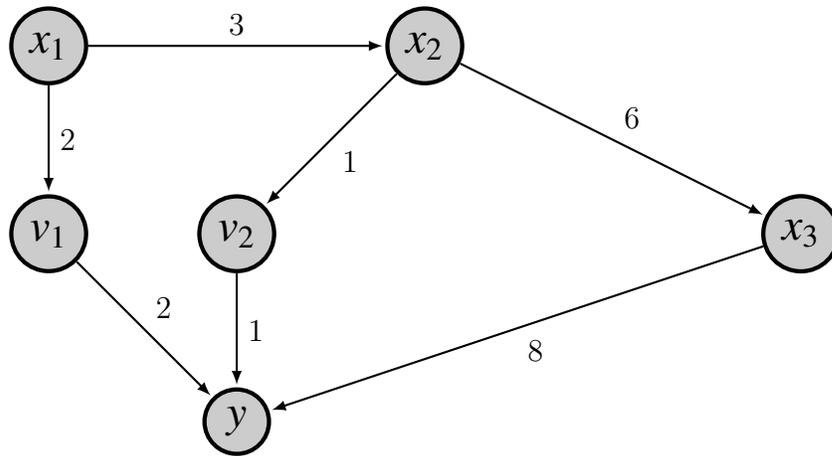


Figure 3.5: Graph T with cycles

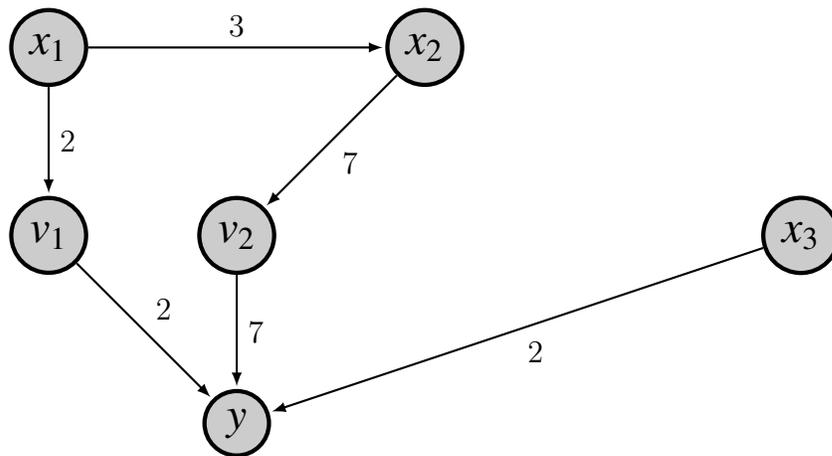


Figure 3.6: Graph from Figure 3.5 with one cycle removed

Starting with the graph in Figure 3.6, we can again consider two orientations on the remaining loop. We set L_1 to be the path $x_1x_2v_2y$, and L_2 to be the path from x_1 to v_1 to y . We get $\sum_{e \in L_1} m(e) = \sum_{e \in L_1} \text{length}(e) = 2.5 + 3.54 = 6.04$ and $\sum_{e \in L_2} m(e) = \sum_{e \in L_2} \text{length}(e) = 5 + 3.54 + 2.5 = 11.04$. Hence, $\sum_{e \in L_1} m(e) \leq \sum_{e \in L_2} m(e)$. Then again applying $w_1 = \min\{w(e) : e \in L_1\} = \{3, 7, 7\} = 3$ the formula $T'' = T' + w_1(L_1 - L_2) = T' + 3(L_1 - L_2)$ we get the new cycle-free transport path shown in Figure 3.7.

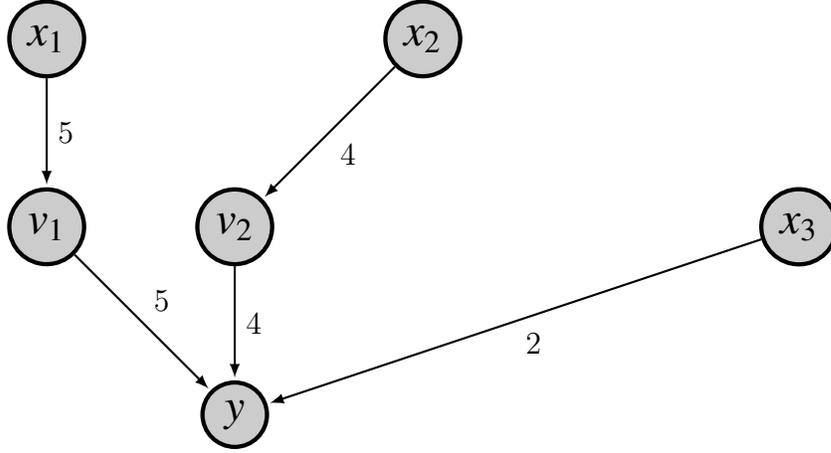


Figure 3.7: Graph from Figure 3.6 with remaining cycle removed

From the above proposition, we may restrict our transport paths to the class of directed trees. For directed trees, we have the following minor but useful lemma.

Lemma 3.5. *Suppose $0 < \Lambda < +\infty$ and $T = \sum_{e \in E(T)} w(e)[[e]] \in \text{Path}(a, b)$ is a directed tree with $a, b, \in \mathcal{A}_\Lambda$ as before. Then for any edge $e \in E(T)$, we have*

$$0 < w(e) \leq \Lambda.$$

Moreover,

$$\frac{H^\alpha(T)}{\Lambda^\alpha} \geq \frac{H^1(T)}{\Lambda}.$$

Proof. It is clear from the definition that $w(e) \leq \Lambda$, so that $\frac{w(e)}{\Lambda} \leq 1$. As $\alpha \in [0, 1]$ we

obtain

$$\left(\frac{w(e)}{\Lambda}\right)^\alpha \geq \frac{w(e)}{\Lambda}.$$

Hence,

$$\begin{aligned} \frac{H^\alpha(T)}{\Lambda^\alpha} &= \sum_{e \in E(T)} \left(\frac{w(e)}{\Lambda}\right)^\alpha \text{length}(e) \\ &\geq \sum_{e \in E(T)} \frac{w(e)}{\Lambda} \text{length}(e) \\ &= \frac{H^1(T)}{\Lambda}. \end{aligned}$$

□

As we use Xia's result for transport paths for atomic measures to build paths for more general measures in the next Section, the following Proposition will allow us to scale paths as necessary.

Proposition 3.6. *For any $T = \sum_{e \in E(T)} w(e)[[e]] \in \text{Path}(a, b)$ and any positive number $r > 0$,*

$$rT := \sum_{e \in E(T)} (rw(e))[[e]]$$

is a transport path from ra to $rb \in \mathcal{A}_{r\Lambda}(X)$, and

$$H^\alpha(rT) = r^\alpha H^\alpha(T).$$

In particular,

$$\frac{T}{\Lambda} \in \text{Path}\left(\frac{a}{\Lambda}, \frac{b}{\Lambda}\right) \quad \text{with} \quad H^\alpha(T) = \Lambda^\alpha H^\alpha\left(\frac{T}{\Lambda}\right).$$

Proof. Note that

$$\begin{aligned} rT &= r \sum_{e \in E(T)} w(e)[[e]] \\ &= \sum_{e \in E(T)} (rw(e))[[e]] \end{aligned}$$

is a transport path from ra to $rb \in \mathcal{A}_{r\Lambda}(X)$, and

$$\begin{aligned} H^\alpha(rT) &= \sum_{e \in E(T)} (rw(e))^\alpha[[e]] \\ &= r^\alpha \sum_{e \in E(T)} (w(e))^\alpha[[e]] \\ &= r^\alpha H^\alpha(T). \end{aligned}$$

Clearly, $\frac{T}{\Lambda} \in \text{Path}\left(\frac{a}{\Lambda}, \frac{b}{\Lambda}\right)$ and

$$\begin{aligned} \Lambda^\alpha H^\alpha\left(\frac{T}{\Lambda}\right) &= \Lambda^\alpha \sum_{e \in E(T)} \left(\frac{w(e)}{\Lambda}\right)^\alpha[[e]] \\ &= \sum_{e \in E(T)} (w(e))^\alpha[[e]] \\ &= H^\alpha(T). \end{aligned}$$

□

Proposition 3.7. (Xia [14], Prop. 1.2) *Let G be a transport path from a to b which contains no cycles. Then the number of vertices in G of degree three or more must be $\leq k+l-2$.*

Proof. It is well known in graph theory (see [9]) that the sum of the degrees of the vertices in a connected graph is twice the number of edges, and the Euler characteristic $\chi_G = |V(G)| - |E(G)|$.

Combining these two facts lets us write

$$\sum_{v \in G} \deg(v) = 2(\# \text{ of vertices in } G) - 2\chi_G \quad (3.10)$$

Now we denote the number of vertices by $\sum_{v \in V(G)} 1$ and split the set of vertices into those of degree 1, 2 and ≥ 3 . This gives

$$\begin{aligned} & \sum_{\substack{v \in V(G) \\ \deg(v)=1}} \deg(V) + \sum_{\substack{v \in V(G) \\ \deg(v)=2}} \deg(V) + \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} \deg(V) \\ &= 2 \sum_{\substack{v \in V(G) \\ \deg(v)=1}} 1 + 2 \sum_{\substack{v \in V(G) \\ \deg(v)=2}} 1 + 2 \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} 1 - 2\chi_G. \end{aligned}$$

$$\begin{aligned} \implies & \sum_{\substack{v \in V(G) \\ \deg(v)=2}} \deg(V) + \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} \deg(V) \\ &= \sum_{\substack{v \in V(G) \\ \deg(v)=1}} 1 + 2 \sum_{\substack{v \in V(G) \\ \deg(v)=2}} 1 + 2 \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} 1 - 2\chi_G. \end{aligned}$$

$$\begin{aligned} \implies & \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} \deg(v) = \sum_{\substack{v \in V(G) \\ \deg(v)=1}} 1 + 2 \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} 1 - 2\chi_G, \\ & \text{since } \sum_{\substack{v \in V(G) \\ \deg(v)=2}} \deg(V) = 2 \sum_{\substack{v \in V(G) \\ \deg(v)=2}} 1 \end{aligned}$$

$$\implies \sum_{\substack{v \in V(G) \\ \deg(v)=1}} 1 - 2\chi_G = \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} \deg(v) - 2 \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} 1$$

$$\implies \sum_{\substack{v \in V(G) \\ \deg(v)=1}} 1 - 2\chi_G = \sum_{\substack{v \in V(G) \\ \deg(v) \geq 3}} [\deg(v) - 2]$$

It follows from this that

$$\begin{aligned}
 (\# \text{ vertices of degree } \geq 3) &\leq \sum_{\substack{v \in V(G) \\ \text{deg}(v) \geq 3}} [\text{deg}(v) - 2] \\
 &= \sum_{\substack{v \in V(G) \\ \text{deg}(v) = 1}} 1 - 2\chi_G \\
 &\leq \sum_{\substack{v \in V(G) \\ \text{deg}(v) = 1}} 1 - 2 \quad \text{since } \chi_G \geq 1 \\
 &= (\# \text{ vertices of degree } = 1) - 2 \\
 &\leq k + l - 2,
 \end{aligned}$$

where the last step follows from the fact that any new internal vertices added in our process have degree at least 2.

□

This Proposition gives a maximum for the number of new interior vertices that can be added, and so helps determine the possible shapes of transport paths. As noted above, given two sources and one sink we can add at most one internal vertex, so we get either a V- or Y-shaped graph. For two sources and two sinks, the maximum $k + l - 2 = 2$, so at most two new vertices can be added, with the shapes shown in Figure 3.2. For three sources and one sink, we have a maximum of two new vertices as shown in Figure 3.3.

Besides giving us information like this about possible scenarios for specific cases, Proposition 3.7 also shows that for a given source-sink combination, the number of new internal vertices which can be added is finite. This means that in the case of two atomic measures, we have a finite collection of possible transport paths to consider, and can always find one with optimal cost.

3.3 Transportation of General Measures

In the previous section we presented Xia's method for constructing a transport path between two atomic measures, along with some examples. Now we show how that method can be extended to produce transport paths for other measures, such as Radon measures. We recall from Definition 2.20 that a Radon measure is a measure which is both inner regular and locally finite. The key fact we use is that any Radon measure on our space X can be approximated by atomic measures, in the sense that the Radon measure can be expressed as a limit, in the weak* topology, of a sequence of atomic measures. We thus get a sequence of transport paths, and can consider the limit of this sequence under the vector measure.

Dyadic Approximation of Radon measures

In this section we will show how any Radon measure can be approximated by a particular sequence of atomic measures. We start with an arbitrary Radon measure μ , which is defined on a set $X \subseteq \mathbb{R}^m$. We can always find a cube Q in \mathbb{R}^m such that $X \subseteq Q \subseteq \mathbb{R}^m$. Suppose that this cube Q has centre point c and edge length l . Our sequence of atomic measures will be determined by a sequence of subdivisions of Q , known as dyadic subdivisions. The notation is complex, but essentially we subdivide Q into a sequence of sub-cubes, by halving the edge length each time.

To start, we let $Q_0 = \{Q\}$. Now for Q_1 , we want to form a set of sub-cubes of Q , by dividing each edge of Q in half. In dimension d , this results in 2^d sub-cubes, and it is easiest to index these by d -tuples with integer coordinates starting at 0. For example, in two-dimensional space, we divide the original cube Q into four sub-cubes, which we can label by coordinates $(0,0), (0,1), (1,0)$ and $(1,1)$.

Next, for Q_2 , we partition each of the 2^d sub-cubes in Q_1 , again by halving each side. In two dimensions, this would result in 16 sub-cubes in Q_2 , each with edge length $l/4$; these 16 can be labelled by (i, j) where i and j are integers between 0 and 3.

With this example in mind, we can now introduce our notation. We define $\mathcal{Q}_0 = \{\mathcal{Q}\}$, and then for $i \geq 1$ we set \mathcal{Q}_i to be the set of the sub-cubes of length $l/2^i$ that make up \mathcal{Q} , indexed by m -tuples with coordinates which are integers between 0 and 2^i . Formally,

$$\mathcal{Q}_i = \{\mathcal{Q}_i^h : h \in \mathbb{Z}^m \cap [0, 2^i]^m\}. \quad (3.11)$$

Note here that each \mathcal{Q}_i consists of a set of sub-cubes of edge length $l/2^i$. For each such sub-cube \mathcal{Q}_i^h , we will denote the centre point of the cube by c_i^h . We also take the weight $m_i^h = \mu(\mathcal{Q}_i^h)$. Now we use this information to form atomic measures, based on the points c_i^h and weights m_i^h . For each $i \geq 0$ we set

$$R_i(\mu) = \sum_{h \in \mathbb{Z}^m \cap [0, 2^i]^m} m_i^h \delta_{c_i^h} \in \mathcal{A}_\Lambda(X).$$

In this way, the original Radon measure induces a sequence $\{R_i(\mu) : i \geq 0\}$ of atomic measures. The key result now is that this sequence converges weakly to μ , in the sense of measure convergence. We can thus consider the sequence $\{R_i(\mu) : i \geq 0\}$ to be an approximation of μ by atomic measures. This is called the dyadic approximation of μ .

Definition 3.8. Let $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$ be any two Radon measures on X with equal total mass Λ . We say a vector measure $T_0 \in \mathcal{M}_\Lambda(X)$ is a transport path from μ^+ to μ^- if there exist two sequences $\{a_i\}$, $\{b_i\}$ of atomic measures in $\mathcal{A}_\Lambda(X)$ with a corresponding sequence of transport Paths $T_i \in \text{Path}(a_i, b_i)$ such that

$$a_i \rightharpoonup \mu^+, b_i \rightharpoonup \mu^- \text{ and } T_i \rightharpoonup T_0,$$

where the measures converge as Radon measures and the T_i converge to T_0 as vector measures.

The sequence of triples $\{a_i, b_i, T_i\}$ is called an approximating graph sequence for T_0 . Note that for any such T_0 , we have $\text{div}(T_0) = \mu^+ - \mu^-$, in the sense of distributions. Let

$$\text{Path}(\mu^+, \mu^-) \subset \mathcal{M}_\Lambda(X)$$

be the space of all transport paths from μ^+ to μ^- .

For the cost function to be used, we set the parameter $\alpha \in [0, 1]$, and for any $T_0 \in \text{Path}(\mu^+, \mu^-)$ we define the H^α cost of T_0 to be

$$H^\alpha(T_0) := \inf \liminf_{i \rightarrow \infty} H^\alpha(T_i),$$

where the infimum is taken over the set of all possible approximating graph sequences $\{a_i, b_i, T_i\}$ of T_0 .

When μ^+ and μ^- are atomic, these new definitions for transport path and H^α cost functions reduce to the previous definitions in the last section.

Using scaling as needed, we can assume that both measures μ^+ and μ^- have total weight $\Lambda = \mu^+(X) = \mu^-(X) = 1$, making them both probability measures. For any probability measure $\mu \in \mathcal{M}_1(X)$, we will now use the dyadic approximation $A_i(\mu)$ of μ to construct a transport path of finite H^α cost from μ to the Dirac measure δ_c , where c is the center of the cube Q containing X with edge length l . We have the following result.

Proposition 3.9. (*[13]*). *Let $\alpha \in (1 - \frac{1}{m}, 1]$. For any $\mu \in \mathcal{M}_1(X)$, there exists a transport path $T \in \text{Path}(\mu, \delta_c)$ such that*

$$H^\alpha(T_0) \leq \frac{1}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{ml}}{2}.$$

Proof. Let $\{R_i(\mu)\}$ be the dyadic approximation of μ . For each $i \geq 0$ and $h \in \mathbb{Z}^m \cap [0, 2^i)^m$, each cube Q_i^h of level i corresponds to 2^m cubes $\{Q_{i+1}^{2^m h + h'} : h' = 0, 1, 2, \dots, 2^m - 1\}$

of level $i+1$ by dyadic subdivision. Then we can construct a transport path

$$T_i^h := \sum_{h'=0}^{2^m-1} m_{i+1}^{2^m h+h'} [(c_{i+1}^{2^m h+h'}, c_i^h)] \in \text{Path} \left(m_i^h \delta_{c_i^h}, \sum_{h'=0}^{2^m-1} m_{i+1}^{2^m h+h'} \delta_{c_{i+1}^{2^m h+h'}} \right),$$

which is a directed tree from the center c_i^h of Q_i^h to the centers $c_{i+1}^{2^m h+h'}$ of 2^m cubes $\{Q_{i+1}^{2^m h+h'} : h' = 0, 1, 2, \dots, 2^m - 1\}$ with suitable weights. Now, for each $n \geq 0$, set

$$T_n = \sum_{i=0}^n \sum_{h \in Z^m \cap [0, 2^i]^m} T_i^h \in \text{Path} (R_0(\mu), R_{n+1}(\mu)).$$

Then

$$\begin{aligned} M^\alpha(T_n) &= \sum_{i=0}^n \sum_{h \in Z^m \cap [0, 2^i]^m} \sum_{h'=0}^{2^m-1} (m_{i+1}^{2^m h+h'})^\alpha \text{length}(c_{i+1}^{2^m h+h'}, c_i^h) \\ &= \sum_{i=0}^n \sum_{h \in Z^m \cap [0, 2^i]^m} \sum_{h'=0}^{2^m-1} (m_{i+1}^{2^m h+h'})^\alpha \frac{\sqrt{ml}}{2^{i+2}} \\ &\leq \sum_{i=0}^n \sum_{h \in Z^m \cap [0, 2^i]^m} \sum_{h'=0}^{2^m-1} \left(\frac{1}{2^{m(i+1)}} \right)^\alpha \frac{\sqrt{ml}}{2^{i+2}} \\ &= \sum_{i=0}^n \sum_{h \in Z^m \cap [0, 2^i]^m} \sum_{h'=0}^{2^m-1} 2^{-m\alpha(i+1)} 2^{-(i+1)} \frac{\sqrt{ml}}{2} \\ &= \sum_{i=0}^n \sum_{h \in Z^m \cap [0, 2^i]^m} 2^m 2^{-m\alpha(i+1)} 2^{-(i+1)} \frac{\sqrt{ml}}{2} \\ &= \sum_{i=0}^n 2^{im} 2^m 2^{-m\alpha(i+1)} 2^{-(i+1)} \frac{\sqrt{ml}}{2} \\ &= \sum_{i=0}^n (2^{i+1})^{m(1-\alpha)-1} \frac{\sqrt{ml}}{2} \\ &\leq \frac{1}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{ml}}{2}, \quad (\text{for } \alpha > 1 - \frac{1}{m}) \end{aligned}$$

where the inequality in the third line of the above equation follows from the fact that

the function

$$f_i(x_1, x_2, \dots, x_{2^{m(i+1)}}) = \sum_{h \in \mathbb{Z}^m \cap [0, 2^i]^m} \sum_{h'=0}^{2^m-1} (x_{2^m h+h'})^\alpha \frac{\sqrt{md}}{2^{i+2}}$$

achieves its maximum at the point $(\frac{1}{2^{m(i+1)}}, \frac{1}{2^{m(i+1)}}, \dots, \frac{1}{2^{m(i+1)}})$.

Since $\|T_n\|(X) = H^1(T_n) \leq H^\alpha(T_n)$ has bounded total variation, by the compactness of vector measures, $\{T_n\}$ sub-converges weakly to a vector measure T_0 with

$$H^\alpha(T_0) = \inf \lim_{i \rightarrow \infty} \inf H^\alpha(G_i) \leq \liminf H^\alpha(G_i) \leq \frac{1}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2}.$$

Thus $T_0 \in \text{Path}(\mu, \delta_c)$ has finite H^α cost. \square

The following Existence Theorem now gives us the solution to Problem 3.1 for Radon measures.

Theorem 3.10. (*Xia's Existence theorem [13]*) *Given two Radon measures $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$ on $X \subset \mathbb{R}^m$ and $\alpha \in (1 - \frac{1}{m}, 1]$, there exists an optimal transport path T with least H^α cost among all transport paths in the family $\text{Path}(\mu^+, \mu^-)$. Moreover*

$$H^\alpha(T) \leq \frac{\Lambda^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{ml}}{2}.$$

Proof. Let $\{P_i\}$ be an H^α minimizing sequence in $\text{Path}(\mu^+, \mu^-)$. For each P_i , there exists a transport path $T_i \in \text{Path}(a_i, b_i)$ such that

$$H^\alpha(T_i) \leq H^\alpha(P_i) + \frac{1}{2^i}.$$

From Proposition 3.4, we may assume that the graph T_i have no cycles in them. Let W be the Wasserstein 1-distance on $\mathcal{M}_1(X)$, a metric which results in the weak* topology on $M_1(X)$. Then the transport path T_i from a_i to b_i also satisfies

$$W(a_i, \mu^+) + W(b_i, \mu^-) < \frac{1}{2^i}.$$

From this we get that the T_i satisfy $\|T_i\|(X) = H^1(T_i)$. Now Lemma 3.5 shows that

$$\frac{H^\alpha(T_i)}{\Lambda^\alpha} \geq \frac{H^1(T_i)}{\Lambda},$$

which means that $H^1(T_i) \leq \Lambda^{1-\alpha} H^\alpha(T_i)$. Finally, this shows that

$$\|T_i\|(X) = H^1(T_i) \leq \Lambda^{1-\alpha} H^\alpha(T_i) \leq (H^\alpha(T_i) + \frac{1}{2^i}) \Lambda^{1-\alpha}.$$

These equations show that the T_i are uniformly bounded. Compactness of vector measures then means that the T_i sequence converges to a vector measure T which gives a path from μ^+ to μ^- . This T is optimal as a transport path, because of the lower semi-continuity of H . Finally, by Proposition 3.9, T satisfies $H^\alpha(T) \leq \frac{\Lambda^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2}$. \square

3.4 A New Distance d_α on the Space of Probability Measures

In [13], Xia also defines a new function giving a distance between two measures from $M_\Lambda(X)$, this time based on the cost function H^α discussed in the previous Section.

The Wasserstein distance was introduced by Russian mathematician Leonid Wasserstein in 1969. The Wasserstein distance between two measures μ and ν depends on the paired mappings from μ to ν , and the distance function on the points of X . Also, we note that for $p = 1$, this Wasserstein distance is often called the “earth mover distance,” as it corresponds to the Monge’s basic model of moving piles of earth from one place to another, with a minimal transport based on the distance between points and a given cost function c .

Definition 3.11. ([?]) Let (X, d) be a Polish metric space and $p \in (0, +\infty)$ (usually $p \geq 1$). For any two probability measures μ, ν on the probability space $P(X)$, the

Wasserstein distance of order p between μ and ν is defined by

$$W_p(\mu, \nu) := \left(\min_{\gamma \in \mathcal{M}} \int_{X \times X} |x - y|^p d\gamma(x, y) \right)^{\min(1, \frac{1}{p})}.$$

Here the set \mathcal{M} of all couplings μ and ν denotes the collection of all measures on $X \times X$ with marginals μ and ν on the first and second factors respectively.

Definition 3.12. (Xia, [13]) Let α be the cost parameter of a transport scheme, with $\alpha \in (1 - 1/m, 1]$. For any two Radon measures μ^+ and μ^- in $M_\Lambda(X)$, we set

$$d_\alpha(\mu^+, \mu^-) := \min\{H^\alpha(T) : T \in \text{Path}(\mu^+, \mu^-)\}.$$

From Xia's Existence Theorem, optimal paths minimizing the H^α cost exist, and this function is well-defined. In this section, we show that the function d_α is indeed a distance function, that is, that it satisfies the following properties for any Radon measures μ_1, μ_2 and μ_3 :

- (i) $d_\alpha(\mu_1, \mu_2) \geq 0$.
- (ii) $d_\alpha(\mu_1, \mu_2) = 0$ iff $\mu_1 = \mu_2$.
- (iii) $d_\alpha(\mu_1, \mu_2) = d_\alpha(\mu_2, \mu_1)$.
- (iv) $d_\alpha(\mu_1, \mu_3) \leq d_\alpha(\mu_1, \mu_2) + d_\alpha(\mu_2, \mu_3)$.

We will also present Xia's proof that this distance function is a natural one, in that it metrizes the weak* topology on $M_\Lambda(X)$.

Xia ([13]) notes that this distance function is different from the Wasserstein distance. In general, the Wasserstein distance measure results in V-shaped paths for optimal transport graphs from two sources to one sink, while Xia's cost function H^α and the resulting metric d_α correspond to Y-shaped paths.

We first show that because of the scaling feature from Proposition 3.6, we may without loss of generality assume in our proof that the total weight $\Lambda = 1$. To see this, we note that

$$\frac{T}{\Lambda} \in \text{Path}\left(\frac{\mu^+}{\Lambda}, \frac{\mu^-}{\Lambda}\right) \quad \text{with} \quad H^\alpha(T) = \Lambda^\alpha H^\alpha\left(\frac{T}{\Lambda}\right).$$

Hence,

$$\begin{aligned} d_\alpha(\mu^+, \mu^-) &= \min\{H^\alpha(T) : T \in \text{Path}(\mu^+, \mu^-)\} \\ &= \min\left\{\Lambda^\alpha H^\alpha\left(\frac{T}{\Lambda}\right) : T \in \text{Path}\left(\frac{\mu^+}{\Lambda}, \frac{\mu^-}{\Lambda}\right)\right\} \\ &= \Lambda^\alpha \min\left\{H^\alpha\left(\frac{T}{\Lambda}\right) : T \in \text{Path}\left(\frac{\mu^+}{\Lambda}, \frac{\mu^-}{\Lambda}\right)\right\} \\ &= \Lambda^\alpha d_\alpha\left(\frac{\mu^+}{\Lambda}, \frac{\mu^-}{\Lambda}\right). \end{aligned}$$

Therefore for any $\Lambda > 0$ and any $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$,

$$d_\alpha(\mu^+, \mu^-) = \Lambda^\alpha d_\alpha\left(\frac{\mu^+}{\Lambda}, \frac{\mu^-}{\Lambda}\right),$$

and thus we may assume that $\Lambda = 1$.

The following Lemma is another preliminary step in the proof that d_α is a distance function.

Lemma 3.13. (*Xia, [13]*) *Let $\mu \in \mathcal{M}_1(X)$, and suppose that $\{a_i\}, \{b_i\} \subset \mathcal{A}_1(X)$ are two sequences of atomic probability measures on X . If $a_i \rightarrow \mu$ and $b_i \rightarrow \mu$, then $d_\alpha(a_i, b_i) \rightarrow 0$.*

Proof. Let $\varepsilon > 0$. Since $\alpha \in (1 - \frac{1}{m}, 1]$ we have $m(1 - \alpha) - 1 < 0$. Then there exists a natural number n large enough so that

$$n^{m(1-\alpha)-1} \frac{1}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2} < \frac{\varepsilon}{3}.$$

For any small number $\beta > 0$, we can find a partition $Q_n = \{Q_n^h : h \in Z^m \cap [0, n]^m\}$ of Q consisting of cubes of edge length between $[(1 - \beta)\frac{d}{n}, (1 + \beta)\frac{d}{n}]$ such that for all i , the finite set $\text{spt}(a_i) \cup \text{spt}(b_i)$ doesn't intersect the boundary of those cubes, where $\text{spt}(a_i)$ is the support of the measure a_i . For each h , let c_n^h be the center of Q_n^h , $p_i^h = a_i(Q_n^h)$ and $q_i^h = b_i(Q_n^h)$. Since $a_i - b_i \rightarrow 0$, we have $p_i^h - q_i^h = (a_i - b_i)(\chi(\text{interior of } Q_n^h)) \rightarrow 0$ as $i \rightarrow \infty$ for all h . Let

$$p_i = \sum_{h \in Z^m \cap [0, n]^m} p_i^h \delta_{c_n^h} \text{ and } q_i = \sum_{h \in Z^m \cap [0, n]^m} q_i^h \delta_{c_n^h}.$$

By Proposition 3.9, there exists an $S_i^h \in \text{Path}(a_i|_{Q_n^h}, p_i^h \delta_{c_n^h})$ with

$$M^\alpha(S_i^h) \leq \frac{(p_i^h)^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2n}.$$

Thus $S_i = \sum_{h \in Z^m \cap [0, n]^m} S_i^h \in \text{path}(a_i, p_i)$ and

$$\begin{aligned} M^\alpha(S_i) &\leq \sum_{h \in Z^m \cap [0, n]^m} M^\alpha(S_i^h) \\ &\leq \sum_{h \in Z^m \cap [0, n]^m} (p_i^h)^\alpha \frac{(1)}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2n} \\ &\leq \sum_{h \in Z^m \cap [0, n]^m} \left(\frac{1}{n^m}\right)^\alpha \frac{(1)^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2n} \\ &\leq n^m \left(\frac{1}{n^m}\right)^\alpha \frac{(1)^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2n} \\ &< \frac{\varepsilon}{3}. \end{aligned}$$

Similarly, we may find some $S_i' \in \text{path}(b_i, q_i)$ with $M^\alpha(S_i') < \frac{\varepsilon}{3}$.

Finally, let G_i be the cone over $p_i - q_i$ with vertex c , the center of Q . Then

$G_i \in \text{path}(p_i, q_i)$ and when i is large enough we have

$$M^\alpha(G_i) \leq \sum_{h \in z^m \cap [0, n]^m} (|p_i^h - q_i^h|)^\alpha \frac{\sqrt{md}}{2} < \frac{\varepsilon}{3}.$$

Therefore, we have $T_i = S_i + G_i + S'_i \in \text{path}(a_i, b_i)$ with $M^\alpha(T_i) < \varepsilon$ when i is large enough. Thus $d_\alpha(a_i, b_i) \rightarrow 0$. \square

The next Lemma shows that the distance d_α between two measures is always greater than or equal to the 1-Wasserstein distance between them, another fact to be used in our distance proof.

Lemma 3.14. *For any $\mu^+, \mu^- \in \mathcal{M}_1(X)$, we have*

$$W(\mu^+, \mu^-) \leq d_\alpha(\mu^+, \mu^-),$$

where W is the Wasserstein distance on $\mathcal{M}_1(X)$.

Proof. By the definition of $d_\alpha(\mu^+, \mu^-)$, there is a sequence T_i of cycle-free transport graphs, built on an optimal sequence $\{c_i, d_i, T_i\}$ of transport paths from μ^+ to μ^- , such that

$$\lim_{i \rightarrow \infty} H^\alpha(T_i) = d_\alpha(\mu^+, \mu^-).$$

Lemma 3.5 shows that $H^1(T_i) \leq H^\alpha(T_i)$. The definition of the Wasserstein distance also shows that $W(a_i, b_i) \leq H^1(T_i)$. Combining these gives us

$$W(a_i, b_i) \leq H^1(T_i) \leq H^\alpha(T_i).$$

Then as $i \rightarrow \infty$, we have

$$W(\mu^+, \mu^-) \leq \liminf_{i \rightarrow \infty} W(a_i, b_i) \leq \lim_{i \rightarrow \infty} H^\alpha(T_i) = d_\alpha(\mu^+, \mu^-).$$

□

Corollary 3.15. (Xia, [13]) *If $d_\alpha(\mu_i, \mu) \rightarrow 0$, then $\mu_i \rightarrow \mu$.*

Proof. Lemma 3.14 shows that if $d_\alpha(\mu_i, \mu) \rightarrow 0$ we also have $W(\mu_i, \mu) \rightarrow 0$. It follows from this that the sequence μ_i also converges to μ .

□

Theorem 3.16. (Xia, [13]) *d_α is a distance on $\mathcal{M}_1(X)$.*

Proof. We show that for any metrics μ_1, μ_2 and μ_3 in $\mathcal{M}_1(X)$, the four required distance properties hold. The first three properties hold by the Definition of d_α and Lemma 3.14, and it remains to show the fourth property. We assume that there are a_i, b_i, c_i and d_i which converge respectively to μ_1, μ_2, μ_2 and μ_3 , with corresponding path sequences T_i for $d_\alpha(\mu_1, \mu_2)$ and P_i for $d_\alpha(\mu_2, \mu_3)$. Since these last two sequences converge as $i \rightarrow \infty$, we may choose a large enough $i \in \mathbb{Z}$ such that

$$\lim_{i \rightarrow \infty} H^\alpha(T_i) \leq d_\alpha(\mu_1, \mu_2) + \frac{\varepsilon}{3}$$

and

$$\lim_{i \rightarrow \infty} H^\alpha(P_i) \leq d_\alpha(\mu_2, \mu_3) + \frac{\varepsilon}{3}.$$

Then, as in the Proof of Lemma 3.13, we can also find a sequence of paths R_i from b_i to c_i , and a large enough value of i to ensure that $H^\alpha(T - I) \leq \frac{\varepsilon}{3}$. Combining these facts, we have a path sequence $T_i + R_i + P_i$ and a large enough i to make

$$\begin{aligned} d_\alpha(\mu_1, \mu_3) &\leq \liminf H^\alpha(T_i + R_i + P_i) \\ &\leq d_\alpha(\mu_1, \mu_2) + d_\alpha(\mu_2, \mu_3) + \varepsilon. \end{aligned}$$

Therefore, $d_\alpha(\mu_1, \mu_3) \leq d_\alpha(\mu_1, \mu_2) + d_\alpha(\mu_2, \mu_3)$, and d_α is a distance on $\mathcal{M}_1(X)$.

□

The following Corollary shows an important feature of our transport graph sequences, that if each T_i is optimal for its path, then the limit path R is also optimal.

Corollary 3.17. *(Xia, [13]) Let R be a transport path between metrics μ^+ and μ^- with approximating sequences a_i, b_i and T_i . If each of the sequence paths T_i is optimal as a path from a_i to b_i for each $i \geq 1$, then the path R is optimal from μ^+ and μ^- .*

Proof. Suppose T is an optimal transport path in $\text{Path}(\mu^+, \mu^-)$ and $\{a'_i, b'_i, F_i\}$ is an approximating graph sequence of T such that

$$\lim_{i \rightarrow \infty} H^\alpha(F_i) = M^\alpha(T) = d_\alpha(\mu^+, \mu^-).$$

Then, by Lemma 3.13,

$$\begin{aligned} H^\alpha(R) &\leq \liminf_{i \rightarrow \infty} H^\alpha(T_i) \\ &= \liminf_{i \rightarrow \infty} d_\alpha(a_i, b_i) \\ &\leq \liminf_{i \rightarrow \infty} d_\alpha(a'_i, b'_i) + d_\alpha(a_i, a'_i) + d_\alpha(b_i, b'_i) \\ &\leq \liminf_{i \rightarrow \infty} H^\alpha(F_i) = d_\alpha(\mu^+, \mu^-). \end{aligned}$$

It follows from the Definition of d_α that $d_\alpha(\mu^+, \mu^-) \leq M^\alpha(T)$. Therefore, $M^\alpha(T) = d_\alpha(\mu^+, \mu^-)$ and T is also optimal. \square

Finally, we can show that Xia's new d_α distance metric, although different from the Wasserstein distance, in fact induces the same topology on $\mathcal{M}_1(X)$ as it does, namely the weak* topology. We first show that atomic measures are dense in $(\mathcal{M}_1(X), d_\alpha)$.

Lemma 3.18. *For each $\mu \in \mathcal{M}_1(X)$, let $\{R_n(\mu)\}$ be the dyadic approximation of μ . Let for some constant $C = \frac{\sqrt{ml}/2}{2^{1-m(1-\alpha)} - 1}$ and $0 < \beta = 2^{1-m(1-\alpha)} < 1$. Then,*

$$d_\alpha(\mu, R_n(\mu)) \leq C\beta^n.$$

Proof. Since we are assuming that $\Lambda = 1$, we can assume that our space X sits within the cube $[0, 1]^m$ in m -space. Setting $Q_0 = Q$, we can form the dyadic representation $Q_n = Q_n^h : h \in \mathbb{Z}^m \cap [0, 2^n]^m$ as above. As before we assume that the cubes have edge length $\frac{1}{2^n}$ and centers c_n^h for the appropriate indices n and h .

Then we define the atomic measure $R_n(\mu) = \sum_h \mu(Q_n^h) \delta_{c_n^h}$, and since d_α is a distance, we have

$$d_\alpha(\mu, R_n(\mu)) \leq \sum_h d_\alpha(\mu|_{Q_n^h}, \delta_{c_n^h}).$$

By the Xia's existence theorem, we have

$$\sum_h d_\alpha(\mu|_{Q_n^h}, \delta_{c_n^h}) \leq \sum_h \frac{\mu(Q_n^h)^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{ml}}{2^{n+1}}.$$

Hence,

$$\begin{aligned} d_\alpha(\mu, R_n(\mu)) &\leq \sum_h \frac{\mu(Q_n^h)^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{ml}}{2^{n+1}} \\ &\leq \sum_h \left(\frac{1}{2^{nm}} \right)^\alpha \frac{1}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{ml}}{2^{n+1}} \\ &= \frac{\sqrt{ml}}{2^{1-m(1-\alpha)} - 1} 2^{n[m(1-\alpha)-1]-1} \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

□

Corollary 3.19. (Xia, [13]) For any $\mu^+, \mu^- \in \mathcal{M}_1(X)$, let $\{R_n(\mu^+)\}$ and $\{R_n(\mu^-)\}$ be the dyadic approximation of μ^+ and μ^- respectively. Then

$$|d_\alpha(\mu^+, \mu^-) - d_\alpha(R_n(\mu^+), R_n(\mu^-))| \leq 2C\beta^n.$$

Proof. We have from Lemma 3.18 that

$$d_\alpha(\mu^+, R_n(\mu^+)) \leq C\beta^n \text{ and } d_\alpha(\mu^-, R_n(\mu^-)) \leq C\beta^n.$$

Then

$$\begin{aligned} d_\alpha(\mu^+, \mu^-) &\leq d_\alpha(\mu^+, R_n(\mu^+)) + d_\alpha(R_n(\mu^+), R_n(\mu^-)) + d_\alpha(\mu^-, R_n(\mu^-)) \\ &\leq C\beta^n + d_\alpha(R_n(\mu^+), R_n(\mu^-)) + C\beta^n \\ &= 2C\beta^n + d_\alpha(R_n(\mu^+), R_n(\mu^-)). \end{aligned}$$

Hence,

$$\begin{aligned} &|d_\alpha(\mu^+, \mu^-) - d_\alpha(R_n(\mu^+), R_n(\mu^-))| \\ &\leq |2C\beta^n + d_\alpha(R_n(\mu^+), R_n(\mu^-)) - d_\alpha(R_n(\mu^+), R_n(\mu^-))| \\ &= 2C\beta^n \end{aligned}$$

□

Theorem 3.20. (Xia, [13]) Xia's distance metric d_α metrizes the weak* topology of $\mathcal{M}_1(X)$.

Proof. We know from Lemma 3.14 that $W(\mu^+, \mu^-)$ is bounded above by $d_\alpha(\mu^+, \mu^-)$, so to prove that $W(\mu^+, \mu^-)$ converges to zero it suffices to show that $d_\alpha(\mu^+, \mu^-)$ does too. For $i \geq 1$, Lemma 3.18 lets us find an atomic probability measure a_i with $d_\alpha(a_i, \mu_i) \leq \frac{1}{2^i}$, so that a_i converges to μ , and a sequence b_i of measures that converges to μ , such that $d_\alpha(b_i, \mu)$ converges to zero. Lemma 3.13 tells us that $d_\alpha(a_i, b_i)$ then converges to zero

too, and thus

$$\begin{aligned}
 d_\alpha(\mu_i, \mu) &\leq d_\alpha(\mu_i, a_i) + d_\alpha(a_i, b_i) + d_\alpha(b_i, \mu) \\
 &\leq \frac{1}{2^i} + d_\alpha(a_i, b_i) + \frac{1}{2^i} \\
 &= \frac{1}{2^{i-1}} + d_\alpha(a_i, b_i) \\
 &\longrightarrow 0.
 \end{aligned}$$

□

Conclusion

This chapter has focused on describing Xia's process for finding optimal transport paths, with optimal paths over various possible paths, as well as explaining Xia's proofs from [13]. The idea of the optimal transport path first came from Monge in 1781, in the setting where the mass moved from any one source cannot be split up in the transport to a sink; Kantorovich later relaxed this condition to allow for a mass to be split and transported along multiple paths. In Chapter 3 we showed how to find an optimal transport path between two probability measures and also set an optimal cost function so that the total transportation cost is minimized. Then we proved that there exists a cycle free path on a transportation problem and showed some examples on this. First we discussed Xia's method for Radon measures, and then showed how to extend this to general measures using sequence approximations. Then we proved Xia's Existence Theorem [13] that there exists an optimal transport path between two measures with minimum cost. Finally, the Wasserstein distance function was introduced and showed that this is a distance function and metrizes the weak* topology.

Bibliography

- [1] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 4, 1991.
- [2] Lawrence C. Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, 1991.
- [3] N. Ghoussoub and A. Moameni. Symmetric monge-kantorovich problems and polar decompositions of vector fields. *Geom. Funct. Anal.*, 24, 2014.
- [4] T. J. Jech. Set theory, 2nd ed. *Berlin: Springer-Verlag*, 1997.
- [5] L. V. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
- [6] G. Monge. Mémoire sur la théorie des déblais et de remblais. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, 1781.
- [7] R. J. McCann P.A. Chiappori and L. P. Nesheim. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Econom. Theory*, 42, 2010.
- [8] Clayton R. Paul. Fundamentals of electric circuit analysis. *John Wiley and Sons*, 2001.
- [9] Keijo Rouhonen. *Graph Theory*. 2013.
- [10] Walter Rudin. Real and complex analysis. *McGraw-Hill*, 1987.
- [11] Michele Pagani Thomas Ehrhard and Christine Tasson. Measurable cones and stable, measurable functions. *Proc. ACM Program. Lang.*, 2, 2018.
- [12] C. Villani. *Optimal transport, Old and new*. Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.
- [13] Qinglan Xia. Optimal paths related to transport problems. *Communications in Contemporary Mathematics*, 5(2):251–279, 2003.
- [14] Qinglan Xia. Motivations, ideas and applications of ramified optimal transportation. *ESAIM, Math*, 51(49):1791–1832, 2015.
- [15] Yoav Zemel. Optimal transportation: Continuous and discrete. Master's Thesis. *Ecole Federale de Lausanne*, 2012.