# Congestion and heterogenous quadratic regularization in optimal transport

Marcelo Gallardo *
marcelo.gallardo@pucp.edu.pe

Manuel Loaiza †
manuel.loaiza@pucp.edu.pe

Jorge Chávez ‡
jrchavez@pucp.edu.pe

February 19, 2025

**Abstract**

In this work, we build upon the optimal transport quadratic regularization model to develop a framework that incorporates congestion costs, particularly in matching within the healthcare and education sectors. Specifically, we introduce a model with heterogeneous quadratic costs. We analyze the model's properties under specific cases, extending the existing literature. Furthermore, we explore key structural characteristics of the model and provide numerical examples illustrating why this formulation more accurately captures real-world phenomena, particularly in the Peruvian context. The main result consists of identifying a specific type of corner solution when matching the same number of clusters, i.e., $N = L$.

**Keywords:** optimal transport, congestion costs, quadratic regularization, matching.
**JEL classifications:** C61, C62, C78, D04, R41.

*Department of Mathematics, Pontificia Universidad Católica del Perú (PUCP). Acknowledges insightful discussions with Professors Federico Echenique (UC Berkeley), Juan Carlos Carbajal (UNSW), and Fidel Jimenez (PUCP). This work is based on the preprint Congestion and Penalization in Optimal Transport; it is an excerpt from that work, focusing specifically on the quadratic model. In this paper, we do not present the penalization and weighting model.

†Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro.

‡Department of Mathematics, Pontificia Universidad Católica del Perú (PUCP). Acknowledges support from The Academic Directorate for Professors at the Pontificia Universidad Católica del Perú.

# 1 Introduction

Matching theory in economics studies how agents are paired based on their preferences and market constraints. The seminal work of Gale and Shapley (1962) introduced the concept of stable matching, where no two agents prefer each other over their assigned partners. This theory was later extended by Hylland and Zeckhauser (1979) in the context of house allocation and by Kelso and Crawford (1982), who incorporated transfers in two-sided matching. Alvin Roth significantly advanced these ideas, applying them to real-world scenarios such as school admissions and organ transplants, demonstrating the effectiveness of matching mechanisms (Roth (1982), Roth and Sotomayor (1990)). He also proved that no stable matching mechanism ensures truthfulness as a dominant strategy (Roth (1982)). In school choice, Abdulkadiroğlu and Sönmez (2003) developed mechanisms that balance stability and individual preferences. Further contributions include Hatfield and Milgrom (2005), Echenique and Yenmez (2015), and the recent book Echenique et al. (2023), which provides a comprehensive overview of these subjects.

In recent years, matching theory has been enriched by Optimal Transport (OT) methods, a mathematical framework introduced by Monge in the 18th century and rigorously formalized by Kantorovich in the 20th century. OT addresses optimal assignment problems by minimizing transportation costs and has been widely applied in matching markets, including student-school, patient-hospital, and worker-firm pairings. The foundational book by Cédric Villani (Villani (2009)) provides an exhaustive treatment of OT, while Ekeland (2010) and Ambrosio et al. (2021) further develop its mathematical aspects. In economics, Alfred Galichon bridged OT with discrete matching problems in markets (Galichon (2016), Galichon (2021)). The main advantage of OT is its ability to model agents as continuous distributions rather than discrete entities, optimizing over distributions. A recent study, Echenique et al. (2024), explores stability in matching through an OT framework applicable to both discrete and continuous settings. They show that utility transformations parameterized by $\alpha$ yield $\varepsilon$-stable, welfare-maximizing, or $\varepsilon$-egalitarian solutions, depending on whether the transformation is convex or concave. This approach aligns with Niederle and Yariv (2009) and Ferdowsian et al. (2023). In a discrete setting, Galichon (2021) employs computational methods (SISTA algorithm) akin to Merigot and Thibert (2020) and Nenna (2020) to recover matching costs in migration problems, with further applications in marriage and labor markets (Dupuy and Galichon (2014), Dupuy et al. (2019)).

In this work, we develop a variant of the OT model with quadratic heterogenous regularization. The model captures congestion costs and provides new insights relative to the existing literature. In particular, after addressing the problem from a theoretical perspective using basic elements of convex nonlinear programming, we present examples that illustrate our approach. The remainder of the paper is organized as follows. In Section 2, we define the notation and present existing models from the literature. Next, in Section 3, we introduce our model. We then derive some new results and conclude with examples and extensions. These examples account for the economic insights brought by our model.

## 2 Preliminaries

We consider two sets, $X = \{x_1, \dots, x_N\}$ and $Y = \{y_1, \dots, y_L\}$. Each element $x_i \in X$ represents a group/type of students or patients, and each $y_j \in Y$ represents a school or hospital[1]. To each $x_i$ ($y_j$), we associate a *mass* $\mu_i$ ($\nu_j$) in $\mathbb{Z}_{++}$, corresponding to the number of individuals in the group (school capacity or hospital bed availability). For simplicity, we refer to elements of $X$ by their index $i$ and elements of $Y$ by their index $j$. We denote by $\pi_{ij}$ the number of individuals of type $i$ assigned to entity $j$ (e.g., students to schools or patients to hospitals).

The problem, from a central planner's perspective, is to decide how many individuals from group $i$ should be assigned to $j = 1, \dots, L$, while minimizing the matching cost, given by a function $C : \mathbb{R}_+^{N,L} \times \mathbb{R}^P \to \mathbb{R}$, depending on the matching $\pi = [\pi_{ij}] \in \mathbb{R}_+^{N,L}$[2], and a vector of parameters $\theta \in \mathbb{R}^P$. The central planner's objective is to minimize this cost[3], subject to the constraints:

$$\Pi(\mu, \nu) = \left\{ \pi_{ij} \geq 0 : \sum_{j=1}^{L} \pi_{ij} = \mu_i, \ \forall\, i \in I \ \wedge \ \sum_{i=1}^{N} \pi_{ij} = \nu_j, \ \forall\, j \in J \right\}. \tag{1}$$

Constraints (1) ensure that all individuals (students, patients, etc.) are assigned and that all entities (schools, hospitals, etc.) fill their available capacity. Thus, the central planner solves:

$$\min_{\pi \in \Pi(\mu,\nu)} C(\pi; \theta). \tag{2}$$

The planner must choose an optimal matching[4] $\pi$ that minimizes the cost function. A solution to (2) is referred to as an optimal matching or optimal (transport) plan, denoted by $\pi^*$.

In the literature, the standard optimal transport model assumes separable linear costs, i.e., $C(\pi, \theta) = \sum_{i,j} c_{ij}\pi_{ij}$. This implies that the marginal cost of assigning an additional individual from $i$ to $j$ is constant, regardless of the existing assignments. Hence, the planner solves:

$$\mathcal{P}_O : \quad \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j} c_{ij}\pi_{ij}.$$

To solve $\mathcal{P}_O$, linear programming techniques such as the simplex method are typically used.

The classical OT model extends to continuous settings, optimizing over distributions rather than discrete entities. However, in this work, we remain in a discrete framework, inspired by the entropic an quadratic regularization approaches (Carlier et al. (2020), Peyré and Cuturi (2019),

[1] In this work we keep in mind these contexts. However, this model can be extended to other situations: labor market, migration etc.

[2] In this work, we mostly assume that the number of individuals matched can take real positive values, not just positive integers. This assumption is analogous to utility maximization problems where goods are treated as divisible.

[3] Matching individuals involves costs beyond «physical» transportation, including implicit factors such as tuition fees, admission criteria, medical specialties, insurance coverage, and personal preferences. Thus, we refer to these as matching costs rather than transportation costs.

[4] In some cases, $\pi$ is treated as a vector instead of a matrix: $\pi = (\pi_{11}, \pi_{12}, \cdots, \pi_{NL})^T$.

Nutz (2024)). With respect to entropic regularization, the problem takes the form:

$$\min_{\pi\in\Pi(\mu,\nu)} \sum_{i=1}^{N}\sum_{j=1}^{L} c_{ij}\pi_{ij} + \sigma\pi_{ij}\ln(\pi_{ij}),$$

with $\sigma > 0$. On the other hand, quadratic regularization Wiesel and Xu (2024); Nutz (2024) changes the last term by $(\varepsilon/2)||\pi||_2^2$. This is the most similar formulation already existing to our model.

Returning to the discrete setting with linear costs, if $\pi$ is taken in $\mathbb{Z}_+^{N,L}$, the central planner's problem always has a solution since there is a finite number of possible assignments.

**Proposition 2.1.** In an integer setting, the number of matchings is at most $L^M$.

*Proof.* The number of ways to assign all $\mu_i$ individuals from group $i$ to entities is given by solutions to:

$$\pi_{i1} + \cdots + \pi_{iL} = \mu_i, \quad 0 \le \pi_{ij} \le \nu_j \quad \forall j. \tag{3}$$

Ignoring the upper bounds $\nu_j$, this reduces to a stars and bars problem. The upper bound for the number of solutions to (3) is $\binom{\mu_i+L-1}{L-1}$. Applying the multiplication principle, the total number of matchings satisfies:

$$\prod_{i=1}^{N}\binom{\mu_i+L-1}{L-1} = \prod_{i=1}^{N}\prod_{j=1}^{\mu_i}\frac{j+L-1}{j} \le \prod_{i=1}^{N}\prod_{j=1}^{\mu_i} L = L^M. \qquad \blacksquare$$

Proposition 2.1 guarantees the existence of a solution in the integer setting. However, $\mathcal{P}_O$ does not enforce $\pi_{ij} \in \mathbb{Z}_+$, potentially allowing non-integer solutions. Despite this, Weierstrass' theorem ensures that $\mathcal{P}_O$ always has a solution in $\mathbb{R}_+^{N,L}$.

**Proposition 2.2.** Given $\mu = (\mu_1, ..., \mu_N)^T \in \mathbb{R}_{++}^N$ and $\nu = (\nu_1, ..., \nu_L)^T \in \mathbb{R}_{++}^L$, $\mathcal{P}_O$ always has a solution $\pi^*$.

*Proof.* The objective function is continuous as it is linear. The constraint set $\Pi(\mu, \nu)$ is compact in $\mathbb{R}^{NL}$ since it is the intersection of closed sets and bounded within $[0, M]^{NL}$. $\blacksquare$

The basic model has been extensively studied, with variations such as entropic (Dupuy and Galichon, 2014; Carlier et al., 2020) or quadratic (Lorenz et al., 2019; González-Sanz and Nutz, 2024; Wiesel and Xu, 2024; Nutz, 2024) regularization, or in the continuous framework (Dupuy and Galichon, 2014; Echenique et al., 2024). To the best of our knowledge, the model and results we deliver are new in the literature.

## 3 Congestion costs

We now introduce a new variant of the optimal transport problem in the discrete setting that explicitly accounts for congestion effects. Traffic congestion and institutional overload are crucial factors affecting the allocation of individuals to entities such as schools and hospitals.

When too many individuals are matched to the same entity, congestion costs escalate, leading to inefficiencies in both physical and bureaucratic dimensions. This phenomenon is observed in various settings:

- **Traffic congestion:** The simultaneous assignment of many students to the same school in urban areas can increase travel times, overload public transport, and generate bottlenecks in key traffic zones. The same happens with patients and hospitals.
- **Medical centers overload:** Large patient inflows can overwhelm hospital resources, creating long waiting times, administrative bottlenecks, and inefficient service delivery.
- **Bureaucratic congestion:** Excess demand for certain institutions may slow down processing times, affecting school admissions, hospital triage, and public service allocation due to outdated systems and inefficient workflows.

To model this phenomenon, we consider a strictly convex cost function with respect to the number of matched individuals, capturing the increasing marginal costs associated with congestion.

### 3.1 Mathematical Formulation

We define the cost function $C(\pi;\theta)$[5] as a separable function:

$$C(\pi;\theta) = \sum_{i=1}^{N}\sum_{j=1}^{L} \varphi(\pi_{ij};\theta_{ij}), \tag{4}$$

where $\varphi_{ij}$ is structurally homogeneous[6]. The central planner's problem then becomes:

$$\min_{\pi\in\Pi(\mu,\nu)} \left\{ \sum_{i=1}^{N}\sum_{j=1}^{L} \varphi(\pi_{ij};\theta_{ij}) \right\}, \tag{5}$$

where $\Pi(\mu,\nu)$ is defined as in (1). Given that congestion leads to increasing costs, $\varphi$ should be strictly increasing and strictly convex, transforming the problem into a convex optimization problem with linear constraints.

Although the Linear Independence Constraint Qualification (LICQ) condition may fail for solutions where non-negativity constraints are not binding, the convexity of the objective function and the linearity of constraints allow us to apply the Karush-Kuhn-Tucker (KKT) conditions (see Boyd (2004)).

[5] As mentioned earlier, $\pi$ can be taken as a vector in $\mathbb{R}_+^{NL}$ rather than a matrix in $\mathbb{R}_+^{N,L}$.

[6] The function $\varphi_{ij}$ does not change structurally across $(i,j)$ pairs; whether logarithmic, exponential, or polynomial, we assume $\varphi_{ij} = \varphi$.

### 3.2 Lagrangian Formulation and KKT Conditions

The Lagrangian function associated with (5) is given by:

$$\mathscr{L}(\pi,\lambda,\xi,\gamma;\theta) = \sum_{i=1}^{N}\sum_{j=1}^{L}\varphi(\pi_{ij};\theta_{ij}) + \sum_{i=1}^{N}\xi_i\left(\mu_i - \sum_{j=1}^{L}\pi_{ij}\right) + \sum_{j=1}^{L}\lambda_j\left(\nu_j - \sum_{i=1}^{N}\pi_{ij}\right) - \sum_{i=1}^{N}\sum_{j=1}^{L}\gamma_{ij}\pi_{ij}. \tag{6}$$

The KKT first-order conditions are:

$$\begin{aligned}
\frac{\partial\mathscr{L}(\pi^*,\xi^*,\lambda^*,\gamma^*;\theta)}{\partial\pi_{ij}} &= \frac{\partial\varphi(\pi_{ij}^*;\theta_{ij})}{\partial\pi_{ij}} - \lambda_j^* - \xi_i^* - \gamma_{ij}^* = 0,\ \forall\ (i,j)\in I\times J,\\
-\pi_{ij}^* &\le 0,\ \forall\ (i,j)\in I\times J,\\
\sum_{j=1}^{N}\pi_{ij}^* - \mu_i &= 0,\ \forall\ i\in I,\\
\sum_{i=1}^{N}\pi_{ij}^* - \nu_j &= 0,\ \forall\ j\in J,\\
\gamma_{ij}^*\pi_{ij}^* &= 0.
\end{aligned}$$

Since the objective function is strictly convex and the constraint set is convex, the solution is unique.

### 3.3 Quadratic Cost Function and Optimal Matching

To carry out a quantitative analysis, we assume a quadratic cost function:

$$\varphi(\pi_{ij};\theta_{ij}) = d_{ij} + c_{ij}\pi_{ij} + a_{ij}\pi_{ij}^2. \tag{7}$$

Thus, the optimization problem simplifies to:

$$\mathcal{P}_1 : \min_{\pi\in\Pi(\mu,\nu)}\left\{\sum_{i=1}^{N}\sum_{j=1}^{L} d_{ij} + c_{ij}\pi_{ij} + a_{ij}\pi_{ij}^2\right\}. \tag{8}$$

The parameters have clear economic interpretations:

- $d_{ij}$ represents fixed costs associated with each matching (e.g., baseline administrative or transportation costs).
- $c_{ij}$ corresponds to constant marginal costs, capturing characteristics such as individual preferences.
- $a_{ij}$ introduces congestion effects, ensuring increasing marginal costs as $\pi_{ij}$ grows.

Applying the KKT conditions, we obtain:

$$\pi_{ij}^* = \frac{\xi_i^* + \lambda_j^* + \gamma_{ij}^* - c_{ij}}{2a_{ij}}. \tag{9}$$

Strict, convexity guarantees the uniqueness of $\pi^*$.

### 3.4 Structural Properties of the Solution

A fundamental issue in this model is determining whether solutions are interior ($\pi_{ij}^* > 0$ for all $(i,j)$) or a corner solutions ($\gamma_{ij}^* > 0$ for some $(i,j)$). The following result characterizes a structural property of the solution:

**Proposition 3.1.** With respect to the problem (5), with costs given by (7), whenever $\gamma_{ij}^* = 0$ for all $(i,j) \in I \times J$, the linear system obtained from (9), with respect to $(\xi^*, \lambda^*)$, leads to a singular $N + L$ linear system.

*Proof.* Since $\gamma_{ij}^* = 0$ for all $(i,j) \in I \times J$, first order conditions lead to

$$\sum_{j=1}^{L} \pi_{ij}^* = \sum_{j=1}^{L} \frac{\xi_i^*}{2a_{ij}} + \sum_{j=1}^{L} \frac{\lambda_j^*}{2a_{ij}} - \sum_{j=1}^{L} \frac{c_{ij}}{2a_{ij}} = \mu_i, \ \forall\, i \in I \tag{10}$$

$$\sum_{i=1}^{N} \pi_{ij}^* = \sum_{i=1}^{N} \frac{\xi_i^*}{2a_{ij}} + \sum_{i=1}^{N} \frac{\lambda_j^*}{2a_{ij}} - \sum_{i=1}^{N} \frac{c_{ij}}{2a_{ij}} = \nu_j, \ \forall\, j \in J. \tag{11}$$

By setting $x = \begin{bmatrix} \xi_1^* & \cdots & \xi_N^* & \lambda_1^* & \cdots & \lambda_L^* \end{bmatrix}^T \in \mathbb{R}^{N+L}$, the linear equalities (10) and (11) on $\xi_i^*$ and $\lambda_j^*$ are described by the linear system $(\Lambda + T)x = b$, where

$$\Lambda = \text{Diag}\left( \sum_{j=1}^{L} \frac{1}{2a_{1j}}, \ldots, \sum_{j=1}^{L} \frac{1}{2a_{Nj}}, \sum_{i=1}^{N} \frac{1}{2a_{i1}}, \ldots, \sum_{i=1}^{N} \frac{1}{2a_{iL}} \right) \in \mathbb{R}^{N+L,N+L}.$$

$$\Upsilon = \left[ \frac{1}{2a_{ij}} \right]_{\substack{1 \le i \le N \\ 1 \le j \le L}} \in \mathbb{R}^{N,L} \text{ and } T = \begin{bmatrix} 0 & \Upsilon \\ \Upsilon^T & 0 \end{bmatrix} \in \mathbb{R}^{N+L,N+L},$$

$$b = \left[ \mu_1 + \sum_{j=1}^{L} \frac{c_{1j}}{2a_{1j}}, \ldots, \mu_N + \sum_{j=1}^{L} \frac{c_{Nj}}{2a_{Nj}}, \nu_1 + \sum_{i=1}^{N} \frac{c_{i1}}{2a_{i1}}, \ldots, \nu_L + \sum_{i=1}^{N} \frac{c_{iL}}{2a_{iL}} \right]^T \in \mathbb{R}^{N+L}.$$

Let $R = \Lambda + T$. If $R_k$ denotes the $k-$th row of $R$, we note that $R_1 = \sum_{k=N+1}^{N+L} R_k - \sum_{k=2}^{N} R_k$. Hence, $\text{Det}(R) = 0$, and the claim follows. ■

As usual in economics, we are interested in perform monotone or smooth comparative statics. With respect to the former (see Milgrom and Shannon (1994)), it can't be performed since $S = \Pi(\mu, \nu)$ is not a sub-lattice of $X = \mathbb{R}_+^{NL}$. Indeed, given $\pi_1, \pi_2 \in S$, in general, $\pi_1 \wedge \pi_2$ and $\pi_1 \vee \pi_2$ do not belong to $S$. With respect to the latter, Proposition 3.2 explains why smooth comparative statics cannot be accomplished.

**Proposition 3.2.** With respect to (6), considering quadratic costs[7]

$$\text{Det}(J_{\pi,(\xi,\lambda)}\overline{\mathscr{L}}(\pi^*, \xi^*, \lambda^*, \overline{\theta})) = 0.$$

*Proof.* First, let $\pi = (\pi_{11}, \ldots, \pi_{1L}, \cdots, \pi_{N1}, \ldots, \pi_{NL})^T$. Then, we define

$$D = \text{Diag}(a_{11}, \ldots, a_{1L}, \cdots, a_{N1}, \ldots, a_{NL}) \in \mathbb{R}_{++}^{NL,NL}$$

and $B = [b_{k\ell}] \in \mathbb{R}^{N+L,N+L}$, where

$$b_{k\ell} = \begin{cases} 1 & \text{if } k \leq N \text{ and } (k-1)L < \ell \leq kL, \\ 1 & \text{if } N < k \leq N+L \text{ and } \ell \equiv k - N \pmod{L}, \\ 0 & \text{otherwise.} \end{cases}$$

Matrix $B$ never has full rank. Indeed, $B_1 = \sum_{k=N+1}^{N+L} B_k - \sum_{k=2}^{N} B_k$, where $B_k$ is row $k$ of $B$. Thus, since

$$J_{\pi,(\xi,\lambda)}\overline{\mathscr{L}}(\pi^*, \xi^*, \lambda^*, \overline{\theta}) = \begin{bmatrix} D & -B^T \\ -B & 0 \end{bmatrix},$$

following Gentle (2017), $\text{Det}(J_{\pi,(\xi,\lambda)}\overline{\mathscr{L}}(\pi^*, \xi^*, \lambda^*, \overline{\theta})) = \text{Det}(D)\text{Det}(0 - BD^{-1}B^T) = 0$. ■

Although we cannot apply smooth comparative statics, the conditions of the Envelope Theorem are satisfied for $\pi^*$ in the interior of $\Pi$. Therefore, by defining $V = V(\pi^*) = \sum_{i=1}^{N}\sum_{j=1}^{L} \varphi_{ij}(\pi_{ij}^*; \overline{\theta}_{ij})$, we can conclude that $\partial V/\partial c_{ij} = \pi_{ij}^* > 0$ and $\partial V/\partial a_{ij} = \pi_{ij}^{*2} > 0$, which is expected, as the cost of the optimal transport plan only increases if the coefficients associated with preference costs and congestion costs rise.

Note that, in general, obtaining the optimal matching $\pi^*$ from (9), is quite complicated. Even if we assume an interior solution, which would simplify the equations since $\gamma_{ij}^* = 0$ automatically, we still cannot solve the linear system systematically. Note also that $R$ not being invertible does not imply that the system has no solution. It only means that, if a solution $(\xi^*, \lambda^*)$ exists, it is either not unique, or there is $\gamma_{ij}^* \neq 0$. What is unique is $\pi^*$ since the objective function is strictly convex. Hence, even if we have several $(\xi^*, \lambda^*)$, at the end, we obtain a unique $\pi^*$. The non uniqueness of $(\xi^*, \lambda^*)$ originates from the fact that the LICQ does not hold for interior solutions.

Nonetheless, when $N = L$, we do can obtain a explicit solution for our model under mild assumptions.

## 4 Analysis for $N = L$

An important issue in our model is to determine whether the solutions will be a corner solutions ($\gamma_{ij}^* > 0$ for some $(i, j) \in I \times J$) or not. The following examples show that under the quadratic setting, both interior and corder solutions can exist.

[7] Following de la Fuente (2000) notation. Here $\overline{\mathscr{L}} = (\nabla_\pi \mathscr{L}, \nabla_\theta \mathscr{L})$.

**Example 4.1.** In this example, we show a case where the solution is interior. Consider

$$a = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, \; c = \begin{bmatrix} 24 & 48 \\ 16 & 24 \end{bmatrix}, \; \mu = (20, 20), \; \nu = (12, 28), \; \text{so that } N = L = 2.$$

Then, we have $\pi^* = (7, 13, 5, 15)$.

**Example 4.2.** To illustrate a case where the solution is a corner solution, consider the following values:

$$a = \begin{bmatrix} 200 & 2 \\ 2 & 200 \end{bmatrix}, \quad c = \begin{bmatrix} 200 & 2 \\ 2 & 200 \end{bmatrix}, \quad \mu = (10, 10), \quad \nu = (10, 10), \quad \text{so that } N = L = 2.$$

In this scenario, the optimal solution is $\pi^* = (0, 10, 10, 0)$, a corner solution.

Note that in Example 4.1, the solution no longer satisfies the property of the classical model where there always exists $(i, j) \in I \times J$ such that $\pi^*_{ij} = 0$ (see Tardella (2010)).

Now, consider adding restrictions to the parameter vector and the sizes of the sets to explicitly obtain a specific corner solutions.

**Assumption 1.** Let $M$ be a positive integer strictly greater than 1. Assume that $N = L = M$ and $\mu_i = \nu_j$ for all $1 \leq i, j \leq M$.

Assumption 1 ensures that each school reaches full capacity with individuals from the same group.

**Assumption 2.** For each $1 \leq i \leq N$, suppose there exists $1 \leq \zeta_i \leq L$ such that $c_{i\,\zeta_i} < c_{i\,j}$ for all $1 \leq j \leq L$ with $j \neq \zeta_i$. Furthermore, assume that $\zeta_i \neq \zeta_j$ for all $1 \leq i, j \leq L$ with $i \neq j$.

Assumption 2 imposes that each individual is optimally matched with their top choice school, ensuring a distinct best fit for each individual. Note that assumptions 1 and 2 imply immediately that the solution to the linear model is:

$$\pi^* = [\pi^*_{ij}] = \begin{cases} \mu_i & \text{if } j = \zeta_i, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Indeed, for any other matching $\pi \in \Pi(\mu, \nu)$,

$$C(\pi, \theta) = \sum_{i=1}^{N} \sum_{j=1}^{L} d_{ij} + c_{ij}\pi_{ij} > \sum_{i=1}^{N} \sum_{j=1}^{L} d_{ij} + \sum_{i=1}^{N} c_{i\,\zeta_i} \sum_{j=1}^{L} \pi_{ij} = C(\pi^*, \theta).$$

**Assumption 3.** Let $\widetilde{c}_i = \min\limits_{\substack{1 \leq j \leq L \\ j \neq \zeta_i}} \{c_{ij}\}$ satisfy $\widetilde{c}_i > c_{i\,\zeta_i} + a_{i\,\zeta_i}\mu_i^2(1 - 1/L)$ for $1 \leq i \leq N$.

Assumption 3 tells us that preferences between student types and schools must be such that the top choice only based on preferences and individual characteristics is at least $a_{i\,\zeta_i}\mu_i^2(1 - 1/L)$ better than the other ones. We now show by combining assumptions 1, 2 and 3 that the solution to $\mathcal{P}_1$, in the integer setting, is given by (12).

**Theorem 4.3.** *Under Assumptions 1, 2 and 3, the optimal matching for the quadratic model in the integer setting is (12).*

*Proof.* Let $\pi$ be an arbitrary matching different from $\pi^*$. Then,

$$C(\pi;\theta) = \sum_{i=1}^{N}\sum_{j=1}^{L} d_{ij} + c_{ij}\pi_{ij} + a_{ij}\pi_{ij}^2 \geq \sum_{i=1}^{N}\sum_{j=1}^{L} d_{ij} + \sum_{i=1}^{N}\left(\sum_{j=1}^{L} c_{ij}\pi_{ij} + a_{i\,\zeta_i}\sum_{j=1}^{L}\pi_{ij}^2\right).$$

Now, consider $i$ such that $\pi_{i\,\zeta_i} < \mu_i$. Due to the integer nature of $\pi$, $\pi_{i\,\zeta_i} \leq \mu_i - 1$. Hence

$$\begin{aligned}
\sum_{j=1}^{L} c_{ij}\pi_{ij} &= c_{i\,\zeta_i}\pi_{i\,\zeta_i} + \sum_{j\neq\zeta_i} c_{ij}\pi_{ij} \\
&\geq c_{i\,\zeta_i}\pi_{i\,\zeta_i} + \widetilde{c}_i(\mu_i - \pi_{i\,\zeta_i}) \\
&= \widetilde{c}_i\mu_i - \pi_{i\,\zeta_i}(\widetilde{c}_i - c_{i\,\zeta_i}) \\
&\geq \widetilde{c}_i\mu_i - (\mu_i - 1)(\widetilde{c}_i - c_{i\,\zeta_i}) \\
&= \mu_i c_{i\,\zeta_i} + \widetilde{c}_i - c_{i\,\zeta_i}.
\end{aligned}$$

On the other hand, consider the function $f : \mathbb{R}^{L-1} \to \mathbb{R}$ defined by

$$f(x_1,\ldots,x_{L-1}) = x_1^2 + \cdots + x_{L-1}^2 + (\mu_i - x_1 - \cdots - x_{L-1})^2.$$

Note that the set $x_j^* = \mu_i/L$ minimizes $f$. As a consequence,

$$\sum_{j=1}^{L}\pi_{ij}^2 = f(\pi_{i1},\ldots,\pi_{i\;L-1}) \geq \sum_{j=1}^{L}\left(\frac{\mu_i}{L}\right)^2 = \frac{\mu_i^2}{L}.$$

Combining these results, we have

$$C(\pi;\theta) \geq \sum_{i=1}^{N}\sum_{j=1}^{L} d_{ij} + \sum_{i=1}^{N}\mu_i c_{i\,\zeta_i} + \widetilde{c}_i - c_{i\,\zeta_i} + a_{i\,\zeta_i}\left(\frac{\mu_i^2}{L}\right) > C(\pi^*;\theta). \qquad \blacksquare$$

**Example 4.4.** The following examples were computed using Mathematica 14.1. For each case, we present the parameter matrices $d$, $c$, and $a$ (where applicable), along with the optimal matching matrix $\pi^*$, obtained using the appropriate optimization method.

For the linear model, with $N = L = 4$ and $\mu_i = \nu_j = 50$, the optimal matching was computed using `LinearOptimization`:

$$d = \begin{bmatrix} 32 & 83 & 82 & 37 \\ 47 & 75 & 56 & 45 \\ 87 & 74 & 79 & 4 \\ 40 & 55 & 94 & 14 \end{bmatrix}, \quad c = \begin{bmatrix} 76 & 77 & 83 & 6 \\ 74 & 98 & 7 & 41 \\ 6 & 86 & 8 & 70 \\ 88 & 17 & 40 & 96 \end{bmatrix}, \quad \pi^* = \begin{bmatrix} 0 & 0 & 0 & 50 \\ 0 & 0 & 50 & 0 \\ 50 & 0 & 0 & 0 \\ 0 & 50 & 0 & 0 \end{bmatrix}.$$

For the quadratic model, with $N = L = 4$ and $\mu_i = \nu_j = 20$,

$$d = \begin{bmatrix} 88 & 88 & 100 & 91 \\ 19 & 42 & 37 & 69 \\ 81 & 87 & 9 & 50 \\ 66 & 18 & 77 & 91 \end{bmatrix}, \quad c = \begin{bmatrix} 989 & 24 & 975 & 941 \\ 673 & 612 & 684 & 9 \\ 20 & 352 & 387 & 380 \\ 675 & 687 & 44 & 697 \end{bmatrix}, \quad a = \begin{bmatrix} 9 & 3 & 8 & 9 \\ 6 & 8 & 3 & 2 \\ 1 & 7 & 8 & 3 \\ 9 & 5 & 2 & 6 \end{bmatrix},$$

the optimal matching, obtained using `QuadraticOptimization`, is

$$\pi^* = \begin{bmatrix} 0 & 20 & 0 & 0 \\ 0 & 0 & 0 & 20 \\ 20 & 0 & 0 & 0 \\ 0 & 0 & 20 & 0 \end{bmatrix},$$

Hence, the result is in accordance with Theorem 4.3

Having explored the specific cases where the solution is either a corner or interior solution, we now turn to the general case for $N = L = 2$, disregarding any assumption. The following calculations were obtained using Mathematica 14.1. By solving (10) and (11), we identified four parametric solution families that require $\mu_1 + \mu_2 = \nu_1 + \nu_2$. Three of these families are discarded because they correspond to degenerate cases: the first case holds when $a_{12} + a_{22} = 0$, the second case holds when $a_{11} + a_{12} + a_{21} + a_{22} = 0$ and $\mu_2 = (2a_{12}(\nu_1 + \nu_2) + 2\nu_1(a_{21} + a_{22}) - c_{11} + c_{12} + c_{21} - c_{22})/(2a_{12} + 2a_{22})$ and the third case holds when $a_{12} + a_{22} = 0$, $a_{11} + a_{21} = 0$ and $\nu_1 = (2\nu_2 a_{22} + c_{11} - c_{12} - c_{21} + c_{22})/(2a_{21})$. These unfeasible conditions leave us with one valid solution family, given by $\xi_2^* = \xi_1^* + (2(a_{11}a_{12} + a_{12}a_{21} + a_{11}a_{22} + a_{21}a_{22})\mu_2 - 2(a_{11}a_{12} + a_{11}a_{22})\nu_1 - 2(a_{11}a_{12} + a_{12}a_{21})\nu_2 + (a_{12} + a_{22})(c_{21} - c_{11}) + (a_{11} + a_{21})(c_{22} - c_{12}))/(a_{11} + a_{12} + a_{21} + a_{22})$, $\lambda_1^* = (-\xi_1^* a_{21} - \xi_2^*(a_{12} + a_{21} + a_{22}) + 2(a_{12}a_{21} + a_{21}a_{22})\mu_2 - 2a_{12}a_{21}\nu_2 + a_{22}c_{21} + a_{21}c_{22} - a_{21}c_{12} - a_{12}c_{21})/(a_{12} + a_{22})$ and $\lambda_2^* = (-\xi_1^* a_{22} - \xi_2^* a_{12} - 2a_{12}a_{22}\nu_2 - a_{22}c_{12} - a_{12}c_{22})/(a_{12} + a_{22})$ where $\xi_1^*$ is free. By plugging these equalities into (9), we obtain the optimal matching when all the resulting expressions are strictly greater than zero. A detailed analysis to guarantee that $\pi_{ij}^* > 0$ was performed by reducing inequalities programmatically, but the numerous inequalities generated are omitted here. This analysis establishes a well-defined parameter space where the solution remains interior.

Given the specific cases analyzed above, it becomes evident that there is little hope of determining analytically whether solutions are interior or corner as $N$ and $L$ increase beyond 2. While the examples for $N = L = 2$ allowed us to identify some conditions under which solutions are either interior or corner, as the dimension of the problem grows, these conditions become increasingly complex and indeterminate. It is nonetheless always possible to obtain a numerical solution to $\mathcal{P}_1$[8].

The case $N = L$ becomes particularly relevant when considering the healthcare sector, where certain hospital networks are designated for specific types of diseases. This also applies to health

[8]In particular, we used QuadraticOptimization in Mathematica 14.1 (also known as quadratic programming (QP), mixed-integer quadratic programming (MIQP) or linearly constrained quadratic optimization).

insurance systems, such as SIS, Essalud, and EPS. Let us see this in detail in Section 5.

# 5 Applications

The formulation in Problem $\mathcal{P}_1$ is particularly relevant in contexts where congestion costs significantly affect the allocation of resources. Unlike models with linear costs, the quadratic cost structure accounts for congestion effects indirectly by making overburdened facilities increasingly costly. This feature is crucial in understanding inefficiencies in the Peruvian healthcare and education sectors, where access is heavily determined by proximity and efficiency in medical care.

## 5.1 Healthcare: The Impact of Bureaucratic and Geographic Congestion

In the Peruvian healthcare system, individuals are theoretically assigned to facilities based on their insurance type—whether EsSalud (public insurance for formal workers), SIS (universal public insurance for low-income individuals), or private insurance (EPS), Anaya-Montes and Gravelle (2024). However, the reality is far more complex due to congestion in high-demand facilities and bureaucratic inefficiencies that prevent patients from accessing appropriate care.

For instance, hospitals in urban centers frequently experience extreme congestion, with reports indicating that some facilities operate at over 60% beyond their intended capacity (EsSalud, 2024b). The quadratic cost structure in $\mathcal{P}_1$ captures this congestion effect: as the number of patients $\pi_{ij}$ assigned to a hospital increases, the marginal cost grows non-linearly, making further allocations increasingly inefficient. This helps explain why, despite having a designated primary hospital, many patients end up in less appropriate facilities due to excessive waiting times or administrative delays. The average referral delay of 37.8 days in EsSalud further exacerbates this inefficiency (EsSalud, 2024a).

Geographic constraints further reinforce these inefficiencies. Traffic congestion in Lima and other major cities significantly increases travel costs, deterring patients from seeking care at facilities that may have better capacity but require longer commutes. Empirical studies show that even when better-equipped hospitals exist within a reasonable distance, a significant portion of patients opt for closer, overcrowded facilities (Anaya-Montes and Gravelle, 2024). This behavior is well captured in our model, where $a_{ij}$ reflects congestion effects that create a de facto access constraint, even in the absence of explicit restrictions.

Moreover, misallocation is not limited to geographic factors. Bureaucratic hurdles often prevent patients from receiving specialized care at appropriate institutions. A recent analysis found that 38.5% of patients requiring specialized treatment were misallocated to general health centers that lacked the necessary expertise, leading to inefficiencies in treatment and additional referral delays (EsSalud, 2025). The quadratic cost formulation accounts for these inefficiencies by internalizing how an increased number of patients in non-specialized centers inflates the overall cost of healthcare provision. We illustrate this in the following examples.

**Example 5.1.** In this example, we aim to represent the healthcare sector scenario, where three groups of patients are theoretically assigned to a specific type of medical center: SIS, EsSalud,

or EPS. However, due to quadratic costs, patient distribution is not strictly confined to these categories, leading to cross-assignments between systems. Given the following parameters:

$$a = \begin{bmatrix} 2.0 & 1.0 & 2.0 \\ 1.0 & 2.0 & 2.0 \\ 2.0 & 1.0 & 2.0 \end{bmatrix}, \; c = \begin{bmatrix} 1.0 & 10.0 & 10.0 \\ 10.0 & 1.0 & 10.0 \\ 10.0 & 10.0 & 1.0 \end{bmatrix}, \; d = I_{3\times 3}$$

and

$$\mu = \begin{bmatrix} 20.0 \\ 20.0 \\ 20.0 \end{bmatrix} = \nu.$$

The optimal solution $\pi^*$ under these conditions is:

$$\pi^* = \begin{bmatrix} 6.8074 & 7.1959 & 5.9966 \\ 8.6351 & 5.6081 & 5.7567 \\ 4.5574 & 7.1959 & 8.2466 \end{bmatrix}$$

This solution highlights the deviations from a strict one-to-one patient allocation, as the quadratic cost terms allow for cross-assignments that would not occur in a purely linear model. For comparison, when $a = 0$, meaning there are no quadratic costs, the optimal assignment is:

$$\pi^* = \begin{bmatrix} 20.0 & 0 & 0 \\ 0.0 & 20.0 & 0.0 \\ 0.0 & 0 & 20.0 \end{bmatrix}.$$

Here, patients are strictly assigned to their designated medical system, as expected in the absence of congestion effects.

**Example 5.2.** In this example, we analyze a scenario where the linear costs $c_{ij}$ are identical across all assignments, meaning there are no inherent preferences between different allocation routes. However, the quadratic congestion costs $a_{ij}$ vary, influencing the final distribution of assignments. Despite the presence of optimal routes under linear costs, congestion effects lead to deviations in the allocation. The parameters are as follows:

$$a = \begin{bmatrix} 1.0 & 2.0 & 2.0 \\ 2.0 & 1.0 & 2.0 \\ 2.0 & 2.0 & 1.0 \end{bmatrix}, \quad c = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{bmatrix}, \quad d = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 20.0 \\ 20.0 \\ 20.0 \end{bmatrix}, \quad \nu = \begin{bmatrix} 20.0 \\ 20.0 \\ 20.0 \end{bmatrix}$$

The optimal solution $\pi^*$ under these conditions is:

$$\pi^* = \begin{bmatrix} 10.0 & 5.0 & 5.0 \\ 5.0 & 10.0 & 5.0 \\ 5.0 & 5.0 & 10.0 \end{bmatrix}$$

This result highlights the impact of congestion costs. Even though all routes have the same linear cost, the quadratic term introduces distortions in the allocation, preventing a strict adherence to any single preferred matching pattern. Instead, the system distributes assignments to mitigate excessive congestion, leading to deviations from what would be optimal under purely linear costs.

**Example 5.3.** We compare the standard quadratic regularization model with our proposed heterogeneous congestion cost model. Both cases share the same linear costs $c_{ij}$ and distance factors $d_{ij}$, as well as the same supply and demand constraints:

$$c = \begin{bmatrix} 1.0 & 5.0 & 5.0 \\ 5.0 & 1.0 & 5.0 \\ 5.0 & 5.0 & 1.0 \end{bmatrix}, \quad d = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 20.0 \\ 20.0 \\ 20.0 \end{bmatrix}, \quad \nu = \begin{bmatrix} 20.0 \\ 20.0 \\ 20.0 \end{bmatrix}$$

In the standard quadratic regularization model, $a_{ij}$ is uniform:

$$a = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{bmatrix}$$

yielding the optimal allocation:

$$\pi^* = \begin{bmatrix} 8.0 & 6.0 & 6.0 \\ 6.0 & 8.0 & 6.0 \\ 6.0 & 6.0 & 8.0 \end{bmatrix}$$

In contrast, our model introduces heterogeneity in congestion costs:

$$a = \begin{bmatrix} 2.0 & 1.0 & 1.0 \\ 1.0 & 2.0 & 1.0 \\ 1.0 & 1.0 & 2.0 \end{bmatrix}$$

leading to a different optimal allocation:

$$\pi^* = \begin{bmatrix} 4.8 & 7.6 & 7.6 \\ 7.6 & 4.8 & 7.6 \\ 7.6 & 7.6 & 4.8 \end{bmatrix}.$$

Unlike the uniform model, this formulation better captures congestion differences, reducing allocations where costs are higher and redistributing demand accordingly. This results in a more realistic representation of congestion-driven inefficiencies.

## 5.2 Education: Congestion Costs and School Choice Constraints

The Peruvian education system is highly fragmented, with only a few centralized subsystems, such as COAR schools. Although our model aligns more closely with centralized systems, it effectively captures essential features of the Peruvian educational landscape.

Congestion costs distort optimal school assignments based on $c_{ij}$, emphasizing the impact of geographic barriers on student placement. While our model does not fully replicate the Peruvian education framework, it offers a more realistic approximation under centralized decision-making and provides useful insights for policymakers in both education and healthcare.

The next example illustrates this fact by showing how introducing heterogenous quadratic costs $a_{ij}\pi_{ij}^2$ distorts the matching under the linear structure. Moreover, we also compare our model to the classic quadratic regularization.

**Example 5.4.** This example illustrates how introducing heterogeneous quadratic costs $a_{ij}\pi_{ij}^2$ distorts student allocation compared to a purely linear preference-based model. In many developed countries, such as France or Switzerland, well-developed metro systems allow students to access top schools regardless of distance. However, in Peru, inadequate public transportation significantly affects school choice, leading to inefficient assignments. We consider three groups of students and three types of schools, where $c_{ij}$ represents student preferences, including perceived school quality and distance constraints. Without congestion costs, students would be perfectly sorted into their most preferred schools. The parameters are as follows:

$$a = \begin{bmatrix} 4.0 & 2.0 & 3.0 \\ 4.0 & 2.0 & 6.0 \\ 3.0 & 4.0 & 3.0 \end{bmatrix}, \quad c = \begin{bmatrix} 1.0 & 5.0 & 100.0 \\ 10.0 & 1.0 & 50.0 \\ 100.0 & 50.0 & 1.0 \end{bmatrix}, \quad d = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 40.0 \\ 40.0 \\ 40.0 \end{bmatrix}, \quad \nu = \begin{bmatrix} 40.0 \\ 40.0 \\ 40.0 \end{bmatrix}$$

When congestion costs are included, the optimal assignment is:

$$\pi^* = \begin{bmatrix} 16.40 & 17.07 & 6.53 \\ 15.07 & 17.65 & 7.29 \\ 8.53 & 5.28 & 26.19 \end{bmatrix}.$$

Here, students are not necessarily assigned to their most preferred schools due to congestion effects. Those who would ideally attend top schools are redirected to lower-ranked institutions, as excessive demand increases quadratic congestion costs. For comparison, when congestion costs

are removed ($a = 0$), the optimal assignment is:

$$\pi^* = \begin{bmatrix} 40.0 & 0.0 & 0.0 \\ 0.0 & 40.0 & 0.0 \\ 0.0 & 0.0 & 40.0 \end{bmatrix}.$$

This result aligns perfectly with the preference-based structure of $c_{ij}$, as all students are assigned to their most desired schools without deviation. This example highlights how transportation inefficiencies and congestion distort the school choice process. Unlike countries with high-quality metro systems, where students can attend their ideal schools regardless of distance, in Peru, traffic congestion and poor infrastructure create a situation where even high-achieving students may not access top-tier institutions. Our model captures these effects by incorporating heterogeneous quadratic costs, providing a more realistic representation of school allocation dynamics in constrained environments.

### 5.3 Insights from the Quadratic Cost Model

Unlike models with explicit constraints or penalization terms, the quadratic formulation in $\mathcal{P}_1$ captures congestion effects endogenously. The parameter $a_{ij}$ adjusts dynamically based on observed congestion levels, allowing the model to replicate key real-world patterns. In the Peruvian context, where access to essential services is heavily influenced by both bureaucratic and geographic congestion, this framework provides a more realistic representation of allocation inefficiencies.

Future empirical work could refine these insights by calibrating $a_{ij}$ to real-world congestion data, similar to the methodological framework proposed by (Agarwal and Somaini, 2023). By extending this approach, policymakers could better understand the trade-offs between accessibility and efficiency, ultimately informing targeted interventions in both the healthcare and education sectors.

## 6 Conclusions

In this paper, we developed an optimal transport model with heterogeneous quadratic regularization to account for congestion effects in matching problems. Unlike classical models that assume linear transportation costs or entropy regularization, our formulation introduces increasing marginal costs, enabling the modeling of real-world inefficiencies caused by overcrowding.

From a theoretical perspective, we demonstrated that the optimization problem retains a convex structure and that the uniqueness of the optimal assignment is guaranteed. However, we showed that analytically characterizing the solutions is nontrivial, as the system of equations derived from the KKT conditions is singular. For the particular case where the number of agent types and entities matches ($N = L$), we provided conditions under which the model produces corner solutions, meaning that each agent type is assigned to a single entity. This result is particularly relevant in sectors where specialization and supply segmentation are crucial, such as

education and healthcare.

In terms of applications, our model is relevant for various allocation problems in both the public and private sectors. In education, it captures the congestion effects that arise when too many students are assigned to specific institutions, leading to infrastructure constraints and quality deterioration. In healthcare, our formulation applies to the distribution of patients across hospitals under segmented healthcare systems, such as the Peruvian case with SIS, EsSalud, and EPS, where excessive demand in certain hospitals leads to long waiting times and service inefficiencies. Additionally, the model can be extended to labor markets where firms face increasing costs when hiring additional workers with similar profiles, a phenomenon observed in industries with capacity constraints.

Our analysis of specific cases suggests that the problem admits both interior and boundary solutions, depending on the cost structure and assignment preferences. While we derived explicit conditions for certain cases, a general analytical solution remains challenging, indicating the need for numerical methods to solve the problem in more complex scenarios.

Future extensions of this work could include incorporating uncertainty in assignment costs and exploring intertemporal congestion dynamics. Moreover, advanced computational techniques, such as mixed-integer quadratic programming and nonlinear constrained optimization methods, could be used to analyze high-dimensional and more intricate cases.

## References

Abdulkadiroğlu, A. and Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *The American Economic Review*, 93(3):729–747.

Agarwal, N. and Somaini, P. (2023). Empirical Models of Non-Transferable Utility Matching. In Echenique, F., Immorlica, N., and Vazirani, V. V., editors, *Online and Matching-Based Market Design*, pages 530–551. Cambridge University Press.

Ambrosio, L., Brué, E., and Semola, D. (2021). *Lectures on Optimal Transport*, volume 130 of *Unitext*. Springer.

Anaya-Montes, M. and Gravelle, H. (2024). Health insurance system fragmentation and covid-19 mortality: Evidence from peru. *PLOS ONE*, 19(8):e0309531.

Boyd, S. (2004). *Convex Optimization*. Cambridge University Press.

Carlier, G., Dupuy, A., Galichon, A., and Sun, Y. (2020). SISTA: Learning Optimal Transport Costs under Sparsity Constraints. *arXiv preprint arXiv:2009.08564*. Submitted on 18 Sep 2020, last revised 21 Oct 2020.

de la Fuente, A. (2000). *Mathematical Methods and Models for Economists*. Cambridge University Press. Digital publication date: 04 June 2012.

Dupuy, A. and Galichon, A. (2014). Personality Traits and the Marriage Market. *Journal of Political Economy*, 122(6):1271–1319.

Dupuy, A., Galichon, A., and Sun, Y. (2019). Estimating Matching Affinity Matrices under Low-Rank Constraints. *Information and Inference: A Journal of the IMA*, 8(4):677–689.

Echenique, F., Immorlica, N., and Vazirani, V. V. (2023). *Online and Matching-Based Market Design*. Cambridge University Press.

Echenique, F., Root, J., and Sandomirskiy, F. (2024). Stable Matching as Transportation. Preprint submitted to arXiv on 12 Feb 2024.

Echenique, F. and Yenmez, M. B. (2015). How to Control Controlled School Choice. *The American Economic Review*, 105(8):2679–2694.

Ekeland, I. (2010). Notes on Optimal Transportation. *Economic Theory*, 42(2):437–459.

EsSalud, C. (2024a). External consultation deferral dashboard power bi. Accessed in December 2024, official source of EsSalud.

EsSalud, D. (2025). Appointment deferral dashboard power bi. Accessed in January 2025, official source of EsSalud.

EsSalud, E. (2024b). Hospital stay dashboard power bi. Accessed in August 2024, official source of EsSalud.

Ferdowsian, A., Niederle, M., and Yariv, L. (2023). Strategic Decentralized Matching: The Effects of Information Frictions. This version: November 13, 2023.

Gale, D. and Shapley, L. S. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1):9–15.

Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton University Press.

Galichon, A. (2021). The Unreasonable Effectiveness of Optimal Transport in Economics. Preprint submitted on 12 Jan 2023.

Gentle, J. E. (2017). *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer, Cham, Switzerland, 2nd edition.

González-Sanz, A. and Nutz, M. (2024). Sparsity of quadratically regularized optimal transport: Scalar case. *arXiv preprint arXiv:2410.03353*.

Hatfield, J. W. and Milgrom, P. R. (2005). Matching with Contracts. *The American Economic Review*, 95(4):913–935.

Hylland, A. and Zeckhauser, R. (1979). The Efficient Allocation of Individuals to Positions. *The Journal of Political Economy*, 87(2):293–314.

Kelso, A. S. and Crawford, V. P. (1982). Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica*, 50(6):1483.

Lorenz, D. A., Manns, P., and Meyer, C. (2019). Quadratically regularized optimal transport. *Applied Mathematics & Optimization*.

Merigot, Q. and Thibert, B. (2020). Optimal Transport: Discretization and Algorithms. Preprint submitted on 2 Mar 2020.

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1):157–180.

Nenna, L. (2020). Lecture 4 entropic optimal transport and numerics.

Niederle, M. and Yariv, L. (2009). Decentralized Matching with Aligned Preferences. NBER Working Paper 14840, National Bureau of Economic Research.

Nutz, M. (2024). Quadratically regularized optimal transport: Existence and multiplicity of potentials. Preprint submitted to arXiv on 10 Feb 2024.

Peyré, G. and Cuturi, M. (2019). Computational Optimal Transport: With Applications to Data Science. Preprint submitted on 4 June 2019.

Roth, A. E. (1982). The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7(4):617–628.

Roth, A. E. and Sotomayor, M. A. O. (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, volume 18 of *Econometric Society Monographs*. Cambridge University Press.

Tardella, F. (2010). The fundamental theorem of linear programming: extensions and applications. *Optimization*, 59(3):283–301.

Villani, C. (2009). *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer.

Wiesel, J. and Xu, X. (2024). Sparsity of quadratically regularized optimal transport: Bounds on concentration and bias. *arXiv preprint arXiv:2410.03425*.