

Non-conservative optimal transport

Gabriela Kováčová*, Georg Menz[†], Niket Patel^{†‡}

August 24, 2026

Abstract

Motivated by optimal re-balancing of a portfolio, we formalize an optimal transport problem in which the transported mass is scaled by a mass-change factor depending on the source and destination. This allows direct modeling of the creation or destruction of mass. We discuss applications and position the framework alongside unbalanced, entropic, and unnormalized optimal transport. On complete, separable metric spaces, we establish the existence of optimal transport plans for costs that are lower semi-continuous and bounded from below and for continuous mass-change factors satisfying two-sided bounds. Under an additional finite-cost condition, we prove equality of the primal and dual values; attainment of the dual problem is treated separately under compactness. Under additional structural assumptions, the existence of optimal maps is established in two central regimes, namely perturbative mass-change and quadratic mass-loss. For power costs $c(x, y) = |y - x|^a$, we derive a dynamical-plan analogue of the Benamou–Brenier formulation.

MSC2020: Primary 49Q22, 49N15; Secondary 91G10, 91G80.

Keywords: Optimal Transport, Duality Theory, Portfolio Rebalancing.

1 Introduction

Optimal transport (OT) originates with Monge’s 1781 sand-moving problem [Mon81] and Kantorovich’s 1942 relaxation [Kan42], and has since blossomed via the Benamou–Brenier dynamic formulation [BB00] into the geometry of Wasserstein spaces. Over the last two decades, advances in duality, regularity, and scalable algorithms have made OT a central tool across analysis, probability, economics, imaging, and machine learning. For a thorough introduction to optimal transport we refer to the standard references [AGS08, V⁺08, San15]. In the classical

*Department of Engineering, Reykjavík University, Iceland
Email: gabrielak@ru.is

[†]Department of Mathematics, University of California Los Angeles, CA, USA
Email: gmenz@math.ucla.edu

[‡]Center for Data Science, New York University, NY, USA
Email: nnp5656@nyu.edu

Kantorovich problem of optimal transport the goal is to find a coupling π of two given probability measures μ and ν that minimizes a given cost functional

$$\inf_{\pi} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy).$$

Coupling means that the first marginal of π is given by μ and the second marginal by ν , i.e. for all test functions ξ and ζ

$$\int \xi(x) \pi(dx, dy) = \int \xi(x) \mu(dx) \quad \text{and} \quad \int \zeta(y) \pi(dx, dy) = \int \zeta(y) \nu(dy). \quad (1)$$

One of the limitations of the original optimal transport formulation is its rigidity in the marginal constraint (1). In particular, this constraint entails that the transport conserves the mass. Many applications require more flexibility and, therefore, variants of optimal transport allowing flexibility in the marginal constraints were developed and successfully applied in recent years, see Section 4.1 for a more detailed discussion including references. Building on this trajectory, we introduce the framework of non-conservative optimal transport and develop its basic theory, rather than aiming for the greatest possible generality.

The framework originated when the authors studied a problem from financial mathematics, namely how to rebalance a portfolio of assets in an optimal way, see Section 2.1. However, studying this problem revealed a deeper structure. When exchanging one asset for another, a small percentage of its value is lost to trading costs (e.g. bid-ask spread and trading fees). Mass is lost as the value of the portfolio after rebalancing is smaller than the value before. The shape of the target distribution is still fixed as one strives for obtaining the correct distribution of wealth across assets.

Motivated by this example, non-conservative optimal transport generalizes the classical optimal transport problem by two tweaks. First, one introduces a mass-change factor $m(x, y)$ that models whether a mass transported from x to y is gained ($m(x, y) > 1$), conserved ($m(x, y) = 1$) or destroyed ($m(x, y) < 1$). Second, one weakens the second marginal condition by prescribing only the shape of the transported measure. this means

$$\int \xi(x) \pi(dx, dy) = \int \xi(x) \mu(dx) \quad \text{and} \quad \frac{\int \zeta(y) m(x, y) \pi(dx, dy)}{\int m(x, y) \pi(dx, dy)} = \int \zeta(y) \nu(dy)$$

for all test functions ξ and ζ . Classical optimal transport is recovered when $m \equiv 1$. This framework allows for simple and direct modeling of transports that do not conserve mass. In Section 3 we discuss more examples like logistics with spoilage/yield, biochemical conversions, as well as economic scenarios like the influence of tariffs on the optimal distribution of goods, or the influence of retention rates onto the optimal distribution of jobs in an organization. To the best of our knowledge, this simple framework to describe non-conservative

optimal transport is new and not contained in the existing literature, though it can be described as an optimization problem over a family of unbalanced optimal transport problems. We refer to Section 4 for more details. As the framework of non-conservative optimal transport stays close to the original Kantorovich optimal transport problem, we are able to study fundamental questions (like duality, existence of optimal maps, dynamic formulations) with similar methods, and come to satisfying answers.

Summarizing, the article provides: In Section 2, a *detailed motivating example* from financial mathematics. In Section 3, the *formulation* of *non-conservative optimal transport* with cost $c(x, y)$ and efficiency $m(x, y)$, reducing to classical OT when $m \equiv 1$ (see Section 3.1). The *existence* of optimal plans on complete, separable metric spaces for lower semi-continuous costs and continuous mass-change factors satisfying two-sided bounds (see Section 3.1). The *strong duality of values* under an additional finite-cost condition and the *attainment of the dual problem* under compactness (see Sections 3.2 and 3.3). The *existence* of *optimal maps* on compact subsets of $\mathbb{R}^n$ in perturbative ($m \approx 1$) and quadratic-cost regimes (see Section 3.4). A *generalized Benamou–Brenier formula* in terms of dynamical plans, without assuming *a-priori* vector-field regularity or the existence of an optimal transport map (see Section 3.5). In Section 4, the relation to other optimal transport frameworks allowing for mass creation or destruction.

The regularity of optimal transport maps in the non-conservative framework is not discussed in this article and remains an open problem.

2 Motivating example: Optimal rebalancing of portfolios.

We open this paper with a motivating example from financial mathematics. Consider an investor whose investment strategy (or a target) is to hold $100 \cdot \nu_i\%$ of their portfolio value in asset i . Throughout, portfolios are long-only, i.e. weights satisfy $\nu_i \geq 0$. If their currently held position does not correspond to this target proportion, then they need to buy and sell assets to achieve their target. We refer to this as rebalancing the investor’s portfolio. Since trading incurs costs, we aim to find an optimal rebalancing trade. Viewed as a transport problem, rebalancing moves portfolio value from the current to the target distribution; trading costs destroy value in transit, and only the *shape* of the target distribution is prescribed. These two features will lead us to *non-conservative optimal transport*.

We do not address the choice of the target portfolio — constructing optimal portfolios is a classical topic in financial mathematics; e.g. see [Won19] for a survey of optimal transport methods in portfolio construction — nor the questions of when and how often to rebalance. These and further extensions are discussed in Remark 2.15.

The rebalancing problem is non-trivial in the presence of fees or transaction costs, [Kab09] discusses the case of separable trading costs (e.g. markets with a

single numéraire). This manuscript extends optimal rebalancing to the case of general markets with multiple numéraires. This is a non-trivial task as trading costs are not separable.

In Subsection 2.1 we introduce the mathematical framework and state the optimal rebalancing problem. Then, in Subsection 2.2 we interpret the problem in the framework of optimal transport. This will motivate us to introduce a notion of non-conservative optimal transport.

2.1 Optimal rebalancing problem

We take a static, snap-shot view of a market modeled as a graph.

Definition 2.1. *A financial market is modeled by a weighted directed graph $\mathcal{G} = (V, E, P)$, where*

- *vertices $V = \{1, \dots, N\}$ represent assets on the market,*
- *a directed edge $(i, j) \in E$ indicates that it is possible to directly trade asset i into asset j ,*
- *and the weight $p : E \rightarrow (0, \infty)$ represents the price quoted in units of asset received per unit of asset sold. For an edge $(i, j) \in E$, one unit of asset i can be exchanged for $p_{(i,j)}$ units of asset j on the market.*

A market with a single numéraire (Example 2.5 below) will be considered as a special case. However, we do not assume the existence of a single numéraire in general. This allows one to consider portfolios held in assets traded across different international markets and currencies or various crypto currencies. We refer to Figure 1 for an illustration.

Assumption 2.2. We make the following assumptions about the market.

- A1 *Strong connectedness:* $\mathcal{G}$ is strongly connected.
- A2 *No arbitrage:* For every directed cycle Γ in $\mathcal{G}$ it holds $\prod_{e \in \Gamma} p_e \leq 1$.
- A3 *Unlimited liquidity and no price impact:* Assets are liquid and there is no limit on number of units that can be traded. Trading does not impact prices.
- A4 *No explicit fees:* Fees, if any, are incorporated in the quoted prices.

Remark 2.3. We are taking a snap-shot view of considering the market at a single time point with the aim of implementing the rebalancing in a single trade executed immediately. Such a trade may consist of several successive exchanges along the edges of the graph, all executed within the same period. This, alongside the assumption of no price impact, is a reasonable approximation when considering small portfolios (for extensions see Remark 2.15).

Regarding A4, it is possible to model linear (per share) fees as part of the bid-ask spread, but not flat fees.

If we start with one unit of asset i and trade along a cycle Γ , we hold $\prod_{e \in \Gamma} p_e$ units of asset i after the trade. In particular, for any pair of directly tradable assets i and j no arbitrage implies $p_{(i,j)} \leq \frac{1}{p_{(j,i)}}$.

For convenience, we extend the quoted prices p to the *best execution prices* $P : V \times V \rightarrow (0, \infty)$ among all ordered pairs of assets. We set $P_{(i,i)} := 1$ and

$$P_{(i,j)} := \max_{\Gamma: i \rightarrow j} \prod_{e \in \Gamma} p_e \quad \text{for } i \neq j,$$

where the maximum is taken over all directed paths Γ from i to j in $\mathcal{G}$. Consequently, $P_{(i,j)}$ is the best price obtainable by routing a trade through the market. Intuitively, by A1 paths between any pair of assets exist and by A2 the maximum is attained on paths without repeated vertices.

Lemma 2.4. *Let the market $\mathcal{G} = (V, E, p)$ satisfy A1 and A2. The best execution prices P are well defined and satisfy no arbitrage, i.e., for any directed cycle $\bar{\Gamma}$ in the completed graph $\bar{\mathcal{G}} = (V, V \times V, P)$ it holds $\prod_{e \in \bar{\Gamma}} P_e \leq 1$.*

Proof and further discussion is given in Appendix A.1.

Example 2.5. Consider a market with a single numéraire (w.l.o.g. asset 1), where assets $i = 2, \dots, N$ can only be traded directly for the numéraire. We refer to this as a *star-shaped market*. It is usual to quote prices in units of the numéraire as the bid and ask prices,

$$P_i^b = p_{(i,1)} \quad \text{and} \quad P_i^a = \frac{1}{p_{(1,i)}} \quad \text{for assets } i = 2, \dots, N.$$

For completeness we set $P_1^b = P_1^a := 1$. Rebalancing between all-long portfolios on such a star-shaped market with proportional transaction costs is analyzed in detail in [RX20]. After completion, this market quotes $P_{(i,j)} = P_i^b / P_j^a$ for all assets $i \neq j$: every trade is routed through the numéraire.

Definition 2.6. *A portfolio is given by a vector $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$, where x_i denotes the number of units of the asset i held in the portfolio.*

A trade (on $\bar{\mathcal{G}}$) is a map $\xi : V \times V \rightarrow [0, \infty)$ satisfying $\xi_{(i,j)} = 0$ for $i = j$. Here $\xi_{(i,j)}$ denotes the number of units of asset i that are traded for asset j at the best-execution price $P_{(i,j)}$. Ξ denotes the set of all trades.

A trade ξ changes the number of units of asset i held by

$$\Delta x_i(\xi) = \sum_{j=1}^N P_{(j,i)} \xi_{(j,i)} - \xi_{(i,j)}, \quad i = 1, \dots, N. \quad (2)$$

The aim is to rebalance the current portfolio $\mathbf{x}$ in such a way that the *target proportions* $\nu = (\nu_1, \dots, \nu_N)$ of value are achieved. To measure portfolio value, the investor fixes (‘true’ or ‘fair’) asset values, which we assume to be consistent with the quoted prices.

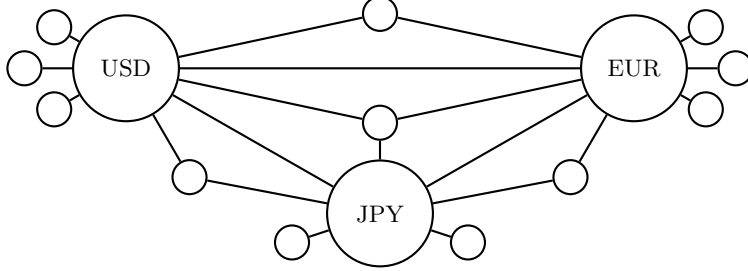


Figure 1: Illustration of a general financial market without a single numéraire. The vertices USD, JPY and EUR denote currencies used as numéraires at national markets. Vertices connected to several numéraires denote cross-traded assets, i.e. assets that are traded on several national markets.

Definition 2.7. A relative price vector $\mathbf{q} : V \rightarrow (0, \infty)$ is consistent with the financial market $\mathcal{G} = (V, E, p)$ if it holds

$$P_{(i,j)} \leq \frac{q_i}{q_j} \quad \text{for all } i, j \in V.$$

Remark 2.8. Since a consistent price vector is scale invariant, we can turn any asset i into a *hypothetical numéraire* by imposing $q_i = 1$.

Example 2.9. Once more, consider the *star-shaped market*, where the no arbitrage condition (A2) simplifies to

$$P_i^b \leq P_i^a \quad i = 2, \dots, N.$$

Any choice of $P_i^b \leq q_i \leq P_i^a$ leads to a consistent price vector, one such example is the mid price. Observe that in presence of a non-trivial bid-ask spread, the consistent price vector is not unique (despite fixed numéraire).

Proposition 2.10. Let the financial market $\mathcal{G} = (V, E, P)$ satisfy Assumption A1 (strong connectedness). A consistent price vector exists if and only if the market satisfies Assumption A2 of no arbitrage.

Proof of Proposition 2.10 is given in Appendix A.1. Moreover, Remark A.2 comments on computation of a consistent price vector.

Assumption 2.11. In addition to A1-A4 we assume the following.

A5 *Long only portfolio:* Current portfolio $\mathbf{x}$ consists only of long positions, $x_i \geq 0$ for $i = 1, \dots, N$, and is non-trivial, $\mathbf{x} \neq 0$.

A6 *Target proportions:* A vector ν satisfying $\sum_{i=1}^N \nu_i = 1$ and $\nu_i \geq 0$ for $i = 1, \dots, N$ is given.

A7 *Consistent price vector:* A consistent price vector $\mathbf{q}$ is fixed. Without loss of generality, we impose $q_1 = 1$ and treat asset 1 as the hypothetical numéraire.

In what follows we use the fixed consistent price vector $\mathbf{q}$ to compute value of a portfolio expressed in units of the hypothetical numéraire. Consider the current portfolio $\mathbf{x}$ and a trade ξ . The value held in asset i before and after the trade, respectively, are $q_i x_i$ and $q_i (x_i + \Delta x_i(\xi))$. Note that while we use the consistent prices $\mathbf{q}$ to evaluate positions, trades are executed at the market prices P . Therefore, a trade decreases the value of the portfolio. We denote the *cost of (rebalancing) trade* ξ by

$$C(\xi) := - \sum_{i=1}^N q_i \Delta x_i(\xi).$$

We aim to find a trade ξ such that the value of portfolio $\mathbf{x} + \Delta \mathbf{x}(\xi)$ is distributed according to the target proportions ν and to do so at the lowest cost possible. For an illustration of the rebalancing problem we refer to Figure 2. This yields the following optimization problem.

Problem 2.12. We identify the optimal rebalancing trade as the solution of

$$\begin{aligned} & \underset{\xi \in \Xi}{\text{minimize}} && C(\xi) \\ & \text{subject to} && x_i + \Delta x_i(\xi) \geq 0, && i = 1, \dots, N, \\ & && \nu_i = \frac{q_i (x_i + \Delta x_i(\xi))}{\sum_{k=1}^N q_k (x_k + \Delta x_k(\xi))} && i = 1, \dots, N. \end{aligned}$$

We denote the feasible set of this problem by $\Xi(\mathbf{x}, \nu)$, highlighting its dependence on the current portfolio $\mathbf{x}$ and the target proportions ν ; it also depends on the consistent price vector $\mathbf{q}$, which we regard as fixed throughout.

Remark 2.13. Since $\sum_{k=1}^N q_k (x_k + \Delta x_k(\xi)) = \sum_{k=1}^N q_k x_k - C(\xi)$, minimizing the rebalancing cost is equivalent to maximizing the post-rebalancing value of the portfolio,

$$\arg \min_{\xi \in \Xi(\mathbf{x}, \nu)} C(\xi) = \arg \max_{\xi \in \Xi(\mathbf{x}, \nu)} \sum_{k=1}^N q_k (x_k + \Delta x_k(\xi)).$$

Theorem 2.14. *Let the assumptions A1-A7 hold. Solution to the optimal rebalancing Problem 2.12 exists.*

In the next subsection, we show that the rebalancing problem corresponds to a *non-conservative optimal transport*: by Proposition 2.18, optimal rebalancing trades correspond to optimal plans of the associated transport problem, whose general theory — including the existence of optimal plans (Theorem 3.3) — is developed in Section 3. The proof of Theorem 2.14 is provided in Appendix A.2 through a series of partial results.

Remark 2.15 (Extensions of optimal re-balancing). Solving the optimal rebalancing Problem 2.12 is only a first (theoretical) step. When trading larger

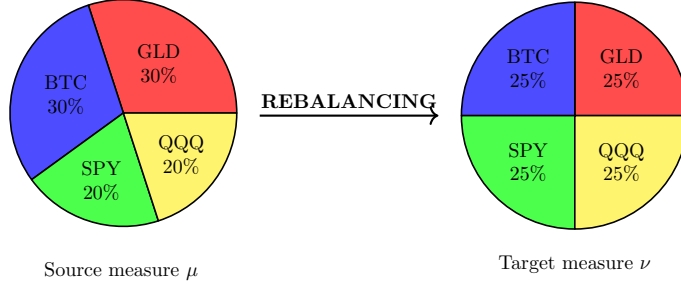


Figure 2: Illustration of the rebalancing process. The goal is to sell and buy assets such that the wealth of the new portfolio is distributed according to the target proportions (or measure) ν .

quantities, price impact cannot be ignored and leads to a non-linear optimization problem that is either convex or nonconvex depending on how price impact is modeled. Re-balancing all at once might then be suboptimal, and spreading trades out over time requires modeling dynamic, correlated asset prices and price impacts. In particular, the price processes should then be modeled stochastically; a proper development of such a dynamic model is beyond the scope of this paper. Optimal rebalancing can then be approached via the dynamic formulation of optimal transport (cf. Section 3.5). Finally, the question of how often and when one should re-balance (see for example [DN90, MP95, HH13]) leads to tracking problems: transaction costs create a no-trade region in which the marginal benefit of re-balancing is smaller than the additional cost, and many tracking problems can be cast as entropic optimal transport, where the second marginal condition is relaxed to trade off precision against cost (cf. Section 4.2).

2.2 Rebalancing and optimal transport

Let us now show that the optimal portfolio rebalancing problem is closely connected to optimal transport on a discrete state space $\Omega = V = \{1, \dots, N\}$: rebalancing transports the current distribution of portfolio value into the target distribution while minimizing trading costs. Since trades destroy value, mass is lost during transport — in contrast to classical optimal transport, where mass is conserved. We call the resulting problem a *non-conservative optimal transport* problem.

Given the current portfolio position $\mathbf{x}$ and investor's consistent price vector $\mathbf{q}$, the proportion of current portfolio value held in asset i is

$$\mu_i = \frac{q_i \cdot x_i}{\sum_{k=1}^N q_k \cdot x_k} \quad \text{for } i = 1, \dots, N.$$

The current proportions $\mu = (\mu_1, \dots, \mu_N)$ as well as the target proportions $\nu = (\nu_1, \dots, \nu_N)$ are probability distributions on Ω . We consider transport

plans $\pi : \Omega \times \Omega \rightarrow [0, 1]$ that distribute the mass of μ ; the target ν is prescribed through the *delivered* values, as formalized in Problem 2.17 below. The value $\pi(i, j)$ should represent *proportion of portfolio's value* that should be transported (traded) from asset i to asset j . The cost functional $c : \Omega \times \Omega \rightarrow [0, 1]$ capturing the proportional loss of portfolio value when exchanging (trading) assets is given by

$$c(i, j) = 1 - \frac{q_j}{q_i} P_{(i, j)};$$

note that $c \geq 0$ by the consistency of $\mathbf{q}$, that $c(i, i) = 0$ (holding an asset is free), and that $c < 1$ since all best-execution prices are positive.

Remark 2.16. Let us relate transport plan π and cost c to trade ξ and its cost $C(\xi)$. In view of (2.1), the cost of a trade equals

$$C(\xi) = \sum_{i=1}^N q_i \sum_{j=1}^N (\xi_{(i, j)} - P_{(j, i)} \xi_{(j, i)}) = \sum_{i, j=1}^N (q_i - q_j P_{(i, j)}) \xi_{(i, j)}. \quad (3)$$

Since the transport plan π replaces the trade ξ as the variable, $\pi(i, j)$ should be proportional to $q_i \xi_{(i, j)}$, the value traded from asset i into asset j . Consequently, the cost associated with $\pi(i, j)$ is $c(i, j) = \frac{1}{q_i} (q_i - q_j P_{(i, j)})$.

To formulate a transport problem corresponding to the optimal portfolio rebalancing, we need to account for losses of mass during the transport. For this purpose, we introduce the mass-change function $m : \Omega \times \Omega \rightarrow (0, 1]$ given by

$$m(i, j) := \frac{q_j}{q_i} P_{(i, j)} = 1 - c(i, j).$$

Indeed, one unit of asset i (with value q_i) can be traded for $P_{(i, j)}$ units of asset j (with value $q_j P_{(i, j)}$), so exactly the proportion $m(i, j)$ of the traded value is preserved. In terms of the plan, the value delivered to asset j is therefore $\sum_i m(i, j) \pi(i, j)$; the marginal condition $\sum_i \pi(i, j) = \nu_j$ of a classical coupling would prescribe the dispatched value instead and ignore the loss. Rebalancing prescribes the proportions of the delivered values, while their total is not fixed a priori.

Problem 2.17. Now, we consider problem

$$\min_{\pi \in \Gamma_m(\mu, \nu)} \sum_{i, j=1}^N c(i, j) \cdot \pi(i, j),$$

where

$$\Gamma_m(\mu, \nu) := \left\{ \pi : \Omega \times \Omega \rightarrow [0, 1] : \sum_{j=1}^N \pi(i, j) = \mu_i \quad \text{and} \quad \frac{\sum_{i=1}^N m(i, j) \pi(i, j)}{\sum_{k=1}^N \sum_{i=1}^N m(i, k) \pi(i, k)} = \nu_j \right\}.$$

Problem 2.17 is a modification of the optimal transport problem with mass being lost during the transport, that is, a non-conservative optimal transport problem. Since $m > 0$, the constraint set $\Gamma_m(\mu, \nu)$ is non-empty and an optimal solution exists, see Remark 3.2 and Theorem 3.3 below. Section 3 studies the general case of a non-conservative optimal transport problem. To conclude this subsection, we relate the optimal solutions of the rebalancing Problem 2.12 and of the transport Problem 2.17.

Proposition 2.18. *Let Assumptions A1-A7 hold and denote $v := \sum_{k=1}^N q_k x_k$. Call a trade $\xi \in \Xi$ direct if $\sum_{j \neq i} \xi_{(i,j)} \leq x_i$ for every asset i , i.e. if it sells no more of any asset than initially held. Then some optimal solution of Problem 2.12 is a direct trade. Moreover, a direct trade $\xi \in \Xi(\mathbf{x}, \nu)$ is an optimal solution of Problem 2.12 if and only if π^ξ given by*

$$\pi^\xi(i, j) = \begin{cases} \frac{q_i}{v} \xi_{(i,j)} & \text{if } i \neq j, \\ \frac{q_i}{v} \left(x_i - \sum_{k \neq i} \xi_{(i,k)} \right) & \text{if } i = j, \end{cases}$$

is an optimal solution of Problem 2.17; every optimal solution of Problem 2.17 arises in this way.

Proof of Proposition 2.18 is given in Appendix A.3 following two auxiliary lemmas. In financial terms, directness is a minimal-turnover property: among the optimal rebalancing trades, those minimizing the total sold value are direct (Lemma A.8 b)).

3 Non-conservative optimal transport

This section develops the framework and theory of non-conservative optimal transport. In Section 3.1 we propose and solve the non-conservative Kantorovich problem and discuss applications. In Section 3.2, we develop and solve the dual-problem. Strong duality is deduced in Section 3.3. The existence of optimal transport maps is discussed in Section 3.4 and the dynamic formulation is treated in Section 3.5.

3.1 The non-conservative Kantorovich Problem.

Let us consider two complete, separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$. For a metric space $\mathcal{Z}$, we denote its Borel σ -algebra by $\mathcal{B}(\mathcal{Z})$, the set of Borel probability measures on $\mathcal{Z}$ by $\mathcal{P}(\mathcal{Z})$, and the space of bounded and continuous real-valued functions on $\mathcal{Z}$ by $\mathcal{C}_b(\mathcal{Z})$. Let the mass-change factor $m : \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty)$ be Borel measurable. It models the percentage of mass that is lost (if $m(x, y) \leq 1$) or gained (if $m(x, y) > 1$) when transporting some mass from x to y . The loss or gain of mass is *proportional* to the amount of transported mass $\pi(x, y)$. The mass-change factor $m(x, y)$ depends only on the starting point x and endpoint y , and not on other factors like the amount of mass transported, the route or the velocity of the transport.

We consider a source probability measure $\mu \in \mathcal{P}(\mathcal{X})$ and transport its mass according to a transport plan $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. As mass is lost or gained during the transport, the transported measure is distributed according to

$$\mathcal{B}(\mathcal{Y}) \ni B \mapsto \int_{\mathcal{X} \times B} m(x, y) \pi(dx, dy),$$

which generically is not a probability measure. However, we require that the remaining mass after transport will be distributed *proportionally* to a given target probability measure $\nu \in \mathcal{P}(\mathcal{Y})$. Hence, the set of non-conservative transport plans w.r.t. source measure μ and target measure ν is given by

$$\Gamma_m(\mu, \nu) = \left\{ \begin{array}{l} \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \text{ such that} \\ 0 < \int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi(dx, dy) < \infty, \\ (\text{Proj}_1)_\# \pi = \mu, \\ (\text{Proj}_2)_\#(m\pi) = \left(\int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi(dx, dy) \right) \nu \end{array} \right\}. \quad (4)$$

This means that the first marginal is distributed as μ , i.e. $(\text{Proj}_1)_\# \pi = \mu$. The second marginal satisfies the relation $(\text{Proj}_2)_\#(m\pi) = \left(\int m d\pi \right) \cdot \nu$.

Definition 3.1 (Kantorovich problem in non-conservative optimal transport). *Given Borel-measurable functions $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ and $m : \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty)$, the non-conservative Kantorovich problem is to minimize the functional*

$$\pi \mapsto \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy)$$

over all $\pi \in \Gamma_m(\mu, \nu)$. Equivalently, its optimal value is

$$\inf_{\pi \in \Gamma_m(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy). \quad (\text{KP})$$

Remark 3.2 (Feasibility of (KP)). Under general non-degeneracy assumptions the set $\Gamma_m(\mu, \nu)$ is non-empty. More precisely, if the probability measure $\tilde{\nu}$ given by

$$\tilde{\nu}(B) := \frac{1}{Z} \int_B \frac{1}{\int_{\mathcal{X}} m(x, y) \mu(dx)} \nu(dy) \quad \text{with } 0 < Z < \infty \text{ such that } \tilde{\nu}(\mathcal{Y}) = 1$$

is well-defined, it holds $\mu \otimes \tilde{\nu} \in \Gamma_m(\mu, \nu)$.

Non-emptiness of $\Gamma_m(\mu, \nu)$ will not be assumed globally: under the two-sided bounds (3.3) below it holds automatically, since $\tilde{\nu}$ is then well-defined (see Theorem 3.3), and in the one setting where these bounds are relaxed it is assumed explicitly (Section 3.5). Let us turn to the solution of (KP). To

find a minimizer via the direct method one needs to establish tightness for a sequence of minimizing transport plans. To transfer tightness of ν to the second marginal we assume that the mass factor is uniformly bounded from above and from below. Under this assumption the non-conservative Kantorovich problem admits a solution on complete, separable metric spaces. It is clear that the assumptions could be relaxed further, but this is not the main objective of this article.

Theorem 3.3 (Solution of (KP)). *Suppose that the cost function c is lower semi-continuous and bounded from below, and that the mass-change factor m is continuous. Assume that there are constants $0 < \underline{m} \leq \overline{m} < \infty$ such that*

$$\underline{m} \leq m(x, y) \leq \overline{m} \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (5)$$

Then the set $\Gamma_m(\mu, \nu)$ is non-empty. Moreover, Problem 3.1 admits a minimizer $\pi^ \in \Gamma_m(\mu, \nu)$, i.e.*

$$\inf_{\pi \in \Gamma_m(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy) = \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi^*(dx, dy).$$

Proof of Theorem 3.3. First, let us note that (3.3) implies that the measure $\tilde{\nu}$ from the preceding remark is well-defined. Therefore, $\mu \otimes \tilde{\nu} \in \Gamma_m(\mu, \nu)$ and the admissible set is non-empty.

Let $(\pi_n)_{n \in \mathbb{N}} \subset \Gamma_m(\mu, \nu)$ be a minimizing sequence for (KP), and set

$$q_n := \int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi_n(dx, dy).$$

It follows from (3.3) that $\underline{m} \leq q_n \leq \overline{m}$. Additionally, for every Borel set $B \in \mathcal{B}(\mathcal{Y})$ it holds

$$\pi_n(\mathcal{X} \times B) \leq \frac{1}{\underline{m}} \int_{\mathcal{X} \times B} m(x, y) \pi_n(dx, dy) = \frac{q_n}{\underline{m}} \nu(B) \leq \frac{\overline{m}}{\underline{m}} \nu(B). \quad (6)$$

Let $\varepsilon > 0$. Since μ and ν are probability measures on complete, separable metric spaces, there are compact sets $K_{\mathcal{X}} \subset \mathcal{X}$ and $K_{\mathcal{Y}} \subset \mathcal{Y}$ such that

$$\mu(\mathcal{X} \setminus K_{\mathcal{X}}) < \frac{\varepsilon}{2} \quad \text{and} \quad \nu(\mathcal{Y} \setminus K_{\mathcal{Y}}) < \frac{\underline{m}}{\overline{m}} \frac{\varepsilon}{2}.$$

Using that the first marginal of π_n is given by μ and applying (3.1), we obtain

$$\pi_n((\mathcal{X} \times \mathcal{Y}) \setminus (K_{\mathcal{X}} \times K_{\mathcal{Y}})) \leq \mu(\mathcal{X} \setminus K_{\mathcal{X}}) + \pi_n(\mathcal{X} \times (\mathcal{Y} \setminus K_{\mathcal{Y}})) < \varepsilon.$$

Hence, the sequence $(\pi_n)_{n \in \mathbb{N}}$ is tight. By Prokhorov's theorem, there is a subsequence, which we again denote by $(\pi_n)_{n \in \mathbb{N}}$, and a probability measure π^* such that π_n converges weakly to π^* .

It remains to verify that $\pi^* \in \Gamma_m(\mu, \nu)$. The first marginal condition follows directly from weak convergence. Set $q^* := \int m d\pi^*$ and let $\zeta \in \mathcal{C}_b(\mathcal{Y})$.

As m is bounded and continuous, both functions $(x, y) \mapsto m(x, y)$ and $(x, y) \mapsto \zeta(y)m(x, y)$ are bounded and continuous. Consequently, $q_n \rightarrow q^*$ and

$$\begin{aligned} \int \zeta(y)m(x, y)\pi^*(dx, dy) &= \lim_{n \rightarrow \infty} \int \zeta(y)m(x, y)\pi_n(dx, dy) \\ &= \lim_{n \rightarrow \infty} q_n \int \zeta(y)\nu(dy) = q^* \int \zeta(y)\nu(dy). \end{aligned}$$

Since $q^* \geq \underline{m} > 0$, this is the second marginal condition in (3.1). Thus, $\pi^* \in \Gamma_m(\mu, \nu)$. Finally, lower semi-continuity of c and the Portmanteau theorem yield

$$\int c(x, y)\pi^*(dx, dy) \leq \liminf_{n \rightarrow \infty} \int c(x, y)\pi_n(dx, dy),$$

which shows that π^* is a minimizer. $\square$

Remark 3.4 (Mass escaping to infinity). The upper bound alone in (3.3) is not sufficient to guarantee tightness. Let $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \mathbb{N} = \{1, 2, \dots\}$, both equipped with the discrete topology. Consider the probability measures

$$\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \quad \text{and} \quad \nu(\{j\}) = 2^{-j}, \quad j \in \mathbb{N},$$

the mass-change factor

$$m(0, j) = 1 \quad \text{and} \quad m(1, j) = 4^{-j},$$

and the cost function

$$c(0, j) = 0 \quad \text{and} \quad c(1, j) = \frac{1}{j}.$$

Both m and c are continuous, the function m is positive and bounded from above, and c is bounded from below. However, $\inf m = 0$.

For $n \in \mathbb{N}$, let $q_n = (1 + 4^{-n})/2$ and define the transport plan π_n by

$$\pi_n(1, j) = \begin{cases} \frac{1}{2}, & j = n, \\ 0, & j \neq n, \end{cases} \quad \pi_n(0, j) = \begin{cases} q_n 2^{-n} - \frac{1}{2} 4^{-n}, & j = n, \\ q_n 2^{-j}, & j \neq n. \end{cases}$$

The values $\pi_n(0, n)$ are non-negative and a direct calculation shows that the first marginal of π_n is μ . Moreover, for every $j \in \mathbb{N}$ it holds

$$\pi_n(0, j) + 4^{-j}\pi_n(1, j) = q_n 2^{-j}.$$

Hence, the weighted second marginal is given by $q_n \nu$ and $\pi_n \in \Gamma_m(\mu, \nu)$. The cost of π_n is

$$\int c(x, y)\pi_n(dx, dy) = \frac{1}{2n} \rightarrow 0.$$

On the other hand, every $\pi \in \Gamma_m(\mu, \nu)$ satisfies $\sum_{j \in \mathbb{N}} \pi(1, j) = 1/2$ and therefore

$$\int c(x, y) \pi(dx, dy) = \sum_{j \in \mathbb{N}} \frac{1}{j} \pi(1, j) > 0.$$

Thus, the infimum in (KP) is zero but is not attained. In the sequence (π_n) , half of the mass escapes to infinity through points where m becomes small.

Remark 3.5 (Non-uniqueness of the optimal transport plan). In general, the optimal transport plan π^* is not unique. Let us consider a mass-change factor such that $m(x, y) = 0$ whenever $|x - y| \geq C > 0$ (see e.g. (3.6) below). If the source measure μ has some mass in a set A such that $\inf_{x \in A, y \in \text{supp } \nu} |x - y| \geq C$, the mass from A cannot reach the support $\text{supp } \nu$ and it does not matter where in $\text{supp } \nu$ it gets transported to. Remark 3.18 in Section 3.4 discusses a situation where the optimal transport plan is unique.

In the remainder of this subsection we present settings and situations where the non-conservative OT may arise or find applications. A detailed example of optimal rebalancing of portfolios corresponding to a discrete non-conservative optimal transport problem was outlined in Section 2.

Example 3.6 (Leaky Monge Problem). The original motivation of Monge was to find the most efficient way to move sand distributed in space according to μ to a desired target distribution ν using buckets of unit mass. Therefore, the cost function $c(x, y)$ is given by the natural (e.g. Euclidean or Manhattan) distance between x and y . The solution of the associated optimal transport problem minimizes the mean dislocation. No particle of sand is making an unnecessary round trip. If the mass is only moved horizontally on a frictional surface with constant resistive force proportional to its weight, this corresponds to minimizing the needed work to transport the sand.

Now, let us assume that buckets are leaky, i.e. they lose sand at a uniform rate due to a small hole in the bucket. Assuming that mass is transported with uniform speed, the mass-change factor is given by

$$m(x, y) = \max\{0, 1 - kd(x, y)\}, \quad (7)$$

where $d(x, y)$ measures the time to transport and the constant $k > 0$ models the size of the hole.

In many physical models, the loss of mass would not happen at a uniform rate but be proportional to the remaining sand in the bucket. Examples would be loss of mass due to radioactive decay, first order chemical reactions such as drug elimination in a host, small gas leaks, or biodegradation of organic matter. In such a situation the loss will be exponential in the time needed to transport the mass from x to y and the mass-change factor is given by

$$m(x, y) = \exp(-kd(x, y)). \quad (8)$$

A situation when mass is gained in the transport could be a vehicle that charges its battery via solar panels or an infrastructure (e.g. satellites or cars). Note that for (3.6) the condition of Remark 3.2 may fail: if the supports of μ and ν are at distance at least $1/k$, no mass survives the transport and $\Gamma_m(\mu, \nu)$ is empty; for the strictly positive factor (3.6) this cannot happen.

Remark 3.7 (Limitations and extensions). Our model cannot account for situations when the loss of mass is not proportional to the amount of transported mass. In reality, the dependency between cost function, the mass-change factor, and the transport map is a lot more subtle and should be carefully modeled with the specific application in mind. The interplay between those quantities gives rise to richer models and strategies. In the application to optimal rebalancing, minimizing the cost function is equivalent to minimizing the loss of mass. Non-conservative optimal transport problems are able to find a balance between competing goals; for example when finding the best speed to transport mass from x to y in order not to lose too much mass but also not to spend too much energy for the transport.

Optimal transport has beautiful interpretations and applications in economics, see e.g. [San15, Section 1.7.3]. We mention two such problems from economics here for which a non-conservative extension has natural applications.

Example 3.8 (Utility maximization). One such application is in optimal allocation of goods to consumers. The source measure $\mu(x)$ and the target measure $\nu(y)$ represent the distributions of various types of goods and consumers, respectively. The aim of the optimal transport problem is to maximize the overall utility $u = -c$ across possible matchings π of goods to consumers. The optimal potential $\varphi(x)$ of the dual problem (cf. Section 3.2 below) determines the optimal prices of the good x , and the potential $\psi(y)$ describes the received utility of the consumer y . One remarkable feature is that in the optimal state, every single utility is maximized, i.e. $\psi(y) = \max_x (u_x - \varphi(x))$.

In a non-conservative setting, the mass-change function $m(x, y)$ can describe the taxes or tariffs consumer y has to pay when buying the good x . Non-conservative optimal transport provides a tool to study the implications of new tariffs. It could also give rise to new tools that allow one to find the optimal design of tariffs given a certain agenda. Galichon [Gal16, Section 10.4] asks how the transport viewpoint extends when transfers are taxed, contrasting the resulting equilibrium transport problems with optimization-based transport. In matching markets the tax falls on the payments between matched agents, and [DGJK20] recover an optimization structure for proportional taxes. Here it falls on the transported mass itself, and the optimization structure survives for proportional, pair-dependent loss; nonlinear frictions lie outside our framework.

Example 3.9 (Productivity maximization). Another economic application is the maximization of the productivity of a company. The source measure $\mu(x)$ is interpreted as the number of workers of type x available, and the target measure $\nu(y)$ the amount of work of type y needed to create the product. The

productivity function $p = -c$ then describes how productive a worker of type x is at performing task y . The associated optimal transport problem then maximizes the productivity of the company and finds the optimal assignment of jobs. Moving to the non-conservative setting, the mass-change factor $m(x, y)$ could model the retention rate, i.e. how many workers of type x quit if they are forced to do job y . Including the retention rates will result in a more productive company.

3.2 The non-conservative Dual Problem.

We open this subsection with a formal derivation that will motivate the formulation of the dual problem to the optimization problem (KP). Under the compactness hypotheses of Theorem 3.13 below, all displayed quantities are well-defined; outside that setting, the calculation only motivates the dual problem. Let $\mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ denote the set of positive finite measures on $\mathcal{X} \times \mathcal{Y}$. We define the function $\chi : \mathcal{M}_+(\mathcal{X} \times \mathcal{Y}) \rightarrow [0, \infty]$ as

$$\begin{aligned} \chi(\pi) = \sup_{\varphi, \psi} & \left\{ \int_{\mathcal{X}} \varphi(x) \mu(dx) - \int_{\mathcal{X} \times \mathcal{Y}} \varphi(x) \pi(dx, dy) \right. \\ & \left. + \int_{\mathcal{Y}} \psi(y) \nu(dy) \int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi(dx, dy) - \int_{\mathcal{X} \times \mathcal{Y}} \psi(y) m(x, y) \pi(dx, dy) \right\}, \end{aligned}$$

where the supremum is over all bounded and continuous functions $\varphi \in \mathcal{C}_b(\mathcal{X})$ and $\psi \in \mathcal{C}_b(\mathcal{Y})$. Under the hypotheses of Theorem 3.13, the function χ coincides with

$$\chi(\pi) = \begin{cases} \infty, & \text{if } \pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y}) \setminus \Gamma_m(\mu, \nu), \\ 0, & \text{if } \pi \in \Gamma_m(\mu, \nu). \end{cases}$$

Hence, we obtain that

$$\begin{aligned} \inf_{\pi \in \Gamma_m(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy) &= \inf_{\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \left\{ \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy) + \chi(\pi) \right\} \\ &= \inf_{\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \sup_{\varphi, \psi} \left\{ \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy) + \int_{\mathcal{X}} \varphi(x) \mu(dx) - \int_{\mathcal{X} \times \mathcal{Y}} \varphi(x) \pi(dx, dy) \right. \\ & \quad \left. + \int_{\mathcal{Y}} \psi(y) \nu(dy) \int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi(dx, dy) - \int_{\mathcal{X} \times \mathcal{Y}} \psi(y) m(x, y) \pi(dx, dy) \right\}. \end{aligned}$$

Consequently,

$$\begin{aligned} \inf_{\pi \in \Gamma_m(\mu, \nu)} \int c(x, y) \pi(dx, dy) &\geq \sup_{\varphi, \psi} \left\{ \int \varphi(x) \mu(dx) \right. \\ & \quad \left. + \inf_{\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \int \left(c(x, y) - \varphi(x) - \left(\psi(y) - \int \psi(b) \nu(db) \right) m(x, y) \right) \pi(dx, dy) \right\}. \end{aligned} \tag{9}$$

For a fixed pair (φ, ψ) , the inner infimum in (3.2) equals $-\infty$ unless

$$c(x, y) \geq \varphi(x) + \left(\psi(y) - \int \psi(b) \nu(db) \right) m(x, y) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

This calculation motivates the following definition.

Definition 3.10 (Dual Kantorovich problem in non-conservative optimal transport). *Given a cost function $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ and a mass-change factor $m : \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty)$, we define the admissibility set $\mathcal{A} := \mathcal{A}(c, m)$ as all pairs $(\varphi, \psi) \in \mathcal{C}_b(\mathcal{X}) \times \mathcal{C}_b(\mathcal{Y})$ such that*

$$c(x, y) \geq \varphi(x) + \left(\psi(y) - \int \psi(b) \nu(db) \right) m(x, y) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (10)$$

The dual problem is given by

$$\sup_{(\varphi, \psi) \in \mathcal{A}(c, m)} \int \varphi(x) \mu(dx). \quad (\text{DP})$$

Remark 3.11. For a constant mass-change factor $m \equiv 1$, one recovers the dual problem of the classical Kantorovich problem

$$\sup_{\varphi, \psi} \left\{ \int \varphi(x) \mu(dx) + \int \psi(y) \nu(dy) \mid \forall x, y : c(x, y) \geq \varphi(x) + \psi(y) \right\}.$$

Here the supremum is taken over $\varphi \in \mathcal{C}_b(\mathcal{X})$ and $\psi \in \mathcal{C}_b(\mathcal{Y})$. One difference in non-conservative optimal transport is that the roles of μ and ν are not symmetric. Consequently, also the roles of φ and ψ become non-symmetric.

The preceding calculation motivates the form of the dual problem. Weak duality itself follows directly from the marginal constraints and is summarized below.

Lemma 3.12 (Weak duality). *For all $\pi \in \Gamma_m(\mu, \nu)$ and all $(\varphi, \psi) \in \mathcal{A}$ it holds*

$$\int c(x, y) \pi(dx, dy) \geq \inf (\text{KP}) \geq \sup (\text{DP}) \geq \int \varphi(x) \mu(dx).$$

Proof. Let $\pi \in \Gamma_m(\mu, \nu)$ and $(\varphi, \psi) \in \mathcal{A}$, and set

$$q_\pi := \int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi(dx, dy).$$

The terms involving φ and ψ are finite since these functions are bounded and $0 < q_\pi < \infty$. Moreover, the admissibility inequality makes $\int c d\pi$ well-defined as an extended real number. Integrating (3.10) with respect to π and using both marginal conditions in (3.1), we obtain

$$\begin{aligned} \int c(x, y) \pi(dx, dy) &\geq \int \varphi(x) \mu(dx) + \int \psi(y) m(x, y) \pi(dx, dy) - q_\pi \int \psi(y) \nu(dy) \\ &= \int \varphi(x) \mu(dx). \end{aligned}$$

Taking the supremum over $(\varphi, \psi) \in \mathcal{A}$ and then the infimum over $\pi \in \Gamma_m(\mu, \nu)$ proves the middle inequality. The remaining inequalities follow directly from the definitions. $\square$

Let us now investigate existence of a maximizer of the dual problem (DP).

Theorem 3.13. *Suppose that the sets $\mathcal{X}$ and $\mathcal{Y}$ are compact and that the functions c and m are continuous with $\inf_{x,y} m(x, y) > 0$. Then, there exists $(\varphi, \psi) \in \mathcal{A}$ satisfying*

$$\varphi(x) := \psi_m^c(x) := \inf_{y \in \mathcal{Y}} \left(c(x, y) - \left(\psi(y) - \int_{\mathcal{Y}} \psi(b) \nu(db) \right) m(x, y) \right) \quad (11)$$

and

$$\psi(y) = \inf_{x \in \mathcal{X}} \frac{c(x, y) - \varphi(x)}{m(x, y)} \quad \text{with} \quad \int \psi(y) \nu(dy) = 0 \quad (12)$$

such that

$$\sup(\text{DP}) = \max(\text{DP}) = \int \varphi(x) \mu(dx).$$

Under the hypotheses of Theorem 3.13, continuous functions are bounded. In particular, $\mathcal{C}(\mathcal{X}) = \mathcal{C}_b(\mathcal{X})$, $\mathcal{C}(\mathcal{Y}) = \mathcal{C}_b(\mathcal{Y})$, and $\mathcal{C}(\mathcal{X} \times \mathcal{Y}) = \mathcal{C}_b(\mathcal{X} \times \mathcal{Y})$.

The main idea of the proof is using the Arzela-Ascoli theorem to show that a maximizing sequence $(\varphi_n, \psi_n) \in \mathcal{A}$ converges to a maximizer $(\varphi, \psi) \in \mathcal{A}$. The equation (3.13) takes on a similar role as the c -transform in classical optimal transport. In contrast to the classical case, the argument gets more involved due to (3.13) and (3.13) being a more complicated analogue of the c -transform that lacks symmetry. Compactness is also a standard convenient assumption in the classical construction of continuous Kantorovich potentials by c -transforms and the Arzela-Ascoli theorem; see, e.g., [San15, Section 1.2]. On non-compact spaces, dual attainment requires additional boundedness, regularity, or integrability assumptions.

Proof of Theorem 3.13. In the first part of this proof we show that the feasible set of the dual problem (DP) can be restricted to admissible pairs $(\varphi, \psi) \in \mathcal{A}$ satisfying (3.13) and (3.13) without changing the optimal value or solution of the problem. We start with deducing (3.13). For this, let us fix a function ψ and consider all φ such that $(\varphi, \psi) \in \mathcal{A}$ is admissible. From the definition of $\mathcal{A}$ and the continuity of c and m it follows that

$$\varphi(x) \leq \inf_{y \in \mathcal{Y}} \left(c(x, y) - \left(\psi(y) - \int \psi(b) \nu(db) \right) m(x, y) \right) =: \psi_m^c(x). \quad (13)$$

As all involved functions are continuous, the function ψ_m^c is also continuous. Because $\mathcal{X}$ is compact, the function ψ_m^c is also bounded. It follows that the pair $(\psi_m^c, \psi) \in \mathcal{A}$ and from (3.2) it follows that

$$\int_{\mathcal{X}} \varphi(x) \mu(dx) \leq \int_{\mathcal{X}} \psi_m^c(x) \mu(dx).$$

Therefore, we can w.l.o.g. impose $\varphi = \psi_m^c$.

Next, let us show that we can w.l.o.g. impose (3.13). Indeed, from the admissibility condition it follows that

$$\psi(y) - \int \psi(y) \nu(dy) \leq \inf_x \frac{c(x, y) - \varphi(x)}{m(x, y)}.$$

Integrating this inequality yields

$$0 = \int_{\mathcal{Y}} \psi(y) \nu(dy) - \int_{\mathcal{Y}} \psi(b) \nu(db) \leq \int_{\mathcal{Y}} \inf_x \frac{c(x, y) - \varphi(x)}{m(x, y)} \nu(dy).$$

By continuity and monotonicity there exists a constant $k \geq 0$ such that

$$0 = \int_{\mathcal{Y}} \inf_x \frac{c(x, y) - \varphi(x) - k}{m(x, y)} \nu(dy).$$

We define now

$$\bar{\varphi} = \varphi + k \quad \text{and} \quad \bar{\psi}(y) = \inf_x \frac{c(x, y) - \varphi(x) - k}{m(x, y)}.$$

It follows from the construction that $(\bar{\varphi}, \bar{\psi}) \in \mathcal{A}$ is admissible and

$$\int \varphi \, d\mu \leq \int \bar{\varphi} \, d\mu.$$

From the original pair (φ, ψ) , we constructed an admissible pair $(\bar{\varphi}, \bar{\psi}) \in \mathcal{A}$ that increased $\int \varphi \, d\mu \leq \int \bar{\varphi} \, d\mu$, and satisfies $\int \bar{\psi} \, d\nu = 0$ and (3.13). To ensure that (3.13) and (3.13) hold simultaneously, it is sufficient to define

$$\tilde{\varphi}(x) := \inf_{y \in \mathcal{Y}} (c(x, y) - \bar{\psi}(y) m(x, y)),$$

and realize that the pair $(\tilde{\varphi}, \bar{\psi}) \in \mathcal{A}$ is admissible, satisfies (3.13) and (3.13), and again increased $\int \bar{\varphi} \, d\mu \leq \int \tilde{\varphi} \, d\mu$. This two-step improvement is the analogue of the double convexification trick of classical optimal transport, see e.g. [San15].

In the next part of this proof, we show that admissible pairs $(\varphi, \psi) \in \mathcal{A}$ satisfying (3.13–3.13) are uniformly bounded, i.e. there is a constant $K > 0$ such that

$$-K \leq \varphi(x) \leq K, \quad -K \leq \psi(y) \leq K \quad \text{for all } x \in \mathcal{X}, \, y \in \mathcal{Y}. \quad (14)$$

- (i) From (3.13) it follows that $\psi(y^*) \geq 0$ for some $y^* \in \mathcal{Y}$. Using (3.2) and $\int \psi \, d\nu = 0$, it follows that φ is uniformly bounded from above, namely

$$\varphi(x) \leq c(x, y^*) \leq \sup_{x, y} c(x, y) < \infty. \quad (15)$$

(ii) From (3.2) it follows that

$$\frac{c(x, y) - \varphi(x)}{m(x, y)} \geq \frac{\inf_{x, y} c(x, y) - \sup_{x, y} c(x, y)}{\inf_{x, y} m(x, y)},$$

which together with (3.13) yields that ψ is uniformly bounded from below.

(iii) Note that the objective value for an admissible pair $(\varphi, \psi) \in \mathcal{A}$ with $\varphi(x) < \inf_{x, y} c(x, y)$ across all $x \in \mathcal{X}$ is dominated by an objective value for the constant admissible pair $(\bar{\varphi}, \bar{\psi}) \equiv (\inf_{x, y} c(x, y), 0)$. Therefore, we can w.l.o.g. exclude such admissible pairs and assume that there exists some $x^* \in \mathcal{X}$ such that $\varphi(x^*) \geq \inf_{x, y} c(x, y)$. In light of this, for all $y \in \mathcal{Y}$ we obtain from (3.13)

$$\psi(y) \leq \frac{c(x^*, y) - \varphi(x^*)}{m(x^*, y)} \leq \frac{\sup_{x, y} c(x, y) - \inf_{x, y} c(x, y)}{\inf_{x, y} m(x, y)}.$$

This shows that ψ is uniformly bounded from above.

(iv) Finally, the uniform upper bound on ψ implies by (3.13) and (3.13) a uniform lower bound on φ .

In the last part of the proof, we argue existence of an optimal solution. By the previous arguments, we restrict ourselves to admissible pairs satisfying (3.13–3.13) and uniform boundedness. Let us consider a maximizing sequence $(\varphi_n, \psi_n) \in \mathcal{A}$ realizing the supremum in (DP). In order to apply Arzela-Ascoli, it remains to show that the functions φ_n and ψ_n are equi-continuous. Let us first consider $\{\varphi_n\}$ and define the family of functions $f_{y, n}(x) := c(x, y) - \psi_n(y)m(x, y)$ and observe $\varphi_n(x) = \inf_y f_{y, n}(x)$. It is a known fact that the modulus of continuity of $f_{y, n}(x)$ carries over to $\inf_y f_{y, n} = \varphi_n$ (see e.g. [San15, Box 1.5]). Direct calculation using (3.13) and (3.2) gives

$$|f_{y, n}(x_2) - f_{y, n}(x_1)| \leq |c(x_2, y) - c(x_1, y)| + K|m(x_2, y) - m(x_1, y)|.$$

So the modulus of continuity of the functions φ_n only depends on the modulus of continuity of c and m . Therefore the functions φ_n are equi-continuous. Let us now consider the family $\{\psi_n\}$ and define the auxiliary functions $g_{n, x} = \frac{c(x, y) - \varphi_n(x)}{m(x, y)}$. As the family φ_n is equi-continuous with modulus of continuity only depending on c and m , a direct calculation shows that the family $g_{n, x}$ also has a uniform modulus of continuity, and therefore $\inf_x g_n = \psi_n$ has a uniform modulus of continuity as well. Hence, the functions ψ_n are equi-continuous.

As a consequence, we can apply Arzela-Ascoli and can extract a subsequence $(\varphi_{n_k}, \psi_{n_k})$ that converges uniformly to a maximizer $(\varphi, \psi) \in \mathcal{A}$. $\square$

3.3 Strong Duality

We prove strong duality by a direct application of the Hahn–Banach separation theorem. The argument follows the approach to classical Monge–Kantorovich duality in [Edw10, Theorem 4.1], adapted here to the weighted marginal constraint. In particular, equality of the primal and dual values does not require the existence of optimal potentials.

Theorem 3.14 (Strong Duality). *Suppose that the assumptions of Theorem 3.3 hold and that there exists $\pi_0 \in \Gamma_m(\mu, \nu)$ such that*

$$\int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi_0(dx, dy) < \infty. \quad (16)$$

Then

$$\min_{\pi \in \Gamma_m(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy) = \sup_{(\varphi, \psi) \in \mathcal{A}(c, m)} \int_{\mathcal{X}} \varphi(x) \mu(dx).$$

The proof of Theorem 3.14 needs some preparation. Throughout the remainder of this subsection, we impose the assumptions of Theorem 3.3; in particular, the cost function c is lower semi-continuous and bounded from below, and the mass-change factor m is continuous and satisfies the two-sided bounds (3.3). For a metric space $\mathcal{Z}$, let $\mathcal{M}(\mathcal{Z})$ denote the space of finite signed Borel measures on $\mathcal{Z}$, equipped with the weak topology induced by $\mathcal{C}_b(\mathcal{Z})$. We write $\mathcal{M}_+(\mathcal{Z})$ for its cone of non-negative measures and

$$\mathcal{M}_0(\mathcal{Y}) := \{\lambda \in \mathcal{M}(\mathcal{Y}) : \lambda(\mathcal{Y}) = 0\}.$$

For $\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$, define

$$q_\pi := \int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi(dx, dy), \quad L\pi := (\text{Proj}_2)_\#(m\pi) - q_\pi \nu.$$

Then $L\pi \in \mathcal{M}_0(\mathcal{Y})$. Moreover, if $(\text{Proj}_1)_\# \pi = \mu$, then π is a probability measure and (3.3) yields $0 < \underline{m} \leq q_\pi \leq \overline{m} < \infty$. In this case, the relation $L\pi = 0$ is equivalent to $\pi \in \Gamma_m(\mu, \nu)$.

We need the following auxiliary lemma.

Lemma 3.15. *Suppose that the assumptions of Theorem 3.3 hold. Consider the locally convex space*

$$E := \mathcal{M}(\mathcal{X}) \times \mathcal{M}_0(\mathcal{Y}) \times \mathbb{R}$$

with the product topology. Then the subset

$$K := \left\{ (\eta, \lambda, r) \in E : \text{there exists } \pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y}) \text{ such that} \right. \quad (17) \\ \left. (\text{Proj}_1)_\# \pi = \eta, \quad L\pi = \lambda, \quad \int c d\pi \leq r \right\}$$

is a closed convex cone in E .

Proof of Lemma 3.15. The convex-cone property follows directly from the linearity of the first marginal, the map L , and the cost integral. We prove closedness. Since the weak topology on the space of signed measures need not be metrizable, we begin with a net. Let (η_i, λ_i, r_i) be a net in K converging to (η, λ, r) in E , and choose corresponding measures $\pi_i \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ as in (3.15). Since η_i converges weakly to η , the masses

$$\pi_i(\mathcal{X} \times \mathcal{Y}) = \eta_i(\mathcal{X})$$

converge to $\eta(\mathcal{X})$ and are therefore eventually bounded. Passing to a tail of the net, we may assume that they are bounded. Hence, by (3.3), the numbers $q_i := \int m d\pi_i$ are bounded. Passing to a subnet, we may assume that $q_i \rightarrow q$ for some $q \geq 0$. The positive measures

$$\rho_i := (\text{Proj}_2)_\#(m\pi_i) = \lambda_i + q_i\nu$$

then converge weakly to the positive measure $\rho := \lambda + q\nu$. On Polish spaces, the weak topology on finite positive measures is metrizable. We may therefore select a sequence of terms of the subnet, which we denote by (η_n, λ_n, r_n) , with corresponding measures π_n , such that

$$\eta_n \rightarrow \eta, \quad \rho_n \rightarrow \rho, \quad q_n \rightarrow q, \quad r_n \rightarrow r.$$

The weakly convergent sequences (η_n) and (ρ_n) are uniformly tight. Moreover, for all Borel sets $A \subset \mathcal{X}$ and $B \subset \mathcal{Y}$,

$$\pi_n(A \times \mathcal{Y}) = \eta_n(A), \quad \pi_n(\mathcal{X} \times B) \leq \frac{1}{\underline{m}} \rho_n(B).$$

It follows that (π_n) is uniformly tight. By Prokhorov's theorem, a subsequence converges weakly to some $\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$.

Passing to the limit in the first marginal gives $(\text{Proj}_1)_\#\pi = \eta$. Since m is bounded and continuous, weak convergence also gives $L\pi = \lambda$. Finally, lower semi-continuity of c , the Portmanteau theorem, and $r_n \rightarrow r$ yield

$$\int c d\pi \leq \liminf_n \int c d\pi_n \leq \lim_n r_n = r.$$

Thus, $(\eta, \lambda, r) \in K$, and K is closed. □

Proof of Theorem 3.14. Let

$$P := \min_{\pi \in \Gamma_m(\mu, \nu)} \int c d\pi,$$

which is finite by (3.14). Fix $\varepsilon > 0$. By the definition of P and (3.15), the point

$$z_\varepsilon := (\mu, 0, P - \varepsilon)$$

does not belong to K . Since K is a closed convex cone, the Hahn–Banach separation theorem yields a non-zero continuous linear functional F on E such that

$$F(z_\varepsilon) < 0 \leq F(k) \quad \text{for all } k \in K. \quad (18)$$

By the definition of the weak topologies on the first two factors of E , there are functions $f \in \mathcal{C}_b(\mathcal{X})$ and $g \in \mathcal{C}_b(\mathcal{Y})$, and a number $\alpha \in \mathbb{R}$, such that

$$F(\eta, \lambda, r) = \int_{\mathcal{X}} f d\eta + \int_{\mathcal{Y}} g d\lambda + \alpha r.$$

Since $(0, 0, s) \in K$ for all $s \geq 0$, it follows that $\alpha \geq 0$. If $\alpha = 0$, then the finite-cost plan π_0 from (3.14) gives a point $(\mu, 0, r_0) \in K$ with $r_0 < \infty$. This would imply

$$0 \leq F(\mu, 0, r_0) = \int f d\mu = F(z_\varepsilon) < 0,$$

a contradiction. Hence, $\alpha > 0$.

Let $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $s > 0$. Taking $\pi = s\delta_{(x,y)}$ in (3.15) shows that

$$(s\delta_x, sm(x, y)(\delta_y - \nu), sc(x, y)) \in K.$$

Applying (3.3) and dividing by s yields

$$f(x) + m(x, y) \left(g(y) - \int_{\mathcal{Y}} g d\nu \right) + \alpha c(x, y) \geq 0.$$

Define

$$\varphi := -\frac{f}{\alpha}, \quad \psi := -\frac{g}{\alpha}.$$

Then $(\varphi, \psi) \in \mathcal{A}(c, m)$. Furthermore, the strict inequality in (3.3) gives

$$\int_{\mathcal{X}} \varphi d\mu > P - \varepsilon.$$

It follows that $\sup(\text{DP}) \geq P - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, weak duality gives $\sup(\text{DP}) = P$. $\square$

3.4 Existence of optimal transport map(s)

In this subsection, we study whether the optimal transport *plan* of problem (KP) can be expressed via optimal transport *map(s)*. As the roles of the measures μ and ν are not symmetric, we define two notions of optimal transport maps. Throughout, $\sharp$ denotes the push-forward measure.

The arguments for the existence results below use optimal Kantorovich potentials and their regularity. We therefore retain compactness of the underlying

spaces $\mathcal{X}$ and $\mathcal{Y}$: it yields dual attainment through Theorem 3.13 and provides the uniform bounds needed in the characteristic-equation argument. In classical optimal transport, compactness can be replaced by suitable moment and growth assumptions in particular settings; we do not pursue the corresponding extension here.

Definition 3.16 (Primal and dual optimal transport map). *Assume that π^* is an optimal transport plan for the problem (KP).*

- (i) *A function $T : \mathcal{X} \rightarrow \mathcal{Y}$ is called a (primal) optimal transport map if $\pi^* = (\text{Id}_{\mathcal{X}}, T)_{\#} \mu$ is given by the push forward of μ under the map $(\text{Id}_{\mathcal{X}}, T)$.*
- (ii) *A function $S : \mathcal{Y} \rightarrow \mathcal{X}$ is called a dual optimal transport map of (KP) if $\pi^* = (S, \text{Id}_{\mathcal{Y}})_{\#} \lambda$. Here, the probability measure λ is given by the Radon-Nikodym derivative*

$$\lambda(dy) = \frac{1}{Z} \frac{1}{m(S(y), y)} \nu(dy),$$

where Z denotes the normalization constant transforming λ into a probability measure.

Using the same argument as in the case of classical optimal transport we get the following statement.

Proposition 3.17 (Criterion for the existence of optimal transport maps). *Let π^* be an optimal transport plan for problem (KP).*

- (i) *If for a function $T : \mathcal{X} \rightarrow \mathcal{Y}$ it holds*

$$\text{supp } \pi^* \subset \{(x, T(x)) \mid x \in \mathcal{X}\},$$

then T is an optimal transport map of (KP).

- (ii) *If for a function $S : \mathcal{Y} \rightarrow \mathcal{X}$ it holds*

$$\text{supp } \pi^* \subset \{(S(y), y) \mid y \in \mathcal{Y}\},$$

then S is a dual optimal transport map of (KP).

Remark 3.18. Suppose that every optimal transport plan of (KP) is induced by a transport map, as in Proposition 3.21 (i) below. Then the optimal transport map is μ -a.e. unique, and consequently so is the optimal transport plan. Indeed, if $\pi_1 := (\text{Id}_{\mathcal{X}}, T_1)_{\#} \mu$ and $\pi_2 := (\text{Id}_{\mathcal{X}}, T_2)_{\#} \mu$ are optimal, then $\pi_3 = \frac{1}{2} \pi_1 + \frac{1}{2} \pi_2$ is optimal as well, since the set $\Gamma_m(\mu, \nu)$ is convex and the objective is linear in the plan. Being induced by a map, $\pi_3 = (\text{Id}_{\mathcal{X}}, T_3)_{\#} \mu$, which is only possible if $T_1 = T_2$ μ -a.e. The same argument applies to dual optimal transport maps, with ν in place of μ .

Remark 3.19. If there exists an optimal transport map T as well as a dual optimal transport map S , then it holds $(x, T(x)) = (S(y), y)$ for π^* -a.e. (x, y) . Therefore, T and S are inverse to each other.

From Proposition 3.17 it becomes clear that in order to show the existence of optimal transport maps, one needs to study the support of the optimal transport plan.

Proposition 3.20. *Assume that $\min(\text{KP}) = \max(\text{DP})$, which means that there is $\pi^* \in \Gamma_m(\mu, \nu)$ and $(\varphi, \psi) \in \mathcal{A}$ such that*

$$\int \varphi(x) \mu(dx) = \int c(x, y) d\pi^*. \quad (19)$$

Then, the support of π^ is contained in the set*

$$\mathcal{S} := \left\{ (x, y) \in \mathcal{X} \times \mathcal{Y} \mid c(x, y) = \varphi(x) + \left(\psi(y) - \int \psi d\nu \right) m(x, y) \right\}, \quad (20)$$

i.e. $\text{supp } \pi^ \subset \mathcal{S}$. Moreover, consider the family of functions $f_x : \mathcal{Y} \rightarrow \mathbb{R}$ and $g_y : \mathcal{X} \rightarrow \mathbb{R}$ given by*

$$f_x(y) := c(x, y) - \left(\psi(y) - \int \psi d\nu \right) m(x, y) \quad (21)$$

and

$$g_y(x) := \frac{c(x, y) - \varphi(x)}{m(x, y)}. \quad (22)$$

Then, the set $\mathcal{S}$ is given by

$$\mathcal{S} = \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid y \text{ minimizes } f_x \text{ and } x \text{ minimizes } g_y\}. \quad (23)$$

Proof. Recall from Theorem 3.13 that we may choose the optimal dual solution $(\varphi, \psi) \in \mathcal{A}$ such that (3.13) and (3.13) hold. Define the non-negative continuous function

$$R(x, y) := c(x, y) - \varphi(x) - \left(\psi(y) - \int \psi d\nu \right) m(x, y).$$

Using both marginal conditions on π^* , we obtain

$$\int R d\pi^* = \int c d\pi^* - \int \varphi d\mu = 0,$$

where the last equality follows from (3.20). Since $R \geq 0$ is continuous, it follows that $R = 0$ on $\text{supp } \pi^*$, which proves (3.20).

It remains to characterize $\mathcal{S}$. For $(x, y) \in \mathcal{S}$, the contact equality and admissibility give

$$\begin{aligned} f_x(y) = \varphi(x) &\leq f_x(y') \quad \text{for all } y' \in \mathcal{Y}, \\ g_y(x) = \psi(y) &\leq g_y(x') \quad \text{for all } x' \in \mathcal{X}. \end{aligned}$$

Thus, y minimizes f_x and x minimizes g_y . Conversely, if y minimizes f_x and x minimizes g_y , then (3.13) and (3.13) give

$$f_x(y) = \varphi(x), \quad g_y(x) = \psi(y),$$

and hence $(x, y) \in \mathcal{S}$. This proves (3.20). $\square$

The theory for the existence of optimal transport maps does not easily carry over from the conservative case and needs further investigation. The next statement shows the existence of (primal or dual) optimal transport maps if the mass-change factor $m \approx 1$, i.e. if the model is close to a classical optimal transport problem.

Proposition 3.21 (Existence of optimal transport maps in the perturbative regime). *Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^n$ be non-empty, compact sets that are equal to the closures of their interiors. Let $h, d : [0, \infty) \rightarrow \mathbb{R}$ be two smooth, convex, Lipschitz functions such that*

- *h is strictly increasing and $\lim_{t \downarrow 0} \frac{h(t) - h(0)}{t} = 0$;*
- *d is non-decreasing and $\lim_{t \downarrow 0} \frac{d(t) - d(0)}{t} = 0$.*

Additionally, we assume that there is $\varepsilon > 0$ such that

$$h''(z) > \varepsilon d''(z) \quad \text{and} \quad h'(z) > \varepsilon d'(z) \quad \text{for all } 0 < z \leq \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} |x - y|. \quad (24)$$

We consider the cost function $c(x, y) := h(|x - y|)$ and the mass-change factor $m(x, y) = \exp(-k d(|x - y|))$ where $k \geq 0$ is a constant.

Then, there is a constant K such that for all $0 \leq k < K$ it holds:

- (i) *If the probability measure $\mu \in \mathcal{P}(\mathcal{X})$ is absolutely continuous w.r.t. the Lebesgue measure and the boundary of $\mathcal{X}$ has Lebesgue-measure 0, then every optimal transport plan is given by a unique optimal transport map $T : \mathcal{X} \rightarrow \mathcal{Y}$.*
- (ii) *If the probability measure $\nu \in \mathcal{P}(\mathcal{Y})$ is absolutely continuous w.r.t. the Lebesgue measure, and the boundary of $\mathcal{Y}$ has Lebesgue-measure 0, then every optimal transport plan is given by a unique dual optimal transport map $S : \mathcal{Y} \rightarrow \mathcal{X}$.*

Proof of Proposition 3.21. Under the current assumptions, Theorems 3.3, 3.13, and 3.14 hold. Consequently, the optimal dual solution $(\varphi, \psi) \in \mathcal{A}$ exists and the assumptions of Proposition 3.20 are satisfied. Hence, for any point $(x, y) \in \text{supp } \pi^*$ it holds that the value x is a minimizer of the function $g_y(x) = \frac{c(x, y) - \varphi(x)}{m(x, y)}$ and the value y is a minimizer of the function $f_x(y) = c(x, y) - \psi(y)m(x, y)$. We now show that f_x and g_y are a.e. differentiable.

Recall from the proof of Theorem 3.13 that φ, ψ are uniformly bounded. Specifically, we can show that

$$-\bar{h}(1 + 2e^{2k\bar{d}}) \leq \varphi(x) \leq \bar{h} \quad \text{and} \quad -2\bar{h}e^{k\bar{d}} \leq \psi(y) \leq 2\bar{h}e^{k\bar{d}}$$

where $\bar{h} := \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} |h(|x - y|)|$ and $\bar{d} := \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} |d(|x - y|)|$. In particular, those bounds are uniform in $|k| \leq K$. As $\mathcal{X}$ and $\mathcal{Y}$ are compact and additionally φ

and ψ are uniformly bounded, the functions φ and ψ inherit the modulus of continuity, up to a universal constant, from c and m (see proof of Theorem 3.13 for more details). Therefore φ and ψ are both Lipschitz, which implies together with our assumptions on c and m that f_x and g_y are also Lipschitz (see (3.20) and (3.20)). It follows from Rademacher's theorem (here we use that $\mathcal{X}$ is the closure of its interior) that if μ is absolutely continuous w.r.t. the Lebesgue measure, the functions g_y are μ -a.e. differentiable. Similarly, the functions f_x are ν -a.e. differentiable if ν is absolutely continuous w.r.t. the Lebesgue measure.

Let us turn to the verification of (i). Recall from the characterization (3.20) that for all $(x, y) \in \text{supp } \pi^*$, the value x minimizes the function g_y . For $(x, y) \in \text{supp } \pi^*$ it holds (recall that boundary points are negligible by assumption)

$$\begin{aligned} 0 &= m(x, y) \nabla_x g_y(x) \\ &= \nabla_x c(x, y) - \nabla_x \varphi(x) - \left(c(x, y) - \varphi(x) \right) \nabla_x \ln m(x, y) \quad \mu \text{ a.e. } x. \end{aligned} \quad (25)$$

For $c(x, y) = h(|x - y|)$ and $m(x, y) = \exp(-kd(|x - y|)) > 0$ we obtain

$$\left(h'(|x - y|) + kd'(|x - y|) (h(|x - y|) - \varphi(x)) \right) \frac{x - y}{|x - y|} = \nabla_x \varphi(x). \quad (26)$$

We now argue that, for fixed x , the equation (3.4) uniquely characterizes y . Since $\mathcal{X}$ and $\mathcal{Y}$ are compact, $D := \sup_{x, y} |x - y|$ is finite, and since (x, y) solves the equation (3.4), it must hold that $x - y = z \frac{\nabla_x \varphi(x)}{|\nabla_x \varphi(x)|}$ for some value $-D \leq z \leq D$. In light of (3.4), the scalar z must satisfy

$$\frac{z}{|z|} \underbrace{\left(h'(|z|) + kd'(|z|) (h(|z|) - \varphi(x)) \right)}_{=: F(|z|)} = |\nabla_x \varphi(x)|. \quad (27)$$

In view of the uniform upper bound $\varphi(x) \leq \bar{h}$ and $z \in [-D, D]$ it holds $h(|z|) - \varphi(x) \geq -2\bar{h}$. Then, for $0 < k < \frac{\varepsilon}{2\bar{h}}$ it holds for $|z| > 0$

$$F(|z|) = h'(|z|) + k d'(|z|) (h(|z|) - \varphi(x)) \geq h'(|z|) - 2\bar{h} k d'(|z|) \stackrel{(3.21)}{>} 0.$$

Additionally, it follows from our assumptions that $F(|z|) = 0$ if and only if $|z| = 0$. Therefore, in the case $|\nabla_x \varphi(x)| = 0$, it follows that $y = x$ is uniquely characterized by x .

In the case $|\nabla_x \varphi(x)| \neq 0$, the equation (3.4) may only have solutions $z > 0$ for sufficiently small $k > 0$. We now show that $F(z)$ is strictly increasing for sufficiently small $k > 0$. In view of the assumptions, it holds that

$$\begin{aligned} F'(z) &= h''(z) + kd''(z) (h(z) - \varphi(x)) + kd'(z) h'(z) \\ &\geq h''(z) - 2\bar{h} k d''(z) > (\varepsilon - 2\bar{h} k) d''(z) \geq 0, \end{aligned}$$

for any $0 \leq k \leq \frac{\varepsilon}{2\bar{h}}$. This concludes the argument for y being uniquely characterized via (3.4) given x , which proves part (i).

The argument for (ii) follows in the footsteps of (i). We have seen above that the functions f_x are ν -a.e. differentiable. Taking the gradient of f_x w.r.t. y yields for $(x, y) \in \text{supp } \pi^*$ (boundary points are again negligible by assumption)

$$\nabla_y c(x, y) - m(x, y) \nabla_y \psi(y) - \psi(y) \nabla_y m(x, y) = 0 \quad \nu \text{ a.e. } y. \quad (28)$$

This equation uniquely characterizes x in terms of y , which can be seen by similar arguments as for step (i). $\square$

In the proof of Proposition 3.21, the equations (3.4) and (3.4) play a special role. They are used to characterize the existence of optimal transport maps motivating the following definition.

Definition 3.22. *Let us consider a pair (φ, ψ) of optimal Kantorovich potentials solving (DP). Then we call the equation (3.4) the characteristic equation for the optimal transport map. Similarly, the equation (3.4) is called the characteristic equation for the dual optimal transport map.*

Remark 3.23 (An open question). We pose as an open question whether the existence and uniqueness of optimal transport maps persist in the perturbative regime beyond radially symmetric and monotone mass-change factors, for example under a smallness condition on $\nabla \ln m$.

Now, let us illustrate with an example that optimal transport maps can also exist when the mass-change factor m is not close to 1, i.e. when one does not expect the non-conservative optimal transport to be close to a classical optimal transport.

Proposition 3.24 (Existence of optimal transport maps in the quadratic leaky-Monge problem). *Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^n$ be non-empty, compact sets that are equal to the closures of their interiors. We consider the quadratic cost function $c(x, y) := \frac{1}{2}|x - y|^2$ and the mass-change factor $m(x, y) = 1 - \frac{k}{2}|x - y|^2$ for a constant $k > 0$ such that $\inf_{x \in \mathcal{X}, y \in \mathcal{Y}} m(x, y) > 0$. Then it holds:*

- (i) *If the probability measure μ is absolutely continuous w.r.t. the Lebesgue measure and the boundary of $\mathcal{X}$ has Lebesgue-measure 0, then there exists an optimal transport map $T : \mathcal{X} \rightarrow \mathcal{Y}$.*
- (ii) *If the probability measure ν is absolutely continuous w.r.t. the Lebesgue measure and the boundary of $\mathcal{Y}$ has Lebesgue-measure 0, then there exists a dual optimal transport map $S : \mathcal{Y} \rightarrow \mathcal{X}$.*

Proof of Proposition 3.24. The first part of the proof follows the same steps as the proof of Proposition 3.21; we do not repeat them. Recall that the functions $g_y(x) = \frac{c(x, y) - \varphi(x)}{m(x, y)}$ and $f_x(y) = c(x, y) - \psi(y)m(x, y)$ are Lipschitz and therefore Lebesgue a.e. differentiable.

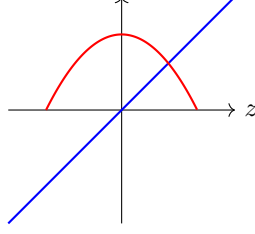


Figure 3: Uniqueness of the optimal transport map for the quadratic leaky Monge problem (see proof of Proposition 3.24).

Let us turn to the verification of (i). The characteristic equation for the optimal transport map is given for $(x, y) \in \text{supp } \pi^*$ by

$$\frac{(x - y) - \nabla_x \varphi(x)}{1 - \frac{k}{2}|x - y|^2} + \frac{\frac{1}{2}|x - y|^2 - \varphi(x)}{(1 - \frac{k}{2}|x - y|^2)^2} k(x - y) = 0 \quad \mu\text{-a.e. } x.$$

Rearranging terms yields

$$(x - y)(1 - k\varphi(x)) = \left(1 - \frac{k}{2}|x - y|^2\right) \nabla_x \varphi(x). \quad (29)$$

We now argue that, for fixed x , the equation (3.4) uniquely characterizes y . By (3.2) we observe that $\varphi(x) \leq \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} \frac{1}{2}|x - y|^2$. Combining this with our assumption

$$\inf_{x \in \mathcal{X}, y \in \mathcal{Y}} m(x, y) = 1 - \frac{k}{2} \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} |x - y|^2 > 0,$$

it follows that

$$1 - k\varphi(x) > 0.$$

In the case of $\nabla_x \varphi(x) = 0$ it follows from (3.4) that $x - y = 0$. Let us consider the case $\nabla_x \varphi(x) \neq 0$. An Ansatz $x - y = z \nabla_x \varphi(x)$ yields the equation

$$\underbrace{1 - \frac{k}{2}|\nabla_x \varphi(x)|^2 z^2}_{=1 - \frac{k}{2}|x - y|^2 = m(x, y) > 0} = (1 - k\varphi(x))z.$$

Let us interpret both sides as functions of z : The left-hand side is a concave quadratic function (as $-\frac{k}{2}|\nabla_x \varphi(x)|^2 < 0$), which is restricted to the area where it attains positive values (as $m(x, y) > 0$). The right-hand side is a linear function with a zero intercept. Consequently, there exists a unique intersection, see Figure 3. Hence, y is uniquely characterized by x and an application of Proposition 3.17 yields the desired statement. The verification of (ii) follows from a similar argument as for (i), based on the characteristic equation for the dual optimal transport map.

□

3.5 Dynamic formulation of non-conservative optimal transport

In the dynamic formulation, the transport of mass is modeled along trajectories interpolating continuously between both marginal constraints. It has the advantage of allowing for finer modeling and interpretation of the mass transport. In our example of optimal re-balancing, the dynamic formulation would allow one to calculate specific trading rates and to include, for example, dynamic price impact models. For simplicity, let us consider cost functions of the form $c(x, y) = |y - x|^a$ for $a \geq 1$.

We use the dynamical-plan point of view of the classical theory (see e.g. Chapter 8 in [AGS08]). Therefore, no a-priori regularity assumptions on a vector field are needed, and the existence of an optimal transport map is not required.

Definition 3.25 (Dynamical-plan formulation of (KP)). *Let $\mathcal{X} \subset \mathbb{R}^n$ be compact, let $m : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ be continuous, and let $\mu, \nu \in \mathcal{P}(\mathcal{X})$. Set $\Theta := C([0, 1]; \mathbb{R}^n)$, equipped with the supremum norm. Denote the evaluation map by $e_t(\gamma) := \gamma_t$. Recall that $\gamma \in AC^a([0, 1]; \mathbb{R}^n)$ if there is a $g \in L^a(0, 1)$ such that $|\gamma_t - \gamma_s| \leq \int_s^t g(r) dr$ for all $0 \leq s \leq t \leq 1$; such a curve is differentiable almost everywhere, satisfies $\gamma_t - \gamma_s = \int_s^t \dot{\gamma}_r dr$, and $|\dot{\gamma}| \in L^a(0, 1)$, see [AGS08, Definition 1.1.1].*

For $Q \in \mathcal{P}(\Theta)$ such that $\gamma_0, \gamma_1 \in \mathcal{X}$ for Q -almost every γ , set

$$q_Q := \int_{\Theta} m(\gamma_0, \gamma_1) Q(d\gamma).$$

We denote by $\mathcal{D}_m(\mu, \nu)$ the set of all such Q with $Q(AC^a([0, 1]; \mathbb{R}^n)) = 1$ and $q_Q > 0$, satisfying

$$(e_0)_\# Q = \mu, \quad (e_1)_\# (m(e_0, e_1)Q) = q_Q \nu. \quad (30)$$

The action of Q is given by

$$\mathcal{E}_a(Q) := \int_{\Theta} \int_0^1 |\dot{\gamma}_t|^a dt Q(d\gamma).$$

The dynamic formulation of (KP) is

$$\inf_{Q \in \mathcal{D}_m(\mu, \nu)} \mathcal{E}_a(Q). \quad (\text{GBB})$$

Remark 3.26. The requirement that $\gamma_0, \gamma_1 \in \mathcal{X}$ for Q -almost every γ makes the integrand defining q_Q meaningful, as the mass-change factor m is defined only on $\mathcal{X} \times \mathcal{X}$. Note that only the endpoints are constrained. The paths themselves are allowed to leave $\mathcal{X}$.

We work on Θ and not on $AC^a([0, 1]; \mathbb{R}^n)$. The reason is that Θ is a Polish space and that $\mathcal{P}(\Theta)$, the Borel probability measures on Θ , carries the usual narrow topology. However, $AC^a([0, 1]; \mathbb{R}^n)$ is a Borel subset of Θ that is not

closed and not preserved under narrow limits. The map $\gamma \mapsto \int_0^1 |\dot{\gamma}_t|^a dt$ is Borel measurable on Θ with values in $[0, \infty]$, with the convention that it equals $+\infty$ outside of $AC^a([0, 1]; \mathbb{R}^n)$. Consequently, $\mathcal{E}_a$ is well defined on all of $\mathcal{P}(\Theta)$ and finite only on plans concentrated on $AC^a([0, 1]; \mathbb{R}^n)$.

Therefore, (GBB) generalizes the classical Benamou-Brenier formulation [BB00] and recovers it in the special case $m \equiv 1$.

Proposition 3.27. *Let us consider the setting of Definition 3.25 and suppose that $\Gamma_m(\mu, \nu)$ is non-empty. Then $\mathcal{D}_m(\mu, \nu)$ is non-empty and*

$$\inf_{Q \in \mathcal{D}_m(\mu, \nu)} \mathcal{E}_a(Q) = \inf_{\pi \in \Gamma_m(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} |y - x|^a \pi(dx, dy). \quad (31)$$

Moreover, if π^* is an optimal transport plan for (KP), then the dynamic problem (GBB) admits a minimizer.

Proof. Let $Q \in \mathcal{D}_m(\mu, \nu)$ and define $\pi_Q := (e_0, e_1)_\# Q$. The endpoint conditions (3.25) show that $\pi_Q \in \Gamma_m(\mu, \nu)$. For every $\gamma \in AC^a([0, 1]; \mathbb{R}^n)$, Jensen's inequality gives

$$|\gamma_1 - \gamma_0|^a = \left| \int_0^1 \dot{\gamma}_t dt \right|^a \leq \int_0^1 |\dot{\gamma}_t|^a dt.$$

Integrating with respect to Q yields

$$\int_{\mathcal{X} \times \mathcal{X}} |y - x|^a \pi_Q(dx, dy) \leq \mathcal{E}_a(Q).$$

Conversely, let $\pi \in \Gamma_m(\mu, \nu)$ and associate to (x, y) the straight path

$$G(x, y)_t := (1 - t)x + ty.$$

The map $G : \mathcal{X} \times \mathcal{X} \rightarrow \Theta$ is continuous and therefore Borel. Hence $Q_\pi := G_\# \pi$ is a well-defined element of $\mathcal{P}(\Theta)$. Its curves are affine, and therefore Q_π is concentrated on $AC^a([0, 1]; \mathbb{R}^n)$ and on curves with endpoints in $\mathcal{X}$. Note that $(e_0, e_1)_\# Q_\pi = \pi$, which yields the endpoint conditions (3.25). Since $\pi \in \Gamma_m(\mu, \nu)$, we also have $q_{Q_\pi} = \int_{\mathcal{X} \times \mathcal{X}} m d\pi \in (0, \infty)$. Consequently, $Q_\pi \in \mathcal{D}_m(\mu, \nu)$ and the dynamic admissible set is non-empty. Moreover,

$$\mathcal{E}_a(Q_\pi) = \int_{\mathcal{X} \times \mathcal{X}} |y - x|^a \pi(dx, dy).$$

Taking the infimum in both constructions proves (3.27). If π^* is optimal, then Q_{π^*} is an optimal dynamical plan. $\square$

Corollary 3.28 (The optimal-map case). *Assume, in addition, that an optimal transport map $T : \mathcal{X} \rightarrow \mathcal{X}$ exists. Then*

$$Q_T := (x \mapsto (t \mapsto (1 - t)x + tT(x)))_\# \mu$$

is an optimal dynamical plan and

$$\inf_{Q \in \mathcal{D}_m(\mu, \nu)} \mathcal{E}_a(Q) = \int_{\mathcal{X}} |T(x) - x|^a \mu(dx).$$

Proof. The assertion follows from Proposition 3.27 applied to $\pi^* = (\text{Id}_X, T)_\# \mu$. $\square$

Let us now relate the dynamical description to an Eulerian one.

Lemma 3.29 (Barycentric velocity field). *Let $Q \in \mathcal{D}_m(\mu, \nu)$ have finite action and set $p_t := (e_t)_\# Q$. Let $(Q^{t,z})_{z \in \mathbb{R}^n}$ be the disintegration of Q with respect to the evaluation map e_t and define the barycentric velocity field*

$$v(t, z) := \int_{\Theta} \dot{\gamma}_t Q^{t,z}(d\gamma),$$

the average velocity of those curves that are at z at time t . Then v is well defined for almost every (t, z) , the pair (p, v) satisfies

$$\partial_t p + \nabla \cdot (pv) = 0, \quad p_0 = \mu, \quad (32)$$

in the sense of distributions, and for almost every t

$$\int_{\mathbb{R}^n} |v(t, z)|^a p_t(dz) \leq \int_{\Theta} |\dot{\gamma}_t|^a Q(d\gamma).$$

If $a > 1$, equality holds if and only if $\dot{\gamma}_t = v(t, z)$ for $Q^{t,z}$ -almost every γ , for p_t -almost every z .

Proof. Since Θ is Polish, the disintegration exists and is uniquely determined for p_t -almost every z . The measure $Q^{t,z}$ is concentrated on the curves passing through z at time t , i.e. $Q^{t,z}(e_t^{-1}(z)) = 1$, and $Q = \int_{\mathbb{R}^n} Q^{t,z} p_t(dz)$. As the action is finite and $a \geq 1$, Jensen's inequality on the probability space $dt \otimes Q$ gives

$$\int_0^1 \int_{\Theta} |\dot{\gamma}_t| Q(d\gamma) dt \leq \mathcal{E}_a(Q)^{1/a} < \infty.$$

Therefore, $\dot{\gamma}$ is integrable with respect to $dt \otimes Q$ and v is well defined for almost every (t, z) .

Let $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$. For Q -almost every γ the curve lies in $AC^a([0, 1]; \mathbb{R}^n)$. Hence $r \mapsto \varphi(\gamma_r)$ is absolutely continuous with derivative $\nabla \varphi(\gamma_r) \cdot \dot{\gamma}_r$. By the definition of p_t as an image measure and Fubini's theorem, which the last estimate licenses,

$$\int \varphi dp_t - \int \varphi dp_s = \int_{\Theta} (\varphi(\gamma_t) - \varphi(\gamma_s)) Q(d\gamma) = \int_s^t \int_{\Theta} \nabla \varphi(\gamma_r) \cdot \dot{\gamma}_r Q(d\gamma) dr.$$

Hence $t \mapsto \int \varphi dp_t$ is absolutely continuous, and for almost every t

$$\frac{d}{dt} \int \varphi dp_t = \int_{\Theta} \nabla \varphi(\gamma_t) \cdot \dot{\gamma}_t Q(d\gamma) = \int_{\mathbb{R}^n} \nabla \varphi(z) \cdot v(t, z) p_t(dz),$$

where the last identity is the disintegration together with the definition of v . As φ was arbitrary, this is (3.29). The initial condition follows from $(e_0)_\# Q = \mu$.

The remaining inequality is Jensen's inequality applied to the probability measure $Q^{t,z}$ and the convex function $|\cdot|^a$, integrated against p_t . For $a > 1$ this function is strictly convex. Therefore, equality forces $\dot{\gamma}_t$ to be $Q^{t,z}$ -almost surely equal to its mean $v(t, z)$, for p_t -almost every z . $\square$

Proposition 3.30 (The Eulerian description is not equivalent). *There are a mass-change factor m , marginals μ, ν and plans $Q_1, Q_2 \in \mathcal{D}_m(\mu, \nu)$ such that*

$$(e_t)_\# Q_1 = (e_t)_\# Q_2 \quad \text{for every } t \in [0, 1], \quad \mathcal{E}_a(Q_1) = \mathcal{E}_a(Q_2),$$

whose barycentric velocity fields coincide, and such that

$$(e_0, e_1)_\# Q_1 \neq (e_0, e_1)_\# Q_2 \quad \text{and} \quad q_{Q_1} \neq q_{Q_2}.$$

In particular, neither the coupling of the endpoints nor the normalisation q_Q is a function of (p, v) , and the conditions (3.25) cannot be expressed in Eulerian variables.

Proof. Let $\mathcal{X} = [-1, 1]$, let M be continuous and positive with $M(2) \neq M(0)$, let $m(x, y) := M(|y - x|)$ and let $\mu = \nu = \frac{1}{2}(\delta_{-1} + \delta_1)$. Let Q_1 give mass $\frac{1}{2}$ to each of the two curves

$$\gamma_t^\pm := \pm(2t - 1),$$

which run from ∓ 1 to ± 1 and cross at the origin at time $\frac{1}{2}$, and let Q_2 give mass $\frac{1}{2}$ to each of the two curves

$$\sigma_t^\pm := \pm|2t - 1|,$$

which run from ± 1 to the origin and back to ± 1 . All four curves are piecewise affine with speed 2. Hence they belong to $AC^a([0, 1]; \mathbb{R})$ and have their endpoints in $\mathcal{X}$.

At every time t the curves of Q_1 occupy the two points $\pm(2t - 1)$ and those of Q_2 the two points $\pm|2t - 1|$. As subsets of $\mathbb{R}$ these coincide, and therefore

$$(e_t)_\# Q_1 = (e_t)_\# Q_2 = \frac{1}{2}(\delta_{-|2t-1|} + \delta_{|2t-1|}).$$

For $t \neq \frac{1}{2}$ exactly one curve of each plan passes through each of these two points, in both cases with velocity 2 at $|2t - 1|$ and -2 at $-|2t - 1|$. At $t = \frac{1}{2}$ both plans put half of their mass on each of the velocities ± 2 at the origin. Consequently, both barycentric fields vanish there, the barycentric velocity fields coincide, and both plans have action $\int_0^1 2^a dt = 2^a$.

Both plans are admissible. Indeed $(e_0)_\# Q_i = \mu$ for $i = 1, 2$. On the support of Q_1 one has $|\gamma_1 - \gamma_0| = 2$, hence $m(e_0, e_1) = M(2)$ and

$$(e_1)_\#(m(e_0, e_1)Q_1) = M(2) \frac{1}{2}(\delta_1 + \delta_{-1}) = q_{Q_1}\nu, \quad q_{Q_1} = M(2) > 0;$$

on the support of Q_2 one has $\gamma_1 = \gamma_0$, hence $m(e_0, e_1) = M(0)$ and, in the same way, $q_{Q_2} = M(0) > 0$. Finally

$$(e_0, e_1)_\# Q_1 = \frac{1}{2}(\delta_{(-1,1)} + \delta_{(1,-1)}), \quad (e_0, e_1)_\# Q_2 = \frac{1}{2}(\delta_{(-1,-1)} + \delta_{(1,1)}),$$

which differ, and $q_{Q_1} = M(2) \neq M(0) = q_{Q_2}$. $\square$

Some trajectories can still be reconstructed. By the superposition principle (see Theorem 8.2.1 in [AGS08]), every solution of (3.29) of finite energy is induced by some plan concentrated on integral curves of v . However, this plan need not be unique. The failure matters here because q_Q and the conditions (3.25) are expressed through $m(\gamma_0, \gamma_1)$, i.e. through both endpoints of one and the same curve. For $m \equiv 1$ no such coupling occurs and only the marginals p_0 and p_1 enter. This is the reason why the classical Benamou-Brenier formulation can be stated in Eulerian variables alone.

In the optimal-map case the situation is different. The curves do not cross and the displacement can be read off from the velocity. Two densities enter. The probability measure p_t records where the transported mass is located at time t , and ϱ_t records how much mass is present there, each particle carrying the mass-change factor that it has accumulated up to time t .

Proposition 3.31 (Eulerian formulation in the optimal-map case). *Let $m(x, y) = M(|y - x|)$ with $M \in C^1([0, \infty); (0, \infty))$ and $M(0) = 1$, let T be a continuously differentiable optimal transport map and let*

$$X_t(x) := (1 - t)x + tT(x)$$

be a diffeomorphism onto its image for every $t \in [0, 1]$. Let Q_T be the dynamical plan of Corollary 3.28, define v by $v(t, X_t(x)) := T(x) - x$ and set

$$p_t := (X_t)_\# \mu, \quad \varrho_t := (X_t)_\# (M(|X_t(x) - x|) \mu(dx)),$$

assuming that both admit Lebesgue densities, which we identify with the measures themselves and denote by $p(t, \cdot)$ and $\varrho(t, \cdot)$. Then, for every $t \in [0, 1]$:

(i) *if $z = X_t(x)$, then $x = z - tv(t, z)$ and $T(x) = z + (1 - t)v(t, z)$; in particular (p, v) determines Q_T and the coupling $(e_0, e_1)_\# Q_T$;*

(ii) *the normalisation is given in Eulerian terms by*

$$q_{Q_T} = \int M(|v(t, z)|) p_t(dz),$$

the right-hand side being independent of t ;

(iii) *the two densities are related by*

$$\varrho(t, z) = M(t|v(t, z)|)p(t, z), \tag{33}$$

and ϱ satisfies the entirely Eulerian equation

$$\partial_t \varrho + \nabla \cdot (\varrho v) = \varrho |v| \frac{M'(t|v|)}{M(t|v|)}, \quad (34)$$

together with $\varrho_0 = \mu$ and $\varrho_1 = q_{Q_T} \nu$;

(iv) conversely, if a pair (p, v) satisfies (3.29) with $p_0 = \mu$ and $\partial_t v + (v \cdot \nabla)v = 0$, and if $x \mapsto x + t v(0, x)$ is injective for every $t \in [0, 1]$, then the map $S(x) := x + v(0, x)$ satisfies $p_t = (x \mapsto (1-t)x + tS(x))_{\#} \mu$, and the plan carried by the straight interpolations of S is the unique dynamical plan with these Eulerian data.

Proof. (i) By definition $z = X_t(x) = (1-t)x + tT(x)$ and $v(t, z) = T(x) - x$. Hence

$$z - t v(t, z) = (1-t)x + tT(x) - t(T(x) - x) = x, \quad \text{and} \quad z + (1-t)v(t, z) = T(x).$$

Both endpoints of the curve passing through z at time t are therefore determined by $(t, z, v(t, z))$. As X_t is a bijection onto its image, every curve of Q_T is recovered in this way, and with it the coupling of the endpoints.

(ii) By (i) we have $|v(t, X_t(x))| = |T(x) - x|$ for every x . Consequently,

$$\int M(|v(t, z)|) p_t(dz) = \int_{\mathcal{X}} M(|T(x) - x|) \mu(dx) = q_{Q_T},$$

which does not depend on t .

(iii) Let $\varphi \in \mathcal{C}_c(\mathbb{R}^n)$ and substitute $z = X_t(x)$. This yields

$$\int \varphi d\varrho_t = \int_{\mathcal{X}} \varphi(X_t(x)) M(|X_t(x) - x|) \mu(dx) = \int \varphi(z) M(|z - X_t^{-1}(z)|) p_t(dz).$$

Since $X_t(x) - x = t(T(x) - x) = t v(t, X_t(x))$, we have $|z - X_t^{-1}(z)| = t|v(t, z)|$, which is (3.31). The velocity is constant along the straight trajectories, that is

$$\partial_t v + (v \cdot \nabla)v = 0.$$

Let us abbreviate $w := M(t|v|)$ and write $D_t := \partial_t + v \cdot \nabla$ for the derivative along the trajectories. By (3.31) we have $\varrho = w p$, and the product rule gives

$$\partial_t \varrho + \nabla \cdot (\varrho v) = w(\partial_t p + \nabla \cdot (pv)) + p(\partial_t w + v \cdot \nabla w) = p D_t w,$$

because the first bracket vanishes by (3.29), which applies by Lemma 3.29. The identity above states precisely that $D_t v = 0$ and hence also $D_t |v| = 0$. Therefore,

$$D_t w = M'(t|v|) D_t(t|v|) = M'(t|v|)(|v| + t D_t |v|) = |v| M'(t|v|).$$

Substituting $p = \varrho/M(t|v|)$ yields (3.31). The endpoint identities are the admissibility conditions (3.25) for Q_T . Indeed, at $t = 0$ the accumulated factor

is $M(0) = 1$ and therefore $\varrho_0 = \mu$. At $t = 1$ the weighted second marginal gives $\varrho_1 = q_{Q_T} \nu$ with $q_{Q_T} = \int_{\mathcal{X}} M(|T(x) - x|) \mu(dx)$.

(iv) The identity $\partial_t v + (v \cdot \nabla)v = 0$ states that v is constant along its own integral curves. Hence $v(t, x + tv(0, x)) = v(0, x)$, and the integral curves are the straight lines $t \mapsto x + tv(0, x)$. These are the straight interpolations between x and $S(x) = x + v(0, x)$. The measure $(x \mapsto (1 - t)x + tS(x))_{\#} \mu$ solves (3.29) with the same initial datum and the same velocity field. By the injectivity assumption, the integral curve through a given point is unique. Consequently, this measure coincides with p_t , and the plan concentrated on those curves is the only one with the given Eulerian data. $\square$

For $M(r) = e^{-kr}$ with $k > 0$ the right-hand side of (3.31) equals $-k|v|\varrho$, and for $M(r) = e^{r^2/2}$ it equals $t|v|^2\varrho$.

4 Relation to existing optimal–transport variants

In this section we compare the framework of this article to existing variants of optimal transport which allow flexibility in the marginal constraints. As the literature on those variants is abundant, we do not claim that this overview is exhaustive. We focus on three popular approaches: unbalanced optimal transport in Section 4.1, entropic optimal transport in Section 4.2, and unnormalized optimal transport in Section 4.3.

4.1 Relation to unbalanced optimal transport

Unbalanced optimal transport became an umbrella term for optimal transport variants that weaken the original marginal conditions of a transport plan. However, here we use the term unbalanced optimal transport to refer to the framework originally developed in [KMV16, CPSV18a] and extended in [CPSV18b]. The starting point is the dynamic formulation of optimal transport via the Benamou-Brenier formula, allowing for mass creation or destruction along the trajectories of mass transport. In [CPSV18b], it was shown that the dynamic formulation is equivalent to a static optimization problem described below.

Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^n$ denote two compact sets. In unbalanced optimal transport, the cost function $\bar{c} = \bar{c}(x_0, m_0, x_1, m_1)$ determines the cost of transporting/transforming a Dirac mass of size m_0 located at x_0 into a Dirac mass of size m_1 located at x_1 . In this context, the cost function $\bar{c}: (\mathcal{X} \times [0, \infty)) \times (\mathcal{Y} \times [0, \infty)) \rightarrow [0, \infty]$ is assumed to be l.s.c. in all its arguments and jointly sublinear in (m_0, m_1) , i.e. positively 1-homogeneous and subadditive.

In classical optimal transport, the decision space consists of couplings, which preserve mass. To allow for creation or destruction of mass, the unbalanced optimal transport introduces the notion of semi-couplings (π_0, π_1) , namely a pair of two couplings. The interpretation is that $\pi_0(x, y)$ represents the amount

of mass taken from μ at point x and then transported to a possibly different amount of mass $\pi_1(x, y)$ at point y of ν .

Definition 4.1 (Semi-couplings). *Given two marginals $\mu \in \mathcal{M}_+(\mathcal{X})$ and $\nu \in \mathcal{M}_+(\mathcal{Y})$, the set of semi-couplings is*

$$\bar{\Gamma}(\mu, \nu) \stackrel{\text{def}}{=} \{(\pi_0, \pi_1) \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})^2 : (\text{Proj}_0)\pi_0 = \mu, (\text{Proj}_1)\pi_1 = \nu\}.$$

This leads to the static Kantorovich formulation of unbalanced optimal transport.

Definition 4.2 (Unbalanced Kantorovich problem). *The static unbalanced optimal transport problem is given by*

$$\inf_{(\pi_0, \pi_1) \in \bar{\Gamma}(\mu, \nu)} \iint_{\mathcal{X} \times \mathcal{Y}} \bar{c}\left(x, \frac{\pi_0}{\gamma}, y, \frac{\pi_1}{\gamma}\right) d\gamma(x, y), \quad (\text{UKP})$$

where $\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ is any measure such that $\pi_0, \pi_1 \ll \gamma$. The integral is well-defined since $\bar{c}$ is jointly 1-homogeneous w.r.t. the mass variables.

The interpretation of (UKP) is that one looks for the most efficient way to transform the measure μ into the measure ν by a combination of transport and of creation or destruction of mass along the way. In [CPSV18b, Proposition 3.4] it was shown that (UKP) admits a minimizer. The dual problem of (UKP) is quite subtle and the existence of a solution of the dual problem is not easily guaranteed. The same is true for strong duality. The main difference between (UKP) and (KP) is the following. In (UKP) the mass difference between μ and ν is fixed. In (KP), only the shape of the second marginal ν is prescribed, allowing for couplings with varying mass differences. There is the following link between both problems.

Proposition 4.3 (Link between (KP) and (UKP)). *Let c and m be the cost and mass-change factor of (KP). Then the function*

$$\bar{c}(x, y, m_0, m_1) = \begin{cases} c(x, y), & \text{if } m_1 = m(x, y)m_0, \\ \infty, & \text{else,} \end{cases}$$

is a cost function $\bar{c}: (\mathcal{X} \times [0, \infty)) \times (\mathcal{Y} \times [0, \infty)) \rightarrow [0, \infty]$ and

$$\inf_{\Gamma_m(\mu, \nu)} \iint c(x, y) \pi(dx, dy) = \inf_{Z > 0} \inf_{\bar{\Gamma}(\mu, Z\nu)} \iint_{\mathcal{X} \times \mathcal{Y}} \bar{c}\left(x, \frac{\pi_0}{\gamma}, y, \frac{\pi_1}{\gamma}\right) d\gamma(x, y).$$

In particular, if $Z = \iint m(x, y) \pi^*(dx, dy)$ for the optimal π^* of (KP), then

$$\inf_{\pi \in \Gamma_m(\mu, \nu)} \iint c(x, y) \pi(dx, dy) = \inf_{\bar{\Gamma}(\mu, Z\nu)} \iint_{\mathcal{X} \times \mathcal{Y}} \bar{c}\left(x, \frac{\pi_0}{\gamma}, y, \frac{\pi_1}{\gamma}\right) d\gamma(x, y).$$

The last proposition states that one can recover the solution (KP) via minimizing over a family of (UKP) problems. If one also knows how much mass is lost or generated in the optimal (KP) transport then the problems (KP) and (UKP) are equivalent. However, this is not known a priori. The verification of the last proposition follows directly from careful examination of the definitions and is left as an exercise. The simpler structure of (KP) compared to (UKP) allows for easier arguments when solving the dual problem, deriving strong duality, or studying optimal transport maps. However, the framework (UKP) allows one to model more general situations.

4.2 Relation to entropic optimal transport in the sense of [LMS18]

The entropic transport problem, introduced in [LMS18], relaxes classical mass-conserving optimal transport by allowing creation and annihilation of mass, captured via entropic penalization. Entropic transport has been a very active area and an overview of recent developments can be found e.g. in [SPV23]. For the discussion, let us briefly introduce the main concepts of entropic optimal transport in a slightly simplified setting.

Definition 4.4 (Entropy functions and relative entropies). *A function $F : [0, \infty) \rightarrow [0, \infty]$ is called an admissible entropy function if it is convex, lower semicontinuous and there is a point $0 < x < \infty$ such that $F(x) < \infty$. Given two finite measures γ_1 and γ_2 on a space $\mathcal{X}$, the relative entropy $\mathcal{F}(\gamma_1|\gamma_2)$ w.r.t. entropy function F is defined as*

$$\mathcal{F}(\gamma_1|\gamma_2) = \begin{cases} \int F\left(\frac{d\gamma_1}{d\gamma_2}\right) d\gamma_2, & \text{if } \gamma_1 \ll \gamma_2, \\ \infty, & \text{else.} \end{cases}$$

Definition 4.5 (Entropic Transport Problem). *We consider two measures μ on $\mathcal{X}$ and ν on $\mathcal{Y}$ and relative entropies $\mathcal{F}_1$ and $\mathcal{F}_2$ on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Given a lower semicontinuous cost function $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ we optimize the expression*

$$\inf_{\pi \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy) + \mathcal{F}_1((\text{Proj}_1)_\# \pi | \mu) + \mathcal{F}_2((\text{Proj}_2)_\# \pi | \nu).$$

Remark 4.6. Entropic transport contains the classical optimal transport problem as a degenerate case, namely choosing the admissible entropy function as

$$F(z) = \begin{cases} 0, & \text{if } z = 1 \\ \infty, & \text{else.} \end{cases}$$

Remark 4.7. The entropy functionals in Definition 4.5 are applied to the marginals of the coupling π . In the optimization problem (KP), the second marginal of the transport is not determined a priori. So the problem of Definition 4.5 does not contain (KP) unless it is generalized to the following degenerate

problem. Given a measure $\pi \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ let us consider the transformation $T(\pi) \in \mathcal{P}(\mathcal{Y})$ defined by

$$T(\pi) = \frac{\int_{\mathcal{X}} m(x, y) \pi(dx, y)}{\int \int_{\mathcal{X} \times \mathcal{Y}} m(x, y) \pi(dx, dy)}.$$

Then (KP) becomes the following generalized entropic transport problem.

$$\inf_{\pi \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \pi(dx, dy) + \mathcal{F}((\text{Proj}_1)_\# \pi | \mu) + \mathcal{F}(T(\pi) | \nu).$$

4.3 Relation to unnormalized optimal transport

Unnormalized optimal transport [GLOP19] uses a computational fluid mechanics approach to allow the transport of two measures with different mass. A spatially homogeneous term is added to the continuity equation to account for the creation or loss of mass. Additionally, a regularization term is added to the minimization of the kinetic energy, leading to a generalization of the Benamou-Brenier formulation of optimal transport.

Definition 4.8 (Unnormalized Optimal Transport). *Let us consider a bounded domain $\mathcal{X} \subset \mathbb{R}^n$, two nonnegative measures $\mu, \nu \in \mathcal{M}(\mathcal{X})$, and parameters $p \geq 1$ and $\alpha > 0$. The unnormalized optimal transport problem is given by*

$$\inf_{v, f} \int_0^1 \int_{\mathcal{X}} \|v(t, x)\|^p \varrho(t, x) dx dt + \frac{1}{\alpha} \int_0^1 |f(t)|^p dt,$$

such that the vector field v satisfies the zero flux condition and $\varrho(t)$ satisfies the dynamical constraint

$$\partial_t \varrho + \nabla \cdot (v \varrho) = f(t) \quad \text{and} \quad \varrho(0) = \mu, \quad \varrho(1) = \nu.$$

The regularization term $\int_0^1 |f(t)|^p dt$ constitutes the main difference to the other variants of optimal transport. The dynamic formulation (GBB) of (KP) does not produce such a regularization term (see Section 3.5) and therefore the two optimal transport variants are distinct.

4.4 Entropic regularization and a Sinkhorn-type algorithm

Let us discuss how the non-conservative Kantorovich problem (KP) can be solved numerically. We follow the approach of entropic regularization, in which a relative entropy is added to the objective and which, in the classical case, leads to the Sinkhorn algorithm of [Cut13]. Note that this is a different use of the word entropic than in Section 4.2, where the marginal constraints are relaxed by entropy functionals in the sense of [LMS18].

On finite state spaces the approach carries over without difficulty, since with the linear structure of Section 3.3 the problem (KP) becomes a linear program; see Proposition 4.9 below.

The continuous setting is more delicate. In the classical case the dual objects are the additively separable functions $\varphi(x) + \psi(y)$, and the existence theory for the associated potentials rests on this tensor structure, see [Nut22] for a modern account. Here they are of the form $\varphi(x) + m(x, y)\psi(y)$, where ψ is multiplied by the x -dependent factor $m(x, y)$ and can therefore not be isolated. The regularized problem is then of Schrödinger type, with the second marginal constraint weighted by m and self-normalizing, so that the mass q_π enters as an additional scalar unknown, without counterpart in the classical Schrödinger system. Moreover, the quantitative bound of Proposition 4.9 (v) depends on the number of states and degenerates in the continuum limit, while its classical counterparts are independent of the discretization. We are not aware of an existence result for potentials of this form in the continuous setting and leave this question open, together with the convergence of the corresponding scaling iteration.

Let us restrict to finite state spaces $\mathcal{X} = \{x_1, \dots, x_N\}$ and $\mathcal{Y} = \{y_1, \dots, y_M\}$, and let us write $c_{ij} = c(x_i, y_j)$, $m_{ij} = m(x_i, y_j)$, $\mu_i = \mu(\{x_i\})$ and $\nu_j = \nu(\{y_j\})$. Furthermore, let C_1 and C_2 denote the sets of nonnegative matrices satisfying $(\text{Proj}_1)_\# \pi = \mu$ and $L\pi = 0$, respectively, where $(L\pi)_j = \sum_i m_{ij}\pi_{ij} - q_\pi \nu_j$ with $q_\pi = \sum_{ij} m_{ij}\pi_{ij}$, so that $\Gamma_m(\mu, \nu) = C_1 \cap C_2$. For nonnegative matrices we write

$$D(\sigma|\pi) = \sum_{ij} \left(\sigma_{ij} \log \frac{\sigma_{ij}}{\pi_{ij}} - \sigma_{ij} + \pi_{ij} \right),$$

which is the usual relative entropy whenever σ and π have the same total mass.

Proposition 4.9 (Entropic regularization and scaling iteration). *Let $m_{ij} > 0$ and let μ and ν be strictly positive. For $\varepsilon > 0$ consider*

$$\inf_{\pi \in \Gamma_m(\mu, \nu)} \sum_{ij} c_{ij} \pi_{ij} + \varepsilon \sum_{ij} \pi_{ij} \log \pi_{ij}.$$

Then the following holds.

- (i) *The problem has a unique minimizer π^ε , which is strictly positive and minimizes $D(\cdot|K)$ over $\Gamma_m(\mu, \nu)$, where $K_{ij} = e^{-c_{ij}/\varepsilon}$.*
- (ii) *There are $\varphi \in \mathbb{R}^N$ and $\psi \in \mathbb{R}^M$ with $\sum_j \nu_j \psi_j = 0$ such that $\pi_{ij}^\varepsilon = \exp((\varphi_i + m_{ij}\psi_j - c_{ij})/\varepsilon)$, and this pair is unique.*
- (iii) *The minimizer of $D(\cdot|\pi)$ over C_1 is the rescaling $\pi_{ij} \mapsto \pi_{ij}\mu_i/\sum_l \pi_{il}$, and the minimizer of $D(\cdot|\pi)$ over C_2 is*

$$\pi_{ij} \mapsto \pi_{ij} b_j^{m_{ij}},$$

where $b_j > 0$ and $q > 0$ are the unique solutions of $\sum_i m_{ij}\pi_{ij} b_j^{m_{ij}} = q\nu_j$ and $\sum_j \nu_j \log b_j = 0$.

- (iv) *The iteration alternating the two minimizations of (iii) converges to π^ε .*

(v) It holds $0 \leq \sum_{ij} c_{ij} \pi_{ij}^\varepsilon - \min(\text{KP}) \leq \varepsilon \log(NM)$.

The proof is given in Appendix B. It combines classical arguments from finite-dimensional convex optimization. Since the constraints are linear and the minimizer is strictly positive, parts (ii) and (iii) follow from the Lagrange multiplier method; the only computation specific to our setting is that of the multipliers for the second marginal constraint. Part (iv) is the classical method of iterative Bregman projections [Bre67, BCC⁺15, vLS23], and part (v) follows by comparing the values of the regularized objective at π^ε and at a minimizer of (KP).

Acknowledgments

We want to express our deepest gratitude to Wilfrid Gangbo for his interaction. His insights and knowledge immensely helped to write this article. We also want to thank Lenaic Chizat, Pablo López Rivera, Bernhard Schmitzer, and Moritz Voss for their valuable comments. We are grateful to Matthias Erbar for helpful discussions and to the anonymous referee for a careful reading and constructive comments. During the revision of this manuscript the authors used Claude (Anthropic; Opus 5 and Fable 5 models) for editing and for drafting assistance under the authors' direction; no AI tools were used in the preparation of the original version. All mathematical statements and proofs were checked by the authors, who reviewed and edited all generated text and take full responsibility for the content of the publication.

Statements and Declarations

Gabriela Kováčová acknowledges support from the Society for Industrial and Applied Mathematics (SIAM) as part of the SIAM Postdoctoral Support Program. The authors declare that no other funds, grants, or other support were received during the preparation of this manuscript. The authors have no relevant financial or non-financial interests to disclose.

A Proofs of results from Section 2

A.1 Quoted prices, best execution prices and the consistent price vector

Proof of Lemma 2.4. Since the market $\mathcal{G} = (V, E, p)$ satisfies assumptions A1 and A2, a graph $\mathcal{H} = (V, E, w)$ defined by setting

$$w_{(i,j)} := -\log p_{(i,j)} \quad \text{for all } (i,j) \in E$$

is strongly connected and for any directed cycle Γ (in $\mathcal{G}$ or $\mathcal{H}$) it holds

$$\prod_{e \in \Gamma} p_e \leq 1, \quad \text{or equivalently,} \quad \sum_{e \in \Gamma} w_e \geq 0.$$

Therefore, graph $\mathcal{H}$ contains no cycles of negative weight and by [CLRS09, Theorem 24.4] the shortest path between any pair of vertices in $\mathcal{H}$ exists. Since $\arg \min_{\Gamma: i \rightarrow j} \sum_{e \in \Gamma} w_e = \arg \max_{\Gamma: i \rightarrow j} \prod_{e \in \Gamma} p_e$, the best execution prices are well defined.

Finally, consider a directed cycle in the completed graph. Such cycle expands into a closed walk in graph $\mathcal{G}$ when edges in $\bar{\mathcal{G}}$ are replaced by the best execution price paths. The no arbitrage property for P follows from A2. $\square$

Definition 2.6 introduces trades on the best execution price market $\bar{\mathcal{G}} = (V, V \times V, P)$. For practical implementation of rebalancing, one needs to know how to implement such trade on the market $\mathcal{G}$. We will call a map $\zeta : V \times V \rightarrow [0, \infty)$ a *trade on $\mathcal{G}$* if it satisfies $\zeta_{(i,j)} = 0$ for $(i,j) \notin E$.

Paths along which the best execution prices P are achieved will be needed to convert a trade ξ on the completed market $\bar{\mathcal{G}}$ into a trade ζ on the underlying market $\mathcal{G}$. Remark A.2 below comments on computation of the best execution price paths.

Example A.1. Consider a trade on the star-shaped market that uses only the exchanges with the numéraire. A positive bid-ask spread ensures that a round-trip exchanges (buying and selling the same asset) results in a net loss, effectively 'burning money'. In terms of a trade ζ on $\mathcal{G}$, imposing

$$\xi_{(1,i)} \cdot \xi_{(i,1)} = 0 \quad i = 2, \dots, N \quad (36)$$

excludes round-trip exchanges. Then, a trade the star-shaped market is fully determined by the change in the portfolio position $\Delta \mathbf{x}$, since $(\Delta x_i)^+ = \frac{1}{P_i^a} \xi_{(1,i)}$ and $(\Delta x_i)^- = \xi_{(i,1)}$ for $i > 1$. Equation (2.1) corresponds to the *self-financing condition*

$$\Delta x_1 = \sum_{i=2}^N P_i^b (\Delta x_i)^- - P_i^a (\Delta x_i)^+.$$

Proof of Proposition 2.10. To prove that no arbitrage is *necessary*, assume that a consistent price vector $\mathbf{q}$ exists. Let $\Gamma = ((v_1, v_2), (v_2, v_3), \dots, (v_M, v_1))$ be an arbitrary directed cycle in the graph. Then, by definition of the best execution price and the consistent price vector, it holds

$$\prod_{e \in \Gamma} p_e \leq \prod_{e \in \Gamma} P_e \leq \frac{q_{v_1}}{q_{v_2}} \cdot \frac{q_{v_2}}{q_{v_3}} \cdot \dots \cdot \frac{q_{v_M}}{q_{v_1}} = 1.$$

Now let us show that no arbitrage is *sufficient*. We apply very similar tools as in the proof of Lemma 2.4. Define a graph $\bar{\mathcal{H}} = (V, V \times V, \bar{w})$ by setting

$$\bar{w}_{(i,j)} := -\log P_{(i,j)} \quad \text{for all } (i,j) \in V \times V.$$

Since $\bar{\mathcal{G}}$ satisfies no arbitrage, $\bar{\mathcal{H}}$ contains no negative cycles and the shortest paths exist by [CLRS09, Theorem 24.4]. Denote by ϕ_i the length of the shortest directed path from 1 to i in graph $\bar{\mathcal{H}}$. Then, for any pair of vertices i, j it holds

$$\phi_j \leq \phi_i + w_{(i,j)}, \quad \text{or equivalently,} \quad \log P_{(i,j)} = -w_{(i,j)} \leq \phi_i - \phi_j.$$

By setting $q_i = e^{\phi_i}$ for $i = 1, \dots, N$ we obtain $P_{(i,j)} \leq q_i/q_j$ for arbitrary $i, j \in V$.

Hence $\mathbf{q}$ is a consistent price vector. $\square$

Remark A.2. The Bellman-Ford algorithm, see [CLRS09, Section 24.1], can be used to check whether the market $\mathcal{G}$ satisfies the no arbitrage condition and a consistent price vector exists. Run with the edge weights $-\log p_e$, the same computation also yields the best-execution prices and corresponding best paths. In practice, a trade $\xi_{(i,j)}$ on completed graph $\bar{\mathcal{G}}$ is therefore executed as several successive exchanges along a best path (cf. Remark 2.3), in the same way as the cost of the classical Monge problem is measured along geodesics; an explicit construction appears after the proof of Proposition A.6.

A.2 Existence of optimal rebalancing

As the first step towards proving Theorem 2.14, let us observe that the rebalancing problem is a linear program with an objective bounded from below.

Lemma A.3. *For a trade $\xi \in \Xi$ it holds $C(\xi) \geq 0$. Moreover, the rebalancing Problem 2.12 is (equivalent to) a linear program.*

Proof. Recall that $C(\xi)$ can be expressed via (2.16). Non-negativity $\xi_{(i,j)} \geq 0$ of the trade, consistency $P_{(i,j)} \leq \frac{q_i}{q_j}$ of $\mathbf{q}$ and non-negative prices P imply $C(\xi) \geq 0$.

To simplify restating the target proportion constraints, let us introduce an additional variable $\eta := C(\xi)$. The rebalancing problem can be reformulated as

$$\begin{aligned} & \underset{\xi \in \Xi, \eta \in \mathbb{R}}{\text{minimize}} && \eta \\ \text{subject to} &&& \eta = \sum_{i=1}^N q_i \sum_{j=1}^N (\xi_{(i,j)} - P_{(j,i)} \xi_{(j,i)}), \\ &&& x_i + \sum_{j=1}^N (P_{(j,i)} \xi_{(j,i)} - \xi_{(i,j)}) \geq 0, \quad i \in V, \\ &&& \nu_i \left(\sum_{k=1}^N q_k x_k - \eta \right) = q_i \left(x_i + \sum_{j=1}^N (P_{(j,i)} \xi_{(j,i)} - \xi_{(i,j)}) \right), \quad i \in V. \end{aligned}$$

Since the objective and all constraints are linear in ξ and η jointly, this is a linear program. The reformulation admits additional feasible points with $\eta = \sum_k q_k x_k$, i.e. with all post-trade positions equal to zero, for which the proportion constraint of Problem 2.12 is undefined. These do not affect optimality: a feasible trade with positive post-trade value exists by Proposition A.6 below, so every minimizer satisfies $\eta < \sum_k q_k x_k$ and corresponds to a feasible rebalancing trade. $\square$

As the next step towards proving Theorem 2.14, we consider the special case of a *star-shaped market*. We use the terminology and notation introduced in Examples 2.5, A.1 and 2.9.

Example A.4. Consider the *star-shaped market*. Given a current portfolio $\mathbf{x} = (x_1, x_2, \dots, x_N)$, we aim to show existence of a *feasible* rebalancing trade yielding the desired target proportions ν . Note we do not aim for the optimal trade here, instead we use a (sub-optimal) strategy consisting of two steps:

1. First, trade all available units of assets $i = 2, \dots, N$ for the numéraire via $\xi_{(i,1)} := x_i$, creating a 'money pot' portfolio

$$\left(x_1 + \sum_{i=2}^N P_i^b x_i, 0, \dots, 0 \right).$$

2. Second, use the 'money pot' held in the numéraire only to buy a portfolio with the desired target proportions. Proposition A.5 below shows that such a trade exists.

Note that a trade constructed in this way will not satisfy (A.1).

Proposition A.5. *Consider the star-shaped market as described in Examples 2.5, A.1 and 2.9. For a portfolio $\mathbf{x} = (1, 0, \dots, 0)$ there exists a feasible rebalancing trade $\xi \in \Xi(\mathbf{x}, \nu)$ achieving the desired target ν .*

Proof. Firstly, set

$$\xi_{(i,j)} = 0 \quad \text{if } i \neq 1 \text{ or } j = 1.$$

Consequently, the trade ξ is fully determined by the quantities $(\xi_{(1,i)})_{i=2,\dots,N}$. For the sake of a simpler notation in the remainder of the proof we denote by $\mathbf{z} := \mathbf{x} + \Delta\mathbf{x}(\xi)$ the portfolio position after the trade. Recall that it holds

$$z_i = P_{(1,i)} \xi_{(1,i)} = \frac{1}{P_i^a} \xi_{(1,i)} \quad i = 2, \dots, N \quad (37)$$

and

$$z_1 = x_1 - \sum_{i=2}^N \xi_{(1,i)} = 1 - \sum_{i=2}^N P_i^a z_i. \quad (38)$$

Note that (A.2) represents the self-financing condition and a trade $\xi \in \Xi$ can be recovered from $\mathbf{z}$ via (A.2).

The constraints of the feasible set $\Xi(\mathbf{x}, \nu)$ expressed in terms of after-trade position $\mathbf{z}$ are

$$z_i \geq 0 \quad i = 1, \dots, N, \quad (39)$$

$$q_i z_i + \nu_i \cdot \sum_{j=1}^N (P_j^a - q_j) z_j = \nu_i, \quad i = 1, \dots, N, \quad (40)$$

where the self-financing condition (A.2) was used to obtain (A.2). Recall that $P_1^a = q_1 = 1$ and $P_i^a \geq q_i$ for $i = 2, \dots, N$. In order to show existence of a vector $\mathbf{z}$ solving the system (A.2)-(A.2) we use the Farkas' Lemma. Let us, therefore, introduce vectors $\mathbf{q} = (q_1, \dots, q_N)^T$ and $\mathbf{P}_a = (P_1^a, \dots, P_N^a)^T$. In matrix notation, the system (A.2)-(A.2) is

$$(\text{diag } \mathbf{q} + \nu \cdot (\mathbf{P}_a - \mathbf{q})^T) \cdot \mathbf{z} = \nu, \quad \mathbf{z} \geq 0. \quad (41)$$

According to Farkas' Lemma, solution of (A.2) exists if and only if the complementary system

$$(\text{diag } \mathbf{q} + \nu \cdot (\mathbf{P}_a - \mathbf{q})^T)^T \cdot \mathbf{y} \geq 0, \quad \nu^T \mathbf{y} < 0 \quad (42)$$

has no solution. Consider (A.2): $\nu^T \mathbf{y} < 0$ implies that there exists $j \in \{1, \dots, N\}$ with $y_j < 0$. Looking at the j -th row of $(\text{diag } \mathbf{q} + \nu \cdot (\mathbf{P}_a - \mathbf{q})^T)^T \cdot \mathbf{y}$, we obtain a contradiction

$$\left[(\text{diag } \mathbf{q} + \nu \cdot (\mathbf{P}_a - \mathbf{q})^T)^T \cdot \mathbf{y} \right]_j = q_j y_j + (P_j^a - q_j) \cdot \nu^T \cdot \mathbf{y} < 0$$

since $P_j^a \geq q_j > 0$. Therefore, the complementary system (A.2) admits no solution and, by Farkas' Lemma, a solution $\mathbf{z}$ to system (A.2) exists. Finally, note that any vector $\mathbf{z}$ satisfying (A.2) also satisfies the self-financing condition (A.2). Therefore, vector $\mathbf{z}$ corresponds to a feasible trade $\xi \in \Xi(\mathbf{x}, \nu)$ that can be recovered via (A.2). $\square$

On the completed market $\bar{\mathcal{G}}$, a feasible trade can be constructed in two single-hop stages.

Proposition A.6. *For a portfolio $\mathbf{x} = (x_1, x_2, \dots, x_N) \geq 0$, $\mathbf{x} \neq 0$, on a (general) market $\mathcal{G}$ there exists a feasible rebalancing trade $\xi \in \Xi(\mathbf{x}, \nu)$ achieving the desired target ν .*

Proof. In the first stage, we convert the portfolio $\mathbf{x}$ into a ‘money pot’ portfolio

$$\tilde{\mathbf{x}} = \left(x_1 + \sum_{i=2}^N P_{(i,1)} x_i, 0, \dots, 0 \right)$$

by the trade $\xi^{(1)}$ with $\xi_{(i,1)}^{(1)} := x_i$ for $i = 2, \dots, N$ and all other entries zero. Since $\mathbf{x} \neq 0$ and all best-execution prices are positive, the money pot is positive. In the second stage, Proposition A.5, applied with the quotes $P_i^b = P_{(i,1)}$ and $P_i^a = 1/P_{(1,i)}$ and combined with scaling, yields a trade $\xi^{(2)}$ supported on the pairs $(1, i)$ that converts $\tilde{\mathbf{x}}$ into a portfolio with the target proportions ν . Since $\Delta \mathbf{x}(\cdot)$ is linear in the trade, applying $\xi := \xi^{(1)} + \xi^{(2)}$ to $\mathbf{x}$ yields the same final portfolio, so $\xi \in \Xi(\mathbf{x}, \nu)$. $\square$

On the underlying graph $\mathcal{G}$, a trade on the completed market is realized by successive exchanges along best paths, cf. Remark A.2: for a pair (i, j) with

$\xi_{(i,j)} > 0$, let $(e_1, \dots, e_M)$ be the edges of a directed path attaining $P_{(i,j)}$ and execute $\zeta_{e_1} := \xi_{(i,j)}$ and $\zeta_{e_k} := P_{e_{k-1}} \zeta_{e_{k-1}}$ for $k = 2, \dots, M$. Each exchange sells exactly the amount delivered by the previous one, so all intermediate positions cancel and the executed exchanges have the same net effect $\Delta \mathbf{x}$ and the same cost as the trade ξ .

Proof of Theorem 2.14. Proposition A.6 has shown that a feasible rebalancing trade exists, therefore, the optimal rebalancing Problem 2.12 is feasible. According to Lemma A.3, Problem 2.12 is (equivalent to) a linear program and its objective function is bounded from below by zero. For a feasible, bounded linear program an optimal solution exists. $\square$

A.3 Relation of rebalancing to non-conservative OT

Lemma A.7. *Let Assumptions A1-A7 hold and denote $v := \sum_{k=1}^N q_k x_k$.*

- a) *If $\xi \in \Xi(\mathbf{x}, \nu)$ is feasible for Problem 2.12 and direct, i.e. $\sum_{j \neq i} \xi_{(i,j)} \leq x_i$ for all i , then π^ξ defined as*

$$\pi^\xi(i, j) = \begin{cases} \frac{q_i}{v} \xi_{(i,j)} & \text{if } i \neq j, \\ \frac{q_i}{v} \left(x_i - \sum_{k \neq i} \xi_{(i,k)} \right) & \text{if } i = j, \end{cases}$$

is feasible for Problem 2.17, i.e. $\pi^\xi \in \Gamma_m(\mu, \nu)$.

- b) *If $\pi \in \Gamma_m(\mu, \nu)$ is feasible for Problem 2.17, then ξ^π defined as*

$$\xi_{(i,j)}^\pi = \begin{cases} \frac{v}{q_i} \pi(i, j) & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases}$$

is a direct trade that is feasible for Problem 2.12, i.e. $\xi^\pi \in \Xi(\mathbf{x}, \nu)$.

Proof. a) Since ξ is direct, $\pi^\xi(i, i) \geq 0$, and all other entries are nonnegative as well. The rows sum to

$$\sum_{j=1}^N \pi^\xi(i, j) = \frac{q_i}{v} \left(x_i - \sum_{k \neq i} \xi_{(i,k)} \right) + \sum_{j \neq i} \frac{q_i}{v} \xi_{(i,j)} = \frac{q_i x_i}{v} = \mu_i.$$

Using $m(j, j) = 1$, $m(i, j) = q_j P_{(i,j)} / q_i$ and (2.1),

$$\begin{aligned} \sum_{i=1}^N m(i, j) \pi^\xi(i, j) &= \frac{q_j}{v} \left(x_j - \sum_{k \neq j} \xi_{(j,k)} + \sum_{i \neq j} P_{(i,j)} \xi_{(i,j)} \right) \\ &= \frac{q_j}{v} (x_j + \Delta x_j(\xi)), \end{aligned}$$

and since $\xi \in \Xi(\mathbf{x}, \nu)$, the right-hand side is proportional to ν_j with positive total mass. Hence $\pi^\xi \in \Gamma_m(\mu, \nu)$.

- b) The trade ξ^π belongs to Ξ , and it is direct since the row constraint in $\Gamma_m(\mu, \nu)$ gives $\sum_{j \neq i} \xi_{(i,j)}^\pi = \frac{v}{q_i} (\mu_i - \pi(i, i)) \leq \frac{v\mu_i}{q_i} = x_i$. As in a), using (2.1) and $x_i = v\mu_i/q_i$,

$$x_i + \Delta x_i(\xi^\pi) = \frac{v}{q_i} \sum_{k=1}^N m(k, i) \pi(k, i) = \frac{v q_\pi}{q_i} \nu_i \geq 0,$$

where $q_\pi = \sum_{k,l} m(k, l) \pi(k, l) > 0$ is the total m -weighted mass of π . Hence the post-trade positions are nonnegative with proportions ν , i.e. $\xi^\pi \in \Xi(\mathbf{x}, \nu)$. $\square$

The second lemma shows that re-selling value received within the same trading period never improves the optimal cost. Part a) is the key device: a portfolio whose value dominates a ν -proportional floor, i.e. $q_i z_i \geq \lambda \nu_i$ for all i , can be traded exactly onto the target proportions without falling below the floor. It is applied in the proof of part b) to turn a strict gain in a single asset into a strictly cheaper feasible trade.

Lemma A.8. *Let Assumptions A1-A7 hold and denote $v := \sum_{k=1}^N q_k x_k$.*

- a) *Let ξ' be a trade with $\mathbf{z} := \mathbf{x} + \Delta \mathbf{x}(\xi') \geq 0$ and suppose that, for some $\lambda > 0$, it holds $q_i z_i \geq \lambda \nu_i$ for all $i = 1, \dots, N$. Then there exists $\xi'' \in \Xi(\mathbf{x}, \nu)$ with*

$$C(\xi'') = v - \lambda - \sum_{i=1}^N \lambda_i \leq v - \lambda, \quad \text{where } \lambda_i := \frac{q_i z_i - \lambda \nu_i}{\sum_{l=1}^N \nu_l / m(i, l)} \geq 0.$$

- b) *Among the optimal solutions of Problem 2.12 there are trades that minimize the total sold value $\sum_{i \neq j} q_i \xi_{(i,j)}$, and every such minimizer is a direct trade.*

Proof. For a): write $e_i := q_i z_i - \lambda \nu_i \geq 0$ for the excess value at asset i . We sell each excess onto the target proportions: from asset i , the value $\lambda_i \nu_l$ is to be delivered to every asset l , where the share $\lambda_i \nu_i$ simply stays at i (recall $m(i, i) = 1$), and delivering $\lambda_i \nu_l$ to $l \neq i$ requires selling the inflated value $\lambda_i \nu_l / m(i, l)$. The requirement that these values exhaust the excess of asset i reads

$$\lambda_i \nu_i + \sum_{l \neq i} \frac{\lambda_i \nu_l}{m(i, l)} = \lambda_i \sum_{l=1}^N \frac{\nu_l}{m(i, l)} = e_i.$$

Since $m \in (0, 1]$, the coefficient $\sum_l \nu_l / m(i, l)$ lies in $[1, \infty)$ — it is finite precisely because every pair of assets is tradable — so this equation has the unique solution λ_i given in the statement, and $0 \leq \lambda_i \leq e_i$. In units, the additional exchanges are

$$\xi_{(i,l)}^e := \frac{\lambda_i \nu_l}{q_i m(i, l)} \quad \text{for } l \neq i,$$

selling the value $e_i - \lambda_i \nu_i$ of asset i and delivering the value $\lambda_i \nu_l$ to asset l . Since $\Delta \mathbf{x}(\cdot)$ is linear, the final value held in asset l under $\xi'' := \xi' + \xi^e$ equals

$$q_l z_l - (e_l - \lambda_l \nu_l) + \sum_{i \neq l} \lambda_i \nu_l = \lambda \nu_l + \lambda_l \nu_l + \sum_{i \neq l} \lambda_i \nu_l = \left(\lambda + \sum_{i=1}^N \lambda_i \right) \nu_l.$$

The final portfolio is nonnegative with proportions ν and positive total value, so $\xi'' \in \Xi(\mathbf{x}, \nu)$, and the cost formula follows from Remark 2.13.

For b): by Theorem 2.14 optimal solutions of Problem 2.12 exist, and by Lemma A.3 they form a non-empty polyhedron; a linear functional bounded below on such a set attains its minimum, so volume-minimal optimal trades exist. Let ξ^* be any of them and suppose, contrary to the claim, that ξ^* is not direct: some asset i sells more than its initial holding, $s_i := \sum_{j \neq i} \xi_{(i,j)}^* > x_i$. We will reroute re-sold through-traffic around asset i so that either the total sold value strictly drops while nothing else changes, contradicting minimality, or some value is strictly gained, contradicting optimality via a). Since the final position $x_i + \sum_{j \neq i} P_{(j,i)} \xi_{(j,i)}^* - s_i$ is nonnegative, asset i also receives value. Hence there are assets $j, k \neq i$ with

$$\xi_{(j,i)}^* > 0 \quad \text{and} \quad \xi_{(i,k)}^* > 0,$$

and we reroute a small amount $\alpha > 0$ of this through-traffic $j \rightarrow i \rightarrow k$ around asset i : for $k \neq j$, define ξ^α from ξ^* by

$$\xi_{(j,i)}^\alpha := \xi_{(j,i)}^* - \alpha, \quad \xi_{(i,k)}^\alpha := \xi_{(i,k)}^* - \alpha P_{(j,i)}, \quad \xi_{(j,k)}^\alpha := \xi_{(j,k)}^* + \alpha \frac{P_{(j,i)} P_{(i,k)}}{P_{(j,k)}},$$

while for $k = j$ only the first two modifications apply (the round trip is reduced). For small α the trade ξ^α lies in Ξ , and its final positions coincide with those of ξ^* except possibly at j : at asset i the lost inflow $\alpha P_{(j,i)}$ matches the reduced outflow; at asset k (if $k \neq j$) the lost delivery $\alpha P_{(j,i)} P_{(i,k)}$ is restored exactly by the new direct leg; and asset j sells less, so its position gains $g\alpha \geq 0$ units, where

$$g := 1 - \frac{P_{(j,i)} P_{(i,k)}}{P_{(j,k)}} \geq 0, \quad \text{respectively} \quad g := 1 - P_{(j,i)} P_{(i,j)} \geq 0 \quad \text{if } k = j;$$

the first inequality holds since the best-execution price $P_{(j,k)}$ dominates the routed execution, the second by the no-arbitrage assumption A2. In both cases the total sold value strictly decreases, by at least $q_i \alpha P_{(j,i)}$.

If $g = 0$, all positions are unchanged, so $\xi^\alpha \in \Xi(\mathbf{x}, \nu)$ is optimal with strictly smaller total sold value, contradicting the choice of ξ^* . If $g > 0$, the final values of ξ^α dominate the ν -proportional floor $\lambda := v - C(\xi^*)$: they equal $\lambda \nu_l$ for $l \neq j$ and exceed $\lambda \nu_j$ by $e_j = q_j g \alpha > 0$. Hence a) yields a trade in $\Xi(\mathbf{x}, \nu)$ of cost $v - \lambda - \lambda_j < C(\xi^*)$, contradicting the optimality of ξ^* . Therefore ξ^* is direct. $\square$

Proof of Proposition 2.18. By Lemma A.8 b), some optimal solution ξ^* of Problem 2.12 is direct. For a direct trade ξ and the corresponding π^ξ , relation (2.16) gives

$$C(\xi) = v \cdot \sum_{i,j=1}^N c(i,j) \pi^\xi(i,j),$$

and by Lemma A.7 the maps $\xi \mapsto \pi^\xi$ and $\pi \mapsto \xi^\pi$ are mutually inverse, cost-preserving (up to the factor v) bijections between the direct trades in $\Xi(\mathbf{x}, \nu)$ and the plans in $\Gamma_m(\mu, \nu)$. In particular, the optimal values satisfy $C(\xi^*) = v \cdot \min_{\Gamma_m(\mu, \nu)} \sum_{i,j} c \pi$, a direct trade in $\Xi(\mathbf{x}, \nu)$ is optimal for Problem 2.12 if and only if its plan π^ξ is optimal for Problem 2.17, and every optimal plan is the image π^ξ of an optimal direct trade. $\square$

B Proof of Proposition 4.9

Proof of Proposition 4.9. For (i), existence and uniqueness follow from compactness of the feasible set and strict convexity of the objective, which up to an additive constant equals $\varepsilon D(\pi|K)$. Strict positivity is the standard entropy argument: at a zero entry the derivative along the segment towards the strictly positive feasible point $\mu \otimes \tilde{\nu}$ of Remark 3.2 is $-\infty$.

For (ii), the constraints are linear and π^ε is strictly positive, so the gradient of the objective lies in the range of the adjoint of the constraint map. The first marginal constraint contributes a term α_i as in the classical case. Pairing the second constraint $\sum_i m_{ij} \pi_{ij} = q_\pi \nu_j$ with a multiplier $\lambda \in \mathbb{R}^M$ gives

$$\begin{aligned} \sum_j \lambda_j \left(\sum_i m_{ij} \pi_{ij} - q_\pi \nu_j \right) &= \sum_{ij} \pi_{ij} m_{ij} \lambda_j - \left(\sum_l \lambda_l \nu_l \right) q_\pi \\ &= \sum_{ij} \pi_{ij} m_{ij} \psi_j, \end{aligned}$$

so that it contributes the term $m_{ij} \psi_j$, where $\psi_j := \lambda_j - \sum_l \lambda_l \nu_l$ satisfies the normalization $\sum_j \nu_j \psi_j = 0$. Stationarity therefore reads

$$c_{ij} + \varepsilon (\log \pi_{ij}^\varepsilon + 1) = \alpha_i + m_{ij} \psi_j,$$

and solving for π_{ij}^ε with $\varphi_i := \alpha_i - \varepsilon$ gives the asserted form. If two normalized pairs (φ, ψ) and (φ', ψ') represent the same minimizer, taking logarithms gives $\varphi_i + m_{ij} \psi_j = \varphi'_i + m_{ij} \psi'_j$, that is, $\psi_j - \psi'_j = -(\varphi_i - \varphi'_i)/m_{ij}$ for all i and j . Fixing i , all these differences have the same sign, so $\sum_j \nu_j (\psi_j - \psi'_j) = 0$ forces $\psi = \psi'$, and in turn $\varphi = \varphi'$.

For (iii), the first rescaling is classical. For the second, the minimizer σ of $D(\cdot|\pi)$ over C_2 exists since $D(\cdot|\pi)$ has compact sublevel sets and C_2 is closed, and as in (i) it is unique and strictly positive. By the computation of (ii) it has the form $\sigma_{ij} = \pi_{ij} b_j^{m_{ij}}$ with $\sum_j \nu_j \log b_j = 0$, and membership in C_2

reads $\sum_i m_{ij} \sigma_{ij} = q \nu_j$ with $q := \sum_{ij} m_{ij} \sigma_{ij}$. It therefore remains to show that the system

$$\sum_i m_{ij} \pi_{ij} b_j^{m_{ij}} = q \nu_j \quad \text{for } j = 1, \dots, M, \quad \text{and} \quad \sum_j \nu_j \log b_j = 0$$

has exactly one solution (b, q) with $b > 0$ and $q > 0$. For fixed $q > 0$ the map $b \mapsto \sum_i m_{ij} \pi_{ij} b^{m_{ij}}$ is a strictly increasing bijection of $(0, \infty)$ onto itself. Hence the first equation has a unique root $b_j(q)$, and $q \mapsto \log b_j(q)$ is a strictly increasing bijection of $(0, \infty)$ onto $\mathbb{R}$, so that $\sum_j \nu_j \log b_j(q)$ vanishes for exactly one q .

For (iv), the sets C_1 and C_2 are affine and contain the strictly positive matrix $\mu \otimes \tilde{\nu}$, so the iteration is the classical method of iterative Bregman projections, whose convergence goes back to [Bre67]; see [BCC⁺15] for its use in entropic optimal transport and [vLS23] for affine constraints of this generality.

For (v), let π^* be a minimizer of (KP). The lower bound holds since π^ε is feasible. Every feasible π has total mass one, so that $-\log(NM) \leq \sum_{ij} \pi_{ij} \log \pi_{ij} \leq 0$. Comparing the regularized objective at π^ε and at π^* gives

$$\sum_{ij} c_{ij} \pi_{ij}^\varepsilon - \sum_{ij} c_{ij} \pi_{ij}^* \leq \varepsilon \left(\sum_{ij} \pi_{ij}^* \log \pi_{ij}^* - \sum_{ij} \pi_{ij}^\varepsilon \log \pi_{ij}^\varepsilon \right) \leq \varepsilon \log(NM).$$

□

References

- [AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.
- [BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [BCC⁺15] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.
- [Bre67] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.
- [CLRS09] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms*. MIT Press, Cambridge, Mass., 3rd edition, 2009.
- [CPSV18a] Lenaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and Fisher–Rao metrics. *Foundations of Computational Mathematics*, 18:1–44, 2018.

- [CPSV18b] L  na  c Chizat, Gabriel Peyr  , Bernhard Schmitzer, and Fran  ois-Xavier Vialard. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.
- [Cut13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, pages 2292–2300, 2013.
- [DGJK20] Arnaud Dupuy, Alfred Galichon, Sonia Jaffe, and Scott Duke Kominers. Taxation in matching markets. *International Economic Review*, 61(4):1591–1634, 2020.
- [DN90] Mark HA Davis and Andrew R Norman. Portfolio selection with transaction costs. *Mathematics of operations research*, 15(4):676–713, 1990.
- [Edw10] D. A. Edwards. A simple proof in Monge-Kantorovich duality theory. *Studia Mathematica*, 200(1):67–77, 2010.
- [Gal16] Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 2016.
- [GLOP19] Wilfrid Gangbo, Wuchen Li, Stanley Osher, and Michael Puthawala. Unnormalized optimal transport. *Journal of Computational Physics*, 399:108940, 2019.
- [HH13] Helge Holden and Lars Holden. Optimal rebalancing of portfolios with transaction costs. *Stochastics An International Journal of Probability and Stochastic Processes*, 85(3):371–394, 2013.
- [Kab09] Yuri Kabanov. *Markets with Transaction Costs Mathematical Theory*. Springer, 2009.
- [Kan42] L Kantorovich. On translation of mass (in Russian), dokl. *AN SSSR*, 37:227–229, 1942.
- [KMV16] Stanislav Kondratyev, L  onard Monsaingeon, and Dmitry Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Advances in Differential Equations*, 21(11/12), 2016.
- [LMS18] Matthias Liero, Alexander Mielke, and Giuseppe Savar  . Optimal entropy-transport problems and the Hellinger–Kantorovich distance. *Inventiones Mathematicae*, 211(3):969–1117, 2018.
- [Mon81] Gaspard Monge. M  moire sur la th  orie des d  blais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.
- [MP95] Andrew J Morton and Stanley R Pliska. Optimal portfolio management with fixed transaction costs. *Mathematical Finance*, 5(4):337–356, 1995.
- [Nut22] Marcel Nutz. Introduction to entropic optimal transport. Lecture notes, Columbia University, 2022.
- [RX20] Johannes Ruf and Kangjianan Xie. The impact of proportional transaction costs on systematically generated portfolios. *SIAM Journal on Financial Mathematics*, 11(3):881–896, 2020.
- [San15] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkh  user, Cham, 2015.

- [SPV23] Thibault Séjourné, Gabriel Peyré, and François-Xavier Vialard. Chapter 12 - Unbalanced Optimal Transport, from theory to numerics. In Emmanuel Trélat and Enrique Zuazua, editors, *Numerical Control: Part B*, volume 24 of *Handbook of Numerical Analysis*, pages 407–471. Elsevier, 2023.
- [V⁺08] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.
- [vLS23] Johannes von Lindheim and Gabriele Steidl. Generalized iterative scaling for regularized optimal transport with affine constraints: Application examples. *arXiv preprint arXiv:2305.07071*, 2023.
- [Won19] Ting-Kam Leonard Wong. Information geometry in portfolio theory. *Geometric Structures of Information*, pages 105–136, 2019.