Stable Matching as Transportation*

Federico Echenique[†] Joseph Root[‡] Fedor Sandomirskiy[§]

February 20, 2024
the latest version: **link**

Abstract

We study matching markets with aligned preferences and establish a connection between common design objectives—stability, efficiency, and fairness—and the theory of optimal transport. Optimal transport gives new insights into the structural properties of matchings obtained from pursuing these objectives, and into the trade-offs between different objectives. Matching markets with aligned preferences provide a tractable stylized model capturing supply-demand imbalances in a range of settings such as partnership formation, school choice, organ donor exchange, and markets with transferable utility where bargaining over transfers happens after a match is formed.

1 Introduction

We study matching markets with aligned preferences, meaning that a matched pair confers the same utility to each agent in the pair, and establish a connection between such markets and the theory of optimal transportation. We argue that markets with aligned preferences model economically relevant environments, and that a rich theory is made possible by exploiting the connection to optimal transport. Our main results concern stable matchings, utilitarian welfare, and egalitarian matchings. Our findings

*The paper has benefited from suggestions and comments by Sophie Bade, Peter Biro, Mark Braverman, Modibo Camara, David Kempe, Axel Niemeyer, Afshin Nikzad, Piyush Panigrahi, Marek Pycia, Doron Ravid, Kirill Rudov, Utku Unver, Rakesh Vohra, Matt Weinberg, audiences at ASU, Caltech, MSRI, Princeton, and the Simons Institute.

[†]Department of Economics, University of California, Berkeley
[‡]Department of Economics, University of Chicago
[§]Department of Economics, Princeton University

serve, among other things, to describe structural properties of stable matching in "spatial" settings, and highlight the role of supply-demand imbalances.

The canonical problem of optimal transport corresponds to maximizing utilitarian welfare, i.e., an aggregate of agents' utilities. We show that if we instead maximize a certain convex transformation of agents' utilities, then the solution to the optimal transport problem is approximately stable. In fact, a single parameter α captures how convex the objective is. We show that when $\alpha > 0$, then a solution to the optimal transport problem is approximately stable, and no blocking pair can generate more than $\ln(2)/\alpha$ additional utility.

For the parameter $\alpha = 0$, we recover the utilitarian welfare objective. Taking negative α, we obtain a concave transformation of agents' utilities. Solving the optimal transport problem with $\alpha < 0$ leads to an approximately egalitarian matching with approximation factor $\frac{1}{|\alpha|} \max\{1, \ln|\alpha|\}$. As a simple corollary, we get the existence of stable and egalitarian matchings in general—potentially infinite—matching markets.

The main insight can be summarized as follows:

> *The same optimal transportation problem leads to approximately stable, welfare-maximizing, or egalitarian matchings, depending on how a single knob (the parameter α) is turned.*

This perspective allows us to recast stability in like terms with efficiency and fairness. A concave transformation of utilities is often used to model a planer concerned with fairness: their objective is improved less by increasing the utility of agents who already have high utility than by increasing the utility of worse-off agents. Stability corresponds to the opposite objective. A planner seeking a stable matching prioritizes high-utility agents, ignoring the negative externalities their matches may impose on the rest of the population. Indeed, as we observe, stable matchings may lead to a loss of fairness and welfare. However, the connection to optimal transport implies a general bound on how dramatic this loss can be.

We specialize our general results to models with a spatial structure, where agents are identified with points in $\mathbb{R}^d$ and want to match with partners who are close to them. For $d = 1$, we identify a "no-crossing" property satisfied by stable matchings; the property allows us to obtain an efficient algorithm for constructing stable matchings. No-crossing is a prominent property in optimal transport, further motivating the connection between stable matching and optimal transport.

Our model calls attention to unfairness that stems from supply and demand imbalances. The most well-studied source of unfairness associated with stable matching is the conflict of interest "across" the market, between the two sides. Firms and workers in a labor market; or men and women in a heterosexual marriage market. This conflict is reflected in the fundamental unfairness of the deferred acceptance algorithm, which chooses the best matching for one of the sides of the market and the worst for the other. The conflict and corresponding unfairness have motivated the consideration of median matchings, max-min criteria, and other mechanisms. In our paper, however, we effectively shut down the problem by focusing on aligned preferences. In our case, the stable matching is often (but not always!) unique, and the usual concern for unfairness plays no role.

Instead, we study the conflict that arises because of supply and demand imbalances, and show that it uncovers a different source of unfairness; one that we believe is of independent interest. In a market with aligned preferences, there may be agents who have the ability to generate large utilities, but that are also in excess supply. These agents can then receive worse matches than other agents who have the same potential for high-utility matches. As a consequence, there can be envy between similar agents, despite the stability of the match.

Our results generalize from two-sided markets to multi-sided markets. It is well known that—without the assumption of preference alignment —stable matching may fail to exist even for three-sided finite markets. Aligned preferences guarantee existence, and our main result generalizes to the multi-sided case. This opens up our results to many economic applications, including models of team formation, supply chains, and organ exchange.

A major advantage of our theory is that it encompasses finite and infinite markets, including markets with countably and uncountably infinite agents. The theory of stable matching has often developed in successive steps. The initial theory was developed for finite markets. Then an extension to infinite markets was studied in order to better understand incentives, which then poses the problem of whether large, but finite, markets behave in ways that are similar to those of infinite markets. The generality of our model means that our ideas are independent of any assumptions regarding the cardinality, and other technical aspects, of the set of agents. For a continuum of agents, our matching markets gain extra tractability.

The connection to optimal transport offers a new perspective on stability, even in finite markets. Ours is not the first to connect stability with linear programming. Existing results show how stability can be written as a finite system of linear inequalities. By contrast, we show how stability results from picking the right objective function and how a modification of this objective leads to egalitarian matchings.

We conclude the introduction with a discussion of aligned preferences. Preferences are aligned when any two agents enjoy the same payoff from being matched. Note that preferences being aligned does not mean that different agents have the same preferences over partners. It means that there exists a utility representation for the agents' preferences for which the utility that each member of a pair enjoys is the same, if the pair were to be matched.

In particular, in a finite matching market without indifferences, there will be a pair of agents that generates the highest utility in the market. Such agents would form a blocking pair unless matched together. The same logic applies to a pair of agents who generate the next-highest level of utility, and so on. As a consequence, it is easy to see that a stable matching exists in finite markets with aligned preferences. Markets with aligned preferences have been mostly overlooked by the matching literature. Notable exceptions are papers by Ferdowsian, Niederle, and Yariv (2023), who study the convergence of decentralized dynamics to a stable matching, and Lee and Yariv (2018), who analyze efficiency loss in stable matchings in markets with random aligned preferences.

Markets with aligned preferences cover a lot of ground. First, many economic applications of matching markets have an important aligned component, which motivates us to consider preference alignment in isolation. For example, in the matching of organ donors to patients, the matching is organized using a priority ranking over the patients that is built up, among other considerations, from the estimated quality of the potential match between the donor and the patient. The quality of the match is, of course, also a key component of the patients' ranking over organs. In other words, the "preference" on one side of the market is the priority over patients, which reflects in part the quality of the matching between the transplanted organ and the patient. The preference on the other side of the market is a ranking of the organs that is also based on the quality of the match. The two preferences are aligned because the quality of the match is the same for both sides of the market.

One may imagine many other examples where the driving force of the agents' motivation to match is a quality component that is common to each member of the pair. In forming partnerships among firms, doubles in tennis, or coauthorships in academia, the quality of a match seems to be a common source of each individual agent's utility from forming a pair. In school choice, proximity is a key component of both student preferences over schools and school priorities (Walters, 2018; Laverde, 2022). Actually, it is hard to think of examples where some idea of match quality is *not* a key driving source of agents' preferences. Of course, the assumption of exact preference alignment is not always perfectly descriptive, but it is arguably an approximation that is worth understanding better than the current state of the literature.

Second, matching markets with transferable utility (matching "with money") are often used to capture various economic problems, ranging from platforms that connect buyers and sellers (Rochet and Tirole, 2003; Armstrong, 2006), to general two-sided markets with heterogeneous goods (Shapley and Shubik, 1971), as well as heterosexual marriage (Becker, 1973; Galichon and Salanié, 2022). Here, a scheme of transfers is determined at the same time as a matching is formed. It is, however, possible that bargaining over transfers takes place *after* a match is formed, and is not part of the matching process. In this case, there are natural circumstances in which the agents' preferences are automatically aligned.

Consider, to fix ideas, the case of a marriage market. In the basic model of Becker (1973)—which is a version of Shapley and Shubik (1971)—there are two finite sets of agents, X and Y; and each pair $(x, y) \in X \times Y$ can generate a surplus $u(x, y)$ that the agents in the pair can divide among themselves in any way they like. This is a model of matching with transferable utility because the utilities of x and y can be any two numbers that add up to $u(x, y)$. In Becker's story, transferable utility arises because potential partners bargain over the allocation of chores and resources within the household; but the model assumes that such bargaining takes place at the same time as a match is formed. If the couple cannot commit to a transfer schedule—meaning that the couple generates a certain surplus, but they cannot write a contract over how the surplus will be shared if the couple is to match—then the bargaining over transfers takes place after the match is formed. Now, if post-match bargaining follows a Nash bargaining protocol with some fixed, exogenous, outside options, then the agents' preferences are aligned because each agent receives a constant fraction of the total surplus $u(x, y)$. See Section 6.1 for a more formal argument.

1.1 Related literature

The economic literature on matching markets with aligned preferences is sparse. A systematic study of markets with aligned preferences was initiated by Ferdowsian, Niederle, and Yariv (2023) and Niederle and Yariv (2009), studying the convergence of decentralized matching dynamics in such markets. The observation that a stable matching for a finite population with aligned preferences can be obtained via a greedy algorithm—iteratively matching the unmatched pair with the highest utility—is implicitly contained in Eeckhout (2000); Clark (2006) who provide conditions for uniqueness of a stable matching; see also Gutin, Neary, and Yeo (2023). Galichon, Ghelfi, and Henry (2023) extend the greedy algorithm to many-to-one stable matching and observe its unfairness in simulations, a phenomenon we explore further.

Our paper contributes to understanding the stability-efficiency trade-off for markets with aligned preferences, complementing insights by Cantillon, Chen, and Pereyra (2022) and Lee and Yariv (2018). Cantillon, Chen, and Pereyra (2022) introduce a school choice environment where preference alignment becomes a descriptive property. Under a condition generalizing alignment, they demonstrate that stable matchings are Pareto optimal. Our paper relies on a quantitative approach to efficiency, and demonstrates that Pareto optimality may entail a loss of utilitarian welfare. Lee and Yariv (2018) consider large markets with random preferences. If a common utility for each pair is drawn independently from a continuous distribution, they show that the stable matching has utilitarian welfare close to optimal. Our results impliy that this phenomenon is specific to having no correlation across agents' preferences: in the presence of correlation, stability may cost up to half of the optimal welfare.

A more general perspective on preference alignment is given by Pycia (2012), who studies a general coalition-formation model that has many-to-one matching models as a special case. He focuses on allowing for complementarities and peer effects and proves the existence of stable outcomes under a richness assumption on feasible coalitions and preferences. Echenique and Yenmez (2007) make a related point for many-to-one matching, based on the earlier ideas of Banerjee, Konishi, and Sönmez (2001) in coalition formation. Pycia's work also precedes our discussion in Section 6.1 on the connection between preference alignment and second-stage bargaining.

Our study touches on several other strains of literature. In computer science, aligned preferences have appeared under the name of "globally ranked pairs." Abraham, Levavi, Manlove, and O'Malley (2008) studies aligned preferences in roommates problem and Lebedev, Mathieu, Viennot, Gai, Reynier, and De Montgolfier (2007) applies them to peer-to-peer networks. For an empirical use of markets with aligned preferences, see, for example, Agarwal (2015) and Sørensen (2007). Our findings are potentially applicable to questions arising from models of international trade, where a large and active recent literature emphasizes the spatial aspects of buyer-seller networks; for example, Chaney (2014); Antras, Fort, and Tintelnot (2017); Panigrahi (2021).

The connection to optimal transport established in our paper supports the common wisdom that matching models with a continuum of agents can be more intuitive than their finite-population counterparts; see, for example, Ashlagi and Shi (2016); Azevedo and Leshno (2016); Leshno and Lo (2021); Arnosti (2022). With optimal transport methods, and for the problems we have focused on, the cardinality of the space of agents is largely irrelevant. Connections to optimal transport have been recently discovered and used in various areas of economic theory, for example, mechanism design (Daskalakis, Deckelbaum, and Tzamos, 2015; Perez-Richet and

Skreta, 2023; Kolesnikov, Sandomirskiy, Tsyvinski, and Zimin, 2022; McCann and Zhang, 2023), information design (Malamud, Cieslak, and Schrimpf, 2021; Arieli, Babichenko, and Sandomirskiy, 2023), and many others; see surveys by Ekeland (2010); Carlier (2012); Galichon (2018). We find a link between stability and a particular area of optimal transportation known as concave transport, pioneered by McCann (1999); Gangbo and McCann (1996). To the best of our knowledge, concave transport has not appeared in economic applications, with the exception of Boerma, Tsyvinski, Wang, and Zhang (2023). Their paper relies on concave transport to study labor market sorting—a matching market with transferable utility—and is the closest to ours in the technical dimension. The literature on markets with and without transfers is almost disjoint as such markets have few similarities. The possibility of approaching both models from a similar concave transport perspective shows its generality and importance.

2 Matching on the Line

To motivate our approach, we first discuss stable matching when agents are identified with points on the real line, $\mathbb{R}$, and all agents prefer partners close to them. The general model is deferred to Section 3. A feature of the real line is that the unique stable matching can be easily constructed and visualized. Its structure hints at the connection to optimal transport explored thereafter.

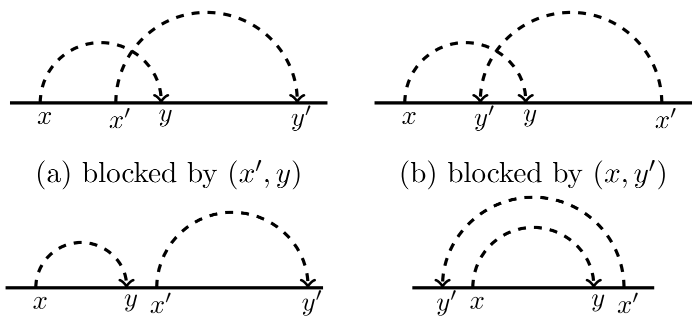
The two sides of the market, X and Y, are described by non-atomic measures, μ and ν, on $\mathbb{R}$. Imagine, for example, horizontally differentiated workers and firms. A location $x \in \mathbb{R}$ represents the type of an agent, and $\mu([a, b])$ the population of X agents with types located in the interval $[a, b]$. We assume that $\mu(X) = \nu(Y)$. Two agents with types x and y get utility $-|x - y|$ from matching together. A matching is a joint distribution π on $\mathbb{R}^2$ with marginals μ and ν. The matching π is stable if, for any two pairs (x_1, y_1) and (x_2, y_2) in the support of π,

$$|x_1 - y_2| \geq \min\{|x_1 - y_1|, |x_2 - y_2|\}.$$

That is, a matching is stable if it is not blocked by some pair of agents x_1 and y_2 who would both strictly prefer to leave their partners and match together.

First, note that any stable matching will exhaust all possible matches x, y such that $x = y$, thus exhausting all mass common to μ and ν. To understand how the rest of the population is matched, we establish a structural property that must be satisfied by stable matchings. For any two real numbers z_1 and z_2, let $O(z_1, z_2)$ be the smallest circle containing the points $(z_1, 0)$ and $(z_2, 0)$ in $\mathbb{R}^2$.

Definition 1. A matching π satisfies **no-crossing** if, for any two pairs (x, y) and (x', y') in the support of π, the circles $O(x, y)$ and $O(x', y')$ do not intersect unless $x = x'$ or $y = y'$.



(a) blocked by (x', y) (b) blocked by (x, y')

(c) blocked by (x', y) (d) blocked by (x, y'), (x', y)

Figure 1: Forbidden patterns in stable matchings.

Lemma 1. *Any stable matching satisfies no-crossing.*

Proof. Suppose that (x, y) and (x', y') are in the support of π where $x \neq x'$, $y \neq y'$ and the pair of circles $O(x, y)$ and $O(x', y')$ intersect. There are, up to symmetry, two cases, depicted in Figures 1a and 1b. First, $x < x' < y < y'$ in which case $|x' - y| < |x - y|$ and $|x' - y| < |x' - y'|$, and so π is blocked by (x', y). Second, we could have $x < y' < y < x'$. Now $|x - y'| < |x - y|$ and $|x' - y| < |x' - y'|$. If π were stable we would have $|x - y| \leq |x' - y|$ and $|x' - y'| \leq |x - y'|$ however by transitivity $|x - y| < |x' - y'|$ and $|x' - y'| < |x - y|$. Thus π is not stable. $\square$

As indicated by Figures 1c and 1d, stable matchings have more structure than just no-crossing. However, the no-crossing property itself is strong enough to reduce finding a stable matching to analysis of a parametric family. We illustrate this in an example. Suppose μ has density 1 on the interval $[-2, -1]$ and density 2 on the half-open interval $(0, 1]$. Likewise, assume that ν has density 1 on the the half-open intervals $(-1, 0]$ and $(1, 3]$. The difference $\rho = \mu - \nu$ is shown in Figure 2.

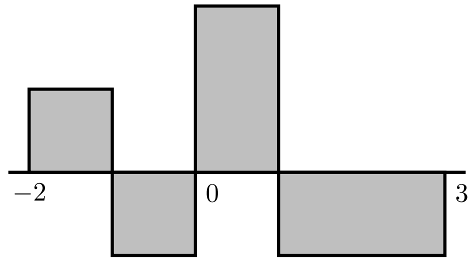


Figure 2: The signed measure ρ.

The set of matchings in this example which satisfy no-crossing are described by a single parameter $\theta \in [0, 1]$ that describes what portion of the agents in the region $[-2, -1]$ are matched with agents to the right non-locally. Three values of θ are shown in Figure 3.

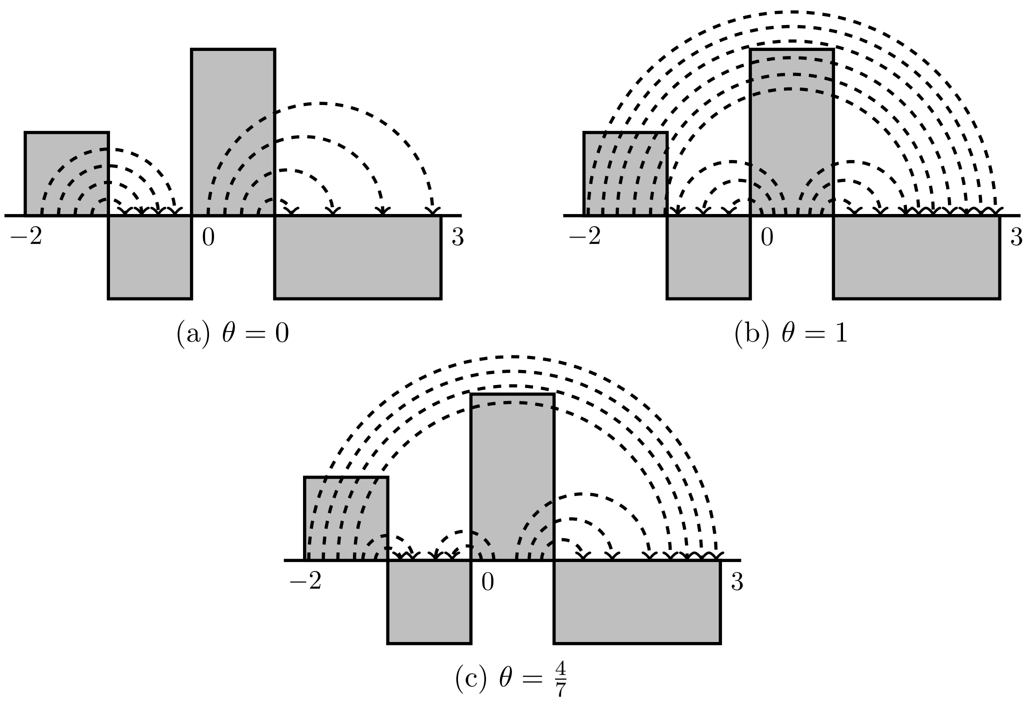


Figure 3: The matchings which satisfy no-crossing for three values of θ.

One can verify that the only value of θ which results in a stable matching is $4/7$. In this stable matching, each type x is matched with $h(x)$ where

$$h(x) = \begin{cases} 1 - x, & x \in [-2,\ -1 - \frac{3}{7}] \\ -2 - x, & x \in (1 - \frac{3}{7},\ -1] \\ -2x, & x \in (0,\ \frac{2}{7}] \\ 3 - 2x, & x \in (\frac{2}{7},\ 1] \end{cases}$$

as depicted by the half-circles in Figure 3c.

The no-crossing property has played a major role in the theory of one-dimensional optimal transport with concave costs developed by McCann (1999).[1] For ρ from Figure 2, the set of no-crossing matchings is one-parametric, thus making it easy to construct a stable matching. McCann (1999) (Theorem 3.11) characterized all non-crossing matchings for non-atomic μ and ν. The set of no-crossing matchings consists from a number of parametric families, but the number of these families grows exponentially with the number of times that $\rho = \mu - \nu$ changes sign.[2] We show that one can avoid the exploding search space by combining insights about no-crossing from the transportation literature with properties specific to stable matchings.

Proposition 1. *If $\rho = \mu - \nu$ is a non-atomic measure changing sign a finite number of times, then there is a unique stable matching π that can be obtained via an algorithm described in Appendix A. For each $x \in X$, there are at most two distinct types $y, y' \in Y$ such that (x, y) and (x, y') are in the support of π. If μ and ν have piecewise-constant density with at most m intervals of constancy, the algorithm runs in time of the order of m^2.*

The proof contained in Appendix A is based on the following idea. For a market represented by $\rho = \mu - \nu$, we can identify a submarket such that agents in the submarket are to be matched with each other no matter what the rest of the population is, the match within the submarket is pinned down uniquely by no-crossing, and the residual market ρ' has one sign change fewer. Matching such simple submarkets sequentially, we construct a stable matching for the original market in the number of steps bounded by the number of sign changes of ρ. For piecewise-constant density,

[1] According to Villani (2009), the no-crossing property dates back to papers by Monge (who initiated the study of optimal transportation); see also the discussion by Boerma, Tsyvinski, Wang, and Zhang (2023).

[2] Roughly speaking, the parametrization is as follows. The real line gets partitioned into intervals where $\rho \geq 0$ or $\rho \leq 0$. For each interval of positivity, we specify what fraction of agents are matched with those to the right non-locally (θ in our example) and what negativity intervals these matches are from. As a result, for each interval of positivity, we get a single parameter augmented with combinatorial data. This combinatorial data grows exponentially with the number of intervals.

the number of sign changes is bounded by m and constructing each simple submarket requires a number of operations proportional to m, hence the quadratic total runtime.

That no-crossing plays an important role in stable matching and optimal transport is no coincidence. In the next section, we establish a deeper connection between stability and transport by showing that approximately stable matchings may be obtained as a solution to certain optimal transport problems, and that exactly stable matchings result as the limit of solutions to a sequence of such optimal transport problems. While the no-crossing condition is specific to matching in $\mathbb{R}$, the connection to optimal transport holds for general problems with aligned preferences. After developing this connection, we will return to matching in $\mathbb{R}^d$, including $d = 1$, in Section 4.

3 Stability, fairness, and optimal transport

Consider a general matching market with two sides X and Y. We assume that X and Y are separable metric spaces endowed with their Borel sigma-algebras. For a space Z, we denote by $\mathcal{M}_+(Z)$ the set of positive measures on Z with a finite total mass. The distributions of agents' types over X and Y is represented by $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$. For simplicity, we assume that the market is balanced, i.e., $\mu(X) = \nu(Y)$.[3]

The classical marriage-market model of Gale and Shapley (1962) corresponds to the case when X and Y are finite. More generally, any market with atomic μ and ν having a finite number of atoms, all of the same mass, can be reduced to the classical model.

A distribution $\pi \in \mathcal{M}_+(X \times Y)$ is a **matching** if it has μ and ν as marginal distributions on, respectively, X and Y. We denote by $\Pi(\mu, \nu)$ the set of all matchings.

Stability. To make the similarity between the classical stability notion and the one in our paper more apparent, we first give a definition without assuming that preferences are aligned, and then refine the definitions in the aligned case.

If agents with types $x \in X$ and $y \in Y$ are matched, they enjoy utilities $u(x, y)$ and $v(x, y)$, respectively, where $u, v \colon X \times Y \to \mathbb{R}$ are measurable. Greinecker and Kah (2021) proposed a way to extend the classical notion of stability from Gale and

[3]It is, however, easy to accommodate unbalanced markets by adding "dummy" agents as in Example 5 from Section 5.

Shapley (1962) to general spaces X and Y. Their definition adapts to our setting as follows.

Definition 2. A matching π is **ε-stable** with a parameter $\varepsilon \geq 0$ if for $\pi \times \pi$-almost all pairs (x_1, y_1) and (x_2, y_2) at least one of the following two inequalities holds

$$\begin{aligned} u(x_1, y_1) - u(x_1, y_2) &\geq -\varepsilon \\ v(x_2, y_2) - v(x_1, y_2) &\geq -\varepsilon \end{aligned} \tag{1}$$

If both inequalities are violated, (x_1, y_2) is called an **ε-blocking pair**. Definition 2 effectively says that, in a random submarket consisting of two couples (x_1, y_1) and (x_2, y_2) sampled from π independently, there is no ε-blocking pair almost surely. For $\varepsilon = 0$, we will refer to ε-stable matchings as **stable**.

If u and v are continuous, Definition 2 becomes equivalent to its more intuitive pointwise version resembling the classical definition of Gale and Shapley (1962). Recall that the support of π, denoted by $\mathrm{supp}(\pi)$, is the minimal closed set of full measure.

Lemma 2. *Let u and v be continuous utility functions. A matching π is ε-stable if and only if for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$, at least one of the two inequalities (1) holds.*

Proof. One direction is straightforward: the pointwise property implies the almost-everywhere property. We prove the opposite direction. Let π be an ε-stable matching of populations $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$ with continuous utilities u and v. Our goal is to show that, for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$, at least one of the two inequalities (1) holds. Towards a contradiction, suppose that there are $(x_1^*, y_1^*), (x_2^*, y_2^*) \in \mathrm{supp}(\pi)$ such that both inequalities are violated. By continuity of u and v, we can find open neighborhoods $U_1 \subseteq X \times Y$ of (x_1^*, y_1^*) and $U_2 \subseteq X \times Y$ of (x_2^*, y_2^*) such that the inequalities are violated for all $(x_1, y_1), (x_2, y_2) \in U_1 \times U_2$. Since $(x_1^*, y_1^*), (x_2^*, y_2^*) \in \mathrm{supp}(\pi)$, we have $\pi(U_1) > 0$ and $\pi(U_2) > 0$. Thus $U_1 \times U_2$ has a positive $\pi \times \pi$-measure, which contradicts the ε-stability of π. We conclude that ε-stability of π implies that at least one of the two inequalities (1) holds for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$. □

Henceforth, we focus on the special case of markets with aligned preferences.

Definition 3. We say that the preferences of agents in X and Y are **aligned** if $u = v$, i.e., the utility of a match is the same for both sides of the market.[4]

[4]We further discuss aligned preferences in Section 6. We discuss how aligned preferences arise naturally in a model with transfers and lack of commitment in Section 6.1, and we provide sufficient conditions for ordinal preferences to be aligned in Section 6.2.

The stability condition for markets with aligned preferences boils down to a single inequality. A matching π is stable, if for $\pi \times \pi$-almost all pairs (x_1, y_1) and (x_2, y_2), the following inequality holds

$$u(x_1, y_2) \leq \max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon. \tag{2}$$

Remark. When agents' types are identified with points on the real line, $X = Y = \mathbb{R}$ and $u(x, y) = -|x - y|$, the condition of stability (2) with $\varepsilon = 0$ reduces to the notion discussed in Section 2. The special case of the real line also illustrates that aligned preferences do not mean common preferences: two distinct types of agents, x and x' in X, will generally disagree on how they rank agents in Y.

Fairness. We now turn to a notion of fairness. To each matching π we assign a number $U_{\min}(\pi)$ equal to the minimal utility of a couple matched under π:

$$U_{\min}(\pi) = \min_{(x,y)\in \mathrm{supp}[\pi]} u(x, y)$$

and denote the best $U_{\min}(\pi)$ over all matchings by

$$U^*_{\min}(\mu, \nu) = \max_{\pi\in\Pi(\mu,\nu)} U_{\min}(\pi).$$

This quantity is the egalitarian lower bound in the spirit of Rawls (1971): it is the highest utility level feasible for every couple in the population simultaneously.

Under the assumption that X and Y are compact and u is continuous, $U_{\min}$ and $U^*_{\min}$ are well-defined. Indeed, the support of π is a closed set, a closed subset of a compact space is compact, and hence, the minimum is attained; the fact that the maximum is attained is established below in Corollary 1 even for non-compact X and Y. If u is discontinuous or X, Y are not compact, we replace the minimum with the essential infimum and the maximum with the supremum.

Definition 4. A matching $\pi \in \Pi(\mu, \nu)$ is ε-**egalitarian** if there is a subset $S \subset X \times Y$ with $\pi(S) \geq (1 - \varepsilon) \cdot \pi(X \times Y)$ such that

$$u(x, y) \geq U^*_{\min}(\mu, \nu) - \varepsilon \qquad \text{for any} \quad (x, y) \in S.$$

In words, for a large fraction of the whole population, the utility resulting from a ε-egalitarian matching almost satisfies the egalitarian lower bound. For $\varepsilon = 0$, we will refer to ε-egalitarian matchings as **egalitarian**.

Optimal transport. The canonical optimal transportation problem involves a measurable cost function $c\colon X \times Y \to \mathbb{R}$, and marginal measures $\mu \in M_+(X)$ and $\nu \in M_+(Y)$ with $\mu(X) = \nu(Y)$. The goal is to find a matching $\pi \in \Pi(\mu, \nu)$ that minimizes the total cost:

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x, y)\ \mathrm{d}\pi(x, y). \tag{3}$$

In this context, a matching $\pi \in \Pi(\mu, \nu)$ is often referred to as a transportation plan. We refer to Villani (2009) and Galichon (2018) for the basic theory of optimal transport.

Solutions to optimal transport problems exhibit a certain "monotonicity" property. Given $c\colon X \times Y \to \mathbb{R}$, a set $\Gamma \subset X \times Y$ is called c**-cyclic monotone** if

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{i+1}) \tag{4}$$

for all $n \geq 2$ and pairs of points $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ with the convention $y_{n+1} = y_1$. Solutions to an optimal transportation problem are known to be supported on a c-cyclic monotone set.[5]

We shall see that the stability condition (2) arises naturally from the c-cyclic monotonicity property of an auxiliary optimal transportation problem.

3.1 The main result

We now describe the link between stability, fairness, and optimal transport. For a matching market with aligned preferences represented by a utility function $u\colon X \times Y \to \mathbb{R}$, we consider a parametric class of cost functions given by

$$c_\alpha(x, y) = \frac{1 - \exp(\alpha \cdot u(x, y))}{\alpha}. \tag{5}$$

The tension between stability and fairness is a general phenomenon. We explore it by recasting both properties in optimal transportation terms. Our main result, Theorem 1, shows how stability and fairness result from taking opposite extreme values of the parameter α.

Theorem 1. *Let π^* be a solution to the optimal transport problem*

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c_\alpha(x, y)\ \mathrm{d}\pi(x, y).$$

[5] See Kausamo, De Pascale, and Wyczesany (2023) for a survey of related results.

1. *If* $\alpha > 0$*, then* π^* *is* ε*-stable, with* $\varepsilon = \frac{\ln 2}{\alpha}$.

2. *If* $\alpha < 0$*, then* π^* *is* ε*-egalitarian, with* $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$.

Taking the limit $\alpha \to 0$ in (5) suggests that the cost c_α for $\alpha = 0$ must be defined by $c_0(x, y) = -u(x, y)$. Hence, for $\alpha = 0$, the transportation problem in Theorem 1 corresponds to maximizing the utilitarian social welfare

$$W(\pi) = \int_{X\times Y} u(x, y)\, \mathrm{d}\pi(x, y) \tag{6}$$

Thus, by changing α gradually from $-\infty$ to $+\infty$, we interpolate between *fairness*, *welfare*, and *stability* objectives. Theorem 2 below connects stability and utilitarian welfare maximization.

Remark (Existence of optimal matchings)*.* To guarantee the existence of a solution in Theorem 1, it is enough to assume that u is continuous and bounded. No compactness assumptions on spaces X and Y are needed. Indeed, an optimal transportation plan π in (3) is known to exist under the assumption that the cost c is lower semicontinuous and bounded from below when X and Y are Polish spaces.[6]

The proof of Theorem 1 is in Appendix B. Proving part 1 requires a result by Beiglböck, Goldstern, Maresch, and Schachermayer (2009), and boils down to checking that the cyclic monotonicity condition (4) for the cost function c_α from (5) implies the ε-stability condition (2). To prove part 2, we show that the contribution of low-utility couples to the transportation objective becomes dramatic for negative α, and thus π^* cannot place substantial weight on such couples.

Theorem 1 implies the existence of stable and egalitarian matchings under a continuity assumption. The idea is to construct these matchings as the weak limits of solutions to the optimal transportation problem from Theorem 1 for $\alpha \to \pm\infty$.

Denote by $\Pi^u_{+\infty}(\mu, \nu)$ the set of matchings π that can be obtained as the weak limit $\pi = \lim_{n\to+\infty} \pi_{\alpha_n}$ of sequences of solutions π_{α_n} to the transportation problem (3), with the cost c_{α_n} from (5), for some sequence of $\alpha_n \to +\infty$. Similarly define $\Pi^u_{-\infty}(\mu, \nu)$ to be the weak limits for some sequence $\alpha_n \to -\infty$.

Corollary 1. *For continuous and bounded utility* u*, the sets* $\Pi^u_{+\infty}(\mu, \nu)$ *and* $\Pi^u_{-\infty}(\mu, \nu)$ *are non-empty, convex, and weakly closed. All matchings in* $\Pi^u_{+\infty}(\mu, \nu)$ *are stable and all matchings in* $\Pi^u_{-\infty}(\mu, \nu)$ *are egalitarian.*

[6]See for instance Theorem 4.1 in Villani (2009).

Corollary 1 provides the existence of stable and egalitarian matchings, and a computationally tractable way of finding them, as a limit of solutions to a sequence of linear programs.[7] The existence of stable matchings for markets with aligned preferences gives a special case of the general existence result proven in Greinecker and Kah (2021). In Appendix B, we check that Corollary 1 follows from Theorem 1.

Remark. For finite markets, it is enough to take α large enough to obtain exact stability. Indeed, if X and Y are finite, denote by δ the minimal change in utility that an agent can experience if they change a partner:

$$\delta = \min\left\{\min_{x,\ y\neq y'} |u(x,y)-u(x,y')|, \quad \min_{y,\ x\neq x'} |u(x,y)-u(x',y)|\right\}.$$

Then any ε-stable matching with $\varepsilon < \delta$ is automatically stable. Combining this observation with Theorem 1, we conclude that the solution of the optimal transportation problem with $\alpha > \delta/\ln 2$ is stable.

The conclusions of Theorem 1 are not specific to the exponential objective in c_α. As we demonstrate in Appendix B, it is enough to take

$$c(x,y) = -h\left(u(x,y)\right), \tag{7}$$

where h is a monotone-increasing function with a large logarithmic derivative. Indeed, we show that if $h>0$ is monotone increasing and

$$\frac{h'(t)}{h(t)} \geq \alpha \tag{8}$$

for some positive α and t in the range of u, then any solution to the transportation problem with cost (7) is ε-stable with $\varepsilon = \frac{\ln 2}{\alpha}$. Similarly, if $h<0$ is monotone increasing and satisfies (8) with $\alpha < 0$, then the solution to the transportation problem is ε-egalitarian with $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$.

The set $\Pi^u_{+\infty}(\mu,\nu)$ of stable matchings in Corollary 1 is convex, but the set of all stable matchings may not have this property. Indeed, the stability condition (2) is not convex. Thus Corollary 1 constructs solutions to a non-convex problem by approximating it via convex problems. In consequence, it provides a convex selection from the set of all stable matchings; but there can exist stable matchings that do

[7]The set of stable matchings does not change after a monotone parametrization of the utility function. Therefore, Corollary 1 implies the existence of stable matchings without boundedness assumption on utility u. Indeed, it is enough to replace u with a reparametrized utility $\tilde{u}(x,y) = \arctan(u(x,y))$.

not correspond to (limits of) solutions to optimal transportation problems (3) with cost (5). This happens when there is a positive measure of agents indifferent between several potential matches, as in the following example.

Example 1 (Non-uniqueness of stable matchings). Consider $X = Y = \mathbb{R}$, the utility function $u(x, y) = -|x-y|$, and two-point populations: μ places equal weight on $x_1 = 1/3$ and $x_2 = 1$ and ν places equal weight on $y_1 = 0$ and $y_2 = 2/3$.

One can check that the following two matchings are stable.

- The deterministic matching, where x_i is matched with y_i, i.e.,

$$\pi = \frac{1}{2}\delta_{(x_1,y_1)} + \frac{1}{2}\delta_{(x_2,y_2)},$$

  where $\delta_{(x,y)}$ denoted the unit mass at a point (x, y).

- The random matching, where each x_i is matched with each y_j equally likely, i.e.,

$$\pi' = \frac{1}{4}\delta_{(x_1,y_1)} + \frac{1}{4}\delta_{(x_2,y_2)} + \frac{1}{4}\delta_{(x_1,y_2)} + \frac{1}{4}\delta_{(x_2,y_1)}.$$

Interestingly, π and π' have different welfare consequences. The matching π is egalitarian, and all agents are matched with a partner $1/3$ away. For π', all couples except (x_2, y_1) are $1/3$ away from their partners and $|x_2-y_1| = 1$. We conclude that the stable matching π Pareto dominates another stable matching π'. The set $\Pi^u_{+\infty}(\mu, \nu)$ in this example consists of the matching π only.[8] Indeed, Pareto-dominated matching π' cannot result from optimizing an objective strictly monotone in agents' utilities.

3.2 Welfare and fairness of stable matchings

Theorem 1 shows that fairness and stability correspond to the two extremes of the α-spectrum, $+\infty$ and $-\infty$, respectively. This result suggests that stability may fail to provide any individual fairness guarantee. As we show here, even though the loss in fairness and welfare may be dramatic, there is an upper bound on it.

To get the intuition of why losses in welfare and fairness are bounded, we reinterpret the definition of stability. Consider a continuous utility u and a matching $\pi \in \mathcal{M}_+(X \times Y)$ with marginals μ and ν. By Lemma 2, matching π is ε-stable

[8] We conjecture that the set of stable matchings is a singleton under a condition ruling out large groups of indifferent agents. For matchings in $\mathbb{R}$, this conjecture holds (Proposition 1); see also Section 4 for $\mathbb{R}^d$ with $d \geq 2$.

with $\varepsilon \geq 0$ if for all (x_1, y_1) and (x_2, y_2) in the support of π,

$$u(x_1, y_2) \leq \max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon. \tag{9}$$

The couple (x_1, y_2) can serve as a generic element of the product space $X \times Y$. This observation allows us to interpret formula (9) as follows:

> A matching π is ε-stable if, for a generic couple (x, y), the utility of at least one of the partners of x or y in π gives at least the utility of (x, y) minus ε.

In other words, the utility of a hypothetical couple (x, y) provides a lower bound to the utility of at least one of the partners in a stable match. In particular, the latter utilities cannot be too small simultaneously. This observation bounds the extent to which welfare and fairness are to be sacrificed in order to get stability.

Recall that utilitarian welfare $W(\pi)$ of a matching π is given by (6) and let $W^*(\mu, \nu)$ be the maximal welfare over all $\pi \in \Pi(\mu, \nu)$. The following result is proved in Appendix B.

Theorem 2. *Consider a market with unit populations ($\mu(X) = \nu(Y) = 1$) and aligned preferences represented by a bounded and continuous utility*[9] $u \geq 0$. *Then any ε-stable matching π satisfies*

$$W(\pi) \geq \frac{1}{2}\left(W^*(\mu, \nu) - \varepsilon\right).$$

Moreover, π is ε'-egalitarian with

$$\varepsilon' = \max\left\{\frac{1}{2},\, \varepsilon\right\}.$$

In particular, we obtain that any stable matching guarantees $1/2$ of the optimal welfare and is $1/2$-egalitarian. For markets with a finite number of agents, we recover the welfare guarantee obtained by Anshelevich, Das, and Naamad (2013). The bounds in Theorem 2 are conservative as they target ε-stable matchings that have the lowest welfare or that are least egalitarian.

[9] If a bounded continuous utility u does not satisfy the non-negativity requirement, one can apply Theorem 2 to $\tilde{u}(x, y) = u(x, y) - \inf_{x', y'} u(x', y')$. For example, this argument applies to the distance-based utility of Section 2 if μ and ν have bounded support.

4 Matching in $\mathbb{R}^d$

We now consider the case of multidimensional type spaces, generalizing the analysis from Section 2. The two sides of the market X and Y are now subsets of $\mathbb{R}^d$ with $d \geq 1$. Each agent prefers to be matched to someone as close as possible, which is captured by the following distance-based utility function

$$u(x, y) = -\|x - y\| = -\sqrt{\sum_{i=1}^{d}(x_i - y_i)^2}.$$

The connection between stability, fairness, welfare, and optimal transport established in Theorem 1 is not sensitive to the dimension d. However, in the case of $d = 1$ discussed in Section 2, further structure can be observed.

Given a cost function of the form $c(x, y) = h(|x - y|)$ where h is strictly convex, it is well-known that the solution to (3) is simple and has little dependence on h. In particular, for any such cost function, the assortative matching is uniquely optimal. Recall that the assortative matching is supported on the curve given by the equation $F_\mu(x) = F_\nu(y)$ where F_μ and F_ν are the cumulative distribution functions of μ and ν, respectively.

In the case $d = 1$, the cost function from (5) specializes to

$$c_\alpha(x, y) = \frac{1 - \exp(-\alpha \cdot |x - y|)}{\alpha}. \tag{10}$$

which is a strictly convex function of $|x - y|$ so that the assortative matching is the unique solution to the optimal transport problem for any $\alpha < 0$. By taking weak limits, and noting that as $\alpha \to 0$, (3) approaches the utilitarian objective, applying Theorem 1, we see that the assortative matching is fair and also attains the maximal welfare level.

Corollary 2. *For $X = Y = \mathbb{R}$ and non-atomic μ, ν with bounded support, the assortative matching is egalitarian and is welfare-maximizing.*

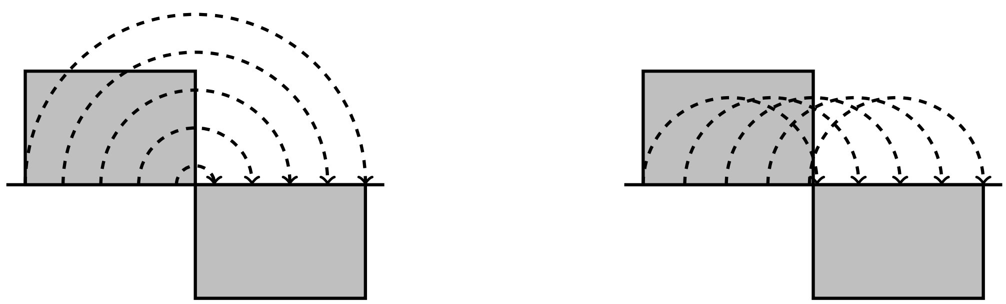
While the assortative matching is both fair and efficient, outside of some degenerate cases it is not stable. To see why, note that whenever there are points $x < x'$ and $y < y'$ matched assortatively, the no-crossing condition cannot be satisfied unless x, x' and y, y' interlace. This observation implies that the assortative matching is never stable for non-atomic μ and ν. Indeed, the following example shows that stability can induce substantial inequality.

Example 2 (Fairness-stability tension). Let $X = Y = \mathbb{R}$, like the example we discussed in Section 2. Let μ be the uniform distribution on $[-1, 0]$ and ν be the uniform distribution on $[0, 1]$, and so $\rho = \mu - \nu$ has density

$$\sigma(t) = \begin{cases} 1, & t \in [-1, 0] \\ -1, & t \in [0, 1] \end{cases}$$

See Figure 4 for an illustration. The unique stable matching π is demonstrated in Figure 4a by the half-circles. Types $-z \in [-1, 0]$ are matched with types $z \in [0, 1]$. Notice that the stable matching induces substantial inequality in the value of the matches. The agents with types close to zero are paired to agents who have similar types to them, while the agents far from zero are matched with agents who have very different types.

Figure 4b, by contrast, exhibits the egalitarian matching. All agents are given a partner whose type is distance one away from their own, and thus all agents obtain the same utility from their match.



(a) Stable

(b) Egalitarian

Figure 4: Two solutions to the matching problem.

Inequality in a stable matching becomes even more stark if the two sides of the market have mass in common. Consider new populations $\mu' = \mu + \lambda$ and $\nu' = \nu + \lambda$, where λ is the uniform distribution on $[-1, 1]$. A stable matching for (μ', ν') matches as many agents x as possible with their ideal partners $y = x$, thus exhausting λ; see Section 2. Then, the residual population (μ, ν) is matched as in Figure 4a. As a result, half of the agents of type $x \in [-1, 0]$ are matched with their ideal partner $y = x$ and half, with $y = -x$. For $x = -1$, this means obtaining either the best or the worst possible partner. We conclude that stable matchings may result in envy within the same type.

The following example illustrates that there is no stability-welfare tension in markets with at most three regions of under/oversupply, but the tension can be present in general.

Example 3 (Welfare-stability tension). Let $X = Y = \mathbb{R}$. Consider a market (μ, ν) from the previous example. One can easily check that the stable matching (Figure 4a) and the egalitarian one (Figure 4b) have the same welfare. Since egalitarian matching is also welfare-maximizing (Corollary 2), so is the stable one. In other words, stable matching results in inequality but not in welfare loss.

More generally, the stability-welfare tension is absent if $\rho = \mu - \nu$ changes sign at most twice (in our example, ρ changes sign once at $t = 0$). Indeed, McCann (1999) showed that a transportation problem on $\mathbb{R}$ with a concave objective admits a no-crossing solution. Moreover, a matching satisfying no-crossing is unique if ρ changes sign at most two times as in Figure 5. Thus, the same no-crossing matching solves the transportation problem with the cost c_α from (10) with any $\alpha \in [0, +\infty)$. Letting α go to $+\infty$, we conclude this matching is stable. Plugging $\alpha = 0$, we get that it is also welfare-maximizing.

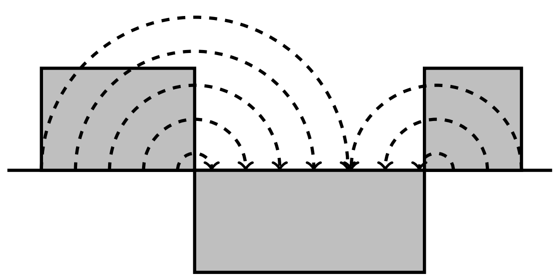


Figure 5: For $\mu - \nu$ changing sign two times, no-crossing matching is unique and thus is simultaneously stable and welfare-maximizing.

For ρ changing sign three or more times, stability may come at welfare cost. For example, consider the market from Figure 3, where ρ changes sign three times. Recall that the set of all no-crossing matchings is parameterized with $\theta \in [0, 1]$, and stable matching corresponds to $\theta = 4/7$. It has welfare of approximately -6.02, while the optimal welfare of -2 is attained at $\theta = 1$. Figure 6 illustrates the dependence of optimal θ on α in the optimal transportation problem with cost c_α.

Consider now $\mathbb{R}^d$ with $d > 1$. There may be a tension between individual and collective welfare, absent in the one-dimensional case.

Example 4 (Fairness-welfare tension). Let $X = Y = \mathbb{R}^2$. Let $B_\delta(x, y)$ denote the ball of radius δ centered at the point (x, y). Let μ have density 1 on $B_\delta(0, 0) \cup$

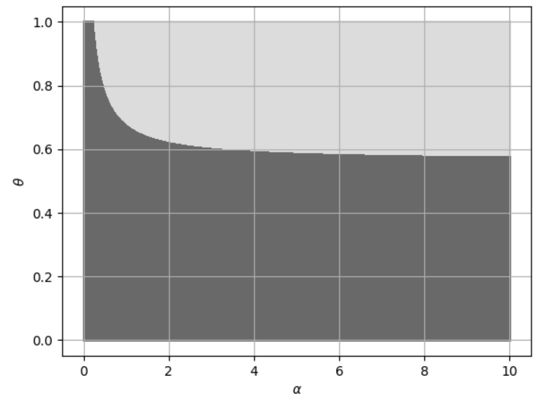


Figure 6: Dependence of θ on α in the optimal transport problem with cost c_α and distribution from Figure 3.

$B_\delta(0,1)$ and ν have density 1 on $B_\delta(0,0) \cup B_\delta(1,0)$. In the welfare maximizing matching, the agents X and Y in $B_\delta(0,0)$ are matched together and the agents in X from $B_\delta(0,1)$ are matched with those from $B_\delta(1,0)$ in Y. This gives an average match distance of $\frac{1}{2}\sqrt{2}$ and the longest distance between matched agents is no less than $\sqrt{2}$. By contrast, the matching which pairs the X agents in $B_\delta(0,1)$ with the Y agents in $B_\delta(0,0)$ and the X agents in $B_\delta(0,0)$ with the Y agents in $B_\delta(0,1)$ gives an average match distance of 1 but has the longest distance between matched pairs no more than $1+2\delta$.

Concave and convex optimal transport are not prone to explicit solutions beyond $d=1$, but some structural results can be obtained for $d>1$. Now the cost function is

$$c_\alpha(x,y) = \frac{1-\exp(-\alpha \cdot \|x-y\|)}{\alpha}. \tag{11}$$

Say that a matching π is deterministic if it is supported on the graph of an invertible map $s\colon X \to Y$. A matching π is diagonal if $x=y$ for π-almost all couples (x,y). Off-the-shelf results from optimal transport (Theorems 1.2 and 1.4 of Gangbo and McCann (1996)) address multidimensional optimal transport with concave or convex costs. Their results imply the following corollary.

Corollary 3. *Consider a market with $X=Y=\mathbb{R}^d$, compactly supported μ and ν absolutely continuous with respect to the Lebesgue measure, and utility $u(x,y) = -\|x-y\|$. The optimal transportation problem with cost (11) admits an optimal matching π^* and the following assertions hold:*

- *For $\alpha < 0$, π^* is unique and deterministic;*
- *For $\alpha > 0$, π^* is unique and is a convex combination of a deterministic matching and a diagonal one.*
- *There is a deterministic matching π^* which maximizes welfare.*

We note that the condition of absolute continuity with respect to the Lebesgue measure can be weakened to the requirement that μ and ν place zero mass on $(d-1)$-dimensional surfaces. Moreover, by imposing this requirement on μ only, we would obtain a version of Corollary 3 where the conclusion that π^* is deterministic is replaced with the Monge property (π^* is supported on a graph of $s\colon X \to Y$, but s may not be invertible).

The egalitarian matching obtained as the limit $\alpha \to -\infty$ corresponds to the so-called L^∞-transportation problem; see, for example, Champion, De Pascale, and Juutinen (2008); Brizzi, De Pascale, and Kausamo (2023). The transportation literature has mostly focused on questions of existence, uniqueness, and an appropriate notion of cyclic monotonicity. We note that, for a distance-based cost, the L^∞-problem is only interesting in the multi-dimensional case since, on the real line, it corresponds to the assortative matching by Corollary 2. To the best of our knowledge, the limit problem for $\alpha \to +\infty$ has not been studied prior to our paper.

4.1 Discussion of applications to school choice

In school choice, student preferences and school priorities can depend on many factors and, in general, will not be aligned. However, distance is a key component of student preferences over schools and of school priorities. For example, Walters (2018) estimates student preferences over charter schools in Boston, and finds that distance traveled is a key driver of student preferences. Other papers find generally similar results (Laverde, 2022; Dinerstein and Smith, 2021; Agostinelli, Luflade, Martellini, et al., 2021). Similarly, school districts typically give priority to local students. For example, in Boston, priority is given to students with siblings who attend the same school and to students who live within a mile of the school (Angrist, Gray-Lobe, Idoux, and Pathak, 2022).

Stable matching in $\mathbb{R}^d$ provides a setting to analyze this component of school choice in isolation. By recasting stable matching as the solution to a particular planner's problem, we can more easily compare the objective implicit in stability with traditional welfare measures. Theorem 1 shows that stability comes from maximizing an objective which places very high weight on matching students to nearby schools, but which places vanishing value on the externalities that these matches can cause to students who will have to travel farther as a result. The stability objective is at the opposite extreme of the commonly used welfare measures in other areas of economics. For example, in public finance, it is common to model the planner as maximizing the average of a concave function—equivalently, minimizing a convex function—of utilities (Saez, 2001). The concave welfare measure reflects a planner's preference for fairness. When α is large, however, (11) corresponds to a planner who loves inequality.

This discussion has two implications for school choice. First, stable matching can generate long average travel times; though Theorem 2 suggests a limit on the severity of this phenomenon. Second, stable matching can entail very long travel times for some students, and short travel times for others, even when these students are ex-ante identical.

Consistent with our findings, Angrist, Gray-Lobe, Idoux, and Pathak (2022) document that, since adopting the deferred acceptance algorithm, both Boston and New York have seen substantial increases in municipal expenditures on transportation and a significant fraction of students travel long distances to school. For example, in the first decade after implementing the deferred acceptance algorithm, Boston saw more than a 50% increase in per-pupil expenditures on travel. They estimate that switching to neighborhood assignment would decrease average travel times by as much as 17 minutes in New York and 13 minutes in Boston. By contrast, Angrist, Gray-Lobe, Idoux, and Pathak (2022) find modest effects on test scores and college attendance.

5 Multi-partner matching

The two-sided matching problem can be generalized to the formation of stable k-tuples. There are k sets of agents, $X_1, \ldots, X_k$, and functions $u_i \colon X_1 \times \ldots \times X_k \to \mathbb{R}$, $i = 1, \ldots, k$. The interpretation is that each tuple $(x_1, \ldots, x_k)$ generates a utility $u_i(x_1, \ldots, x_k)$ to agent x_i.

First, suppose that all X_i are finite and have the same cardinality. A **matching** is a collection μ of k-tuples $(x_1, \ldots, x_k)$ such that each x_i appears in exactly one tuple. We write $\mu(x_i)$ for the tuple that x_i belongs to. A k-tuple $(x_1, \ldots, x_k)$ **blocks** a matching μ if $u_i(x_1, \ldots, x_k) > u_i(\mu(x_i))$ for all $i = 1, \ldots, k$. A matching is **stable** if there is no blocking tuple.

The multi-partner matching problem is well known to be intractable. Even with three partners ($k = 3$) and additively separable preferences $u_i(x) = \sum_{j \neq i} v_i(x_i, x_j)$, there may not exist a stable matching (Alkan, 1988). Moreover, deciding whether a stable matching exists is NP-hard (Ng and Hirschberg, 1991). The case of aligned preferences—utilities $u_i = u$ for all agents i—turns out to be an exception: a stable matching is guaranteed to exist, and the theory we have developed for two-sided markets readily extend to multi-partner matching.

We turn to a generalization of the model introduced in Section 3. The sets X_i are assumed to be Polish spaces, each X_i is endowed with a measure $\mu_i \in \mathcal{M}_+(X_i)$ representing the distribution of agents in X_i, and the total mass $\mu_i(X_i)$ is the same for all i. A matching π is a positive measure on $X = X_1 \times \ldots \times X_k$ with marginals μ_i on X_i, and $\Pi(\mu_1, \ldots, \mu_k)$ denotes the set of all matchings. A measurable function $u\colon X \to \mathbb{R}$ is the agents' common utility: each agent in a k-tuple $x = (x_1, \ldots, x_k)$ enjoys utility of $u(x)$.

For a matching π, let $\pi^{\times k}$ denote the product measure on X^k. So $\pi^{\times k}$ corresponds to k independent draws of a tuple from X. A matching π is ε-**stable** if, for $\pi^{\times k}$-almost all collections of tuples $x^1, \ldots, x^k$,

$$u(x_1^1, \ldots, x_k^k) \leq \max_{i=1,\ldots,k} u(x^i) + \varepsilon.$$

The notions of an ε-egalitarian matching and that of the optimal welfare W^* also straightforwardly extend to the multi-partner case.

Given a cost function $c\colon X \to \mathbb{R}$, the k-marginal optimal transport problem is to find a matching π minimizing the cost

$$\min_{\pi \in \Pi(\mu_1, \ldots, \mu_n)} \int_X c(x) \; \mathrm{d}\pi(x). \tag{12}$$

As in the two-sided case, we consider the following cost function

$$c_\alpha(x) = \frac{1 - \exp(\alpha \cdot u(x))}{\alpha}. \tag{13}$$

Theorems 1 and 2 generalize to the k-partner case as follows.

Theorem 3. *Consider a k-partner matching market with aligned preferences and bounded continuous utility $u\colon X_1 \times \ldots \times X_k \to \mathbb{R}_+$. Let π^* be a solution to the optimal transportation problem (12) with the cost c_α from (13). The following assertions hold:*

1. *If $\alpha > 0$, then π^* is ε-stable with $\varepsilon = \frac{\ln k}{\alpha}$.*

2. *If $\alpha < 0$, then π^* is ε-egalitarian with $\varepsilon = \frac{\max\{1, \ln |\alpha|\}}{|\alpha|}$.*

3. *For unit populations ($\mu_i(X_i) = 1$ for all $i = 1, \ldots, k$), any ε-stable matching π satisfies $W(\pi) \geq \frac{1}{k}\left(W^* - \varepsilon\right)$. Moreover, π is ε'-egalitarian with $\varepsilon' = \max\left\{\frac{1}{k}, \varepsilon\right\}$.*

The theorem is proved in Appendix C. As we see, the factor 2 in Theorems 1 and 2 is equal to the number of sides of the market and so gets replaced with k in Theorem 3. This theorem allows one to capture team, club, or coalition formation. The following example illustrates that it can also capture organ exchanges under compatibility constraints.

Example 5 (Organ exchanges). Consider the problem of organ exchanges, as in (Roth, Sönmez, and Ünver, 2005, 2007). An agent is a patient-donor pair, in which the donor is willing to donate an organ to the patient, but the donation is not feasible because the two are incompatible. For practical reasons—see (Roth, Sönmez, and Ünver, 2005, 2007)—we rule out complicated organ trades, and focus attention on exchanges. These are trades that are done in pairs. At first glance, the organ exchange problem appears to be a one-sided problem, where any pair can be matched with any other pair. However, we can recast the model as a multi-sided matching problem.

Suppose that there are k types of agents. These could, for example, correspond to the blood types of the patient and the donor: (O,A), (A,B), and so on. Depending on the organ, other biological markers may be used in defining the type of an agent. Let the total number of types be k. In addition, assume that agents differ in the size $(s_1, s_2) \in \mathbb{R}^2$ of the organs of the patient and donor, respectively. Let μ_i represent the distribution of agents of type i in $\mathbb{R}^2$. Let $G = (\{1, \dots, k\}, E)$ be a compatibility graph, so that $(i, j) \in E$ whenever the type of the donor from i is compatible with the type of the patient from j, and vice versa. Now for any compatible pair of types (i, j) and for any two vectors of sizes $s^i = (s^i_1, s^i_2)$ and $s^j = (s^j_1, s^j_2)$, assume that the match quality between the pairs can be captured by a function $q_{i,j}(s^i, s^j)$.

To make this into a k-sided matching market, we augment each type by adding a "dummy" agent d_i, and add a sufficient quantity of copies of the dummy agent so that $\mu_i(\mathbb{R}^2)$ is the same for all types i.

Now define a utility function $u : \bigtimes_{i=1}^{k} \mathbb{R}^2 \cup \{d_i\} \to \mathbb{R}$ by

$$u(x_1, \dots, x_k) = \begin{cases} q_{i,j}(x_i, x_j), & \text{if } (i, j) \in E, \text{ and } x_i, x_j \text{ are the only non-dummy types} \\ 0, & \text{if there is at most one non-dummy type} \\ -M, & \text{otherwise} \end{cases},$$

where $M > 0$ is a large number. This now forms a multi-sided matching market as considered above, and the theory developed can be used to find stable and egalitarian matchings.

6 Aligned preferences

In this section, we discuss some foundations for preference alignment. Thus far, we have defined aligned preferences via utility functions. However, utility functions are just a convenient way to represent underlying preferences, and passing from preferences to utilities involves some arbitrariness unless there is a natural common scale, as in the presence of transfers. First, we argue that alignment is natural in the case of matching with transfers if agents bargain over transfers after matches are formed as in Pycia (2012). Second, we discuss conditions on agents' preferences sufficient for alignment, i.e., for the existence of a common utility function representing these preferences. The key requirement is a certain acyclicity property.

6.1 Aligned preferences in markets with transfers

We argue that aligned preferences are a natural assumption in models of matching with transfers and lack of commitment ability. Namely, we show that alignment arises when agents cannot commit to utility transfers at the stage where they bargain over matches, and the post-match surplus is shared according to Nash bargaining.

Consider the matching market with transferable utility introduced by Shapley and Shubik (1971) and used, for example, in marriage-market models of Becker (1973) and Galichon and Salanié (2022). If a couple (x, y) forms a match, then they generate a surplus $s(x, y)$ that they may share: x getting $\hat{u}(x, y)$ and y getting $\hat{v}(x, y)$, with $\hat{u}(x, y)+\hat{v}(x, y) = s(x, y)$. The model with transferable utilities assumes that the shares $\hat{u}(x, y)$ and $\hat{v}(x, y)$ are determined at the same time as the match. Effectively, the model assumes that transfers between couples are negotiated, agreed upon, and committed to, as part of the bargaining over who matches with whom.

We consider instead the possibility that agents cannot commit to a specific share of the surplus when they agree to form a pair with another agent. Specifically, we assume that, once a match has been formed, it cannot be broken without significant cost, and that the two members of a couple bargain over how to share the surplus according to the Nash bargaining model with fixed weights, β and $1 - \beta$, for the X and Y sides, respectively. This means that if a couple (x, y) forms, then $\hat{u}(x, y) = \beta \cdot s(x, y)$ while $\hat{v}(x, y) = (1 - \beta) \cdot s(x, y)$.

For a matching π to be ε-stable we would need that, for $\pi\times\pi$-almost all pairs (x_1, y_1), (x_2, y_2), either $\hat{u}(x_1, y_2) \leq \hat{u}(x_1, y_1) + \varepsilon$ or $\hat{v}(x_1, y_2) \leq \hat{v}(x_2, y_2) + \varepsilon$. Under Nash bargaining surplus-sharing, we get that π is ε-stable if at least one of the following inequalities hold

$$s(x_1, y_2) \le s(x_1, y_1) + \frac{\varepsilon}{\beta}$$
$$s(x_1, y_2) \le s(x_2, y_2) + \frac{\varepsilon}{1-\beta}$$

Defining $\varepsilon' = \varepsilon / \min\{\beta,\ 1-\beta\}$, we obtain

$$s(x_1, y_2) \le \max\{s(x_1, y_1),\ s(x_2, y_2)\} + \varepsilon'$$

and conclude that a matching π is ε-stable in the model with Nash bargaining surplus-sharing if and only if it is ε'-stable in a market with aligned preferences represented by $s(x, y)$.

6.2 Ordinal conditions for aligned preferences

In the absence of a monetary scale pinning down a particular utility representation of preferences, it becomes especially important to identify requirements on agents' preferences that enable the existence of a utility representation common to both sides of the market.

Recall that a binary relation is called a preference if it is complete and transitive. We denote preferences by $\succeq$. The two sides of the market are given by sets X and Y. For each $x \in X$ we have a preference $\succeq_x$ over Y. Likewise, for each $y \in Y$ we have a preference $\succeq_y$ over X. We collect these data into a tuple $(X, Y, \succeq_X, \succeq_Y)$ where $\succeq_X = (\succeq_x)_{x \in X}$ and $\succeq_Y = (\succeq_y)_{y \in Y}$.

Definition 5. A function $u : X \times Y \to \mathbb{R}$ is a **potential** for $(X, Y, \succeq_X, \succeq_Y)$ if

$$\begin{array}{l} u(x, y) \ge u(x, y') \iff y \succeq_x y' \\ u(x, y) \ge u(x', y) \iff x \succeq_y x' \end{array} \qquad \text{for all} \quad x, x', y, y'.$$

In other words, the potential is a single utility function representing the preferences of each of the sides of the market. The key condition for the existence of a potential turns out to be a form of acyclicity. A sequence of couples,

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n),$$

with $n > 2$ forms a cycle if $(x_n, y_n) = (x_1, y_1)$ and each couple (x_{i+1}, y_{i+1}) has exactly one agent in common with the preceding couple (x_i, y_i). The tuple $(X, Y, \succeq_X, \succeq_Y)$ is **acyclic** if, for any cycle $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where each common agent prefers their partner in (x_{i+1}, y_{i+1}) to their partner in (x_i, y_i), all common agents are, in fact, indifferent between the two partners.

For finite markets, acyclicity is necessary and sufficient for a potential to exist. This result is due to Niederle and Yariv (2009) who established the connection with potential games where Monderer and Shapley (1996) and Voorneveld and Norde (1997) used a similar condition. We extend this result to arbitrary topological spaces. This generalization requires additional technical requirements, similar to those necessary for the existence of a utility function on a general topological space.

For a preference $\succeq$ on a set Z, we denote by $\succ$ the strict part of $\succeq$, i.e., $z \succ z'$ if and only if $z \succeq z'$ but $z' \not\succeq z$. For a topological space Z, a preference $\succeq$ is said to be continuous if the upper contour sets

$$U_{\succeq}(z) = \{z' \in Z : z' \succ z\}$$

and lower contour sets

$$L_{\succeq}(z) = \{z' \in Z : z \succ z'\}$$

are open. In addition to the continuity of each agent's preferences, we need continuity with respect to changing the agent. We consider the following requirements on of $(X, Y, \succeq_X, \succeq_Y)$:

1. If $b \succ_a b'$, then there is a neighborhood N_a of a for which $b \succ_{\tilde{a}} b'$ for all $\tilde{a} \in N_a$.

2. If $b \succ_a b'$ then there are neighborhoods N_a of a, N_b of b, and $N_{b'}$ of b' such that $\tilde{b} \succ_{\tilde{a}} \tilde{b}'$ for all $(\tilde{a}, \tilde{b}, \tilde{b}') \in N_a \times N_b \times N_{b'}$.

3. If $b' \succeq_a b$ and $b \succeq_{a'} b''$, where $b \neq b', b''$ and $a \neq a'$, then, in any neighborhood of b, there exists $\tilde{b}$ such that $b' \succ_a \tilde{b}$ and $\tilde{b} \succ_{a'} b''$.

Property 1 is a continuity assumption relating how one agents' preferences must be similar to the preferences of agents who are close by on the same side of the market. Property 2 is a stronger joint continuity property, and, under some additional assumptions on the spaces X and Y, amounts to saying that $a \mapsto \succeq_a$ is a continuous map in the topology of closed convergence; see Kannai (1970). Property 3 is in the spirit of local strictness by Border and Segal (1994).

Theorem 4. *Let $(X, Y, \succeq_X, \succeq_Y)$ be such that X and Y are complete, separable, and connected topological spaces. Suppose that preferences are continuous, satisfy acyclicity, and properties (1) and (3). Then there is a potential u. If, moreover, Property 2 is satisfied, then u can be taken to be upper semicontinuous.*

The proof of Theorem 4 is in Appendix D. The proof uses a construction borrowed from Debreu (1954, 1964), which works by extending a primitive ordering defined on a countable dense subset to the whole space. In the case of Theorem 4, the ordering admits a potential on a countable dense subset, but the extension is more challenging than in the canonical utility representation problem of Debreu. One has to show the existence of a strict comparison for the utility of each agent, even for those who are not included in the countable dense subset. This necessitates the continuity and local strictness properties listed as the hypotheses of the theorem.

References

D. J. Abraham, A. Levavi, D. F. Manlove, and G. O'Malley. The stable roommates problem with globally ranked pairs. *Internet Mathematics*, 5(4):493–515, 2008.

N. Agarwal. An empirical model of the medical match. *American Economic Review*, 105(7):1939–1978, 2015.

F. Agostinelli, M. Luflade, P. Martellini, et al. On the spatial determinants of educational access. Technical report, 2021.

A. Alkan. Nonexistence of stable threesome matchings. *Mathematical social sciences*, 16(2):207–209, 1988.

J. Angrist, G. Gray-Lobe, C. M. Idoux, and P. A. Pathak. Still worth the trip? school busing effects in boston and new york. Technical report, National Bureau of Economic Research, 2022.

E. Anshelevich, S. Das, and Y. Naamad. Anarchy, stability, and utopia: creating better matchings. *Autonomous Agents and Multi-Agent Systems*, 26(1):120–140, 2013.

P. Antras, T. C. Fort, and F. Tintelnot. The margins of global sourcing: Theory and evidence from us firms. *American Economic Review*, 107(9):2514–2564, 2017.

I. Arieli, Y. Babichenko, and F. Sandomirskiy. Persuasion as transportation. *arXiv preprint arXiv:2307.07672*, 2023.

M. Armstrong. Competition in two-sided markets. *The RAND journal of economics*, 37(3):668–691, 2006.

N. Arnosti. A continuum model of stable matching with finite capacities. *arXiv preprint arXiv:2205.12881*, 2022.

I. Ashlagi and P. Shi. Optimal allocation without money: An engineering approach. *Management Science*, 62(4):1078–1097, 2016.

E. M. Azevedo and J. D. Leshno. A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268, 2016.

S. Banerjee, H. Konishi, and T. Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18:135–153, 2001.

G. S. Becker. A theory of marriage: Part i. *Journal of Political economy*, 81(4): 813–846, 1973.

M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *Journal of Functional Analysis*, 256(6):1907–1927, 2009.

J. Boerma, A. Tsyvinski, R. Wang, and Z. Zhang. Composite sorting. *arXiv preprint arXiv:2303.06701*, 2023.

K. C. Border and U. Segal. Dynamic consistency implies approximately expected utility preferences. *Journal of Economic Theory*, 63(2):170–188, 1994.

C. Brizzi, L. De Pascale, and A. Kausamo. L^∞-optimal transport for a class of strictly quasiconvex cost functions. *Journal of Mathematical Analysis and Applications*, 527(1):127331, 2023.

E. Cantillon, L. Chen, and J. S. Pereyra. Respecting priorities versus respecting preferences in school choice: When is there a trade-off? *arXiv preprint arXiv:2212.02881*, 2022.

G. Carlier. Optimal transportation and economic applications. 2012.

T. Champion, L. De Pascale, and P. Juutinen. The ∞-Wasserstein distance: Local solutions and existence of optimal transport maps. *SIAM Journal on Mathematical Analysis*, 40(1):1–20, 2008.

T. Chaney. The network structure of international trade. *American Economic Review*, 104(11):3600–3634, 2014.

S. Clark. The uniqueness of stable matchings. *Contributions in Theoretical Economics*, 6(1), 2006.

C. Daskalakis, A. Deckelbaum, and C. Tzamos. Strong duality for a multiple-good monopolist. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 449–450, 2015.

G. Debreu. Representation of a preference ordering by a numerical function. In *Decision processes*, pages 159–165. Wiley New York, 1954.

G. Debreu. Continuity properties of paretian utility. *International Economic Review*, 5(3):285–293, 1964.

M. Dinerstein and T. D. Smith. Quantifying the supply response of private schools to public policies. *American Economic Review*, 111(10):3376–3417, 2021.

F. Echenique and M. B. Yenmez. A solution to matching with preferences over colleagues. *Games and Economic Behavior*, 59(1):46–71, 2007.

J. Eeckhout. On the uniqueness of stable marriage matchings. *Economics Letters*, 69(1):1–8, 2000.

I. Ekeland. Notes on optimal transportation. *Economic Theory*, pages 437–459, 2010.

A. Ferdowsian, M. Niederle, and L. Yariv. Strategic decentralized matching: The effects of information frictions. *Working Paper*, 2023.

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

A. Galichon. *Optimal transport methods in economics*. Princeton University Press, 2018.

A. Galichon and B. Salanié. Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629, 2022.

A. Galichon, O. Ghelfi, and M. Henry. Stable and extremely unequal. *Economics Letters*, 226:111101, 2023.

W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

M. Greinecker and C. Kah. Pairwise stable matching in large economies. *Econometrica*, 89(6):2929–2974, 2021.

G. Z. Gutin, P. R. Neary, and A. Yeo. Unique stable matchings. *Games and Economic Behavior*, 141:529–547, 2023.

Y. Kannai. Continuity properties of the core of a market. *Econometrica*, 88:791–815, 1970.

A. Kausamo, L. De Pascale, and K. Wyczesany. 60 years of cyclic monotonicity: a survey. *arXiv preprint arXiv:2308.07682*, 2023.

Y.-H. Kim and B. Pass. A general condition for monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46 (2):1538–1550, 2014.

A. V. Kolesnikov, F. Sandomirskiy, A. Tsyvinski, and A. P. Zimin. Beckmann's approach to multi-item multi-bidder auctions. *arXiv preprint arXiv:2203.06837*, 2022.

M. Laverde. Distance to schools and equal access in school choice systems. *Working Paper*, 2022.

D. Lebedev, F. Mathieu, L. Viennot, A.-T. Gai, J. Reynier, and F. De Montgolfier. On using matching theory to understand p2p network design. In *INOC 2007, International Network Optimization Conference*, 2007.

S. Lee and L. Yariv. On the efficiency of stable matchings in large markets. *Working Paper*, 2018.

J. D. Leshno and I. Lo. The cutoff structure of top trading cycles in school choice. *The Review of Economic Studies*, 88(4):1582–1623, 2021.

S. Malamud, A. Cieslak, and A. Schrimpf. Optimal transport of information. *arXiv preprint arXiv:2102.10909*, 2021.

R. J. McCann. Exact solutions to the transportation problem on the line. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455(1984):1341–1380, 1999.

R. J. McCann and K. S. Zhang. A duality and free boundary approach to adverse selection. *arXiv preprint arXiv:2301.07660*, 2023.

D. Monderer and L. S. Shapley. Potential games. *Games and economic behavior*, 14 (1):124–143, 1996.

C. Ng and D. S. Hirschberg. Three-dimensional stable matching problems. *SIAM Journal on Discrete Mathematics*, 4(2):245–252, 1991.

M. Niederle and L. Yariv. Decentralized matching with aligned preferences. *National Bureau of Economic Research*, 2009.

P. Panigrahi. Endogenous spatial production networks: Quantitative implications for trade and productivity. CESifo Working Paper, 2021.

E. Perez-Richet and V. Skreta. Fraud-proof non-market allocation mechanisms. *Working Paper*, 2023.

M. Pycia. Stability and preference alignment in matching and coalition formation. *Econometrica*, 80(1):323–362, 2012.

J. Rawls. Atheory of justice. *Cambridge (Mass.)*, 1971.

J.-C. Rochet and J. Tirole. Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029, 2003.

A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic theory*, 125(2):151–188, 2005.

A. E. Roth, T. Sönmez, and M. U. Ünver. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *American Economic Review*, 97(3):828–851, 2007.

E. Saez. Using elasticities to derive optimal income tax rates. *The review of economic studies*, 68(1):205–229, 2001.

L. S. Shapley and M. Shubik. The assignment game i: The core. *International Journal of game theory*, 1:111–130, 1971.

M. Sørensen. How smart is smart money? a two-sided matching model of venture capital. *The Journal of Finance*, 62(6):2725–2762, 2007.

C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

M. Voorneveld and H. Norde. A characterization of ordinal potential games. *Games and Economic Behavior*, 19(2):235–242, 1997.

C. R. Walters. The demand for effective charter schools. *Journal of Political Economy*, 126(6):2179–2223, 2018.

A Exact algorithm for stable matching on $\mathbb{R}$

Recall that when X and Y are finite, a simple greedy algorithm will result in a stable matching: let (x, y) be the pair with the highest utility from matching. Match (x, y), remove the matched agents, and repeat. In this section, we construct an analogous procedure for the case where X and Y are possibly continuous.

Let $X = Y = \mathbb{R}$ and consider two populations represented by $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$ with equal total measure $\mu(X) = \nu(Y)$ and utility $u(x, y) = -|x-y|$. We will refer to the pair (μ, ν) as a market and assume that μ and ν are non-atomic measures, and the difference $\rho = \mu - \nu$ changes sign a finite number of times. We say that ρ changes sign K times if there is a collection of K intervals $I_0, \ldots, I_K$ such that the restriction of ρ to I_k is either a positive or a negative measure, and K is the minimal number with this property. We will refer to the intervals I_j as the imbalance regions.

We will construct a stable matching for (μ, ν) by sequentially simplifying the market. The idea is to identify submarkets (μ', ν') that are to be matched together in any stable matching the same way they would be matched if the rest of the population $(\mu - \mu', \nu - \nu')$ did not exist. As we will see, there is always a way to find such (μ', ν') that are to be matched in a simple monotone way. Repeating this procedure, we construct the stable matching for (μ, ν). This requires only a finite number of steps since the number of imbalance regions is finite and monotonically decreases.

We will start with two lemmas describing benchmark markets for which a stable matching can be obtained in one step. For general markets (μ, ν), these benchmark cases will correspond to the submarkets of interest.

We say that (μ, ν) is a *diagonal market* if $\mu = \nu$. In such markets, everyone can be matched with their ideal partner $x = y$. The opposite extreme is when nobody can be matched with their hypothetical ideal partner $x = y$. Populations (μ, ν) form a *disjoint market* if μ and ν are mutually singular, i.e., there is a measurable subset $A \subset \mathbb{R}$ such that $\mu(\mathbb{R} \setminus A) = 0$ and $\nu(A) = 0$. An important special case corresponds to $A = (-\infty, t]$ or $A = [t, \infty)$ for some $t \in \mathbb{R}$. In this case, the supports of μ and ν are ordered on the line, i.e., either $\mathrm{supp}(\mu) \leq \mathrm{supp}(\nu)$ or $\mathrm{supp}(\nu) \leq \mathrm{supp}(\mu)$. We call such (μ, ν) an *anti-diagonal market*. Equivalently, (μ, ν) is an anti-diagonal market if μ and ν are mutually singular and ρ changes sign only once.

Lemma 3. *If (μ, ν) are diagonal or anti-diagonal markets, then the stable matching π is unique and is assortative or anti-assortative, respectively.*

In other words, for diagonal markets, π is supported on the solution $F_\mu(x) =$

$F_\nu(y)$, where F_μ and F_ν are the cumulative distribution functions of μ and ν. Since $\mu = \nu$, π is supported on the 45-degree line. In the anti-diagonal case, π is supported on the solution to $F_\mu(x) = 1 - F_\nu(y)$.

Proof. For diagonal and anti-diagonal markets, $\rho = \mu - \nu$ changes sign only once. As shown by McCann (1999), a matching satisfying the no-crossing condition is unique whenever the sign changes at most two times. Assortative and anti-assortative matchings satisfy the no-crossing condition and, thus, are unique matchings with this property. By Lemma 1, they are unique stable ones. □

A market (μ', ν') is a *submarket* of (μ, ν) if $\mu' \le \mu$ and $\nu' \le \nu$. Note that $\mu'(\mathbb{R}) = \nu'(\mathbb{R})$ by the assumption that (μ', ν') is a market. For a submarket (μ', ν'), the *residual submarket* is defined by $(\mu'', \nu'') = (\mu - \mu', \nu - \nu')$.

Lemma 4. *For any market (μ, ν), there exists a unique diagonal submarket (μ', ν') such that the residual submarket $(\mu'', \nu'') = (\mu - \mu', \nu - \nu')$ is disjoint.*

A matching π is a stable matching of (μ, ν) if and only if $\pi = \pi' + \pi''$, where π' is the (unique) stable matching of (μ', ν') and π'' is a stable matching of (μ'', ν'').

Proof. We first construct a diagonal submarket (μ', ν') such that the residual submarket (μ'', ν'') is disjoint. Let $\rho = \mu - \nu$. We get that ρ can also be expressed as $\mu'' - \nu''$. By Hahn's theorem, there is a unique way to represent a signed measure τ as $\tau_+ - \tau_-$, where $\tau_\pm$ are positive measures that are mutually singular. We get that (μ'', ν'') must provide the Hahn decomposition of ρ, i.e., $\mu'' = \rho_+$ and $\nu'' = \rho_-$. Thus (μ', ν') is given by $\mu' = \mu - \rho_+$ and $\nu' = \nu - \rho_-$. It is a diagonal submarket since

$$\mu' - \nu' = (\mu - \nu) - (\rho_+ - \rho_-) = \rho - \rho = 0.$$

Its uniqueness follows from the uniqueness of the Hahn decomposition.

Let π' be the stable matching of the diagonal submarket (μ', ν'). By Lemma 3, matching π' is unique and is given by the assortative matching. If π'' is a stable matching of (μ'', ν''), then $\pi' + \pi''$ is a stable matching of (μ, ν). Indeed, by combining the two markets, we do not create any cross-market blocking pairs as any agent in π' is matched with their best partner.

It remains to show that any stable matching π of (μ, ν) can be represented as $\pi' + \pi''$ with stable π' and π'', where π' is the stable matching of the diagonal submarket (μ', ν'). We adapt an argument from the optimal transport literature with metric costs (see Gangbo and McCann, 1996, Proposition 2.9). Let π_{diag} be the restriction of π to the diagonal $\{(x, y) \colon x = y\}$, i.e., $\pi_{\mathrm{diag}}(A) = \pi'(A \cap \{(x, y) \colon x = y\})$ for any measurable A. Our goal is to show that $\pi' = \pi_{\mathrm{diag}}$. Towards contradiction, assume this equality does not hold. Hence, the marginals μ_{diag} and ν_{diag} of π_{diag} satisfy $\mu_{\text{diag}} \neq \mu'$ and $\nu_{\text{diag}} \neq \nu'$. By the uniqueness of the Hahn decomposition, $\mu'' = \mu - \mu_{\text{diag}}$ and $\nu'' = \nu - \nu_{\text{diag}}$ do not form a disjoint market, i.e., there is a set $B \subset \mathbb{R}$ with $\mu''(B) > 0$ and $\nu''(B) > 0$. Consider a set $S = \text{supp}(\pi) \setminus \{(x, y) : x = y\}$. Let X_S and Y_S be the projections of the set S. These are sets of full measure with respect to μ'' and ν'', and thus, the intersection $X_S \cap Y_S$ is non-empty. Pick $t \in X_S \cap Y_S$. There are two couples $(x_1, y_1), (x_2, y_2) \in \text{supp}(\pi)$ with $x_1 = t$ and $y_2 = t$ and $x_1 \neq y_1$ and $x_2 \neq y_2$ which violate the no-crossing condition. By Lemma 1, π cannot be stable. This contradiction implies that $\pi_{\text{diag}} = \pi'$. We conclude that any stable π can be represented as $\pi' + \pi''$, where π' is a stable matching of the diagonal submarket and π'' is a stable matching of a disjoint submarket. □

Lemma 4 reduces the problem of constructing and showing the uniqueness of a stable matching for general (μ, ν) to these questions for disjoint markets.

A submarket (μ', ν') is *independent* if members of (μ', ν') top-rank each other within (μ, ν), i.e., $|x - y'| > |x - y|$ for μ'-almost all x, ν'-almost all y, and $(\nu - \nu')$-almost all y' and $|x' - y| > |x - y|$ for μ'-almost all x, ν'-almost all y, and $(\mu - \mu')$-almost all x'. Independent submarkets are important because they can be matched myopically, i.e., without worrying about the residual populations.

Lemma 5. *Let (μ', ν') be an independent submarket of (μ, ν) and let $(\mu'', \nu'') = (\mu - \mu', \nu - \nu')$ be the residual submarket. A matching π is a stable matching of (μ, ν) if and only if $\pi = \pi' + \pi''$, where π' is a stable matching of (μ', ν') and π'' is a stable matching of (μ'', ν'').*

Proof. Let π' be a stable matching of an independent submarket (μ', ν') and π'' be a stable matching of the residual submarket (μ'', ν''). Then $\pi = \pi' + \pi''$ is a stable matching of (μ, ν) since combining the two markets cannot create cross-market blocking pairs by the independence of (μ', ν').

Now, let π be a stable matching of (μ, ν). We show that π can be represented as $\pi' + \pi''$, where π' is a stable matching of (μ', ν') and π'' be a stable matching of (μ'', ν''). Note that the stability of a submarket follows from the stability of a market, and thus we only need to check that $\pi = \pi' + \pi''$, where π' and π'' are matchings.

Strict inequalities in the definition of independence imply that μ' and μ'' cannot have any mass in common. In other words, μ', μ'' and ν', ν'' are mutually singular. By Hahn's theorem, there are disjoint sets X' and X'' such that μ' and μ'' are given by restricting μ on X' and X'', respectively. The disjoint sets Y' and Y'' are constructed analogously.

We now show that π cannot place positive mass on $X' \times Y''$ and $X'' \times Y'$. First, observe that $\pi(X' \times Y'') = \pi(X'' \times Y')$. Indeed,

$$\pi(X' \times Y'') - \pi(X'' \times Y') = \pi(X' \times (Y' \cup Y'')) - \pi((X' \cup X'') \times Y') = \mu(X') - \nu(Y') = 0.$$

Second, we show that the common value is zero. Towards contradiction, assume that $\pi(X' \times Y'') = \pi(X'' \times Y') > 0$. Thus $\mathrm{supp}(\pi)$ contains $(x_1, y_2) \in X' \times Y''$ and $(x_2, y_2) \in (X'', Y')$. By independence of (μ', ν'), the couple (x_1, y_2) is a blocking pair which contradicts stability of π. We conclude that π has no mass outside of $X' \times Y'$ and $X'' \times Y''$. Consequently, $\pi = \pi' + \pi''$, where π' is the restriction of π to $X' \times Y'$ and π'' is the restriction to $X'' \times Y''$. By the construction, π' is a matching of (μ', ν') and π'' is a matching of (μ'', ν''), which are both stable by the stability of π. □

We say that a matching π is Monge if π is supported on a graph of some function $f : \mathbb{R} \to \mathbb{R}$. The following lemma is the key step in the proof of Proposition 1. Starting from disjoint (μ, ν) such that $\rho = \mu - \nu$ changes sign $K \geq 2$ times, this lemma allows us to reduce K sequentially until we reach an anti-diagonal market ($K = 1$).

Lemma 6. *Let (μ, ν) be a disjoint market such that ρ changes sign $K \geq 2$ times. Then μ and ν can be represented as $\mu = \mu_1' + \mu_2' + \mu''$ and $\nu = \nu_1' + \nu_2' + \mu''$ so that*

1. *(μ_1', ν_1') and (μ_2', ν_2') are independent anti-diagonal submarkets of (μ, ν);*
2. *a matching π is a stable matching of (μ, ν) if and only if $\pi = \pi_1' + \pi_2' + \pi''$, where π_i' is the (unique) stable matching of (μ_i', ν_i'), $i = 1, 2$, and π'' is a stable matching of (μ'', ν'')*
3. *such a stable matching π is Monge if and only if π'' is Monge;*
4. *the residual submarket (μ'', ν'') is a disjoint market with $\rho' = \mu' - \nu'$ changing sign at most $K - 1$ times.*
5. *if μ and ν have piecewise constant density with m intervals of constancy, then (μ_1', ν_1') and (μ_2', ν_2') can be constructed in time of the order of m.*

Proof. Since (μ, ν) is a non-atomic disjoint market with $K \geq 2$, we can find points $a_0 < a_1 < a_2 < \ldots < a_{K-1} < a_{K+1}$ with $a_0 = -\infty$ and $a_K = +\infty$ such that either μ is supported on intervals $I_k = (a_k, a_{k+1})$ with even k and ν is supported on I_k with odd k, or the other way around. By the minimality of K, each interval I_k carries strictly positive μ-mass or strictly positive ν-mass.

Consider $\rho = \mu - \nu$ and let δ be a non-negative number. In each closed interval $\bar{I}_k = [a_k, a_{k+1}]$ we aim to pick a pair of points $a_k^+ \leq a_{k+1}^-$ so that the following conditions are satisfied:

1. equal-weight condition: $\rho([a_k^-, a_k^+]) = 0$ for $k = 1, \ldots, K$;

2. equal-distance condition: $|a_k^+ - a_k^-| = \delta$ for $k = 1, \ldots, K$.

The conditions on the collection of points $a_k^{\pm}$ and δ are closed. Consider the maximal $\delta \geq 0$ such that points satisfying the conditions exist and let $a_k^{\pm}$ be the corresponding collection of points.

As δ cannot be increased further, there is an interval $I_{k^*} = (a_{k^*}, a_{k^*+1})$ with $k^* = 1, \ldots, K-1$ such that $a_{k^*}^+ = a_{k^*+1}^-$. In other words, the two points $a_{k^*}^+$ and $a_{k^*+1}^-$ hit each other, which does not allow us to increase δ.

Consider the two submarkets (μ_1', ν_1') and (μ_2', ν_2') cut from (μ, ν) by the intervals $J_1^* = [a_{k^*}^-, a_{k^*}^+]$ and $J_2^* = [a_{k^*+1}^-, a_{k^*+1}^+]$, i.e.,

$$\mu_i'(A) = \mu(A \cap J_i^*), \qquad \nu_i'(A) = \mu(A \cap J_i^*) \qquad i = 1, 2$$

for any measurable $A \subset \mathbb{R}$. Denote by (μ'', ν'') the residual submarket $(\mu'', \nu'') = (\mu - \mu_1' - \mu_2', \nu - \nu_1' - \nu_2')$.

The submarket (μ_1', ν_1') is an independent anti-diagonal submarket of (μ, ν). To show this, assume without loss of generality that the interval I_{k^*} carries a positive μ-weight. Hence, μ_1' is supported on $[a_{k^*}, a_{k^*}^+]$ and ν_1' on $[a_{k^*}^-, a_{k^*}]$. By the equal-weight condition, $\mu'(\mathbb{R}) = \nu'(\mathbb{R})$ and so (μ_1', ν_1') is an anti-diagonal submarket of (μ, ν). The equal-distance condition implies that any $x \in [a_{k^*}, a_{k^*}^+]$ and $y \in [a_{k^*}^-, a_{k^*}]$ are within δ from each other, while the distances between x and $y' \in \mathrm{supp}(\nu'')$ and between x' and $y \in \mathrm{supp}(\mu'')$ are at least δ. Hence, (μ_1', ν_1') is independent. Similarly, (μ_2', ν_2') is an independent anti-diagonal submarket of (μ, ν). We get assertion 1.

Applying Lemma 5, we conclude that π is a stable matching of (μ, ν) if and only if $\pi = \pi_1' + \pi_2' + \pi''$, where π_i' is the unique stable matching of (μ_i', ν_i'), $i = 1, 2$, and π'' is a stable matching of (μ'', ν''). We obtain assertion 2.

The stable matching for anti-diagonal markets is Monge (Lemma 3). Since μ_1', μ_2', and μ'' are mutually singular (and similarly for ν), we conclude that a stable matching π is Monge if and only if π'' is Monge. Assertion 3 is proved.

The residual submarket (μ'', ν'') is a disjoint market, and $\rho' = \mu' - \nu'$ changes sign at most $K - 1$ times. Indeed, the interval I_{k^*} is removed from both μ and ν, and so the number of imbalance regions is reduced by at least 1. Thus assertion 4 holds.

Finally, μ and ν having piecewise constant densities with at most m intervals of constancy. Without loss of generality, these intervals are common and consecutive, i.e., there is a collection of points $b_0 < b_1 < \ldots < b_m$ such that the densities of μ and ν on an interval $J_i = (b_i, b_{i+1})$ are constants f_i and g_i, respectively. We can also assume that $(b_i)_{i=0,\ldots,m}$ is a subsequence of the sequence $(a_k)_{k=0,\ldots,k+1}$, which defines intervals I_k.

To construct the submarket (μ_1', ν_1') and (μ_2', ν_2'), we need to identify an interval I_{k^*} defined above. This can be done as follows. For each three consecutive intervals I_{k-1}, I_k, I_{k+1}, we look for a triplet of points $\alpha_k^- \in \bar{I}_{k-1}$, $\alpha_k \in \bar{I}_k$ and $\alpha_k^+ \in \bar{I}_{k+1}$ such that

$$\rho([\alpha_k^-, \alpha_k]) = \rho([\alpha_k, \alpha_k^+]) = 0 \qquad \text{and} \qquad |\alpha_k - \alpha_k^-| = |\alpha_k - \alpha_k^+|. \tag{14}$$

If $I_{k-1},\ I_k,\ I_{k+1}$ contain m_k intervals of constancy J_i, then finding a solution $(\alpha_k^-, \alpha_k, \alpha_k^+)$ or checking that no such solution exists can be done in $O(m_k)$ operations. Indeed, to solve the system (14), it is enough to test each of the subintervals of constancy for whether it contains a solution. Indeed, the cumulative distribution function $R_k(x) = \rho([a_{k-1}, x])$ for $x \in I_{k-1} \cup I_k \cup I_{k+1}$ is piecewise linear with m_k intervals of linearity. Testing corresponds to solving a linear system with three unknowns. We conclude that solving (14) boils down to solving of the order of m_k linear systems of given size and thus requires $O(m_k)$ operations. After finding a solution for each triplet of consecutive intervals, we pick $k = k^*$ that minimizes $\delta_k = |\alpha_k - \alpha_k^-|$. The interval I_{k^*} and points $(a_{k^*}^-, a_{k^*}^+, a_{k^*+1}^-, a_{k^*+1}^+) = (\alpha_{k^*}^-, \alpha_{k^*}, \alpha_{k^*}, \alpha_{k^*}^+)$ determine (μ_1', ν_1') and (μ_2', ν_2').

We need $O(m_k)$ operations per each interval I_k, and thus, the total number of operations is of the order $\sum_k m_k$, i.e., of the order of the number of intervals of constancy m. Assertion 5 is proved. $\square$

We are now ready to prove Proposition 1.

Proof of Proposition 1. Consider (μ, ν) and assume that $\rho = \mu - \nu$ changes sign K times. By Lemma 4, any stable matching $\pi = \pi_{\text{diag}} + \pi_1$, where π_{diag} is a unique stable matching of a diagonal market $(\mu - \rho_+, \nu - \rho_-)$ and π_1 is a stable matching of the residual submarket (μ_1, ν_1) with $\mu_1 = \rho_+$ and $\nu_1 = \rho_-$.

The residual submarket (μ_1, ν_1) is disjoint and $\rho_1 = \mu_1 - \nu_1$ changes sign K times since $\rho_1 = \rho$. We construct a sequences of submarkets (μ_k, ν_k) inductively starting from (μ_1, ν_1). Suppose (μ_k, ν_k) is already constructed and let (μ_k', ν_k') to be its independent submarket from Lemma 5. We then define (μ_{k+1}, ν_{k+1}) by

$$\mu_{k+1} = \mu_k - \mu_k' \quad \text{and} \quad \nu_{k+1} = \nu_k - \nu_k'.$$

Let π_k' be a stable matching of (μ_k', ν_k') which is unique since $\rho_k' = \mu_k' - \nu_k'$ changes sign at most two times. Thus any stable matching of (μ_ν) can be represented as

$$\pi = \pi_{\text{diag}} + \pi_1' + \pi_2' + \ldots + \pi_{k-1}' + \pi_k,$$

where π_k is a stable matching of (μ_k, ν_k). By Lemma 5, ρ_k changes sign at most $K - k + 1$ times, there is $L \le K$ such that ρ_L changes sign at most one time. Abusing the notation, denote the stable matching of (μ_L, ν_L) by π'_L is unique by Lemma 3.

We conclude that a stable matching π of (μ, ν) is unique and has the following form

$$\pi = \pi_{\mathrm{diag}} + \underbrace{\pi'_1 + \pi'_2 + \ldots \pi'_{L-1} + \pi'_L}_{L \le K \text{ terms}}.$$

By Lemma 6, $pi'_1 + \pi'_2 + \ldots \pi'_{L-1} + \pi'_L$ is a Monge matching. Thus π is a convex combination of a diagonal matching and a Monge matching. Consequently, for each x there are at most two distinct y such that (x, y) is in the support of π. Moreover, if there are two such y, one of them necessarily equals x.

We now consider the computational complexity of constructing π for μ and ν having piecewise constant densities with at most m intervals of constancy. Without loss of generality, these intervals are common and consecutive, i.e., there is a collection of points $b_0 < b_1 < \ldots < b_m$ such that the densities of μ and ν on an interval $J_i = (b_i, b_{i+1})$ are constants f_i and g_i, respectively. The collection of all these numbers is the input of the algorithm.

Constructing the diagonal π_{diag} requires a linear number of operations in m. Indeed, $\mu - \rho_+ = \nu - \rho_-$ have density $\min\{f_i, g_i\}$ on J_i and π_{diag} has the corresponding density on the diagonal of $J_i \times J_i$.

The complexity bottleneck corresponds to finding independent submarkets from Lemma 5. By Assertion 5, this requires of the order of m operations for each (μ_k, ν_k). The number of steps k is bounded by the number of times ρ changes its sign, and this number cannot exceed $m - 1$. We conclude that stable matching can be constructed in $O(m^2)$ operations.

B Proofs for Section 3

We prove two more general results applicable to

$$c(x, y) = -h\left(u(x, y)\right), \tag{15}$$

and then deduce Theorem 1 as their corollary. The first result addresses the case of positive α.

Theorem 5. *For a market with aligned preferences, assume that the utility function u takes values in an open interval $I \subset \mathbb{R}$, possibly infinite. Let $h\colon I \to (0, +\infty)$ be a* *differentiable function and assume that there exists $\alpha > 0$ such that*

$$\frac{h'(t)}{h(t)} \geq \alpha \quad \textit{for any} \quad t \in I. \tag{16}$$

Then any solution to the transportation problem with cost (15) is ε-stable with

$$\varepsilon = \frac{\ln 2}{\alpha}.$$

An example of h satisfying the requirements of the theorem is $h(t) = \exp(\alpha \cdot t)$.

We will need the following result by Beiglböck, Goldstern, Maresch, and Schachermayer (2009). Consider an optimal transportation problem

$$\min_{\pi} \int_{X \times Y} c(x, y) \, \mathrm{d}\pi(x, y).$$

with Polish spaces X and Y, measurable cost c. Assume that the value of this transportation problem is finite and is attained at some π^*. Beiglböck, Goldstern, Maresch, and Schachermayer (2009) establish the existence of a c-cyclic monotone Γ such that $\pi^*(\Gamma) = 1$, i.e., the optimal transportation plan is supported on a cyclically-monotone set. Recall that c-cyclic monotonicity of Γ means that

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

for all $n \geq 2$ and pairs of points $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ with the convention $y_{n+1} = y_1$.

Proof of Theorem 5. Consider now the optimal transportation problem (3) with c from (15) let π^* be its solution. By the result of Beiglböck, Goldstern, Maresch, and Schachermayer (2009), π^* is supported on some c-monotone set Γ. The requirement of cyclic monotonicity with $n = 2$ implies that for any $(x_1, y_1), (x_2, y_2) \in \Gamma$

$$h\big(u(x_1, y_2)\big) + h\big(u(x_2, y_1)\big) \leq h\big(u(x_1, y_1)\big) + h\big(u(x_2, y_2)\big).$$

Dropping the second term on the left-hand side and replacing both terms on the right-hand side with their maximum, we obtain

$$h\big(u(x_1, y_2)\big) \leq 2 \cdot h\Big(\max\{u(x_1, y_1), u(x_2, y_2)\}\Big). \tag{17}$$

Consider a pair of points t, t' from I. Integrating the bound on the derivative (6) from t to t', we get

$$\ln\left(\frac{h(t')}{h(t)}\right) \geq \alpha \cdot (t' - t) \quad \text{for} \quad t \leq t'. \tag{18}$$

Pick $t = \max\{u(x_1, y_1), u(x_2, y_2)\}$ and $t' = u(x_1, y_2)$. Consider two cases depending on whether $t \leq t'$ or $t > t'$. If $t \leq t'$, plugging t and t' into into (18) and taking into account the bound (17), we obtain

$$\ln 2 \geq \alpha \cdot (u(x_1, y_2) - \max\{u(x_1, y_1), u(x_2, y_2)\}).$$

If $t > t'$, this inequality holds trivially as the right-hand side is negative. We conclude that

$$u(x_1, y_2) \leq \max\{u(x_1, y_1), u(x_2, y_2)\} + \frac{\ln 2}{\alpha}$$

for a set of (x_1, y_1) and (x_2, y_2) of full π^*-measure. Thus π^* is ε-stable with $\varepsilon = \frac{\ln 2}{\alpha}$. $\square$

The following theorem deals with negative α.

Theorem 6. *For a market with aligned preferences, assume that the utility function u takes values in an open interval $I \subset \mathbb{R}$, possibly infinite. Let $h\colon I \to (-\infty, 0)$ be a differentiable increasing function and assume that there exists $\alpha < 0$ such that*

$$\frac{h'(t)}{h(t)} \geq \alpha \quad \textit{for any} \quad t \in I.$$

Then any solution to the transportation problem with cost (15) is ε-egalitarian with

$$\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}.$$

For example, a function $h(t) = -\exp(-|\alpha| \cdot t)$ satisfies the requirements of Theorem 6.

Proof. Without loss of generality, we assume that the populations μ and ν are given by probability measures. Recall that $U_{\min}(\pi)$ is the essential infimum of $u(x, y)$ with respect to measure π, i.e.,

$$U_{\min}(\pi) = \inf\big\{\lambda \in \mathbb{R} : \pi(\{u(x, y) < \lambda\}) > 0\big\}.$$

The egalitarian lower bound is given by

$$U^*_{\min}(\mu,\nu) = \sup_{\pi\in\Pi(\mu,\nu)} U_{\min}(\pi). \tag{19}$$

Consider the set C of all hypothetical couples whose utility is below the egalitarian lower bound by more than ε

$$C = \left\{(x,y)\in X\times Y\colon u(x,y) < U^*_{\min}(\mu,\nu)-\varepsilon\right\}.$$

Let π^* be a solution to an optimal transportation problem with cost (15). Our goal is to show that $\pi^*(C)$ cannot be too big. Fix small $\delta>0$ and find a matching $\pi'\in\Pi(\mu,\nu)$ such that

$$U_{\min}(\pi') \geq U^*_{\min}(\mu,\nu)-\delta.$$

Note that we may not be able to find such a matching for $\delta=0$ since the supremum in (19) may not be attained. We get

$$\int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi'(x,y) \leq \int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi^*(x,y)$$

since π^* is the optimal matching for the transportation problem with cost $-h(u(x,y))$. Since h is increasing, $h(u(x,y)) \geq h(U^*_{\min}(\mu,\nu)-\delta)$ for π'-almost all pairs (x,y). Thus, the left-hand side admits the following bound

$$h(U^*_{\min}(\mu,\nu)-\delta) \leq \int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi'(x,y)$$

and thus

$$h(U^*_{\min}(\mu,\nu)-\delta) \leq \int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi^*(x,y).$$

Since h and u are continuous and $\delta>0$ was arbitrary, we get

$$h(U^*_{\min}(\mu,\nu)) \leq \int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi^*(x,y). \tag{20}$$

Using monotonicity and negativity of h and the definition of C, we obtain

$$\begin{aligned}\int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi^*(x,y) &= \int_C -h(u(x,u))\,\mathrm{d}\pi^*(x,y) + \int_{X\times Y\setminus C} h(u(x,u))\,\mathrm{d}\pi^*(x,y)\\ &\leq \int_C h(u(x,u))\,\mathrm{d}\pi^*(x,y)\\ &\leq h(U^*_{\min}(\mu,\nu)-\varepsilon)\cdot\pi^*(C).\end{aligned}$$

Combining this bound with (20) gives

$$h(U^*_{\min}(\mu,\nu)) \le h(U^*_{\min}(\mu,\nu)-\varepsilon)\cdot \pi^*(C)$$

and thus

$$\pi^*(C) \le \frac{h(U^*_{\min}(\mu,\nu))}{h(U^*_{\min}(\mu,\nu)-\varepsilon)}.$$

Note that the inequality changes the direction because of the negativity of h. Similarly to (18), integrating the bound on the logarithmic derivative of h, we obtain

$$\ln\left(\frac{|h(t')|}{|h(t)|}\right) \le |\alpha|\cdot(t'-t) \quad \text{for} \quad t \le t'$$

and conclude that

$$\pi^*(C) \le \exp(-|\alpha|\cdot\varepsilon).$$

Plugging in $\varepsilon = \frac{\max\{1,\ln|\alpha|\}}{|\alpha|}$, we get

$$\pi^*(C) \le \min\left\{\exp(-1),\, \frac{1}{|\alpha|}\right\}$$

and thus

$$\pi^*(C) \le \varepsilon.$$

We conclude that π^* is ε-egalitarian with $\varepsilon = \frac{\max\{1,\ln|\alpha|\}}{|\alpha|}$. □

Proof of Theorem 1. Theorem 1 immediately follows from Theorems 5 and 6. Consider $h(t) = \exp(\alpha\cdot t)$. For $\alpha > 0$, Theorem 5 implies that the solution to the optimal transportation problem with the cost $c(x,y) = -h\big(u(x,y)\big)$ is ε-stable with $\varepsilon = \frac{\ln 2}{\alpha}$. For $\alpha < 0$, Theorem 5 gives ε-egalitarianism with $\varepsilon = \frac{\max\{1,\ln|\alpha|\}}{|\alpha|}$. Multiplying a cost function by a positive factor and adding a constant does not affect the optimum. Thus a solution to the transportation problem with cost

$$c_\alpha(x,y) = \frac{1-\exp(\alpha\cdot u(x,y))}{\alpha}$$

has the same properties.

Proof of Corollary 1. Recall the setting: utility u is continuous and bounded, but the spaces X and Y are not assumed to be compact.

We first show that the sets $\Pi^u_{+\infty}(\mu,\nu)$ and $\Pi^u_{-\infty}(\mu,\nu)$, corresponding to $\alpha \to \pm\infty$, are non-empty. The argument for the two cases is identical, and so we focus on $\Pi^u_{+\infty}(\mu,\nu)$. Consider a sequence $\alpha_n \to +\infty$ and let π_n be a solution to the optimal transportation problem from Theorem 1 with $\alpha = \alpha_n$. Such a solution is guaranteed to exist under our assumptions on u; see Remark 3.1. The set of all transportation plans $\Pi(\mu,\nu)$ for $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$ with Polish X and Y is sequentially compact in the topology of weak convergence; this is a corollary of Prokhorov's theorem, see Lemma 4.4 in Villani (2009). Thus, possibly passing to a subsequence, we conclude that the sequence π_n converges weakly to some $\pi_{+\infty} \in \Pi(\mu,\nu)$. Thus $\Pi^u_{+\infty}(\mu,\nu)$ and $\Pi^u_{-\infty}(\mu,\nu)$ are non-empty.

We now show that $\Pi^u_{+\infty}(\mu,\nu)$ consists of stable matchings. In other words, we demonstrate that $\pi_{+\infty}$ is stable. Consider a continuous function $f\colon (X\times Y)^2 \to \mathbb{R}$ given by

$$f(x_1,y_1,x_2,y_2) = \max\Big\{0,\quad u(x_1,y_2) - \max\left\{u(x_1,y_1),\, u(x_2,y_2)\right\}\Big\}.$$

A matching π is stable if and only if $\int f\, \mathrm{d}\pi \times \mathrm{d}\pi = 0$. Note that for an ε-stable matching, this integral does not exceed $\varepsilon \cdot \mu(X)\cdot \nu(Y)$. We obtain

$$\int f\, \mathrm{d}\pi_{+\infty} \times \mathrm{d}\pi_{+\infty} = \lim_{n\to\infty} \int f\, \mathrm{d}\pi_n \times \mathrm{d}\pi_n \leq \lim_{n\to\infty} \frac{\ln 2}{\alpha_n}\cdot \mu(X)\cdot\nu(Y) = 0.$$

Since the left-hand side is non-negative, we get that it equals zero. Thus $\pi_{+\infty}$ is stable. We conclude that $\Pi^u_{+\infty}(\mu,\nu)$ consists of stable matchings, and so a stable matching exists.

To show that $\Pi^u_{-\infty}(\mu,\nu)$ consists of egalitarian matchings, consider a weak limit $\pi_{-\infty}$ of a sequence of matchings π_{α_n}, where $\alpha_n \to -\infty$. Fix $\varepsilon > 0$ and let C be the set of hypothetical couples whose utility is below the egalitarian lower bound by more than ε

$$C_\varepsilon = \left\{(x,y) \in X\times Y\colon u(x,y) < U^*_{\min}(\mu,\nu) - \varepsilon\right\}.$$

By the continuity of u, the set C is open and thus

$$\pi_{-\infty}(C_\varepsilon) \leq \liminf_{n\to\infty} \pi_{\alpha_n}(C_\varepsilon)$$

By Theorem 1, the right-hand side goes to zero, and thus $\pi_{-\infty}(C_\varepsilon) = 0$ for any $\varepsilon > 0$. Sets C_ε are decreasing in ε. Hence,

$$\pi_{-\infty}(C_0) = \pi_{-\infty}\left(\cup_{\varepsilon>0} C_\varepsilon\right) = \lim_{\varepsilon\to 0} \pi_{-\infty}(C_\varepsilon) = 0.$$

Since $\pi_{-\infty}(C_0) = 0$, for $\pi_{-\infty}$-almost all couples (x, y) the utility $u(x, y)$ is at least $U^*_{\min}(\mu, \nu)$, i.e., $\pi_{-\infty}$ is an egalitarian matching. Therefore, all elements of $\Pi^u_{-\infty}(\mu, \nu)$ are egalitarian matchings.

Finally, the convexity of $\Pi^u_{+\infty}(\mu, \nu)$ and $\Pi^u_{-\infty}(\mu, \nu)$ follows from the convexity of the set of solutions to the optimal transport problem. The weak closedness of $\Pi^u_{+\infty}(\mu, \nu)$ and $\Pi^u_{-\infty}(\mu, \nu)$ follows from the weak closeness of the set of solutions via the standard diagonal procedure.

Proof of Theorem 2. Let π be an ε-stable matching with marginals μ and ν. Since u is continuous, Lemma 2 and formula (9) imply that, for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$,

$$u(x_1, y_2) \le \max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon.$$

By non-negativity of u, we get

$$u(x_1, y_2) \le u(x_1, y_1) + u(x_2, y_2) + \varepsilon.$$

Let π' be any other matching with marginals μ and ν. Consider a distribution $\lambda \in \mathcal{M}_+\big((X \times Y) \times (X \times Y)\big)$ such that the marginals of λ on (x_1, y_1) and on (x_2, y_2) are equal to π and the marginal on (x_1, y_2) is π'. We get

$$\begin{aligned}
W(\pi') &= \int_{X\times Y} u(x_1, y_2)\ \mathrm{d}\pi'(x_1, y_2) = \int_{(X\times Y)\times(X\times Y)} u(x_1, y_2)\ \mathrm{d}\lambda(x_1, y_1, x_2, y_2) \\
&\le \int_{(X\times Y)\times(X\times Y)} (u(x_1, y_1) + u(x_2, y_2) + \varepsilon)\ \mathrm{d}\lambda(x_1, y_1, x_2, y_2) = \\
&= \int_{X\times Y} u(x_1, y_1)\ \mathrm{d}\pi(x_1, y_1) + \int_{X\times Y} u(x_2, y_2)\ \mathrm{d}\pi(x_2, y_2) + \varepsilon = \\
&= 2W(\pi) + \varepsilon.
\end{aligned}$$

We thus obtain

$$W(\pi) \ge \frac{1}{2}\left(W(\pi') - \varepsilon\right)$$

for any matching π'. In particular, this inequality holds for π' maximizing welfare. Thus $W(\pi) \ge \frac{1}{2}\left(W^*(\mu, \nu) - \varepsilon\right)$.

Now we show that a substantial fraction of agents in an ε-stable matching π have utilities above the egalitarian lower bound $U^*_{\min}(\mu, \nu)$. Consider the set of hypothetical couples whose utility is more than ε below $U^*_{\min}(\mu, \nu)$

$$C = \left\{(x, y) \in X \times Y\colon u(x, y) < U^*_{\min}(\mu, \nu) - \varepsilon\right\}.$$

Our goal is to show that $\pi(C)$ cannot be too big. Let π' be the egalitarian matching, which exists by Corollary 1. Take λ as in the construction above for the pair π and π'. In other words, λ is a distribution on $(X \times Y) \times (X \times Y)$ with marginals π on (x_1, y_1) and (x_2, y_2) and π' on (x_1, y_2). Thus $u(x_1, y_2) \geq U^*_{\min}(\mu, \nu)$ on a set of full λ-measure. By ε-stability, $u(x_1, y_2) \leq \max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon$ and thus

$$\max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon \geq U^*_{\min}(\mu, \nu).$$

Towards a contradiction, assume that $\pi(C) > \frac{1}{2}\pi(X \times Y)$. Then $\lambda(C \times C) > 0$ by Dirichlet's pigeonhole principle. On the other hand,

$$\max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon < U^*_{\min}(\mu, \nu)$$

on $C \times C$. This contradiction implies that $\pi(C) \leq \pi(X \times Y)/2$ and thus any ε-stable matching π is ε'-egalitarian with $\varepsilon' = \max\{1/2,\, \varepsilon\}$.

C Proof of Theorem 3

We extend the proof techniques used in the two-sided case.

Proof of Assertion 1. Kim and Pass (2014) generalize two-marginal c-cyclic monotonicity condition that we used in the proof of Theorem 1. They consider a k-marginal optimal transportation problem with a bounded continuous cost $c\colon X_1 \times \ldots \times X_k \to \mathbb{R}$ and show that π is optimal if and only if $\mathrm{supp}(\pi)$ is c-cyclically monotone. Here, a set $T \subset X_1 \times \ldots \times X_k$ is c-cyclically monotone if for any collection of k-tuples $(x_1^i, \ldots, x_k^i) \in T$, $i = 1, \ldots, n$, and a collection of k permutations $\sigma_1, \ldots, \sigma_k$ of the set $\{1, \ldots, n\}$, the following inequality holds:

$$\sum_{i=1}^{n} c(x_1^i, \ldots, x_k^i) \leq \sum_{i=1}^{n} c(x_1^{\sigma_1(i)}, \ldots, x_k^{\sigma_k(i)}).$$

We now apply this result to $c = c_\alpha$ with

$$c_\alpha(x) = \frac{1 - \exp(\alpha \cdot u(x))}{\alpha}.$$

Let π^* be a solution to the optimal transportation problem with this cost. Take the number of points n equal to the number of marginals k and put $\sigma_i(1) = i$. We obtain that for any any collection $x^i = (x_1^i, \ldots, x_k^i) \in \mathrm{supp}(\pi^*)$, $i = 1, \ldots, k$

$$\begin{aligned}\sum_{i=1}^{k} \exp\left(\alpha \cdot u(x^i)\right) &\geq \sum_{i=1}^{k} \exp\left(\alpha \cdot u(x_1^{\sigma_1(i)}, \ldots, x_k^{\sigma_k(i)})\right) \\ &= \exp\left(\alpha \cdot u(x_1^1, \ldots, x_k^k)\right) + \sum_{i=2}^{k} \exp\left(\alpha \cdot u(x_1^{\sigma_1(i)}, \ldots, x_k^{\sigma_k(i)})\right).\end{aligned}$$

Thus

$$k \cdot \max_i \exp\left(\alpha \cdot u(x^i)\right) \geq \exp\left(\alpha \cdot u(x_1^1, \ldots, x_k^k)\right).$$

Taking logarithm, we get

$$\max_i u(x_i) \geq u(x_1^1, \ldots, x_k^k) + \frac{\ln k}{\alpha}$$

and conclude that π^* is ε-stable with $\varepsilon = \frac{\ln k}{\alpha}$.

Proof of Assertion 2. Without loss of generality, we assume that there is a unit amount of agents on each side of the market, i.e., $\mu_i \in \Delta(X_i)$ for all $i = 1, \ldots, k$.

Let $U^*_{\min}(\mu_1, \ldots, \mu_k)$ be the egalitarian lower bound and π' be a matching such that $u(x) \geq U^*_{\min} - \delta$ where $\delta > 0$ is small. Consider a set

$$C = \left\{x \in X : u(x) < U^*_{\min} - \varepsilon\right\}.$$

A solution π^* to the optimal transportation problem with cost c_α maximizes $\int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi$ over all matchings and thus

$$\int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \geq \int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi'(x) \geq -\exp\left(\alpha \cdot (U^*_{\min} - \delta)\right).$$

Since δ is arbitrary, we obtain

$$\begin{aligned}-\exp\left(\alpha \cdot U^*_{\min}\right) &\leq \int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \\ &= \int_C -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) + \int_{X \setminus C} -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \\ &\leq \int_C -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \\ &\leq -\exp\left(\alpha \cdot (U^*_{\min} - \varepsilon)\right) \cdot \pi(C).\end{aligned}$$

Thus

$$\alpha \cdot U^*_{\min} \geq \alpha \cdot (U^*_{\min} - \varepsilon) + \ln(\pi(C)),$$

and so $\alpha \cdot \varepsilon \geq \ln(\pi(C))$. Plugging in $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$, we get

$$\pi(C) \leq \min\left\{\exp(-1), \frac{1}{|\alpha|}\right\}$$

and thus $\pi(C) \leq \varepsilon$. We conclude that π is ε-egalitarian with $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$.

Proof of Assertion 3. Let π be an ε-stable matching. Since u is continuous, ε-stability implies that, for all collections of points $x^i = (x^i_1, \ldots, x^i_k) \in \mathrm{supp}(\pi)$, $i = 1, \ldots, k$,

$$u(x^1_1, \ldots, x^k_k) \leq \max_{i=1,\ldots,k} u(x^i) + \varepsilon \leq u(x^1) + \ldots + u(x^k) + \varepsilon.$$

Let π' be any other matching. Denote $X = X_1 \times \ldots \times X_k$ and consider a distribution

$$\lambda \in \mathcal{M}_+\big(\underbrace{X \times \ldots \times X}_{k \text{ times}}\big)$$

such that the marginals of λ on each X are equal to π, and the marginal on the set $\{(x^1_1, \ldots, x^k_k) : (x^1, \ldots, x^k) \in X \times \ldots \times X\}$ coincides with π'. We get

$$\begin{aligned} W(\pi') &= \int_X u(x^1_1, \ldots, x^k_k)\ \mathrm{d}\pi'(x^1_1, \ldots, x^k_k) = \int_{X\times\ldots\times X} u(x^1_1, \ldots, x^k_k)\ \mathrm{d}\lambda(x^1, \ldots, x^k) \\ &\leq \int_{X\times\ldots\times X} \big(u(x^1) + \ldots + u(x^k) + \varepsilon\big)\ \mathrm{d}\lambda(x^1, \ldots, x^k) = \\ &= k \cdot W(\pi) + \varepsilon. \end{aligned}$$

We thus obtain $W(\pi) \geq \frac{1}{k}\left(W(\pi') - \varepsilon\right)$ for any matching π'. In particular, this inequality holds for π' maximizing welfare. Thus $W(\pi) \geq \frac{1}{k}\left(W^* - \varepsilon\right)$.

We now show that a substantial fraction of agents in an ε-stable matching π have utilities above the egalitarian utility level $U^*_{\min}$. Consider

$$C_\delta = \left\{x \in X \colon u(x) < U^*_{\min} - \varepsilon - \delta\right\},$$

where δ is a small positive number. Our goal is to bound $\pi(C_0)$, but we will bound $\pi(C_\delta)$ with $\delta > 0$ first. Fix $\delta > 0$. As above, let λ be a distribution on $X \times \ldots \times X$ with marginals π on each copy of X and π' on $(x^1_1, \ldots, x^k_k)$. Let π' in this construction be a matching such that $u(x) \geq U^*_{\min} - \delta$ for π'-almost all x. Thus $u(x^1_1, \ldots, x^k_k) \geq U^*_{\min} - \delta$ on a set of full λ-measure. By ε-stability, $u(x^1_1, \ldots, x^k_k) \leq \max_i u(x^i) + \varepsilon$ and thus

$$\max_i u(x^i) + \varepsilon \geq U^*_{\min} - \delta.$$

Towards a contradiction, assume that $\pi(C_\delta) > \frac{1}{k}\pi(X)$. Then

$$\begin{aligned} \lambda(C_\delta \times \ldots \times C_\delta) &\geq \lambda(X \times \ldots \times X) - \sum_i \lambda(\{(x^1, \ldots, x^i) \colon x_i \notin C_\delta\}) \\ &= \pi(X)^{k-1} \left(\pi(X) - k \cdot \pi(C_\delta)\right) > 0. \end{aligned}$$

On the other hand,

$$\max_i u(x_i) + \varepsilon < U^*_{\min} - \delta$$

on $C_\delta \times \ldots \times C_\delta$. This contradiction implies that $\pi(C_\delta) \leq \frac{1}{k}\pi(X)$. The sequence of sets C_δ is decreasing in δ and $C_0 = \cup_{\delta>0} C_\delta$. Hence,

$$\pi(C_0) = \lim_{\delta \to 0} \pi(C_\delta) \leq \frac{1}{k}\pi(X)$$

and thusany ε-stable matching π is ε'-egalitarian with $\varepsilon' = \max\{\frac{1}{k}, \varepsilon\}$.

D Proof of Theorem 4

The existence of a potential without continuity guarantees. Let $\geq$ be the binary relation on $Z = X \times Y$ defined by $(x, y) \geq (x', y')$ if and only if $x = x'$ and $y \succeq_x y'$, or $y = y'$ and $x \succeq_y x'$. Our goal is to show the existence of a potential, i.e., a function $u \colon X \times Y \to \mathbb{R}$ such that for any $(x, y), (x', y')$ with $x = x'$ or $y = y'$,

$$u(x, y) \geq u(x', y') \iff (x, y) \geq (x', y').$$

We write $>$ for the strict part of $\geq$.

Define the binary relation $\geq^*$ as the transitive closure of $\geq$, which is constructed as follows. Say that $z \geq^* z'$ if there exists a sequence $z^1, \ldots, z^k$ in Z with

$$z = z^1 \geq z^2 \geq \ldots \geq z^{k-1} \geq z^k = z'.$$

The binary relation $\geq^*$ is a preorder, i.e., it is transitive and reflexive but may not be complete. We also define $>^*$ by $z >^* z'$ if $z \geq^* z'$ and at least one of the $\geq$ comparisons in the above sequence is $>$.

A function u represents $\geq^*$ if

$$u(x,y) \geq u(x',y') \Longleftrightarrow (x,y) \geq^* (x',y')$$

for any (x,y) and (x',y'). It is easy to see that any such function is a potential, and thus, it is enough to establish a representation of $\geq^*$.

We construct the representation of $\geq^*$ as follows. Fix two countable dense subsets $\bar{X}$ of X and $\bar{Y}$ of Y and denote $\bar{X} \times \bar{Y}$ by $\bar{Z}$. Consider

$$L_{\geq^*}(z) = \{z' \in \bar{Z} : z \geq^* z'\}$$

and let $r\colon \bar{Z} \to \mathbb{N}$ be an enumeration of $\bar{Z}$. Define a function $u\colon X \times Y \to \mathbb{R}$ by

$$u(z) = \sum_{z' \in L_{\geq^*}(z)} \left(\frac{1}{2}\right)^{r(z')}. \tag{21}$$

To show that u represents $\geq^*$, we need the following technical lemma, which is the key step in the proof.

Lemma 7. *If* $(x,y) >^* (x',y')$ *then there exists* $(\bar{x},\bar{y}) \in \bar{Z}$ *with*

$$(x,y) >^* (\bar{x},\bar{y}) >^* (x',y').$$

Proof. By definition of $>^*$, if $(x,y) >^* (x',y')$ then there exists $(\hat{x},\hat{y})$ and $(\hat{x}',\hat{y}')$ with $(x,y) \geq^* (\hat{x},\hat{y}) > (\hat{x}',\hat{y}') \geq^* (x',y')$. We may without loss assume that $\hat{x} = \hat{x}'$ and that $\hat{y} \succ_{\hat{x}} \hat{y}'$.

In fact, by transitivity of the agents' preferences, we may assume that no consecutive terms in a sequence that defines the transitive closure correspond to the same agents' preference. Indeed, if $z^{k-1} \geq z^k \geq z^{k+1}$ and x is the common agent to all three couples, then we would also have $z^{k-1} \geq z^{k+1}$ and so the sequence can be shortened by dropping z^k. Thus the comparisons in the sequence must then alternate between the common agent being from X and from Y.

So there exists x^1 and x^2 with

$$x^1 \succeq_{\hat{y}} \hat{x}, \quad \hat{y} \succ_{\hat{x}} \hat{y}', \quad \hat{x} \succeq_{\hat{y}'} x^2.$$

By the continuity of agent's preferences with respect to the agent (Property 1), there exists a neighborhood N of $\hat{x}$, so that $\hat{y} \succ_w \hat{y}'$ for all $w \in N$.

Now, by Property 3, $x^1 \succeq_{\hat{y}} \hat{x} \succeq_{\hat{y}'} x^2$ means that there is $\bar{x} \in N$ with $x^1 \succ_{\hat{y}} \bar{x} \succ_{\hat{y}'} x^2$. The condition $\hat{y} \neq \hat{y}'$ needed to apply Property 3 follows from $\hat{y} \succ_{\hat{x}} \hat{y}'$.

The continuity of preferences then implies that there is a neighborhood of $\bar{x}$ which consists of alternatives that are strictly worse for $\hat{y}$ than x^1, and one neighborhood with alternatives that are strictly better for $\hat{y}'$ than x^2. Thus there is a neighborhood N' of $\bar{x}$ with $x^1 \succ_{\hat{y}} w \succ_{\hat{y}'} x^2$ for all $w \in N'$.

Since $\bar{X}$ is dense and $N' \cap N$ is open and non-empty (as it contains $\bar{x}$), there exists $x^* \in \bar{X} \cap N' \cap N$. Thus,

$$x^1 \succ_{\hat{y}} x^*, \quad \hat{y} \succ_{x^*} \hat{y}', \quad x^* \succ_{\hat{y}'} x^2.$$

Denote by $L_{\succeq_{x^*}}(\hat{y}) = \{\tilde{y} \in Y : \hat{y} \succ_{x^*} \tilde{y}\}$ the set of agents in Y that are strictly worse than $\hat{y}$ for x^*, and by $U_{\succeq_{x^*}}(\hat{y}') = \{\tilde{y} \in Y : \tilde{y} \succ_{\hat{x}} \hat{y}'\}$ the set of agents that are strictly better than $\hat{y}'$. These sets are open sets by the continuity of $\succeq_{x^*}$ and non-empty as they contain, respectively, $\hat{y}'$ and $\hat{y}$.

For any $y \in Y$, by completeness of $\succeq_{x^*}$, either $y \succeq_{x^*} \hat{y} \succ_{x^*} \hat{y}'$ and thus $y \in U_{\succeq_{x^*}}(\hat{y}')$, or $\hat{y} \succ_{x^*} y$ and thus $y \in L_{\succeq_{x^*}}(\hat{y})$. Therefore, $Y = U_{\succeq_{x^*}}(\hat{y}') \cup L_{\succeq_{x^*}}(\hat{y})$. Then $U_{\succeq_{x^*}}(\hat{y}') \cap L_{\succeq_{x^*}}(\hat{y}) \neq \emptyset$, as Y is connected and therefore cannot be the union of disjoint, non-empty, open sets.

Since $\bar{Y}$ is dense in Y, and $U_{\succeq_{x^*}}(\hat{y}') \cap L_{\succeq_{x^*}}(\hat{y})$ is open, there exists $y^* \in \bar{Y} \cap U_{\succeq_{x^*}}(\hat{y}') \cap L_{\succeq_{x^*}}(\hat{y})$. So we obtain $(x^*, y^*) \in \bar{X} \times \bar{Y}$ for which

$$(x, y) \geq^* (x^1, \hat{y}) \geq (x^*, \hat{y}) > (x^*, y^*) > (x^*, \hat{y}') \geq (x^2, \hat{y}') \geq^* (x', y'),$$

completing the proof of the lemma. □

Lemma 7 shows that $\geq^*$ admits an order-dense countable subset $\bar{Z}$ in the terminology of Voorneveld and Norde (1997). Once this property is established, one can use Theorem 3.1 and Lemma 2.3 from their paper to deduce the existence of a potential u (without continuity guarantees). We include a short proof not relying on the results by Voorneveld and Norde (1997) for the reader's convenience.

Lemma 8. *Function u given by (21) represents $\geq^*$.*

Proof. If $z \geq^* z'$ then $L_{\geq^*}(z') \subseteq L_{\geq^*}(z)$, and so $u(z) \geq u(z')$. And if $z >^* z'$ then, by Lemma 7, there exists $\bar{z} \in \bar{Z}$ with $z >^* \bar{z} >^* z'$. This means that $\bar{z} \in L_{\geq^*}(z)$ while $\bar{z} \notin L_{\geq^*}(z)$ because, by acyclicity, $\bar{z} >^* z'$ implies that $z' \geq^* \bar{z}$ cannot hold. Thus $L_{\geq^*}(z') \subsetneq L_{\geq^*}(z)$, and so $u(z) > u(z')$.

Finally, by definition of $\geq^*$, if $y \succeq_x y'$ then $(x, y) \geq^* (x, y')$ and thus $u(x, y) \geq u(x, y')$. Similarly, if $y \succ_x y'$ then $(x, y) >^* (x, y')$ and thus $u(x, y) > u(x, y')$. For $\succeq_y$, the argument is analogous and thus omitted. □

The existence of an upper semicontinuous potential. Here, we prove the second statement of Theorem 4 establishing the existence of an upper semicontinuous potential.

As before, we consider a countable dense set $\bar{Z} \subseteq X \times Y$ and its enumeration $r\colon \bar{Z} \to \mathbb{N}$. We define u by

$$u(z) = (-1) \cdot \sum_{z' \in \bar{Z}\colon z' > z} \left(\frac{1}{2}\right)^{r(z')}, \tag{22}$$

and $u(z) = 0$ if there is no $z' > z$. Note that, in contrast to (21), this formula uses on $>$ but not $>^*$. We will need several technical lemmas to demonstrate that u represents $\geq^*$ and is upper semicontinuous.

Lemma 9. *If $(x_1, y) \geq (x, y) > (x, y') \geq (x_2, y')$, then there exists $z \in X$ such that $(x_1, y) > (z, y) > (z, y') > (x_2, y')$.*

Proof. First, $(x, y) > (x, y')$ and, hence, Property 1 implies that there is a neighborhood N_x of x in X such that $(z, y) > (z, y')$ for all $z \in N_x$. Second, $(x_1, y) \geq (x, y)$ and $(x, y') \geq (x_2, y')$ combined with Property 3 imply that there exists $z \in N_x$ with $(x_1, y) > (z, y)$ and $(z, y') > (x_2, y')$. Note that $y \neq y'$ follows from $(x, y) > (x, y')$, and so Property 3 applies. $\square$

Lemma 10. *If $(x, y) >^* (x', y')$ then there are neighborhoods N_x, $N_{x'}$, N_y and $N_{y'}$ of, respectively, x in X, x' in X, y in Y, and y' in Y such that $(\tilde{x}, \tilde{y}) > (\tilde{x}', \tilde{y}')$ for all $(\tilde{x}, \tilde{x}', \tilde{y}, \tilde{y}') \in N_x \times N_{x'} \times N_y \times N_{y'}$.*

Proof. By definition of $>^*$, we may without loss assume that there is a sequence (x_i, y_i), $i = 1, \dots, k$, with

$$\begin{aligned} (x, y) = (x_1, y_1) \geq (x_2, y_1) \geq \dots \\ \dots \geq (x_{t-1}, y_{t-1}) \geq (x_t, y_{t-1}) > (x_t, y_t) \geq (x_{t+1}, y_t) \geq \dots \\ \dots \geq (x_{k-1}, y_{k-1}) \geq (x_k, y_k) = (x', y'). \end{aligned} \tag{23}$$

By repeatedly applying Lemma 9, we may replace all $\geq$ in (23) with $>$. Then, for each t, we have

$$(x_{t-1}, y_{t-1}) > (x_t, y_{t-1}) > (x_t, y_t) > (x_{t+1}, y_t).$$

By Property 2, there are neighborhoods N_z^1, N_z^2, and N_z^3 of $z \in \{x_{t-1}, x_t, x_{t+1}, y_{t-1}, y_t\}$ with the property that $(\tilde{x}, \tilde{y}) > (\tilde{x}', \tilde{y})$ for all $(\tilde{x}, \tilde{x}', \tilde{y}) \in N_{x_{t-1}}^3 \times N_{x_t}^1 \times N_{y_{t-1}}^2$

and all $(\tilde{x}, \tilde{x}', \tilde{y}) \in N^3_{x_t} \times N^1_{x_{t+1}} \times N^2_{y_t}$, while $(\tilde{x}, \tilde{y}) > (\tilde{x}, \tilde{y}')$ for all $(\tilde{x}, \tilde{y}, \tilde{y}') \in N^2_{x_t} \times N^3_{y_{t-1}} \times N^1_{y_t}$.

Denote $N_z = N^1_z \cap N^2_z \cap N^3_y$ for each $z \in \cup_{t=1}^k \{x_t, y_t\}$. Then (23) holds for any selection of $\tilde{x}_t \in N_{x_t}$ and $\tilde{y}_t \in N_{y_t}$, $t = 1, \ldots, k$, completing the proof. $\square$

The second statement of Theorem 4 is the content of the following lemma.

Lemma 11. *A function u given by (22) represents $\geq^*$ and is upper semicontinuous.*

Proof. The argument that u represents $\geq^*$ repeats that in Lemma 8 and thus is omitted. We prove that u is upper semicontinuous.

Consider a sequence $z_k = (x_k, y_k)$ converging to $(x, y) = z$. Let $V_k = \{\tilde{z} \in X \times Y : \tilde{z} > z_k\}$ and $V = \{\tilde{z} \in X \times Y : \tilde{z} > z\}$. First note that $V \subseteq \cup_n \cap_{k \geq n} V_k$. Indeed, by Lemma 10, for $z' \in V$, there is a neighborhood N_z of z for which $z' > \tilde{z}$ for all $\tilde{z} \in N_z$. Hence, for k large enough, we have $z' > z_k$ since $z_k \to z$. Thus $z' \in \cap_{k \geq n} V_k$ for n large enough. We obtain

$$u(z) = (-1) \cdot \sum_{z \in \bar{Z} \cap V} \left(\frac{1}{2}\right)^{r(z)} \geq (-1) \cdot \sum_{z \in \bar{Z} \cap (\cup_n \cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)}.$$

On the other hand,

$$u(z_n) = (-1) \cdot \sum_{z \in \bar{Z} \cap V_n} \left(\frac{1}{2}\right)^{r(z)} \leq (-1) \cdot \sum_{z \in \bar{Z} \cap (\cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)}.$$

Observe that

$$\sum_{z \in \bar{Z} \cap (\cup_n \cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)} - \sum_{z \in \bar{Z} \cap (\cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)} \leq 2 \cdot \left(\frac{1}{2}\right)^{r^*_n},$$

where

$$r^*_n = \min \left\{ r(z) \ : \ z \in \bar{Z} \cap (\cup_m \cap_{k \geq m} V_k) \setminus \bar{Z} \cap (\cap_{k \geq n} V_k) \right\}$$

under the convention $\min \emptyset = +\infty$. For any decreasing sequence of sets $C_n \subseteq \mathbb{N}$ such that $\cap C_n = \emptyset$, the minimum $\min C_n$ converges to $+\infty$ as n gets large. Therefore, $r^*_n \to +\infty$ as $n \to \infty$ because r takes values in $\mathbb{N}$ and is injective. We obtain that

$$\begin{aligned} u(z) &\geq \sum_{z \in \bar{Z} \cap (\cup_n \cap_{k \geq n} V_k)} (-1) \left(\frac{1}{2}\right)^{r(z)} \\ &= \lim_{n \to \infty} \sum_{z \in \bar{Z} \cap (\cap_{k \geq n} V_k)} (-1) \left(\frac{1}{2}\right)^{r(z)} \\ &\geq \limsup_{n \to \infty} u(z_n), \end{aligned}$$

and thus u is upper semicontinuous. $\square$