

# STABILITY AND MEDIAN RATIONALIZABILITY FOR AGGREGATE MATCHINGS

ABSTRACT. We develop the theory of stability for aggregate matchings used in empirical studies and establish fundamental properties of stable matchings including the result that the set of stable matchings is a non-empty, complete, and distributive lattice. Aggregate matchings are relevant as matching data in revealed preference theory: We present a result on rationalizing a matching data as the median stable matching.

## 1. INTRODUCTION

Following the seminal work of Gale and Shapley (1962), an extensive literature has developed regarding matching markets with non-transferable utility. This literature assumes that there are agent-specific preferences, and studies the existence of *stable matchings* in which each agent prefers her assigned partner to the outside option of being unmatched, and there are no pairs of agents who would like to match with each other rather than keeping their assigned partners.

In this paper, we develop the theory of stability for aggregate matchings, which we interpret as a dataset of matchings, and consider its testable implications. In an aggregate matching, agents are classified into *types* according to their observable characteristics: for example age, employment status, gender, or race. The aggregate matching is then a table, a matrix, that records how many women of each type are matched with men of each type. Our notion of matching data is more general than just deterministic matchings and includes randomized matchings of agents to objects (a probabilistic assignment of children to schools, for example, as in Abdulkadiroğlu et al. (2005a) and Abdulkadiroğlu et al. (2005b)) as well. Therefore, our results are also applicable to randomized matching environments.

We assume that preferences are at the type-specific level, rather than the individual-level. In our setting, accommodating individual-level preferences would entail losing all empirical bite. Thus we adopt instead the assumption that the definition of types already carries all the information needed to determine agents' preferences. This differs from the approach in empirical applications of matching theory, which assume transferable utilities (see, e.g., Choo and Siow (2006) or Galichon and Salanie (2009)), but allow for heterogeneous preferences at the individual-level. When our results are taken to data, one needs to introduce an appropriate error structure that allows for individual-level randomness but without individual-level

preference heterogeneity. One way of doing this is via measurement error, as in the classical revealed preference theory for consumption (Varian, 1985; Echenique et al., 2011). One illustration can be found in Echenique et al. (2013a).

We establish fundamental results on the set of stable matchings in Theorem 1. First of all, there exists a stable matching. Then we introduce two partial orders on matchings using preferences of the two sides of the market and show that for every set of stable matchings there exists the greatest lower bound and the least upper bound with respect to the two partial orders in the set of stable matchings, i.e., the set of stable matchings endowed with either partial order is a *complete lattice*. Furthermore, we show an additional structural result that this lattice is *distributive*.<sup>1</sup> Using the lattice structure on the set of stable matchings, we show that there exists a *median stable matching*, which assigns each agent the median outcome among all stable matching outcomes (Corollary 1). Theorem 1 also provides other structural results including “polarity of interests,” saying that a stable matching is more preferred for men if and only if it is less preferred for women. Furthermore, the outcomes in two stable matchings are always comparable using the partial orders for an agent. The “rural hospitals theorem,” which states that for each type of agent the number of partners in any two stable matchings are the same, also holds.

Once stability of matchings is understood, we turn to the main question in the paper: when we interpret an aggregate matching as a table of matching data, and when we do not have access to the agents’ preferences, when can we say that the data is consistent with the theory? This is a revealed preference question that has been addressed by Echenique, Lee, Shum, and Yenmez (2013b) (see also Echenique (2008) and Chambers and Echenique (2016)). Here we focus on a particular selection of stable matchings—the *median* stable matching, which is defined for markets with deterministic outcomes. We provide a sufficient condition under which a matching is rationalizable as the median stable matching (Theorem 3). We show that the condition is not necessary in Appendix C.

Theorem 1 on the structure of stable matchings generalizes well-known results in the literature to allow for randomization, aggregate matchings, and continuum of agents (see Roth and Sotomayor (1990) for a review). Proposition 1 establishes the existence of generalized median stable matchings in our environment. For earlier existence results of median stable matchings see Teo and Sethuraman (1998); Fleiner (2003); Klaus and Klijn (2006b, 2010); Schwarz and Yenmez (2011); Chen et al. (2016). Revealed preference studies include Echenique (2008); Echenique et al. (2013b); Chambers and Echenique (2015); Demuynck and Salman (2020). The works of Echenique et al. (2013b) and Demuynck and Salman (2020) take aggregate matchings as the source of data, but none of these papers considers the question of median-rationalizability.

---

<sup>1</sup>The definition of distributive lattices is given in Section 2.

## 2. PRELIMINARY DEFINITIONS

**2.1. Lattice theory.** If  $S$  is a set and  $\leq$  is a partial order on  $S$ , we say that the pair  $(S, \leq)$  is a **partially ordered set**. We say that  $x, y \in S$  are **comparable** if  $x \leq y$  or  $y \leq x$ . A partially ordered set  $(S, \leq)$  is a **lattice** if, for every  $x, y \in S$ , the least upper bound and the greatest lower bound of  $\{x, y\}$  with respect to the partial order  $\leq$  exist in  $S$ . We denote the least upper bound of  $\{x, y\}$  by  $x \vee y$ ; and the greatest lower bound of  $\{x, y\}$  by  $x \wedge y$ . Similarly, if for every  $S' \subseteq S$ , the least upper bound and the greatest lower bound of  $S'$  exist in  $S$ , the lattice  $(S, \leq)$  is called **complete**.

We say that a lattice  $(S, \leq)$  is **distributive** if the following holds for all  $x, y, z \in S$ :

- $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ , and
- $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ .

**2.2. Graph theory.** An (undirected) **graph** is a pair  $G = (V, L)$ , where  $V$  is a set and  $L$  is a subset of  $V \times V$ . A **path** in  $G$  is a sequence  $p = \langle v_0, \dots, v_N \rangle$  such that for  $n \in \{0, \dots, N-1\}$ ,  $(v_n, v_{n+1}) \in L$ . We write  $v \in p$  to denote that  $v$  is a vertex in  $p$ . A path  $\langle v_0, \dots, v_N \rangle$  **connects** the vertices  $v_0$  and  $v_N$ . A path  $\langle v_0, \dots, v_N \rangle$  is **minimal** if there is no proper subsequence of  $\langle v_0, \dots, v_N \rangle$  which is also a path connecting the vertices  $v_0$  and  $v_N$ . The **length** of path  $\langle v_0, \dots, v_N \rangle$  is  $N$ .

A **cycle** in  $G$  is a path  $c = \langle v_0, \dots, v_N \rangle$  with  $v_0 = v_N$ . A cycle is **minimal** if for any two vertices  $v_n$  and  $v_{n'}$  in  $c$ , the paths in  $c$  from  $v_n$  to  $v_{n'}$  and from  $v_{n'}$  to  $v_n$  are distinct and minimal. We call  $v$  and  $w$  **adjacent in**  $c$  if there is  $n$  such that  $v_n = v$  and  $v_{n+1} = w$  or  $v_n = w$  and  $v_{n+1} = v$ . If  $c$  and  $c'$  are two cycles, and there is a path from a vertex of  $c$  to a vertex of  $c'$ , then we say that  $c$  and  $c'$  are **connected**.

## 3. MODEL

The primitives of the model are represented by a four-tuple  $\langle M, W, P, K \rangle$ , where

- $M$  and  $W$  are finite and disjoint sets of, respectively **types of men**, and **types of women**.
- $P$  is a **preference profile**: a list of preferences  $P_m$  for every type- $m$  man and  $P_w$  for every type- $w$  woman. Each  $P_m$  is a linear order over  $W \cup \{w_0\}$ , and each  $P_w$  is a linear order over  $M \cup \{m_0\}$ . Here,  $w_0$  and  $m_0$  represent the outside option of being unmatched. The weak order associated with  $P_m$  is denoted by  $R_m$  and the weak order associated with  $P_w$  is denoted by  $R_w$ .
- $K$  is a list of non-negative real numbers  $K_m$  for each  $m \in M$  and  $K_w$  for each  $w \in W$ . There are  $K_m$  type- $m$  men and  $K_w$  type- $w$  women.

We enumerate  $M$  as  $\{m_1, \dots, m_{|M|}\}$  and  $W$  as  $\{w_1, \dots, w_{|W|}\}$ . A **matching** is a  $|M| \times |W|$  matrix  $X = (x_{m_i, w_j})$  such that  $x_{m_i, w_j} \in \mathbf{R}_+$ ,  $\sum_j x_{m_i, w_j} \leq K_{m_i}$  for all  $i$ , and  $\sum_i x_{m_i, w_j} \leq K_{w_j}$

for all  $j$ .<sup>2</sup> We denote the mass of single  $m$ -agents in  $X$  by  $x_{m,w_0}$ , and denote the mass of single  $w$ -agents in  $X$  by  $x_{m_0,w}$ .

The  $i^{th}$  row of a matching  $X$  is denoted by  $X_{m_i,\cdot}$ , and the  $j^{th}$  column by  $X_{\cdot,w_j}$ . When it is not ambiguous, we write  $X_{m_i}$ , or  $X_i$ , for  $X_{m_i,\cdot}$ , and  $X_{w_j}$ , or  $X_j$ , for  $X_{\cdot,w_j}$ . Similarly, we use  $x_{i,j}$  for  $x_{m_i,w_j}$ .

Our solution concept is stability:

**Definition 1.** A matching  $X$  is **individually rational** if  $x_{m,w} > 0$  implies that  $w P_m w_0$  and  $m P_w m_0$ .

A pair  $(m, w)$  is a **blocking pair for**  $X$  if there exist  $m' \in M \cup \{m_0\}$  and  $w' \in W \cup \{w_0\}$  such that  $m P_w m'$ ,  $w P_m w'$ ,  $x_{m,w'} > 0$ , and  $x_{m',w} > 0$ .

A matching  $X$  is **stable** if it is individually rational and there are no blocking pairs for  $X$ .

We denote by  $S(M, W, P, K)$  the set of all stable matchings in  $\langle M, W, P, K \rangle$ .

Individual rationality states that if  $x_{m,w} > 0$ , then for type- $m$  men type- $w$  women are more preferred than being unmatched and for type- $w$  women type- $m$  men are more preferred than being unmatched. A pair  $(m, w)$  is a blocking pair if type- $m$  men have a positive measure of partners who are less preferred than type- $w$  women and type- $w$  women have a positive measure of partners who are less preferred than type- $m$  men.

Two special cases of our model are worth emphasizing: The model of **random matching** is obtained when  $K_m = 1$  for all  $m$ , and each  $K_w$  is a natural number. The interpretation of random matching is that men are “students” and women are “schools.” Students are assigned to schools at random, and each school  $w$  has  $K_w$  seats available for students. In real-life school choice, the randomization often results from indifferences in schools’ preferences over students (Abdulkadiroğlu et al., 2005b): Matching theory often requires strict preferences, so a random “priority order” is produced for the schools in order to break indifferences. Random matchings arise in many other situations as well, because random assignment is often a basic consequence of fairness considerations (if two children want the same toy, fairness demands that they flip a coin to determine who gets it).

The second model is that of **integral complete matching**, where all  $K_m$  and  $K_w$  are natural numbers, and all entries of a matching  $X$  are natural numbers. The interpretation is that there are  $K_m$  type- $m$  men and  $K_w$  type- $w$  women, and that a matching  $X$  exhibits in  $x_{m,w}$  how many type  $m$  men matched to type- $w$  women. This special case captures actual observations in marriage models (see, for example, Echenique et al. (2013a)). We observe that men and women are partitioned into types according to their observable characteristics (age, income, education, etc.); and we are given a table showing how many type- $m$  men

---

<sup>2</sup> $\mathbf{R}_+$  denotes the set of non-negative real numbers.

married type- $w$  women. These observations are essentially “flow” observations (marriages in a given year), so the integral complete matchings do not have any single agents.

#### 4. STRUCTURE OF STABLE MATCHINGS

In this section, we study the structure of stable matchings. For each  $m \in M$ , define

$$\mathcal{X}_m = \left\{ x \in \mathbf{R}_+^{|W|} : \sum_{1 \leq j \leq |W|} x_j \leq K_m \right\},$$

and a partial order  $\leq_m$  on  $\mathcal{X}_m$  by

$$y \leq_m x \quad \text{iff} \quad \forall w \in W \cup \{w_0\}, \quad \sum_{\substack{0 \leq j \leq |W| \\ w_j R_m w}} y_j \leq \sum_{\substack{0 \leq j \leq |W| \\ w_j R_m w}} x_j,$$

where  $x_0 = K_m - \sum_{1 \leq j \leq |W|} x_j$ , and  $y_0 = K_w - \sum_{1 \leq j \leq |W|} y_j$ .

Note that  $\leq_m$  is defined by analogy with the first-order stochastic dominance order on probability distributions. In the case where  $K_m = 1$ , vectors  $x$  and  $y$  in  $\mathcal{X}_m$  represent probability distributions over the types of women that  $m$  may match to. In that case  $y \leq_m x$  if and only if the lottery induced by  $y$  over  $W \cup \{w_0\}$  is worse than the lottery induced by  $x$ , for any von Neumann-Morgenstern utility function that is compatible with  $R_m$ .

Letting  $\mathcal{X}_w = \left\{ x \in \mathbf{R}_+^{|M|} : \sum_{1 \leq i \leq |M|} x_i \leq K_w \right\}$ , we define  $\leq_w$  in an analogous way.

We introduce two partial orders on matchings. Suppose  $X$  and  $Y$  are matchings, then:

- $X \leq_M Y$  if, for all  $m \in M$ ,  $X_m \leq_m Y_m$
- $X \leq_W Y$  if, for all  $w \in W$ ,  $X_w \leq_w Y_w$ .

**Theorem 1.** *( $S(M, W, P, K), \leq_M$ ) and ( $S(M, W, P, K), \leq_W$ ) are nonempty, complete, and distributive lattices; in addition, for  $X, Y \in S(M, W, P, K)$*

- (1)  $X \leq_M Y$  if and only if  $Y \leq_W X$ ;
- (2) for all  $a \in M \cup W$ , either  $X_a \leq_a Y_a$  or  $Y_a \leq_a X_a$ ;
- (3) for all  $m$ ,  $\sum_{w \in W} x_{m,w} = \sum_{w \in W} y_{m,w}$  and for all  $w$ ,  $\sum_{m \in M} x_{m,w} = \sum_{m \in M} y_{m,w}$ .

Theorem 1 presents a version, adapted to our model, of the standard results on matching theory. The proof is in Appendix A. Statement (1) is a “polarity of interests” result, saying that a stable matching  $X$  is better for men if and only if it is worse for women. Statement (2) says that the outcomes in two stable matchings are always comparable for an agent: note that  $\leq_a$  is an incomplete preference relation. Statement (3) is the “rural hospitals theorem,” which says that the same number of agents of a given type are matched in any two stable matchings.

Theorem 1 also implies that there are two stable matchings,  $X^M$  and  $X^W$ , such that for all stable matchings  $X$ ,

$$X^W \leq_M X \leq_M X^M, \text{ and} \\ X^M \leq_W X \leq_W X^W.$$

We refer to  $X^M$  as the **man-optimal** (M-optimal) stable matching, and to  $X^W$  as the **woman-optimal** (W-optimal) stable matching. We also call  $X^M$  and  $X^W$  **extremal** stable matchings. A matching  $X$  is the unique stable matching if  $S(M, W, P, K) = \{X\}$ ; in this case  $X$  coincides with the man-optimal and the woman-optimal stable matchings.

Perhaps extremal stable matchings are unlikely to arise because they favor one side of the market over the other. One may instead be interested in matchings that present a compromise. The median stable matching provides such a compromise by assigning each agent a partner who is median ranked amongst all of his/her stable matching partners. Indeed in an experimental study on decentralized market, Echenique and Yariv (2012) find that the median stable matching tends to emerge among stable matchings. As such, we may want to test whether an observed matching in field data is rationalizable as a median stable matching.

It is not obvious that median stable matchings exist. We shall first show that median stable matchings exist by only considering integral matchings that have integer entries, and then present a sufficient condition for rationalizability as a median stable matching.

We consider a market  $\langle M, W, P, K \rangle$ , where all numbers  $K_m$ ,  $K_w$ , and entries of matchings  $X$  are non-negative integers. Consequently, the number of stable matchings is finite, say  $n$ . Let  $S(M, W, P, K) = \{X^1, \dots, X^n\}$  be the set of stable matchings. For each agent  $a$  we consider all stable outcomes and rank them according to  $\geq_a$ . All the outcomes are comparable by (2) of Theorem 1. More formally, let  $\{X_a^{(1)}, \dots, X_a^{(n)}\} = \{X_a^1, \dots, X_a^n\}$  and  $X_a^{(1)} \geq_a \dots \geq_a X_a^{(n)}$ . Using these ranked outcomes and (1) of Theorem 1, we construct the matrices  $Y^{(l)}$  for all  $l = 1, \dots, n$  such that  $Y_m^{(l)} = X_m^{(l)}$  and  $Y_w^{(l)} = X_w^{(n+1-l)}$ . Matrix  $Y^{(l)}$  assigns each type of men the  $l^{th}$  best outcome, and each type of women the  $l^{th}$  worst outcome among stable matching partners.

**Proposition 1.**  $Y^{(l)}$  is a stable matching.

The proof of Proposition 1 follows from the lattice structure of the set of stable matchings. It is essentially the same as the proofs of Theorem 3.2 in Klaus and Klijn (2006a) and Theorem 1 of Schwarz and Yenmez (2011).

If  $n$  is odd we term  $Y^{(\frac{n+1}{2})}$  the **median stable matching**. If  $n$  is even we refer to  $Y^{(\frac{n}{2})}$  (or  $Y^{(\frac{n}{2}+1)}$ ) as the median stable matching (of course the choice is arbitrary).

**Corollary 1.** The median stable matching exists.

## 5. TESTABLE IMPLICATIONS

Suppose that we are given a data of a matching  $X$ , and the corresponding  $M$ ,  $W$  and  $K$ . Similar to Echenique et al. (2013b), we assume that there are no “singles,” or unmatched agents, and so  $\sum_m K_m = \sum_w K_w$ , and  $\sum_w x_{m,w} = K_m$  for all  $m$ , and  $\sum_m x_{m,w} = K_w$  for all  $w$ . We call such a matching in which there are no singles **complete**. We ask when there are preferences for the different types of men and women such that  $X$  is a stable matching or an extremal stable matching (which is shown to exist in Section 4).

**Definition 2.** A matching  $X$  is **rationalizable** if there exists a preference profile  $P = ((P_m)_{m \in M}, (P_w)_{w \in W})$  such that  $X$  is a stable matching in  $\langle M, W, P, K \rangle$ . A matching  $X$  is  **$M$ -optimal ( $W$ -optimal) rationalizable** if there is a preference profile  $P$  such that  $X$  is the  $M$ -optimal ( $W$ -optimal) stable matching in  $\langle M, W, P, K \rangle$ .

We also want to know when a matching can be rationalized as the median stable matching.

**Definition 3.** An integral matching  $X$  is **median-rationalizable** if there is a preference profile  $P$  such that  $X$  is the median stable matching in  $\langle M, W, P, K \rangle$ .

To answer the questions, we define a graph  $(V, L)$ , where  $V$  consists of all the positive entries in  $X$ , and there is an edge  $(v, v') \in L$  if and only if  $v$  and  $v'$  are either in the same row or in the same column in  $X$ . More formally, for each matching  $X$  we associate a graph  $(V, L)$  defined as follows. The set of vertices  $V$  is  $\{(m, w) : m \in M, w \in W \text{ such that } x_{m,w} > 0\}$ , and an edge  $((m, w), (m', w')) \in L$  is formed for every pair of vertices  $(m, w)$  and  $(m', w')$  with  $m = m'$  or  $w = w'$ .

A previous study finds that the rationalizability of a matching depends on minimal cycles (defined in Section 2.2) of its associated graph:<sup>3</sup>

**Theorem 2** (Echenique, Lee, Shum, and Yenmez (2013b)). (1) A complete matching  $X$  is rationalizable if and only if its associated graph does not contain two connected, distinct, minimal cycles.  
 (2) A complete matching  $X$  is  $M$ -optimal ( $W$ -optimal) rationalizable if and only if its associated graph has no minimal cycle.

For example, let matching  $X$  be the matrix on the left, with three types of men and women. Its associated graph is depicted on the right:



<sup>3</sup>The result in Echenique et al. (2013b) is more general than what is stated here.

The following two minimal cycles in this graph



are connected; thus the matching is not rationalizable.

The rationalizability by stable matchings and extremal stable matchings only depends on whether each number  $x_{m,w}$  is positive or 0. However, we find that median rationalizability depends on the value of  $x_{m,w}$  as well as whether  $x_{m,w}$  is positive or zero.

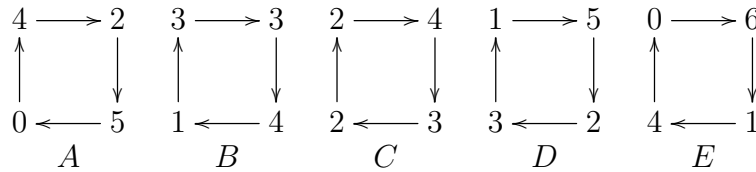
If  $\langle v_0, \dots, v_N \rangle$  is a minimal cycle, then  $N$  is an even number. We say that a minimal cycle  $c$  is **balanced** if

$$\min \{v_0, v_2, \dots, v_{N-2}\} = \min \{v_1, v_3, \dots, v_{N-1}\}.$$

**Theorem 3.** *An integral complete matching  $X$  is median rationalizable if it is rationalizable and if all minimal cycles of  $(V, L)$  are balanced.*

The proof is in Appendix B. It is worth observing that, in our proof, we construct preferences such that the resulting number of stable matchings is odd. Therefore, our results do not depend on the choice of  $Y^{(\frac{n}{2})}$  or  $Y^{(\frac{n}{2}+1)}$  as the median stable matching when  $n$  is even.

FIGURE 1. Five Stable Matchings (Matching C is Median-Stable)



To provide some intuition for this result, consider the following example:

- $M = \{m_1, m_2\}$ ,
- $W = \{w_1, w_2\}$ ,
- $k_{m_1} = 6, k_{m_2} = 5, k_{w_1} = 4, k_{w_2} = 7$ ,
- $w_2 P_{m_1} w_1, w_1 P_{m_2} w_2, m_1 P_{w_1} m_2$ , and  $m_2 P_{w_2} m_1$ ,

where each man prefers all women to being unmatched and each woman prefers all men to being unmatched. In this market, there are five stable matchings as shown in Figure 1. Note that the preferences of each type of agent is shown by an arrow, for example the arrow in the first row represents the preference ranking of  $m_1$  that type- $w_2$  woman is more preferred to type- $w_1$  woman. Therefore, the preferences are “clockwise” cyclical. These matchings differ from each other in that, starting with matching A, matchings B-E are formed by clockwise



“rotations”—that is, by dissolving two couples  $\{(m_1, w_1), (m_2, w_2)\}$  and repairing them as  $\{(m_1, w_2), (m_2, w_1)\}$ . Given the clockwise preferences, moving from matching A to E, the matchings become increasingly preferred by the men, and at the same time less preferred by the women.

Hence, matching C is the median stable matching, as the two stable matchings obtained by rotating this matching counterclockwise (i.e., A and B) are more preferred by the women, while the two obtained by rotating clockwise (i.e., D and E) are more preferred by the men.<sup>4</sup> Note also that only in the third matching is the cycle balanced: In the minimal cycle there are four vertices with entries 2, 4, 3, 2 and when we take the minimum of the entries of the odd-numbered vertices and the minimum of the entries of the even-numbered vertices for any numbering of vertices preserving the order in which they appear in the cycle we get:

$$\min\{2, 3\} = 2 = \min\{4, 2\}.$$

Therefore, the minimal cycle for matching C is balanced. The balancedness condition ensures that one can rotate an equal number of times both in the clockwise direction and in the counterclockwise direction. Furthermore, note that the balancedness condition fails for the other matchings. For example, in matching B,  $\min\{3, 4\} = 3 \neq 1 = \min\{3, 1\}$ , there the minimal cycle in matching B is not balanced.

Note that the stability and rationalizability of a matching only depend on which entries are zero or positive. Thus if an observed matching  $X$  is *canonicalized* into  $X^C$  ( $x_{m,w}^C = 0$  if  $x_{m,w} = 0$ , and  $x_{m,w}^C = 1$  if  $x_{m,w} > 0$ ), then obviously  $X$  is rationalizable if and only if  $X^C$  is rationalizable. Given the role of canonical matchings in rationalizability, the following corollary is of some interest.

**Corollary 2.** *A canonicalized matching  $X^C$  is either not rationalizable or it is median-rationalizable.*

Unfortunately, Theorem 3 provides only a sufficient condition for median-rationalizability. Corollary 2 gives a characterization of necessary conditions, but only for the case of canonical matchings. To sketch the boundaries of this result, we present two examples in Appendix C. Example C.1 shows that there are indeed matchings that are not rationalizable as median stable matchings, so Corollary 2 on canonical matchings does not extend to all aggregate matchings. Example C.2 shows that the sufficient condition in Theorem 3 is not necessary.

## APPENDIX A. PROOF OF THEOREM 1

We split up the proof in short propositions.

**Lemma 1.**  *$(\mathcal{X}_m, \leq_m)$  is a complete lattice, with a largest and smallest element.*

<sup>4</sup>In addition, one cannot rotate further clockwise beyond E, nor further counterclockwise beyond A, as one “runs out” of pairs in both cases.

*Proof.* That  $(\mathcal{X}_m, \leq_m)$  is a partially ordered set follows from the definition of  $\leq_m$ . Take a subset  $S_m$  of  $\mathcal{X}_m$ . We need to show that  $S_m$  has the least upper bound and the greatest lower bound in  $(\mathcal{X}_m, \leq_m)$  to complete the proof.

For every  $x \in \mathcal{X}_m$ , define  $x_0 = K_m - \sum_{1 \leq j \leq |W|} x_j$ . By definition, for every  $x \in \mathcal{X}_m$ ,  $\sum_{0 \leq j \leq |W|} x_j = K_m$ . Assign a new index  $\{(j) : j = 0, 1, 2, \dots, |W|\}$  over  $W \cup \{w_0\}$  such that

$$w_{(0)} P_m w_{(1)} P_m \dots P_m w_{(|W|)}$$

Let  $z_{(0)} = \sup\{x_{(0)} : x \in S_m\}$ . Define  $z_{(i)}$  inductively as follows:  $z_{(i)} = \sup\{x_{(0)} + \dots + x_{(i)} : x \in S_m\} - (z_{(0)} + \dots + z_{(i-1)})$ .

Note that  $z_{(j)} \geq 0$  and  $\sum_{0 \leq j \leq |W|} z_{(j)} = K_m$ . We define  $z$  as a  $|W|$ -vector,  $(z_1, \dots, z_{|W|})$ , with respect to the original index. Clearly,  $z_j \geq 0$  and  $\sum_{1 \leq j \leq |W|} z_j \leq K_m$ , so  $z \in \mathcal{X}_m$ .

**Claim 1.**  $z$  is the least upper bound of  $S_m$ .

*Proof.* By definition,  $\sum_{j=0}^i z_{(j)} = \sup\{x_{(0)} + \dots + x_{(i)} : x \in S_m\}$  which is greater than  $x_{(0)} + \dots + x_{(i)}$  for all  $x \in S_m$  and  $0 \leq i \leq |W|$ . Therefore,  $z \geq_m x$  for all  $x \in S_m$ , which means that  $z$  is an upper bound.

Suppose that  $z'$  is another upper bound of  $S_m$ . Therefore,  $z'_{(0)} + \dots + z'_{(i)} \geq x_{(0)} + \dots + x_{(i)}$  for all  $x \in S_m$  and  $0 \leq i \leq |W|$ . If we take the supremum of the right hand side, then we get  $z'_{(0)} + \dots + z'_{(i)} \geq \sup\{x_{(0)} + \dots + x_{(i)} : x \in S_m\}$  for all  $0 \leq i \leq |W|$ . On the other hand,  $z_{(0)} + \dots + z_{(i)} = \sup\{x_{(0)} + \dots + x_{(i)} : x \in S_m\}$ . The last two impressions imply  $z'_{(0)} + \dots + z'_{(i)} \geq z_{(0)} + \dots + z_{(i)}$  for all  $i$ , so  $z' \geq_m z$ . Thus,  $z$  is the least upper bound of  $S_m$ .  $\square$

Similarly, we can construct the greatest lower bound of  $S_m$  as follows:  $u_{(0)} = \inf\{x_{(0)} : x \in S_m\}$ . Define  $u_{(i)}$  inductively:  $u_{(i)} = \inf\{x_{(0)} + \dots + x_{(i)} : x \in S_m\} - (u_{(0)} + \dots + u_{(i-1)})$ . The proof that  $u$  is the greatest lower bound of  $S_m$  is similar to the proof that  $z$  is the least upper bound of  $S_m$ , so it is omitted.  $\square$

For each  $m$ , let the choice  $C_m$  be defined as follows. For a vector  $x \in \mathbf{R}_+^{|W|}$ , let  $C_m(x)$  be the vector in

$$\{y \in \mathcal{X}_m : y_j \leq x_j, j = 1, \dots, |W|\}$$

that is maximal for  $\leq_m$ . In other words, if  $x$  represents the quantities of type of women available for  $m$ ,  $C_m$  chooses according to  $P_m$  from the best choice downwards until filling quota  $K_m$ . Note that if  $w_0 P_m w_j$  then  $y_j = 0$ , and if  $\sum_{j:w_j P_m w_0} y_j \leq K_m$  then  $y_0 = K_m - \sum_{j:w_j P_m w_0} y_j$ . Define  $C_w$  analogously.

**Proposition 2.**  $(S(M, W, P, K), \leq_M)$  is a nonempty and complete lattice.

*Proof.* A *man pre-matching* is a matrix  $A = (a_{m,w})_{M \times W}$  such that  $a_{m,w} \in \mathbf{R}_+$  and  $\sum_w a_{m,w} \leq K_m$ . A *woman pre-matching* is a matrix  $B = (b_{m,w})_{M \times W}$  such that  $b_{m,w} \in \mathbf{R}_+$  and  $\sum_m b_{m,w} \leq K_w$ .

We consider pairs  $(A, B)$ , where  $A$  is a man prematching, and  $B$  is a woman prematching, ordered by a partial order  $\leq$ . The order  $\leq$  is defined as  $(A, B) \leq (A', B')$  if

$$\forall m \in M, \forall w \in W, (A_m \leq_m A'_m \text{ and } B'_w \leq_w B_w).$$

The order  $\leq$  is a product order of complete lattices by Lemma 1, so that the set of all pairs  $(A, B)$  ordered by  $\leq$  is a complete lattice.

We define a function  $C$ , mapping pairs  $(A, B)$  of prematchings into pairs of prematchings. Fix  $(A, B)$ : For a type- $m$  man, the number of type- $w$  women who are willing to match with  $m$  at  $B$  is  $\theta_{m,w} = K_w - \sum_{i: m_i P_w m} b_{i,w}$ . Let  $\Theta = (\theta_{m,w})$ , i.e., the  $|M| \times |W|$ -matrix such that entry  $\theta_{m,w}$  is the number of type- $w$  women who are willing to match with type  $m$  men at  $B$ . Similarly, let  $\Psi = (\psi_{m,w})$  be the  $|M| \times |W|$ -matrix for which entry  $\psi_{m,w}$  is the number of type- $m$  men who are willing to match with type- $w$  women at  $A$ . Now let  $C(A, B) = (\tilde{A}, \tilde{B})$  where  $\tilde{A}_m = C_m(\Theta_m)$  and  $\tilde{B}_w = C_w(\Psi_w)$ . Note that  $\tilde{A}$  is a man prematching and  $\tilde{B}$  is a woman prematching.

We now prove that  $C$  is isotone. Assuming that  $(A, B) \leq (A', B')$ , we prove that  $C(A, B) \leq C(A', B')$ .

Fix  $m$ . For any  $w$ , we have

$$\sum_{i: m_i P_w m} b_{i,w} \geq \sum_{i: m P_w m_i} b'_{i,w},$$

because  $B'_w \leq_w B_w$ . As a consequence, a type- $m$  man has weakly more women of each type willing to match with him in  $B'$  than in  $B$ : i.e.,  $\Theta_m \leq \Theta'_m$ . Thus  $C_m(\Theta_m) \leq_m C_m(\Theta'_m)$ . Similarly, for a type  $w$  woman,  $C_w(\Psi'_w) \leq_w C_w(\Psi_w)$ . It follows that  $C(A, B) \leq C(A', B')$ . By Tarski's fixed point theorem, there is a fixed point of  $C$ , and the set of fixed points of  $C$  is a complete lattice when ordered by  $\leq$ .

Let  $(A, B)$  be a fixed point of  $C$ , i.e.,  $(A, B) = C(A, B)$ . Assume that  $a_{m,w} > b_{m,w}$  for some  $m$  and  $w$ .  $C_w(\Psi_w) = B_w$  and  $a_{m,w} > b_{m,w}$  implies that although  $a_{m,w}$  number of type- $m$  men were available, only  $b_{m,w}$  of them are chosen by  $C_w$ . Therefore, all nonnegative entries in  $B_w$  are at least as good as  $m$  with respect to  $R_w$ . This implies that  $\theta_{m,w} = b_{m,w}$ . Since  $b_{m,w} < a_{m,w}$ , we get  $\theta_{m,w} < a_{m,w}$  which contradicts  $A_m = C_m(\Theta_m)$ . Therefore,  $a_{m,w} = b_{m,w}$  for all  $m$  and  $w$ . Hence the fixed point has the property that  $A = B$ , and they are not only a prematching but a matching as well since a man prematching that is also woman prematching is a matching.

Finally we prove that the set of fixed points of  $C$  is the set of stable matchings. More precisely,  $(A, A)$  is a fixed point of  $C$  if and only if  $A$  is a stable matching.

Suppose that a fixed point  $(A, A)$  of  $C$  is not stable. By construction of  $C_m$  and  $C_w$ ,  $A$  is individually rational. Then there is a blocking pair  $(m, w)$ . By definition of blocking pairs, there exist  $m' \in M \cup \{m_0\}$  and  $w' \in W \cup \{w_0\}$  such that  $m P_w m'$ ,  $w P_m w'$ ,  $a_{m,w'} > 0$ , and  $a_{m',w} > 0$ . Now, the number of type- $w$  women who are willing to match with  $m$  at  $A$  is

$$\theta_{m,w} = K_w - \sum_{i: m_i P_w m} a_{i,w} \geq a_{m,w} + a_{m',w} > a_{m,w},$$

as  $a_{m',w} > 0$ . But  $\theta_{m,w} > a_{m,w}$  and  $a_m = C_m(\Theta_m)$  contradicts that there is  $w'$  with  $w P_m w'$  and  $a_{m,w'} > 0$ .

Suppose that  $A$  is a stable matching. We fix  $m$  and show that  $a_m = C_m(\Theta_m)$  where  $\theta_{m,w} = K_w - \sum_{i: m_i P_w m} a_{i,w}$ . Suppose, for contradiction, that  $a_m \neq C_m(\Theta_m)$ . Denote  $w_j$  as the most preferred type of women (with respect to  $R_m$ ) such that  $a_{m,j} \neq (C_m(\Theta_m))_j$ . For all  $w_{j'}$  preferred to  $w_j$ ,  $a_{m,j'} = (C_m(\Theta_m))_{j'}$ . Thus,  $a_{m,j} > (C_m(\Theta_m))_j$  implies either  $a_{m,j} > \theta_{m,j}$  or  $\sum_{j': w_{j'} R_m w_j} a_{m,j'} > K_m$ , neither of which is possible. On the other hand,  $a_{m,j} < (C_m(\Theta_m))_j$  contradicts that  $A$  is stable. Although there are type  $w_j$  women available more than  $a_{m,j}$ , some type  $m$  men are matched to less preferred women. Then, there is  $j'$  such that  $w_j P_m w_{j'}$  and  $a_{m,j'} > 0$ , so  $(m, w_j)$  is a blocking pair. Similarly, we can show that  $a_w = C_w(\Theta_w)$ , and therefore  $(A, A) = C(A, A)$ .  $\square$

**Proposition 3.** *Suppose that  $X$  and  $Y$  are stable matchings, then  $X \leq_M Y$  if and only if  $Y \leq_W X$ .*

*Proof.* Suppose that  $X$  and  $Y$  are stable matchings such that  $X \leq_M Y$ , we show that  $Y \leq_W X$ . The other direction of the claim can be proved analogously.

Consider the construction of  $C$  in the proof of Proposition 2: Let  $\Psi$  and  $\Psi'$  correspond to the matrices for  $X$  and  $Y$  as in the proof. Since  $X$  and  $Y$  are stable matchings we have  $C(X, X) = (X, X)$  and  $C(Y, Y) = (Y, Y)$ . Since  $X \leq_m Y$  for all  $m$ ,  $\Psi'_{m,w} \leq \Psi_{m,w}$  for all  $m$  and  $w$ . Therefore,  $Y_w = C_w(\Psi'_w) \leq_w C_w(\Psi_w) = X_w$  for all  $w$ , which implies that  $Y \leq_W X$ .  $\square$

**Proposition 4.** *Suppose that  $X$  and  $Y$  are two stable matchings. Then for any men or women type  $a$ , either  $X_a \leq_a Y_a$  or  $Y_a \leq_a X_a$ . Consequently,  $X_a \vee_a Y_a = \max_{\leq_a} \{X_a, Y_a\}$  and  $X_a \wedge_a Y_a = \min_{\leq_a} \{X_a, Y_a\}$ .*

*Proof.* We only prove the first part that either  $X_a \leq_a Y_a$  or  $Y_a \leq_a X_a$ , in three steps depending on whether  $X$  and  $Y$  have integer, rational, and real entries. The second part follows immediately from the first part.

Case 1 (Integer Entries): We first start with the case when  $X$  and  $Y$  have integer entries as well as  $K_w$  and  $K_m$ . From  $\langle M, W, P, K \rangle$ , we create a many-to-one matching market (of colleges and students) as follows.

The set of types of men remains the same; interpreted as the set of colleges. A college  $m$  has a capacity of  $K_m$ . Whereas, each type- $w$  woman is split into  $K_w$  copies, all of which have the same preferences  $P_w$  over colleges; women types are interpreted as students. On the other hand,  $m$ 's preferences  $P'_m$  replace  $w$  in  $P_m$  with her copies enumerated from 1 to  $K_w$ , in increasing order. Denote  $w_j$ 's  $l^{th}$  copy by  $w_j^l$ . So  $w_j^l P'_m w_j^k$  if and only if  $k > l$ . In addition, each college has responsive preferences over groups of students. The new matching market with  $|M|$  colleges and  $\sum_w K_w$  students is a many-to-one matching market where an outcome for a college is a group of students, and an outcome for a student is either a college or being single.

Now, we construct a new matching,  $X'$ , in the new market from  $X$ .<sup>5</sup> It is enough to describe the matches of students in  $X'$ . Rank woman  $w$ 's outcomes in  $X$  in decreasing order according to her preference  $P_w$ . Let the  $l^{th}$  copy of  $w_j$ ,  $w_j^l$ , match to the  $l^{th}$  highest outcome of  $w_j$  in  $X$ . Similarly, construct  $Y'$  from  $Y$ .

We claim that  $X'$  and  $Y'$  are stable matchings in the new market. Suppose for contradiction that  $X'$  is not a stable matching. Since  $X'$  is individually rational by construction, there exists a blocking pair  $(m, w_j^l)$ . This means that  $w_j^l$ 's match is worse than  $m$ . Similarly,  $m$ 's match includes a student worse than  $w_j^l$ : this agent cannot be  $w_j^k$  where  $k < l$  by definition of  $P'_m$ , and it cannot be  $w_j^k$  where  $k > l$  because by construction  $w_j^k$ 's match is worse than  $w_j^l$ 's match with respect to  $P_{w_j}$ . Hence, one of  $m$ 's matches is worse than  $w_j^l$ , and not a copy of  $w_j$ . This means that  $(m, w_j)$  forms a blocking pair in  $X$ : A contradiction to the stability of  $X$ . Therefore,  $X'$  must be stable. Similarly,  $Y'$  is also stable.

Now, by Theorem 5.26 of Roth and Sotomayor (1990), for any college  $m$  the outcomes in  $X'$  and  $Y'$  are comparable. This means that the responsive preferences over groups of students inherited from  $P'_m$ , which is equivalent to the first order stochastic dominance, can compare the outcomes of  $m$  in these two stable matchings. Therefore,  $\leq_m$  can compare the outcomes in  $X$  and  $Y$  since  $P_m$  is a coarser order than  $P'_m$ .

An analogous argument shows that  $\leq_w$  can compare the matching outcomes in  $X$  and  $Y$ .

Case 2 (Rational Entries): Suppose for now that all entries of  $X$  and  $Y$  are rational numbers as well as  $K_w$  and  $K_m$ . Define a new matching market from  $\langle M, W, P, K \rangle$  as follows.

$M$ ,  $W$ , and  $P$  are the same. We change the capacities as follows. Find a common factor of denominators in all entries in  $X$  and  $Y$ , say  $r$ , and multiply all capacities by  $r$ . Therefore, the new market is  $\langle M, W, P, rK \rangle$ . Define  $X' = rX$  and  $Y' = rY$  with non-negative integer entries. By the argument above, for any agent  $a$ , the matchings in  $rX$  and  $rY$  can be compared with respect to  $\leq_a$  which implies that the outcomes in  $X$  and  $Y$  can also be compared.

---

<sup>5</sup>The new matching is a many-to-one matching, which is not an aggregate matching. Nevertheless, we use the aggregate matching notation to stress the relations between  $X'$  and  $X$ , and  $Y'$  and  $Y$ .

Case 3 (Real Entries): Suppose now that entries of  $X$  and  $Y$  are real numbers. We construct two sequences of matrices,  $X^{(n)}$  and  $Y^{(n)}$ , as follows:

- (1)  $X^{(n)}$  and  $Y^{(n)}$  have rational entries,
- (2)  $x_{ij}^{(n)} = 0 \iff x_{ij} = 0$  and  $y_{ij}^{(n)} = 0 \iff y_{ij} = 0$  for all  $i, j$ ,
- (3) the sum of entries in row  $i$  and column  $j$  is the same for  $X^{(n)}$  and  $Y^{(n)}$ , and
- (4)  $x_{ij}^{(n)} \rightarrow x_{ij}$  and  $y_{ij}^{(n)} \rightarrow y_{ij}$  as  $n \rightarrow \infty$  for all  $i, j$ .

By construction, stability of  $X$  and  $Y$  imply stability of  $X^{(n)}$  and  $Y^{(n)}$  for the same market with adjusted capacities. By the argument above, for each type  $a$ ,  $x_a^{(n)}$  and  $y_a^{(n)}$  can be compared with respect to  $\leq_a$ . Take a subsequence  $\{n_l\}$  such that the ordering is the same for all entries. Therefore,  $x_a^{(n_l)} \leq_a y_a^{(n_l)}$  for all  $l$  or  $y_a^{(n_l)} \leq_a x_a^{(n_l)}$  for all  $l$ . By taking  $l$  to  $\infty$  we get that either  $x_a \leq_a y_a$  in the former case, or  $y_a \leq_a x_a$  in the latter since (4) implies  $x_{i,j}^{(n)} \rightarrow x_{i,j}$  and  $y_{i,j}^{(n)} \rightarrow y_{i,j}$  as  $n \rightarrow \infty$  for all  $i, j$ .  $\square$

We have shown that  $(S(M, W, P, K), \leq_M)$  is a lattice in Proposition 2. Using the proposition above, we show that the lattice,  $(S(M, W, P, K), \leq_M)$ , is distributive.

**Proposition 5.**  $(S(M, W, P, K), \leq_M)$  is a distributive lattice.

*Proof.* Suppose that  $X, Y$ , and  $Z$  are stable matchings. We are going to prove that type  $a$  has the same matching in  $X \wedge (Y \vee Z)$  and  $(X \wedge Y) \vee (X \wedge Z)$  for all agent types  $a$ :

$$\begin{aligned}
 (X \wedge (Y \vee Z))_a &= \min\{X_a, \max\{Y_a, Z_a\}\} \\
 &= \max\{\min\{X_a, Y_a\}, \min\{X_a, Z_a\}\} \\
 &= \max\{(X \wedge Y)_a, (X \wedge Z)_a\} \\
 &= ((X \wedge Y) \vee (X \wedge Z))_a,
 \end{aligned}$$

where min and max operators are defined with respect to  $\leq_a$  and where we repeatedly use Proposition 4.

The proof that  $X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z)$  is analogous.  $\square$

**Proposition 6.** Suppose that  $X$  and  $Y$  are stable matchings. Then  $\sum_{1 \leq j \leq |W|} x_{ij} = \sum_{1 \leq j \leq |W|} y_{ij}$  for all  $i$  and similarly  $\sum_{1 \leq i \leq |M|} x_{ij} = \sum_{1 \leq i \leq |M|} y_{ij}$  for all  $j$ .

*Proof.* The proof has the same structure as in the proof of Proposition 4: it has three steps depending on whether  $X$  and  $Y$  have integer, rational, and real entries.

Case 1 (Integer Entries): We first start with the case when  $X$  and  $Y$  have integer entries as well as  $K_m$  for all  $m$  and  $K_w$  for all  $w$ . From  $\langle M, W, P, K \rangle$ , we create a many-to-one matching market and also stable matchings  $X'$  and  $Y'$  in this new market as in the proof of Proposition 4.

Now, by Theorem 5.12 of Roth and Sotomayor (1990), the set of positions filled for any  $m$  in  $X'$  and  $Y'$  are the same. Therefore,  $\sum_{1 \leq j \leq |W|} x_{ij} = \sum_{1 \leq j \leq |W|} y_{ij}$  for all  $i$ . Similarly,  $\sum_{1 \leq i \leq |M|} x_{ij} = \sum_{1 \leq i \leq |M|} y_{ij}$  for all  $j$ .

Case 2 (Rational Entries): Suppose for now that all entries of  $X$  and  $Y$  are rational numbers as well as  $K_a$  for all  $a$ . Define a new matching market from  $\langle M, W, P, K \rangle$  as follows.

$M$ ,  $W$ , and  $P$  are the same. We change the capacities as follows. Find a common factor of denominators in all entries in  $X$  and  $Y$ , say  $r$ , and multiply all capacities by  $r$ . Therefore, the new market is  $\langle M, W, P, rK \rangle$ . Define  $X' = rX$  and  $Y' = rY$  with non-negative integer entries.

By the argument above,  $\sum_{1 \leq j \leq |W|} rx_{ij} = \sum_{1 \leq j \leq |W|} ry_{ij}$  for all  $i$  and  $\sum_{1 \leq i \leq |M|} rx_{ij} = \sum_{1 \leq i \leq |M|} ry_{ij}$  for all  $j$ . The conclusion follows.

Case 3 (Real Entries): Suppose now that entries of  $X$  and  $Y$  are real numbers. We construct two sequences of matrices,  $X^{(n)}$  and  $Y^{(n)}$ , as follows:

- (1)  $X^{(n)}$  and  $Y^{(n)}$  have rational entries,
- (2)  $x_{ij}^{(n)} = 0 \iff x_{ij} = 0$  and  $y_{ij}^{(n)} = 0 \iff y_{ij} = 0$  for all  $i, j$ ,
- (3) the sum of entries in row  $i$  and column  $j$  is the same for  $X^{(n)}$  and  $Y^{(n)}$ , and
- (4)  $x_{ij}^{(n)} \rightarrow x_{ij}$  and  $y_{ij}^{(n)} \rightarrow y_{ij}$  as  $n \rightarrow \infty$  for all  $i, j$ .

By construction, stability of  $X$  and  $Y$  imply stability of  $X^{(n)}$  and  $Y^{(n)}$  for the same market with adjusted capacities. By the argument above,  $\sum_{1 \leq j \leq |W|} x_{ij}^{(n)} = \sum_{1 \leq j \leq |W|} y_{ij}^{(n)}$  for all  $i$  and

$\sum_{1 \leq i \leq |M|} x_{ij}^{(n)} = \sum_{1 \leq i \leq |M|} y_{ij}^{(n)}$ . If we take the limit of these equalities as  $n \rightarrow \infty$ , we get the desired equalities.  $\square$

## APPENDIX B. PROOF OF THEOREM 3

Let  $X$  be a rationalizable integral complete matching. Suppose that all cycles of the associated graph  $(V, L)$  are balanced. Direct the edges of  $(V, L)$  such that each cycle is oriented as follows: if  $\langle v_0, \dots, v_N \rangle$  is a cycle, then the edge  $(v_n, v_{n+1}) \in L$  is oriented such that  $d(v_{n+1}, v_n) = 1$ , which we denote by  $v_n \rightarrow v_{n+1}$ . For each path  $\langle v_0, \dots, v_N \rangle$ , direct the edges in a similar way. If the matching  $X$  is rationalizable, then such an orientation of the edges exists and defines a rationalizing preference profile  $P$  (Theorem 2, Part 1). The rationalizing preferences have the property that if  $x_{i,j} = 0$  and  $x_{i',j'} > 0$  then  $w_{j'} P_{m_i} w_j$  if  $i = i'$ , and  $m_{i'} P_{w_j} m_i$  if  $j = j'$ .

First, if  $X$  has no cycles, then it is rationalizable as the unique stable matching (Theorem 2, Part 2), so there is nothing to prove, as a unique stable matching is also the median stable matching. Suppose then that  $X$  has at least one cycle  $c = \langle v_0, \dots, v_N \rangle$ . Enumerate the

vertexes of the cycle such that  $v_n \rightarrow v_{n+1}$  in the orientation (directed graph) of  $(V, L)$  above, and  $v_0$  lies in the same row as  $v_1$ . Let

$$\Theta = \min \{v_0, v_2, \dots, v_{N-2}\} = \min \{v_1, v_3, \dots, v_{N-1}\}.$$

Let  $\mathcal{E}$  be the set of all  $|M| \times |W|$  matrices  $E$  of integer numbers such that

- $e_{i,j} = 0$  if  $(i, j)$  is not in the cycle  $c$ ;
- for all  $i$  and  $j$ ,  $\sum_h e_{i,h} = 0$  and  $\sum_l e_{l,j} = 0$ ; and
- $|e_{i,j}| \leq \Theta$ .

We want to make two observations about the matrices in  $\mathcal{E}$ . First,  $(X + E)_{i,j} > 0 \Rightarrow x_{i,j} > 0$ , so  $X + E$  is a stable matching for all  $E \in \mathcal{E}$ . Second,  $E \in \mathcal{E}$  if and only if  $-E \in \mathcal{E}$ ; and  $E \in \mathcal{E}$  is such that  $X_i \leq_{m_i} (X + E)_i$  if and only if  $(X - E)_i \leq_{m_i} X_i$ . Similarly,  $E \in \mathcal{E}$  is such that  $X_j \leq_{w_j} (X + E)_j$  if and only if  $(X - E)_j \leq_{w_j} X_j$ .

We need to prove that any stable matching  $Y \neq X$  in the resulting market  $\langle M, W, P, K \rangle$  must be obtained from  $X$  through matrices in  $\mathcal{E}$ : Then  $X$  is a median stable matching.

First, we prove that  $y_{i,j} \neq x_{i,j}$  only if  $x_{i,j}$  is a vertex in a minimal cycle of  $X$ . The number of single agents of each type is the same in  $X$  as in  $Y$  (Proposition 6; in this case it is zero, as  $X$  has no single agents). So, if  $x_{i,j} < y_{i,j}$  then there is  $h \neq j$  and  $l \neq i$  such that  $y_{i,h} < x_{i,h}$  and  $y_{l,j} < x_{l,j}$ . Similarly, if  $x_{i,j} > y_{i,j}$  then there is  $h \neq j$  and  $l \neq i$  such that  $y_{i,h} < x_{i,h}$  and  $y_{l,j} < x_{l,j}$ . Hence, we can apply this observation repeatedly, starting from any  $(i, j)$  with  $x_{i,j} \neq y_{i,j}$  and obtain a sequence  $(i_1, j_1), \dots, (i_N, j_N)$  with  $(i_1, j_1) = (i_N, j_N)$  such that for each  $n \pmod N$ :

- (1)  $(x_{i_n, j_n} - y_{i_n, j_n})(x_{i_{n+1}, j_{n+1}} - y_{i_{n+1}, j_{n+1}}) < 0$
- (2)  $i_n \neq i_{n+1} \iff j_n = j_{n+1}$ .

**Claim 2.** For  $n = 1, \dots, N$ ,  $x_{i_n, j_n} > 0$ .

*Proof.* Suppose for contradiction that  $0 = x_{i_1, j_1} < y_{i_1, j_1}$ . Without loss of generality, assume that  $i_1 = i_2$ . By definition of the rationalization  $P$ , we have that  $w_{j_2} P_{m_{i_1}} w_{j_1}$ , as  $x_{i_1, j_1} = 0$  and  $x_{i_1, j_2} > y_{i_1, j_2} \geq 0$ . We can now show that if  $(i_n, j_n)$  and  $(i_{n+1}, j_{n+1})$  differ in  $i$ , then  $j_n$  prefers  $m_{i_{n+1}}$  to  $m_{i_n}$ ; and that if they differ in  $j$ , then  $i_n$  prefers  $w_{j_{n+1}}$  to  $w_{j_n}$ . This fact, which we prove in the next paragraph, establishes the contradiction:  $i_N \neq i_{N-1}$ , but  $m_{i_{N-1}} P_{w_{i_N}} m_{i_N}$  by definition of  $P$  and because  $0 = x_{i_N, j_N} = x_{i_1, j_1}$ .

To prove the fact, we reason by induction. We have already established that  $w_{j_2} P_{m_{i_1}} w_{j_1}$ . Suppose that  $w_{j_n} P_{m_{i_n}} w_{j_{n-1}}$ . By Property 1 of the sequence  $\langle (i_n, j_n) \rangle_{n=1}^N$ , either  $x_{i_{n-1}, j_{n-1}} > 0$  and  $x_{i_{n+1}, j_{n+1}} > 0$ ; or  $y_{i_{n-1}, j_{n-1}} > 0$  and  $y_{i_{n+1}, j_{n+1}} > 0$  (or both hold). Then the stability of  $X$  and  $Y$  implies that  $m_{i_{n+1}} P_{w_{j_n}} m_{i_n}$ . The proof for the case when  $w_{i_n} P_{w_{i_n}} m_{i_{n-1}}$  is similar.  $\square$



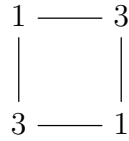
The claim implies that the sequence  $\langle v_n \rangle = \langle x_{i_n, j_n} \rangle$  is a cycle in  $(V, L)$ . Thus a stable matching  $Y$  can only differ from  $X$  in vertexes that are part of a cycle of  $(V, L)$ .

Second, we prove that  $E \equiv Y - X \in \mathcal{E}$ . We established above that  $e_{i,j} \neq 0$  only if  $x_{i,j}$  is a vertex in a cycle. We now prove that  $|e_{i,j}| \leq \Theta$ . Clearly,  $e_{i,j} \geq -x_{i,j} \geq -\Theta$ . We show that if  $e_{i,j} > 0$  then there is  $h$  such that  $e_{i,j} + e_{i,h} = 0$ .

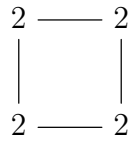
If  $e_{i,j} > -e_{i,h}$  for all  $h \neq i$  then there is  $h_1$  and  $h_2$  such that some type- $m_i$  men of type  $m_i$  who are married to women of type  $h_1$  and  $h_2$  in  $X$  are married to women of type  $w_j$  in  $Y$ . Then we can define two cycles, and  $x_{i,j}$  would be a vertex in both of them. The first cycle has  $(x_{i,j}, x_{i,h_1})$  as the first edge, and the remaining edges defined inductively, by the definition of  $\langle (i_n, j_n) \rangle$  above. The second cycle has  $(x_{i,j}, x_{i,h_2})$  as the first edge, and the remaining edges defined inductively. The resulting two cycles would be connected, which contradicts the hypothesis that  $X$  is rationalizable. So there must exist some  $h$  with  $e_{i,j} \leq -e_{i,h}$ . An analogous argument applied to  $e_{i,h}$  implies that  $e_{i,j} \geq -e_{i,h}$ ; so  $e_{i,j} = -e_{i,h}$ . Then,  $e_{i,j} \leq \Theta$ , as  $e_{i,h} \geq -\Theta$ .

## APPENDIX C. EXAMPLES

**C.1. There are matchings rationalizable, but not median-rationalizable.** A rationalizable aggregate matching that cannot be rationalized as a median stable matching:



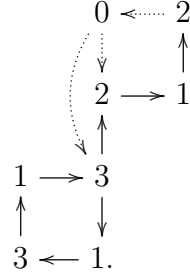
The only way to rationalize this aggregate matching is to have a cycled preference profile, i.e., if  $w_1 P_{m_1} w_2$  then  $m_2 P_{w_1} m_1$ ,  $w_2 P_{m_2} w_1$ , and  $m_1 P_{w_2} m_2$ . Similarly, if  $w_2 P_{m_1} w_1$  then the preferences of agents are cycled in the opposite direction. Regardless of the choice, all the feasible aggregate matchings are stable, so there are 5 stable aggregate matchings in total. Therefore, the median aggregate stable matching is the one where each agent is matched to both possible partners twice, which is



**C.2. Median-rationalizability condition is not necessary.** The following example shows that the sufficient condition in Theorem 3 is not necessary. Suppose that there are four types of men, three types of women, and consider an aggregate matching

$$X = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 1 & 3 & 0 \\ 3 & 1 & 0 \end{pmatrix}.$$

We choose the preferences such that the graph corresponding to  $X$  is as follows. Note that we indicate preference with an arrow, so that for example  $x_{i,j} \rightarrow x_{i,h}$  means that  $w_h P_{m_i} w_j$ .



Note that there is a cycle  $\{(m_3, w_1), (m_3, w_2), (m_4, w_2), (m_4, w_1)\}$  and that this cycle is not balanced.

By “rotating” the cycle  $\{(m_3, w_1), (m_3, w_2), (m_4, w_2), (m_4, w_1)\}$  in a clockwise direction we obtain the aggregate matching

$$\begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{pmatrix},$$

which is better for the men. However, by counterclockwise rotations we obtain the matchings

$$\begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 2 & 0 \\ 2 & 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 3 & 1 & 0 \\ 1 & 3 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 4 & 0 & 0 \\ 0 & 4 & 0 \end{pmatrix},$$

which are all better for women than  $X$ .

Now, *with the rationalization in the arrows above*, the following “joint” rotation of the cycle and the upper entries of the matchings is stable as well:

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 3 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{pmatrix}.$$

These two matchings are better for the men. In all, then, under the rationalizing preferences as in the arrows, there are 7 stable matchings, and  $X$  is the median stable matching.

There is a crucial aspect of the example that makes this possible. Note that, if we are to rotate the upper part of the graph, we need preferences to be as indicated by the arrows. In particular, we must have  $x_{2,3} \rightarrow x_{1,3}$  for the graph to be rationalizable; then, to accommodate a  $> 0$  entry in  $x_{1,2}$  after the rotation, we must have  $x_{1,3} \rightarrow x_{1,2}$ , or otherwise we would get a blocking pair  $(m_1, w_3)$ . However, positive entries in  $x_{1,3}$  and in  $x_{3,2}$  imply that we need  $x_{1,2} \rightarrow x_{3,2}$  (the long dotted arrow in the graph in order to satisfy that  $x_{1,3} \rightarrow x_{1,2}$ . Now, a modification of  $X$  that has a positive entry in  $x_{1,2}$  is only possible if we simultaneously set  $x_{3,1} = 0$ , as  $x_{3,1} \rightarrow x_{3,2}$  and  $x_{1,2} \rightarrow x_{3,2}$ . Hence the rotation of the upper side of the graph *is not feasible under any of the modification of  $X$  that improve the matches of the women*.

There is, therefore, an asymmetry in the graph that allows us to offset the unbalancedness of the number of men and women in the cycle.

## REFERENCES

- ABDULKADIROĞLU, A., P. PATHAK, A. ROTH, AND T. SÖNMEZ (2005a): “The Boston Public School Match,” *American Economic Review*, 95, 368–371.
- ABDULKADIROĞLU, A., P. A. PATHAK, AND A. E. ROTH (2005b): “The New York City High School Match,” *The American Economic Review*, 95, pp. 364–367.
- CHAMBERS, C. P. AND F. ECHENIQUE (2015): “The Core Matchings of Markets with Transfers,” *American Economic Journal: Microeconomics*, 7, 144–64.
- (2016): *Revealed preference theory*, vol. 56, Cambridge University Press.
- CHEN, P., M. EGESDAL, M. PYCIA, AND M. B. YENMEZ (2016): “Median stable matchings in two-sided markets,” *Games and Economic Behavior*, 97, 64–69.
- CHOO, E. AND A. SIOW (2006): “Who Marries Whom and Why,” *Journal of Political Economy*, 114, 175–201.
- DEMUYNCK, T. AND U. SALMAN (2020): “On the Revealed Preference Analysis of Stable Aggregate Matchings,” Working Paper, ECARES.
- ECHENIQUE, F. (2008): “What Matchings Can Be Stable? The Testable Implications of Matching Theory,” *Mathematics of Operations Research*, 33, 757–768.
- ECHENIQUE, F., S. LEE, AND M. SHUM (2011): “The money pump as a measure of revealed preference violations,” *Journal of Political Economy*, 119, 1201–1223.
- (2013a): “Partial identification in two-sided matching models,” in *Structural Econometric Models*, Emerald Group Publishing Limited, 117–139.
- ECHENIQUE, F., S. LEE, M. SHUM, AND M. B. YENMEZ (2013b): “The Revealed Preference Theory of Stable and Extremal Stable Matchings,” *Econometrica*, 81, 153–171.

- ECHENIQUE, F. AND L. YARIV (2012): “An experimental study of decentralized matching,” Mimeo, California Institute of Technology.
- FLEINER, T. (2003): “A Fixed-Point Approach to Stable Matchings and Some Applications,” *Mathematics of Operations Research*, 28, 103–126.
- GALE, D. AND L. S. SHAPLEY (1962): “College Admissions and the Stability of Marriage,” *The American Mathematical Monthly*, 69, 9–15.
- GALICHON, A. AND B. SALANIE (2009): “Matching with Trade-offs: Revealed Preferences over Competing Characteristics,” Mimeo., Ecole Polytechnique.
- KLAUS, B. AND F. KLIJN (2006a): “Median stable matching for college admissions,” *International Journal of Game Theory*, 34, 1–11.
- (2006b): “Procedurally fair and stable matching,” *Economic Theory*, 27, 431–447.
- (2010): “Smith and Rawls share a room: stability and medians,” *Social Choice and Welfare*, 35, 647–667.
- ROTH, A. AND M. SOTOMAYOR (1990): *Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis*, vol. 18 of *Econometric Society Monographs*, Cambridge University Press, Cambridge England.
- SCHWARZ, M. AND M. B. YENMEZ (2011): “Median Stable Matching for Markets with Wages,” *Journal of Economic Theory*, 146, 619–637.
- TEO, C.-P. AND J. SETHURAMAN (1998): “The geometry of fractional stable matchings and its applications,” *Mathematics of Operations Research*, 23, 874–891.
- VARIAN, H. R. (1985): “Non-Parametric Analysis of Optimizing Behavior with Measurement Error,” *Journal of Econometrics*, 30, 445–458.