UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Shortest Path Approximation and Optimal Transport with Flow-rate Constraints

Permalink

https://escholarship.org/uc/item/4dt0b39d

Author

Dong, Anqi

Publication Date

2023

Copyright Information

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Shortest Path Approximation and
Optimal Transport with Flow-rate Constraints

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mechanical and Aerospace Engineering

by

Anqi Dong

Dissertation Committee:
Distinguished Professor Tryphon T. Georgiou, Chair
Professor Faryar Jabbari
Associate Professor Solmaz Kia
Assistant Professor Yanning Shen

2023

DEDICATION

To my parents
Yunxiu Wang and Guicheng Dong

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the following individuals and organizations who have contributed to the completion of this thesis:

First and foremost, I would like to thank Professor Tryphon T. Georgiou. I consider myself extremely lucky and greatly honored to have Tryphon as my Ph.D. advisor. For me, Tryphon is not only a great academic advisor who has provided me with patient guidance and taught me invaluable things during these six years, but also a close friend. Words are so pale to truly express the depth of my gratitude for Tryphon's unwavering care and support, treating me like his own child. The time we have spent together is, and undoubtedly will be, the most precious memories I will never forget.

I would also like to express my gratitude to Prof. Amirhossein Taghvaei, who has mentored me and helped me produce my first paper. I sincerely appreciate his generous support and all the suggestions he has provided.

I would love to thank Prof. Faryar Jabbari, Prof. Solmaz Kia, and Prof. Yanning Shen for their presence on my committee. I would like to express my appreciation for the time and effort that each committee member dedicated to reviewing and evaluating my work. Their commitment to scholarly excellence and their willingness to provide guidance and feedback have been invaluable. I am deeply grateful for their exceptional support, expertise, valuable input, and thoughtful suggestions. The discussion between us is very constructive and educational. I would also like to thank Prof. Vijay Vazirani for his series of courses in matching and the ACO seminars.

VITA

Anqi Dong

EDUCATION

Doctor of Philosophy in Mechanical Engineering   **2023**
University of California, Irvine   *USA*

Master of Science in Mechanical Engineering   **2020**
University of California, Irvine   *USA*

Bachelor of Science in Mechanical Engineering   **2017**
Harbin Institute of Technology   *China*

RESEARCH EXPERIENCE

Graduate Research Assistant   **2021–2023**
University of California, Irvine   *California*

TEACHING EXPERIENCE

Teaching Assistant   **2017–2021**
University of California, Irvine   *California*

REFEREED JOURNAL PUBLICATIONS

A lasso-alternative to Dijkstra algorithm for identifying   **2022**
short paths in networks
Submitted to Automatica

Optimal transport through a toll   **2022**
Submitted to European journal of applied mathematics

Monge-Kantorovich Optimal Transport Through Con-   **2023**
strictions and Flow-rate Constraints
Submitted to Automatica

REFEREED CONFERENCE PUBLICATIONS

Lasso formulation of the shortest path problem   **2020**
59th IEEE Conference on Decision and Control

SOFTWARE

Python   `https://github.com/dytroshut`
TensorFlow, PyTorch, Keop

MATLAB

ABSTRACT OF THE DISSERTATION

Shortest Path Approximation and
Optimal Transport with Flow-rate Constraints

By

Anqi Dong

Doctor of Philosophy in Mechanical and Aerospace Engineering

University of California, Irvine, 2023

Distinguished Professor Tryphon T. Georgiou, Chair

In an increasingly interconnected world, the efficient and economical transportation of individuals and commodities has emerged as a cornerstone of modern society. Optimizing transportation plans has a huge potential for journey planning, congestion reduction, supply chain management, and data exchanges. These strategies hold immense relevance not only in the realm of engineering and transportation, but also in other fields, such as physics, computer science, economics, and several subject areas in mathematics. The present thesis aims to elucidate the optimization of transportation strategies, with a particular focus on two classical problems, finding short paths in large networks and solving optimal transport problems with flow-rate constraints.

The so-called shortest path problem seeks an optimal path of transporting one unit of mass between pairs of vertices on graphs. We present a novel formulation of the problem as an l_1-regularized regression, often referred to as lasso (Least Absolute Shrinkage and Selection Operator). Based on this formulation, we draw a connection specifically between trees that grow as active edge-sets in the least angle regression (LARS) algorithm of the lasso problem, and respective shortest-path trees that emerge using the bi-directional Dijkstra algorithm. Then, to overcome the dimensionality challenge in large graphs, we explore the alternating direction method of multipliers (ADMM) in the lasso formulation. The resulting derivative

proximal algorithm speeds up the search for the short paths, trading off optimality (i.e., finding shortest paths) that may not be absolutely essential in a variety applications.

The basic transport problem is motivated by the need to transport resources/mass between end-point distributions (supply and demand). We consider the classical Monge-Kantorovich optimal transport problem with a quadratic cost functional to penalize distance of transport, with an added constraint that transported mass is required to pass through constriction points while abiding by specified allowable flow-rate; constriction points may be conceptualized as toll stations with limited throughput. Our contributions in this topic are as follows: **(1)** we provide a precise Monge formulation for the optimal transport problem with flux constraint at constriction sites along the path that is amenable to generalization in higher dimensions. We work out in detail the case of transport in one dimension by proving existence and uniqueness of solutions. Under suitable regularity assumptions we give an explicit construction of the transport plan; **(2)** we provide a Kantorovich-type re-formulation of the problem by introducing a marginal probability density for the time that mass-elements cross toll stations –a probability density that is to be determined so as to meet given flow-rate constraints. Interestingly, the Kantorovich-type formalism leads to multi-marginal optimal transport problem that is readily solvable by using linear programming. Moreover, existence and uniqueness of solutions are also established in this setting. Then, **(3)** we propose an entropic penalty term to regularize and reduce the computational cost of resulting multi-marginal problems. Entropic regularization of standard optimal transport leads to an efficient algorithm, the *Sinkhorn* algorithm, which applies in the present case as well. Leveraging the splittable nature of the cost in our formulation, we proposed a *Gluing Sinkhorn* algorithm for the multi-marginal optimal transport problem, which reduces the computational cost to a level comparable to that in standard two-marginal problems.

Notation

For the convenience of the readers, frequently used notations are listed

Symbols	Definitions
$\mathcal{G}$	graph $G := (\mathcal{V}, \mathcal{E})$
$\mathcal{V}$	vertex set with
$\mathcal{E}$	edge set
$\mathcal{W}$	weight set
A	adjacency matrix
D	incidence matrix
L	graph Laplacian
μ, ν, σ	marginal
$\text{Supp}(\cdot)$	support of probability distribution
T	push-forward map $T : x \to y$
$c(x, y)$	cost between x and y
$\pi(x, y), \pi(\mu, \nu)$	coupling/ transportation plans
ρ	probability density
X_t	flow at time t
$\mathcal{W}_2(\mu, \nu)$	Wasserstein-2 distance between μ and ν
$v_i/v_t(x)$	vertex i of a graph/ velocity at time t
$\mathbb{1}_n/\mathbb{1}_{\{\cdot\}}$	all-one vector/ indicator function
$\mathcal{M}$	Computational domain
h/r	momentum/flow-rate bound
ξ	location of the toll
$\mathcal{H}(\cdot)$	entropy

Chapter 1

Introduction to transportation

In this thesis, we would like to address a common (maybe one of the most frequently asked) question:

> *–Does there exist an optimal path/map between given origin and destination?*
>
> *–If so, how can I find it?*

The significance of such a problem not only shows up in our daily life (especially when we are using Google Maps or Uber Eats), but echoes in control, economics, and computer science research. Transportation can happen between (a group of) robots/drones, goods, information and data, so that the theory we developed can be applied in numerous fields. In particular, we discuss two well-known problems over the past century,

1. *The shortest path problem*

2. *Optimal transport problem*

Before we formulate the problem and talk about technical results, let us expand on the fundamental idea of these two problems:

The problem of determining paths was first considered by Leonhard Euler [39], who proved that the solution of the *Seven Bridges of Königsberg* does not exist in 1735. The characterization of the city, using a graph with vertices and edges, laid the foundation of graph theory.

Figure 1.1: Seven Bridges of Königsberg (Historic Cities Research Project)

The shortest path problem, by its name, seeks the shortest path between two vertices (starting/targeting) over a graph. The *length* of the path is defined as the summation of weights of all the edges along the path. Thus, the shortest path problem is a *point-wise*[1] transportation from one location to another along the shortest route. Even though the problem has a long history, the solution of finding the shortest path emerged in the last century with the contributions (almost at the same time) of Ford [43], Edsger W. Dijkstra [35], Richard Bellman [7], Dantzig [34], and many others [45, 93].

Forty-six years later, the optimal transport problem was formalized by the giant Gaspard Monge in 1781. The Monge problem [74], seeks an optimal plan for transporting the soil(e.g.,

---

[1]The original problem is from one vertex to another (*one-pair*), the *all-pair* shortest path problem has also been considered, but is known as an NP-hard problem

from a hill to a hole) when constructing forts and roads. Distinguished from our first problem, the optimal transport problem is an inherently *group-wise* problem, i.e., transportation takes place between a group of supplies and demands. An optimal plan is assigned with a given cost and varies with the choice of the cost. For example, the cost between every supply and demand can be the length of the path. Similar to the shortest path problem, after 155 years of waiting, the original Monge problem was advanced by Leonid Kantorovich [59] in 1942, with the Kantorovich relaxation and Kantorovich duality, which allows the mass to be split.

Our work is organized as follows: In Chapter 1, we introduce the two transportation problems with some preliminaries. In Chapter 2, we briefly review some fundamental concepts in the topic of optimization. In Chapter 3, the l_1-regularized shortest path problem is presented. In Chapter 4, the optimal transport problem with tolls is solved in its Monge formulation. In Chapter 5, the toll problem abiding the flow-rate constraint is formulated as a Kantorovich relaxation of the Monge problem under the framework of the multi-marginal transport problem. In Chapter 6, the entropic regularization of the Kantorovich problem is proposed with a modified *gluing Sinkhorn* algorithm for efficient computation.

In the following sections, we formally introduce the two problems in their standard form, along with the corresponding notations.

1.1   Transportation on a graph

Transportation happens in multiple forms over graphs in different settings. Besides the pair-wise shortest path problem we studied, a close variant will be the all-pair shortest path [93]. Moreover, transportation like the traveling salesman problem [67, 4] considers visiting all the vertices with minimum cost, and the idea of moving with randomness is introduced as random walks on graph [69, 22], and basically building the foundation of Google's searching engine [13].

1.1.1 Graph and matrices

A graph $\mathcal{G}$ is specified by a pair of sets,

$$\mathcal{G} := \{\mathcal{V}, \mathcal{E}\}$$

where $\mathcal{V}$ represents the set of nodes/vertices, herein always assumed to have finite cardinality $|\mathcal{V}| = n$, i.e.,

$$\mathcal{V} := \{v_1, v_2, \ldots, v_n\},$$

and where $\mathcal{E}$ represents the set of edges

$$\mathcal{E} = \{e_1, e_2, \ldots, e_m\},$$

with cardinality $|\mathcal{E}| = m$. Graphs are specified further to be directed or undirected. For specificity, we often denote edges with a double index, to indicate the vertices they link, i.e., $e_{ij} = (v_i, v_j)$ represents an edge that connects vertices v_i and v_j, that may also encode the directionality $v_i \rightarrow v_j$ with $v_i \ v_j$ being the *head* and *tail* of the edge in the case of a directed graph, respectively. In the case of an undirected graph, the order is inconsequential as transport along both directions is permitted.

Edge-weights (e.g., w_{ij} corresponding to edge e_{ij}) are often used to quantify cost or ease of flow along the corresponding edge. A graph with uniform edge-weights is known as an unweighted graph. A weighted graph is specified by a triple

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{W}\},$$

by adjoining the weight set $\mathcal{W} = \{w_1, w_2, \ldots, w_m\}$, corresponding to the edge set $\mathcal{E}$.

The graph can coincide with more complex scenarios, e.g., we may consider the *self-loop* for vertices and *multi-edges* between one pair of vertices. Another important concept is the connectivity of the graph, the graph is said to be connected if there always exists a path between any pair of vertices. A connected subgraph is also known as a *component*. Herein, we always consider positive-weighted connected graphs without self-loops and multi-edges.

Alternatively, the graph can be characterized by associated matrices, which is understood under the framework of *linear algebraic graph theory*. We refer to [5] for a comprehensive survey on this topic. Herein, we hand out the definition and some nice properties of the *adjacency matrix*, *incidence matrix*, and *graph Laplacian*.

Definition 1.1 (Incidence matrix). *The incidence matrix D is a n-by-m matrix with elements*

$$[D]_{ij} = \begin{cases} -1 & \text{if } v_i \text{ is the head of edge } e_j, \\ 1 & \text{if } v_i \text{ is the tail of edge } e_j, \\ 0 & \text{otherwise.} \end{cases}$$

The incidence matrix directly characterizes the relationship between vertices and edges. A pair of vertices is *incident* if they are connected by an edge. By its definition, each column of the incidence matrix has only two nonzero elements, and thus the matrix is sparse. The incidence matrix is generically defined for a directed graph since the direction of the edges is considered (head and tail). For the undirected graph, a $0-1$ incidence matrix can be similarly defined with only 0 and 1.

The adjacency of the graph can be also represented by the adjacency matrix as follows

Definition 1.2 (Adjacency matrix). *The adjacency matrix is an n-by-n (square) matrix*

6

with elements

$$[A]_{ij} = \begin{cases} 1 & \textit{if vertex } v_i \textit{ and } v_j \textit{ are adjacent,} \\ 0 & \textit{otherwise.} \end{cases}$$

It can be instantly observed that the adjacency matrix is symmetric, i.e., $[A]_{ij} = [A]_{ji}$. Moreover, the row/column sum of the ith row/column is the *degree* (the number of neighbors) of vertices v_i. The power of the adjacency matrix is fruitful as well, in the sense of understanding how well the graph is connected.

Lemma 1.1 (Power of A). *The value of i, j-th elements of $A^x, \forall\ x \in \mathbb{N}$, i.e., the x power of the adjacency matrix A, equals to the number of paths between v_i and v_j with x edges.*

Finally, with arbitrarily assigned orientation, the graph Laplacian (also, Laplacian matrix) can be then obtained according to the following definition.

Definition 1.3 (Graph Laplacian). *The graph Laplacian L is a n-by-n matrix so that $L = DD^T$ where D is the incidence matrix.*

$$[L]_{ij} = \begin{cases} \text{degree of vertex } i, & \text{if } i = j \\ -1, & \text{if } e_{ij} \in \mathcal{E} \\ 0, & \text{if } e_{ij} \notin \mathcal{E} \end{cases}$$

From the Definition 1.1, we can easily observe some properties of the graph Laplacian. For instance, L is a symmetric matrix and also positive semidefinite. Moreover, the graph Laplacian can capture the Laplacian quadratic form, i.e.,

$$x^T L x = \sum_{i,j} (x_i - x_j)^2.$$

where $x \in \mathbb{R}^n$ is the state vector of vertices. The state vector x can capture physical features such as voltage assigned to every vertex. Also, for general states, the Laplacian characterizes a first-order system as

$$\dot{x} = -Lx$$

Below, we also summarize some graphs with specific structures to give a better view of this topic.

Remark 1.1. *There exist some graphs with a specific structure, that have more fruitful properties than general ones. Herein we list some of them for readers' interests.*

1) *Random graph: the random graph, by its name, introduces a certain level of randomness to the graph structure. For instance, an edge can be assigned to a pair of vertices with respect to some probability. One of the most famous examples is the Erdős–Rényi model. See, e.g., [38].*

2) *Tree: the graph is known as a tree when loops do not exist. Equivalently, for a connected tree with only one component, the path between any pair of vertices is unique.*

3) *Planar graph: The planar graph is the group of the graph that can be embedded in a plane.*

4) *Regular graph: the graph is a regular graph if each vertex has the same number of neighbors. More specifically, the graph is a k-regular graph is every vertex has k neighbors.*

There are many good references to understanding graph theory/network science. For example, [78] gives a nice overview of various applications; [1] provides a comprehensive study of network work problems; [5] focus more on the matrices of the graph and their properties; in [73] the multi-agent network are considered in the language of control. Also, the spectral graph theory [29].

1.1.2 Network flow problems

Network flow problems are well-studied in the field of computer science and have a natural connection to transportation. In this section, we would like to explain two kinds of flow problems and their variants in a nutshell. The flow problems can be categorized into two main streams, i.e., the minimum cost flow problem and the maximum flow problem. Herein, we will briefly review them for the readers. Needless to say, both problems are salient in computer science and graph theory. We refer the readers to [1] for a comprehensive study on them. The readers can always consider the example (Fig.1.2) below, where transportation starts from vertex $v_s(= v_1)$ and arrive at $v_t(= v_9)$. The weights on each edge can be interpreted as the capacity in the maximum flow problem, the transportation cost in the minimum cost flow problem, and the length in the shortest path problem.
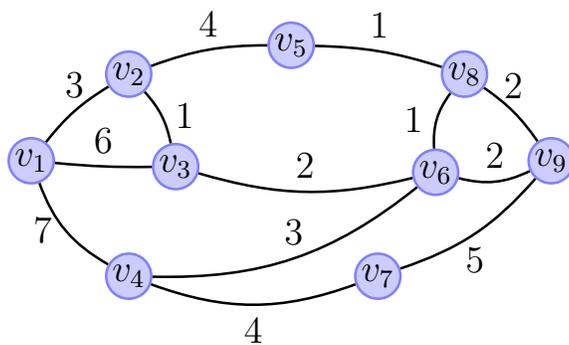
Figure 1.2: Example: Nicholson's graph

The maximum flow problem/ minimum cut problem

The Max-flow problem [42] of a network is to consider a digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, for each edge $e \in \mathcal{E}$, there exists a capacity $c(e)$, and we consider maximizing the flow that can go through the whole network from a given vertex v_s to a target vertex v_t.

Problem 1.1. *Denote $f_{i,j}$ as the flow goes through v_i to v_j, we can then propose the linear*

programming formulation to maximize the flow $f_{s,t}$:

$$\max \quad f_{s,t}$$

$$\sum_i f_{i,j} = \sum_j f_{j,i}, \quad \forall \quad v_i, v_j \in \mathcal{V}/\{v_s, v_t\} \qquad \text{(Capacity conditions)}$$

$$0 \le f_{i,j} \le c(e_{ij}) \qquad \qquad \text{(Conservation conditions)}$$

where the flow is defined as

$$f_{s,t} = \sum_{v_j \in \mathcal{V}} f^{out}(v_s, v_j) - \sum_{v_j \in \mathcal{V}} f^{in}(v_s, v_j) \qquad (1.1)$$

with $f_{in}(v_s, v_j)$ is the in-flow from neighborhood of v_s and $f^{out}(v_s, v_j)$ is the out-flow.

The constraints guarantee the following: first, the quality of the out-flow is equal to the in-flow for the rest vertices except v_s and v_t; secondly, the flow is bounded by the capacity of the edge; finally, only positive flow is allowed. The well-known algorithm to solve the max-flow problem is the Ford–Fulkerson algorithm.

Algorithm 1 Ford–Fulkerson algorithm

Input: Graph $\mathcal{G}$, $\mathcal{G}_f$
Output: flow f
 1: **while** there exists an (s, t) path P in $\mathcal{G}_f$ **do**
 2:    Consider (f, c, P), do $\delta \leftarrow$ bottleneck capacity of augmenting path
 3:    **for each** $e \in P$
 4:    **if** $e \in \mathcal{E}$ **then**
 5:       $f(e) + \delta \to f(e)$
 6:    **else**
 7:       $f(e^{reverse}) - \delta \to f(e^{reverse})$
 8:    **end if**
 9:    Update $\mathcal{G}_f$
10: **end while**
11: **return** f

The residual graph $\mathcal{G}_f$ allowed us to undo the bad path and adjust it by the bottleneck

capacity so that we can choose another path P. The result from [1, Theorem 6.5 (Integrality Theorem)] for the maximum flow problem guarantees the existence of an integer maximum flow.

Example 1.1. *Considering Nicholson's graph and seeking the maximum flow from v_1 to v_t, we have $f_{s,t} = 8$. The edges are highlighted in red in Fig. 1.3.*
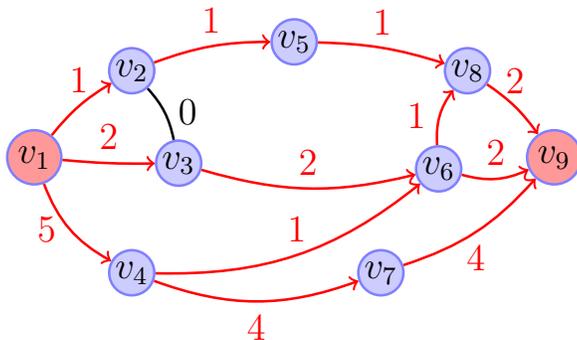
Figure 1.3: Maximum flow from v_1 to v_9 over Nicholson's graph

The minimum cost flow problem

The minimum cost flow, considering the case of transporting a certain amount of mass (say, b) from a supply vertex v_s to a demand vertex v_t. The weight c_{ij} assigned to each is the cost of passing one unit of mass through e_{ij}. The problem then has the form

Problem 1.2 (Minimum cost flow). *Given the transporting cost of every edge $e \in \mathcal{E}$ and mass b, determine the flow $f(e_{ij}), \forall\ e_{ij} \in \mathcal{E}$ that minimizes*

$$\sum_{i,j} c_{ij} f(e_{ij})$$

obeying the constraints

$$f(e_{ij}) + f(e_{ji}) = 0, \sum_{v_s, v_t \notin \mathcal{V}} f(e_{ij}) = 0, \sum_{v_j \in \mathcal{V}} f(e_{s,j}) = d, \sum_{v_j \in \mathcal{V}} f(e_{t,j}) = d.$$

11

The integrality of the minimal flow can be then obtained by [1, Theorem 9.10 (Integrality property)] using the result of [1, Theorem 6.5 (Integrality Theorem)], and the cycle-canceling algorithm connects the two problems.

1.1.3 Shortest path problem

Now we are in the position to define a *path* for undirected graphs[2], and hence formulate the shortest path problem.

The shortest path problem can be considered as a special case of the minimum cost flow problem by assuming the supply of s equals 1 and the demand of $t = -1$ (transporting one unit of mass). For this problem, we aim to find a path of minimum cost (or length) from a specified vertex v_s to a targeting vertex v_t. Some of the simplest applications of the shortest path problem are to determine a path between two specified vertices of a network that has minimum length, a path that takes the least time to traverse, or a path that has the maximum reliability. A path $p(s, t)$ between v_s and v_t can be defined as a set of edges,

$$p(s, t) = \{(v_s, v_1), (v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_t)\}.$$

Thus, the shortest path problem is then defined as follows,

$$d(s, t) = \min_{p(s,t) \in \mathcal{P}_{st}} \text{length} \ p(s, t) \tag{1.2}$$

where $\mathcal{P}_{st}$ denotes the set of all the admissible paths and $d(s, t)$ is known as the *distance* between the two vertices. Even though the formula seems trivial, finding the optimizer is not so easy. We will detail this problem and our new approach in Chapter 3.

---

[2]We will show that the definition of path can be easily generalized for a directed graph in Chapter 3

Dijkstra's algorithm

Dijkstra's algorithm [35] to find the shortest path from v_s to v_t requires maintaining and updating values for the following variables,

$dist$:   an array of distances from the root vertex v_s to all

         the other vertices in the graph,

$\mathcal{S}$:   the set of visited vertices up to the present step, and

$\mathcal{Q}$:   the queue of vertices to be visited.

The algorithm begins with an initial value of $+\infty$ for all distances. It updates the values at each step as follows:

Algorithm 2 Dijkstra's algorithm

Input: source vertex v_s and target vertex v_t.
Output: the shortest path and the length of the path.
1:   $dist^{(0)}[v_s] = 0$, $dist^{(0)}[v_i] = \infty, \forall v_i \neq v_s$,
      $\mathcal{S} = \emptyset$, $\mathcal{Q} = \mathcal{V}$.
2:   **while** $v_t \notin S$ **do**
3:     pick u from Q with minimum distance:
         $u = \underset{v \in Q}{\arg\min}\, dist[v]$
4:     Remove u from Q: $Q \leftarrow Q \setminus \{u\}$
5:     Add u to S: $S \leftarrow S \cup \{u\}$
6:     **for** $v_i \in neighbors[u]$ **do**
7:       **if** $dist[v_i] > dist[u] + w_{u,v_i}$ **then**
8:         $dist[v_i] \leftarrow dist[u] + w_{u,v_i}$
9:       **end if**
10:    **end for**
11: **end while**
12: **return** $dist[v_t]$

1.2   Transportation between distributions

The topic of optimal transport (OT), as we previously discussed, has a long history and raises interest in diverse fields, e.g., machine learning and thermodynamics. The goal of optimal transport is to find a transportation map T that takes mass from the supply distribution μ

to the demand distribution ν. Of course, transportation is not free in general, and thus a cost c is always introduced to our problem.

In this section, we will briefly review the three most classical formulations of the problem: the Monge problem, its Kantorovich relaxation, and an equivalent dynamical formulation when having a quadratic cost. In the end, we briefly review a generalized multi-marginal problem.

Now we are in the position to formulate the optimal transport problem. Denote the probability spaces as $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$, where μ, ν are the probability measures and $\mathcal{X}, \mathcal{Y}$ are the measure space. The *cost function* is then defined on a product space $\mathcal{X} \times \mathcal{Y}$, which is a measurable map from $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+ \cup \{+\infty\}$.

1.2.1 Monge problem

The Monge problem considers the transportation of non-splittable mass, i.e., the mass at location x must be transported to a unique destination y. The transportation map is thus introduced as the push-forward map $T : \mathcal{X} \mapsto \mathcal{Y}$, so that[3]

$$\nu = T \sharp \mu,$$

meaning the push-forward map transport μ onto ν and ν is known as the push-forward. Moreover, the push-forward map is defined as

$$\nu(B) = \mu(T^{-1}(B)) \ \text{ for any measurable set } \ B \subset \mathcal{Y}. \tag{1.3}$$

Let the transportation cost be $c(x, T(x))$, the Monge transport problem reads

Problem 1.3 (Monge transport). *The Monge problem minimizes the transportation cost*

---

[3]As it is standard, the notation $\sharp$ denotes the "push forward".
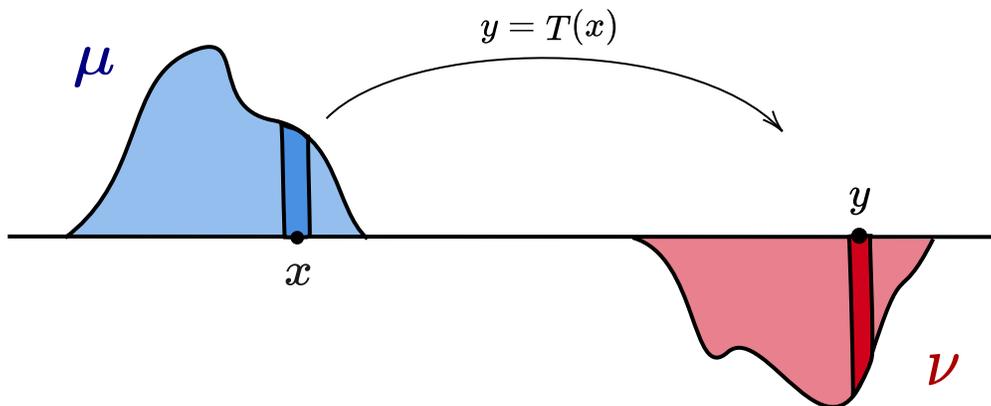
Figure 1.4: Monge problem

functional

$$\arg\min_{T} \int_{\mathcal{X}} c\big(x, T(x)\big) d\mu(x) \tag{1.4}$$

over all the admissible T satisfying (1.3).

Two fundamental questions of the Monge transport can be thus answered:

1. *Does the Monge solution (T) always exist?*

2. *If the Monge solution exists, is it unique?*

Unfortunately, the answers to both of the questions are negative, and we give two classic counter-examples (Remark 1.2 and 1.3) for the readers. We will revisit these questions for different formulations (multiple times) throughout.

Remark 1.2 (**Existence of Monge solution**). *The existence of an optimizer for the Monge problem is not guaranteed. For instance, consider the supply is a Dirac delta at x while the demands are two Dirac deltas at y_1 and y_2, since the mass is not splittable at x, the map T is not well-defined.*

On the other side, even if the optimal map exists, it may not be unique. To see this, we introduce the case of *book shifting* from [91, Example 2.16]

15

Remark 1.3 (Book shifting). *In the one-dimensional book shifting, we consider the trans-portation cost $c(x, y) = |y - x|$, and let $\mathcal{X} = [0, 2]$ and $\mathcal{Y} = [1, 3]$. The maps*

$$T(x) := x + 1 \quad and \quad T'(x) := \begin{cases} x + 2, & if \ x \in [0, 1] \\ x, & if \ x \in (1, 2] \end{cases}$$

are both optimal so that for (not strictly convex) costs $c(x, y)$, the optimal transport map is not unique.

Since the Monge problem "matches" the mass between x and y, a related problem is the stable matching problem (also, stable marriage problem [50]), where the cost function is replaced by preference lists[4].

1.2.2   Kantorovich problem

The Kantorovich formulation, as a relaxation of the original Monge problem, allows the mass to split by defining the *coupling* π on the product $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν. Thus, $d\pi(x, y)$ represents the amount of mass transported from x to y, and $c(x, y)$ is the corresponding cost. Next, we formally define the coupling and formulate the Kantorovich problem.

Definition 1.4 (Coupling). *The coupling $\pi(x, y)$ is defined as a probability measure of the product space $\mathcal{X} \times \mathcal{Y}$. Moreover, the projection of $\pi(x, y)$ onto $\mathcal{X}$ coincides with the marginal $d\mu(x)$, and the projection of $\pi(x, y)$ onto $\mathcal{Y}$ coincides with marginal $d\nu(y)$, meaning*

$$\int_{\mathcal{Y}} d\pi(x, y) = d\mu(x), \quad \int_{\mathcal{X}} d\pi(x, y) = d\nu(y).$$

---

[4]In the stable marriage problem matching between a two side market, e.g., men and women, with respect to their individual preference list, the matching is considered to be stable if and only if the no pair of men and women can be both matched to better candidates in their preference list. Similar to the Monge problem, the matching is, in general, not unique. The well-known solver for the problem is the Gale-Shapley algorithm [46]

We can then define the set Π of all the admissible couplings with marginals μ and ν by

$$\Pi(\mu,\nu) = \left\{ \pi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+ \,\middle|\, \int_{\mathcal{Y}} d\pi(x,y) = d\mu(x), \quad \int_{\mathcal{X}} d\pi(x,y) = d\nu(y) \right\}.$$
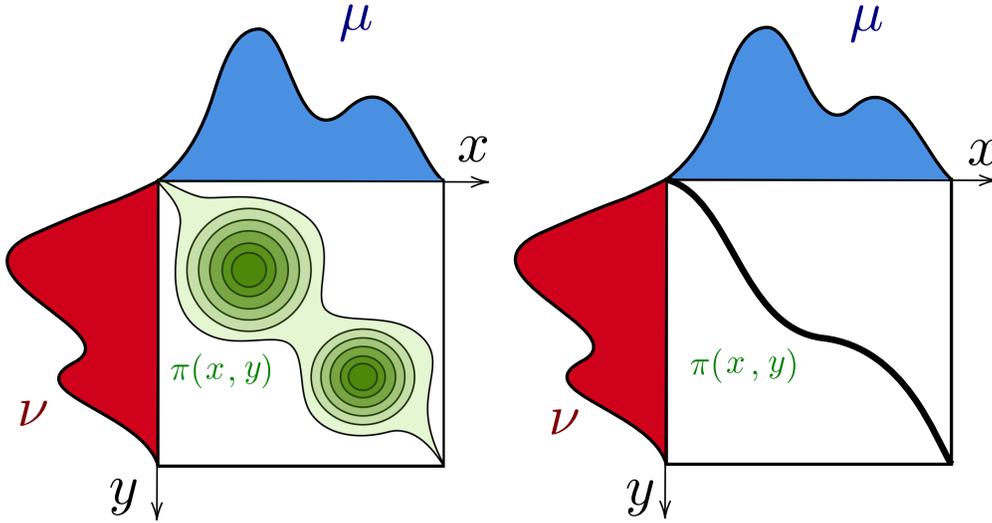
Figure 1.5: Coupling with marginal μ and ν

The gluing lemma below allows to *glue* two Kantorovich maps sharing a common marginal together and becomes a useful tool in proving the triangle inequality of the Wasserstein distance, and its significance shows up in the analysis of the multi-marginal transportation as we will show later

Lemma 1.2 (Gluing lemma [105, Lemma 7.6]). *Let μ, ν, σ be the three probability measures, with support in spaces $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$, respectively. Let the couplings*

$$\pi \in \Pi(\mu,\sigma) \text{ and } \pi' \in \Pi(\nu,\sigma),$$

which share the common marginal σ. Then there exists a probability measure $\pi^\star$ on the product space $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ with marginals π on $\mathcal{X}_1 \times \mathcal{X}_3$ and π' on $\mathcal{X}_2 \times \mathcal{X}_3$.

Problem 1.4 (Kantorovich transport). *Given the cost $c(x,y)$, the Kantorovich problem*

can then be written as

$$\operatorname*{arg\,min}_{\pi \in \Pi} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y) d\pi(x,y)$$

with given marginals μ and ν.

Remark 1.4 (Relaxation). *The Kantorovich formulation can be considered as a relaxation of the Monge problem. In the sense that the formulation allows the mass to be split, and the variable changes from $T(x)$ to $\pi(x,y)$, the coupling π can be rewritten in terms of the map T (when T exists) as $d\pi(x,y) = d\pi_T = d\mu(x)\delta_{y=T(x)}$ with δ denotes the Dirac delta function.*

Next, we see the dual of the Kantorovich problem reads

Problem 1.5 (Dual problem). *For integrable functions $\varphi \in L^1(\mathcal{X}.\mu)$ and $\psi \in L^1(\mathcal{Y}.\nu)$, the dual problem has the form*

$$\operatorname*{arg\,min}_{\varphi,\psi} \int_{\mathcal{X}} \varphi(x)d\mu + \int_{\mathcal{Y}} \psi(y)d\nu$$

with φ and ψ satisfying

$$\varphi(x) + \psi(y) \leq c(x,y).$$

The Kantorovich dual problem has an insightful economics/transportation interpolation, known as the *shipper's problem* [105, Section 1.1.3.], in which the function φ and ψ are interpreted as the loading and unloading prices for the supply and demand. Now we formally introduce the Kantorovich duality

Theorem 1.1 (Kantorovich duality). *For functions $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ and lower*

semi-continuous cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+ \cup \{+\infty\}$, *we have*

$$\inf_{\pi \in \Pi} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y) d\pi(x,y) = \sup_{\varphi, \psi} \int_{\mathcal{X}} \varphi(x) d\mu + \int_{\mathcal{Y}} \psi(y) d\nu.$$

Proposition 1.1 (**Existence**). *Kantorovich transport always admits a minimizer.*

Proof. For detailed proof, see, e.g., [105, Proposition 2.1]. The idea is basically to show that the set Π of admissible couplings is non-empty. $\square$

Knowing the optimizer of Kantorovich transport always exists, we now discuss the condition for a unique optimizer. If the minimizer π is unique, then π must be supported on a graph of a function, i.e., a thin line in Fig. 1.5. We start from the simplest one-dimensional case, i.e., when $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, and define the Monge condition (also known as the Spencer-Mirrlees condition in economics [27, 91]) of the cost function c is as follows

Definition 1.5 (**Monge condition**). *In one dimension, the cost function $c(x,y) : \mathbb{R}^2 \to \mathbb{R}$ satisfies the "Monge" condition [88] if $\forall (x, x', y, y') \in \mathbb{R}^4$, with $x' \geq x$, $y' \geq y$, the cross difference*

$$\Delta(c) := c(x,y) + c(x',y') - c(x,y') - c(x',y)$$

is either quasi-antitone ($\Delta(c) \leq 0$), or quasi-monotone ($\Delta(c) \geq 0$).
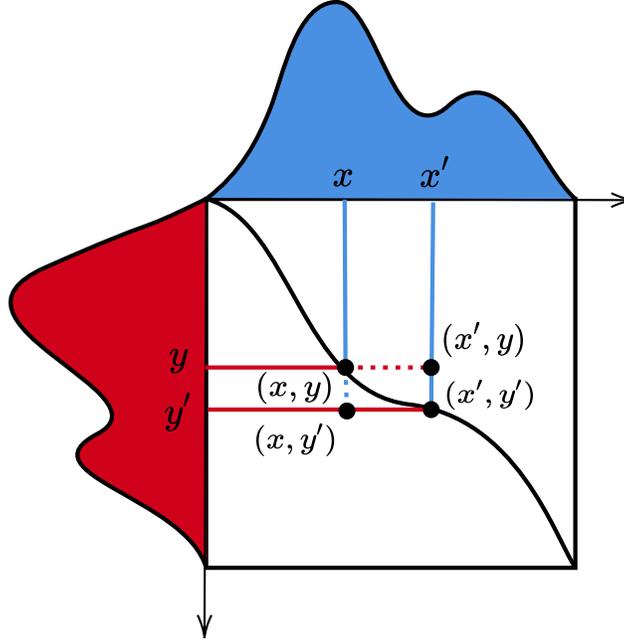
Figure 1.6: For a matching (x', y) and (x, y'), where $x' > x$ and $y' > y$, if the cost c satisfies the Monge condition, there always exist a better pairing, namely, (x, y) and (x', y') having a lower cost.

Let μ and ν be two probability measures on $\mathbb{R}$ with cumulative distribution functions F and G,

$$F(x) = \int_{-\infty}^{x} d\mu(x), \ G(y) = \int_{-\infty}^{y} d\nu(y),$$

and also the inverse of F and G with variable $u \in [0, 1]$,

$$F^{-1}(u) = \inf\{x, F(x) > u\}, \ G^{-1}(u) = \inf\{y, G(y) > u\}$$

The Hoeffding-Fréchet Theorem states that function $H(x, y) : \mathbb{R}^2 \to \mathbb{R}_+$ is in the class

of joint cumulative distribution functions with marginal $F(x)$ and $G(y)$ if and only if[5]

$$F(x) + G(y) - 1 \leq H(x,y) \leq \min\{F(x), G(y)\}. \tag{1.5}$$

Define $H_-(x,y) := F(x) + G(y) - 1$ and $H_+(x,y) := \min\{F(x), G(y)\}$, and thus arrive at the first important result.

Theorem 1.2. *If the cost $c(x,y)$ satisfies the Monge condition, for all the admissible $H(x,y)$ that satisfy* (1.5), *we have*

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)dH_+(x,y) \leq \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)dH(x,y) \leq \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)dH_-(x,y),$$

so that the transportation map π associated with the joint cumulative distribution functions $H_+(x,y) := \min\{F(x), G(y)\}$ is optimal.

The above result can be found from [105, Theorem 2.18] and [88, Theorem 1.4.11 and Chapter 3], and the π associated to H_+ is optimal for convex costs that take the form $c(x-y)$. The key idea is to apply *monotone arrangement* of μ onto ν.

For the higher dimensional case $(\mathcal{X}, \mathcal{Y} = \mathbb{R}^n, \ n > 1)$, an alternative argument for transportation on the graph of a function is based on the cyclic monotonicity as below:

Definition 1.6 (Cyclic monotonicity ([104, Definition 5.1],[105, Definition 2.22])). *A subset $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is a cyclic monotonic set if, for any family of points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ and $N \geq 1$, the inequality*

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

holds with the convection $y_{N+1} = y_1$

---

[5]We use the notation $d\pi(x,y) = dH(x,y)$ and H is non-decreasing,right-continuous in both of the arguments.

From now on, we will restrict ourselves to the *Monge-Kantorovich* problem with the strictly convex cost $c(x, y) = |x - y|^2$. Optimizing with respect to the quadratic cost is equivalent to finding the Wasserstein-2 distance between the measures μ and ν

$$\mathcal{W}_2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X} \times \mathcal{Y}} |x - y|^2 d\pi(x, y)$$

1.2.3 Benamou-Brenier formulation

In the Benamou-Brenier formula of the OMT problem [8], the transportation happens between $\rho_0 dx = d\mu(x)$ and $\rho_{t_f} dy = d\nu(y)$ over the space $\mathcal{M}$. The following dynamical formulation, where the transportation happens in a fixed time interval $[0, t_f]$ so that all the particles leave at time $t = 0$ and arrive at $t = t_f$

$$\begin{aligned}
\mathcal{W}_2(\rho_0, \rho_1) = &\inf_{\rho, v} \int_{\mathcal{M}} \int_0^{t_f} \rho(t, x) \|v(t, x)\|^2 dx dt \\
&\text{s.t.} \quad \frac{\partial}{\partial t}\rho(t, x) + \nabla(\rho(t, x)v(t, x)) = 0, \ \rho(0, x) = \rho_0(x), \ \ \rho(t_f, x) = \rho_{t_f}(x), \quad (1.6) \\
&\rho(t, x) \geq 0, \ \forall \, t \in [0, t_f].
\end{aligned}$$

Consider $\rho(t, x)v(t, x)$ as the momentum $m(t, x) = \rho(t, x)v(t, x)$ of particles, then 1.6 can be rewritten as

$$\begin{aligned}
\mathcal{W}_2(\rho_0, \rho_1) = &\inf_{\rho, v} \int_{\mathcal{M}} \int_0^{t_f} \frac{m(t, x)^2}{\rho(t, x)} dx dt \\
&\text{s.t.} \quad \frac{\partial}{\partial t}\rho(t, x) + \nabla(m(t, x)) = 0, \ \rho(0, x) = \rho_0(x), \ \ \rho(t_f, x) = \rho_{t_f}(x), \quad (1.7) \\
&\rho(t, x) \geq 0, \ \forall \, t \in [0, t_f].
\end{aligned}$$

Noticing the current two formulations are in the Eulerian form, we can convert the problem in the Lagrangian form by introducing the trajectory X_t of the particle labeled by the starting location $x = X_0$, and we have $\partial_t X_t = v(t, X_t(x))$.

The Benamou-Brenier formulation, in its Lagrangian form, now reads

$$\inf_{X_t} \int_{\mathcal{X}} \int_0^{t_f} |\partial X_t(x)|^2 \rho_0(x) dx dt$$

$$\text{s.t. } X_0(x) = x, \ X_{t_f}(x) = y.$$

Moreover, the optimal velocities $\partial_t X_t$ are constant, i.e., $\partial_t X_t = \big(T(x) - x\big)/t_f$, and thus

$$X_t = x + \frac{t}{t_f}(T(x) - x)$$

In general, we always assume $t_f = 1$ without loss of generality. An alternative version of the above property is also characterized by McCann's interpolation (also, displacement interpolation)[71], [105, Section 5.1.3].

1.2.4 Multi-marginal Optimal transport

A newly developed branch is the *multi-marginal* optimal transport (also known as multi-variate optimal transport) [87, 82, 16], where a coupling of more than two marginals. A standard multi-marginal optimal transport problem, with three marginals, can be characterized as follows

Problem 1.6 (Multi-marginal optimal transport). *Given n marginals $\mu_1, \mu_2, \ldots, \mu_n$ and cost $c(x_1, x_2, \ldots, x_n)$, the optimal transport problem minimizes the objective*

$$\iiint_{x_1, x_2, \ldots, x_n} c(x_1, x_2, \ldots, x_n) d\pi(x_1, x_2, \ldots, x_n)$$

over all the admissible couplings $\pi \in \Pi(\mu_1, \mu_2, \ldots, \mu_n)$, where[6]

$$\Pi := \left\{ \pi \in \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \mathcal{X}_n \to \mathbb{R}_+ \ \middle| \ P_{x_i}\big(\pi(x_1, x_2, \ldots, x_n)\big) = \mu_i(x_i), \ i = 1, 2, \ldots n \right\}.$$

---

[6]The operator $P_{x_i}(\pi)$ project the coupling π onto the x_i marginal.

The difficulty of solving the multi-marginal optimal transport problem is twofold: determining the uniqueness of the coupling may be hard for a general cost; numerically solving the problem is expensive (just think about the size of the variable π). Specifically, for the one-dimensional case, i.e., $x_1, x_2, \ldots, x_n \in \mathbb{R}$, the uniqueness can be determined by a *generalized Monge condition*.

Proposition 1.2 (General Monge condition). *To have a unique minimizer, the cost function $c(x_1, x_2, \ldots, x_n)$ needs to satisfy the generalized Monge condition [88, Page 24] which says that any two of the arguments in $c(x_1, x_2, \ldots, x_n)$ have to satisfy the Monge condition.*[7]

Bibliographical notes

Beyond the formulations of the classical two marginal optimal transport problems and the multi-marginal one, more general settings and constrictions have been considered. We list a couple of interesting ideas to spark imagination. Whereas the transportation between marginals with equal mass, [14, 41] consider the partial transportation between marginal with unequal mass. Also, [27, 72, 28] considers the transportation between unequal dimensions. We list some of the nice pedagogic literature for the readers' consideration, for instance, the earlier works by Rachev and Ruüschendorf [88, 89], Villani [104, 105], Santambrogio [91], Carlier [17], Peyré [83]. The optimal transport can surely happen on graphs. For instance, [98] considers continuous flow over graphs, and more general, on discrete surfaces [66]. On the practical side, the idea of optimal transport has a wide application in image processing [80], matching [26].

---

[7]A simple example to check is the cost $c(x_1, x_2, \ldots, x_n) = (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^2 + \cdots + (x_{n-1} - x_n)^2$

Chapter 2

Introduction to optimization

In this chapter, we will go through elementary mathematical optimization in order to refresh the readers with some fundamental ideas. The chapter can be skipped if the readers are familiar with the topic. The outline is as follows: we first formalize the linear and convex programming; secondly, we discuss the l_1 norm and entropic regularization; finally, we derive two popular optimization solvers.

Throughout this chapter, we denote $x \in \mathbb{R}^n$ as the unknown optimizer, and x can be a vector such that $x = [x_1, \ x_2, \ \ldots, \ x_n]$ with a slight abuse of notation. Additionally, assume the function $f(x) : \mathbb{R}^n \to \mathbb{R}$ is the objective function (the function to be minimized) and function $g(x) : \mathbb{R}^n \to \mathbb{R}$ is the subjective function (the constraint to be obeyed).

2.1 Linear and convex programming

Herein, we consider the linear and convex problem. The problem has a general form with the objective function $f(x)$ and m constraints as

$$\min_{x} \quad f(x) \quad \text{s.t.} \quad g_i(x) \leq b, \quad i = 1, \ldots, m.$$

where, for $x \in \mathbb{R}^n$, $f(x) : \mathbb{R}^n \to \mathbb{R}$ and $g_i(x) : \mathbb{R}^n \to \mathbb{R}$, which can satisfy either equality or inequality criteria.

2.1.1 Linear programming

For linear programming (LP), both the objective and subjective functions are linear, and the optimization problem can be written in the explicit form.

Problem 2.1 (Linear programming). *Given a cost vector $c = [c_1, c_2, \ldots, c_n]$, the linear programming has the form*

$$\min \quad c^T x \quad s.t. \quad Ax \leq b$$

Accordingly, by its definition, the maximum flow and the minimum cost flow (shortest path) can be formulated as linear programs. Similarly, the discrete version of the Kantorovich formulation can be solved by linear programming

Problem 2.2 (Discrete Kantorovich problem). *Given the cost matrix c, where the i,j-th element c_{ij} is the cost transporting from x_i to y_j, the discrete optimal transport plan π_{ij} can be obtained by minimizing*

$$\sum_i \sum_j c_{ij} \pi_{ij}$$

26

with respect to the marginal (linear equality constraint) and subject to element-wise inequality

$$\sum_i \pi_{ij} = \nu_j, \ \sum_j \pi_{ij} = \mu_i, \quad and \ \ \pi_{ij} \geq 0.$$

However, the discrete Monge problem can not be formalized as linear programming since the mass is not splittable, and thus the nonzero elements in π must be 1 (the problem minimizes over all the permutation matrix). This problem is known as *integer programming* [108], which is an NP-hard problem [60]. We recommend the readers to [70] for a survey on the early works.

2.1.2 Convex programming

Now we explore linear programming in a more general setting by introducing the connectivity of the function.

Definition 2.1 (Convex function). *A real-valued function $f(x) : \mathbb{R}^n \to \mathbb{R}$ is convex if, for all $x_1, x_2 \in \mathbb{R}^n$ and $\alpha \geq 0$, the condition*

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2),$$

holds.

The convex optimization problem reads

Problem 2.3 (Convex programming). *Given the convex functions $f(x)$, $g(x)$, minimizing $f(x)$ with respect to the $g_i(x) \leq r_i, \quad i = 1, 2, \ \ldots, m$.*

The applications of convex programming can be found everywhere in the field of static, machine learning, and control. We list a few examples of the objectives below:

1. Least-square cost: $\|Ax - b\|_2^2$.

2. 1-norm cost in [101] [1]: $\|x\|_1 = \sum_i |x_i|$.

3. 2-norm cost: $\|x\|_2^2$.

4. Elastic Net cost in [111]: $\|Ax - b\|_2^2 + \lambda_1 \|x\|_2^2 + \lambda_2 |x|_1$.

2.2 Regularization

An extra penalty term is frequently added to the cost as the *regularization term*, which is used later. We revisit two kinds of regularization which are used in the later of our discussion.

2.2.1 l_1-regularization

An interesting property of the optimizer is the *sparsity*, i.e., we would like the optimizer to have as many zeros as possible so that x is sparse. Such a property is extremely important for certain groups of problems. For instance, in feature selection and model reduction [94, 55], one crucial question to answer is "what are the most important features/parameters?". However, seeking the sparest optimizer requires minimizing the 0-norm, which is NP-hard [76] and the problem becomes impractical. Due to the convexity of the l_1 norm, the following problem is proposed as a relaxation for sparse solutions.

Problem 2.4 (l_1-norm minimization). *Given a matrix A and a vector r, the problem reads*

$$\arg\min_x \quad \|x\|_1$$

$$s.t. \quad Ax = b$$

---

[1] Note that the l_1 norm is convex but not differentiable, dealing with such function requires the usage of the *splitting operator* and the sub-differentibility introduced in Chapter 3

The l_1 regularization (also, lasso) is to penalize the constraints with a weight of $1/\lambda$ and the l_1-norm cost, i.e.,

$$\arg\min_x \ \|x\|_1 + \frac{1}{\lambda}\|Ax - b\|_2^2, \quad \lambda > 0.$$

The problem above is known as lasso (least absolute shrinkage and selection operator) in statistics and later gained a lot of attention in machine learning.

2.2.2 Entropic regularization

Compared to convex optimization, solving linear programming comes with a higher price tag, and one may add a regularization term and arrives at a convex relaxation and suitable for proximal algorithms.

Entropic regularization considers an additional convex penalty term, which is the entropy of the coupling $\iint_{\mathcal{X},\mathcal{Y}} \pi(dx, dy)dxdy = 1$ as

$$\mathcal{H}(\pi) = \iint_{\mathcal{X},\mathcal{Y}} -\pi \log(\pi)dxdy,$$

or the discrete version for the stochastic matrix π_{ij}

$$\mathcal{H}(\pi) = \sum_i \sum_j -\pi_{ij}(\log(\pi_{ij}) - 1),$$

see also [31] for more. The entropic regularization of Problem 1.4 has the form

$$\arg\min_\pi \iint_{\mathcal{X},\mathcal{Y}} \left(\frac{1}{\epsilon}c\pi - \mathcal{H}(\pi)\right)dxdy = \iint_{\mathcal{X},\mathcal{Y}} \pi \log(\frac{\pi}{\hat{\pi}})dxdy$$

where $\hat{\pi} = \exp(-c(x,y)/\epsilon)$ is the prior and

$$\mathrm{KL}(\pi\|\hat{\pi}) := \iint_{\mathcal{X},\mathcal{Y}} \pi(dx, dy) \log(\frac{\pi(dx, dy)}{\hat{\pi}(dx, dy)})dxdy$$

29

is the Kullback-Leibler divergence [65] between $\pi(dx, dy)$ and $\hat{\pi}(dx, dy)$. Similarly, for the discrete problem, the optimal coupling can be approximated by solving

$$\arg\min_{\pi \in \Pi} \mathrm{KL}(\pi_{ij} \| \hat{\pi}_{ij}) = \arg\min \sum_{i,j} -\pi_{ij} \log(\frac{\pi_{ij}}{\hat{\pi}_{ij}}). \tag{2.1}$$

over π in the set

$$\Pi(\mu, \nu) = \left\{ \pi \in \mathbb{R}_+^{n \times n} \,\middle|\, \sum_i \pi_{ij} = \nu_j, \;\; \sum_j \pi_{ij} = \mu_i, \;\; \sum_{ij} \pi_{ij} = 1 \right\}.$$

2.3 Optimization methods

Knowing the optimization problem has a unique optimizer, the optimization methods offer a chance to solve/approximate the optimal solution. The framework of optimization is not only for linear/convex programming, but also for more complicated scenarios like non-convex programming. There are numerous methods have been proposed, for example, the gradient descent (GD) method [11, Section 9.3], the stochastic gradient descent (SGD) method [77], the Alternating Direction Method of Multipliers (ADMM) [9], Lagrange multipliers, Augmented Lagrangian Method, etc.

2.3.1 Karush–Kuhn–Tucker (KKT) conditions

The Karush–Kuhn–Tucker (KKT) condition can be traced back to the work [64], which was first discovered by William Karush and rediscovered by the giants in game theory Harold W. Kuhn and Albert W. Tucker. To understand the KKT condition (also, first-order optimality condition), we start with a standard form of convex programming with equality and inequality constraints.

$$\min \quad f(x) \quad \text{s.t.} \quad g_1(x) \le 0, \; g_2(x) = 0.$$

Introducing the Lagrangian multipliers (dual variables) for the two constraints λ_1 and λ_2, the Lagrangian of the problem reads

$$\mathcal{L}(x, \lambda_1, \lambda_2) = f(x) + \lambda_1 g_1(x) + \lambda_2 g_2(x),$$

and thus the corresponding dual problem has the form

$$\max_{\lambda_1, \lambda_2} \ f(x^\star) + \lambda_1 g_1(x^\star) + \lambda_2 g_2(x^\star)$$

where $\lambda_1 \in \mathbb{R}_+$, $\lambda_2 \in \mathbb{R}$, and minimizer $x^\star$ can be obtained according to

$$x^\star = \arg\min \mathcal{L}(x, \lambda_1, \lambda_2).$$

The KKT condition leads to

$$\partial_x f(x) + \lambda_1 \partial_x g_1(x) + \lambda_2 \partial_x g_2(x) \ni 0.$$

Note that the objective function $f(x)$ is not necessarily convex and differentiable. For example, the KKT condition can also be derived for the sub-differentiable function $f(x) = |x|_1$.

Example 2.1. *To see the KKT condition of Problem 2.1, we first derive the Lagrangian, introducing the Lagrangian multipliers u and v for the marginal equality constraints as*

$$\mathcal{L}(\pi, \lambda_1, \lambda_2) = \sum_{i,j} c_{ij} \pi_{ij} + \mathcal{H}(\pi) + u^T(\pi \mathbb{1} - \mu) + v^T(\pi^T \mathbb{1} - \nu),$$

and thus the first-order optimality condition gives

$$\frac{\partial \mathcal{L}(\pi, \lambda_1, \lambda_2)}{\partial \pi_{ij}} = c_{ij} + \epsilon \mathcal{H}(\pi_{ij}) + v_i + u_j = 0.$$

which has a closed-form expression of π, and further leads to the well-known Sinkhorn algo-

rithm by updating u and v with respect to the marginals μ and ν.

The above result leads to the famous Sinkhorn algorithm – a powerful tool for efficient approximate the map π by repetitively updating a prior map given by the cost. The history of such a method can be traced back to the works of Sinkhorn [95, 96], and further developed by [33, 25].

2.3.2 Augmented Lagrangian Method

Without loss of generality, we consider the optimization problem with only the equality constraint, i.e.,

$$\min \quad f(x), \quad \text{s.t.} \quad g(x) = 0.$$

The augmented Lagrangian method, originally discussed in [86, 56], includes a weighted augmentation term, and hence the augmented Lagrangian is

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x) + \frac{r}{2}\|g(x)\|^2$$

where $\rho > 0$ is the weight of the augmentation term. The variables are updated by the method of multipliers

$$x_{k+1} := \arg\min_{x} \mathcal{L}(x_k, \lambda)$$

$$\lambda_{k+1} := \lambda_k + rg(x_{k+1})$$

A novel connection between such a method and optimal transport is addressed by Benamou and Brenier [8] to numerically solve the Monge-Kantorovich mass transfer problem.

2.3.3 Alternating direction method of multipliers

Consider the problem with the objective function $f_1(x) + f_2(x)$, where $f_1(x)$ and $f_2(x)$ are proper convex functions. Its *consensus form* reads

$$\min \quad f_1(x) + f_2(z) \quad \text{s.t.} \quad x = z.$$

The corresponding augmented Lagrangian of the above problem is

$$\mathcal{L}(x, z, \lambda, r) = f_1(x) + f_2(z) + \lambda(x - z) + \frac{r}{2}\|x - z\|_2^2$$

The updates in ADMM can be expressed as

$$x_{k+1} := \arg\min_x \mathcal{L}(x_k, z_k, \lambda_k, r),$$

$$z_{k+1} := \arg\min_x \mathcal{L}(x_{k+1}, z_k, \lambda_k, r),$$

$$\lambda_{k+1} := \lambda_k + r(x_{k+1} - z_{k+1}).$$

The advantage of the ADMM is that the variables x, z split the objectives into two and are updated independently.

Chapter 3

The Shortest path

We shall not cease from exploration,

And the end of all our exploring,

Will be to arrive where we started,

And know the place for the first time.

– T.S. Eliot, from "Little Gidding," Four Quartets, 1943.

The main contribution of this work is to formulate the shortest path problem as ℓ_1-regularized regression, a convex optimization problem [102, 101]. This formulation is, to the best of the authors' knowledge, original.

A second contribution stems from exploring at depth a popular ℓ_1-regularized-regression solver, known as *Least Angle Regression (LARS)*, as applied to the shortest path problem. Specifically, we have shown (Theorem 3.1) that the LARS implementation of our "lasso-shortest-path" formulation replicates a defining feature of the so-called *bi-directional Dijkstra algorithm*, to iteratively build two shortest-path trees, starting from the two specified vertices and until the two trees connect. Through this connection, we present a new perspective of the Dijkstra's algorithm that is completely different than the common presentation as a

greedy algorithm or a dynamic programming viewpoint [97].

Lastly, we explore the *Alternating Direction of Multiplier Method* (ADMM)[9, 11], and a variant (InADMM) [109], for reducing the computational cost in identifying an approximate shortest path for very large graphs. A useful feature of ADMM is that it admits distributed implementation, initialized with any suitable path, if one is available. This feature, which speeds up convergence, is especially useful when a short path needs to be updated following topological changes in the graph. A comparison of the computational cost of these implementations to those of Dijkstra's algorithm is given in Table 3.2.

The ADMM algorithm proposed here is completely different than earlier proposals on the subject, e.g., the self-stabilizing approach in [19] and the consensus-based approach in [110]. The proposed algorithm aims at identifying an (approximate) shortest path between specified vertices, allowing for better computational complexity and relatively efficient updates to topological changes.

The outline of the work is as follows. Section 3.1 introduces notation along with basic concepts and a brief account of Dijkstra's algorithm. Section 3.2 casts the search for short paths in a network as a convex optimization problem as discussed. Section 3.3 details the algorithmic steps for updating state-values on edges, that turn out to coincide with the so-called *lasso solution* in the LARS algorithm. Section 3.4 highlights the commonality of features between the LARS algorithm and bi-directional Dijkstra algorithm. Finally, Section 3.5 explores the application of the ADMM method to our lasso formulation, and highlights its relevance in identifying short, but not necessarily shortest, paths in very large graphs.

3.1 Preliminaries

3.1.1 Graph theoretic notations and definitions

Throughout we consider a weighted *undirected* graph $\mathcal{G}$ that is connected and has no self-loops or multi-edges. We write $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V} = \{v_1, \ldots, v_n\}$ is the set of vertices/nodes, $\mathcal{E} = \{e_1, \ldots, e_m\}$ is the set of edges, and $\mathcal{W} = \{w_1, \ldots, w_m\}$ a set of weights corresponding to the edges.

We will consistently use $n = |\mathcal{V}|$ and $m = |\mathcal{E}|$ for the cardinality of these two sets. When labeling edges, we also use the notation (v_i, v_j) for the edge that connects vertices v_i and v_j, without significance to the order.

However, as is common, in defining the *incidence matrix* of the graph, denoted by $D(\mathcal{G})$ an arbitrary but fixed orientation is assigned to edges that has no bearing on the results. To this end, the incidence matrix is defined as the $n \times m$ matrix with (i, j)th entry

$$
[D]_{ij} = \begin{cases} +1 & \text{if the } i\text{th vertex is the tail of edge } e_j, \\ -1 & \text{if the } i\text{th vertex is the head of edge } e_j, \\ 0 & \text{otherwise.} \end{cases}
$$

It is convenient to define the *weight matrix* $W = \mathrm{diag}(w_1, \ldots, w_m)$ as the diagonal matrix formed by the weights in $\mathcal{W}$, consistent with the ordering in $\mathcal{E}$.

A path from vertex v_s to vertex v_t is a sequence of connected edges

$$
p = \{(v_{i_0}, v_{i_1}), (v_{i_1}, v_{i_2}), \ldots, (v_{i_{l-1}}, v_{i_\ell})\}
$$

that "starts" at $v_{i_0} = v_s$ and "terminates" at $v_{i_l} = v_t$. An alternative representation of the path, which now encodes edge-orientation that is consistent with that in specifying

D, is in terms of the *incidence vector* $x^{(p)}$. This is an m-dimensional vector defined as follows: the ith entry $(x^{(p)})_i$ is $+1$, or -1, depending on whether an edge $(v_{i_{k-1}}, v_{i_k})$ (for some $k \in \{1, \dots, \ell\}$) in the path is the ith edge in $\mathcal{E}$ and is listed with orientation consistent or not with the tail/head designation in specifying D, respectively; if the ith edge is not in the path, $(x^{(p)})_i = 0$. The *length of the path* is defined as the sum of edge-weights, i.e.,

$$\text{length}(p) \triangleq \sum_{e_i \in p} w_i = \|W x^{(p)}\|_1,$$

where $\| \cdot \|_1$ denotes the ℓ_1-norm.

A graph is said to be a *tree* if it has no cycle, i.e., it has no path where $s = t$. If $\mathcal{G}$ is connected, there is always a subgraph which is a tree. When $\mathcal{G}$ is a tree, $m = n - 1$ and there is a unique path from any given vertex to any other. Any vertex can be designated as *root*, and the structure of graph encapsulated by all paths connecting vertices to the root $(n - 1$ paths). The $(n - 1) \times (n - 1)$ matrix of incidence vectors of all such paths is referred to as *path matrix* (often with reference to the root) and denoted by P_{v_1}, or simply P, when the root is clear from the context. Interestingly, P is closely connected to the incidence matrix D. This is the content of the following lemma which is key for results in Section 3.4 but also of independent interest.

Lemma 3.1. *Let $\mathcal{G}$ be a tree rooted at v_1 with n vertices and P its path matrix. The pseudoinverse of its incidence matrix D, denoted as D^+, is[1]*

$$D^+ = \left[-\tfrac{1}{n} P \mathbb{1}_{n-1}, \quad PJ \right],$$

where

$$J = \left(I_{n-1} - \frac{1}{n} \mathbb{1}_{n-1} \mathbb{1}_{n-1}^T \right)$$

---

[1] Throughout, $\mathbb{1}_k$ denotes the k-column vector with entries equal to 1, and I_k the $k \times k$ identity matrix.

Proof. See [5, Theorem 2.10 & Lemma 2.15]. □

3.1.2 Shortest path problem and Dijkstra's algorithm

Let $\mathcal{P}_{s,t}$ denote the set of all paths between v_s and v_t. This set is non-empty because the graph is connected. The shortest path problem is to find a path with minimum length over all the paths between v_s and v_t, i.e.,

$$\arg \min_{p \in \mathcal{P}_{s,t}} \text{length}(p). \tag{3.1}$$

The minimum value is known as the *distance* between v_s and v_t. A well-known search algorithm - Dijkstra's algorithm, has been proposed for this problem.

Dijkstra's algorithm [35] begins with the "starting" vertex v_s, and initially assigns a distance of 0 to v_s and $+\infty$ to all other vertices. It iteratively labels the vertex with the lowest *distance estimate* as *visited* and updates the distance estimates of its neighbors that have not yet been visited (*unvisited*). The distance estimates are updated by summing up the distances of visited vertices and the weights of the edges linking these vertices to their unvisited neighbors. Dijkstra's algorithm terminates when v_t is visited and produces the shortest path from source vertex v_s to others vertices in the form of shortest-path tree.

The essential feature of Dijkstra's algorithm is that it iteratively constructs the shortest-path tree rooted at v_s to all the visited vertices before reaching the target v_t. Similarly, bi-directional Dijkstra algorithm constructs two shortest-path trees rooted at v_s and v_t and terminates when the two trees connect.

Later on in Section 3.4, we will point out analogies between the bi-directional Dijkstra algorithm and properties of the LARS algorithm applied to the shortest-path problem advocated herein.

The main contribution in this work is to point out that short paths in graphs, sought via formulation as a lasso problem, has computational and implementation advantages. Specifically, the use of ADMM algorithms for large graphs reduces computational complexity and allows for distributed implementation.

Extension of the framework to one that can cope with negative weights is desirable, but at present, not available.

3.2   Problem formulation

We now cast the shortest path problem (3.1) as a linear program. To this end we will use a well-known technique for finding sparse solutions to linear equations by minimizing the ℓ_1 norm of a vector as a surrogate for the count of its non-vanishing entries [101].

In our setting, constraints are expressed in terms of D (incidence, also constraint matrix) and x, the incidence vector of a sought path p from v_s to v_t, in that,

$$p \in \mathcal{P}_{s,t} \quad \Rightarrow \quad Dx^{(p)} = y^{(s,t)}. \tag{3.2}$$

Here, $y_i^{(s,t)} = \mathbf{1}_{\{s\}} - \mathbf{1}_{\{t\}}$ is the *indicator vector* in $\mathbb{R}^n$ of a *virtual edge* directly connecting v_s to v_t; the path together with the virtual edge form a *closed cycle*. Throughout, $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function of set $\{\cdot\}$.

Remark 3.1. *An alternative justification can be provided by noting that closed cycles, i.e., paths that begin and end at the same node, form a basis for the null space of the incidence matrix of the graph [5, 73]. If we attach a* virtual direct link (i.e., a new edge) *between vertices s and t, we need to update the incidence matrix to* $\begin{bmatrix} D & -y^{(s,t)} \end{bmatrix}$ *so that this virtual edge is included. Now a path from s to t "closes" into a cycle by including this extra virtual edge. Any cycle that includes the virtual edge corresponds to a null vector of* $\begin{bmatrix} D & -y^{(s,t)} \end{bmatrix}$

with a 1 as the last entry (indicating that the virtual edge is included), and therefore, to a
solution of

$$\begin{bmatrix} D & -y^{(s,t)} \end{bmatrix} \begin{bmatrix} x^{(p)} \\ 1 \end{bmatrix} = 0.$$

This is precisely (3.2), while the first component $x^{(p)}$ of the solution vector corresponds to a
sought path from s to t.

Note that linear combinations $x = ax^{(p_1)} + (1-a)x^{(p_2)}$ with $a \in (0,1)$ of incidence vectors
of two distinct paths p_1 and p_2 between v_s and v_t, also satisfy the constraint $Dx = y^{(s,t)}$.
That is, although an exact correspondence between the two sides of (3.2) does not hold,
it does hold between the *shortest path* and a corresponding integer vertex of the polytope
defined by the constraint matrix D). Thus, we *propose*

$$\arg\min_{x \in \mathbb{R}^m} \|Wx\|_1, \quad \text{s.t.} \quad Dx = y^{(s,t)} \tag{3.3a}$$

as a way to solve the shortest path problem.

To gain insight as to the nature of the minimizer, problem (3.3a) can be recast as the
linear program:

$$\arg\min_{\xi \geq 0} \sum_{i=1}^{2m} \xi_i w_{i \bmod(m)}, \quad \text{s.t.} \quad \mathcal{D}\xi = y^{(s,t)} \tag{3.3b}$$

with $\mathcal{D} := [D, -D]$ and $\xi \in \mathbb{R}^{2m}$. The solutions to (3.3a) and (3.3b) correspond via

$$\xi = \begin{bmatrix} x'_+, & x'_- \end{bmatrix}',$$

where x_+ (x_-, resp.) is the vector of positive (negative, resp.) entries of x, setting zero for
the negative (positive, resp.) entries, i.e., in Matlab notation, $x_+ = x(x \geq 0)$, $x_- = x(x \leq 0)$.
The constraint matrix $\mathcal{D}$ in (3.3b) is *totally unimodular* (i.e., all minors have determinant

in $\{0, \pm 1\})$ [5, Lemma 2.6], and therefore, application of [1, Theorem 11.11 (Unimodularity Theorem)] shows that, provided the shortest path is unique, the solution ξ (and, hence, x) is integer-valued, actually, $\{0, \pm 1\}$-valued. Thus, x in (3.3a) corresponds to a valid incidence vector.

3.2.1 Lasso formulation

Returning to (3.3a), rewritten in the form

$$\arg\min_{\beta}\{\|\beta\|_1 \mid \|Q\beta - y\|_2^2 = 0\} \tag{3.3a$'$}$$

in new variables $\beta = Wx$, $y = y^{(s,t)}$, and $Q = DW^{-1}$, leads us a relaxation as the ℓ_1-regularized regression

$$\boldsymbol{\beta}(\lambda) := \arg\min_{\beta \in \mathbb{R}^m} \frac{1}{2}\|y - Q\beta\|_2^2 + \lambda\|\beta\|_1, \tag{3.4}$$

with (regularization parameter) $\lambda > 0$. The formulation (3.4) is known as *lasso* [102].

The limit $\boldsymbol{\beta}_0 := \lim_{\lambda \to 0} \boldsymbol{\beta}(\lambda)$ from (3.4), for $\lambda > 0$, provides the indicator vector $\boldsymbol{x}_0 = W^{-1}\boldsymbol{\beta}_0$ of a path. For $\lambda > 0$, $\boldsymbol{\beta}(\lambda)$ may not correspond to a path. However, due to continuity and the fact that $\boldsymbol{x}_0$ in the limit must be $\{0, \pm 1\}$-valued, for sufficiently small λ, $\boldsymbol{x}(\lambda) = W^{-1}\boldsymbol{\beta}(\lambda)$ reveals the shortest path (e.g., by rounding the values to the nearest integer).

Thus, the lasso formulation represents a viable and attractive approach for solving the shortest path problem. Interestingly, as we show in Section 3.4, the LARS algorithm – a popular solver for lasso (3.4), shares features of the bi-directional Dijkstra algorithm. Most importantly, the lasso formulation (3.4), as discussed in Section 3.5, allows the use of proximal optimization methods for obtaining satisfactory approximations of the shortest path in large graph settings.

3.2.2 Uniqueness of the lasso solution

A sufficient condition for uniqueness of solution to (3.4) is that $\text{rank}(Q) = m$, the size of β and number of edges [102, Lemma 2]. However, recall that $WQ = D$, the incidence matrix. It follows that $\text{rank}(Q) = m$ only holds when the graph is a tree (or possibly, a disjoined set of trees, cf. [5, Theorem 2.3]).

Evidently, such an assumption is too restrictive, also since the shortest path problem in this case becomes trivial.

Herein we introduce a fairly general sufficient condition for the uniqueness of solution to (3.4) (as we claim next), that is in fact generic, for generic weights.

Assumption 3.1. *The shortest path between vertex v_s and any other vertex is unique, and the same applies to v_t.*

We note that $\boldsymbol{\beta}(\lambda)$ from (3.4) turns out to be piece-wise linear (see Section 3.3). The values of λ where the slope changes are referred to as *breakpoints*. With this in place, the implications of the assumption to our problem can be stated as follows.

Lemma 3.2 (Uniqueness). *Under Assumption 3.1, Problem (3.4) admits a unique solution for all $\lambda > 0$.*

It is important to note that the quadratic expression in (3.4) is not strictly convex, since Q may have a nontrivial null space. The key idea in proving uniqueness under the conditions of the lemma requires a discussion of the LARS algorithm that is explained next. Hence the proof is deferred to Appendix B.4.

3.3 The LARS algorithm

3.3.1 Karush-Kuhn-Tucker (KKT) conditions

The solution $\boldsymbol{\beta}(\lambda)$ of (3.4) must satisfy the *KKT condition* [102, Section 2.1]

$$Q^T(y - Q\beta(\lambda)) = \lambda\gamma, \tag{3.5}$$

where the vector γ is in the sub-differential of $\|\beta(\lambda)\|_1$, with jth component given by

$$\gamma_j \in \begin{cases} \{\text{sign}(\beta_j(\lambda))\} & \text{if } \beta_j(\lambda) \neq 0, \\ [-1, 1] & \text{if } \beta_j(\lambda) = 0. \end{cases}$$

The KKT condition in (3.5) motivates us to divide the indices $\{1, 2, \ldots, m\}$ into two set: an *active* set (also, *equicorrelation* set [102]) $\mathcal{A}$ and, a *non-active* set $\mathcal{A}^c$, i.e.,

$$\begin{aligned} \mathcal{A} &\triangleq \{j \mid \boldsymbol{\beta}_j(\lambda) \neq 0\}, \\ \mathcal{A}^c &\triangleq \{j \mid \boldsymbol{\beta}_j(\lambda) = 0\}. \end{aligned} \tag{3.6}$$

Let $\boldsymbol{\beta}_{\mathcal{A}}(\lambda)$ denote the vector $\boldsymbol{\beta}(\lambda)$ with the non-active (or equivalently, zero) entries removed and, likewise, $Q_{\mathcal{A}}$ be the matrix Q with the columns corresponding to the non-active set removed. Then, the KKT condition (3.5) can be expressed as

$$Q_j^T(y - Q_{\mathcal{A}}\boldsymbol{\beta}_{\mathcal{A}}(\lambda)) = s_j\lambda, \quad \forall j \in \mathcal{A}, \tag{3.7a}$$

$$|Q_j^T(y - Q_{\mathcal{A}}\boldsymbol{\beta}_{\mathcal{A}}(\lambda))| \leq \lambda, \quad \forall j \in \mathcal{A}^c, \tag{3.7b}$$

where Q_j denotes the jth column of Q and the *sign vector*

$$s := \text{sign}\left(Q_{\mathcal{A}}^T(y - Q_{\mathcal{A}}\boldsymbol{\beta}_{\mathcal{A}}(\lambda))\right) = \text{sign}\left(\boldsymbol{\beta}_{\mathcal{A}}\right), \tag{3.8}$$

with jth element the sign (± 1) of the jth entry of $\boldsymbol{\beta}_{\mathcal{A}}$.

3.3.2   The LARS algorithm

The LARS algorithm, as formulated in [102, Section 3.1] to solve lasso, finds the *solution path* of $\boldsymbol{\beta}(\lambda)$ that meets the KKT condition (3.5) for all $\lambda > 0$. The algorithm is initialized with $\lambda_0 = \infty$, $\mathcal{A}_0 = \emptyset$, and $s_0 = \emptyset$. The solution path $\boldsymbol{\beta}_{\mathcal{A}}(\lambda)$, in (3.9) below, is computed for decreasing λ, and is piece-wise linear and continuous with breakpoints $\lambda_0 > \lambda_1 > \ldots > 0$. Breakpoints are successively computed at each iteration of the algorithm, and the linear segment of $\boldsymbol{\beta}_j(\lambda)$ is determined to satisfy the element-wise KKT condition as detailed in (3.7). As λ crosses breakpoints, the active (non-active) set (3.6) and the sign vector (3.8) are updated accordingly.

We now detail the kth iteration of the LARS algorithm, initializing $\lambda = \lambda_k$, $\mathcal{A} = \mathcal{A}_k$, $s = s_k$, and seeking the next breakpoint λ_{k+1}. The lasso variable $\boldsymbol{\beta}_{\mathcal{A}_k}(\lambda)$, as a function of λ, is calculated as the minimum ℓ_2-norm solution of (3.7a), and is given by:

$$\boldsymbol{\beta}_{\mathcal{A}_k}(\lambda) = (Q_{\mathcal{A}_k}^T Q_{\mathcal{A}_k})^+ (Q_{\mathcal{A}_k}^T y - \lambda s_k) = a^{(k)} - b^{(k)} \lambda, \tag{3.9}$$

where $(\cdot)^+$ denotes the pseudoinverse and

$$\begin{aligned} a^{(k)} &:= (Q_{\mathcal{A}_k}^T Q_{\mathcal{A}_k})^+ Q_{\mathcal{A}_k}^T y, \\ b^{(k)} &:= (Q_{\mathcal{A}_k}^T Q_{\mathcal{A}_k})^+ s_k. \end{aligned} \tag{3.10}$$

The next breakpoint λ_{k+1} is determined as the largest value at which the KKT condition

$$\left(Q_j^T (y - Q_{\mathcal{A}_k} \beta_{\mathcal{A}_k}(\lambda)) = \lambda \gamma_j \right)$$

is violated by λ. Such violation occurs in two circumstances ("crossing/joining" in the

language of [102]), either (3.7a) or (3.7b) fails.

i) The "joining" case is when condition (3.7b) is violated for some $j \in \mathcal{A}_k^c$, i.e., $|Q_j^T(y - Q_{\mathcal{A}_k}\boldsymbol{\beta}_{\mathcal{A}_k}(\lambda))| \leq \lambda$ no longer holds. For each index $j \in \mathcal{A}_k^c$, this happens at $\lambda = t_j^{\text{join}}$ given by

$$t_{j,k}^{\text{join}} = \frac{Q_j^T(Q_{\mathcal{A}_k}a^{(k)} - y)}{Q_j^T Q_{\mathcal{A}_k}b^{(k)} \pm 1}, \tag{3.11}$$

where the choice $\pm$ is the one for which $t_{j,k}^{\text{join}} \in [0, \lambda_k]$. We set "joining time" $\lambda_{k+1}^{\text{join}} := \max_{j \in \mathcal{A}_k^c}\{t_{j,k}^{\text{join}}\}$.

ii) The "crossing" case is when condition (3.7a) is violated for some $j \in \mathcal{A}_k$ so that one of the element of $\boldsymbol{\beta}_{\mathcal{A}_k}(\lambda)$ crosses zero (changes its sign), i.e. $a_j^{(k)} - \lambda b_j^{(k)} = 0$ for some $\lambda < \lambda_k$. For each index $j \in \mathcal{A}_k$, the crossing happens at $\lambda = t_j^{\text{cross}}$ given by

$$t_{j,k}^{\text{cross}} = (a_j^{(k)}/b_j^{(k)}) \cdot \mathbf{1}_{\{0 < (a_j^{(k)}/b_j^{(k)}) < \lambda_k\}}. \tag{3.12}$$

We set "crossing time" $\lambda_{k+1}^{\text{cross}} = \max_{j \in \mathcal{A}_k}\{t_{j,k}^{\text{cross}}\}$.

The next breakpoint is

$$\lambda_{k+1} = \max\{\lambda_{k+1}^{\text{join}}, \lambda_{k+1}^{\text{cross}}\}. \tag{3.13}$$

If the joining occurs, the joining index is added to the active set and the sign vector is updated. If a crossing happens, the crossing index is removed from the active set. The overall algorithm is summarized as blow.

3.4   Theoretical result

Both the LARS as well as the bi-directional Dijkstra algorithm iteratively construct shortest-path trees with roots at v_s and v_t, and terminate when the two trees meet. We prove our main result below by induction. The induction hypothesis is specified next.

Assumption 3.2 (Induction hypothesis). *At $(k-1)th$ iteration of LARS, the edges in the*

Algorithm 3 LARS algorithm for the lasso

Input: matrix $Q = DW^{-1}$, vector $y = y^{(s,t)}$.

Output: incidence vector $\boldsymbol{x}_0 = W^{-1}\boldsymbol{\beta}_0$, distance $\|\boldsymbol{\beta}_0\|_1$.

1: $k = 0$, $\lambda_0 = \infty$, $\mathcal{A} = \emptyset$, $s = 0$, $a^{(0)} = 0$ and $b^{(0)} = 0$.

2: **while** $\lambda_k > 0$ **do**

3:     Compute the joining time $\lambda_{k+1}^{\text{join}}$ (3.11) for $j \in \mathcal{A}_k^c$.

4:     Compute the crossing time $\lambda_{k+1}^{\text{cross}}$ (3.12) for $j \in \mathcal{A}_k$.

5:     Compute λ_{k+1} according to (3.13) and

      i)   if join happens, i.e., $\lambda_{k+1} = \lambda_{k+1}^{\text{join}}$, add the joining index j to $\mathcal{A}_k$ and its sign to s_k;

      ii)   if cross happens, i.e., $\lambda_{k+1} = \lambda_{k+1}^{\text{cross}}$, remove the crossing index j from $\mathcal{A}_k$ and its sign from s_k.

6:     $k = k + 1$.

7:     Compute $a^{(k)}$ and $b^{(k)}$ according to (3.10)

8:     Set $\boldsymbol{\beta}_{\mathcal{A}_k} = a^{(k)} - \lambda_k b^{(k)}$ and $\boldsymbol{\beta}_{\mathcal{A}_k^c} = 0$.

9: **end while**

---

active set $\mathcal{A}_{k-1} = \mathcal{A}_{k-1}^{(s)} \cup \mathcal{A}_{k-1}^{(t)}$, where $\mathcal{A}_{k-1}^{(s)}$ and $\mathcal{A}_{k-1}^{(t)}$ are disjoint subsets of edges forming trees on vertices $T_{k-1}^{(s)}, T_{k-1}^{(t)} \subset \mathcal{V}$, rooted at v_s and v_t, respectively. The two trees are the shortest-path trees from the roots. Moreover, crossing does not occur at this iteration, i.e., no edges are removed from the active set.

Our induction starts with the *base case $k = 1$*, where the active set is empty and both trees consist of a single root vertex, $T_0^{(s)} = \{s\}$ and $T_0^{(t)} = \{t\}$. The crossing does not occur since the active set is empty. The proof is based on the following two propositions that provide simplified expressions for the joining and crossing times, derived from Lemma 3.1 and Assumption 3.2 that the graph is a tree. The proofs are given in Appendices B.2 and B.3.

3.4.1 Joining and crossing times

Proposition 3.1 (Joining time). *For the edge $e_j = (v_1, v_2)$ where $e_j \in \mathcal{A}_k^c$, the element-wise joining time is*

$$
t_{j,k}^{join} = \begin{cases}
0 & \text{if } (v_1, v_2) \in \Omega_k^2 \cup T_k^{(s)^2} \cup T_k^{(t)^2} \\[2ex]
\left(|T_k^{(s)}| l_{v_2}^{(s)} - \displaystyle\sum_{v \in T_k^{(s)}} l_v^{(s)} \right)^{-1} & \text{if } (v_1, v_2) \in T_k^{(s)} \times \Omega_k \\[2ex]
\left(|T_k^{(t)}| l_{v_2}^{(t)} - \displaystyle\sum_{v \in T_k^{(t)}} l_v^{(t)} \right)^{-1} & \text{if } (v_1, v_2) \in T_k^{(t)} \times \Omega_k \\[2ex]
\left(|T_k^{(s)}| + |T_k^{(t)}| \right) / \gamma & \text{if } (v_1, v_2) \in T_k^{(s)} \times T_k^{(t)}
\end{cases}
$$

where $\Omega_k = \mathcal{V} \setminus (T_k^{(s)} \cup T_k^{(t)})$, $l_v^{(s)}$ and $l_v^{(t)}$ denote the distance of vertex v to the root s and t respectively, and

$$
\gamma = |T_k^{(s)}||T_k^{(t)}| l_t^{(s)} - |T_k^{(t)}| \sum_{v \in T_k^{(s)}} l_v^{(s)} - |T_k^{(s)}| \sum_{v \in T_k^{(t)}} l_v^{(t)}.
$$

Proposition 3.2 (Crossing time). *For an edge $e_j = (v_1, v_2)$ where $e_j \in \mathcal{A}_k$, the expression $a_j^{(k)} / b_j^{(k)}$ that appears in the definition of crossing time (3.12) is*

$$
\frac{a_j^{(k)}}{b_j^{(k)}} = \begin{cases}
\left((|T_k^{(s)}|/|R_j^{(s)}|) \displaystyle\sum_{v \in R_j^{(s)}} l_v^{(s)} - \sum_{v \in T_k^{(s)}} l_v^{(s)} \right)^{-1} & \text{if } (v_1, v_2) \in T_k^{(s)^2} \\[2ex]
\left((|T_k^{(t)}|/|R_j^{(t)}|) \displaystyle\sum_{v \in R_j^{(t)}} l_v^{(t)} - \sum_{v \in T_k^{(t)}} l_v^{(t)} \right)^{-1} & \text{if } (v_1, v_2) \in T_k^{(t)^2}
\end{cases}
$$

where $R_j^{(s)}$ and $R_j^{(t)}$ are the subsets of vertices in the tree $T_k^{(s)}$ and $T_k^{(t)}$ respectively, whose path to the root contains the edge e_j.

3.4.2 Toward the equavalence

Assuming the induction hypothesis in Assumption 3.2 holds, we show that the hypothesis also holds at iteration k through the following lemmas.

Lemma 3.3 (Edge adding). *At iteration k, either the edge connecting $v_{min}^{(s)}$ to tree $T_k^{(s)}$, or the edge $v_{min}^{(t)}$ to tree $T_k^{(t)}$ will be added to the active set, where $v_{min}^{(s)}$ and $v_{min}^{(t)}$ are vertices with minimum distance to roots v_s and v_t among all other vertices outside the two trees, respectively.*

Proof. First, assume no edge connects the two trees, i.e., the last case of joining time does not happen (the case when it does is studied in Lemma 3.5). The joining time $\lambda_{j,k}^{\text{join}}$ now reads

$$\max\left\{ \left(|T_k^{(s)}|l_{v_{\min}^{(s)}} - \sum_{v \in T_k^{(s)}} l_v^{(s)}\right)^{-1}, \left(|T_k^{(t)}|l_{v_{\min}^{(t)}} - \sum_{v \in T_k^{(s)}} l_v^{(s)}\right)^{-1}\right\},$$

where the first expression is achieved by the edge that connects $v_{\min}^{(s)}$ to tree $T_k^{(s)}$ and the second is achieved by the edge that connects $v_{\min}^{(t)}$ to tree $T_k^{(t)}$. Hence, one of these two edges is joined to the active set, if crossing does not occur. $\square$

Remark 3.2 (No cycles). *Cycles may be created in the following two scenarios: (i) An edge that connects two vertices of a tree is joined; (ii) Two edges that connect the tree to a single vertex, say v, are joined simultaneously. Scenario (i) can not happen because $t_j^{join} = 0$ for such edges (first case of joining time). Scenario (ii) can not happen, because in order for two edges to join simultaneously, we must have two distinct shortest paths from v to the root, which is not possible according to Assumption 3.1.*

Lemma 3.4 (No edge removed). *At iteration k and if $\lambda > 0$, crossing does not take place.*

Proof. To prove that crossing does not take place before the algorithm terminates, we show that $a_j^{(k)}/b_j^{(k)} \geq \lambda_{k-1}$ and hence $a_j^{(k)}/b_j^{(k)} \geq \lambda_k$ for all e_j in the active set, so that crossing

time is zero according to its definition (3.12).

First, we obtain an expression for λ_k and then compare it to crossing times. The value λ_k is determined by the maximum of joining time and crossing time at $(k-1)$th iteration according to (3.13). Under Assumption 3.2, the breakpoint λ_k is the maximum joining time, which takes two possible values, corresponding to the edge that connects to either tree $T_{k-1}^{(s)}$ or tree $T_{k-1}^{(t)}$ as proved in Lemma 3.3.

Without loss of generality, we now assume the joining happens to the tree $T_{k-1}^{(s)}$. Then,

$$\lambda_{k-1} = \left(|T_{k-1}^{(s)}|l_{v_{\min}^{(s)}} - \sum_{v \in T_{k-1}^{(s)}} l_v^{(s)}\right)^{-1}. \tag{3.14}$$

Next, we show $a_j^{(k)}/b_j^{(k)} \geq \lambda_{k-1}$ for all e_j that belong to the tree $T_k^{(s)}$. From Proposition 3.2, we can write $(a_j^{(k)}/b_j^{(k)})$ as

$$\frac{a_j^{(k)}}{b_j^{(k)}} = \left(\frac{1 + |T_{k-1}^{(s)}|}{|R_j^{(s)}|} \sum_{v \in R_j^{(s)}} l_v^{(s)} - l_{v_{\min}^{(s)}} - \sum_{v \in T_{k-1}^{(s)}} l_v^{(s)}\right)^{-1}$$

$$\geq \left(|T_{k-1}^{(s)}|l_{v_{\min}^{(s)}} - \sum_{v \in T_{k-1}^{(s)}} l_v^{(s)}\right)^{-1},$$

where we used $|T_k^{(s)}| = |T_{k-1}^{(s)}| + 1$, $\sum_{v \in T_k^{(s)}} l_v^{(s)} = l_{v_{\min}^{(s)}} + \sum_{v \in T_{k-1}^{(s)}} l_v^{(s)}$, and $l_{v_{\min}^{(s)}} \geq l_v$ for all $v \in T_k^{(s)}$. The inequality above holds because $v_{\min}^{(s)}$ is the latest vertex that is added to the tree and other vertices that have been already added have a shorter distance to the root.

The proof of $a_j^{(k)}/b_j^{(k)} \geq \lambda_{k-1}$ for all e_j that belong to the other tree $T_k^{(t)}$ is conceptually similar. One needs to compare $a_j^{(k)}/b_j^{(k)}$ with the joining time of the last edge that has been added to the tree $T_k^{(t)}$ at a certain past iteration, say $k' < k$, and use the fact that $\lambda_k < \lambda_{k'}$. The details are omitted on account of space. □

Lemma 3.5 (Termination Criteria). *The LARS algorithm terminates when the two trees*

connect.

Proof. Assume the two trees $T_k^{(s)}$ and $T_k^{(t)}$ become connected at iteration k. This happens when the last expression of joining achieves the maximum joining time, hence $\lambda_k = \left(|T_k^{(s)}| + |T_k^{(t)}|\right)/\gamma$, as depicted in Fig 3.1. The objective is to show that the algorithm terminates after this, i.e. $\lambda_{k+1} = 0$. We show this by proving the joining time and crossing time are both zero.

The derivation of element-wise joining time in Proposition 3.1 reveals that $t_{j,k+1}^{join} = 0$, $\forall e_j \in \mathcal{A}_{k+1}^c$. For the crossing time, the derivation of Proposition 3.2 yields that for all $e_j \in \mathcal{A}_{k+1}$,

$$
\frac{a_j^{(k+1)}}{b_j^{(k+1)}} = \begin{cases} 0, \text{ if } e_j \notin p_{s,t} \\ \left(\displaystyle\sum_{v \in R_j} l_v^{(s)} - \frac{|R_j|}{|T_k^{(t)}| + |T_k^{(s)}|} \sum_{v \in T_k^{(s)} \cup T_k^{(t)}} l_v^{(s)} \right)^{-1}, \end{cases}
$$

where $p_{s,t} \subset \mathcal{A}_{k+1}$ is the path from v_s to v_t. Therefore, it remains to show that the crossing time for the edges in $\mathcal{A}_{k+1} \cap p_{s,t}$ are zero. We show this by proving $a_j^{(k+1)}/b_j^{(k+1)} \geq \lambda_k$ for all edges $e_j \in \mathcal{A}_{k+1} \cap p_{s,t}$. Considering the edges that belong to the tree $\mathcal{A}_k^{(s)}$, we have

$$
\frac{|T_k^{(s)}| + |T_k^{(t)}|}{(a_j^{(k+1)}/b_j^{(k+1)})} = \gamma + |T_k^{(s)}| \sum_{v \in R_j^{(s)}} l_v^{(s)} - |R_j^{(s)}| \sum_{v \in T_k^{(s)}} l_v^{(s)}
$$
$$
- |T_k^{(t)}| \sum_{v \in R_j^{(s)}} l_v^{(t)} + |R_j^{(s)}| \sum_{v \in T_k^{(t)}} l_v^{(t)}
$$

where $R_j^{(s)}$ are the vertices in the tree $T_k^{(s)}$ such that their path to the root v_s contains e_j. Because $l_v^{(s)} \leq l_{v_2}^{(s)}$ and $l_v^{(t)} \geq l_{v_2}^{(t)}$ for all $v \in R_j^{(s)}$, we have the inequality

$$
\frac{|T_k^{(s)}| + |T_k^{(t)}|}{(a_j^{(k+1)}/b_j^{(k+1)})} - \gamma \leq |R_j|\left(|T_k^{(s)}|l_{v_2}^{(s)} - \sum_{v \in T_k^{(s)}} l_v^{(s)} - |T_k^{(t)}|l_{v_2}^{(t)} + \sum_{v \in T_k^{(t)}} l_v^{(t)}\right).
$$

50

We claim that the expression in parentheses is negative

$$|T_k^{(s)}|l_{v_2}^{(s)} - \sum_{v \in T_k^{(s)}} l_v^{(s)} - |T_k^{(t)}|l_{v_2}^{(t)} + \sum_{v \in T_k^{(t)}} l_v^{(t)} \leq 0, \tag{3.15}$$

and if the claim is true, we have

$$a_j^{(k+1)}/b_j^{(k+1)} \geq \left(|T_k^{(s)}| + |T_k^{(t)}|\right)/\gamma = \lambda_k,$$

so that the crossing time is zero for edges $e_j \in \mathcal{A}_k^{(s)}$.
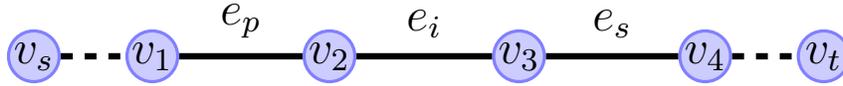
Figure 3.1: Path $\mathcal{P}_{s,t}$

Finally, we show the claim (3.15) is true by considering the two possible cases of the orders of edges are added:

i) The edges are added in the order $e_s \to e_p \to e_i$, the joining time for e_p and e_i are:

$$t_{p,k-1}^{\text{join}} = (|T_k^{(s)}|l_{v_2}^{(s)} - \sum_{v \in T_k^{(s)}} l_v^{(s)})^{-1}, \quad t_{i,k-1}^{\text{join}} = (|T_k^{(t)}|l_{v_2}^{(t)} - \sum_{v \in T_k^{(t)}} l_v^{(t)})^{-1}.$$

The assumption that e_p is added before e_i implies $t_{p,k-1}^{\text{join}} > t_{i,k-1}^{\text{join}}$ concluding the claim (3.15).

ii) The edges are added in the order $e_p \to e_s \to e_i$, the joining time for e_p and e_s are:

$$t_{p,k-2}^{\text{join}} = (|T_k^{(s)}|l_{v_2}^{(s)} - \sum_{v \in T_k^{(s)}} l_v^{(s)})^{-1}, \quad t_{s,k-2}^{\text{join}} = (|T_k^{(t)}|l_{v_2}^{(t)} - \sum_{v \in T_k^{(t)}} l_v^{(t)})^{-1}.$$

The order e_p is added before e_s concludes the claim (3.15) because $t_{p,k-1}^{\text{join}} > t_{s,k-1}^{\text{join}}$.

The proof that the crossing times for the edges that belong to the tree $\mathcal{A}_k^{(t)}$ is by symmetry

and interchanging s and t. □

We now establish a parallel between the LARS algorithm and the bi-directional Dijkstra algorithm, in that they share the defining feature, terminating when two trees built from opposite directions meet.

Theorem 3.1 (**Equivalence**). *The LARS algorithm iteratively builds two shortest-path trees, starting from roots v_s and v_t, and terminates when these two trees connect.*

Proof. Assuming the induction hypothesis in Assumption 3.2 holds true, we have shown that in the kth iteration: i) edges connecting to vertices with the shortest distance to v_s and v_t are added to the active set (Lemma 3.3); ii) crossing, where edges are removed from the active set, does not occur when $\lambda > 0$ (Lemma 3.4). Finally, iii) the algorithm terminates when the two shortest-path trees connect and $\lambda = 0$ (Lemma 3.5). □

3.4.3 Numerical example

We consider Nicholson's graph [84, p. 6] and seek the shortest path between vertex v_1 and v_9. The iterations of the LARS algorithm (with breakpoints $\lambda_0 = +\infty$, $\lambda_1 = 1/2$, $\lambda_2 = 1/3$, $\lambda_3 = 1/5$, $\lambda_4 = 0.1489$) are depicted in Fig 3.2a. It is observed that, at each iteration, edges highlighted in the same color as the corresponding λ are added to the active set and are never removed prior to the last step $\lambda_5 = 0$, as we prove later on in Lemma 3.4. The algorithm terminates after four iterations when $\lambda_5 = 0$ and a path between vertex v_1 and v_9 is formed. The element-wise path of the lasso solution $\boldsymbol{\beta}(\lambda)$ as λ decreases, is drawn in Fig 3.2b.

Example 3.2 highlights the correlation between the LARS algorithm and Dijkstra's algorithm. Specifically, the LARS algorithm builds two shortest-path trees, with roots at vertices v_1 and v_9. This echoes the steps in the bi-directional Dijkstra algorithm discussed in Section 3.1.2. In Theorem 3.1, we highlight similarities between the LARS and Dijkstra's
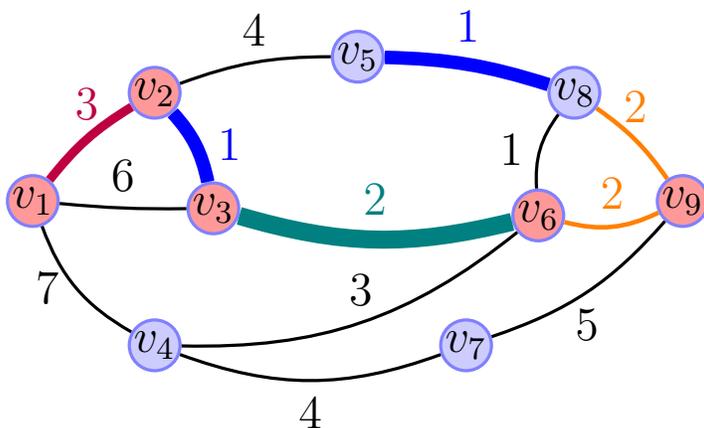
(a) Nicholson's graph: Edges are added according to breakpoints λ in the same color as in Fig. 3.2b.
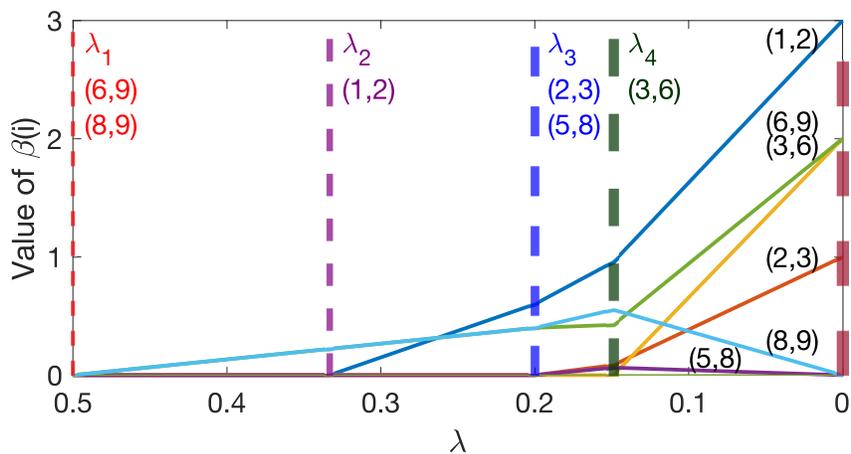
(b) Entries $(\boldsymbol{\beta}(\lambda))(i)$ (corresponding to edges) dawn as functions of λ. Values of λ, where the active edge-set changes, are marked with dashed vertical lines.

Figure 3.2: The LARS algorithm successively identifies a set of active edges while reducing the tuning/control parameter λ. A vector $\beta(\lambda)$ with information of the length-contribution of the active edge-set is also successively being updated.

algorithms.

3.5 Proximal algorithm

The LARS algorithm may provide a computational advantage if the desired path only contains a few edges. For large graphs, the number of breakpoints (proportional to the number of edges in the active set), tends to be very large, rendering $\beta(\lambda)$ intractable. The ADMM

algorithm, on the other hand, is particularly well-suited for solving large-scale convex optimization problems in a distributed manner due to its scalability [11, 9, 109].

3.5.1 ADMM and InADMM algorithm

Application of the ADMM to the lasso, as presented in [9, Section 6.4], is based on the reformulation of the lasso problem (3.4) as follows:

$$\min_{\beta, \alpha \in \mathbb{R}^m} \frac{1}{2}\|y - Q\beta\|_2^2 + \lambda\|\alpha\|_1 + \frac{\rho}{2}\|\beta - \alpha\|_2^2, \text{ s.t } \alpha = \beta, \tag{3.16}$$

where $\alpha \in \mathbb{R}^m$ is an additional optimization variable, and ρ is a positive constant.

The *Lagrangian* $L_\rho(\beta, \alpha, u)$ corresponding to the constrained optimization problem (3.16) is

$$L_\rho = \frac{1}{2}\|y - Q\beta\|_2^2 + \lambda\|\alpha\|_1 + u^T(\beta - \alpha) + \frac{\rho}{2}\|\beta - \alpha\|_2^2,$$

where $u \in \mathbb{R}^m$ is the *Lagrange multiplier*. Let $v \triangleq u/\rho$, the ADMM algorithm consists of the α, β-minimization steps and the step updating the dual variable v. Specifically, the optimal variables β, α, v is computed as

$$\beta^k := \arg \min_\beta L_\rho(\beta, \alpha^{k-1}, v^{k-1}) = (Q^TQ + \rho I)^{-1}\big(Q^Ty + \rho(\alpha^{k-1} - v^{k-1})\big) \tag{3.17}$$

$$\alpha^k := \arg \min_\alpha L_\rho(\beta^k, \alpha, v^{k-1}) = S_{\lambda/\rho}(\beta^k + \frac{1}{\rho}v^{k-1})$$

$$v^k := v^{k-1} + \rho(\beta^k - \alpha^k),$$

in the $(k-1)$th iteration. The β-update can be found in [9, Section 4.2] and $S_{\lambda/\rho}$ is the element-wise interpreted proximity operator of the ℓ_1 norm, known as the *soft-thresholding* operator in [9, Section 4.4.3].

To reduce the size and complexity of ADMM, we replace the matrix inversion $(Q^T Q + \rho I)^{-1}$ by

$$(Q^T Q + \rho I)^{-1} = \frac{1}{\rho}(I - Q^T(QQ^T + \rho I)^{-1}Q), \tag{3.18}$$

using the *matrix identity*, which instead involves $(Q^T Q + \rho I)^{-1}$. We will show this step significantly reduces the complexity of the algorithm in the next section.

To further reduce the complexity of ADMM for large-scale graphs, we use the InADMM algorithm introduced in [109] whose key idea is to approximately solve a system of linear equations instead of evaluating the matrix inversion exactly, using the matrix identity (3.18) to replace the β update in (3.17) with

$$h^{k-1} := Q^T y + \rho(\alpha^{k-1} - w^{k-1})$$
$$\eta^k := (QQ^T + \rho I)^{-1} Q h^{k-1}),$$
$$\beta^k := \frac{1}{\rho}(h^{k-1} - Q^T \eta^k)$$

and computes η^k approximately using the *conjugate gradient* (CG) method [57].

3.5.2   Worst-case complexity analysis and comparison

Throughout, we use the *worst-case complexity*, denoted as $\mathcal{O}(\cdot)$, to evaluate the algorithms' performance. Worst-case complexity quantifies the operations of the algorithm in the worst case, see [11, Appendix C], [30].

The complexity of ADMM is dominated by the matrix inversion $(Q^T Q + \rho I)^{-1}_{m \times m}$, which is of order $\mathcal{O}(m^3)$ (e.g. using *Cholesky decomposition*). However, by utilizing the matrix identity (3.18), this complexity is reduced to $\mathcal{O}(n^3)$. The most costly step in the CG method is the matrix-vector multiplication $(QQ^T + \rho I)x$ where $x \in \mathbb{R}^n$. This has a complexity of $\mathcal{O}(m)$ because the weighted incidence matrix Q has $2m$ nonzero elements.

Assuming the CG algorithm terminates in T_{CG} iterations, the complexity of the CG step in InADMM algorithm is of order $\mathcal{O}(mT_{CG})$. Noticing the complexity of other operations in InADMM is at most $\mathcal{O}(m)$, we summarize the complexities of ADMM and InADMM in Table 3.1 as below.

Variables	η	β(Cholesky)	α	v
ADMM	$\mathcal{O}(nm)$	$\mathcal{O}(n^3)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$
InADMM	$\mathcal{O}(mT_{CG})$	$\mathcal{O}(m)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$

Table 3.1: Complexities of ADMM and InADMM per iteration.

The comparison in complexity of the above methods with the Dijkstra's algorithm and its best improvement by *Fibonacci heap* is shown in Table 3.2.

ADMM[6]	InADMM[6]	Dijkstra	Fibonacci
$\mathcal{O}(n^3)$	$\mathcal{O}(mT_{CG})$[7]	$\mathcal{O}((m+n)\log_2 n)$	$\mathcal{O}(m+n\log_2 n)$

Table 3.2: Complexity comparison with Dijkstra and its variant.

3.5.3 Numerical experiments

We use "*intelligent scissors*"(also, "*live-wire*"), an image segmentation technique that is commonly used in computer vision [75]. It creates a *grid graph* over the image with each pixel represented by a vertex, and is connected to its 8 neighboring pixels through edges. The edge weights are assigned by a cost function [75, Section 3] based on image features. The objective is to identify a boundary between the portraits and the background, which turns out to be equivalent to finding the shortest path between selected pixels (marked in green) over the generated graph. The segmentation is obtained by identifying the edges in the shortest

---

[6]Per iteration, both methods convergence within 40 steps for a graph with 17291 edges, see Fig 3.4

[7]$T_{CG} \ll n$ and scalable with respect to the size of the graph. Empirically, we have $T_{CG} = 350$ for graph with 4422 vertices and $T_{CG} = 320$ for graph with 5694 vertices.

path. Herein, we use images with 5694 and 4422 pixels respectively. The edges, detected by constructing shortest paths using three different methods and colored in red, are drawn in Fig. 3.3 (Code is available at `https://github.com/dytroshut/Lasso-shortest-path.git` with functions from [10]).

For the InADMM, we utilized the CG method from [57], see also [109] for more details. The convergence rates of ADMM, InADMM, and Basis Pursuit[9] [9, Section 6.1], are shown in Fig 3.4. Further, to highlight the efficiency of lasso's distributed implementation, we used InADMM (with a pre-specified path as input).

(a) Dijkstra   (b) ADMM   (c) InADMM
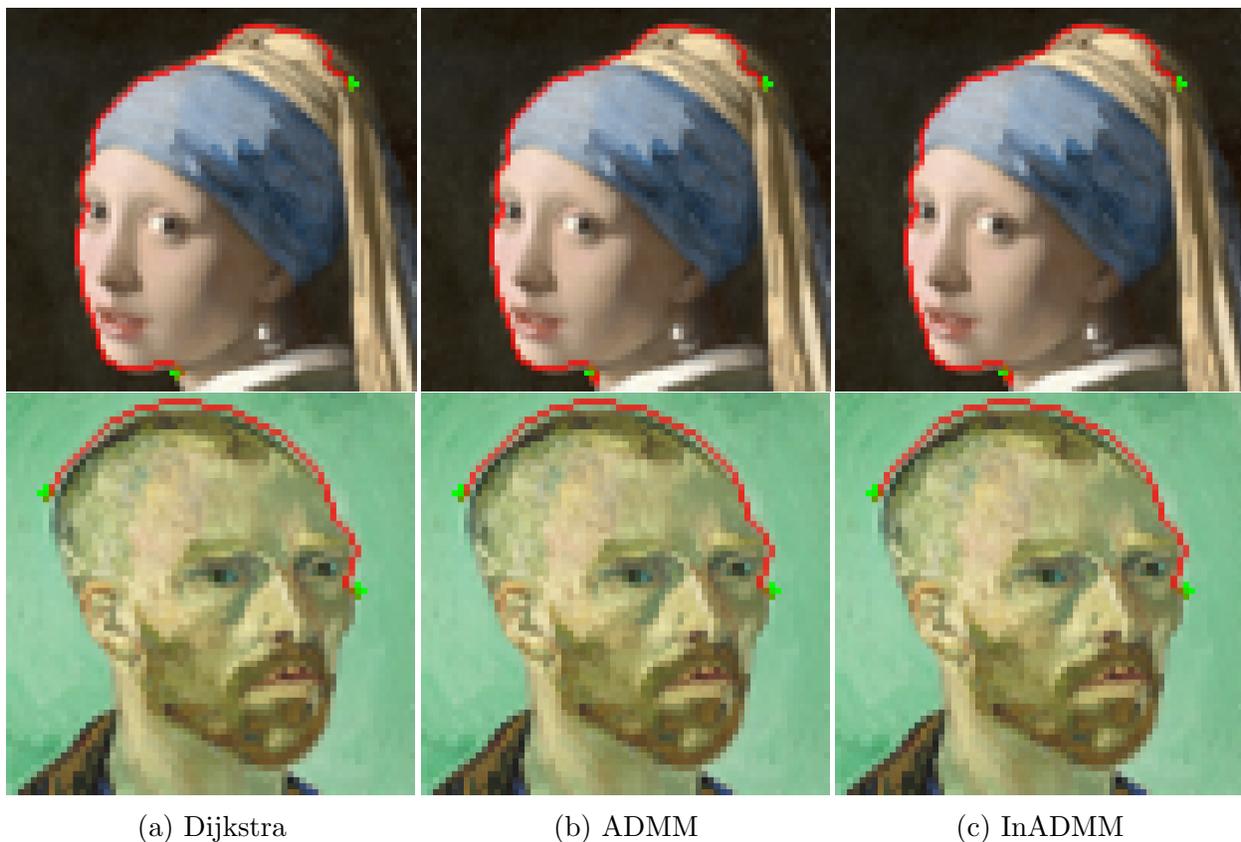
Figure 3.3: "Edge detection" in an image as a short path, highlighted in red, and obtained using Dijkstra's algorithm (Fig. 3.3a), as well as using the lasso solution β in ADMM and InADMM in Fig. 3.3b and Fig. 3.3c, respectively.

---

[9]Basis Pursuit is a technique to obtain sparse solution to an underdetermined system of linear equations.
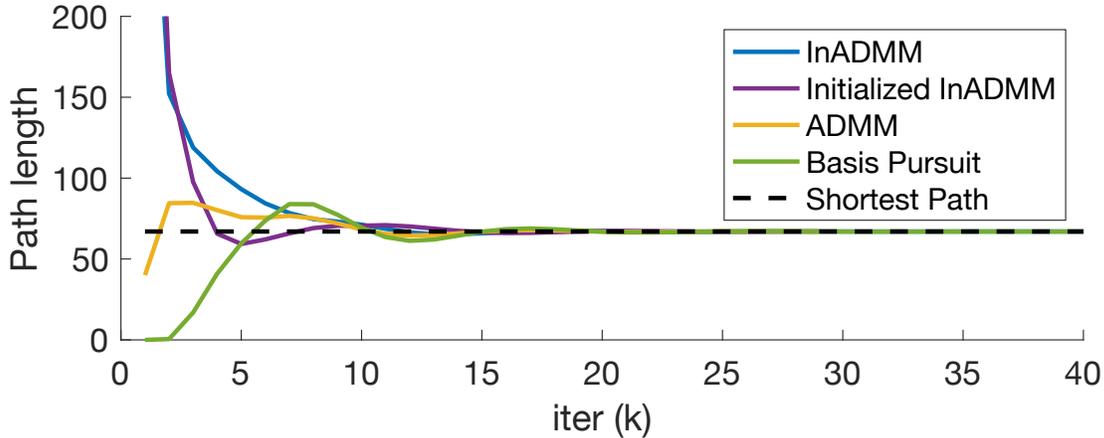
Figure 3.4: Example based on the *Van Gogh* painting (4422 vertices and 17291 edges): "Edge in image" identified via InADMM, InADMM with initializer, ADMM, and Basis pursuit. These converge in 36, 34, 29, 47 steps, respectively, with running times 1.7308, 1.7618, 4.0249, 35.7527 seconds in MATLAB clock. Also, the running times for solving the linear program (3.3b) using three methods (*dual-simplex and interior-point(-legacy)*) are 26.6145, 29.3866, 32.2845 seconds.

Bibliographical notes

The problem of finding the shortest path between two vertices of a graph has a long history[107, 100] and a wide range of applications [106, 75]. A classical algorithm to determine a shortest path is due to Dijkstra [35]. Since Dijkstra's early work, a variety of alternative methods have been developed to reduce complexity and address variants of the problem [6, 2, 19, 103, 45, 110]. A salient issue in applications involving graphs of considerable size, which motivated the present work, is that identifying a shortest path is not absolutely essential, whereas identifying a reasonably short path may suffice [106, 85].

Driven by such considerations and inspired by the effectiveness of convex optimization techniques to address large-scale problems [11, 9], we introduce a formulation of the shortest path problem as an ℓ_1-regularized regression, known as the *Least Absolute Shrinkage and Selection Operator (lasso)*[101]. This type of regularization/relaxation is ubiquitous in inverse problems throughout engineering, statistics and mathematics, with a rich library of numerical implementations for high-dimensional problems.

Chapter 4

Monge transportation through toll

In the chapter, we formulate and address a natural variant of the standard optimal mass transport problem by imposing a hard constraint on the flux rate at a point along the path between distributions. Specifically, we pose and resolve the most basic such problem where the restriction on throughput of the transport plan takes place at a single point. With this constraint in place, we seek to minimize a usual quadratic cost functional.

The analysis we provide focuses on one-dimensional distributions, with transport taking place on $\mathbb{R}$. We prove existence and uniqueness of an optimal transport plan, and under suitable regularity conditions, give an explicit construction. A slight generalization of our formulation, where the distributions have support on $\mathbb{R}^d$ but the transportation is to take place through a specified "constriction" point, with a similar throughput constraint, can be worked out in the same manner and it is sketched in the concluding remarks. The more general case where the transport takes place on higher dimensional manifolds with the throughput through possibly multiple points, curves, or surfaces similarly restricted is substantially more challenging and much remains open.

The problem formulation and ideas in the mathematical analysis that follows can be

visualized by appealing to Figure 4.1. We begin with two probability densities ρ_0, ρ_1 having support on $\mathbb{R}$ and finite second-order moments, and seek to transport one to the other, ρ_0 to ρ_1, within a window of time (herein, of duration normalized to 1) while minimizing a quadratic cost in the local velocity. That is, we seek to minimize the action integral of kinetic energy along the transport path. The minimal cost of the unrestricted transport is the so-called Wasserstein distance $\mathcal{W}_2(\rho_0, \rho_1)$ (a metric on the space of probability measures); we refer to standard references [104, 105] for the unconstrained optimal transport problem. The schematic in Figure 4.1 exemplifies a constraint at a pre-specified point, x_0, that can be seen as the location of constriction, or, of a toll along the transport, where throughput is bounded. That is, the flow rate across x_0 for mass times velocity is bounded by a value h. A vertical axis pointing downwards at x_0 marks the time when a specific mass-element crosses the toll, necessitating at least $1/h$ duration for the unit mass of the probability density ρ_0 to go through, in the most favorable case where the throughput rate is maintained for the duration (that is normalized to 1 time unit).

In the body of the chapter, we prove existence and uniqueness of an optimal transport plan and, assuming suitable regularity of the distributions, we provide an explicit construction for the solution. We further explore consequences of the toll being kept maximally "busy" while mass is being transported through, in conjunction with minimizing the quadratic cost criterion on the kinetic energy, and we highlight ensuing properties of the optimal plan.

Specifically, in Section 4.1 we develop the formulation of the flux constraint and give a precise definition of the problem (Problem 4.1). In Section 4.2 we prove existence and uniqueness (Theorem 4.1) of solution, while conveniently recasting the problem in terms of a flux variable (Problem 4.2). Section 4.4 deals with the structural form of the transport and properties of solutions. We summarize the basic elements that allow an explicit construction of the solution in Theorem 4.2. Section 4.5 provides a rudimentary example of transporting

between uniform distributions, that highlights the essential property that speed needs to be suitably adjusted so as to fully utilize the throughput of the toll, while minimizing the quadratic cost. We close (Section 4.6) with a discussion on possible extension of the problem to higher dimensions and multiple tolls. While the theory may be readily extended in certain cases, much remains to be understood. Such problems are of natural engineering and scientific interest.
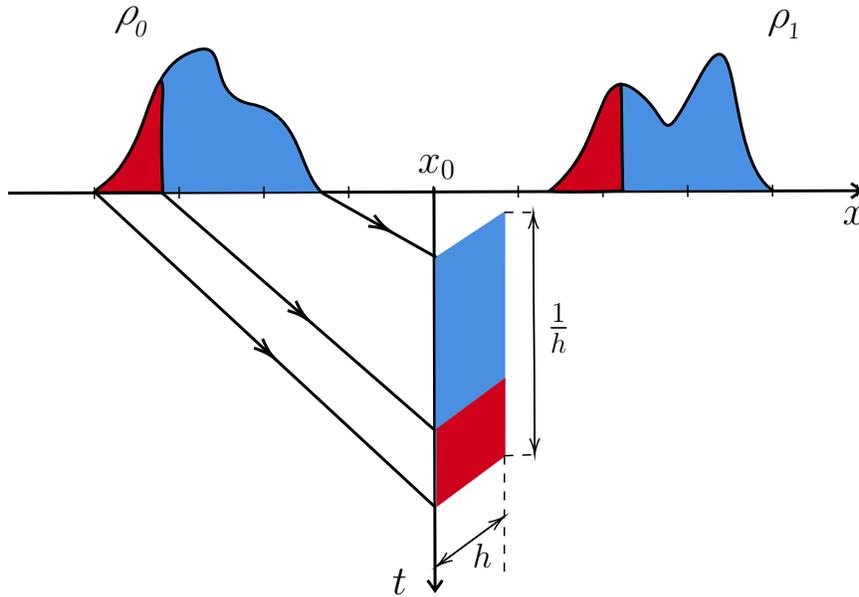
Figure 4.1: Illustration of optimal transport through a toll with finite throughput

4.1 Problem formulation

We consider two probability densities ρ_0, ρ_1 on $\mathbb{R}$ having finite second-order moments. For $X : [0, 1] \times \mathbb{R} \to \mathbb{R}$ such that[1] $X_{0\#\rho_0} = \rho_0$ and $X_{1\#\rho_0} = \rho_1$, we are interested in minimizing

$$J(\partial_t X) := \int_0^1 \int_{\mathbb{R}} (\partial_t X_t(x))^2 \rho_0(x) dx dt. \tag{4.1}$$

---

[1] As is common, $X_{t\#\rho_0}$ denotes the push-forward of ρ_0 under X_t, see [104].

In the absence of any additional constraint on X, the solution is

$$X_t^\star(x) = x + t(T(x) - x)$$

for T the optimal transport map between ρ_0 and ρ_1 and $J(\partial_t X_t^\star) = \mathcal{W}_2^2(\rho_0, \rho_1)$, the squared Wasserstein-2 distance between the two [104]. Here however, for a certain $x_0 \in \mathbb{R}$, we introduce a constraint on the flux passing through x_0, as explained below. Throughout the paper, T will always denote the optimal transportation plan in the absence of any such constraint. The purpose of the present work of course is to develop theory that addresses the case of transport with a bound on the flux through x_0.

When all functions are smooth and well defined, a flux constraint at x_0 can be expressed as

$$|\rho_t(x_0) v_t(x_0)| \leq h \quad \forall t \in (0, 1)$$

for ρ_t the density of $X_{t\#\rho_0}$ and $v_t(x_0) = \partial_t X_t(X_t^{-1}(x_0))$. However in the general case, if ρ_t is not continuous (or doesn't even exist), this constraint is not well defined. One way to deal with such a situation is to recast the constraint as requiring that[2], $\forall t \in (0, 1)$,

$$\limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \frac{1}{|\alpha_2 - \alpha_1|} \int \mathbb{1}_{\{x_0 \in (X_t(x) + \alpha_1, X_t(x) + \alpha_2)\}} |\partial_t X_t(x)| \rho_0(x) dx \leq h. \tag{4.2}$$

Then, if ρ_0 is continuous and X_t is a C^1 diffeomorphism, the left hand side (LHS) of (4.2) amounts to

$$\text{LHS (4.2)} = \limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \frac{1}{|\alpha_2 - \alpha_1|} \int \mathbb{1}_{\{x_0 \in (y + \alpha_1, y + \alpha_2)\}} |\partial_t X_t(X_t^{-1}(y))| \rho_t(y) dy$$

$$= \rho_t(x_0) v_t(x_0),$$

---

[2]We use the standard notation $\mathbb{1}_A$ for the characteristic function of the set A.

Interestingly, when X_t fails to be a C^1 diffeomorphism, special care is needed. For instance, take $x_0 = 0$ and $X_t(x) = \mathbb{1}_{\{x \in [-2,-1]\}}(1 - 2t)^3 x$. The constraint (4.2) is satisfied since $\partial_t X_t(x) = 0$ at $t = 1/2$, and no mass sits near the toll for any $t \neq 1/2$. Thus, the formulation (4.2) fails to capture the situation where infinite mass passes through with zero velocity. We reformulate so as to avoid this technicality.

Consider the modified constraint that bounds the flux passing through x_0, expressed as requiring that $\forall t \in (0,1)$

$$\limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \frac{1}{|\alpha_2 - \alpha_1|} \int \mathbb{1}_{\{x_0 \in (X_{t+\alpha_1}(x), X_{t+\alpha_2}(x))\}} \rho_0(x) dx \le h. \tag{4.3}$$

In the case where X_t is C^1, using the Taylor expansion of X in time, the left hand side (LHS) of (4.3) amounts to

$$
\begin{aligned}
\text{LHS (4.3)} =& \limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \frac{1}{|\alpha_2 - \alpha_1|} \int \mathbb{1}_{\{x_0 \in (X_t(x) + \partial_t X_t(x)\alpha_1 + o(\alpha_1), X_t(x) + \partial_t X_t(x)\alpha_2 + o(\alpha_2))\}} \rho_0(x) dx \\
=& \limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \int \Bigg(\mathbb{1}_{\{X_t(x) \in (x_0 - \partial_t X_t(x)\alpha_2 + o(\alpha_2), x_0 - \partial_t X_t(x)\alpha_1 + o(\alpha_1))\}} \frac{|\partial_t X_t(x)| \mathbb{1}_{\{|\partial_t X_t(x)| > 0\}}}{|\alpha_2 - \alpha_1| |\partial_t X_t(x)|} \\
& + \mathbb{1}_{\{X_t(x) \in (x_0 + o(\alpha_1), x_0 + o(\alpha_2))\}} \frac{\mathbb{1}_{\{\partial_t X_t(x) = 0\}}}{|\alpha_2 - \alpha_1|} \Bigg) \rho_0(x) dx.
\end{aligned}
$$

Using a change of variables, we readily see that (4.3) implies (4.2) and that if X_t is a C^1 diffeomorphism, the two constraints are identical. Note also that, $\forall t \in (0,1)$, condition (4.3) is equivalent to

$$\forall \alpha_1, \alpha_2 \in \mathbb{R}, \quad \int \mathbb{1}_{\{x_0 \in (X_{t+\alpha_1}(x), X_{t+\alpha_2}(x))\}} \rho_0(x) dx \le h |\alpha_2 - \alpha_1|. \tag{4.4}$$

Define $\Omega = \{x \in \text{Supp}(\rho_0) \mid x_0 \in (x, T(x)) \text{ or } x_0 \in (T(x), x)\}$, where T is the optimal transport map of the unconstrained problem. Thus, Ω contains the support of mass that needs to cross the toll station, at some point in time, in either direction. From (4.4) it is

evident that $h \geq \rho_0(\Omega)$ is necessary for the existence of a map satisfying the constraint (since the transport will take place over the time interval $[0, 1]$). Typically, $h > \rho_0(\Omega)$ is required, except in some special cases where $h = \rho_0(\Omega)$ may suffice, as for example when $\rho_0 = \mathbb{1}_{\{[0,1]\}}$, $x_0 = 1$ and $\rho_1 = \mathbb{1}_{\{[1,2]\}}$. From here on we assume that $h > \rho_0(\Omega)$.

We are now in a position to cast our optimization problem in terms of a velocity field $v_t(x)$ that will effect the transport; formally, $v_t(x) = \partial_t X_t(x)$ relates to our earlier notation when functions are smooth. For any $v \in L^2([0, 1] \times \mathbb{R}, \mathbb{R})$, define the map $X^v : [0, 1] \times \mathbb{R} \to \mathbb{R}$ as the flow of v:

$$X_t^v = \text{Id} + \int_0^t v_\tau d\tau, \tag{4.5}$$

with Id denoting the identity map in $\mathbb{R}$. Our problem can now be stated as follows.

Problem 4.1. *Consider*

$$\inf_{v \in \Lambda} J(v), \tag{4.6}$$

over the class Λ of functions $v \in L^2([0, 1] \times \mathbb{R}, \mathbb{R})$ defined so that X^v, the flow of v, satisfies

i) $X_{1\#\rho_0}^v = \rho_1$

ii) $\forall t \in (0, 1)$, X_t^v *satisfies the constraint* (4.4).

Determine existence, uniqueness, and a functional form for a minimizing solution v.

4.2   Existence of a solution

We say that a map $X : [0, 1] \times \mathbb{R} \to \mathbb{R}$ is in the set Γ if there exist $v \in \Lambda$ such that X is the flow of v, i.e. $X_t = X_t^v = \text{Id} + \int_0^t v_\tau d\tau$. From here on, the v in the notation X_t^v is suppressed as we are truly interested in the transport map. We first derive certain useful properties of

candidate minimizers of our problem. To this end, for any $X \in \Gamma$ and $x \in \Omega$, we define

$$toll_X(x) = \inf\{t \mid x_0 = X_t(x)\}.$$

Thus, the function $toll_X$ specifies the times of transit through the toll station of mass that is initially located at x and then transported via X.

It is clear that the function $toll_X$ must be injective[3] for a minimizing solution, and that mass flow takes place always in the same direction across the toll station. Then, $\forall t \in (0,1)$, (4.3) is equivalent to

$$\limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \frac{1}{|\alpha_2 - \alpha_1|} toll_{X \# \rho_0}((t + \alpha_1, t + \alpha_2)) \leq h,$$

and so, if $toll_{X \# \rho_0}$ (the measure on $[0,1]$ that weighs the mass that goes through x_0 at different times $t \in [0,1]$) admits a continuous density ϱ_{toll}, the constraint amounts to $\varrho_{toll}(t) \leq h$. Note also that this condition is different than simply stating $\rho_t(x_0) \leq h$, as the latter doesn't take into account the speed of transport. Then we see that for $x \notin \Omega$, we can restrict ourselves to considering maps $X \in \Gamma$ such that $X_t(x) = x + t(T(x) - x)$ for T the optimal transport map between ρ_0 and ρ_1. Thus, in the sequel, without loss of generality we always suppose that $\Omega = \mathrm{Supp}\,(\rho_0)$ and that $\sup \mathrm{Supp}\,(\rho_0) \leq x_0 \leq \inf \mathrm{Supp}\,(\rho_1)$.

For $X \in \Gamma$ and $toll_X$ its corresponding transit-time function, define the map $\overline{X} : [0,1] \times \mathbb{R} \to \mathbb{R}$ by

$$\overline{X}_t(x) = \begin{cases} x + t\frac{x_0 - x}{toll_X(x)} & \text{if} t \leq toll_X(x) \\[2mm] x_0 + (t - toll_X(x))\frac{T(x) - x_0}{1 - toll_X(x)} & \text{if } t \geq toll_X(x) \end{cases} \tag{4.7}$$

and note $\overline{\Gamma} = \{\overline{X} \mid X \in \Gamma\}$ the set of functions of this type. The next statement states that we can restrict our minimization problem to functions of the form (4.7). Specifically, it states

---

[3]This follows by cyclic monotonicity since the cost is convex, see [105, Section 2.3].

that for any candidate minimizer $X \in \Gamma$, the speed of transport needs to remain constant at all times prior to transit, and again, constant at all times after transit. In addition, from the functional form, we see that $X_1 = T(x)$ for all x. This last statement says that the final destination of mass originally located at x is the same, whether we apply T or the optimal plan that abides by the constraint; the only thing that changes in the two cases is the speed while the mass traverses the segment before x_0 and after (cf. example in Section 4.5).

Proposition 4.1. *We have*

$$\inf_{X \in \Gamma} J(\partial_t X) = \inf_{\overline{X} \in \overline{\Gamma}} J(\partial_t \overline{X})$$

Proof. For $X \in \Gamma$ and $toll_X$, define

$$X^c(x) = \begin{cases} x + t\frac{x_0 - x}{toll_X(x)} & \text{if } t \leq toll_X(x) \\ x_0 + (t - toll_X(x))\frac{X_1(x) - x_0}{1 - toll_X(x)} & \text{if } t \geq toll_X(x) \end{cases} \tag{4.8}$$

Thus, X^c maintains the terminal destination $X_1(x)$ and the crossing time $toll_X(x)$, for the mass that was initially at x, while it ensures constancy of speed before and after crossing. It follows that $X^c \in \Gamma$ and that $J(\partial_t X^c) \leq J(\partial X)$, by convexity, so we can restrict $\mathcal{X}$ to the set of functions that are of the form (4.8) since candidate minimizers will always be of that form.

As the position $X_1(x)$ in (4.8) doesn't impact the constraint (4.3), we consider how $X_1(x)$ may depend on the time of crossing $toll_X(x)$. Specifically, X_1 must be a minimum for the cost

$$\int_{toll_X(x)}^1 \int_{\mathbb{R}} \left(\frac{X_1(x) - x_0}{1 - toll_X(x)}\right)^2 \rho_0(x)dxdt = \int_{\mathbb{R}} \frac{(X_1(x) - x_0)^2}{1 - toll_X(x)}\rho_0(x)dx.$$

From this we deduce that $toll_X(x) \leq toll_X(y)$ iff $X_1(x) \geq X_1(y)$. Furthermore, as the problem is reversible (we can switch ρ_0 and ρ_1), we can deduce in the same way that $toll_X(x) \geq toll_X(y)$ iff $x \leq y$. Therefore $X_1(x)$ is increasing and we conclude that it is

identical to T the optimal transport map between ρ_0 and ρ_1. $\qquad\square$

From Proposition 4.1 we also deduce that for X, the flow of a (candidate) optimal solution, the map $x \mapsto toll_X$ is strictly decreasing on the support of ρ_0, and that X_t is one to one, for all t.

Let us write $v(x) = \frac{x_0 - x}{toll_X(x)}$ for the velocity of transport *prior* to crossing the toll, for the mass initially located at x at the start. Then, in light of Proposition 4.1, our problem is reduced to finding

$$ v \in \arg\min \int_0^1 \int_{\mathbb{R}} \left(v(x)^2 \mathbb{1}_{\{t \leq \frac{x_0 - x}{v(x)}\}} + \left(\frac{T(x) - x_0}{1 - \frac{x_0 - x}{v(x)}} \right)^2 \mathbb{1}_{\{t \geq \frac{x_0 - x}{v(x)}\}} \right) \rho_0(x) dx dt $$

$$ = \int_{\mathbb{R}} \left(v(x)(x_0 - x) + \frac{(T(x) - x_0)^2}{1 - \frac{x_0 - x}{v(x)}} \right) \rho_0(x) dx, \tag{4.9} $$

subject to $x \mapsto \frac{x_0 - x}{v(x)} = toll_v(x)$ being decreasing and bounded between 0 and 1, and

$$ \limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \frac{1}{|\alpha_2 - \alpha_1|} \int \mathbb{1}_{\{(t + \alpha_1)v(x) < x_0 - x < (t + \alpha_2)v(x)\}} \rho_0(x) dx \leq h. \tag{4.10} $$

We now argue the existence of a minimizer $v^\star$.

Proposition 4.2 (Existence). *Supposing that the two probabilities densities ρ_0, ρ_1 have finite second-order moments, Problem 4.1 admits a solution.*

Proof. Let $(v_n)_n$ be a minimizing sequence of (4.9) and write $toll_n : \text{Supp}(\rho_0) \to (0, 1)$ the associated toll function: $toll_n(x) = \frac{x_0 - x}{v_n(x)}$. Let $(\alpha_k)_k$ be a dense sequence in $\text{Supp}(\rho_0)$ (for example the rational numbers). By compactness, we have that $\forall k \in \mathbb{N}$, $toll_n(a_k)$ admits a converging subsequence in n. Then using a diagonal argument, there exist a subsequence

$(v_{\varphi(n)})_n$ and $\beta_k \in [0,1]$ such that,

$$\forall k \in \mathbb{N}, \quad toll_{\varphi(n)}(a_k) \xrightarrow[n \to +\infty]{} \beta_k \text{ and } \alpha_k \leq \alpha_l \iff \beta_k \leq \beta_l.$$

For $x \in \mathrm{Supp}\,(\rho_0)$, and $(\alpha_{\psi(k)})_k$ a decreasing subsequence converging to x, let be $toll(x) = \lim_k \beta_{\psi(k)}$, which is well defined as $\beta_{\psi(k)}$ is decreasing. Then $toll_{\varphi(n)}(x)$ converges to $toll(x)$ for any x being a point of continuity of $toll$. As $toll$ is a nonincreasing map, it has at most a countable number of points of discontinuity, therefore $toll_{\varphi(n)}$ converges to $toll$ a.e. In partic-ular we get that $toll_{\varphi(n)\#\rho_0}$ converges weakly to $toll_{\#\rho_0}$. For $x \in \mathrm{Supp}\,(\rho_0) \setminus \{x \mid toll(x) = 0\}$, define $v(x) = \frac{x_0 - x}{toll(x)}$, it is well defined a.e. because $\{x \mid toll(x) = 0\}$ has measure 0 as $(v_n)_n$ is a minimizing sequence. Then $v_{\varphi(n)}$ converges a.e. to v and as the constraint (4.10) is equivalent to

$$\forall \alpha_1, \alpha_2 \in \mathbb{R}, \quad toll_v((t+\alpha_1, t+\alpha_2)) \leq h|\alpha_2 - \alpha_1|,$$

v verifies the constraint. Finally, by lower semi continuity of the cost, v is a minimizer of (4.9). $\qquad\square$

4.3 Uniqueness of the solution

Before we proceed with the proof of uniqueness of the minimizer, we recast our problem in terms of flux as the optimization variable. For $u \in L^1([0,1] \times \mathbb{R}, \mathbb{R})$, a candidate flux (i.e., mass times velocity), define a corresponding mass-measure ρ_t^u on $\mathbb{R}$ by duality via: $\forall \phi \in C_c^\infty(\mathbb{R}, \mathbb{R})$,

$$\int_{\mathbb{R}} \phi(x) d\rho_t^u(x) = \int_{\mathbb{R}} \phi(x) \rho_0(x) dx + \int_0^t \int_{\mathbb{R}} (\nabla \phi(x)) u_r(x) dx dr.$$

Equivalently, we have that ρ^u solves in the weak sense the continuity equation

$$\begin{cases} \partial_t \rho_t^u = -\nabla \cdot u \\ \rho_0^u = \rho_0 \end{cases}.$$

For a flux u such that $\forall t \in (0,1)$, ρ_t^u admits a positive density, let us express the cost of u as

$$\mathrm{J}(u) = \int_0^1 \int_{\mathbb{R}} \frac{u_t(y)^2}{\rho_t^u(y)} dy dt \tag{4.11}$$

In the above, by a slight abuse of notation as it is often done, we used ρ^u to denote both the measure and the corresponding density, allowing these to be distinguished by the specific usage and context.

Problem 4.2. *Consider*

$$\inf_{u \in \mathcal{U}} \mathrm{J}(u). \tag{4.12}$$

over the class $\mathcal{U}$ defined as the set of functions $u \in L^1([0,1] \times \mathbb{R}, \mathbb{R})$ a.e. such that

i') $\forall t \in (0,1)$, ρ_t^u admits a positive density and $\rho_1^u = \rho_1$

ii') satisfy

$$\forall t \in (0,1), \quad \limsup_{x_1 \to x_0, \ x_2 \to x_0} \frac{1}{|x_2 - x_1|} \int_{x_1}^{x_2} |u_t(y)| dy \le h. \tag{4.13}$$

Determine existence, uniqueness, and a functional form for a minimizing solution u.

We will first prove the equivalence of the above formulation in Problem 4.2 with that in Problem 4.1. The advantage of Problem 4.2 is that the constraint is now convex which will be convenient in proving uniqueness. Note that here we use roman J with argument the flux field, to echo the earlier usage in (4.1) where the action integral J first appeared with argument the velocity.

69

Proposition 4.3. *Problems 4.1 and 4.2 are equivalent.*

Proof. Let $X \in \bar{\Gamma}$ be a solution of Problem 4.1, $v_t(\cdot) = \partial_t X_t(X_t^{-1}(\cdot))$ the associated velocity (defined everywhere except at the points $(toll(x), x)$, for all $x \in \text{Supp}(\rho_0)$) and $\rho_t = X_{t\#\rho_0}$ the associated mass flow. Then for $\phi \in C_c^\infty(\mathbb{R}, \mathbb{R})$ we have

$$
\begin{aligned}
\int_{\mathbb{R}} \phi(x) d\rho_t(x) &= \int_{\mathbb{R}} \phi(X_t(x)) \rho_0(x) dx \\
&= \int_{\mathbb{R}} \left(\phi(X_0(x)) + \int_0^t \partial_t \phi(X_r(x)) dr \right) \rho_0(x) dx \\
&= \int_{\mathbb{R}} \left(\phi(X_0(x)) + \int_0^t \nabla \phi(X_r(x)) v_r(X_r(x)) dr \right) \rho_0(x) dx \\
&= \int_{\mathbb{R}} \phi(x) \rho_0(x) dx + \int_0^t \int_{\mathbb{R}} \nabla \phi(x) v_r(x) d\rho_r(x) dr.
\end{aligned}
$$

Therefore X defines a unique flux $u \in L^1([0,1] \times \mathbb{R}, \mathbb{R})$ (u is L^1 by Jensen inequality) by $u_t(x) = v_t(x)\rho_t(x)$ with $J(u) = J(\partial_t X)$. Furthermore, for $v(x) = \partial_t X_0(x) = \frac{x_0 - x}{toll_v(x)}$ we have that the left hand side of (4.10) amounts to

$$
\begin{aligned}
\text{LHS (4.10)} &= \limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \int \mathbb{1}_{\{X_t(x) + \alpha_1 v(x) < x_0 < X_t(x) + \alpha_2 v(x)\}} \frac{v(x)}{|\alpha_2 - \alpha_1| v(x)} \rho_0(x) dx \\
&= \limsup_{\substack{\epsilon_1 \to 0 \\ \epsilon_2 \to 0}} \frac{1}{|\epsilon_2 - \epsilon_1|} \int \mathbb{1}_{\{y + \epsilon_1 < x_0 < y + \epsilon_2\}} v_t(y) \rho_t(y) dy.
\end{aligned}
$$

Therefore $u \in \mathcal{U}$ and we conclude that $\inf_{u \in \mathcal{U}} J(u) \leq \min_{X \in \Gamma} J(\partial_t X)$.

For establishing the reverse direction, let $u \in \mathcal{U} \cap C([0,1], C_c^1(\mathbb{R}, \mathbb{R}))$ with $J(u) < \infty$ and define T_t^u the optimal transport map between ρ_0 and ρ_t^u. For $F_t(x) = \int_{-\infty}^x \rho_t^u(x) dx$ the cumulative distribution function of ρ_t^u, it is well known that $T_t^u(x) = F_t^{-1}(F_0(x))$, see [105, Chapter 1]. Since $\forall t \in [0,1]$ we have $F_t(F_t^{-1}(F_0(x))) = F_0(x)$, differentiating this expression

70

we have

$$\partial_t F\Big|_{(t,x)=(t,F_t^{-1}(F_0(x)))} + \partial_x F\Big|_{(t,x)=(t,F_t^{-1}(F_0(x)))} \partial_t F_t^{-1}(F_0(x)) = 0$$

$$\int_{-\infty}^{F_t^{-1}(F_0(x))} \partial_t d\rho_t^u(x) + \rho_t^u(F_t^{-1}(F_0(x)))\partial_t F_t^{-1}(F_0(x)) = 0$$

$$\int_{-\infty}^{T_t^u(x)} -\nabla u_t(z)dz + \rho_t^u(T_t^u(x))\partial_t T_t^u(x) = 0$$

$$\Rightarrow \partial_t T_t^u(x) = \frac{u_t(T_t^u(x))}{\rho_t^u(T_t^u(x))}.$$

Therefore T_t^u defines a map in Γ such that $\mathrm{J}(u) = J(\partial_t T_t^u)$. Then, since the space $C([0,1], C_c^1(\mathbb{R}, \mathbb{R}))$ is dense in $L^1([0,1] \times \mathbb{R}, \mathbb{R})$, we deduce that $\min_{u \in \mathcal{U}} \mathrm{J}(u) = \min_{X \in \Gamma} J(\partial_t X)$. □

Using the equivalence of Problem 4.1 and Problem 4.2, we can now prove the uniqueness of the minimizer.

Theorem 4.1 (**Uniqueness**). *Problem 4.2 (and so Problem 4.1) admits a unique solution.*

Proof. Suppose that we have u_1 and u_2, two solutions of (4.9). For $\lambda \in (0,1)$, by convexity we have that $\frac{(\lambda u_1 + (1-\lambda)u_2)^2}{\lambda \rho^{u_1} + (1-\lambda)\rho^{u_2}} \leq \lambda \frac{u_1^2}{\rho^{u_1}} + (1-\lambda)\frac{u_2^2}{\rho^{u_2}}$, but as they are both solutions, this is an equality. However the polynomial $\lambda \mapsto (\lambda(u_1 - u_2) + u_2)^2 - (\lambda(\rho^{u_1} - \rho^{u_2}) + \rho^{u_2})(\lambda(\frac{u_1^2}{\rho^{u_1}} - \frac{u_2^2}{\rho^{u_2}}) + \frac{u_2^2}{\rho^{u_2}})$ is identically zero iff $\frac{u_1}{\rho_1} = \frac{u_2}{\rho_2}$, and iff $v_1 = v_2$. □

4.4 Properties and structural form of the solution under smoothness assumption

We are now in a position to build explicitly the solution $v^\star$ of Problem (4.9) in the case when ρ_0 and ρ_1 have additional smoothness assumptions. All along this section, we will assume that ρ_0 and ρ_1 are continuous, have bounded convex support, and are bounded from below on the interior of their support. In the process of building the solution, we also establish

structural properties of the solution.

Under the stated assumptions on ρ_0, ρ_1, by using the closed-form expression for the optimal transport map T in dimension one [105, Chapter 1], it is immediate to see that T is C^1.

Recall first that, without loss of generality, we assume that $\sup \text{Supp}(\rho_0) \leq x_0 \leq \inf \text{Supp}(\rho_1)$. For $v : \text{Supp}(\rho_0) \to \mathbb{R}$ such that[4] $x \mapsto toll_v(x) = \frac{x_0 - x}{v(x)}$ is decreasing and bounded between 0 and 1 on $\text{Supp}(\rho_0)$, the expression

$$C_y(v) = \limsup_{\substack{\alpha_1 \to 0 \\ \alpha_2 \to 0}} \frac{1}{|\alpha_2 - \alpha_1|} \int \mathbb{1}_{\{(toll_v(y)+\alpha_1)v(x) < x_0 - x < (toll_v(y)+\alpha_2)v(x)\}} \rho_0(x) dx$$

gives the value of the flux passing through the toll station when the mass initially at y is crossing. Let first prove that from the additional assumptions on ρ_0 and ρ_1, we have that the solution is continuous.

Proposition 4.4. *The solution* $v^\star \in L^2$ *admits a continuous representative.*

Proof. From section 1, we know that the solution $v^\star \in L^2$ admits a representative such that the function $x \mapsto toll_{v^\star}(x) = \frac{x_0 - x}{v^\star(x)}$ is decreasing. Now by absurd, suppose that $v^\star$ is not continuous. Then there exists $x_0 \in \text{Supp}(\rho_0)$ and $\epsilon > 0$ such that $\forall \delta > 0, \exists x_\delta \in \text{Supp}(\rho_0)$ with $|x_0 - x_\delta| < \delta$ and $|v^\star(x_0) - v^\star(x_\delta)| > \epsilon$. As $toll_{v^\star}$ is decreasing, we have that for δ small enough,

$$\frac{v^\star(x_0) - v^\star(x_\delta)}{|v^\star(x_0) - v^\star(x_\delta)|} = \frac{x_0 - x_\delta}{|x_0 - x_\delta|}$$

so $toll_{v^\star}$ is not continuous in x_0 neither. Suppose now that $\forall \delta > 0, x_\delta - x_0 > 0$ (the proof would be the same for $x_\delta - x_0 < 0$). Then we have that

$$\lim_{\substack{x \to x_0 \\ x > x_0}} toll_{v^\star}(x) < toll_{v^\star}(x_0).$$

---

[4]The notation $toll_v$ signifies $toll_X$, for the corresponding X obtained via (4.5).

If $v^\star(x_0) < T(x_0) - x_0$, then as $toll_{v^\star}$ is decreasing and T is continuous, we have that for $\gamma > 0$ small enough, $v^\star(x) + 2\gamma \le T(x) - x$ for all $x \in (x_0 - \gamma, x_0]$. Then by strict convexity of J, the function

$$v_2(x) = v^\star(x) + \gamma \mathbb{1}_{\{x \in (x_0 - \gamma, x_0]\}}$$

verifies that $J(v_2) < J(v^\star)$. Furthermore for γ small enough, we have that $C_x(v_2) < h$ for all $x \in (x_0 - \gamma, x_0)$, as ρ_0 is continuous and $toll_{v^\star}$ is decreasing so $C_x(v^\star) < C_{x_\delta}(v^\star)$ for δ small enough. Therefore we have that v_2 is a better solution to the problem.

If $v^\star(x_0) \ge T(x_0) - x_0$, then by continuity of T we have that for $\gamma > 0$ small enough, $v^\star(x) + 2\gamma \ge T(x) - x$, for all $x \in (x_0, x_0 + \gamma]$. As previously we can find a better solution $v_2(x) = v^\star(x) - \gamma \mathbb{1}_{\{x \in (x_0, x_0 + \gamma]\}}$ to the problem which contradicts the fact that $v^\star$ is the minimizer. $\qquad\square$

The next proposition states that at the points where $v^\star$ doesn't saturate the constraint, $v^\star$ is equal to the unconstrained transport $T - \mathrm{Id}$.

Proposition 4.5. *If there exist $y \in \mathrm{Supp}\,(\rho_0)$ such that $C_y(v^\star) < h$, then we have $v^\star(y) = T(y) - y$.*

Proof. Suppose $\exists y \in \mathrm{Supp}\,(\rho_0)$ such that $C_y(v^\star) < h$ and $v^\star(y) \ne T(y) - y$. Define $g_\epsilon(x) = \mathbb{1}_{\{x \in (y - \epsilon, y + \epsilon)\}} \epsilon^3 \exp(-\frac{1}{\epsilon^2 - (x - y)^2} + \frac{1}{\epsilon^2})$. Then there exist $\epsilon \ne 0 \in \mathbb{R}$ and $\delta > 0$ such that $\forall x \in (y - \delta, y + \delta)$ we have $|v^\star(x) + g_\epsilon(x) - (T(x) - x)| < |v^\star(x) - (T(x) - x)|$ and $C_x(v^\star + g_\epsilon) < h$, since g_ϵ introduces a vanishingly small bump at a suitable location. By strict convexity of J we have that $J(v^\star) > J(v^\star + g_\epsilon)$ which contradicts the optimality of $v^\star$. $\qquad\square$

We can now deduce some regularity of the function $v^\star$.

Corollary 4.1. *The optimal solution $v^\star$ of (4.9) is C^1 almost everywhere.*

Proof. As T is C^1, then $v^\star$ is also C^1 at points y that lie in the interior of the closed set

$\{y \in \text{Supp}(\rho_0) \mid v^\star(y) = T(y) - y\}$. Otherwise if for some y it holds that $v^\star(y) \neq T(y) - y$, then $\exists \delta > 0$ such that $\forall x \in (y - \delta, y + \delta)$, $v^\star(x) \neq T(x) - x$ which implies by Proposition 4.5 that $C_x(v^\star) = h$. Solve the ordinary differential equation

$$
\begin{cases}
\partial_x v(x) = \frac{v(x)^2 \rho_0(x) - h v(x)}{h(x_0 - x)} & \text{for } y - \delta \leq x \leq y_1 \\
v(y + \delta) = v^\star(y + \delta)
\end{cases}
\tag{4.14}
$$

for $v(x)$. It can be shown that the function v is well defined by establishing existence and uniqueness of the solution to (4.14) using the Cauchy-Lipschitz theorem and inherent boundedness. Indeed, if

$$
v(x) > \frac{h}{\inf\{\rho_0(y) \mid y \in \text{Supp}(\rho_0)\}},
$$

then $\partial_x v(x) > 0$, and so v is decreasing with decreasing value of its argument on a small interval $[x - \epsilon, x]$, and if

$$
0 < v(x) < \frac{h}{\sup\{\rho_0(y) \mid y \in \text{Supp}(\rho_0)\}},
$$

then $\partial_x v(x) < 0$, and so v is increasing (again with decreasing value of its argument) on a small interval $[x - \epsilon, x]$. As $v^\star(y + \delta) > 0$, and $v \mapsto \frac{v^2 \rho_0(x) - h v}{h(x_0 - x)}$ is Lipschitz on any compact set, we can apply the Cauchy-Lipschitz theorem to establish existence and uniqueness. From the definition of v, it follows that $C_x(v) = h$, and therefore v has the same flux as $v^\star$. By uniqueness, v which is C^1 on $[y - \delta, y_1]$, is optimal, i.e., $v = v^\star$. Finally, as $v^\star$ is C^1 on the interior of the set $\{y \in \text{Supp}(\rho_0) \mid v^\star(y) = T(y) - y\}$ and is also C^1 on the set $\{y \in \text{Supp}(\rho_0) \mid v^\star(y) \neq T(y) - y\}$, we deduce that $v^\star$ is C^1 almost everywhere as the boundary of those two sets is at most countable. $\qquad \square$

Now that we have established that $v^\star$ is C^1 a.e., we can write the constraint (4.10) for

functions $v \in C^1(\mathrm{Supp}\,(\rho_0), \mathbb{R})$ as: for $x \in \mathrm{Supp}\,(\rho_0)$, a.e.

$$C_x(v) = \frac{v(x)\rho_0(x)}{1 + \frac{x_0 - x}{v(x)}\partial_x v(x)} \leq h. \tag{4.15}$$

For $v : \mathrm{Supp}\,(\rho_0) \to \mathbb{R}$, define

$$J(v) = \int_{\mathbb{R}} \left(v(x)(x_0 - x) + \frac{(T(x) - x_0)^2}{1 - \frac{x_0 - x}{v(x)}} \right) \rho_0(x) dx.$$

We can then rewrite Problem 4.1 in the present case where ρ_0 and ρ_1 are continuous, have bounded convex support and are bounded from below on the interior of their support, as follows.

Problem 4.3. *Consider*

$$\min_{v \in V} J(v) \tag{4.16}$$

over a class V of functions $v : \mathrm{Supp}\,(\rho_0) \to \mathbb{R}$, that are C^1 a.e. and are such that

i) the map $x \mapsto \frac{x_0 - x}{v(x)}$ is decreasing and bounded between 0 and 1

ii) v verifies condition (4.15) a.e.

To solve Problem 4.3, we define velocity fields v on all of $\mathbb{R}$, even outside $\mathrm{Supp}\,(\rho_0)$, as this suitably defined *prolongation* of v will be conveniently expressed as a solution of a differential equation. To this end, we note that the constraint (4.15) can be alternatively expressed in the form

$$C_x^{\mathrm{alt}}(v) := \frac{v(x)^2 \rho_0(x) - h(x_0 - x)\partial_x v(x)}{v(x)} \leq h. \tag{4.17}$$

This alternative formulation applies even for points x where $\rho_0 = 0$, and will help define the sought prolongation for $v^\star$.

Let us first prolong on all of $\mathbb{R}$ the optimal transport map between ρ_0 and ρ_1. To this end, define $\alpha_0 = \inf \text{Supp}(\rho_0), \beta_0 = \sup \text{Supp}(\rho_0), \alpha_1 = \inf \text{Supp}(\rho_1), \beta_1 = \sup \text{Supp}(\rho_1)$, and set

$$
T^+(x) = \begin{cases} T(x) & \text{when } x \in \text{Supp}(\rho_0) \\ \beta_1 + x - \beta_0 & \text{when } x \geq \beta_0 \\ \alpha_1 + x - \alpha_0 & \text{when } x \leq \alpha_0. \end{cases}
$$

Let $\gamma_0, \gamma_1 \in \mathbb{R}$ be the uniquely defined points such that $\frac{x_0 - \alpha_0}{v^\star(\alpha_0)} = \frac{x_0 - \gamma_0}{T^+(\gamma_0) - \gamma_0}$ and $\frac{x_0 - \beta_0}{v^\star(\beta_0)} = \frac{x_0 - \gamma_1}{T^+(\gamma_1) - \gamma_1}$. The point γ_1 is the point that, when transported by $T - \text{Id}$, crosses the toll at the same time β_0 crosses the toll when being transported by $v^\star$. Note that we have $\gamma_0 \leq \alpha_0$ and $\gamma_1 \geq \beta_0$. We also prolong $v^\star$ on the whole $\mathbb{R}$ as

$$
v^{\star +}(x) = \begin{cases} T(x) - x & \text{when } x \leq \gamma_0 \text{ or } x \geq \beta_0 \\ v^\star(\alpha_0)\frac{x_0 - x}{x_0 - \alpha_0} & \text{when } \gamma_0 \leq x \leq \alpha_0 \\ v^\star(x) & \text{when } x \in \text{Supp}(\rho_0) \\ v^\star(\beta_0)\frac{x_0 - x}{x_0 - \beta_0} & \text{when } \beta_0 \leq x \leq \gamma_1 \end{cases}
$$

For notational simplicity, in the sequel, we suppress the labeling on $T^+, v^{\star +}$ and use T, $v^\star$ instead for the prolonged versions as well. To build $v^\star$, we first establish that on the points where $T - \text{Id}$ doesn't satisfy the constraint, $v^\star$ actually saturates the constraint.

As an immediate consequence of Proposition 4.5, we have the following lemma:

Lemma 4.1. *For all $x \in \text{Supp}(\rho_0)$ such that $C_x(T - \text{Id}) > h$ we have $C_x(v^\star) = h$.*

We next characterize a leading segment of the distribution corresponding to points with velocity faster than that of the optimal unconstrained transport. It is essential that the leading edge "speeds up" to allow the trailing portion to pass through and meet the time constraint. Specifically, we show that $v^\star$ is greater than $T - \text{Id}$ at the points to the right of points where $T - \text{Id}$ doesn't satisfy the constraint.

Lemma 4.2. *For $x_1 = \sup\{x \in \text{Supp}\,(\rho_0) \mid C_x(T-\text{Id}) > h\}$ and $y_1 = \sup\{x \in \mathbb{R} \mid C_x^{\text{alt}}(\text{v}^\star) = h\}$ we have that $\forall x \in (x_1, y_1)$, $\text{v}^\star(x) \geq T(x) - x$.*

Proof. First note that $y_1 \geq x_1$ by Lemma 4.1. Suppose that

$$\{\text{v}^\star(x) < T(x) - x\} \cap (x_1, y_1) \neq \emptyset$$

and let $a = \sup\{x \in (x_1, y_1) \mid \text{v}^\star(x) < T(x) - x\}$. We consider separately the two cases $\rho_0(a) = 0$ and $\rho_0(a) > 0$ below:

i) If $\rho_0(a) = 0$ then $\forall x \geq a$, $\rho_0(x) = 0$, so

$$\text{v}^\star(x) = \frac{T(y_1) - y_1}{y_1 - x_0}(x - x_0),$$

as $C_x^{\text{alt}}(\text{v}^\star) = h$ for all $x \in [a, y_1]$. Furthermore, $T(a) - a = \beta_1 + a - \beta_0 - a = T(y_1) - y_1$ and $T(a) - a = \text{v}^\star(a)$ so necessarily $a = y_1$ and $\text{Supp}\,(\rho_0) = [\alpha_0, y_1]$. Then $\exists z \in (x_1, y_1)$ such that, $\rho_0(z) > 0$, $\text{v}^\star(z) < T(z) - z$ and $\partial_x \text{v}^\star(z) > T'(z) - 1$.

ii) If $\rho_0(a) > 0$, then by convexity of $\text{Supp}\,(\rho_0)$ we also have existence of that $z \in (x_1, y_1)$ with the same properties. In both cases we have

$$\frac{\text{v}^\star(z)\rho_0(z)}{1 + \frac{x_0 - z}{\text{v}^\star(z)}\partial_x \text{v}^\star(z)} < \frac{(T(z) - z)\rho_0(z)}{1 + \frac{x_0 - z}{T(z) - z}(T'(z) - 1)} \leq h$$

which contradicts the definition of y_1. $\qquad\square$

The following lemma states that if $\text{v}^\star$ saturates the constraint on a maximal interval (i.e., such that, the points just outside do not saturate the constraint), then either $\text{v}^\star = T - \text{Id}$ throughout, or it is strictly greater than $T - \text{Id}$ on a portion of the interval and strictly less than $T - \text{Id}$ on another portion of the interval. This property is inherited by the convexity of the cost.

Lemma 4.3. *For $[a, b] \subset \{x \in \mathbb{R} \mid C_x^{\mathrm{alt}}(\mathrm{v}^\star) = h\}$ with $[a, b]$ of maximal size, $\exists x \in [a, b]$ such that $\mathrm{v}^\star(x) > T(x) - x$ if and only if $\exists y \in [a, b]$ such that $\mathrm{v}^\star(y) < T(y) - y$.*

Proof. Suppose that $\forall x \in [a, b]$ we have $\mathrm{v}^\star(x) \geq T(x) - x$ and we don't have equality on the whole interval. Define

$$\Psi_a(\epsilon) = \int_a^b \left((x_0 - x)(\mathrm{v}^\star(x) + \epsilon) + \frac{(T(x) - x_0)^2}{1 - \frac{x_0 - x}{\mathrm{v}^\star(x) + \epsilon}} \right) \rho_0(x) dx.$$

Then we have $\partial_x \Psi_a(0) = \int_a^b (x_0 - x)(1 - \frac{(T(x) - x_0)^2}{(\mathrm{v}^\star(x) - (x_0 - x))^2}) \rho_0(x) dx > 0$. Let be $c < a$ such that $\partial_x \Psi_c(0) > 0$ and $\exists \delta > 0$ with $\frac{\mathrm{v}^\star(c)^2 \rho_0(c) - h \mathrm{v}^\star(c)}{h(x_0 - c)} - \partial_x \mathrm{v}^\star(c) = -\delta$. Then there exist $d \in (c, a)$ such that $\partial_x \Psi_d(0) > 0$ and $\exists \delta > 0$ with $\frac{\mathrm{v}^\star(d)^2 \rho_0(d) - h \mathrm{v}^\star(d)}{h(x_0 - d)} - \partial_x \mathrm{v}^\star(d) = -\delta/2$. Let us define k_ϵ as the function solving the ODE

$$\begin{cases} \partial_x k_\epsilon(x) = -\partial_x \mathrm{v}^\star(x) + \frac{(\mathrm{v}^\star(x) + k_\epsilon)^2 \rho_0(x) - h(\mathrm{v}^\star(x) + k_\epsilon(x))}{h(x_0 - x)} & \text{for } x \leq d, \\ k_\epsilon(d) = -\epsilon. \end{cases}$$

Then for $\epsilon > 0$ small enough we have $\partial_x k_\epsilon(x) < -\delta/4$, $\forall x \in (c, d)$. Therefore for $\epsilon > 0$ small enough $\exists y \in (c, d)$ such that $k_\epsilon(y) = 0$. Define

$$\mathrm{v}_\epsilon(x) = \begin{cases} \mathrm{v}^\star(x) & \text{if } x \notin (y, b), \\ \mathrm{v}^\star(x) - \epsilon & \text{if } x \in (d, b), \\ \mathrm{v}^\star(x) + k_\epsilon(x) & \text{if } x \in (y, d]. \end{cases}$$

Then for $\epsilon > 0$ small enough, v_ϵ verifies the constraint and $J(\mathrm{v}_\epsilon) < J(\mathrm{v}^\star)$. Using the same method we can prove that having $\mathrm{v}^\star(x) \leq T(x) - x$ for all $x \in [a, b]$ is impossible. $\square$

We are now in a position to build explicitly $\mathrm{v}^\star$ using the lemmas. The process of building

v* consists of determining its value successively on intervals $[z_{y_i}, y_i]$ and $[y_{i+1}, z_{y_i}]$, with

$$\ldots > y_i > z_{y_i} > y_{i+1} > z_{y_{i+1}} > \ldots$$

such that $v^\star(x) \neq T(x) - x$ for $x \in [z_{y_i}, y_i]$ a.e., while $v^\star(x) = T(x) - x$ on the complement where $x \notin \bigcup_i [z_{y_i}, y_i]$. By Proposition 4.5 we know that $C_x^{\text{alt}}(v^\star) = h$ on intervals $[z_{y_i}, y_i]$, a fact that will help us determine $v^\star$ and the succession of points that define these intervals.

We explain the process in Figures 4.2-4.4 with an example. This example presents a situation where the behavior of the corresponding optimal solution $v^\star$ is characterized by two distinct intervals $[z_{y_i}, y_i]$ $i = 1, 2$, where the constraint saturates. Thus, for this example, we identify three intervals of interest, $[z_{y_2}, y_2]$, $[y_2, z_{y_1}]$, and $[z_{y_1}, y_1]$. In the first and the last, the constraint saturates, whereas in the middle interval it does not. We proceed by working our way from right to left, always assuming that $\text{Supp}\,(\rho_0)$ is to the left of the toll, as in the figures.

In general, the process begins by first computing the optimal transport map T, without involving the constraint. Then, we identify x_1 as the rightmost point where the throughput hits the limit set at x_0. Naturally, if the optimal transport map satisfies the throughput constraint, then it is the optimal map and specifies v^* throughout. Assuming that x_1 is finite, then a search to the right of x_1, that we explain later on, identifies y_1 as the rightmost point where v needs to be adjusted so as to abide by the throughput constraint while minimizing the transportation cost. In the example depicted in Fig. 4.2, y_1 is shown located to the right of β_0 (= the supremum of the support of ρ_0), though this is not always the case, and depends on the terminal distribution ρ_1 via the optimization problem that specifies y_1. We choose to explain this case, where y_1 is to the right of β_0 so as to highlight that this is indeed possible.

Continuing on with our specific example, for the interval $[z_{y_1}, y_1]$, we have $v^\star = v_{y_1}$, with v_y defined in equation (4.19) explained below, which ensures that $C_x^{\text{alt}}(v) = h$. Then,
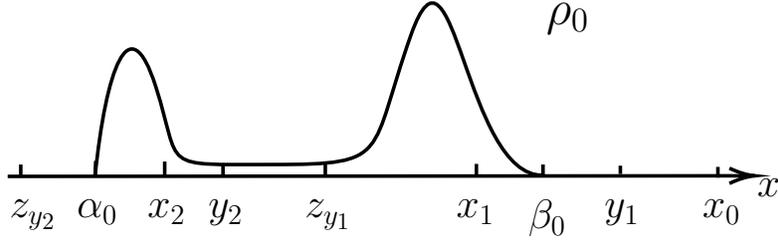
Figure 4.2: Density $\rho_0(x)$ vs. x
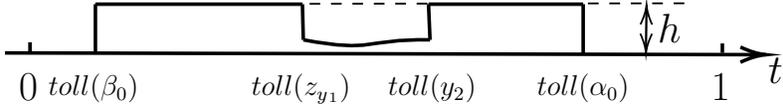
Figure 4.3: Flux $\rho_t(x_0)v_t(x_0)$ at crossing.
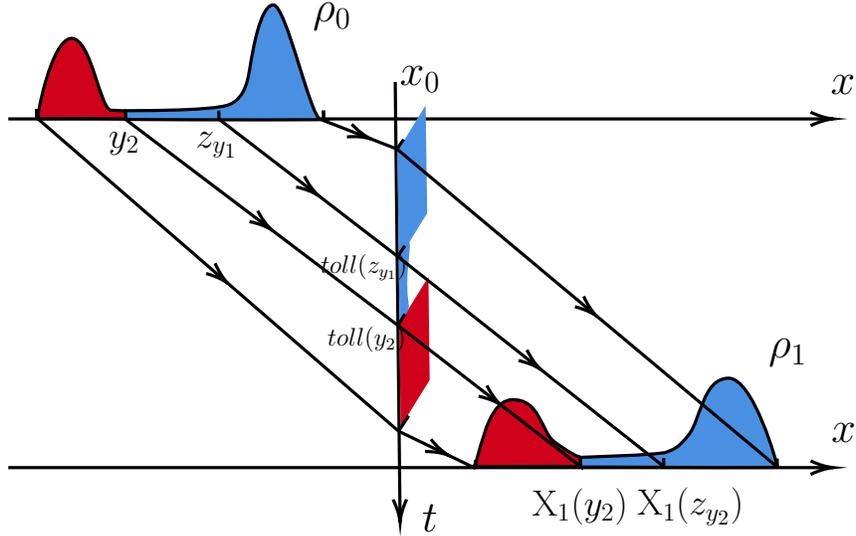
Figure 4.4: Illustration of the flow through the toll. The middle segment $[y_2, z_{y_1}]$ transports through the toll unimpeded by the constraint towards the final destination, via the optimal transport map T, designed for unconstrained transport; each point in this interval maintains the same velocity before and after the toll. In contrast, the segments to the left and right, $[z_{y_2}, y_2]$ and $[z_{y_1}, y_1]$, respectively, are adjusted accordingly so as to saturate the constraint. The exact position of their respective end points (that may even be outside the support of ρ_0, as a matter of computational simplicity, in which case they correspond to zero density) are computed via the solution of an optimization problem and depend on the terminal distribution ρ_1 as well.

on $[y_2, z_{y_1}]$ we have once again that the velocity is specified by the "unconstrained" optimal map T, i.e, that $v^\star = T - \mathrm{Id}$, and so $C_x^{\mathrm{alt}}(v) = C_x^{\mathrm{alt}}(T - \mathrm{Id})$. Finally on $[z_{y_2}, y_2]$, we have $v^\star = v_{y_2}$ as $C_x^{\mathrm{alt}}(v) = h$. Note that in this specific example where $y_1 \geq \beta_0$ and $z_{y_2} \leq \alpha_0$, we

have $\forall x \in [z_{y_2}, \alpha_0]$, $toll(x) = toll(\alpha_0)$ and $\forall x \in [\beta_0, y_1]$, $toll(x) = toll(\beta_0)$.

We now detail how to build explicitly $\mathrm{v}^\star$ in the general case. As noted, if $T - \mathrm{Id}$ verifies the constraint throughout, which can now be explicitly stated as in (4.15), then $\mathrm{v}^* = T - \mathrm{Id}$ is the optimal solution. Otherwise define $x_1 = \sup\{x \in \mathrm{Supp}\,(\rho_0) \mid C_x(T - \mathrm{Id}) > h\}$, and thereby we determine $y_1 \in [x_1, x_0]$ (cf. Lemma 4.1) such that

$$ y_1 = \sup\{x \in \mathbb{R} \mid C_x^{\mathrm{alt}}(\mathrm{v}^\star) = h\}. \tag{4.18} $$

For any $y \in \mathbb{R}$ with $x_1 \le y < x_0$, define the velocity $v_y(x)$ as the solution of the differential equation

$$ \begin{cases} \partial_x \mathrm{v}_y(x) = \frac{\mathrm{v}_y(x)^2 \rho_0(x) - h\mathrm{v}_y(x)}{h(x_0 - x)} & \text{for } x \le y. \\ \mathrm{v}_y(y) = T(y) - y \end{cases} \tag{4.19} $$

Note that this equation is solved backwards, starting from a terminal condition at y. This value for the velocity ensures that the transport will saturate the constraint to left of y (i.e., $C_x^{\mathrm{alt}}(\mathrm{v}_y) = h$ will hold for $x \le y$). The functional form of $v_y(x)$ will be used next to identify the first interval $[z_{y_1}, y_1]$, where the velocity will depart form that of the unconstrained transport T, via solving a suitable optimization problem to determine y_1. Since we know that the equality $C_x^{\mathrm{alt}}(\mathrm{v}^\star) = h$ will be true on a certain interval $[z_{y_1}, y_1]$, on that interval we will have $\mathrm{v}^\star = v_{y_1}$.

Let $w_y^{x_1} = \inf\{x \le x_1 \mid \forall s \in (x, x_1), \mathrm{v}_y(s) \ge T(s) - s\}$ (well defined by Lemma 4.2) and $z_y^{x_1} = \inf\{x \le w_y^{x_1} \mid \forall s \in (x, w_y^{x_1}), \; \mathrm{v}_y(s) < T(s) - s\}$. Then we have that $\mathrm{v}^\star(x) = \mathrm{v}_{y_1}(x)$, $\forall x \in (z_{y_1}^{x_1}, x_1)$ by Lemma 4.3 and Proposition 4.5.

We now determine y_1 by solving a suitable optimization problem. For $x \le y < x_0$,

define

$$J_x(y) = \int_0^1 \int_{\mathbb{R}} \Bigg((T(s) - s) \mathbb{1}\{s \notin (z_y^x, y)\}$$
$$+ ((x_0 - s)\mathrm{v}_y(s) + \frac{(T(s) - x_0)^2}{1 - \frac{x_0 - s}{\mathrm{v}_y(s)}}) \mathbb{1}\{s \in (z_y^x, y)\} \Bigg) \rho_0(s) ds dt.$$

We have $J_x(y) = J(\mathrm{v}_y^+)$ for the function v_y^+ such that $\mathrm{v}_y^+ = \mathrm{v}_y$ on $[z_y^x, y]$ and $\mathrm{v}_y^+ = T - \mathrm{Id}$ on $\mathbb{R} \setminus [z_y^x, y]$. Then the first step of the building process of $\mathrm{v}^\star$ is to find y_1 solution of

$$y_1 = \arg\min_{y \geq x_1} J_{x_1}(y).$$

Such a y_1 is well defined as J_{x_1} is continuous on $[x_1, x_0]$. Once y_1 has been determined, we define $x_2 = \sup\{x < z_{y_1} \mid C_x(T - \mathrm{Id}) > h\}$. If x_2 is not defined then

$$\mathrm{v}^\star(x) = \begin{cases} \mathrm{v}_{y_1}(x) & \text{if } x \in (z_1, y_1), \\ \\ T(x) - x & \text{if } x \notin (z_1, y_1), \end{cases}$$

otherwise we start again the same process to determine y_2 as

$$y_2 = \arg\min_{y > x_2} J_{x_2}(y).$$

If $y_2 < z_{y_1}$, it suggests that there is an interval $[y_2, z_{y_1}]$ where the transport follows the unconstrained map T, and we continue in the same way.

However, it is possible that the condition $y_i \leq z_{y_{i-1}}$ fails at some point, for some $i \geq 2$. In that case, intervals where the velocity departs from being $T(x) - x$, will merge. For instance, if we obtain $y_i > z_{y_{i-1}}$ then as $(y, y') \mapsto J_{x_{i-1}}(y) + J_{x_i}(y')$ is convex on $\{(y, y') \mid y' \leq z_y\}$, it means that $C_x^{\mathrm{alt}}(\mathrm{v}^\star) = h$, $\forall x \in (x_i, x_{i-1})$ and therefore we have to start the optimization

again and determine y_{i-1} as

$$y_{i-1} = \arg\min_{y > x_{i-1}} J_{x_i}(y).$$

If we obtain a value $y_{i-1} > z_{y_{i-2}}$, we reset x_{i-1} as being equal to x_i and, once again, we have to redetermine

$$y_{i-2} = \arg\min_{y > x_{i-2}} J_{x_{i-1}}(y).$$

Otherwise, i.e., if we obtain a value $y_{i-1} \leq z_{y_{i-2}}$, we reset x_i as $x_i = \sup\{x < z_{y_{i-1}} \mid C_x(T - \mathrm{Id}) > h\}$ for this updated value y_{i-1}. Once again, if x_i is well defined we continue the process by finding

$$y_i = \arg\min_{y > x_i} J_{x_i}(y).$$

We continue this iterative process until $\mathrm{v}^\star$ is defined on all of the support of ρ_0. We finally remark that

$$E := \bigcup_{i=1}^{n} (z_{y_i}, y_i) = \{x \in \mathbb{R} \mid \exists \delta > 0, \forall y \in (x - \delta, x) \cup (x, x + \delta), T(x) - x \neq \mathrm{v}^\star(x)\}.$$

Note that we have that $n \in \mathbb{N} \cup \{+\infty\}$, so the process doesn't necessarily terminate. If one absolutely wants the process to terminate, they have to be careful to the oscillations of $C_x(T - \mathrm{Id})$ around the value h. Indeed, if the process doesn't terminate, it implies that $x_i = \sup\{x < z_{y_{i-1}} \mid C_x(T - \mathrm{Id}) > h\}$ always exists $\forall i$, so the function $x \mapsto C_x(T - \mathrm{Id})$ oscillates indefinitely around h as x is moving backward. Supposing that the densities ρ_0 and ρ_1 are Lipschitz, then T has Lipschitz derivative so $x \mapsto C_x(T - \mathrm{Id})$ is also Lipschitz. This implies that the oscillations around h become smaller and smaller (in size) so it suffices to lower the value of h of any $\epsilon > 0$ to avoid the infinite oscillations.

We summarize our conclusions on the shape of $\mathrm{v}^\star$ in the following statement.

Theorem 4.2. *The solution* $v^\star$ *of Problem 4.3 satisfies:*

a) *For* $x \in (z_{y_i}, y_i)$, *we have* $v^\star(x) = v_{y_i}(x)$ *(defined in (4.19)) and* $C_x^{\mathrm{alt}}(v^\star) = h$.

b) *There exists* $w_i \in (z_{y_i}, y_i)$ *such that* $\forall y \in [w_i, y_i)$, $v^\star(y) \geq T(y) - y$ *and* $\forall y \in (z_{y_i}, w_i]$, $v^\star(y) \leq T(y) - y$.

c) $\forall x \notin E$, $v^\star(x) = T(x) - x$.

d) *The building process of* $v^\star$ *consists in solving iteratively* $y_i = \underset{y > x_i}{\arg\min}\, J_{x_i}(y)$ *and if* $y_j > z_{y_{j-1}}$, *then starting again by setting* $y_{j-1} = \underset{y > x_{j-1}}{\arg\min}\, J_{x_j}(y)$. *The process may contain an infinite number of steps.*

4.5   Numerical example

We provide an example to highlight the departure of the optimal transport plan through a toll with a bound on the flux, from the ideal unconstrained transport T. The example we have selected is basic, with uniform probability densities $\rho_0(x) = \mathbb{1}\{x \in [0,1]\}$, $\rho_1(x) = \mathbb{1}\{x \in [2,3]\}$, and a toll at $x_0 = 3/2$ with a bound h on the flux, with $1 < h \leq 2$. The stringent constraint on the flux, that necessitates varying velocities so as to redistribute the mass flow as it traverses the toll, is clearly seen in the succession of distributions $X_{t\sharp\rho_0}$ displayed in Fig. 4.5. Evidently, these readily contrast with the unconstrained transport that pushes forward ρ_0 with constant speed giving $\rho_t(x) = \rho_0(x - 2t)$.

Specifically, with the flux-constraint in place, we obtain that the optimal transport is effected by

$$
X_t(x) = \begin{cases} x + t\mathrm{v}(x) & \text{for} \quad t \leq toll(x) = \frac{3/2 - x}{\mathrm{v}(x)}, \\[2em] 3/2 + (t - toll(x))g(x) & \text{for} \quad t \geq toll(x). \end{cases}
$$

Then, the constraint (4.15) gives that v solves the ODE

$$\frac{\mathrm{v}(x)}{1 + \frac{3/2 - x}{\mathrm{v}(x)} \partial_x \mathrm{v}(x)} = h.$$

It follows that $\mathrm{v}(x) = \frac{h(2x-3)}{2x-3+\alpha}$ for a certain value $\alpha \in \mathbb{R}$. Using the fact that the optimal solution must be symmetric in time ($\mathrm{v}(x) = g(1-x)$) and that $g(x) = \frac{x+0.5}{1-toll(x)}$, we finally obtain that $\mathrm{v}(x) = \frac{h(2x-3)}{2x-1-h}$. Snapshots of the flow along the path from ρ_0 to ρ_1 are depicted in Figure 4.5.
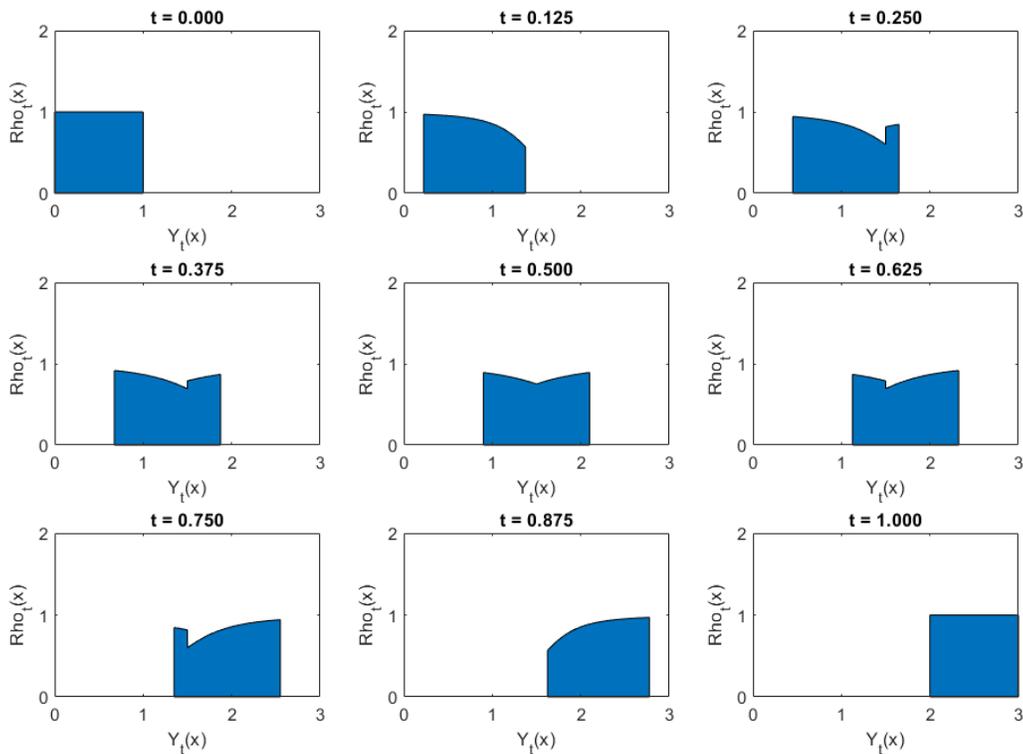
Figure 4.5: Example of transporting a uniform distribution through a constriction (with $h = 1.5$) to a similar uniform terminal distribution. While the optimal unconstrained transport will preserve the shape of the marginals at each time t, the flux constraint necessitates an optimal velocity that changes with x, stretching the leading edge of the distribution as it approaches the toll. Note that the snapshots of the transported distributions $X_{t\sharp\rho_0}$ "squeeze" while crossing the toll, and that the flow is symmetric with time.

4.6 Discussion and conclusion

We have presented theory for the most basic optimal transport problem in $\mathbb{R}$, through a constriction where a throughput constraint is imposed. We modeled the formulation after the standard Monge-Kantorovich optimal transport with a quadratic cost. We have shown that an optimal transport exists and is unique under general assumptions. Under some suitable assumption on the densities to be transported to one another, we have shown explicitly how to construct the transport plan. Moreover, we have highlighted natural properties of the transport plan.

More generally, in the case where ρ_0 and ρ_1 are densities on $\mathbb{R}^d$ and that all the trajectories have to pass through a single point $x_0 \in \mathbb{R}^d$, we can readily extend the result presented as follows. For $\lambda_{\alpha S^{d-1}}$ the Lebesgue measure on the sphere of radius α and center x_0, define

$$\nu^0(\alpha) = \int_{\alpha S^{d-1}} \rho_0(x) d\lambda_{\alpha S^{d-1}}(x)$$

and ν^1 the same way. Then the problem in $\mathbb{R}^d$ is equivalent to solving the problem in dimension 1 between the measure ν^0, ν^1 defined as $\nu^0(x) = \mathbb{1}\{x < 0\}\nu^0(-x)$ and $\nu^1(x) = \mathbb{1}\{x > 0\}\nu^1(x)$.

A significant departure from the current setting arises in the case of multiple tolls, or of a continuum of tolls, where the flux-rate is bounded on a curve, surface, etc. The case where a sequence of tolls, possibly even zero-dimensional (points), where mass has to flow through all in succession, is of particular interest in engineering applications. Indeed, in the modern information age, knowledge of obstructions "down the road" can undoubtedly be used to optimize transportation cost upstream. On the other hand, the paradigm of multiple alternative tolls that one can choose to cross, is expected to have a more combinatorial flavor. Lastly, one could generalize the problem presented in this paper to transport of densities in dimension d, with a flux constraint on a measurable set with respect to the p-

dimensional Hausdorff measure $\mathcal{H}^p$ (with $p \leq d$). For instance, an analogous flux constraint on a measurable set $A \subset \mathbb{R}^d$ with $0 < \mathcal{H}^p(A) < \infty$ can be cast as: $\forall B \subset A$ measurable with $\mathcal{H}^p(B) > 0$ and $t \in (0, 1)$

$$\forall \alpha_1, \alpha_2 \in \mathbb{R}, \quad \frac{\mathcal{H}^p(A)}{\mathcal{H}^p(B)} \int \mathbb{1}_{\{\exists \tau \in (t+\alpha_1, t+\alpha_2) \mid X_\tau(x) \in B\}} \rho_0(x) dx \leq h|\alpha_2 - \alpha_1|.$$

The proof of existence and uniqueness of a solution should follow using similar arguments. However, to completely characterise the behavior of the solution as in the simpler case treated herein, is expected to be considerably more challenging; one would need a finer description of how the mass distributes while traversing the toll.

Transport problems with a throughput restriction are quite natural in a variety of scientific disciplines. Of course, transportation through tolls on highways represents perhaps the most rudimentary paradigm in an engineering setting. Likewise, throughput through servers with a throughput bound is common in queuing systems. A continuum theory as envisioned herein, in higher dimension and with multiple serial tolls, may produce useful practical insights. Finally, while fluid flow, passing through constrictions or porous media, though not directly abiding by the rigid setting of bounded throughput, could provide an idealized pertinent model in certain situations. Evidently, for an accurate model for fluid past constrictions, besides distinguishing between compressible and incompressible, throughput must be dictated by pressure, which in turn may be introduced in a suitable cost functional to be optimized for a further broadening of the general program.

Bibliographical notes

In recent years, the Monge-Kantorovich theory of Optimal Mass Transport has impacted a wide range of mathematical and scientific disciplines from probability theory to geophysics and from thermodynamics to machine learning [47, 32, 3, 104, 105]. Indeed, the Monge-

Kantorovich paradigm of transporting one distribution to another, by seeking to minimize a suitable cost functional, has proved enabling in many ways. It gave rise to a class of control problems [25, 24], underlies variational principles in physics [58, 79], provided natural regularization penalties in inverse problems [12], led to new identification techniques in data science [54, 83], in graphical models [40], and was linked to large deviations in probability theory [68, 23].

Historically, the Monge-Kantorovich theory proved especially relevant in economics when physical commodities were the object to be transported–a fact that contributed to L. Kantorovich receiving the Noble prize. Extensions that pertain to physical constraints along the transport naturally were soon brought up. For instance, moment-type constraints, have been considered in [88, Section 4.6.3] and, more recently, far generalized in [37]. Congestion being a significant impediment to transport has also drawn the attention of theorist and practitioners alike. For instance, besides optimizing for a transportation, considerations of an added path-dependent cost to alleviate congestion has been considered in [18], see also [91, Section 4] for a comprehensive study of this research direction (compared with the Monge-Kantorovich OMT where the cost $c(x, y)$ is only dependent on source and destination, the traffic congestion in [18] is captured by the cost $c_{\pi, p}(x, y)$ for transportation strategy (π, p) with π is the admissible couplings and p characterizes the portion of mass through certain path). Along a different direction, constraints have been introduced for probability densities as part of the optimization problem. Such bounds can capture the capacity of the transportation medium and the optimized among admissible couplings with given marginals μ, ν is now defined as

$$\Pi = \left\{ \pi \in \mathcal{P}(\mathbb{R}^n, \mathbb{R}^n) \mid \int_y \pi = \mu(dx), \int_x \pi = \nu(dy), \ \pi \leq \bar{\pi} \right\}$$

with the capacity constraint $\bar{\pi} \in \mathcal{P}(\mathbb{R}^n, \mathbb{R}^n)$ so that $\pi \leq \bar{\pi}$ almost everywhere which have

been studied in [62], or dynamical flow constraints

$$\int_{S_1 \times S_2} f_k(x, y)\pi(d(x, y)) \leq b_k, \ \forall k = 1, \ldots, m.$$

given the measurable maximal density $h : \Omega \to [0, \infty)$ and Ω is a manifold with boundary, as in [48].

Chapter 5

Kantorovich transport with toll

Simplicity is the ultimate sophistication.

– Leonardo da Vinci

The key idea in the present work for dealing with flow-rate constraints is to seek a distribution for the times of crossing the toll-stations, suitably bounded to ensure meeting the constraints. Our choice for a cost to be minimized is the mean-square of the velocity. This choice effectively orders the flow, in that particles do not overtake each other, and can be expressed as a convex functional of the optimization variables. Moreover, this cost can be conveniently decomposed into a sum of costs, each engaging only space and timing variables that pertain to segments of the transport path. We ensure flux constraints by imposing bounds on the variables representing the time of toll-crossing. Under a fairly general setting our formulation amounts to a multi-marginal OMT problem with the time-crossing marginals as unknown and constrained parameters of the problem.

Below, in Section 5.1 we discuss the formulation of Monge-Kantorovich transport through tolls with flow-rate constraint. In Section 5.3 we specialize to measures with support on $\mathbb{R}$ and we highlight the nature of solutions with representative examples. In Section 5.4, we

discuss two generalizations of the basic problem, the first concerns partial transport through tolls where mass needs only to clear certain tolls on the way to the respective destination, while the second brings in a new dimension to the transport problem by considering arrival and departure times in the formulation.

5.1 Problem formulation

We consider the classical optimal mass transport problem with a quadratic cost, albeit with a flow-rate constraint on the flow across tolls at specified locations. Specifically, we consider distributions μ, ν, that are assumed throughout as being probability measures (i.e., with mass normalized to 1), and seek a transportation plan from one to the other. The "starting" distribution μ represents the mass of a single commodity that needs to be transported accordingly and matches the demand that is specified by ν. Along the way, the mass has to clear tolls abiding by corresponding constraints on throughput.

Let us first review the classical Monge-Kantorovich optimal mass transport with a quadratic cost (on $\mathbb{R}^n$). The Monge formulation of transport seeks a transportation map $T : x \mapsto y = T(x)$ so as to minimize the transportation cost functional

$$J(T) = \int_{\mathbb{R}^n} \|T(x) - x\|^2 \mu(dx),$$

over the choice of T, subject to the transportation map T pushing the starting to the target distribution, i.e., μ to ν. We denote this by writing $T_\sharp \mu = \nu$. As it is standard, the notation $\sharp$ denotes the "push forward," meaning that for any Borel set $S \subset \mathbb{R}^n$, $\mu(T^{-1}(S)) = \nu(S)$.

In general, such a map T may not always exist. The Kantorovich relaxation seeks instead a measure π on the product space $\mathbb{R}^n \times \mathbb{R}^n$ so that the marginals on the two components

coincide with μ, ν, while π minimizes the functional

$$J(\pi) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \|y - x\|^2 \pi(dx, dy).$$

The measure π is referred to as coupling of the two marginal distributions.

Assuming that μ is absolutely continuous with density ρ_0 then the optimal transport map T always exists and is unique [105, Theorem 2.12], and the optimal cost is the minimal of

$$J(\partial_t X) := \int_0^1 \int_{\mathbb{R}^n} \big(\partial_t X_t(x)\big)^2 \rho_0(x) dx dt, \tag{5.1}$$

for $X : [0,1] \times \mathbb{R}^n \to \mathbb{R}^n$ such that[1] $X_{0\sharp}\rho_0 = \rho_0$ and $X_{1\sharp}\rho_0 = \nu$. That is, the minimization is over velocity fields $\partial_t X$, for flows X that correspond the starting and ending marginal densities. It turns out that the optimal flow is effected by

$$X_t^{\text{opt}}(x) = x + t\big(T(x) - x\big),$$

for T the optimal transport map. The minimal value of these functionals is designated as the Wasserstein distance $\mathcal{W}_2^2(\mu, \nu)$ [104].

One observes that, in (5.1), what is actually being minimized is the average kinetic energy (modulo a factor of 2), as the transport takes place over the interval $[0,1]$. In general, having $d\mu = \rho_0 dx$ and carrying out the transport over the interval $[0, t_f]$,

$$
\begin{aligned}
\mathcal{W}_2^2(\mu, \nu) &= J(\partial_t X_t^{\text{opt}}) \\
&= \int_0^{t_f} t_f \int_{\mathbb{R}^n} \big(\partial_t X_t^{\text{opt}}(x)\big)^2 \rho_0(x) dx dt \\
&= \int_0^{t_f} \int_{\mathbb{R}^n} \frac{\|T(x) - x\|^2}{t_f} \rho_0(x) dx dt
\end{aligned}
$$

---

[1] Following a common abuse of notation, for simplicity, we write interchangeably the density ρ_0 and measure μ, and allow the notation to be understood from the context.

since $\partial_t X_t^{\text{opt}}(x) = (T(x) - x)/t_f$, recovering $J(T)$ for the optimal transport map. Most importantly, *the average kinetic energy can also be written*, for the relaxed Kantorovich formulation, as

$$\mathcal{W}_2^2(\mu, \nu) = \int_0^{t_f} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{\|y - x\|^2}{t_f} \pi(dx, dy) dt.$$

The derivation is immediate since $\int_0^{t_f} dt = t_f$. The expression helps highlight the average kinetic energy as the minimal of the convex cost $\|y - x\|^2/t_f$, when transporting mass from x to y over a time interval of duration t_f.

We are now in a position to *formulate the analogue of the Monge-Kantorovich problem for the case where the mass needs to clear tolls*, initially a single toll, at specified locations and with a bound on the flow-rate.

We first assume that a single toll is located at ξ and that mass flowing through cannot exceed a given flow-rate r. Thus, the flow-rate through the toll must satisfy

$$\sigma(t) \leq r.$$

The flow-rate represents the mass density over time as the mass is transported through the toll, i.e., $\sigma(t)dt$ represents mass that clears the toll in the interval $[t, t + dt]$.

The insight that allows us to formulate optimal transport through toll(s) as a multi-marginal optimal transport problem is to view the *as-yet-undermined* mass density $\sigma(t)$ as a time marginal distribution, together with the specified spatial initial and final marginals $\mu(dx)$ and $\nu(dy)$. Thereby, the Kantorovich formulation of the problem seeks a coupling $\pi(dx, dy, dt)$ between the two given marginals $\mu(dx)$, $\nu(dy)$, and the sought marginal $\sigma(t)dt$, that specifies the portion of mass at $[x, x + dx]$, heading towards $[y, y + dy]$, is transported through the toll in the interval $[t, t + dt]$.

In light of the coupling between a starting point x at time $\tau = 0$, location of the toll ξ

that mass is to be transported at time $\tau = t$, and terminal destination y to arrive at $\tau = t_f$, the average kinetic energy[2] over the segment $\tau \in [0, t]$ is minimized when particles travel at constant velocity $(\xi - x)/t$, and thus, equal to

$$\int_0^t \|(\xi - x)/t\|^2 d\tau = \|\xi - x\|^2/t.$$

Likewise, the average kinetic energy over the remaining interval $[t, t_f]$ is $\|y - \xi\|^2/(t_f - t)$.

Thus, we arrive at the formulation of our first problem.

Problem 5.1. *Given probability measures μ, ν and $r > 0$, determine a probability measure $\pi(dx, dy, dt)$ on $\mathbb{R}^n \times \mathbb{R}^n \times [0, t_f]$ that minimizes*

$$\iiint_{x,y,t} \left(\frac{\|\xi - x\|^2}{t} + \frac{\|y - \xi\|^2}{t_f - t} \right) \pi(dx, dy, dt), \tag{5.2}$$

subject to the marginals $\iint_{y,t} \pi(dx, dy, dt) = \mu(dx)$ and $\iint_{x,t} \pi(dx, dy, dt) = \nu(dy)$, and the flow-rate constraint $\iint_{x,y} \pi(dx, dy, dt) \leq r dt$.

Throughout, notation as in $\iint_{x,y} \pi(dx, dy, dt)$, indicates integration over the space of the variables that are subscribed to the integral.

Note that the above formulation allows mass that is initially concentrated x and is heading towards the same terminal destination y, to split and transport over different paths clearing the toll at different times t, as prescribed by the coupling, in order to abide by the imposed bound on flow-rate. This will necessarily be the case when μ assigns finite mass at a point (contains Dirac deltas or singular part, in higher dimensions).

A somewhat more ambitious scheduling may require optimizing the average kinetic energy, from a starting measure μ to the terminal ν, clearing two or multiple tolls in succession.

---

[2]The average kinetic energy is often referred to as *action integral* in physics.

For the case of two tolls, define $\pi(dx, dy, dt_1, dt_2)$ as the coupling of mass at x heading to y that clears the tolls at times t_1, t_2, respectively, where $0 \leq t_1 \leq t_2 \leq t_f$. The coupling, as before, is a probability measure that specifies the respective amount of mass that is transported as prescribed.

The problem with two or multiple tolls is analogous to Problem 5.1. For instance, in the case of two tolls, the cost c^{ξ_1, ξ_2} to be minimized over the choice of admissible couplings is

$$\iiiint_{x,y,t_1,t_2} \left(\frac{\|\xi_1 - x\|^2}{t_1} + \frac{\|\xi_2 - \xi_1\|^2}{t_2 - t_1} + \frac{\|y - \xi_2\|^2}{t_f - t_2} \right) \pi.$$

Admissibility of π amounts to consistency with the problem data, i.e., it amounts to satisfying the usual marginal constraints on x and y, as well as the flow-rate constraints $\iiint_{x,y,t_2} \pi(dx, dy, dt_1, dt_2) \leq r_1 dt_1$ and $\iiint_{x,y,t_1} \pi(dx, dy, dt_1, dt_2) \leq r_2 dt_2$.

We remark that flow-rate constraints can be time-varying without any significant overhead in the difficulty of the problem. Specifically, in the condition $\iint_{x,y} \pi(dx, dy, dt) \leq r dt$ the flow-rate bound can be specified by a time-dependent density $r(t)$ that regulates permissible *throughput* at different times. Evidently, the bound could also be a measure, but this is deemed of minimal practical relevance and not followed here.

5.2   Existence and uniqueness of the solution

Our first technical result establishes in a straightforward manner existence of solutions.

Proposition 5.1 (Existence). *Provided $rt_f > 1$, Problem 5.1 admits a (minimizing) solution π.*

Proof. Denoting with $\sigma(t)$ the mass density that crosses the toll at time t, as before, the

transport of the total mass through the toll over the interval $[0, t_f]$ subject to $\sigma(t) \leq r$, requires that

$$1 = \int_0^{t_f} \sigma(t)dt \leq \int_0^{t_f} r dt = r t_f,$$

Thus, $r t_f \geq 1$ is a necessary condition[3]. The space of admissible measures π in Problem 5.1, i.e.,

$$\Pi = \left\{ \pi \in \mathcal{P}\big(\mathbb{R}^n, \mathbb{R}^n, [0, t_f]\big) \mid \iint_{y,t} \pi(dx, dy, dt) = \mu(dx), \right.$$
$$\left. \iint_{x,t} \pi(dx, dy, dt) = \nu(dy), \iint_{x,y} \pi(dx, dy, dt) \leq r dt \right\}$$

is non-empty. It is also compact for the weak topology as it is tight and closed for the narrow convergence. Indeed, the set of coupling measures between a finite number of probability measures is tight [91, Theorem 1.4], and each of the three constraints in Π is closed for the narrow convergence. Then, as the cost function (integrand in (5.2)) is lower semi-continuous, we have the existence of a minimizer. To see that Π is non-empty, we postulate a uniform distribution on the crossing times, $u = (1/t_f)dt$, and couplings $\tilde{\pi}_{xt}, \tilde{\pi}_{ty}$ that are consistent with the marginals (μ, u) and (u, ν), respectively. Then, the existence of an element $\tilde{\pi} \in \Pi$ with marginals

$$\int_y \tilde{\pi}(dx, dy, dt) = \tilde{\pi}_{xt}, \text{ and } \int_x \tilde{\pi}(dx, dy, dt) = \tilde{\pi}_{yt},$$

is guaranteed by the gluing lemma [104, page 11]. □

Note that the convexity of the cost is not used in the proposition. We next consider

---

[3]The case $r t_f = 1$ is only feasible in the non-generic situation where $\max \operatorname{Support}(\mu) = 0 = \min \operatorname{Support}(\nu)$. In general, when e.g., $\max \operatorname{Supp}(\mu) < 0$, the "rightmost" mass on the support of μ must be transported with infinite velocity so as to allow $\sigma(t) = 1/t_f$ over $[0, t_f]$ for the total mass to have enough time to be transported through. Such limiting cases are non-physical, leading to diverging transportation cost, and thereby excluded.

whether the optimal solution is unique. We first discuss the special case where $n = 1$. Moreover, for this case where locations on the underlying space can be ordered (e.g., from left to right), we assume that the toll sits between the two distributions μ, ν, specifically that the support of μ is to the left of ξ and that the support of ν is to the right. Without loss of generality, we let $\xi = 0$ and we thus consider the following problem.

Problem 1′ (Simplification). *We consider probability measures μ, ν on $\mathbb{R}$, with support on $[-M, 0)$ and $(0, M]$ for sufficiently large M, respectively. Let $r, t_f > 0$ such that $rt_f > 1$. Determine a probability measure $\pi(dx, dy, dt)$ as the minimizer of*

$$\iiint_{x,y,t} \left(\frac{x^2}{t} + \frac{y^2}{t_f - t} \right) \pi(dx, dy, dt)$$

subject to the flow-rate constraint $\iint_{x,y} \pi(dx, dy, dt) \leq rdt$, and the marginals $\iint_{y,t} \pi(dx, dy, dt) = \mu(dx)$ and $\iint_{x,t} \pi(dx, dy, dt) = \nu(dy)$.

We begin with a technical lemma that establishes a correspondence between the time and location of mass as this is transported past the toll. A schematic that exemplifies the statement of the lemma below is shown in Fig. 5.2.

Lemma 5.1 (Monotonicity). *Let $c_{xt} := \frac{x^2}{t}$, with $(x, t) \in [-M, 0) \times (0, t_f]$, and μ, σ measures with support on $[-M, 0)$ and $(0, t_f]$, respectively, with $\sigma(t)dt$ absolutely continuous with respect to the Lebesgue measure. The minimizer of the Kantorovich problem*

$$\min_{\pi} \iint c_{xt}\pi,$$

where π represents a coupling of the two marginals $\mu(dx), \sigma(t)dt$, is unique with support on the graph of a non-increasing function $T^{\mathcal{X}}(t)$.

Proof. We first observe that for any two pairs $x, x' \in [-M, 0)$ and $t, t' \in (0, t_f]$ for which

$0 > x > x' \geq -M$ and $t_f \geq t > t' > 0$, it holds that

$$\frac{x^2}{t} + \frac{(x')^2}{t'} > \frac{x^2}{t'} + \frac{(x')^2}{t}.$$

The ordering in this inequality characterizes c_{xt} as being quasi-monotone, in the language of [15], see also [88, Section 3.1].

It follows from [105, Theorem 2.18, and Remark 2.19] that the optimal coupling π exists and is given by the monotone rearrangement of μ, σ, that is, for a suitable function $T^{\mathcal{X}}(t)$

$$\int_{T^{\mathcal{X}}(t)}^{0} \mu(dx) = \int_0^t \sigma(s)ds.$$

Since $T^{\mathcal{X}}(t) < 0$, it is non-increasing (and is constant on time-intervals that correspond to possible Dirac components of μ). This completes the proof. $\qquad\square$

For similar reasons, the cost c_{yt} (with $t \in (0, t_f]$ and $y \in (0, M]$, and M as in Problem $\mathbf{1}'$) is quasi-monotone. Hence, once again, for a suitable function $T^{\mathcal{Y}}(t)$,

$$\int_{T^{\mathcal{Y}}(t)}^{M} \nu(dy) = \int_0^t \sigma(s)ds,$$

for $T^{\mathcal{Y}}(t) > 0$, so that $T^{\mathcal{Y}}(t)$ is non-increasing. In light of the monotonicity of $T^{\mathcal{X}}(t)$ and $T^{\mathcal{Y}}(t)$, we establish the following proposition.

Proposition 5.2 (Uniqueness). *Under the assumptions of Problem $\mathbf{1}'$ the minimizer is unique. Moreover, there are functions $T^{\mathcal{X}}(t), T^{\mathcal{Y}}(t)$ are monotonically non-increasing such that*

$$\pi = (T^{\mathcal{X}}, T^{\mathcal{Y}}, \mathrm{Id})_{\sharp\sigma},$$

where $\mathrm{Id}(t) = t$ is the identity map and $\sigma(t)dt$ is an absolutely continuous measure on $[0, t_f]$

98

with $\sigma(t) \leq r$.

Proof. Let π be a minimizer as claimed in Proposition 5.1, and let $\pi_{xt} := \int_y \pi$, $\pi_{yt} := \int_x \pi$, and $\sigma := \iint_{xy} \pi$. Since

$$\iiint_{x,y,t} c\pi = \iint_{xt} c_{xt}\pi_{xt} + \iint_{yt} c_{yt}\pi_{yt},$$

π_{xt} is a minimizer of $\iint_{xt} c_{xt}\pi_{xt}$, and the same applies to π_{yt}. If this was not the case, there would be couplings $\hat{\pi}_{xt}, \hat{\pi}_{yt}$ with strictly lower costs $\iint_{xt} c_{xt}\hat{\pi}_{xt}$, and $\iint_{yt} c_{yt}\hat{\pi}_{yt}$. These two couplings share the same marginal on the t-axis, namely,

$$\int_x \hat{\pi}_{xt}(dx, dt) = \int_y \hat{\pi}_{yt}(dy, dt) = \sigma(t)dt.$$

Then, by the gluing lemma [104, page 11], there is a coupling $\hat{\pi}$ on $\mathbb{R} \times \mathbb{R} \times [0, t_f]$ that agrees with the given marginals and has a lower cost. Thus, both π_{xt}, π_{yt} are optimal for the respective problems. We next argue that σ is unique, and therefore, the conclusion follows by Lemma 5.1.

To establish the uniqueness of the density on the t-axis, assume that there are two different minimizers π^a and π^b to start with. Then, as above, each gives rise to a density on the t-axis, $\sigma^a(t)$ and $\sigma^b(t)$, respectively, as well as corresponding marginals and maps $(T^{\mathcal{X},a}, T^{\mathcal{Y},a})$ and $(T^{\mathcal{X},b}, T^{\mathcal{Y},b})$. Since both $\sigma^a(t)$ and $\sigma^b(t)$ satisfy the constraint of being $\leq r$, so does any convex linear combination, say $\bar{\sigma} = \frac{1}{2}\sigma^a + \frac{1}{2}\sigma^b$, and the convex combination $\bar{\pi} = \frac{1}{2}\pi^a + \frac{1}{2}\pi^b$ is also optimal. But then, the marginal

$$\bar{\pi}_{xt} = (T^{\mathcal{X},a}, \mathrm{Id})_{\sharp \frac{1}{2}\sigma^a} + (\hat{T}^{\mathcal{X},b}, \mathrm{Id})_{\sharp \frac{1}{2}\sigma^b}$$

is supported on a set that is not the graph of a function[4], unless of course $T^{\mathcal{X}} = T^{\mathcal{X},a} = T^{\mathcal{X},b}$.

---

[4]Note that $\bar{\pi}_{xt}$ satisfies the marginal constraints since, by virtue of $T^{\mathcal{X}}_{\sharp\sigma^i} = \mu$, for $i \in \{a, b\}$, $T^{\mathcal{X}}_{\sharp\frac{1}{2}\sigma^a} + T^{\mathcal{X}}_{\sharp\frac{1}{2}\sigma^b} =$

But if $\bar{\pi}_{xt}$ is not supported on a graph of a function, there exists a more "economical" coupling with strictly lower cost, obtained by monotone rearrangement of $\bar{\pi}_{xt}$. A similar statement holds for $\bar{\pi}_{yt}$. This contradicts the nonuniqueness and completes the proof.[5] □

In the setting of Problem 5.1, when $n > 1$, the transport cost of all mass that resides at a distance $d = \|x - \xi\|$ is the same. Thus, the problem to transport through the toll cannot distinguish equidistant points from the toll.
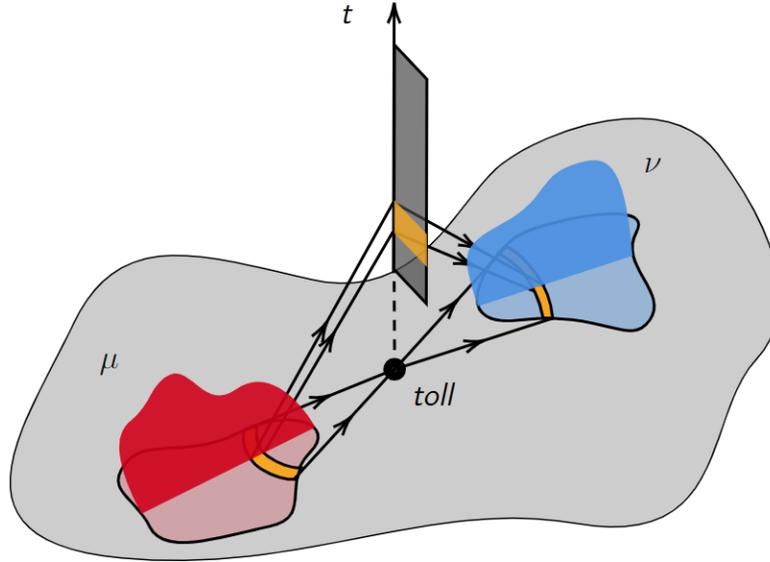
Figure 5.1: Illustration of the optimal mass when $n = 2$.

In this case where the distributions, μ, ν sit in $\mathbb{R}^n$ for $n \geq 1$, Proposition 5.1 and 5.2 can be readily extended as the cost in Problem 5.1 only depends on the distance of the points to ξ. Indeed, the problem is equivalent to solving the 1-dimensional problem between $\widetilde{\mu}, \widetilde{\nu}$

μ. A similar statement holds for the coupling $\bar{\pi}_{yt}$.

[5] A more concise proof by general Monge condition: The cost function $c(x, y, t)$ has to satisfy the *generalized Monge condition* [88, Page 24] in the case $\pi \in \mathcal{P}([-M, 0), (0, -M], [0, t_f])$, i.e., any two of the arguments in $c(x, y, t)$ satisfies the Monge condition. Additionally, the transport has to follow the Monge protocol so that $c^{x,t} := c(x, y^*, t)$ is quasi-monotone and $c^{x,y} := c(x, y, t^*)$ is quasi-antitone when y^*, t^* are fixed. We can easily check that the generalized Monge condition and the monotonicity are satisfied for the cost $c(x, y, t)$.

the measures such that for all $A \subset \mathbb{R}$ measurable set[6],

$$\widetilde{\mu}(A) = \int_{\mathbb{R}^n} \mathbb{1}_{\{||x-\xi|| \in A\}} \mu(dx).$$

Therefore we have existence of a unique solution $\widetilde{\pi} = (T^{\mathcal{X}}, T^{\mathcal{Y}}, \mathrm{Id})_{\sharp\sigma}$ to the 1-dimensional problem which gives rise to solutions π to the n-dimensional problem in the following way: For $\mu_{\tilde{x}}, \nu_{\tilde{y}}$ the disintegrated measures [20] such that

$$\mu(dx) = \int \mu_{\tilde{x}}(dx) \widetilde{\mu}(d\tilde{x})$$

$$\nu(dy) = \int \nu_{\tilde{y}}(dy) \widetilde{\nu}(d\tilde{y}),$$

the solutions π will be of the form

$$\pi(dx, dy) = \int \pi_t(dx, dy) \sigma(t) dt$$

for π_t any coupling[7] measure between $\mu_{T^{\mathcal{X}}(t)}$ and $\nu_{T^{\mathcal{Y}}(t)}$.

Remark 5.1 (Generalization). *A further interesting generalization is when the toll through which the mass is to be transported is no longer a point but a set $\mathcal{T} \subset \mathbb{R}^n$, typically a curve or a manifold of higher dimension, with a Hausdorff measure $\mathcal{H}(dz)$ integrating to 1. The transport problem for such a situation becomes one of minimizing*

$$\inf_{\pi \in \Pi} \int \left(\frac{||x - z||^2}{t} + \frac{||y - z||^2}{t_f - t} \right) \pi(dx, dy, dt, dz),$$

---

[6]The notation $\mathbb{1}_S(x)$, or $\mathbb{1}_S$ for simplicity, signifies the indicator function that takes the value 1 when $x \in S$ and zero otherwise.

[7]The coupling between $\mu_{T^{\mathcal{X}}(t)}$ and $\nu_{T^{\mathcal{Y}}(t)}$ cannot be specified from the problem setting, since it does not affect the cost.

over couplings in

$$\Pi = \left\{ \pi \in \mathcal{P}(\mathbb{R}^n, \mathbb{R}^n, [0, t_f], \mathcal{T}) \mid \iiint_{y,t,z} \pi = \mu(dx), \right.$$

$$\left. \iiint_{x,t,z} \pi = \nu(dy), \iint_{x,y} \pi \le r dt \otimes \mathcal{H}(dz) \right\}.$$

The term to the right of the last inequality, representing the (normalized) Hausdorff measure of $\mathcal{T}$ can be further suitably modified to account for preference/ease of transporting through specific portions of the set $\mathcal{T}$. Physically such a problem may model flow through media, where $\mathcal{T}$ represents porous section that the mass must go through, from source μ to destination ν. Developing theory for this generality is beyond the scope of the current paper. □

5.3 Case studies: transport through tolls in 1D

We now present case studies that help visualize the general scheme for Monge-Kantorovich transport through tolls in $\mathbb{R}$, that is, in dimension 1. Since the formalism in Section 5.1 casts the problem as a multi-marginal one, the coupling with marginal in 1D is already a measure in $\mathbb{R}^3$, with one of the axes the time that mass crosses the toll. The computational aspects and the code using the optimization toolbox-CVX [49] to conduct all the experiments can be found at `https://github.com/dytroshut/OMT-with-Flux-rate-Constraint`.

We discuss four examples that help visualize the effect flow-rate constraints and the nature and support of the transportation coupling π.

Our first example is displayed in Fig. 5.2. The source distribution is Dirac (i.e., concentrated at one point) located at a point $x < 0$, the toll is located at $\xi = 0$, and the terminal distribution is absolutely continuous with respect to the Lebesgue with the density that has support on $\{y \in \mathbb{R} \mid y \ge 0\}$. The figure highlights the maps $T^{\mathcal{X}}$ and $T^{\mathcal{Y}}$ that couple (x, t) and (y, t), respectively, where t denotes the time that mass originally at x crosses the toll on

its way to location y. Both maps are monotonically non-increasing.
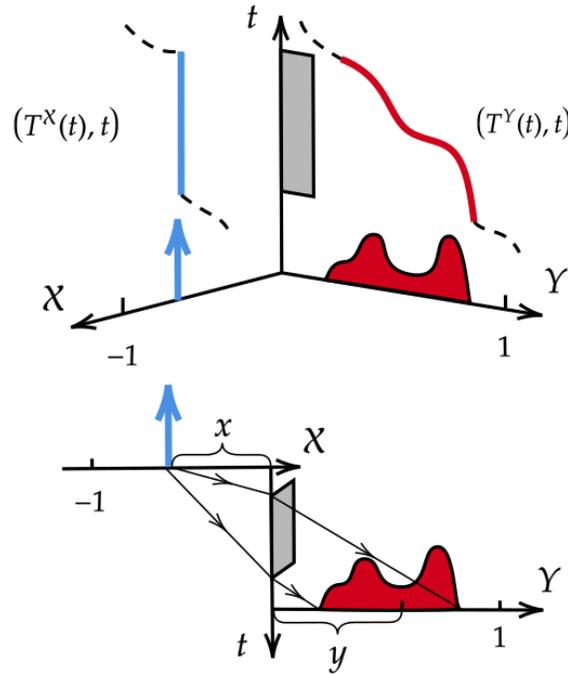
Figure 5.2: Illustration of solution to Problem 5.1: the maps $T^{\mathcal{X}}(t)$, $T^{\mathcal{Y}}(t)$ are monotonically non-increasing, the t-marginal density $\sigma(t)$ is bounded by r.

In our second example, in Fig. 5.3, the marginals are Gaussian mixtures. The flow is visualized via shadowing the paths. Three instances of different bounds on flow-rate are depicted, highlighting how the bound affects the flow and the distribution of crossing times.

(a) $r = \infty$ (no flow-rate constraint)

(b) $r = 2$

(c) $r = 1.2$
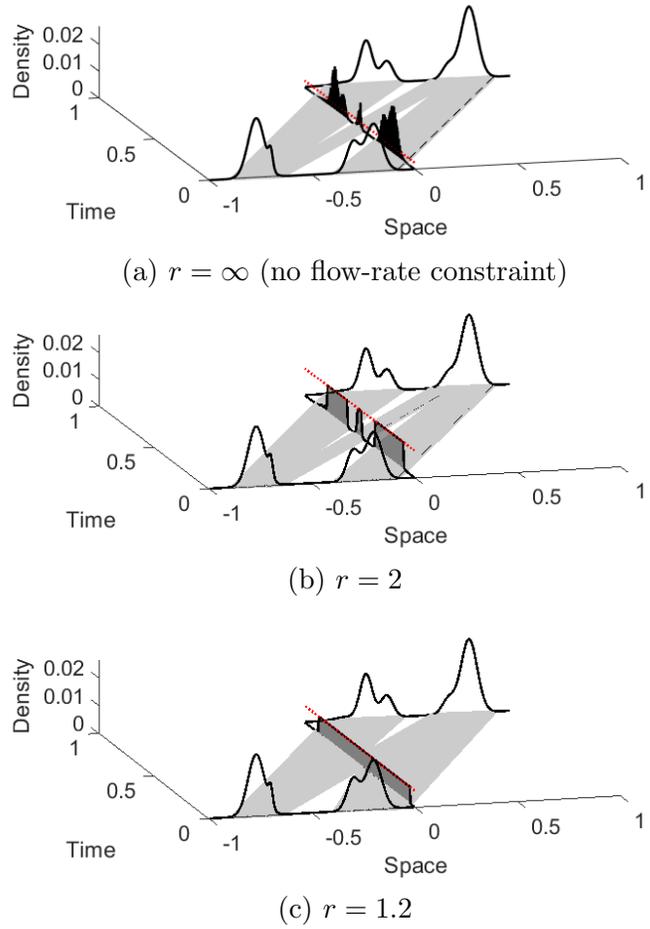
Figure 5.3: Case (a), having no flow-rate constraint at the toll, corresponds to standard Monge-Kantorovich transport with particles moving at constant speeds depending on origin/destination. Cases (b,c) depict the situation where a flow-rate bound at the toll necessitates that mass is transported with different speeds at the two sides of the toll at $\xi = 0$, so as to meet the imposed bound on the t-marginal.

Our third example in Fig. 5.4 helps visualize the coupling in $\mathbb{R}^3$ between (x, y, t), for smooth marginals; the coupling is supported on a curve. In general, the mass on this curve is not uniform and the density is depicted with circles of suitable radius around corresponding points on the curve. Couplings for two different values for acceptable flow-rate are shown $(r \in \{1.5, 3\})$ and it is seen that as r is decreased the t-density tends to become more uniform.
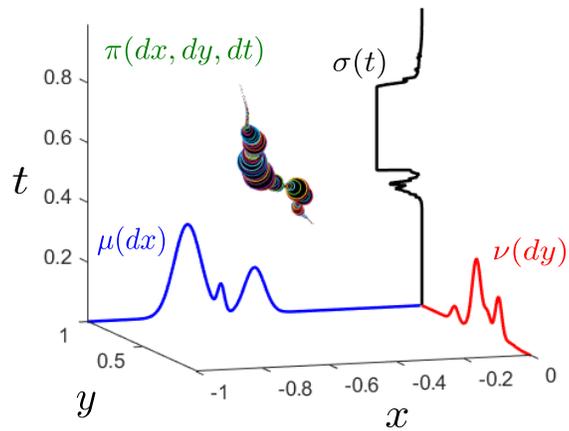
(a) $r = 3$
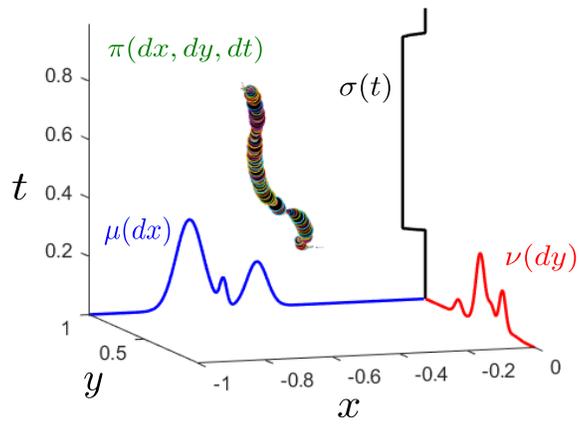
(b) $r = 1.5$

Figure 5.4: Support of the coupling measure $\pi(dx, dy, dt)$ on a curve in $\mathbb{R}^3$; the density is depicted by circles of size proportional to magnitude.

Our fourth example is drawn with a sketch in Fig. 5.5a, and a simulation of transport between two Gaussian mixtures, through two tolls, in Fig. 5.5b. The tolls are positioned at $\xi_1 = -0.4$ and $\xi_2 = 0.4$. The density of the respective times of crossing, t_1 and t_2, are bounded by $r_1 = 1.5$ and $r_2 = 3$, respectively.
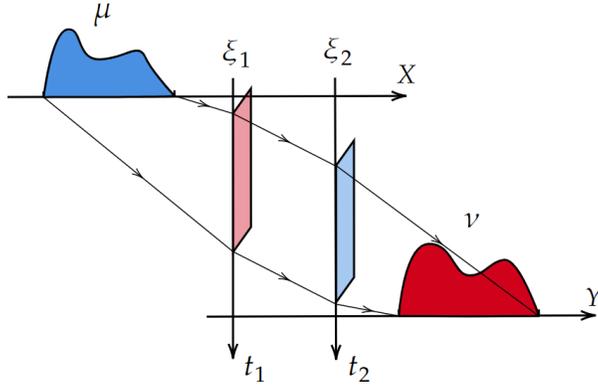
(a) Schematic of transport through two tolls at ξ_1, ξ_2.
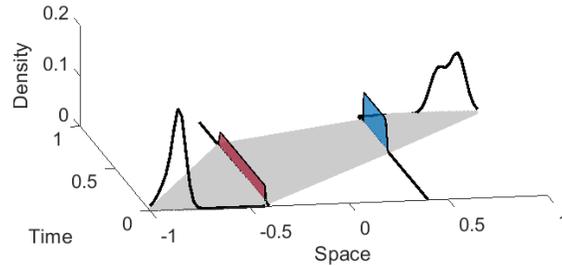
(b) Transport of Gaussian mixtures through two tolls with density-bounds $r_1 = 1.5$ and $r_2 = 3$, at crossing times t_1 and t_2, respectively.

Figure 5.5: Schematic and simulation of 2-toll transport plans.

5.4   Concluding Remarks

The basic idea presented in this paper for dealing with flow-rate constraints has been to introduce a time-variable for when mass transits certain locations. Then, the density of the corresponding marginal distribution, quantifies the amount of mass clearing the toll over a time interval, hence, flow-rate. Such a marginal distribution, as yet to be determined, represents a design parameters to be specified so as to meet flow-rate constraints. Thereby, such problems can be cast in the form of *multi-marginal optimization.*

We note that the present work builds on, and extends our earlier study [99], where under strong regularity assumptions on the marginals for the supply and demand, we developed a Benamou-Brenier approach for Monge transport through a *single* toll. In contrast to this

earlier work, the present formulation allows dealing with more general measures and multiple tolls, and in addition, it casts the problem as a linear program and allows efficient approximation using e.g., entropic regularization as in other timely multi-marginal optimization formulations [51, 53].

We conclude by showing how the basic framework of utilizing marginals to quantify timing information, applies to transportation problems with more complicated structure. Specifically we discuss two cases. First we explain the case where portion of the mass is not constrained to clear the toll, and second, a case of how ordering of arrival and departure times can be incorporated in the same framework.
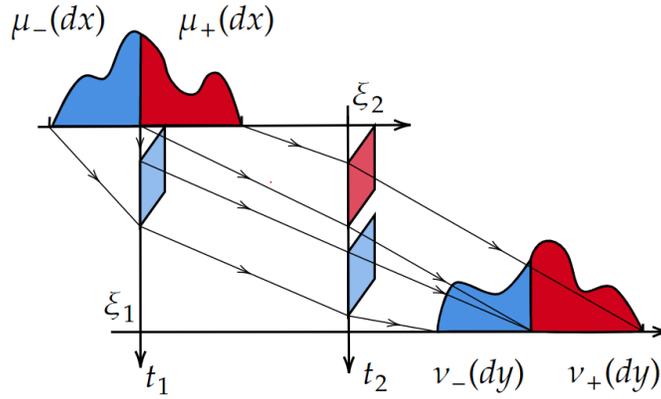
Figure 5.6: Illustration of the optimal mass transport through two successive tolls with separating mass.

For our first case, consider the schematic in Fig. 5.6 where a fraction of a source distribution μ needs to clear two tolls, while the remaining fraction only needs to clear a single toll, due to its relative position with respect to the toll. For instance, in this one-dimensional schematic, the source distribution μ is split by the toll at ξ_1 into $\mu_+ := \mu(x > \xi_1)$ and $\mu_- := \mu(x < \xi_1)$. Accordingly, the target distribution ν needs to be split into $\nu_{+/-}$ corresponding to the two masses, $m_{+/-}$, so that

$$T_{\sharp\mu} = T^{\xi_1,\xi_2}_{\sharp\mu_-} + T^{\xi_2}_{\sharp\mu_+} = \nu_- + \nu_+,$$

107

where T^{ξ_1,ξ_2} transports through two tolls whereas T^{ξ_2} transports through a single toll. With the flow-rate constraints on crossing time marginals, the Kantorovich formulation process is exactly as before, *provided the fractions of corresponding masses can be delineated.* For one-dimensional distributions, specifying the corresponding portions is straightforward. In higher spatial dimensions, when there is no clear separation as in the one-dimensional schematic, the problem of selecting "what fraction of mass needs to clear what toll" is coupled to the optimization problem and has a combinatorial nature.

For our second case, we bring in timing to prioritize departure and arrival, so as to meet objectives and possibly mediate congestion along the flow. To see this, we briefly discuss how to modify the standard Monge-Kantorovich setting in which all particles/mass transport at constant (that depends on the particle) speed along geodesics from source x to destination y according to the McCann flow

$$\rho_t = \big[(1-t)\mathrm{Id}(x) + tT^{\mathcal{Y}}(x)\big]\sharp\mu,$$

over the window $t \in [0,1]$ with transport map $T^{\mathcal{Y}}(x)$. For simplicity we retain arrival time $t_a = 1$, i.e., fixed, and only allow the departure time t_d to vary. The marginal distribution of t_d now represents a design parameter. The formalism, once again, seeks a coupling measure π that satisfies constraints and marginals. For simplicity, we introduce the Monge transport map $T^{\mathcal{Y}}(x)$, to specify the source-destination pairing. Then, the coupling of the variables x, y, t_d gives that $\pi = (\mathrm{Id}, T^{t_d}, T^{\mathcal{Y}})_{\sharp\mu}$. And, if $\tau(t,x)$ denotes the portion of time that a particle at x is "on the move," i.e., $\tau(t,x) = (t - t_d(x))\mathbb{1}_{\{t>t_d(x)\}}$, the McCann's displacement reads

$$\rho_t = \big[(1-\tau(t,x))\mathrm{Id}(x) + \tau(t,x)T^{\mathcal{Y}}(x)\big]\sharp\mu.$$

Figure. 5.7 shows an example where both departure and arrival times are variables (t_d

and t_a, respectively), with marginals selected to minimize a cost functional of the form $\iiiint_{x,y,t_a,t_d} c(x,y,t_a,t_d)\pi(dx,dy,dt_a,dt_d)$, with cost $c(x,y,t_a,t_d) = t_d/x^2 + (y-x)^2 - t_a/y^2$. This choice ensures a natural order in departure and arrival[8].
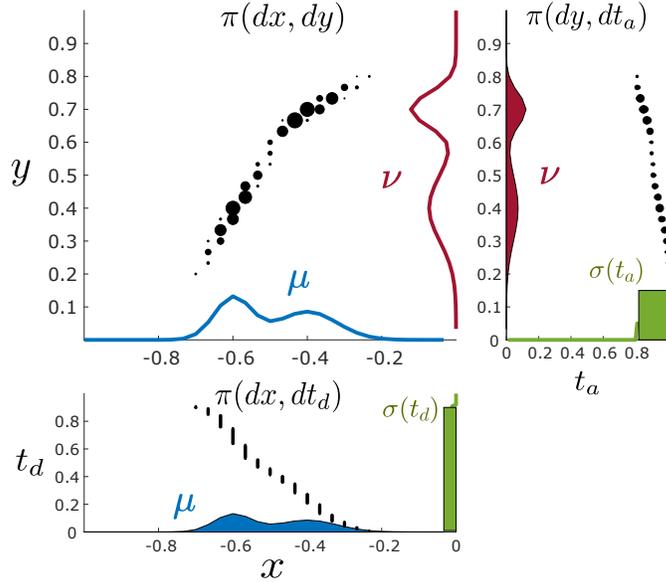
Figure 5.7: In earlier formulations, all particles/agents departed at the same time $t = 0$ and arrived at the same time $t = t_f$. Here, besides meeting flow-rate constraints, we stratify departure and arrival so that particles/agents closer to the toll depart first, and arrive at the most distant target location again first (as it would be natural for a convoy transferring goods). Departure and arrival rate bounds are set to $r_d = 1.1$ and $r_a = 5$, respectively. Coupling measures $\pi(dx,dy)$, $\pi(dx,dt_d)$, and $\pi(dy,dt_a)$ between timing variable are computed; it is seen that these are supported on graphs of maps (showing a monotonic correspondence). For instance, $\pi(dx,dt_d)$ couples $\mu(dx)$ and $\sigma(t_d)$, with $\sigma(t_d) \leq r_d$ being the departure-time marginal. Due to the monotonicity of the cost $c(x,t_d) = t_d/x^2$, the coupling $\pi(dx,dt_d)$ indicates that mass closer to the toll departs earlier, while abiding by the departure rate bound r_d. For the timing of arrival t_a, properties of the coupling $\pi(dy,dt_a)$ are completely analogous.

In closing, we mention that distributed flux constraints may fruitfully capture properties of matter through which transport takes place, e.g., in transport of pollutants through porous media. A generalization of the framework herein to a distributed setting would be desirable and at present open.

---

[8]The agents/particles that are closer to the destination leave first.

Bibliographical notes

Starting at the waning years of the 20th century, contributions by Brenier, McCann, Otto, Gangbo, Evans, Villani, and many others, sparked a new phase of rapid development with significant inroads of OMT as an enabling tool in mathematics and physics [105].

In the present work, we revisit a formulation in our earlier work [99] that aims to limit flow-rate. Specifically, we seek to transport mass through constriction points/tolls abiding by flow-rate bounds. This earlier work proved the existence and uniqueness of flux-limited Monge maps (scheduling maps) that effect transport while minimizing the $\mathcal{W}_2$ Wasserstein length of trajectories. In the present work, we take an alternative view by casting the problem in a Kantorovich-type in the form of multi-marginal optimization. We note that, while multi-marginal problems have been studied for a while [82, 61], issues related to computation for large size problems continue to be of great interest to the present day – we refer to [51, 52, 53, 81] for structured multi-marginal problems and to [83] for the computational aspects of OMT in general.

The formulation can be easily generlized for more general setting. For instance, inspired by [14, 21, 41], we introduce the flow-rate constraint (toll) to the partial optimal mass transport problem (OMT). Compared with the standard optimal mass transport problem, the partial OMT considers transporting a fraction of mass

$$m \in \left(0, \min\{\|\mu\|_1, \|\nu\|_1\} \right]$$

from measure $\mu(x)$ to $\nu(y)$ with the same quadratic cost $c(x, y) = \|x - y\|_2^2$. Noticing that $\|\mu\|_1 \neq \|\nu\|_1$, such a problem can be considered as an unbalance OMT problem.

Chapter 6

Entropic regularization

We already studied the Monge-Kantorovich problem with flow-rate constraint at a specific point, namely, a *toll*. Specifically, the Kantorovich formulation of the problem utilizes a multi-marginal OT framework with upper bound relaxation on the crossing time marginal. Herein, to solve it efficiently, we apply the convex entropic regularization on an equivalent split problem in the sense of the Kullback-Leibler (KL) divergence. A variant of the standard Sinkhorn algorithm (Gluing Sinkhorn) is presented accordingly, and is shown to provide a fast and accurate approximation of the solution.

The outline of this chapter is as follows: In Section 6.1, we revisit the Monge Kantorovich problem with flow-rate constraints and its entropic regularization. In Section 6.2, we approach the entropic regularization with a standard Sinkhorn solver. In Section 6.3, we propose the gluing Sinkhorn specifically for the multi-marginal optimal transport problem with splittable cost. In Section 6.4, we present experimental experience to show the efficiency of our method.

Problem 6.1 (**Kantorovich**)**.** *Given probability measures* μ, ν *and* $r > 0$, *determine a*

probability measure $\pi(dx, dy, dt) \in \Pi$ as the minimizer of

$$\iiint_{x,y,t} \left(\frac{\|\xi - x\|^2}{t} + \frac{\|y - \xi\|^2}{t_f - t} \right) \pi(dx, dy, dt) \tag{6.1}$$

over the feasible set $\Pi \in C^0(X, Y, [0, t_f])$ of all the admissible π can be defined as[1]

$$\Pi(\mu, \nu, r) := \{\pi \mid P_x(\pi) = \mu(dx), \ P_y(\pi) = \nu(dy), \ P_t(\pi) \leq rdt\}$$

The formulation enables mass initially concentrated at point x and headed towards a common destination y, to be divided and transported along different paths, clearing the toll at different times t while adhering to the imposed limit on flow rate. The above problem admits an optimizer as stated in [36, Proposition 1].

6.1 Entropic regularization

Noticing that the cost is splittable and has the form

$$c(x, y, t) = \frac{x^2}{t} + \frac{y^2}{1 - t} = c_{xt} + c_{yt}.$$

and correspondingly define the feasible sets $\Pi_{xt} \in C^0(X, [0, 1])$ and $\Pi_{yt} \in C^0(Y, [0, 1])$ of all the admissible two-dimensional marginals $\pi_{xt} := \int_y \pi(dx, dy, dt)$ and $\pi_{yt} := \int_x \pi(dx, dy, dt)$ as

$$\Pi_{xt}(\mu, r) := \{\pi_{xt} \mid P_x(\pi_{xt}) = \mu(dx), \ P_t(\pi_{xt}) \leq rdt\},$$

$$\Pi_{yt}(\nu, r) := \{\pi_{yt} \mid P_y(\pi_{yt}) = \nu(dy), \ P_t(\pi_{yt}) \leq rdt\}.$$

We then propose to solve the equivalent problem to reduce the size of the variable as below

---

[1]Throughout this chapter, we utilize $P_x(\pi_{xy})$ denotes the projection onto x, i.e., $P_x(\pi) = \int_y \pi dx$ and similar for $P_y(\pi)$.

Problem 6.2 (Splittable Cost). *Determine the probability measures* $\pi_{xt} \in \Pi_{xt}$ *and* $\pi_{yt} \in \Pi_{yt}$ *as the minimizers of*

$$\iint_{x,t} c_{xt}\pi_{xt}(dx, dt) + \iint_{yt} c_{yt}\pi_{y,t}(dy, dt) \tag{6.2}$$

Proposition 6.1 (Equivalence). *Problem 5.1 and Problem 6.2 are equivalent and admit the same unique minimizer.*

Proof. Assuming π_{xt} and π_{yt} are the minimizer of Problem 6.2 and $\pi(dx, dy, dt)$ is the minimizer of Problem 5.1, the equivalence between Problem 5.1 and Problem 6.2 can be proved by showing fact that

$$\int_x \pi_{xt} = \int_y \pi_{yt} = \iint_{xy} \pi(dx, dy, dt) = \sigma(dt),$$

where $\sigma(t)$ is the unique optimal time marginal for both of the problems.

If this was not the case, then there exists another pair of minimizer π'_{xt} and π'_{yt} has the optimal $\sigma'(dt)$ with the same cost as $\iint_{xt} c_{xt}\pi_{xt} + \iint_{yt} c_{yt}\pi_{yt}$, and moreover,

$$\int_x \pi'_{xt} = \int_y \pi'_{yt} = \sigma'(dt)$$

By the gluing lemma [104, page 11], there is a coupling $\pi'(dx, dy, dt)$ with $\int_y \pi'(dx, dy, dt) = \pi'_{xt}$ and $\int_x \pi'(dx, dy, dt) = \pi'_{yt}$, which shares the same cost as $\pi(dx, dy, dt)$. This is a contradiction of the uniqueness of the solution of Problem 5.1, as proved in [36, Proposition 2] □

Adding an entropic regularization to the original objective of the optimal transport problem has garnered attention due to its advantages in numerical approximation of the solution [83, Section 4]. The regularized OT in the continuous framework is often referred

to as the static Schrödinger problem since it was initially considered by Schrödinger, which gains increasing interest in the field of control [23].

Considering the objective with an additional convex penalty term, which is the entropy of $\pi(dx, dy)$ as

$$\mathcal{H}(\pi) = \iint_{x,y} -\pi \log(\pi) dx dy,$$

the entropic regularization of (5.2) has the form

$$\iint_{x,y} \left(\frac{1}{\epsilon} c\pi - \mathcal{H}(\pi)\right) dx dy = \iint_{x,y} \left(\frac{1}{\epsilon} c\pi + \pi \log(\pi)\right) dx dy$$
$$= \iint_{x,y} \pi \log(\frac{\pi}{\hat{\pi}}) dx dy$$

where $\hat{\pi} = \exp(-c(x, y)/\epsilon)$ is the prior, and

$$\mathrm{KL}(\pi \| \hat{\pi}) := \iint_{x,y} \pi(dx, dy) \log(\frac{\pi(dx, dy)}{\hat{\pi}(dx, dy)}) dx dy$$

is the Kullback-Leibler divergence between $\pi(dx, dy)$ and $\hat{\pi}(dx, dy)$. Therefore, the entropic regularization of Problem 5.1 now reads

Problem 1'. *Given probability measures μ, ν and $r > 0$, determine a probability measure $\pi(dx, dy, dt) \in \Pi$ as the minimizer of*

$$\iiint_{x,y,t} \left(c - \epsilon \log(\pi)\right) \pi(dx, dy, dt) = \iiint_{x,y,t} \mathrm{KL}(\pi \| \hat{\pi}) \tag{6.3}$$

with $\hat{\pi} = \exp(-c(x, y, t)/\epsilon)$ as the proir.

Similarly, for Problem 6.2, we have

Problem 2'. *Given μ, ν and r, determine the probability measures $\pi_{xt} \in \Pi_{xt}$ and $\pi_{yt} \in \Pi_{yt}$*

as the minimizers of

$$\iint_{x,t} (c_{xt}\pi_{xt} + \mathcal{H}(\pi_{xt})) + \iint_{yt} (c_{yt}\pi_{y,t} + \mathcal{H}(\pi_{xt}))$$

$$= \iint_{x,t} \mathrm{KL}(\pi_{xt}\|\hat{\pi}_{xt}) + \iint_{y,t} \mathrm{KL}(\pi_{yt}\|\hat{\pi}_{yt}) \tag{6.4}$$

where $\hat{\pi}_{xt} = \exp(-\frac{c_{xt}}{\epsilon})$ and $\hat{\pi}_{yt} = \exp(-\frac{c_{yt}}{\epsilon})$.

By splitting the cost and proposing Problem 2′, we obtain a twist using Schrödinger bridge and KL divergence which splits the coupling and interpret Problem 2′ as a two-sided Schrödinger Bridge problem.

Proposition 6.2. *The minimizer $\pi(x, y, t)$ can be factorized by the marginals $\pi_{xt} := \int_y \pi(x, y, t)$ and $\pi_{yt} := \int_x \pi(x, y, t)$*

$$\pi(x, y, t) = \frac{\pi_{xt}\pi_{yt}}{\sigma(t)}, \tag{6.5}$$

with $\sigma(t) := \iint_{x,y} \pi(x, y, t)$.

Proof. Denote $\mathbb{P}(x|y)$ as the conditional probability of x given y, we have

$$\pi(x, y, t) = \mathbb{P}(x|y, t)\mathbb{P}(y|t)\sigma(t) = \mathbb{P}(x|t)\mathbb{P}(y|t)\sigma(t)$$

$$= \frac{\mathbb{P}(x|t)\mathbb{P}(y|t)\sigma(t)^2}{\sigma(t)} = \frac{\pi_{xt}\pi_{yt}}{\sigma(t)}$$

by the matrix factorization and the result of Proposition 6.1. $\qquad\square$

The factorization we introduce in 6.5 is also discovered and used in [92, 44]. Now we are in the position to prove the equivalence between the two problems above.

Theorem 6.1. *Problem 1′ and Problem 2′ are equivalent.*

Proof. The splittable cost leads to write the prior $\hat{\pi}$ as a production of $\hat{\pi}_{xt}$ and $\hat{\pi}_{yt}$, i.e.,

$$\hat{\pi} = \exp(-\frac{c(x,y,t)}{\epsilon}) = \exp(-\frac{c_{xt}}{\epsilon}) \cdot \exp(-\frac{c_{yt}}{\epsilon}) = \hat{\pi}_{xt}\hat{\pi}_{yt}.$$

Thus, the objective (6.3) becomes

$$\begin{aligned}
&\mathcal{KL}\left(\pi \parallel \hat{\pi}_{xt}\hat{\pi}_{yt}\right) \\
&= \iiint_{x,y,t} \log\left(\frac{\pi_{xt}\pi_{yt}}{\sigma(t)\hat{\pi}_{xt}\hat{\pi}_{yt}}\right)\pi(dx,dy,dt) \\
&= \iiint_{x,y,t} \left(\log(\frac{\pi_{xt}}{\hat{\pi}_{xt}}) + \log(\frac{\pi_{yt}}{\hat{\pi}_{yt}}) - \log(\sigma(t))\right)\pi(dx,dy,dt) \\
&= \iint_{x,t} \log(\frac{\pi_{xt}}{\hat{\pi}_{xt}})\pi_{xt} + \iint_{y,t} \log(\frac{\pi_{yt}}{\hat{\pi}_{yt}})\pi_{yt} - \int_t \log(\sigma(t))\sigma(dt)
\end{aligned}$$

The last term vanishes since

$$\int_t \log(\sigma(t))\sigma(dt) = \sigma(t)\log(\sigma(t)) - \sigma(t)\Big|_{\sigma(t_0)}^{\sigma(t_f)} = 0$$

for compactly supported measure $\sigma(t)$, which concludes the proof. $\square$

6.2 Discretization and solver

To solve the entropic regularization problem numerically, we consider its discretized version where the cost and the coupling are represented as tensors, denoted by $c \in \mathbb{R}_+^{n_x \times n_y \times n_t}$ and $\pi \in \mathbb{R}_+^{n_x \times n_y \times n_t}$, respectively. To simplify notation, we use $P_x(\pi) = \sum_{y,t}\pi$, and similar notation for $P_y(\pi)$ and $P_t(\pi)$ with a slight abuse of notation. The *tensor (outer) product, elementwise product*, and *elementwise division* division are denoted by $\otimes$, $\odot$, and $./$, respectively. We begin by introducing the discretized entropic and the following propositions.

The discrete entropy $\mathcal{H}(\pi)$ of the coupling π is thus introduced as

$$\mathcal{H}(\pi) := -\sum_{i,j,k} \pi_{i,j,k} \log(\pi_{i,j,k}) + \pi_{i,j,k},$$

with the convention that $\mathcal{H}(\pi_{i,j,k}) = 0$ if $\pi_{i,j,k} = 0$, and $\mathcal{H}(\pi_{i,j,k}) = -\infty$ if $\pi_{i,j,k} < 0$.

Given the marginals $\mu \in \mathbb{R}_+^{n_x}$, $\nu \in \mathbb{R}_+^{n_y}$, and $\sigma \in \mathbb{R}_+^{n_t}$. The discretized objective (6.3) reads

$$\arg\min_{\pi} \langle c, \pi \rangle - \epsilon\mathcal{H}(\pi). \tag{6.6}$$

for $\epsilon > 0$, and subject to the same marginal constraints as in Problem 2′. The unique solution π_ϵ converge to the minimizer π of Problem 5.1 as $\epsilon \to 0$, and such convergence follows by the strict convexity of $-\mathcal{H}(\pi)$, and can be done in a similar procedure as in [83, Proposition 4.1].

Proposition 6.3. *The solution to the entropic regularization of Problem 5.1 has the form*

$$\pi_\epsilon = e^{-c/\epsilon} \odot \left(e^{-u(x)/\epsilon} \otimes e^{-v(y)/\epsilon} \otimes e^{-w(t)/\epsilon}\right)$$

$$= \hat{\pi} \odot (\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}),$$

for scaling variables[2] $(u, v, w) \in \mathbb{R}_+^{n_x} \times \mathbb{R}_+^{n_y} \times \mathbb{R}_{\geq 1}^{n_t}$.

Proof. Introducing the dual variable $u \in \mathbb{R}^{n_x}$, $v \in \mathbb{R}^{n_y}$, and $w \in \mathbb{R}_+^{n_t}$. The Lagrangian has the form

$$\mathcal{L}(\pi, u, v, w) = \langle c, \pi \rangle - \epsilon\mathcal{H}(\pi) + \langle w, P_t(\pi) - r \rangle$$

$$+ \langle u, P_x(\pi) - \mu \rangle + \langle v, P_y(\pi) - \nu \rangle.$$

---

[2] $\mathbb{R}_{\geq 1}^n$ denotes n- dimensional vector element-wise larger than 1

The first-order optimality condition gives the optimizer

$$\pi_\epsilon := \hat{\pi} \odot (\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w})$$

with $\hat{\pi} := e^{(-c/\epsilon)}$, $\mathbf{u} := e^{(-u(x)/\epsilon)}$, $\mathbf{v} := e^{(-v(y)/\epsilon)}$, and $\mathbf{w} := e^{(-w(t)/\epsilon)}$. □

6.2.1 Sinkhorn-Knopp algorithm

The updating scheme of the dual variables $u(x)$, $v(y)$, and $w(t)$ can be considered as a generalization of the well-known Sinkhorn-Knopp algorithm [96, 83] for multi-marginal optimal transport problem, which iteratively computes the tensor product of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ such that π_ϵ converges to π with the fix marginals μ and ν, and a upper-bounded marginal σ, which can be found in [52] where a generalized Sinkhorn framework for solving multi-marginal optimal transport problems with relaxed marginals are proposed.

More specifically, the updates in the Sinkhorn framework can be written as[3]

$$\begin{aligned}
\mathbf{u} &:= \frac{\mu}{\hat{\pi} \odot (\mathbf{v} \otimes \mathbf{w})}, \\
\mathbf{v} &:= \frac{\nu}{\hat{\pi} \odot (\mathbf{u} \otimes \mathbf{w})}, \\
\mathbf{w} &:= \min\left\{\frac{r\mathbb{1}}{\hat{\pi} \odot (\mathbf{u} \otimes \mathbf{v})}, 1\right\}
\end{aligned} \tag{6.7}$$

where marginal μ and ν is given and the time marginal is bounded by $r\mathbb{1}$. For the convergence and the linear convergence rate of (6.7), we refer to [52, Section 4].

The generalized Sinkhorn algorithm is expected to have wide-ranging applications due to its superior performance in solving the optimal transport problem. The algorithm can be extended to handle additional constraints and complex cost functions, making it applicable to a broad range of practical problems. Its scalability and computational efficiency

---

[3]Throughout, we denote $\mathbb{1}_n$ for all-one (column) vector.

make it particularly useful for large-scale optimization problems encountered in many fields, including computer vision, image processing, machine learning, and economics.

6.2.2 Illustrative Example

Applying the framework proposed in [52], the the entropic regularization can be solved with $\epsilon = 10^{-2},\ 10^{-3},\ 10^{-4},\ 10^{-5}$, and the results are visualized in Fig. 6.1.

(a) $\epsilon = 0.01$

(b) $\epsilon = 0.001$
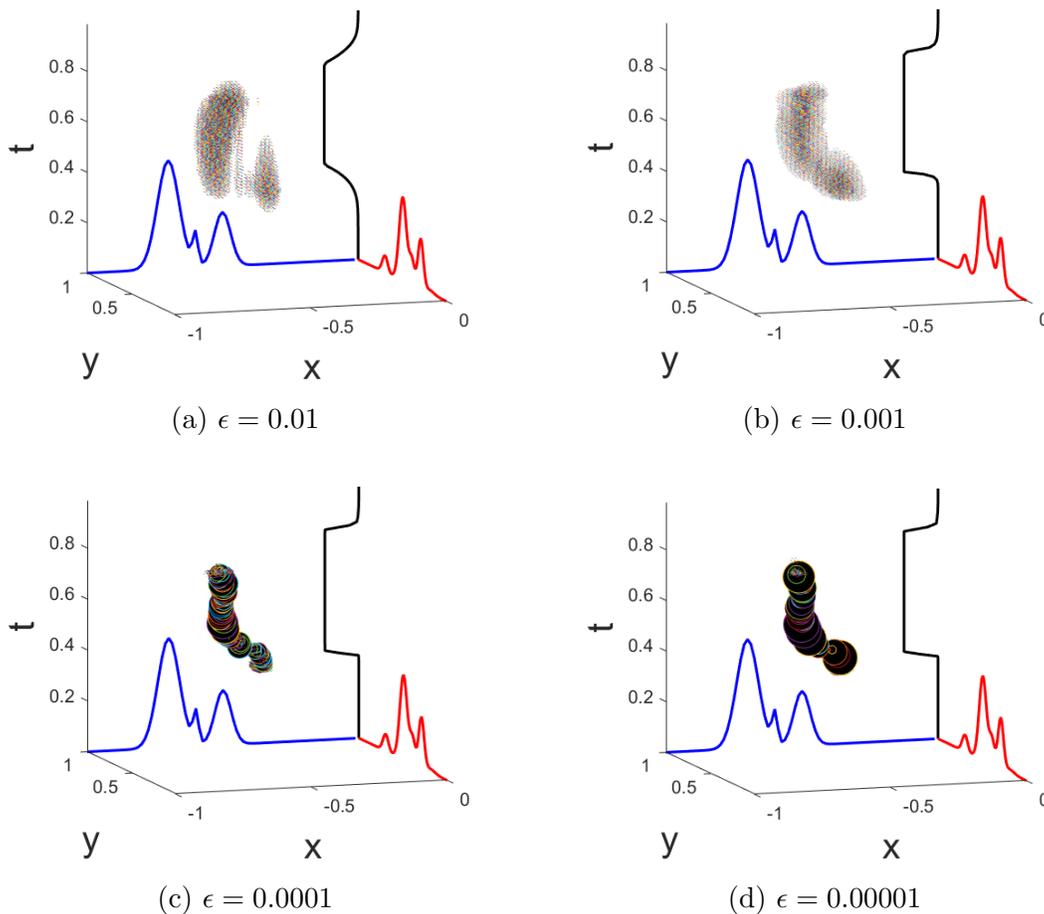
(c) $\epsilon = 0.0001$

(d) $\epsilon = 0.00001$

Figure 6.1: The optimal coupling with $\epsilon = 0.01,\ 0.001,\ 0.0001,$ and 0.00001. The radius of the point (x, y, t) is proportional to the quantity of mass $\pi_\epsilon(x, y, t)$ according to the coupling

However, in all approaches above, the size of the variable and the computational complexity for each iteration $\mathcal{O}(n^3)$ is considerably large, which brings us to further reduce the size of Problem 6.2 in the next section.

6.3 The Gluing Sinkhorn

From Theorem 6.1, the minimizer of Problem 5.1 can be construct by the minimizer π_{xt} and π_{yt} of Problem 6.2 using the factorization (6.5). Herein, aiming to solve Problem 6.2 by the scheme of the Sinkhorn algorithm, we propose a gluing Sinkhorn.

Denoting the two-dimensional couplings between the x, y marginals and t marginal as π_{xt} and π_{yt}, the discretization of the objective of Problem 6.2 is thus

$$\langle c_{xt}, \pi_{xt} \rangle + \langle c_{yt}, \pi_{yt} \rangle - \epsilon \mathcal{H}(\pi_{xt}) - \epsilon \mathcal{H}(\pi_{yt}) \tag{6.8}$$

over $\pi_{xt} \in \Pi_{xt}$ and $\pi_{yt} \in \Pi_{yt}$.

Proposition 6.4. *The minimizer of Problem 6.2, π_{xt} and π_{yt} converge to the exact solution in 6.2 as $\epsilon \to 0$.*

Proof. The convergence can be obtained by the strong convexity of the entropic penalty term, i.e., $-\mathcal{H}(\pi_{xt})$ and $-\mathcal{H}(\pi_{yt})$ are strictly convex, and the proof can be done similarly as in [83, Proposition 4.1]. □

The factorization (6.5) provides an approach to accelerate the standard multi-marginal Sinkhorn algorithm as we summarized in (6.7), whereby the updating of π_ϵ is decomposed into π_{xt} and π_{yt} with the additional equality constraint $P_t(\pi_{xt}) = P_t(\pi_{yt})$ for the time marginal. Such decomposition reduced the size of the variables from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ with interesting insights for the Sinkhorn algorithm as shown later.

Proposition 6.5. *The optimizer of Problem 3 is unique and has the form:*

$$\pi_{xt} = \text{diag}(\mathbf{u}_1)\hat{\pi}_{xt}\,\text{diag}(\mathbf{v}_1 \odot \mathbf{w}),$$

$$\pi_{yt} = \text{diag}(\mathbf{u}_2)\hat{\pi}_{yt}\,\text{diag}(\mathbf{v}_2 \,./\mathbf{w})$$

where $\hat{\pi}_{xt} = \exp(-c_{xt}/\epsilon)$, $\hat{\pi}_{yt} = \exp(-c_{yt}/\epsilon)$ and the five (unknown) scaling variables $(u_{1,2}, v_{1,2}, w) \in (\mathbb{R}^{n_x})^2 \times (\mathbb{R}^{n_y})^2 \times \mathbb{R}^{n_t}$, are optimal (dual) variables for the dual of Problem 6.2

$$\arg\max_{u \in \mathbb{R}^n, v \in \mathbb{R}^n_+, w \in \mathbb{R}^n} \quad -\epsilon\langle\hat{\pi}_{xt}, U_{xt}\rangle - \epsilon\langle\hat{\pi}_{yt}, U_{yt}\rangle - \langle u_1, \mu\rangle$$

$$- \langle u_2, \nu\rangle - \langle v_1, r\rangle - \langle v_2, r\rangle.$$

where $U_{xt} = u_1 \otimes (v_1 \odot w)$ and $U_{yt} = u_2 \otimes (v_2./w)$.

Proof. To see this, we first derive the Lagrangian

$$\mathcal{L}(\pi_{xt}, \pi_{yt}, u_1, u_2, v_1, v_2, w) = \langle c_{xt}, \pi_{xt}\rangle + \langle c_{yt}, \pi_{yt}\rangle$$

$$- \epsilon\mathcal{H}(\pi_{xt}) - \epsilon\mathcal{H}(\pi_{yt}) + \langle u_1, P_x(\pi_{xt}) - \mu\rangle + \langle u_2, P_y(\pi_{yt}) - \nu\rangle$$

$$+ \langle v_1, P_t(\pi_{xt}) - r\rangle + \langle v_2, P_t(\pi_{yt}) - r\rangle + \langle w, P_t(\pi_{xt}) - P_t(\pi_{yt})\rangle,$$

and according to the first-order optimality condition, the $(i,j),(j,k)$-th element of π_{xt} and π_{yt}, denoted as $\pi_{xt}(i,j)$ and $\pi_{yt}(j,k)$, have the form:

$$\pi_{xt}(i,j) = e^{-u_1(i)/\epsilon} \cdot e^{-c_{xt}(i,j)/\epsilon} \cdot e^{-v_1(j)/\epsilon} \cdot e^{-w(j)/\epsilon},$$

$$\pi_{yt}(j,k) = e^{-u_2(k)/\epsilon} \cdot e^{-c_{yt}(j,k)/\epsilon} \cdot e^{-v_2(j)/\epsilon} \cdot e^{w(j)/\epsilon},$$

or equivalently, in the matrix form using the tensor operator

$$\pi_{xt} = \exp(-\frac{c_{xt}}{\epsilon}) \odot \left(\exp(-\frac{u_1}{\epsilon}) \otimes \left(\exp(-\frac{v_1}{\epsilon}) \odot \exp(\frac{-w}{\epsilon}) \right) \right)$$

$$= \text{diag}(\mathbf{u}_1)\hat{\pi}_{xt}\,\text{diag}(\mathbf{v}_1 \odot \mathbf{w}) = \hat{\pi}_{xt} \odot (\mathbf{u}_1 \otimes (\mathbf{v}_1 \odot \mathbf{w})),$$

$$\pi_{yt} = \exp(-\frac{c_{yt}}{\epsilon}) \odot \left(\exp(-\frac{u_2}{\epsilon}) \otimes \left(\exp(-\frac{v_2}{\epsilon}) \odot \exp(\frac{+w}{\epsilon}) \right) \right)$$

$$= \text{diag}(\mathbf{u}_2)\hat{\pi}_{yt}\,\text{diag}(\mathbf{v}_2./\mathbf{w}) = \hat{\pi}_{yt} \odot (\mathbf{u}_2 \otimes (\mathbf{v}_2./\mathbf{w})).$$

Then, by bringing the expression of optimal π_{xt} and π_{yt} above into the Lagrangian, we derive

the dual problem, which concludes the proof. □

The updates scheme for vector $\mathbf{u}_1$ and $\mathbf{u}_2$ for the fixed marginals μ and ν can be obtain by

$$\text{diag}(\mathbf{v}_1 \odot \mathbf{w})\hat{\pi}_{xt}^T \text{diag}(\mathbf{u}_1)\mathbb{1}_{n_t} = \mu,$$
$$\text{diag}(\mathbf{v}_2 ./\mathbf{w})\hat{\pi}_{yt}^T \text{diag}(\mathbf{u}_2)\mathbb{1}_{n_t} = \nu,$$
(6.9)

and the updates of $\mathbf{v}_1$ and $\mathbf{v}_2$ for relaxed marginal, i.e., $\sigma(dt) \leq rdt$, can be found from [52, Section 4] so that

$$\text{diag}(\mathbf{u}_1)\hat{\pi}_{xt} \text{diag}(\mathbf{v}_1 \odot \mathbf{w})\mathbb{1}_{n_x} \leq r\mathbb{1}$$
$$\text{diag}(\mathbf{u}_2)\hat{\pi}_{yt} \text{diag}(\mathbf{v}_2 ./\mathbf{w})\mathbb{1}_{n_y} \leq r\mathbb{1}.$$
(6.10)

Herein, due to the existence of the additional constraint, the *Gluing variable* ω is introduced and follows

$$\text{diag}(\mathbf{u}_1)\hat{\pi}_{xt} \text{diag}(\mathbf{v}_1 \odot \mathbf{w})\mathbb{1} = \text{diag}(\mathbf{u}_2)\hat{\pi}_{yt} \text{diag}(\mathbf{v}_2 ./\mathbf{w})\mathbb{1}$$
(6.11)

Thus, the Gluing Sinkhorn can be summarized as follows.

Theorem 6.2 (Convergence). *Assume Problem 3 admits a feasible minimizer, the rounding scheme described in (6.12), (6.13), and (6.14) converges, and in the limit point, the optimal solution of Problem 6.2 is given by their definitions in Proposition 7.*

Proof. The rounding scheme of the Gluing Sinkhorn is to maximize the objective with respect to one set of dual variables while keeping the other dual variables fixed, i.e., to perform the

Algorithm 4 Gluing Sinkhorn-Knopp algorithm

Input: Costs c_{xt}, c_{yt}, and regularization parameter $\epsilon > 0$.
Initialization: $\mathbf{v}_1 = \mathbb{1}_{n_t}$, $\mathbf{v}_2 = \mathbb{1}_{n_t}$, $\mathbf{w} = \mathbb{1}_{n_t}$.
Output: Couplings π_{xt} and π_{yt}.

1: **while** not converge **do**
2:    Compute $\mathbf{u}_1$ and $\mathbf{u}_2$ according to

$$\mathbf{u}_1 = \frac{\mu}{\hat{\pi}_{xt}(\mathbf{v}_1 \odot \mathbf{w})}, \tag{6.12a}$$

$$\mathbf{u}_2 = \frac{\nu}{\hat{\pi}_{yt}(\mathbf{v}_2 \ ./\mathbf{w})}. \tag{6.12b}$$

3:    Compute $\mathbf{v}_1$ and $\mathbf{v}_2$ according to

$$\mathbf{v}_1 = \min\{\frac{r\mathbb{1}}{\mathbf{w} \odot (\hat{\pi}_{xt}^T \mathbf{u}_1)}, 1\}, \tag{6.13a}$$

$$\mathbf{v}_2 = \min\{\frac{\mathbf{w} \odot r\mathbb{1}}{(\hat{\pi}_{xt}^T \mathbf{u}_1)}, 1\}. \tag{6.13b}$$

4:    Compute $\mathbf{w}$ according to

$$\mathbf{w} = \sqrt{\frac{\mathbf{u}_2 \odot (\hat{\pi}_{yt}\mathbf{v}_2)}{\mathbf{u}_1 \odot (\hat{\pi}_{xt}\mathbf{v}_1)}}. \tag{6.14}$$

5: **end while**
6: **return** π_{xt} and π_{yt} according to

$$\pi_{xt} = \hat{\pi}_{xt} \odot (\mathbf{u}_1 \otimes (\mathbf{v}_1 \odot \mathbf{w})), \pi_{yt} = \hat{\pi}_{yt} \odot (\mathbf{u}_2 \otimes (\mathbf{v}_2./\mathbf{w})).$$

updates[4]

$$u := \underset{\psi_x \in \mathbb{R}^n}{\arg\max} -\epsilon\langle\hat{\pi}_{xt}, U_{xt}\rangle - \langle u_1, \mu\rangle, \tag{6.15a}$$

$$v := \underset{\lambda_1 \in \mathbb{R}^n_+}{\arg\max} -\epsilon\langle\hat{\pi}_{xt}, U_{xt}\rangle - \langle v_1, r\rangle \tag{6.15b}$$

$$w := \underset{w \in \mathbb{R}^n}{\arg\max} -\epsilon\langle\hat{\pi}_{xt}, U_{xt}\rangle - \epsilon\langle\hat{\pi}_{yt}, U_{yt}\rangle \tag{6.15c}$$

First, to prove the correctness of the routing scheme in (6.12)-(6.14), we divided the updates

---

[4]In the updates of u and v, we assume $n_x = n_t = n$ without loss of generality and focus on the updates of π_{xt} with subscript omitted. The updating scheme of π_{yt} is conceptually the same with an orientation of signs due to the last equality constraint on the time marginal.

of the dual variables as follows:

(i) The objective of the unconstrained problem (6.15a) is strictly concave, and thus a necessary and sufficient condition for optimality is that the respective gradient vanishes. The gradient of (6.15a) with respect to $\psi(x)$ is

$$\exp(-\psi(x)/\epsilon) \otimes (\hat{\pi} \odot (v \otimes w)) - \mu,$$

and (6.12) can be obtained by setting (6.15a) to zero.

(ii) The objective in (6.15b) can be written as

$$\forall\, i \in n, \quad \exp(-\lambda_1(i)/\epsilon) \otimes (\hat{\pi} \odot (u \otimes w)) - r(i),$$

The maximization in (6.15b) can be performed element-wise for λ. If the derivative of the objective in (6.15b) w.r.t λ vanishes for a feasible, i.e., non-negative, point, then this is the global maximizer. Otherwise, the maximizer is the projection on the feasible set, i.e., $\lambda(i) = 0$. This yields (6.13).

(iii) Similarly, the strong convexity holds for the objective in 6.15c and its gradient w.r.t w reads

$$\exp(-\frac{w}{\epsilon}) \otimes \hat{\pi}_{xt} \odot (u_1 \otimes v_1) - \exp(\frac{w}{\epsilon}) \otimes \hat{\pi}_{yt} \odot (u_2 \otimes v_2)$$

Thus, (6.14) can be obtained by setting the above to zero. $\qquad\square$

The proof of linear convergence is based on the Hilbert matrix $d_H(\mu, \nu)$ (see, e.g., [83, Remark 4.12]) defined as

$$d_H(\mu, \nu) := \| \log(\mu) - \log(\nu) \|_V, \quad \forall\, \mu, \nu \in \mathbb{R}_+^n, \tag{6.16}$$

with $\|x\|_V = \max_i x_i - \min_i x_i$. Moreover, for $\mu, \nu \in \mathbb{R}^n_+$ and $K \in \mathbb{R}^{n \times n}_+$, we have

$$d_H(K\mu, K\nu) \le \lambda(K)d_H(\mu, \nu) \tag{6.17}$$

where $\lambda(K) = \frac{\sqrt{\eta(K)}-1}{\sqrt{\eta(K)}+1} < 1$, and $\eta(K) = \max_{i,j,k,l} \frac{K_{ij}K_{jl}}{K_{jk}K_{il}}$, as stated in [83, Theorem 4.1]

Theorem 6.3 (Convergence rate). *The gluing Sinkhorn has a linear convergence rate.*

Proof. The linear convergence rate for the marginals μ follows the standard proof of Sinkhorn, i.e.[5],

$$\begin{aligned}
d_H(D_u^k, \ D_u^\star) &= d_H\left(\frac{\mu}{\hat{\pi}_{xt}D_{v \cdot w}^k}, \ \frac{\mu}{\hat{\pi}_{xt}D_{v \cdot w}^\star}\right) \\
&= d_H(\hat{\pi}_{xt}D_{v \cdot w}^k, \ \hat{\pi}_{xt}D_{v \cdot w}^\star) \le \lambda(\hat{\pi}_{xt})d_H(D_{v \cdot w}^k, \ D_{v \cdot w}^\star),
\end{aligned} \tag{6.18}$$

using the result of (6.17). Moreover, using the triangular inequality, we have

$$\begin{aligned}
d_H&(D_u^k, \ D_u^\star) \\
&\le d_H(D_u^{k+1}, \ D_u^k) + d_H(D_u^{k+1}, \ D_u^\star) \\
&\le d_H\left(\frac{\mu}{\hat{\pi}_{xt}D_{v \cdot w}^k}, \ D_u^k\right) + \lambda(\hat{\pi}_{xt})^2 d_H(D_u^k, \ D_u^\star) \\
&= d_H(\mu, D_u^k \odot (\hat{\pi}_{xt}D_{v \cdot w}^k)) + \lambda(\hat{\pi}_{xt})^2 d_H(D_u^k, \ D_u^\star) \\
&= d_H(\mu, \mu^k) + \lambda(\hat{\pi}_{xt})^2 d_H(D_u^k, \ D_u^\star)
\end{aligned}$$

where $\mu^k = D_u^k \odot (\hat{\pi}_{xt}D_{v \cdot w}^k)$, and thus

$$d_H(D_u^k, \ D_u^\star) \le \frac{d_H(\mu, \mu^k)}{1 - \lambda(\hat{\pi}_{xt})^2}. \tag{6.19}$$

The Hilbert metric rates (6.18) on the scaling variable D_u give a linear rate on the dual

---

[5]for the sake of simplicity, we omitted the subscript for u_1 to be u. Also, we denote $\mathrm{diag}(u)$ by D_u, and $\mathrm{diag}(u)$ in kth iteration by D_u^k in the proof.

variable $\epsilon \log(u)$ for the variation norm $\| \cdot \|_V$. The inequality (6.19) suggests error measures on the violation of marginal constraints, e.g., $\|\mu^k - \mu\|_1$.

The result obtained in (6.18) and (6.19) is also true for u_2, $v_1 \odot w$, and $v_2./w$. The proof follows the procedure as u_1.

$\square$

Remark 6.1 (**Complexity**). *Considering the sparsity of admissible couplings, the computation of the projection can be further optimized. In particular, we only need to compute projections on the indices where $\pi_{i,j,t} > 0$ and can skip the rest. This leads to significant computational savings, from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$, especially for large-scale problems.*

Complexity	Variables' Size	Variable's updates	Output's Size
Sinkhorn	$\mathcal{O}(n^3)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n^3)$
Linear program	$\mathcal{O}(n^2)$	$\mathcal{O}(n^3 \log(n))$	$\mathcal{O}(n^2)$
Gluing Sinkhorn	$\mathcal{O}(n^2)$	$\mathcal{O}(n)$	$\mathcal{O}(n^2)$

Table 6.1: Complexity comparision.

6.4   Numerical experiments

The code used to conduct all the experiments is publicly available at `github.com/dytroshut/GluingSinkhorn.git`. The numerical example setting is the same as the one described in Section 3. In this example, we seek two joint distributions or couplings, namely π_{xt} and π_{yt}, instead of the three-dimensional coupling.

The linear convergence rate of fixed marginals μ and ν and the sub-linear (logarithmic) convergence rate of time marginal σ are shown below.

The running times of the six methods we applied are summarized below to show the ad-
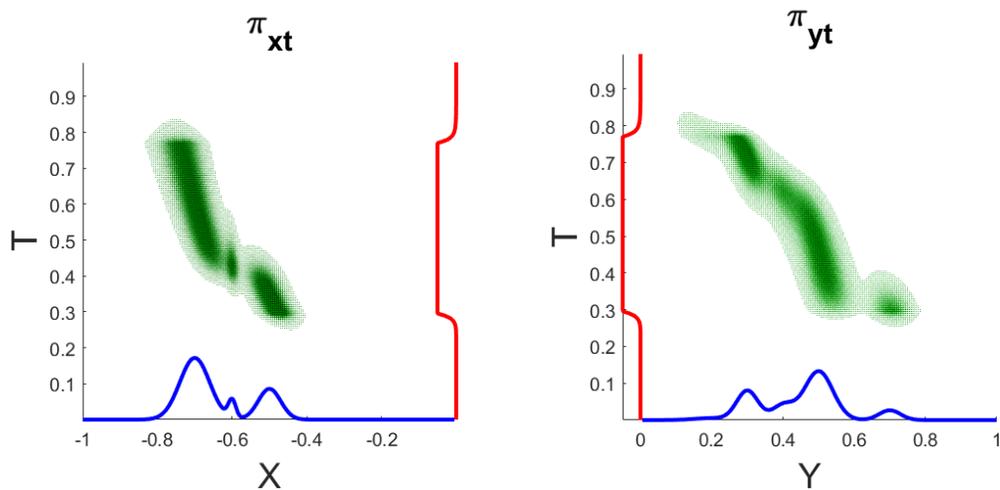
(a) $\epsilon = 0.001$
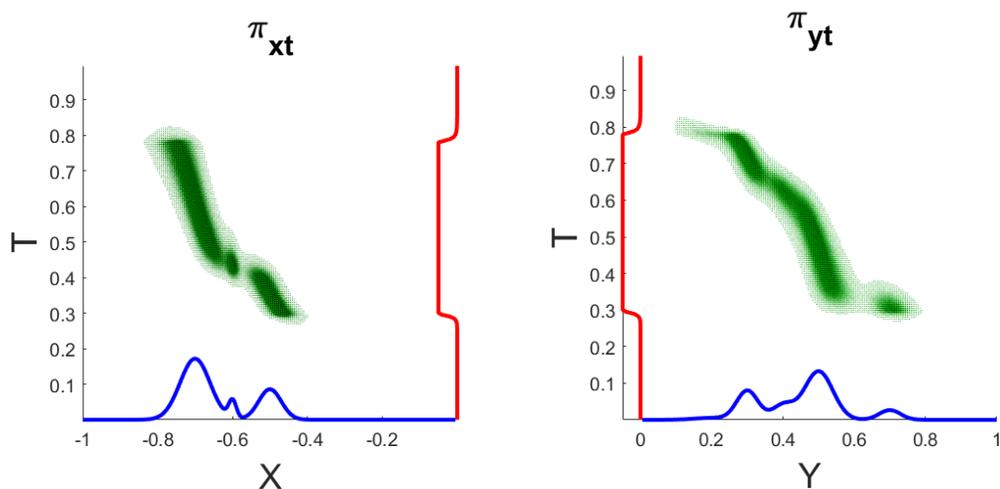
(b) $\epsilon = 0.0005$

Figure 6.2: Visualization of the outputs of the gluing Sinkhorn with $\epsilon = 0.001$ and 0.0005.

vantages of our Gluing Sinkhorn algorithm, wherein we utilized the standard CVX solver [49], linear programming (LP), Sinkhorn (6.7) and gluing Sinkhorn. The running time is measured in seconds using the MATLAB built-in function (tic/toc).

Methods	CVX	CVX-split	LP	Sinkhorn	Gluing
Running time	164.34s	3.73s	$1s$	2.64s	0.99s

Table 6.2: Running time comparison.
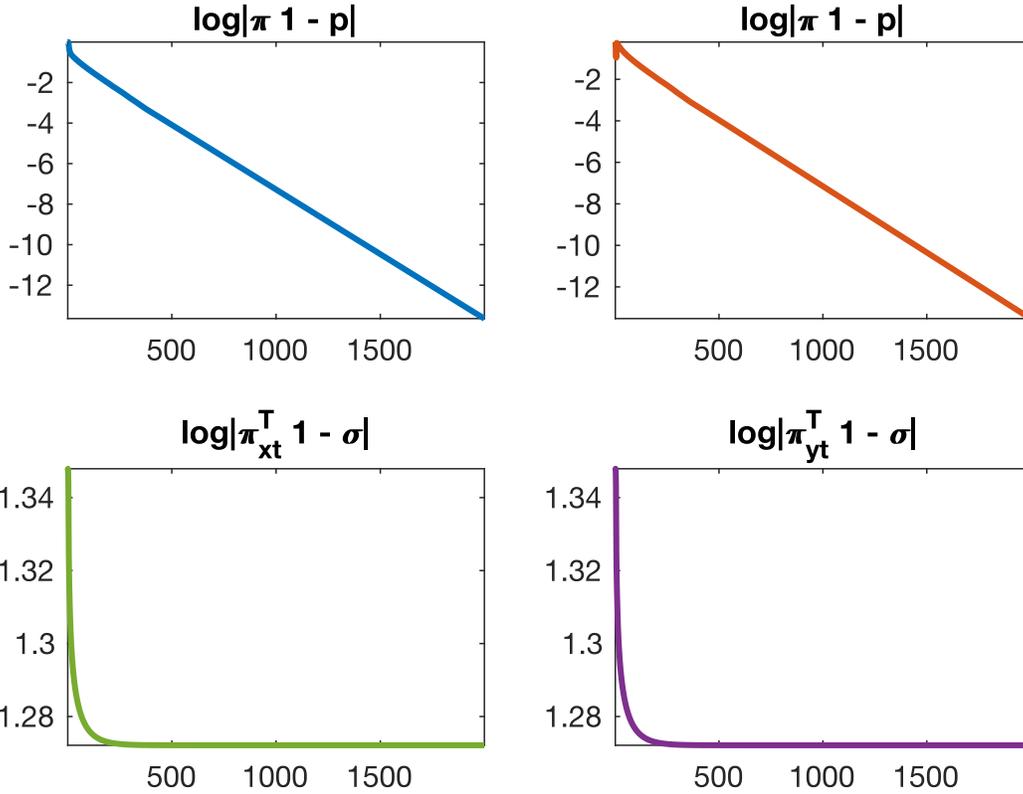
Figure 6.3: The convergence rate of the algorithm is measured with respect to the number of iterations, and the y-axis of the corresponding plot shows the value of $\log(|P_x(\pi_{xt}) - \mu|)$, $\log(|P_x(\pi_{yt}) - \nu|)$, $\log(|P_t(\pi_{xt}) - \sigma^\star|)$, and $\log(|P_t(\pi_{yt}) - \sigma^\star|)$. Here, $\sigma^\star$ denotes the optimal time marginal, and the values in the plot represent the logarithm of the absolute difference between the computed marginal values and the optimal values.

Bibliographical notes

We believe that the formulation of the constricted Monge-Kantorovich problem and the framework for solving the problem has a huge potential in the study of congested transportation analysis, image processing and also time-varying matching/transportation.

Besides the standard OT problem with given marginals (supply and demand), many other constrained variations have been explored. For instance, the marginal equality constraints are relaxed to inequalities in [87, 89]; An upper bound on the coupling, referred to as the *capacity*, is considered in [63].

On the practical side, the computational aspect of OT is widely applied in fields such as mathematics, economics, and engineering with numerous applications in motion planning, matching, image processing, and many other areas. A comprehensive tutorial survey on the subject can be found in [83]. However, solving the Monge-Kantorovich problem with linear programming for an exact solution has $\mathcal{O}(n^3 \log(n))$ complexity. To overcome the high computational cost, entropic regularization of the original problem is often utilized and solved by a rounding scheme known as the *Sinkhorn-Knopp* algorithm [33]. The algorithm has a linear convergence rate and operates through iterative projections. For modifications using matrix factorization, refer to [92], and for prime generalizations of the multi-marginal case, cf. [52].

Bibliography

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network flows. Theory, Algorithms, and Applications.* Prentice Hall, 1988.

[2] R. K. Ahuja, K. Mehlhorn, J. Orlin, and R. E. Tarjan. Faster algorithms for the shortest path problem. *Journal of the ACM*, 37(2):213–223, 1990.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, 2005.

[4] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook. The traveling salesman problem. In *The Traveling Salesman Problem.* Princeton university press, 2011.

[5] R. B. Bapat. *Graphs and matrices*, volume 27. Springer, 2010.

[6] H. Bast, D. Delling, A. Goldberg, M. Müller-Hannemann, T. Pajor, P. Sanders, D. Wagner, and R. F. Werneck. Route planning in transportation networks. In *Algorithm engineering*, pages 19–80. Springer, 2016.

[7] R. Bellman. On a routing problem. *Quarterly of applied mathematics*, 16(1):87–90, 1958.

[8] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning*, 3(1):1–122, 2011.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Matlab scripts for alternating direction method of multipliers, 2011.

[11] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[12] K. Bredies, M. Carioni, S. Fanzon, and F. Romero. A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization. *Foundations of Computational Mathematics*, pages 1–66, 2022.

[13] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.

[14] L. A. Caffarelli and R. J. McCann. Free boundaries in optimal transport and monge-ampere obstacle problems. *Annals of mathematics*, pages 673–730, 2010.

[15] S. Cambanis, G. Simons, and W. Stout. Inequalities for $\mathcal{E}k(x, y)$ when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 36(4):285–294, 1976.

[16] G. Carlier. On a class of multidimensional optimal transportation problems. *Journal of convex analysis*, 10(2):517–530, 2003.

[17] G. Carlier. Optimal transportation and economic applications. *Lecture Notes*, 18, 2012.

[18] G. Carlier, C. Jimenez, and F. Santambrogio. Optimal transportation with traffic congestion and wardrop equilibria. *SIAM Journal on Control and Optimization*, 47(3):1330–1350, 2008.

[19] S. Chandrasekar and P. K. Srimani. A self-stabilizing distributed algorithm for all-pairs shortest path problem. *Parallel Algorithms and Applications*, 4(1-2):125–137, 1994.

[20] J. T. Chang and D. Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317, 1997.

[21] L. Chapel, M. Z. Alaya, and G. Gasso. Partial optimal tranport with applications on positive-unlabeled learning. *Advances in Neural Information Processing Systems*, 33:2903–2913, 2020.

[22] Y. Chen, T. Georgiou, M. Pavon, and A. Tannenbaum. Robust transport over networks. *IEEE transactions on automatic control*, 62(9):4675–4682, 2016.

[23] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

[24] Y. Chen, T. T. Georgiou, and M. Pavon. Controlling Uncertainty. *IEEE Control Systems Magazine*, 41(4):82–94, 2021.

[25] Y. Chen, T. T. Georgiou, and M. Pavon. Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge. *SIAM Review*, 63(2):249–313, 2021.

[26] P.-A. Chiappori, R. J. McCann, and L. P. Nesheim. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, pages 317–354, 2010.

[27] P.-A. Chiappori, R. J. McCann, and B. Pass. Multidimensional matching. *arXiv:1604.05771 preprint*, 2016.

[28] P.-A. Chiappori, R. J. McCann, and B. Pass. Multi-to one-dimensional optimal transport. *Communications on Pure and Applied Mathematics*, 70(12):2405–2444, 2017.

[29] F. R. Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[30] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms.* MIT press, 2022.

[31] T. M. Cover. *Elements of information theory.* John Wiley & Sons, 1999.

[32] M. Cullen and R. Douglas. Applications of the monge-ampere equation and monge transport problem to meteorology and oceanography. *Contemporary Mathematics*, 226:33, 1999.

[33] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[34] G. B. Dantzig. Discrete-variable extremum problems. *Operations research*, 5(2):266–288, 1957.

[35] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[36] A. Dong, A. Stephanovitch, and T. T. Georgiou. Monge-kantorovich optimal transport through constrictions and flow-rate constraints. *arXiv preprint arXiv:2212.14509*, 2022.

[37] I. Ekren and H. M. Soner. Constrained optimal transport. *Archive for Rational Mechanics and Analysis*, 227(3):929–965, 2018.

[38] P. Erdős, A. Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[39] L. Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 8:128–140, 1736.

[40] J. Fan, I. Haasler, J. Karlsson, and Y. Chen. On the complexity of the optimal transport problem with graph-structured cost. In *International Conference on Artificial Intelligence and Statistics*, pages 9147–9165. PMLR, 2022.

[41] A. Figalli. The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2):533–560, 2010.

[42] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian journal of Mathematics*, 8:399–404, 1956.

[43] L. R. Ford Jr. Network flow theory. Technical report, Rand Corp Santa Monica Ca, 1956.

[44] A. Forrow, J.-C. Hütter, M. Nitzan, P. Rigollet, G. Schiebinger, and J. Weed. Statistical optimal transport via factored couplings. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2454–2465. PMLR, 2019.

[45] M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM*, 34(3):596–615, 1987.

[46] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[47] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[48] P. Gladbach and E. Kopfer. Limits of density-constrained optimal transport. *Calculus of Variations and Partial Differential Equations*, 61(2):1–31, 2022.

[49] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1, 2014.

[50] D. Gusfield and R. W. Irving. *The stable marriage problem: structure and algorithms*. MIT press, 1989.

[51] I. Haasler, A. Ringh, Y. Chen, and J. Karlsson. Multimarginal optimal transport with a tree-structured cost and the Schroedinger bridge problem. *SIAM Journal on Control and Optimization*, 59(4):2428–2453, 2021.

[52] I. Haasler, A. Ringh, Y. Chen, and J. Karlsson. Scalable computation of dynamic flow problems via multi-marginal graph-structured optimal transport. *arXiv:2106.14485, preprint*, 2021.

[53] I. Haasler, R. Singh, Q. Zhang, J. Karlsson, and Y. Chen. Multi-marginal optimal transport and probabilistic graphical models. *IEEE Transactions on Information Theory*, 67(7):4647–4668, 2021.

[54] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal of computer vision*, 60(3):225–240, 2004.

[55] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[56] M. R. Hestenes. Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303–320, 1969.

[57] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.

[58] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[59] L. V. Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

[60] R. M. Karp. *Reducibility among combinatorial problems.* Springer, 2010.

[61] Y.-H. Kim and B. Pass. A general condition for monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46(2):1538–1550, 2014.

[62] J. Korman and R. J. McCann. Insights into capacity-constrained optimal transport. *Proceedings of the National Academy of Sciences*, 110(25):10064–10067, 2013.

[63] J. Korman and R. J. McCann. Optimal transportation with capacity constraints. *Transactions of the American Mathematical Society*, 367(3):1501–1521, 2015.

[64] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Traces and emergence of nonlinear programming*, pages 247–258. Springer, 2013.

[65] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[66] H. Lavenant, S. Claici, E. Chien, and J. Solomon. Dynamical optimal transport on discrete surfaces. *ACM Transactions on Graphics (TOG)*, 37(6):1–16, 2018.

[67] E. L. Lawler. The traveling salesman problem: a guided tour of combinatorial optimization. *Wiley-Interscience Series in Discrete Mathematics*, 1985.

[68] C. Léonard. A large deviation approach to optimal transport. *arXiv preprint arXiv:0710.1461*, 2007.

[69] L. Lovász. Random walks on graphs. *Combinatorics, Paul erdos is eighty*, 2(1-46):4, 1993.

[70] D. G. Luenberger, Y. Ye, et al. *Linear and nonlinear programming*, volume 2. Springer, 1984.

[71] R. J. McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.

[72] R. J. McCann and B. Pass. Optimal transportation between unequal dimensions. *Archive for Rational Mechanics and Analysis*, 238(3):1475–1520, 2020.

[73] M. Mesbahi and M. Egerstedt. *Graph theoretic methods in multiagent networks.* Princeton University Press, 2010.

[74] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

[75] E. N. Mortensen and W. A. Barrett. Intelligent scissors for image composition. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191–198, 1995.

[76] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[77] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 19(4):1574–1609, 2009.

[78] M. Newman. *Networks*. Oxford university press, 2018.

[79] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 2001.

[80] N. Papadakis, G. Peyré, and E. Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[81] B. Pass. Uniqueness and monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.

[82] B. Pass. Multi-marginal optimal transport: theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015.

[83] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[84] I. Pohl. Bidirectional and heuristic search in path problems. Technical report, Technical Report SLAC-104, Stanford Linear Accelerator Center, Stanford, California, 1969.

[85] M. Potamias, F. Bonchi, C. Castillo, and A. Gionis. Fast shortest path distance estimation in large networks. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 867–876, 2009.

[86] M. J. Powell. A method for nonlinear constraints in minimization problems. *Optimization*, pages 283–298, 1969.

[87] S. Rachev and L. Rüschendorf. Solution of some transportation problems with relaxed or additional constraints. *SIAM Journal on Control and Optimization*, 32(3):673–689, 1994.

[88] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory*, volume 1. Springer Science & Business Media, 1998.

[89] S. T. Rachev and L. Rüschendorf. *Mass transportation problems: Applications*. Springer Science & Business Media, 2006.

[90] R. T. Rockafellar. *Convex analysis*, volume 11. Princeton university press, 1997.

[91] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

[92] M. Scetbon, M. Cuturi, and G. Peyré. Low-rank sinkhorn factorization. In *International Conference on Machine Learning*, pages 9344–9354. PMLR, 2021.

[93] A. Shimbel. Structural parameters of communication networks. *The bulletin of mathematical biophysics*, 15:501–507, 1953.

[94] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. A sparse-group lasso. *Journal of computational and graphical statistics*, 22(2):231–245, 2013.

[95] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.

[96] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[97] M. Sniedovich. Dijkstra's algorithm revisited: the dynamic programming connexion. *Control and cybernetics*, 35(3):599–620, 2006.

[98] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher. Continuous-flow graph transportation distances. *arXiv preprint arXiv:1603.06927*, 2016.

[99] A. Stephanovitch, A. Dong, and T. T. Georgiou. Optimal transport through a toll. *arXiv:2208.03587, preprint*, 2022.

[100] G. Tarry. Le problème des labyrinthes. *Nouvelles Annales de Mathématiques: Journal des candidats aux Écoles Polytechnique et Normale*, 14:187–190, 1895.

[101] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[102] R. J. Tibshirani. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013.

[103] P. van Emde Boas, R. Kaas, and E. Zijlstra. Design and implementation of an efficient priority queue. *Mathematical Systems Theory*, 10(1):99–127, 1976.

[104] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

[105] C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[106] B. M. Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, 1988.

[107] C. Wiener. Ueber eine aufgabe aus der geometria situs. *Mathematische Annalen*, 6(1):29–30, 1873.

[108] L. A. Wolsey. *Integer programming*. John Wiley & Sons, 2020.

[109] H. Yue, Q. Yang, X. Wang, and X. Yuan. Implementing the alternating direction method of multipliers for big datasets: A case study of least absolute shrinkage and selection operator. *SIAM Journal on Scientific Computing*, 40(5):A3121–A3156, 2018.

[110] Y. Zhang and S. Li. Distributed biased min-consensus with applications to shortest path planning. *IEEE Transactions on Automatic Control*, 62(10):5429–5436, 2017.

[111] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.

Appendix A

Legendre transformation

For the Benamou-Brenier formulation of the OMT problem, the objective function in the problem reads

$$f(m, \rho) = \frac{m^2}{2\rho}, \ \rho \geq 0$$

in which m is the momentum and ρ is the density. Equivalently, $f(m, \rho)$ can be written as

$$f(m, \rho) = \frac{1}{2} \frac{m^2}{\rho} + \sup_{\alpha \leq 0}(\alpha\rho)$$

such that the second term is 0 when only $\rho \geq 0$ and $+\infty$ when $\rho < 0$. With the dual variable $\lambda = (a, b)$, we have

$$
\begin{aligned}
f^*(a, b) &= \sup_{m, \rho} \left\{ am + b\rho - \frac{m^2}{2\rho} - \sup_{\alpha \leq 0}(\alpha\rho) \right\} \\
&= \sup_{m, \rho} \left\{ am + b\rho - \frac{m^2}{2\rho} + \inf_{\alpha \leq 0}(-\alpha\rho) \right\} \qquad\qquad (A.1) \\
&= \sup_{m, \rho} \inf_{\alpha \leq 0} \left\{ am + b\rho - \frac{m^2}{2\rho} - \alpha\rho \right\}
\end{aligned}
$$

By the first-order optimality condition, we have

$$
\begin{cases}
a - \frac{m}{\rho} = 0, \\
b + \frac{m^2}{2\rho^2} - \alpha = 0.
\end{cases}
$$

so that we have

$$
am + b\rho - \frac{m^2}{2\rho} - \alpha\rho = 0, \ \ b + \frac{|a|^2}{2} = \alpha \le 0.
$$

Recall the definition of the Legendre transformation [90, Chapter 26], $f(x) = f^*(f^*(x))$, the objective function $f(x)$ can be rewritten by

$$
f(x) = \sup_{b + \frac{|a|^2}{2} \le 0} \{a\rho + bm\}
$$

Appendix B

The shortest path problem

B.1   Proof of Lemma 3.1

Theorem B.1. *If $P_{\mathcal{A}}$ is the path matrix of graph $\mathcal{G}(\mathcal{A})$ then*

$$D_{\mathcal{A}'}^{-1} = P_{\mathcal{A}}$$

Proof. The proof is to show that

$$D_{\mathcal{A}'}\ P_{\mathcal{A}} = I$$

By considering the $(i,\ j)$-element $(\sum_{k=1}^{n-1} d_{ik} p_{kj})$ of $D_{\mathcal{A}(n)}\ P_{\mathcal{A}}$,

$$\sum_{k=1}^{n-1} d_{ik} p_{kj} = d_{iw} p_{wj} + d_{iz} p_{zj}$$

so that $\sum_{k=1}^{n-1} d_{ik} p_{kj} = 0$ if $i \neq j$, and $\sum_{k=1}^{n-1} d_{ik} p_{kj} = 1$ if $i = j$. □

Lemma B.1. [5, Lemma 2.15] *If $\mathcal{G}$ is connected, then*

$$I - D_{\mathcal{A}} \, D_{\mathcal{A}}^+ = \frac{1}{n} \mathbb{1}\mathbb{1}^T$$

Lemma B.2. *The element $l_v \in \mathcal{L}$ is the length of the path from v to the root.*

Proof. Recall the definition of s and P, if $p_{ik} \neq 0$, then $p_{ik} = \pm 1$, $\forall \, 1 \leq i \leq n - 1$.

$$\text{sign}(p_{ik}) = -\text{sign}(s_i)$$

and $s_i = \pm 1$, consider the i-element in $\mathcal{L}$, which is $p_{ik}s_i$. Then $p_{ik} = 0$ unless edge e_i exists in the path from v_k to the root. Thus,

$$p_{ik}s_i = -1, \quad \text{if} \;\; p_{ik} \neq 0.$$

and

$$l_k = \sum_{i=1}^{n-1} w_i p_{ik} s_i$$

Then we can say that $\mathcal{L} = -P^T W s$ is a vector such that all the elements inside are positive and l_k is the length of the path from v_k to the root. $\qquad\square$

Without loss of generality, let v_1 be the root of the tree $\mathcal{G}(\mathcal{A})$ and n is the number of connected vertices in $\mathcal{G}$. We have

$$J = I - \frac{1}{n} \mathbb{1}\mathbb{1}^T = D_{\mathcal{A}} D_{\mathcal{A}}^+. \tag{B.1}$$

Denote $D_{\mathcal{A}'}^+$ is the pseudoinverse of the reduced incidence matrix $\mathcal{G}$, where the column cor-

responding to the root is removed. From Eq.(B.1), we have

$$D^+_{\mathcal{A}'} = D^{-1}_{\mathcal{A}'}(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T) = P_{\mathcal{A}}J.$$

Therefore, the pseudoinverse of $D_{\mathcal{A}}$ and it transpose reads

$$D^+_{\mathcal{A}} = \begin{bmatrix} -\frac{1}{n} P\mathbb{1} & PJ \end{bmatrix}_{(n-1)\times n}, \quad (D^+_{\mathcal{A}})^T = \begin{bmatrix} (-\frac{1}{n} P\mathbb{1})^T \\ (PJ)^T \end{bmatrix}_{n\times(n-1)}.$$

B.2 Proof of Proposition 3.2

For simplicity, we drop the iteration subscript k in our derivations. $D_{\mathcal{A}}$ is the incidence matrix formed by the edges in the active set. The graph formed by $\mathcal{A}$ consists of two disjoint trees $T^{(s)}$, $T^{(t)}$, and the set of isolated vertices Ω. We decompose the rows of matrix $D_{\mathcal{A}}$ into rows corresponding to these three subsets and express $D_{\mathcal{A}}$ and $D^+_{\mathcal{A}}$ according to

$$D_{\mathcal{A}} = \begin{bmatrix} D_{\mathcal{A}^{(s)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D_{\mathcal{A}^{(t)}} \end{bmatrix}, \quad \text{and } D^+_{\mathcal{A}} = \begin{bmatrix} D^+_{\mathcal{A}^{(s)}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & D^+_{\mathcal{A}^{(t)}} \end{bmatrix},$$

with $D_{\mathcal{A}^{(s)}}, D_{\mathcal{A}^{(t)}}$ are the incidence matrix for the tree $T^{(s)}$ and $T^{(t)}$ respectively, and $\mathbf{0}$ represents all-zeros matrices of appropriate dimensions. We use this expression and Lemma 3.1 to compute a and b by their definition (3.10)

$$a = (Q^T_{\mathcal{A}}Q_{\mathcal{A}})^+ Q^T_{\mathcal{A}}y = Q^+_{\mathcal{A}}y = W_{\mathcal{A}}D^+_{\mathcal{A}}y$$

$$= W_{\mathcal{A}} \begin{bmatrix} D^+_{\mathcal{A}^{(s)}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & D^+_{\mathcal{A}^{(t)}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{|T^{(s)}|}W_{\mathcal{A}^{(s)}}P^{(s)}\mathbb{1}_{|T^{(s)}|} \\ +\frac{1}{|T^{(t)}|}W_{\mathcal{A}^{(t)}}P^{(t)}\mathbb{1}_{|T^{(t)}|} \end{bmatrix},$$

142

where $P^{(s)}$ and $P^{(t)}$ are the path matrix for tree $T^{(s)}$ and $T^{(t)}$ respectively. Then,

$$
\begin{aligned}
b = (Q_{\mathcal{A}}^T Q_{\mathcal{A}})^+ s &= W_{\mathcal{A}} D_{\mathcal{A}}^+ (D_{\mathcal{A}}^T)^+ W_{\mathcal{A}} s \\
&= \begin{bmatrix} -W_{\mathcal{A}^{(s)}} \left(P^{(s)} \mathcal{L}^{(s)} - \frac{1}{|T^{(s)}|} P^{(s)} \mathbb{1}_s \mathbb{1}_s^T \mathcal{L}^{(s)} \right) \\ +W_{\mathcal{A}^{(t)}} \left(P^{(t)} \mathcal{L}^{(t)} - \frac{1}{|T^{(t)}|} P^{(t)} \mathbb{1}_t \mathbb{1}_t^T \mathcal{L}^{(t)} \right) \end{bmatrix},
\end{aligned}
$$

where $\mathcal{L}^{(s)} \triangleq -(P^{(s)})^T W_{\mathcal{A}^{(s)}} s$ is a vector of size $|T^{(s)}|$ corresponding to vertices in the tree $T^{(s)}$. The component of $\mathcal{L}^{(s)}$, corresponding to vertex $v \in T^{(s)}$, is equal to $l_v^{(s)}$, i.e. the length of the path from v to the root s. The vector $\mathcal{L}^{(t)} \triangleq (P^{(t)})^T W_{\mathcal{A}^{(t)}} s$ has a similar interpretation, but for vertices of tree $T^{(t)}$ and the difference of the definitions of $\mathcal{L}^{(s)}$ and $\mathcal{L}^{(t)}$ is resulted in the different signs of the two elements in b.

Putting the results for vectors a and b together, the ratio a_j/b_j for $e_j \in \mathcal{A}^{(s)}$ is

$$
\frac{a_j}{b_j} = \frac{\frac{1}{|T^{(s)}|} w_j |R_j|}{w_j \left(\sum_{v \in R_j} l_v^{(s)} - \frac{|R_j|}{|T^{(s)}|} \sum_{v \in T^{(s)}} l_v^{(s)} \right)},
$$

where R_j is the set of non-zero components of the jth row of $P^{(s)}$ since the same edge e_j shares the same direction along all the paths in a tree. This concludes our proof for $e_j \in \mathcal{A}^{(s)}$. The derivation for $e_j \in \mathcal{A}^{(t)}$ is similar.

B.3 Proof of Proposition 3.1

By definition of joining time (3.11), for all $e_j \in \mathcal{E}$, reads

$$
t_j^{\text{join}} = \frac{\frac{1}{w_j} D_j^T (Q_{\mathcal{A}} a - y)}{\frac{1}{w_j} D_j^T (Q_{\mathcal{A}} b) \pm 1} = \frac{D_j^T (Q_{\mathcal{A}} a - y)}{D_j^T (Q_{\mathcal{A}} b) \pm w_j}, \tag{B.2}
$$

with the choice $\pm$ dictated by t_j^{join} taking positive value. We now obtain expressions for the terms in parentheses. First, the term $(Q_{\mathcal{A}}a - y)$ in the numerator equals to

$$D_{\mathcal{A}}D_{\mathcal{A}}^+ y - y = \begin{bmatrix} D_{\mathcal{A}^{(s)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D_{\mathcal{A}^{(t)}} \end{bmatrix} \begin{bmatrix} D_{\mathcal{A}^{(s)}}^+ & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & D_{\mathcal{A}^{(t)}}^+ \end{bmatrix} y - y$$

$$= \left[-\frac{1}{|T^{(s)}|} \mathbb{1}_s, \ \mathbf{0}_{|\Omega|}, \ +\frac{1}{|T^{(t)}|} \mathbb{1}_t \right]^T, \tag{B.3}$$

where we used $DD^+ = I - \frac{1}{\mathbb{1}^T \mathbb{1}} \mathbb{1}\mathbb{1}^T$ for incidence matrices $D_{\mathcal{A}^{(s)}}$ and $D_{\mathcal{A}^{(t)}}$ of the two disjoint trees [5, Lemma 2.15], and $\mathbb{0}_{|\Omega|}$ denotes all-zero (column) vectors of size $|\Omega|$. Second, the term $(Q_{\mathcal{A}}b)$ in the denominator equals to $(D_{\mathcal{A}}D_{\mathcal{A}}^+ (D_{\mathcal{A}}^T)^+ W_{\mathcal{A}}s)$ and moreover, it equals to

$$(D_{\mathcal{A}}^+)^T W_{\mathcal{A}}s = \begin{bmatrix} -\mathcal{L}^{(s)} + \frac{1}{|T^{(s)}|} \mathbb{1}_s \mathbb{1}_s^T \mathcal{L}^{(s)} \\ \mathbb{0}_{|\Omega|} \\ \mathcal{L}^{(t)} - \frac{1}{|T^{(t)}|} \mathbb{1}_t \mathbb{1}_t^T \mathcal{L}^{(t)} \end{bmatrix}. \tag{B.4}$$

Evaluating the expression in (B.2) for edge $e_j = (v_1, v_2) \in \mathcal{A}^c$ and without loss of generality, assume the assigned direction is from v_1 to v_2, the joining time is categorized into three separate cases.

i) If $(v_1, v_2) \in \Omega^2 \cup T^{(s)^2} \cup T^{(t)^2}$, i.e., the two ends of e_j belongs to or isolated from the trees, we have $t_j^{\text{join}} = 0$ by the numerator (B.3) and the definition of D_j^T.

ii) If $(v_1, v_2) \in (T^{(s)} \times \Omega) \cup (T^{(t)} \times \Omega)$, Without loss of generality, assume $v_1 \in T^{(s)}$, the joining now reads

$$t_j^{\text{join}} = \left(|T^{(s)}| l_{v_1}^{(s)} - \mathbb{1}^T \mathcal{L}_v^{(s)} \pm |T^{(s)}| w_j \right)^{-1}$$

$$= \frac{1}{|T^{(s)}| (l_{v_1}^{(s)} + w_j) - \mathbb{1}^T \mathcal{L}_v^{(s)}}$$

144

where we always add weight w to l_{v_1} to guarantee that $t_{\text{join}} > 0$ by its definition in Eq. (3.11).

iii) If $(v_1, v_2) \in T^{(s)} \times T^{(t)}$, we have

$$t_j^{\text{join}} = \frac{1/|T^{(s)}| + 1/|T^{(t)}|}{(l_{v_1}^{(s)} + l_{v_2}^{(t)} + w_j) - \frac{1}{|T^{(s)}|} \mathbb{1}^T \mathcal{L}_v^{(s)} - \frac{1}{|T^{(t)}|} \mathbb{1}^T \mathcal{L}_v^{(s)}}.$$

B.4   Proof of Lemma 3.2

The proof is based on the sufficient condition for the uniqueness of the lasso solution[102, Lemma 2], which states that *"For any β, Q, and $\lambda > 0$, if $\text{null}(Q_{\mathcal{A}}) = \{0\}$ (or equivalently if $\text{rank}(Q_{\mathcal{A}}) = |\mathcal{A}|$), the lasso solution and the active set $\mathcal{A}$ are always unique."* Thus, we prove the lasso solution $\beta(\lambda)$ is unique for every $\lambda > 0$ by showing that $\text{rank}(Q_{\mathcal{A}}) = |\mathcal{A}|$ is true for every $Q_{\mathcal{A}}$.

In Section 3.4, we showed that the active set forms two disjoint trees (Lemma 3.3 and 3.4). Hence,

$$Q_{\mathcal{A}} = \begin{bmatrix} Q_{\mathcal{A}^{(s)}}, & Q_{\mathcal{A}^{(t)}} \end{bmatrix} = \begin{bmatrix} D_{\mathcal{A}^{(s)}} W_{\mathcal{A}^{(s)}}^{-1}, & D_{\mathcal{A}^{(t)}} W_{\mathcal{A}^{(t)}}^{-1} \end{bmatrix},$$

with $D_{\mathcal{A}_k^{(t)}}$ and $D_{\mathcal{A}_k^{(t)}}$ are incidence matrices of two trees positively weighted by the weighted matrices $W_{\mathcal{A}_k^{(s),(t)}}$. The null space (*kernel)* of the incidence matrix of a tree is empty because there is no cycle in the tree by its definition, so the rank of its incidence matrix is equal to the number of columns, i.e., the rank of $Q_{\mathcal{A}}$ is equal to $|\mathcal{A}|$. Therefore, the sufficient condition for the unique lasso solution is satisfied.