# Congestion and Penalization in Optimal Transport

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

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

Jorge Chávez *
jrchavez@pucp.edu.pe[†]

October 31, 2024

**Abstract**

In this paper we introduce two novel models derived from the discrete optimal transport problem. The first model extends the traditional transport problem by adding a quadratic congestion factor directly into the cost function, while the second model replaces conventional constraints with weighted penalization terms. We present theoretical results for the characterization of interior and corner solution of specific cases, and perform smooth comparative statics analysis. We also propose an $O((N+L)(NL)^2)$ algorithm for computing the optimal plan for the penalized model. Additionally, in the appendices, we have added examples that illustrate important results of the paper.

**Keywords:** optimal transport, quadratic regularization, matching, penalization, Neumann series.
**JEL classifications:** C61, C62, C78, D04, R41.

[*]Department of Mathematics, Pontificia Universidad Católica del Perú (PUCP).
[†]Acknowledges support from la Dirección Académica del Profesorado de la PUCP.

# 1 Introduction

Matching theory in economics examines the processes through which agents are paired based on their preferences and the constraints within a market. The seminal work due to Gale and Shapley (1962) introduced the concept of stable matching, defined as a set of pairs where no two individuals, each from different pairs, would both prefer each other over their current partners. Matching theory was further expanded by the work of Hylland and Zeckhauser (1979) in the context of the house allocation problem, and the work of Kelso and Crawford (1982), which introduced transfers in two-sided matching. Alvin Roth significantly extended these works by applying matching theory to practical scenarios such as school admissions and organ donations, demonstrating its effectiveness in solving real-world allocation problems, see for instance Roth (1982) or Roth and Sotomayor (1990). Roth also highlighted that no stable matching mechanism exists where stating true preferences is a dominant strategy (Roth (1982)). Abdulkadiroğlu and Sönmez (2003) has been pivotal in developing school choice mechanisms that consider individual preferences while aiming for stability. Further contributions in the area include, for instance Hatfield and Milgrom (2005) or Echenique and Yenmez (2015). A comprehensive and detailed overview of market design can be found in the recent book Echenique et al. (2023). This book presents many of the fundamental aspects and state-of-the-art developments in matching theory: stability, Pareto efficiency, social welfare, mechanisms, and more.

In recent years, matching theory has gained a substantial enrichment through the integration of Optimal Transport (OT) methods, a mathematical framework originally formulated by Gaspard Monge in the 18th century. OT essentially addresses the problem of optimally pairing two distributions of agents or resources, minimizing the cost of transportation. This powerful tool has found widespread applications in economics, particularly in matching markets where it is used to analyze pairings such as students with schools, patients with hospitals, or workers with firms.

The modern development of OT theory owes much to the groundbreaking work of Nobel laureate Leonid Kantorovich, who transformed it into a rigorous mathematical discipline in the last century. His contributions laid the groundwork for a plethora of subsequent research. Later, Fields medalist Cédric Villani expanded on these foundations in his comprehensive text, Optimal Transport: Old and New (Villani (2009)), which has become a definitive reference in the field, covering both classical and modern developments.

Further contributions have been made by notable scholars such as Ekeland (2010) and Ambrosio et al. (2021), who explored various mathematical aspects and extended the theory's applications. In 2016, Alfred Galichon published Optimal Transport Methods in Economics (Galichon (2016)), bridging the gap between these sophisticated mathematical techniques and economic theory. His work focuses particularly on discrete market applications within matching theory, offering a practical approach to understanding economic pairings. For a more concise exploration of these ideas, Galichon provides a summarized version in his later work Galichon (2021). Collectively, these advancements have cemented OT as an essential tool for economists, enabling nuanced analysis and insights in matching theory and beyond.

The main advantage of optimal transport is that it allows for addressing cases where the types of agents are represented by a continuum rather than discretely. In such cases, the optimization is done over distributions. In this sense, the use of OT in economic theory spans both discrete and continuous settings. A recent work that explores stability in matching through an OT framework, valid for both discrete and infinite cases, is that of Echenique et al. (2024). They demonstrate how various transformations of utilities, characterized by a parameter $\alpha$, can lead to solutions that are $\varepsilon$-stable, achieve welfare maximization, or are $\varepsilon$-egalitarian, depending on whether the transformation is convex or concave. This is mainly done in a context of aligned preferences, as in Niederle and Yariv (2009) or Ferdowsian et al. (2023). The work of Echenique et al. (2024) is thus a clear example of the intersection of OT theory with matching. On the other hand, focused on a discrete setting, Galichon (2021) employs computational methods (SISTA algorithm) similar to those discussed by Merigot and Thibert (2020) and Nenna (2020), aiming to recover matching costs in problems such as migration. This approach is further explored in Dupuy and Galichon (2014) and Dupuy et al. (2019), but in another context (marriage market, labor market etc.). The estimation of preferences with the goal of estimating the parameters of the transportation cost can be done by, for instance, using Agarwal and Somaini (2023).

In this work, inspired by the discrete optimal transport problem, we develop two models that maintain a structure similar to the classic problem but diverge from it by incorporating a quadratic term in the cost structure to reflect transportation costs due to congestion. These models lead to finite-dimensional optimization problems of the Karush-Kuhn-Tucker (KKT) type with one featuring mixed constraints, and the other including penalization terms, similar to Izmailov and Solodov (2023). Our first model builds upon the quadratically regularized optimal transport problem, which has been recently studied (see Lorenz et al. (2019), González-Sanz (2024) Wiesel and Xu (2024) or Nutz (2024)). While the mathematical structure of the problem is similar, our approach diverges significantly, leading to results that differ from those found in the existing literature.

We begin by briefly presenting the classical linear model and establishing the necessary notation for the remainder of the paper. We then introduce the first model, which incorporates a quadratic congestion term into the cost function while preserving the original constraints. We characterize the solutions and demonstrate that the standard smooth comparative statics techniques are not applicable in this context. Subsequently, we consider the second model, which eliminates the equality constraints by incorporating penalty terms directly into the objective function. In this model, we analyze the properties of the optimal solution and conclude with a discussion on smooth comparative statics. Our analysis centers on computing the inverse of a matrix that arises from the first-order conditions. We present results on the bounds of the matrix coefficients and provide explicit calculations for specific cases. For interior solutions, we utilize Neumann series and an algorithmically computable closed-form formula to efficiently compute the solution.

# 2 Preliminaries

We will consider two sets, $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_L\}$. For our purposes, each element $x_i \in X$ represents a group/type of students, and each $y_j \in Y$ represents a school[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 (capacity of the school). When referring to an element of $X$, instead of denoting it by $x_i$, we usually, to simplify the notation, refer to it by $i$. Analogously, the elements of $Y$ are referred to by the index $j$, instead of $y_j$. Moreover, we denote the set of indices $\{1, \ldots, N\}$ by $I$ and the set of indices $\{1, \ldots, L\}$ by $J$. We denote by $\pi_{ij}$ the number of students of type $i$ going to school $j$.

The problem addressed, from the perspective of a central planner, is to decide how many individuals from group $i$ should go to school $j = 1, \ldots, L$, and so forth for each $i$, minimizing the matching cost, which is given by means of 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 problem is to minimize this cost[3], subject to the constraints

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

The restrictions given by (1) tells us that the central planner must ensure that all students receive education and that all schools fill their quotas. Hence, the central problem solves

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

The central planner has to chose an optimal matching [4] $\pi$ such that the cost function

[1]Although we think in students and schools, the model applies also to workers and firms, patients and hospitals, etc.

[2]In this work, we will mostly assume that the number of students matched can take values in the real positive line and not only in the positive integers. Note that this is the same issue that arises when one solves the utility maximization problem in the classical framework assuming divisible goods.

[3]Matching individuals incurs a cost that is not limited solely to «physical» transportation costs, which certainly accounts for both ways (round trip), but also encompasses implicit costs linked to specific characteristics of $i$ and $j$ such as tuition fee, admission exam, languages, sex, age, etc. This is why we refer to them as matching costs instead of transportation costs.

[4]Sometimes, $\pi$ is going to be treated as a vector instead of a matrix. In such case, we consider $\pi = (\pi_{11}, \pi_{12}, \cdots, \pi_{NL})^T$.

is minimized. A solution to (2) will be from now referred to as an optimal matching or optimal (transport) plan, and will be denoted by $\pi^*$.

In the literature, separable linear costs are assumed in the standard optimal transport model. This is, $C(\pi, \theta) = \sum_{i,j} c_{ij}\pi_{ij}$. It is therefore assumed that the marginal cost of matching one more individual from $i$ with $j$ is always the same, regardless of how many people are already matched and independent of any other variable. In such a model, it is considered that the number of students is equal to the number of available slots, that all student are assigned to a school, and that the number of individuals assigned to $j$ is exactly $\nu_j$. Mathematically, $\sum_j \pi_{ij} = \mu_i$, $\sum_i \pi_{ij} = \nu_j$ and $\sum_{i=1}^N \mu_i = \sum_{j=1}^L \nu_j = M$. Therefore, the central planner seeks to solve

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

To solve $\mathcal{P}_O$, one typically employs linear programming techniques, such as the simplex method, which are designed to find the optimal matching that minimizes the total cost subject to the constraints given by $\Pi(\mu, \nu)$.

As discussed in the classical literature, the most general form of the OT problem allows for the existence of infinite types, and in such cases, the optimization is done over distributions. In this paper, however, we are not going to study continuous distributions. What we do focus on, in line with the entropic regularization problem (see, for example, Carlier et al. (2020) and Peyré and Cuturi (2019)), is working with a variation of the optimization problem in the discrete setting. In the case of entropic regularization, the problem addressed is

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

with $\sigma > 0$. Given the strict convexity of $f(x) = x \ln x$ and that the analogous of Inada's conditions are satisfied ($\lim_{x \downarrow 0} f'(x) = -\infty$), the solution is interior, i.e. $\pi^*_{ij} > 0$ (see a detailed argument in Nenna (2020)).

Returning to our discrete setting assuming linear costs, for the specific case where $\pi$ is taken in $\mathbb{Z}_+^{N,L}$, we can assert that the central planner's problem always has a solution since there is a finite number of ways to assign students to schools.

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

*Proof.* The number of ways to assign all $\mu_i$ students from group $i$ to schools is equivalent to counting the number of solutions to the equation

$$\pi_{i1} + \cdots + \pi_{iL} = \mu_i \quad \text{with} \quad 0 \le \pi_{ij} \le \nu_j \quad \text{for} \quad 1 \le j \le L. \tag{3}$$

If we ignore the upper bounds $\nu_j$ on $\pi_{ij}$, this becomes a classic stars and bars problem.

Thus, an upper bound for the number of solutions to (3) is $\binom{\mu_i+L-1}{L-1}$ for $1 \leq i \leq N$. Next, ignoring the constraint $\pi_{1j} + \cdots + \pi_{Nj} = \nu_j$ and applying the multiplication principle, an upper bound for the total number of matchings is

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

Proposition 2.1 ensures the existence of a solution to the discrete allocation problem considered in the integer setting. Nonetheless, the problem $\mathcal{P}_O$ does not take into account the fact that $\pi_{ij} \in \mathbb{Z}_+$. The problem $\mathcal{P}_O$ may actually have a non-integer solution. Thus, there are infinitely many feasible $\pi$. Nevertheless, what we establish in the following proposition is that the problem $\mathcal{P}_O$ always has a solution, even when $\pi$ is taking in $\mathbb{R}_+^{N,L}$. This follows immediately from the Weierstrass Theorem.

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

*Proof.* The objective function is continuous since it is linear. On the other hand, $\Pi(\mu, \nu)$ is compact in $\mathbb{R}^{NL}$. Indeed, it is closed since it is the intersection of $N + L + NL$ closed sets, and bounded since $\Pi(\mu, \nu) \subset [0, M]^{NL}$. $\blacksquare$

The basic model has been studied thoroughly in the literature, and variations such as the entropic regularization or the infinite case, have been analyzed as well (see for instance Dupuy and Galichon (2014), Carlier et al. (2020), Dupuy and Galichon (2022) or Echenique et al. (2024)). Moreover, since the basic linear model fits into the large class of linear programming models, it is known (see Munkres (1957) or Tardella (2010)) that a solution always lies in a vertex, i.e., $\pi_{ij}^* = 0$ for some $(i, j) \in I \times J$. More recently, the quadratic regularization problem (Nutz (2024)) has also been addressed, but with results of a different nature from those presented below (see González-Sanz (2024) or Wiesel and Xu (2024)), and with a different quadratic term[5]. For this reason, we quickly move on to the models we aim to study. To the best of our knowledge, both our formulation of the model and the analysis we provide are novel contributions to the literature.

# 3 Congestion costs

We now derive a new variant of the optimal transport problem in the discrete setting. Traffic is a crucial phenomenon with considerable impact on the allocation of resources or people. Since it becomes increasingly costly to match individuals living in the same location with the same school, to model this situation, we consider a strictly convex

[5]They consider $||\pi||_{L_2}$, while we focus on congestion costs of the form $\sum_{i,j} a_{ij}\pi_{ij}^2$.

function with respect to the number of matched individual. In some cities arround the world this congestion phenomenon is quite relevant: the mere fact that some schools start classes significantly increases congestion.

In general, as commented before, the cost function $C(\pi;\theta)$ is defined in $\mathbb{R}_+^{NL} \times \mathbb{R}^{P}$[6]. Besides, it is reasonable to assume that this function is separable and structural homogeneous[7]. That is,

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

Under this new framework, the problem that concerns the central planner becomes

$$\min_{\pi\in\Pi(\mu,\nu)} \left\{ \sum_{i=1}^{N}\sum_{j=1}^{L} \varphi(\pi_{ij};\theta_{ij}) \right\}, \tag{4}$$

where $\Pi(\mu,\nu)$ is the same set from (1). As argued before, $\varphi$ should be a strictly increasing and strictly convex function. Hence, this problem is framed as a convex optimization problem with linear equality constraints and non negativity conditions. Thus, *a priori*, it could be approached through KKT theory. However, the Linear Independence Constraint Qualification (LICQ) is generally not satisfied for solutions where non negativity constraints are not binding.

Although we cannot guarantee the LICQ, we can still apply the KKT conditions, since the objective function is convex and the constraints are linear (see Boyd (2004)).

The Lagrangian associated to (4) is the function $\mathscr{L} : \mathbb{R}^{NL} \times \mathbb{R}^{L} \times \mathbb{R}^{N} \times \mathbb{R}^{NL} \to \mathbb{R}$ given by

$$\begin{aligned}\mathscr{L}(\pi,\lambda,\xi,\gamma;\theta) = &\sum_{i=1}^{N}\sum_{j=1}^{L} \varphi(\pi_{ij};\theta_{ij}) + \sum_{i=1}^{N} \xi_i \left(\mu_i - \sum_{j=1}^{L}\pi_{ij}\right) + \sum_{j=1}^{L}\lambda_j\left(\nu_j - \sum_{i=1}^{N}\pi_{ij}\right) \\ &- \sum_{i=1}^{N}\sum_{j=1}^{L}\gamma_{ij}\pi_{ij}.\end{aligned} \tag{5}$$

The KKT first order conditions applied to (5) are

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

[6] As mentioned before, sometimes $\pi$ is taken as a vector in $\mathbb{R}_+^{NL}$ instead of a matrix in $\mathbb{R}_+^{N,L}$.

[7] The structure of $\varphi_{ij}$ does not change with the pair $(i,j)$; whether logarithmic, exponential, or polynomial, $\varphi_{ij} = \varphi$.

$$\sum_{j=1}^{N} \pi_{ij}^* - \mu_i = 0, \ \forall \ i \in I$$
$$\sum_{i=1}^{N} \pi_{ij}^* - \nu_j = 0, \ \forall \ j \in J$$

where $\xi_1, \cdots, \xi_N, \lambda_1, \cdots, \lambda_L \in \mathbb{R}$. Moreover, since the objective function is strictly convex, uniqueness of the solution is ensured. In order to carry out a quantitative analysis, we assign a specific form to the function $\varphi$:

$$\varphi(\pi_{ij}; \theta_{ij}) = d_{ij} + c_{ij}\pi_{ij} + a_{ij}\pi_{ij}^2. \tag{6}$$

Thus, the optimization problem becomes

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

In this model, $\theta_{ij} = (c_{ij}, a_{ij})$. For the sake of simplicity, since the constant term $d_{ij}$ does not take any role in the optimization problem, it is not considered in the vector $\theta_{ij}$. Nonetheless, it has a specific and non-negligible meaning; it represents fixed costs. Regarding the parameters $c_{ij}$, they represent, as in the linear model, constant marginal costs associated with the characteristics of student types and schools. These costs can also be understood as a transformation of the utility generated by the matching pair $(i, j)$, as in Echenique et al. (2024). Finally, the novelty with respect to the classical linear model lies in the incorporation of the quadratic term, which represents the cost associated with congestion. For $\varphi$ given by (6), KKT conditions yield

$$\pi_{ij}^* = \frac{\xi_i^* + \lambda_j^* + \gamma_{ij}^* - c_{ij}}{2a_{ij}}. \tag{7}$$

Convexity arguments ensure the optimality of (7).

In general, solving the first-order KKT conditions is a case-by-case process. Proposition 3.1 shows that in the scenario where the non-negativity multipliers are zero, we obtain a singular linear system. Thus, if the solution lies within the interior of the constraint set $\Pi(\mu, \nu)$, solving the singular system becomes necessary.

**Proposition 3.1.** *With respect to problem (4), with costs given by (6), whenever $\gamma_{ij}^* = 0$ for all $(i, j) \in I \times J$, from (7), 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

$$\sum_{j=1}^{L} \pi_{ij}^* = \sum_{j=1}^{L} \frac{\xi_i^*}{2a_{ij}} + \sum_{j=1}^{L} \frac{\lambda_j^*}{2a_{ij}} - \sum_{j=1}^{L} \frac{c_{ij}}{2a_{ij}} = \mu_i, \ \forall\, i \in I \tag{8}$$

$$\sum_{i=1}^{N} \pi_{ij}^* = \sum_{i=1}^{N} \frac{\xi_i^*}{2a_{ij}} + \sum_{i=1}^{N} \frac{\lambda_j^*}{2a_{ij}} - \sum_{i=1}^{N} \frac{c_{ij}}{2a_{ij}} = \nu_j, \ \forall\, j \in J. \tag{9}$$

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 (8) and (9) can be written in the compact form $(\Lambda + T)x = b$, where

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

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

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

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, $\det(R) = 0$, 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 cannot 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 assuming interior solution.

**Proposition 3.2.** *With respect to (5), considering quadratic costs and assuming an interior solution*

$$det(J_{\pi,(\xi,\lambda)} \overline{\mathscr{L}}(\pi^*, \xi^*, \lambda^*, \overline{\theta})) = 0^{8},$$

*where* $\overline{\mathscr{L}} = (\nabla_\pi \mathscr{L}, \nabla_\theta \mathscr{L})$.

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

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

[8] In here we follow de la Fuente (2000) notation.

and $B = [b_{k\ell}] \in \mathbb{R}^{N+L,NL}$, 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), $\det(J_{\pi,(\xi,\lambda)}\overline{\mathscr{L}}(\pi^*, \xi^*, \lambda^*, \overline{\theta})) = \det(D)\det(0 - BD^{-1}B^T) = 0$. ■

Although we cannot apply smooth comparative statics techniques from de la Fuente (2000), 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 $\partial V/\partial a_{ij} = \pi^{*2}_{ij} > 0$, which is expected, as the cost of the optimal transport plan only increases if the coefficients associated with preference costs and congestion costs rise.

Notice that, in general, obtaining the optimal matching $\pi^*$ from (7) is quite complicated. Even if we assume an interior solution, 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 $(\xi^*, \lambda^*)$ is due to 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.

## 3.1 Analysis for $N = L$

An important issue in our model is to determine whether or not the solutions will be a corner solution. The following examples show that under the quadratic setting, both interior and corner solutions can exist.

**Example 3.3.** In this example we show a case where the solution is interior. Consider

$$a = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \; c = \begin{bmatrix} 12 & 24 \\ 8 & 12 \end{bmatrix}, \; \mu = [10, 10]^T, \; \nu = [6, 14]^T, \text{ so that } N = L = 2.$$

Then, we have $\pi^* = [4, 6, 2, 8]^T$.

**Example 3.4.** To illustrate a case where the solution is a corner solution, consider the following values:

$$a = \begin{bmatrix} 100 & 1 \\ 1 & 100 \end{bmatrix}, \quad c = \begin{bmatrix} 100 & 1 \\ 1 & 100 \end{bmatrix}, \quad \mu = [5,5]^T, \quad \nu = [5,5]^T, \quad \text{so that } N = L = 2.$$

In this scenario, the optimal solution is $\pi^* = [0,5,5,0]^T$, a corner solution.

Note from Example 3.3, that a 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 solution.

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

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 $1 \leq t_i \leq L$ such that $c_{i\,t_i} < c_{i\,j}$ for all $1 \leq j \leq L$ with $j \neq t_i$. Furthermore, assume that $t_i \neq t_j$ 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:

$$\pi^* = [\pi^*_{ij}] = \begin{cases} \mu_i & \text{if } j = t_i, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

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

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

**Assumption 3.** Let $\tilde{c}_i = \min\limits_{\substack{1 \leq j \leq L \\ j \neq t_i}} \{c_{ij}\}$ satisfy $\tilde{c}_i > c_{i\,t_i} + a_{i\,t_i}\mu_i^2(1 - 1/L)$ 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 $a_{i\,t_i}\mu_i^2(1 - 1/L)$ 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 (10).

**Theorem 3.5.** *Under Assumptions 1, 2 and 3, the optimal matching for the quadratic model in the integer setting is (10).*

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

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

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

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

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

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

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 by using Mathematica 14.1. By solving (8) and (9), 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 (7), 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 $\mathcal{P}_1$[9].

# 4 Congestion with penalization

In this section, we present our second model where equality constraints are dismissed. Instead, we now introduce penalization terms[10] in the cost function as follows;

$$\mathcal{P}_2 : \min_{\pi_{ij} \geq 0} \underbrace{\left\{ \underbrace{\alpha \sum_{i=1}^{N} \sum_{j=1}^{L} \varphi(\pi_{ij}; \theta_{ij})}_{\text{matching direct cost}} + \underbrace{(1-\alpha) \left( \sum_{i=1}^{N} \epsilon_i \left( \sum_{j=1}^{L} \pi_{ij} - \mu_i \right)^2 + \sum_{j=1}^{L} \delta_j \left( \sum_{i=1}^{N} \pi_{ij} - \nu_j \right)^2 \right)}_{\text{costs of social objectives}} \right\}}_{=F(\pi;\theta,\alpha,\epsilon,\delta,\mu,\nu)}. \tag{11}$$

In (11), $\alpha$, $\epsilon_1, ..., \epsilon_N$, $\delta_1, ..., \delta_L$ and $\mu_1, \cdots, \mu_N$, $\nu_1, \cdots, \nu_L$ are all non negative, and $\varphi$ and $\theta_{ij}$ are the same as those in Section 3.

The problem $\mathcal{P}_2$ is set once again in the context of a central planner whose objectives are to minimize the matching direct cost, while also attempting to fill the available seats in schools, and at the same time, ensuring that everyone receives an education (cost of social objectives). The coefficients $\epsilon_i$ and $\delta_j$ weigh how much the central planner cares about educating individuals/filling schools or avoiding overcrowding, while the parameter $\alpha \in [0, 1]$ weighs which objective is the most important. Furthermore, in this new formulation, the population is assumed to be arbitrarily large, so the optimization is carried out over the entire space $\mathbb{R}_+^{NL}$. Thus, the new model proposes a trade-off between minimizing the cost of matching and deviations from the social objectives imposed by the central planner. In particular, if there were no penalty, given that $\pi_{ij} \geq 0$, the minimum would be obtained at $0 \in \mathbb{R}^{NL}$. Conversely, if there were no cost for matching, the solution

[9]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).

[10]The idea of incorporating the restrictions into the objective function, and how the constrained and unconstrained problems relate, is studied in Izmailov and Solodov (2023).

would satisfy $\sum_j \pi^*_{ij} = \mu_i$ for all $i \in I$ and $\sum_i \pi^*_{ij} = \nu_j$ for all $j \in J$. For the sake if simplicity, we take $\alpha = 1/2$. KKT first order conditions applied to (11) yield

$$\frac{\partial F}{\partial \pi_{ij}} = \frac{1}{2}\left(\varphi'(\pi^*_{ij};\theta_{ij}) + 2\epsilon_i\left(\sum_{\ell=1}^{L}\pi^*_{i\ell} - \mu_i\right) + 2\delta_j\left(\sum_{k=1}^{N}\pi^*_{kj} - \nu_j\right) - \gamma^*_{ij}\right) = 0,\ \forall\,(i,j) \in I\times J. \tag{12}$$

Here, as before, $\gamma_{ij}$ is the associated multiplier to the inequality constraint $\pi_{ij} \geq 0$. Determining whether or not the solution is interior, as in the previous problem, is not trivial at all. For corner solutions, we have to iterate all possible combinations of $\gamma^*_{ij}$ equal or not to zero. Formally, $2^{NL}$ possibilities.

Regarding the existence of a solution to $\mathcal{P}_2$, in order to apply Weierstrass theorem to overcome the potential issue that the optimization is carried over an unbounded set, we can actually restrict the optimization to $\mathbb{R}^{NL}_+ \cap \Omega$, where

$$\Omega = [0, R]^{NL},\ \text{with } R = N \max_{1\leq i\leq N}\{\mu_i\} + L \max_{1\leq j\leq L}\{\nu_j\}.$$

Indeed, it is clear from the cost function $F$ that $F$ is strictly lower over the interior of $\Omega$ than when evaluated on $\partial\Omega$ or outside $\Omega$.

In what follows, we will address the case where there solution is interior. In this case, from KKT we know that $\gamma^*_{ij} = 0$ for all $(i,j) \in I\times J$. Hence, from (12), we have $\nabla F(\pi^*) = 0$. This set of equations can be written in the compact form $A\begin{bmatrix}\pi^*_{11} & \pi^*_{12} & \cdots & \pi^*_{NL}\end{bmatrix}^T = b$, where

$$A = \underbrace{\mathrm{Diag}(a_{11}, a_{12}, \ldots, a_{NL})}_{=D} + \underbrace{\mathrm{Diag}(\epsilon_1, \ldots, \epsilon_N) \otimes \mathbf{1}_{L\times L}}_{=E} + \underbrace{\mathbf{1}_{N\times N} \otimes \mathrm{Diag}(\delta_1, \ldots, \delta_L)}_{=F}, \tag{13}$$

and $b = \left[\epsilon_1\mu_1 + \delta_1\nu_1 - \frac{c_{11}}{2}, \epsilon_1\mu_1 + \delta_2\nu_2 - \frac{c_{12}}{2}, \cdots, \epsilon_N\mu_N + \delta_L\nu_L - \frac{c_{NL}}{2}\right]^T$. The following lemma shows that $A$ is an invertible matrix.

**Lemma 4.1.** *The determinant of $A$ is strictly positive whenever all parameters are strictly positive.*

*Proof.* First, $\det(D) = \prod_{(i,j)\in I\times J} a_{ij} > 0$, $\det(E) = \det(F) = 0$. On the other hand, the eigenvalues of $E$ are non-negative since the eigenvalues of $\mathrm{Diag}(\epsilon_1, ..., \epsilon_N)$ are $\epsilon_i > 0$ and the eigenvalues of $\mathbf{1}_{L\times L}$ belong to $\{0, L\}$. Hence, the products of eigenvalues $\epsilon_i \cdot 0$ and $\epsilon_i \cdot L$ are non-negative, and so, $E$ is positive semi-definite. Similarly, $F$ is positive semi-definite. Thus, $A$ is the sum of a diagonal and positive definite matrix and two other symmetric and semi-positive definite matrices. According to Zhan (2005)[11]

$$\det(A) = \det(D + E + F) \geq \det(D + E) + \det(F) \geq \det(D) + \det(E) + \det(F) > 0. \blacksquare$$

[11] For Minkowski's determinant inequality and its generalizations, see Marcus and Gordon (1970), Artstein-Avidan et al. (2015).

We have already shown that the linear system $A\pi = b$ has a unique solution. What we still don't know is whether or not this solution belongs to $\mathbb{R}^{NL}_{++}$. If so, given the strict convexity of $F$, we would have determined, through an ex-post analysis, the unique solution to $\mathcal{P}_2$. However, it may not always be the case that $A^{-1}b \in \mathbb{R}^{NL}_{++}$. For instance, if $A^{-1}$ has strictly positive entries and $c_{ij} > 2(\epsilon_i\mu_i + \delta_j\nu_j)$ for all $(i,j) \in I \times J$, then $A^{-1}b$ will have strictly negative entries. Therefore, it is of great interest to determine under what conditions on the parameters the solution is interior. We will see a specific case where we can do this.

## 4.1 Neumann's series approach

**Assumption 4.** Let $a_{ij} > 0$ for all $(i,j) \in I \times J$. Assume that

$$\max_{1\leq i\leq N}\{\epsilon_i\} \cdot L + \max_{1\leq j\leq L}\{\delta_j\} \cdot N < \min_{(i,j)\in I\times J}\{a_{ij}\}.$$

Assumption 4 tells us that convex transport costs are large. Moreover, $\epsilon_i, \delta_j$ can be interpreted as normalized weights, i.e., $\epsilon_i, \delta_j \in [0,1]$ and $\sum_i \epsilon_i = \sum_j \delta_j = 1$.

**Lemma 4.2.** *Under Assumption 4, the following holds*

$$A^{-1} = \left(\sum_{k=0}^{\infty}(-1)^k(D^{-1}X)^k\right)D^{-1}.$$

*Proof.* Let $A = D + X$, where $X = E + F$. Then,

$$A^{-1} = (D+X)^{-1} = (I - (-1)D^{-1}X)^{-1}D^{-1}.$$

Then, for all $\lambda \in \sigma(D^{-1}X)$, $\lambda \leq \max_{i,j}\{1/a_{ij}\} \cdot (\lambda^E_{\max} + \lambda^F_{\max})$, where $\lambda^E_{\max} = \max_i\{\epsilon_i\} \cdot L$ and $\lambda^F_{\max} = \max_j\{\delta_j\} \cdot N$. Thus, $\|D^{-1}X\|_\sigma < 1$ [12],

$$(I - (-1)D^{-1}X)^{-1} = \sum_{k=0}^{\infty}(-1)^k(D^{-1}X)^k.$$

Then, by multiplying the series on the right hand side by $D^{-1}$, the claim follows. ∎

**Theorem 4.3.** *Under Assumption 4, the sequence defined by*

$$\pi_n = \left(\sum_{k=0}^{n}(-1)^k(D^{-1}X)^k\right)D^{-1}b = S_n b$$

*converges to $\pi^* = A^{-1}b$.*

[12] $\|\cdot\|_\sigma$ denotes the spectral norm.

*Proof.* Define

$$\mathcal{E}_n = A^{-1} - S_n = \left(\sum_{k=n+1}^{\infty} (-1)^k (D^{-1}X)^k\right) D^{-1}.$$

On one hand $\|\pi_n - \pi^*\|_\infty = \|\mathcal{E}_n b\|_\infty \leq \|\mathcal{E}_n b\|_2$. On the other hand,

$$\|\mathcal{E}_n b\|_2 \leq \sqrt{NL} \left\| \sum_{k=n+1}^{\infty} (-1)^k (D^{-1}X)^k \right\|_\sigma \left\|D^{-1}b\right\|_\infty \leq \frac{\sqrt{NL}\,\|D^{-1}X\|_\sigma^{n+1}\,\|D^{-1}b\|_\infty}{1 - \|D^{-1}X\|_\sigma}.$$

Given $\varepsilon > 0$, let

$$N_\varepsilon = \max\left\{1, \left\lceil \left| \log_{\|D^{-1}X\|_\sigma} \left( \frac{\varepsilon\left(1 - \|D^{-1}X\|_\sigma\right)}{\sqrt{NL}\,\|D^{-1}b\|_\infty} \right) \right| \right\rceil \right\}.$$

For $n \geq N_\varepsilon$, we have $\|\pi_n - \pi^*\|_\infty < \epsilon$. ■

For the aim to explicitly compute $A^{-1}$, we need to impose the following assumptions.

**Assumption 5.** Assume that $\delta_j = 0$ for all $j \in J$ and $D = \beta I$ for some $\beta > 0$.

Assumption 5 represents the scenario where the central planner has no interest in whether the schools are overfilled or underfilled ($F = 0$). This assumption also tell us that distances or other characteristics of the paths between the geographical location of each type of student and each school are the same (reasonable in small cities or villages).

**Assumption 6.** Assume that $L\epsilon_i < \min\{1, \beta\}$ for all $1 \leq i \leq N$.

In line with Assumption 4, Assumption 6 applies when convex transport costs are large.

**Theorem 4.4.** *Under Assumptions 5 and 6, $A^{-1}$ is given as follows*

$$A^{-1} = \frac{I}{\beta} + \frac{1}{\beta} Diag\left(-\frac{\epsilon_1}{\beta + L\epsilon_1}, \ldots, -\frac{\epsilon_N}{\beta + L\epsilon_N}\right) \otimes \mathbf{1}_{L\times L}. \tag{14}$$

*Proof.* By using classical properties of Kronecker product, we have

$$\begin{aligned}
A^{-1} &= \frac{I}{\beta} + \left[\sum_{k=1}^{\infty} (-1)^k \left(\frac{1}{\beta}\right)^k (\mathrm{Diag}(\epsilon_1, \ldots, \epsilon_N) \otimes \mathbf{1}_{L\times L})^k\right] D^{-1} \\
&= \frac{I}{\beta} + \frac{1}{\beta L} \sum_{k=1}^{\infty} (-1)^k \left(\frac{L}{\beta}\right)^k (\mathrm{Diag}(\epsilon_1^k, \ldots, \epsilon_N^k) \otimes \mathbf{1}_{L\times L}) \\
&= \frac{I}{\beta} + \frac{1}{\beta L} \mathrm{Diag}\left(\sum_{k=1}^{\infty} (-1)^k \left(\frac{L\epsilon_1}{\beta}\right)^k, \ldots, \sum_{k=1}^{\infty} (-1)^k \left(\frac{L\epsilon_N}{\beta}\right)^k\right) \otimes \mathbf{1}_{L\times L} \\
&= \frac{I}{\beta} + \frac{1}{\beta} \mathrm{Diag}\left(-\frac{\epsilon_1}{\beta + L\epsilon_1}, \ldots, -\frac{\epsilon_N}{\beta + L\epsilon_N}\right) \otimes \mathbf{1}_{L\times L}.
\end{aligned}$$

■

A similar result could be obtained by taking $E = 0$. However, from the central planner's perspective, priority is reasonably given to the number of students being educated rather than focusing on whether schools are filled to capacity or left with vacancies.

**Corollary 4.5.** *Under Assumptions 5 and 6, the solution of $\mathcal{P}_2$ is given by*

$$\pi_{ij}^* = \frac{b_{ij}}{\beta} - \sum_{\ell=1}^{L} \frac{b_{i\ell}\epsilon_i}{\beta^2 + L\epsilon_i\beta}, \tag{15}$$

*provided that the right-hand side of (15) is positive.*

*Proof.* This result follows directly from the computation of $A^{-1}b$ by using (14). ■

The following assumption let us establish Proposition 4.6, which in turn is used to compare the new model solution with the classical one.

**Assumption 7.** Assume that $\delta_j = 0$, $c_{ij} = c_i < 2\epsilon_i\mu_i$ for all $1 \le j \le L$, and $a_{ij} = 0$ for all $(i, j) \in I \times J$.

Assumption 7 implies that the central planner is indifferent to filling schools and also that costs are linear. Furthermore, it is assumed that, for a given type, the cost of attending any school is the same, except for fixed costs.

**Proposition 4.6.** *Let $x_i^* = \sum_{j=1}^{L} \pi_{ij}^*$. Then, under Assumption 7, $0 < x_i^* < \mu_i$.*

*Proof.* Given Assumption 7, the optimization problem becomes

$$\min_{\pi_{ij} \ge 0} \left\{ \sum_{i,j} d_{ij} + \sum_{i=1}^{N} c_i \sum_{j=1}^{L} \pi_{ij} + \sum_{i=1}^{N} \epsilon_i \left( \sum_{j=1}^{L} \pi_{ij} - \mu_i \right)^2 \right\}. \tag{16}$$

Denote $x_i = \sum_{j=1}^{L} \pi_{ij}$. Then, (16) can be re-written as

$$\min_{x_i \ge 0} \left\{ \sum_{i=1}^{N} c_i x_i + \epsilon_i (x_i - \mu_i)^2 \right\}.$$

Since

$$c_i x_i + \epsilon_i (x_i - \mu_i)^2 = \left( \sqrt{\epsilon_i} x_i + \frac{c_i - 2\epsilon_i\mu_i}{2\sqrt{\epsilon_i}} \right)^2 + \epsilon_i \mu_i^2 - \left( \frac{c_i - 2\epsilon_i\mu_i}{2\sqrt{\epsilon_i}} \right)^2,$$

it follows that

$$x_i^* = \sum_{j=1}^{L} \pi_{ij}^* = \frac{2\epsilon_i\mu_i - c_i}{2\epsilon_i} > 0 \text{ for } 1 \le i \le N. \tag{17}$$

■

Equation 17 shows that, under the unrestricted regime, the solution no longer satisfies $\sum_{j=1}^{L} \pi_{ij} = \mu_i$. Indeed, since $c_i > 0$, $x_i^* < \mu_i$. This is consistent with the implications of the new model since there is a trade off between making everyone of each type of students go to some school study, and matching direct costs.

In what follows we will establish an upper bound for the number of students from any type matched with any school. For this purpose, we first set another assumption.

**Assumption 8.** Let $\rho$ and $\zeta$ be real numbers such that $\rho > 2NL\zeta > 0$, with $a_{ij} = \rho$ and $\epsilon_i = \delta_j = \zeta$ for all $(i, j) \in I \times J$.

Assumption 8 implies that the central planner weights the same for each social objective. Under this assumption, we have $D = \rho I$ and $X = \zeta Y$, where the entries of $Y$ are given by

$$Y_{ij} = \begin{cases} 2 & i = j, \\ 1 & i \neq j \wedge (\lceil i/N \rceil = \lceil j/N \rceil \vee i \equiv j \pmod{N}), \\ 0 & \text{otherwise.} \end{cases}$$

This allows us to rewrite Theorem 4.2 as

$$A^{-1} = \frac{1}{\rho} \left( \sum_{k=0}^{\infty} \left( -\frac{\zeta}{\rho} \right)^k Y^k \right).$$

**Lemma 4.7.** *Let $k \geq 1$ be a positive integer. Then*

$$\max_{1 \leq i,j \leq NL} \left\{ \left( Y^k \right)_{ij} \right\} \leq \frac{(2NL)^k}{NL}.$$

*Proof.* The claim certainly holds for $k = 1$. Now, assuming it holds for $k \geq 1$, it follows by induction that

$$\max_{1 \leq i,j \leq NL} \left\{ \left( Y^{k+1} \right)_{ij} \right\} = \max_{1 \leq i,j \leq NL} \left\{ \sum_{\ell=1}^{NL} \left( Y^k \right)_{i\ell} Y_{\ell j} \right\} \leq \sum_{\ell=1}^{NL} \frac{(2NL)^k}{NL} \cdot 2 = \frac{(2NL)^{k+1}}{NL}. \qquad \blacksquare$$

**Lemma 4.8.** *Let $k \geq 2$ be a positive integer. Then*

$$\frac{(NL)^{\lfloor k/2 \rfloor}}{NL} \leq \min_{1 \leq i,j \leq NL} \left\{ \left( Y^k \right)_{ij} \right\}.$$

*Proof.* We have two distinct possibilities. **Case** $k = 2m$ with $m \geq 1$ **.** We now proceed by induction. We will manually verify that each $(Y^2)_{ij} = \sum_{\ell=1}^{NL} Y_{i\ell} \cdot Y_{\ell j}$ satisfies the inequality. On the diagonal we have

$$\left( Y^2 \right)_{ii} = \sum_{\substack{\ell=1 \\ \ell \neq i}}^{NL} Y_{i\ell} \cdot Y_{\ell i} + Y_{ii} \cdot Y_{ii} \geq 4.$$

For $i \neq j$, set

$$\ell_0 = N\left(\left\lceil \frac{j}{N} \right\rceil - \left\lfloor \frac{i-1}{N} \right\rfloor - 1\right) + i.$$

Then $\ell_0 \equiv i \pmod N$ and so $Y_{i\ell_0} \geq 1$. On the other hand,

$$\ell_0 \in \left[N\left(\left\lceil \frac{j}{N} \right\rceil - 1\right) + 1, N\left\lceil \frac{j}{N} \right\rceil\right]$$

implies $\lceil \ell_0/N \rceil = \lceil j/N \rceil$. So, $Y_{\ell_0 j} \geq 1$. It follows that

$$\left(Y^2\right)_{ij} = \sum_{\substack{\ell=1 \\ \ell \neq \ell_0}}^{NL} Y_{i\ell} \cdot Y_{\ell j} + Y_{i\ell_0} \cdot Y_{\ell_0 j} \geq 1.$$

Assuming $\min_{1 \leq i,j \leq NL} \left\{(Y^{2m})_{ij}\right\} \geq (NL)^m/NL$ holds for $m \geq 1$, we obtain

$$\min_{1 \leq i,j \leq NL} \left\{\left(Y^{2m+2}\right)_{ij}\right\} = \min_{1 \leq i,j \leq NL} \left\{\sum_{\ell=1}^{NL} \left(Y^{2m}\right)_{i\ell} \cdot \left(Y^2\right)_{\ell j}\right\} \geq \sum_{\ell=1}^{NL} \frac{(NL)^m}{NL} = \frac{(NL)^{m+1}}{NL}.$$

**Case $k = 2m+1$ with $m \geq 1$ .** We prove this by induction on $m$ starting with the base case $Y^3$:

$$\left(Y^3\right)_{ij} = \sum_{\ell=1}^{NL} \left(Y^2\right)_{i\ell} \cdot Y_{\ell j} = \sum_{\substack{\ell=1 \\ \ell \neq j}}^{NL} \left(Y^2\right)_{i\ell} \cdot Y_{\ell j} + \left(Y^2\right)_{ij} \cdot Y_{jj} \geq 2.$$

Assume the statement holds for $m \geq 1$, then

$$\min_{1 \leq i,j \leq NL} \left\{\left(Y^{2m+3}\right)_{ij}\right\} = \min_{1 \leq i,j \leq NL} \left\{\sum_{\ell=1}^{NL} \left(Y^{2m+1}\right)_{i\ell} \cdot \left(Y^2\right)_{\ell j}\right\} \geq \sum_{\ell=1}^{NL} \frac{(NL)^m}{NL} = \frac{(NL)^{m+1}}{NL}.$$

This completes the proof. ■

**Theorem 4.9.** *Under Assumptions 4 and 8, the lower and the upper bounds of* $(A^{-1})_{ij}$ *can be expressed in terms of* $N, L, \zeta$ *and* $\rho$,

$$C_1(N, L, \zeta, \rho) \leq (A^{-1})_{ij} \leq C_2(N, L, \zeta, \rho), \tag{18}$$

*where*

$$C_1 = \frac{\zeta\left(4\zeta N^3 L^3\left(2\zeta^3 - 2\zeta\rho^2 - \rho^3\right) + 8N^2L^2\rho^2\left(\rho^2 - \zeta^2\right) + \zeta NL\rho^2(2\zeta + \rho) - 2\rho^4\right)}{\rho^4\left(\zeta^2 NL - \rho^2\right)(2NL-1)(2NL+1)}$$
$$C_2 = \frac{\zeta^2 NL\rho(4NL-1)}{\left(\rho^2 - \zeta^2 NL\right)(\rho - 2NL\zeta)(\rho + 2NL\zeta)}.$$

*Proof.* We write $A^{-1}$ in terms of $Y$

$$A^{-1} = \frac{1}{\rho}\left(I - \left(\frac{\zeta}{\rho}\right) Y + \sum_{m\geq 1}\left(\frac{\zeta}{\rho}\right)^{2m} Y^{2m} - \sum_{m\geq 1}\left(\frac{\zeta}{\rho}\right)^{2m+1} Y^{2m+1}\right)$$

and apply Lemmas 4.7 and 4.8 to bound the series as follows,

$$\frac{\zeta^2 NL}{\rho^2 - \zeta^2 NL} \leq \sum_{m\geq 1}\left(\frac{\zeta}{\rho}\right)^{2m} \left(Y^{2m}\right)_{ij} \leq \frac{4\zeta^2 N^2 L^2}{\rho^2 - 4\zeta^2 N^2 L^2}$$

and

$$\frac{\rho^3}{\rho(\rho^2 - \zeta^2 NL)} \leq \sum_{m\geq 1}\left(\frac{\zeta}{\rho}\right)^{2m+1} \left(Y^{2m+1}\right)_{ij} \leq \frac{8\zeta^3 N^2 L^2}{\rho(\rho^2 - 4\rho^2 N^2 L^2)}.$$

Therefore, $(A_{ij})^{-1}$ is bounded from above by

$$\frac{1}{\rho}\left(1 + \frac{4\zeta^2 N^2 L^2}{\rho^2 - 4\zeta^2 N^2 L^2} - \frac{\rho^3}{\rho(\rho^2 - \zeta^2 NL)}\right),$$

and from below by

$$\frac{1}{\rho}\left(-2\left(\frac{\zeta}{\rho}\right) + \frac{\zeta^2 NL}{\rho^2 - \zeta^2 NL} - \frac{8\zeta^3 N^2 L^2}{\rho(\rho^2 - 4\rho^2 N^2 L^2)}\right).$$

From here, (18) follows. ■

**Corollary 4.10.** *Under Assumptions 4 and 8, it follows that* $\pi^*_{ij} \leq NL\tilde{C}$, *for all* $(i,j) \in I \times J$, *where*

$$\tilde{C} = \max\{|C_1|, C_2\} \cdot \max_{\substack{1\leq i\leq N\\ 1\leq j\leq L}} \left\{\left|(\epsilon_i \mu_i + \delta_j \nu_j) - \frac{c_{ij}}{2}\right|\right\}.$$

*Proof.* By triangle inequality,

$$\begin{aligned}
\pi^*_{ij} &\leq ||\pi^*||_\infty \\
&= \max_{\substack{1\leq i\leq N\\ 1\leq j\leq L}} \left\{\left|\sum_{k=1}^{NL} \left(A^{-1}\right)_{(i-1)L+j \quad k} \cdot b_{\lceil k/L\rceil \quad k-L\lfloor (k-1)/L\rfloor}\right|\right\} \\
&\leq \sum_{k=1}^{NL} \max_{\substack{1\leq i\leq N\\ 1\leq j\leq L}} \left|\left(A^{-1}\right)_{ij}\right| \cdot \max_{\substack{1\leq i\leq N\\ 1\leq j\leq L}} |b_{ij}| \\
&= NL\tilde{C}.
\end{aligned}$$

■

## 4.2 Algorithm for computing $\pi^*$

We now provide an efficient algorithm to compute $\pi^*$. To this aim let us re-write matrix $A$ given in (13) as follows:

$$A = \mathrm{Diag}(a_{11}, \ldots, a_{NL}) + \sum_{i=1}^{N} \left(\epsilon_i^{1/2}\mathbf{e}_i \otimes \mathbf{1}_{L\times 1}\right)\left(\epsilon_i^{1/2}\mathbf{e}_i^T \otimes \mathbf{1}_{1\times L}\right) \tag{19}$$
$$+ \sum_{j=1}^{L} \left(\delta_j^{1/2}\mathbf{e}_j \otimes \mathbf{1}_{N\times 1}\right)\left(\delta_j^{1/2}\mathbf{e}_j^T \otimes \mathbf{1}_{1\times N}\right).$$

**Theorem 4.11.** *For interior solutions $\pi^*$, there is a deterministic algorithm that computes it in $O((N+L)(NL)^2)$ time.*

*Proof.* Consider Algorithm 1. First, it is easy to see that each prefix sum of $A$ is invertible. Hence, we can iteratively apply the Sherman-Morrison formula with a rank-1 update at each step. Then, it is clear that Lines 3 and 12 take $O((NL)^2)$. First, the number of iterations for the for-loops on Lines 4-7 and 8-11 is $N+L$. We then show that each time we enter any for-loop, the time spent is $O((NL)^2)$. Computing $1 + w^T A^{-1} w$ takes $O((NL)^2)$, so the only possible optimization is finding the optimal parenthesization for the product $A^{-1}ww^TA^{-1}$. Since there are only five possible ways to parenthesize the expression, we determine by brute force that computing $(A^{-1}w)(w^TA^{-1})$ also takes $O((NL)^2)$. This implies the desired time complexity of $O((N+L)(NL)^2)$. ■

**Algorithm 1** OPTIMIZE $(a, b, \epsilon_1, \ldots, \epsilon_N, \delta_1, \ldots, \delta_L)$

1: **Input:** Matrices $a \in \mathbb{R}_{++}^{NL}$, $b \in \mathbb{R}^{NL}$ and parameters $\epsilon_1, \ldots, \epsilon_N, \delta_1, \ldots, \delta_L \in \mathbb{R}_{++}$
2: **Output:** $\pi^* \in \mathbb{R}^{NL}$
3: Initialize $A^{-1} \leftarrow \mathrm{Diag}(1/a_{11}, \ldots, 1/a_{NL}) \in \mathbb{R}^{NL,NL}$
4: **for** $i \leftarrow 1, \ldots, N$ **do**
5: Define $u^{(i)} \in \mathbb{R}^{NL}$ by $u^{(i)} := \epsilon_i^{1/2}\mathbf{e}_i \otimes \mathbf{1}_{L\times 1}$
6: $A^{-1} \leftarrow A^{-1} - \dfrac{A^{-1}u^{(i)}u^{(i)T}A^{-1}}{1 + u^{(i)T}A^{-1}u^{(i)}}$ via Sherman-Morrison formula
7: **end for**
8: **for** $j \leftarrow 1, \ldots, L$ **do**
9: Define $v^{(j)} \in \mathbb{R}^{NL}$ by $v^{(j)} := \delta_j^{1/2}\mathbf{e}_j \otimes \mathbf{1}_{N\times 1}$
10: $A^{-1} \leftarrow A^{-1} - \dfrac{A^{-1}v^{(j)}v^{(j)T}A^{-1}}{1 + v^{(j)T}A^{-1}v^{(j)}}$ via Sherman-Morrison formula
11: **end for**
12: **return** $A^{-1}b$

Note that folklore algorithms require $O((NL)^3)$ arithmetic operations. Moreover, when $L$ is similar to $N$ (i.e. $L \in \Theta(N)$), we can compute $A^{-1}$ in $O(N^5)$ which is significantly faster than the naive approach.

### 4.3 Comparative statics

Although we know how to compute $\pi^*$ though Neumann's series or Algorithm 1, it is not straightforward to obtain a closed formula for $\pi^*_{ij}$ from these techniques. Therefore, with the aim to perform comparative statics, a path that can be followed is to approximate matrix $A^{-1}$ by the Neumann's series. First, assume that $A^{-1} \simeq D^{-1}$. This allows us to obtain a simple closed-form formula for $\pi^*_{ij}$ that provides initial insights. Under the framework $A^{-1} \simeq D^{-1}$, $\pi^*_{ij} \simeq (2(\epsilon_i\mu_i + \delta_j\nu_j) - c_{ij})/(2a_{ij})$. Hence, $\partial\pi^*_{ij}/a_{ij}, \partial\pi^*_{ij}/c_{ij} < 0$ and $\partial\pi^*_{ij}/\partial\epsilon_i, \partial\pi^*_{ij}/\partial\delta_j, \partial\pi^*_{ij}/\partial\mu_i, \partial\pi^*_{ij}/\partial\nu_j > 0$. These results could be anticipated. However, note that under this rough approximation, $\partial\pi^*_{ij}/\partial\theta_{k\ell} = 0$ for $(k,j) \neq (i,j)$. This is not very realistic since we expect a substitution effect.

Let us consider now the approximation $A^{-1} \sim D^{-1} - D^{-1}XD^{-1} = D^{-1} - (D^{-1})^2X$. We know from smooth comparative statics that, if $\pi^* \in \mathbb{R}^{NL}_{++}$ is an interior solution to $\mathcal{P}_2$ associated to the parameter vector $(\overline{\theta}, \epsilon, \delta, \mu, \nu) \in \mathbb{R}^{2NL}_{++} \times \mathbb{R}^N_{++} \times \mathbb{R}^L_{++} \times \mathbb{R}^N_{++} \times \mathbb{R}^L_{++}$, then

$$\left[\frac{\partial\pi^*_{ij}}{\partial\theta_{k\ell}}\right] = -\underbrace{A^{-1}_{(\overline{\theta},\epsilon,\delta,\mu,\nu)}}_{HF(\pi^*,\overline{\theta},\epsilon,\delta,\mu,\nu)^{-1}} [I_{NL\times NL}\,|\,2\mathrm{Diag}(\pi^*_{11}, \cdots, \pi^*_{NL})]. \tag{20}$$

Thus, under the approximation $A^{-1} \sim D^{-1} - (D^{-1})^2X$

$$\left[\frac{\partial\pi^*_{ij}}{\partial\theta_{k\ell}}\right] = \left[\frac{\partial\pi^*_{ij}}{\partial c_{k\ell}} \,\middle|\, \frac{\partial\pi^*_{ij}}{\partial a_{k\ell}}\right] \simeq -[D^{-1} - (D^{-1})^2X \,|\, A^{-1}_{\Pi,2}]. \tag{21}$$

From (21), if $\max_{i,j}\{\epsilon_i + \delta_j\} < 1$, then $\partial\pi^*_{ij}/\partial\theta_{ij} < 0$ for all $(i,j) \in I \times J$, $\partial\pi^*_{ij}/\partial\theta_{k\ell} > 0$ for $i \neq k$ and $j = \ell$ or $i = k$ and $j \neq \ell$, and $\partial\pi^*_{ij}/\partial\theta_{k\ell} = 0$ if $i \neq k$ and $j \neq \ell$. Indeed, since multiplying $D^{-1} - (D^{-1})^2X$ by $[I_{NL}|2\mathrm{Diag}(\pi^*_{ij})]$ generates the partitioned matrix $[D^{-1} - (D^{-1})^2X|A^{-1}_{\Pi,2}]$, where $A^{-1}_{\Pi,2}$ consists of multiplying the column $k$ of $D^{-1} - (D^{-1})^2X$ by $\pi^*_k$, from (20) we can conclude that $\partial\pi^*_{ij}/\partial c_{ij} = -(1 - (\epsilon_i + \delta_j))/a^2_{ij} < 0$, $\partial\pi^*_{ij}/\partial c_{i\ell} = \epsilon_i/a^2_{ij} > 0$, $\partial\pi^*_{ij}/\partial c_{kj} = \delta_j/a^2_{ij} > 0$, $\partial\pi^*_{ij}/\partial c_{k\ell} = 0$ if $i \neq k$ and $j \neq \ell$, $\partial\pi^*_{ij}/\partial a_{ij} = -2\pi^*_{ij}(1 - (\epsilon_i + \delta_j))/a^2_{ij} < 0$, $\partial\pi^*_{ij}/\partial a_{i\ell} = 2\pi^*_{i\ell}\epsilon_i/a^2_{ij} > 0$, $\partial\pi^*_{ij}/\partial a_{kj} = 2\pi^*_{kj}\delta_j/a^2_{ij} > 0$, and $\partial\pi^*_{ij}/\partial a_{k\ell} = 0$ if $i \neq k$ and $j \neq \ell$.

These results are much closer to what we would expect to observe. Indeed, we now have a substitution effect: if the cost of matching individuals of type $i$ with $j$ increases, then the number of students of type $i$ matched with $\ell$, different from $j$, increases. However, note that these results are obtained under the truncation of the Neumann series and should be interpreted only in that context: as an approximation.

Under Assumptions 4, 5 and 6, it possible to compute the effects of the parameters directly by using (15).

# 5 Conclusions

We have proposed and analyzed two models derived from the quadratically regularized optimal transport problem. Our contribution, in line with the previous work related to entropic and quadratic regularization, introduces a modification to the objective function. This modification addresses a recurring observation in cities where congestion affects transportation costs in a *superlinear* manner in relation to the mass being transported, i.e., the number of matched individuals. In both models, we investigate on the characterization of interior or corner solutions, methods for solving the problem, and comparative statics.

In the first model, we derive the first-order conditions and analyze the case of interior solutions. We demonstrate that this leads to a singular system, and for $N = L = 2$, we provide general conditions for the solution to be interior. Regarding the analysis of monotone or smooth comparative statics, we show that it is not possible to perform the former in the general case, and the later under the assumption of an interior solution. Finally, in the integer setting, under certain assumptions on the parameters, we explicitly obtain the optimal matching.

For the second model, we provide an exhaustive analysis for the case of interior solutions. Under the approach of the Neumann series and a set of assumptions, we obtain results regarding the convergence to the optimal matching, bounding $\pi^*_{ij}$ for all $(i, j) \in I \times J$, and computing the optimal matching for specific cases. Given the properties of the objective function, an ex-post analysis allows us to conclude the optimality of a critical point for the function $F$. Finally, we present an algorithm that efficiently computes the optimal solution based on Sherman and Morrison (1950). We conclude the model's analysis with a study of smooth comparative statics, which provides important insights into the behavior of the solution. In particular, we highlight the substitution effects considering a first order approximation, and, for the specific case where $\pi^*_{ij}$ is given by (15), the partial derivatives can be computed straightforwardly.

Both models present theoretical challenges in terms of calculating optimal solutions due to the quadratic term, which is key for modeling congestion costs. In both models, a numerical solution can be obtained using computational tools such as Wolfram Mathematica.

Finally, in line with both the entropic and quadratic regularized problem, our models yield distinct properties for the solution compared to the classical approach, particularly regarding the fact that in the linear model, a solution always lies at a vertex.

The studied models can be applied to various contexts where convexity due to congestion provides a good representation of the matching cost structure. By estimating the parameters, following for example Agarwal and Somaini (2023), the model can be used to draw empirical conclusions.

## CRediT authorship contribution statement

**Marcelo Gallardo:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing.
**Manuel Loaiza:** Formal analysis, Investigation, Software, Validation, Writing – original draft.
**Jorge Chávez:** Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review and editing.

## Declaration of competing interest

None.

## Data availability

No data was used for the research described in the article.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT-4o in order to assist with grammar correction and to make paragraphs more concise. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

## Acknowledgments

Chávez acknowledges financial support from Pontificia Universidad Católica del Perú.

## Appendix A. Optimal Matchings for $\mathcal{P}_O$ and $\mathcal{P}_1$

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 3.5.

# Appendix B. Inverse of Matrix $A$

We consider the case where $N = 3$, $L = 2$, $a = 2 \cdot 1_{N \times L}$, $\epsilon = [0.107014, 0.163166, 0.102569]^T$, and $\delta = 0$. Under this parameters configuration, Assumptions 4, 5, and 6, are satisfied. The inverse of the matrix $A$, as defined in (13), computed using `Inverse`, coincides with the formula derived in Theorem 4.4 computed using `KroneckerProduct`:

$$\begin{bmatrix} 0.462037 & -0.0379635 & 0. & 0. & 0. & 0. \\ -0.0379635 & 0.462037 & 0. & 0. & 0. & 0. \\ 0. & 0. & 0.472764 & -0.0272361 & 0. & 0. \\ 0. & 0. & -0.0272361 & 0.472764 & 0. & 0. \\ 0. & 0. & 0. & 0. & 0.460069 & -0.0399315 \\ 0. & 0. & 0. & 0. & -0.0399315 & 0.460069 \end{bmatrix}.$$

# Appendix C. Optimal Matching for $\mathcal{P}_2$

We consider the penalized model with $N = 3$ and $L = 2$. The following parameters satisfy Assumption 4.

$$d = \begin{bmatrix} 3 & 4 \\ 2 & 5 \\ 5 & 5 \end{bmatrix}, \quad c = \begin{bmatrix} 1.30436 & 1.72858 \\ 1.5623 & 1.20598 \\ 1.10019 & 1.2187 \end{bmatrix}, \quad a = \begin{bmatrix} 1.02308 & 1.45588 \\ 1.36407 & 1.1021 \\ 1.16638 & 1.22178 \end{bmatrix},$$

$$\mu = \begin{bmatrix} 26 \\ 27 \\ 47 \end{bmatrix}, \quad \nu = \begin{bmatrix} 61 \\ 39 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} 0.130457 \\ 0.132428 \\ 0.191539 \end{bmatrix}, \quad \delta = \begin{bmatrix} 0.196703 \\ 0.158533 \end{bmatrix}.$$

Then, the optimal matching $\pi^*$ utilizing `QuadraticOptimization` and computing $A^{-1}b$ coincides with

$$\pi^* = \begin{bmatrix} 8.17174 & 3.29304 \\ 6.19868 & 4.79052 \\ 10.4517 & 7.18412 \end{bmatrix}.$$

The optimal transport plan does not result in all individuals being matched, meaning the targets $\mu_i$ and $\nu_j$ are not fully met. This is consistent with the structure of the objective function in problem $\mathcal{P}_2$.

# 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.

Artstein-Avidan, S., Giannopoulos, A., and Milman, V. D. (2015). *Asymptotic Geometric Analysis, Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society.

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. and Galichon, A. (2022). A Note on the Estimation of Job Amenities and Labor Productivity. *Quantitative Economics*, 13:153–177.

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.

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. (2024). Quantitative convergence of quadratically regularized linear programs. *arXiv preprint arXiv:2408.04088*.

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.

Izmailov, A. F. and Solodov, M. V. (2023). Convergence rate estimates for penalty methods revisited. *Computational Optimization and Applications*, 85(3):973–992.

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*.

Marcus, M. and Gordon, W. R. (1970). An extension of the minkowski determinant theorem. Received 21st September 1970.

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.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38.

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.

Sherman, J. and Morrison, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127.

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*.

Zhan, S. (2005). On the determinantal inequalities. *Journal of Inequalities in Pure and Applied Mathematics*, 6(4):Article 105.