On the Convergence of Swap Dynamics to Pareto-Optimal Matchings

Felix Brandt and Anaëlle Wilczynski

Technical University of Munich, 80538 Munich, Germany
{brandtf,wilczyns}@in.tum.de

Abstract. We study whether Pareto-optimal stable matchings can be reached via pairwise swaps in one-to-one matching markets with initial assignments. We consider housing markets, marriage markets, and roommate markets as well as three different notions of swap rationality. Our main results are as follows. While it can be efficiently determined whether a Pareto-optimal stable matching can be reached when defining swaps via blocking pairs, checking whether this is the case for *all* such sequences is computationally intractable. When defining swaps such that all involved agents need to be better off, even deciding whether a Pareto-optimal stable matching can be reached via *some* sequence is intractable. This confirms and extends a conjecture made by Damamme et al. (2015), who have furthermore shown that convergence to a Pareto-optimal matching is guaranteed in housing markets with single-peaked preferences. We show that in marriage and roommate markets, single-peakedness is not sufficient for this to hold, but the stronger restriction of one-dimensional Euclidean preferences is.

1 Introduction

One-to-one matchings, where individuals are matched with resources or other individuals, are omnipresent in everyday life. Examples include the job market, assigning offices to workers, pairing students in working groups, and online dating. The formal study of matching procedures is fascinating because it leads to challenging mathematical and algorithmic problems while being of immediate practical interest [see, e.g., 22, 24].

One typically distinguishes between three different types of abstract one-to-one matching settings. In *housing markets* [28], each agent is matched with an object (usually referred to as a house). In *marriage markets* [16], agents are partitioned into two groups—say, males and females—and each member of one group is matched with an agent from the other group. Finally, in *roommate markets* [16], all agents belong to the same group and each agent is matched with another agent. In many applications, it is reasonable to assume that there is an initial assignment because agents already live in a house, are engaged in a relationship, and are employed by a company [see, e.g., 1, 25]. Under these assumptions, an important question is whether sequences of individual agreements between small groups of agents can lead to socially optimal outcomes. In this paper, we focus on atomic agreements which require the least coordination: *pairwise swaps*.

In general, we consider three different types of individual rationality for pairwise swaps. In housing markets, there is only one meaningful notion of swap rationality: two agents will only exchange objects if both of them are better off. By contrast, when matching agents with each other, one could require that all four agents involved in a swap or only two of them are better off. The latter requirement allows for two kinds of swap rationality: two agents who exchange their match are better off (e.g., a company and its subsidiary exchange employees without asking their consent) or two agents who decide to form a new pair are better off (e.g., two lovers leave their current partners to be together).

Social optimality in settings with ordinal preferences like that of matching markets is measured in terms of Pareto-optimality. We therefore study whether there exists a sequence of pairwise swaps that results in a Pareto-optimal matching that does not allow for further swaps (and hence is called stable). Whenever all sequences of pairwise swaps are of this kind, we say that the given type of swap dynamics converges.

It turns out that in all three types of matching markets and all three notions of swap rationality, it may not be possible to reach a Pareto-optimal stable matching from the initial assignment. We prove that deciding whether this is the case is NP-hard for two types of swap rationality while it can be solved in polynomial time for swaps based on blocking pairs. However, in the latter case, checking convergence is co-NP-hard. On the other hand, we show that when preferences are one-dimensional Euclidean—a natural but demanding restriction—swap dynamics for two types of swap rationality will always converge.

2 Related Work

Damamme et al. [14] investigated the dynamics of individually rational pairwise swaps in housing markets, where two agents are better off by exchanging their objects. Recently, variants of this problem that further restrict the agents' interactions using underlying graph structures have been examined [17, 20, 27].

In marriage and roommate markets, most of the literature focuses on deviations based on blocking pairs, where two agents decide to leave their old partners in order to be matched with each other. Blocking pairs are best known for their role in the definition of stability [16], but some papers also studied the dynamics of blocking pair swaps [2, 26]. The notion of exchange stability, where two agents agree to exchange their partners has been investigated in both roommate markets [5, 11] and marriage markets [12]. We consider both types of swaps, i.e., blocking pair swaps and exchange rational swaps, but focus on the study of dynamics that reach Pareto-optimal matchings.

In contrast to our definition of Pareto-optimality, some papers on swap dynamics have investigated matchings that are Pareto-optimal *among all reachable matchings* [7, 17]. Other types of dynamics that have been considered in matching markets include pairwise swaps without local rationality constraints [7], Pareto improvements [8, 25], local dynamics based on underlying graphs [18, 19], and exchanges among more than two agents [7, 9].

Perhaps closest to our work is a result by Damamme et al. [14] who proved that swap dynamics always converge to a Pareto-optimal matching in housing markets when the preferences of the agents are single-peaked. However, they left open the computational problem of deciding whether a Pareto-optimal stable matching can be reached for unrestricted preferences and conjectured this problem to be intractable. We solve this problem and extend it to marriage and roommate markets. Moreover, we prove that their convergence result for housing markets under single-peaked preferences does not extend to marriage and roommate markets, but can be restored when restricting preferences even further.

3 The Model

We are given a set N of agents $\{1, \dots, n\}$ and a set O of objects $\{a, b, \dots\}$ such that $|N| = |O| = n$. Each agent $i \in N$ has strict ordinal preferences, represented by a linear order $\succ_i$, over a set A_i of alternatives to be matched with. In the matching markets we consider, A_i is either a subset of the set of agents N or the set of all objects O. A tuple of preference relations $\succ = (\succ_1, \dots, \succ_n)$ is called a *preference profile*.

We consider two restricted preference domains in this article: single-peaked preferences [10] and its subdomain of one-dimensional Euclidean preferences [13]. A preference profile $\succ$ is *single-peaked* if there exists a linear order $>$ over the alternatives in $A := \bigcup_{i \in N} A_i$ such that for each agent i in N and each triple of alternatives $x, y, z \in A_i$ with $x > y > z$ or $z > y > x$, $x \succ_i y$ implies $y \succ_i z$. A preference profile $\succ$ is *one-dimensional Euclidean (1-Euclidean)* if there exists an embedding $E : N \cup O \to \mathbb{R}$ on the real line such that for every agent $i \in N$ and any two alternatives $x, y \in A_i$, $x \succ_i y$ iff $|E(i) - E(x)| < |E(i) - E(y)|$.

One-dimensional Euclidean preferences form a subdomain of single-peaked preferences because every 1-Euclidean preference profile is singled-peaked for the linear order $>$ given by $x > y$ iff $E(x) > E(y)$. However, a single-peaked preference profile may not be 1-Euclidean, as illustrated in the example below.

Example 1. Consider an instance with four agents. Each agent $i \in N$ has preferences over the same set of alternatives $A_i = O = \{a, b, c, d\}$.

$$\begin{aligned} 1 &: \; a \succ b \succ c \succ d \\ 2 &: \; d \succ c \succ b \succ a \\ 3 &: \; b \succ c \succ d \succ a \\ 4 &: \; c \succ b \succ a \succ d \end{aligned}$$

Observe that this preference profile is single-peaked only w.r.t. the linear order $a < b < c < d$ (or its reverse order) because of the preferences of Agents 1 and 2. Suppose that this preference profile is 1-Euclidean w.r.t. an embedding E on the real line. Then, without loss of generality, we can assume that $E(a) < E(b) < E(c) < E(d)$. Since Agent 3 prefers b to c and Agent 4 prefers c to

b, it must hold that $E(3) < E(4)$. However, $d \succ_3 a$, therefore $E(d) - E(3) < E(3) - E(a)$. It follows that $E(d) - E(4) < E(4) - E(a)$, implying that Agent 4 prefers d to a, a contradiction.

While assuming that all agents have 1-Euclidean preferences certainly represents a strong restriction, there are nevertheless some applications where this assumption is not unreasonable. For example, in job markets, preferences could be 1-Euclidean because employees prefer one workplace to another if it is closer to their home, or when pairing workers in offices with a joint thermostat, workers could prefer co-workers whose most preferred room temperature is closer to their own.

3.1 Matching Markets

In this article, we are considering three different settings where the goal is to match the agents either with objects—like in housing markets—or with other agents—like in marriage or roommate markets. In all cases, we assume that there is an initial matching. More formally,

- a *housing market* consists of a preference profile where $A_i = O$ for all $i \in N$, and an initial endowment given as a bijection $\mu : N \to O$,
- a *marriage market* consists of a preference profile where $N = W \cup M$ with $W \cap M = \emptyset$, $A_i = M$ for all $i \in W$ and $A_i = W$ for all $i \in M$, and an initial matching given as a bijection $\mu : W \to M$, and
- a *roommate market* consists of a preference profile with even n and $A_i = N \setminus \{i\}$ for all $i \in N$, and an initial matching given as an involution $\mu : N \to N$ such that $\mu(i) \neq i$ for all $i \in N$.

In marriage markets, we will sometimes denote the inverse function μ^{-1} of matching μ by μ for the sake of simplicity.

When allowing for indifferences as well as unacceptabilities in the preferences, the three settings form a hierarchy: housing markets are marriage markets where the "objects" are indifferent between all agents, and marriage markets are roommate markets where all agents of the same type are considered unacceptable. In this paper, however, we do not make either assumption and therefore these inclusion relationships do not hold.

The key question studied in this paper is whether Pareto-optimal matchings can be reached from the initial matching via local modifications. A matching is *Pareto-optimal* if there is no other matching μ' such that for every agent $i \in N$, $\mu'(i) \succeq_i \mu(i)$ and for at least one agent $j \in N$, $\mu'(j) \succ_j \mu(j)$.

3.2 Rational Swaps

We study sequences of matchings in which two pairs of the current matching are permuted. More formally, we assume that a swap w.r.t. two agents (i, j) transforms a matching μ into a matching μ' where agents i and j have exchanged

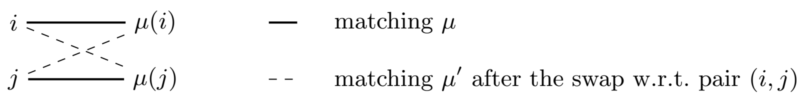


Fig. 1. Two matchings μ and μ' that differ in one swap.

their matches, i.e., $\mu'(i) = \mu(j)$ and $\mu'(j) = \mu(i)$, while the rest of the matching remains unchanged, i.e., $\mu'(k) = \mu(k)$ for every $k \notin \{i, j, \mu(i), \mu(j)\}$ (see Fig. 1).

We furthermore require these swaps to be *rational* in the sense that they result from an agreement among agents, and thus make the agents involved in the agreement better off.

The most natural notion of rationality in our definition of a swap is exchange-rationality, which requires that the two agents who exchange their matches are better off [5]. A swap w.r.t. agents (i, j) from matching μ is *exchange rational (ER)* if the agents who exchange their matches are better off, i.e.,

$$\mu(j) \succ_i \mu(i) \text{ and } \mu(i) \succ_j \mu(j). \qquad (ER\text{-swap})$$

Exchange-rationality is the only meaningful notion of swap rationality in housing markets because only one side of the market has preferences. However, several notions of rationality emerge in marriage and roommate markets, where agents are matched with each other. One could demand that only two of the agents who agree to form a new pair need to be better off. This notion of rational swaps is based on the classic idea of *blocking pairs*, which forms the basis of the standard notion of stability [16]. A swap w.r.t. agents (i, j) from matching μ between agents is *blocking pair (BP) rational* if one of the new pairs in μ' forms a blocking pair, where both agents are better off, i.e.,

$$\big[\mu(j) \succ_i \mu(i) \text{ and } i \succ_{\mu(j)} j\big] \quad \text{or} \quad \big[\mu(i) \succ_j \mu(j) \text{ and } j \succ_{\mu(i)} i\big]. \qquad (BP\text{-swap})$$

We usually refer to a *BP*-swap by mentioning the associated blocking pair ($(i, \mu(j))$ or $(j, \mu(i))$). The old partners of the blocking pair are also assumed to be matched together.[1]

Finally, in marriage and roommate markets, a stronger notion of rationality is that of a *fully rational swap*, which makes all four involved agents better off. A swap w.r.t. agents (i, j) from matching μ is *fully rational (FR)* if all four agents involved in the swap are better off, i.e.,

$$\mu(j) \succ_i \mu(i), \quad \mu(i) \succ_j \mu(j), \quad j \succ_{\mu(i)} i, \text{ and } \quad i \succ_{\mu(j)} j. \qquad (FR\text{-swap})$$

Note that for marriage and roommate markets, an *FR*-swap w.r.t. pair of agents (i, j) from a matching μ is an *ER*-swap w.r.t. pair (i, j) or $(\mu(i), \mu(j))$ and also a *BP*-swap w.r.t blocking pair $(i, \mu(j))$ or $(j, \mu(i))$. We thus obtain the following implications:

[1] Once the old partners are alone, they have an incentive to form a new pair. Roth and Vande Vate [26] therefore decompose *BP*-swaps into two steps. We do not explicitly consider these steps in order to always maintain a perfect matching [cf. 23].

$$BP\text{-swap} \quad \Leftarrow \quad FR\text{-swap} \quad \Rightarrow \quad ER\text{-swap}$$

The different types of swap rationality are illustrated in the following example.

Example 2. Consider a roommate market with six agents. The preferences of the agents are given below, where the initial assignment is marked with frames.

$$\begin{array}{rl} 1: & 4 \succ \boxed{3} \succ 6 \succ 5 \succ 2 \\ 2: & 3 \succ 1 \succ \boxed{4} \succ 6 \succ 5 \\ 3: & 6 \succ 2 \succ \boxed{1} \succ 5 \succ 4 \\ 4: & 5 \succ 1 \succ 3 \succ \boxed{2} \succ 6 \\ 5: & 2 \succ \boxed{6} \succ 4 \succ 1 \succ 3 \\ 6: & 4 \succ 3 \succ 1 \succ 2 \succ \boxed{5} \end{array}$$

The swap w.r.t. pair $(1, 2)$, which matches Agent 1 with Agent 4 and Agent 2 with Agent 3, is an *FR*-swap because every involved agent is better off. Hence, this is also an *ER*-swap for pair $(1, 2)$ or $(3, 4)$ because they both prefer to exchange their partner. It is also a *BP*-swap for blocking pair $(2, 3)$ or $(1, 4)$ because they both prefer to be together than with their current partner.

The swap w.r.t. pair $(1, 6)$ is a *BP*-swap for blocking pair $(3, 6)$ because Agent 3, the old partner of Agent 1, prefers to be with Agent 6, as well as Agent 6 who prefers 3 to her old partner 5. This is not an *ER*-swap (and hence not an *FR*-swap) because neither the agents in pair $(1, 6)$ nor in pair $(3, 5)$ want to exchange their partners.

The swap w.r.t. pair $(4, 6)$ is an *ER*-swap for $(4, 6)$ because Agent 4 prefers the current partner of 6, i.e., Agent 5, to her current partner and 6 prefers the current partner of 4, i.e., Agent 2, to her current partner. This is not a *BP*-swap (and hence not an *FR*-swap) because it matches Agent 4 with Agent 5, who prefers to stay with her current partner, and Agent 6 with Agent 2, who prefers to stay with her current partner.

Stability can now be defined according to the different notions of rational swaps. A matching μ is *σ-stable*, for $\sigma \in \{FR, ER, BP\}$, if no σ-swap can be performed from matching μ. A sequence of σ-swaps, for $\sigma \in \{FR, ER, BP\}$, corresponds to a sequence of matchings $(\mu^0, \mu^1, \ldots, \mu^r)$ such that a σ-swap transforms each matching μ^t into matching μ^{t+1} for every $0 \leq t < r$. Then, matching μ is *σ-reachable* from initial matching μ^0 if there exists a sequence of σ-swaps $(\mu^0, \mu^1, \ldots, \mu^r)$ such that $\mu^r = \mu$. When the context is clear, we omit σ and the initial matching μ^0.

A *σ-dynamics* is defined according to initial matching μ^0 and a type σ of rational swaps. The σ-dynamics is *finite* if all associated sequences of σ-swaps terminate in a σ-stable matching, and it is said to *converge* if it is finite for every initial matching μ^0.

In this article, we consider the following two decision problems related to the convergence of dynamics to a Pareto-optimal matching.

∃-σ-ParetoSequence / ∀-σ-ParetoSequence	
Input:	Matching market, type σ of rational swaps
Question:	Does there exist a sequence of σ-swaps terminating in a Pareto-optimal σ-stable matching? / Do all sequences of σ-swaps terminate in a Pareto-optimal σ-stable matching?

In order to tackle these questions, we also study the stability and convergence properties of the considered dynamics in the three types of matching markets.

4 Exchange Rational Swaps

In housing markets, every *ER*-swap represents a Pareto improvement. Hence, since the number of agents and objects is finite, *ER*-dynamics always converges and the existence of *ER*-stable matchings is guaranteed (simply because every Pareto-optimal matching happens to be *ER*-stable). However, it may be impossible to reach a Pareto-optimal matching from a given matching by only applying *ER*-swaps.

Proposition 1. *ER-dynamics may not converge to a Pareto-optimal matching in housing markets.*

Proof. Consider a housing market with three agents. The preferences of the agents are given below, where the initial assignment is marked with frames.

$$1: \enclose{circle}{a} \succ \boxed{\text{b}} \succ c$$
$$2: \enclose{circle}{b} \succ \boxed{\text{c}} \succ a$$
$$3: \enclose{circle}{c} \succ \boxed{\text{a}} \succ b$$

Observe that no *ER*-swap is possible in this instance, therefore the initial matching (framed objects) is the unique *ER*-reachable matching. However, there exists a unique Pareto-optimal matching (circled objects), and this matching is different from the initial one. □

Nevertheless, Damamme et al. [14] have shown that *ER*-dynamics always converges to a Pareto-optimal matching in housing markets when the agents' preferences are single-peaked.

In marriage and roommate markets, an *ER*-stable matching may not exist, even for single-peaked preferences (Cechlárova [11] and Alcalde [5] provide counterexamples). However, it turns out that, when restricting preferences even further to 1-Euclidean preferences, an *ER*-stable matching always exists, and, moreover, the convergence to such a matching is guaranteed.

Proposition 2. *ER-dynamics always converges in marriage and roommate markets for 1-Euclidean preferences.*

Proof. Denote by $E : N \to \mathbb{R}$ the embedding of the agents on the real line such that their preferences are 1-Euclidean w.r.t. this embedding. Define as a potential function $f : \mu \to \mathbb{R}$ the function which assigns to each matching the sum of the distances on the real line between all the assigned pairs in the matching, i.e., $f(\mu) = \sum_{(i,j) s.t. \mu(i)=j} |E(i) - E(j)|$. Now consider a sequence of *ER*-swaps given by the sequence of matchings $(\mu^0, \mu^1, \ldots, \mu^r)$. Between each μ^t and μ^{t+1}, with $0 \le t < r$, an *ER*-swap is performed, say w.r.t. pair (i, j) of agents. By definition of an *ER*-swap, agents i and j prefer to exchange their partners in μ^t, and thus, $\mu^t(j) \succ_i \mu^t(i)$ and $\mu^t(i) \succ_j \mu^t(j)$. This implies that $|E(i) - E(\mu^t(j))| < |E(i) - E(\mu^t(i))|$ and $|E(j) - E(\mu^t(i))| < |E(j) - E(\mu^t(j))|$. But i and $\mu^t(j)$ are matched in μ^{t+1}, as well as j and $\mu^t(i)$. Since the rest of the pairs remains unchanged between μ^t and μ^{t+1}, we get that $f(\mu^{t+1}) < f(\mu^t)$. Because the number of different matchings is finite, we can conclude that *ER*-dynamics always converges. □

In general, an *ER*-stable matching may not be Pareto-optimal, and thus the convergence to a Pareto-optimal matching is not guaranteed even when an *ER*-stable matching exists (note that determining whether there exists an *ER*-stable matching is NP-hard in both marriage and roommate markets [11, 12]).

Proposition 3. *ER-dynamics may not converge to a Pareto-optimal matching, in marriage and roommate markets, even when an ER-stable matching exists.*

Proof. Consider a marriage market with three women and three men. The preferences are given below and the initial assignment is marked with frames.

$$\begin{array}{ll} w_1 : \enclose{circle}{m_1} \succ \boxed{m_2} \succ m_3 & \qquad m_1 : \enclose{circle}{w_1} \succ \boxed{w_3} \succ w_2 \\ w_2 : \enclose{circle}{m_2} \succ \boxed{m_3} \succ m_1 & \qquad m_2 : \enclose{circle}{w_2} \succ \boxed{w_1} \succ w_3 \\ w_3 : \enclose{circle}{m_3} \succ \boxed{m_1} \succ m_2 & \qquad m_3 : \enclose{circle}{w_3} \succ \boxed{w_2} \succ w_1 \end{array}$$

No *ER*-swap is possible from the initial matching (framed agents), therefore the initial matching is the unique *ER*-reachable matching. However, there is another matching (circled agents) which is the unique Pareto-optimal matching.

Now, consider a roommate market with six agents. Preferences of the agents are given below, where the initial partner of each agent is marked with frames and "[...]" denotes an arbitrary order over the rest of the agents.

$$\begin{array}{ll} 1: \enclose{circle}{3} \succ \boxed{2} \succ [\ldots] & \qquad 4: \enclose{circle}{6} \succ \boxed{3} \succ [\ldots] \\ 2: \enclose{circle}{5} \succ \boxed{1} \succ [\ldots] & \qquad 5: \enclose{circle}{2} \succ \boxed{6} \succ [\ldots] \\ 3: \enclose{circle}{1} \succ \boxed{4} \succ [\ldots] & \qquad 6: \enclose{circle}{4} \succ \boxed{5} \succ [\ldots] \end{array}$$

No *ER*-swap is possible from the initial matching (framed agents), thus the initial matching is the unique *ER*-reachable matching. However, there is another matching (circled agents) which is the unique Pareto-optimal matching. □

Note that the above preference profiles are not 1-Euclidean. In fact, they are not even single-peaked. Again, more positive results can be obtained by restricting the domain of admissible preferences.

Proposition 4. *Every ER-stable matching is Pareto-optimal when preferences are single-peaked in marriage and roommate markets.*

Proof. Let μ be an *ER*-stable matching. For any two agents i and j (in N for roommate markets, or both in either W or M for marriage markets) it holds that $\mu(i) \succ_i \mu(j)$ or $\mu(j) \succ_j \mu(i)$. Suppose there is another matching μ' such that $\mu'(i) \succeq_i \mu(i)$ for every $i \in N$ and there exists $j \in N$ such that $\mu'(j) \succ_j \mu(j)$. Then, there exists a Pareto improving cycle from μ to μ' along agents $(n_1, \ldots, n_k)$ such that each agent n_i, $1 \leq i \leq k$, is matched in μ' with agent $\mu(n_{(i \bmod k)+1})$. For marriage markets, the agents in $(n_1, \ldots, n_k)$ are restricted by definition to only one side of the market, but it impacts both sides since the agents exchange agents of the other side. But there is no problem of preferences of the matched agents because no agent is worse off in μ' compared to μ. The same holds for roommate markets. Since μ is *ER*-stable, it holds that $k > 2$. However, for single-peaked preferences, one can prove, by following the same proof by induction as Damamme et al. [14], that a Pareto improving cycle of any length cannot occur, contradicting the fact that μ is Pareto dominated. □

Propositions 2 and 4 allow us to conclude that sequences of *ER*-swaps will always terminate in Pareto-optimal matchings when preferences are 1-Euclidean.

Corollary 1. *ER-dynamics always converges to a Pareto-optimal matching in marriage and roommate markets for 1-Euclidean preferences.*

An interesting computational question is whether, given a preference profile and an initial assignment, a Pareto-optimal matching can be reached via *ER*-swaps. In the context of housing markets, the complexity of this question was mentioned as an open problem by Damamme et al. [14]. It turns out that this problem is computationally intractable for all kinds of matching markets considered in this paper.

Theorem 1. ∃-ER-ParetoSequence *is* NP*-hard in housing, marriage, and roommate markets.*

The proof is omitted due to space restrictions.

5 Blocking Pair Swaps

BP-swaps cannot occur in housing markets because objects can never be better off. We therefore focus on matching markets that match agents with each other in this section.

By definition of a blocking pair, any *BP*-stable matching is Pareto-optimal. Moreover, a *BP*-stable matching always exists in marriage markets by the Deferred Acceptance algorithm [16]. However, the convergence to such a state is not guaranteed, even for single-peaked preferences [23]. Nevertheless, there always exists a sequence of *BP*-swaps leading to a stable matching [26].[2] In roommate markets, even the existence of a *BP*-stable matching is not guaranteed [16], and

[2] Assuming that the old partners also form a new pair does not alter this result.

actually this is the case even for single-peaked preferences. Nevertheless, checking the existence of a stable matching in a roommate market can be done in polynomial time [21], and there always exists a sequence of *BP*-swaps leading to a stable matching when there exists one [15]. Therefore, by combining these facts with the observation that every *BP*-stable matching is Pareto-optimal, we get the following corollary.

Corollary 2. ∃-BP-ParetoSequence *is solvable in polynomial time in marriage and roommate markets.*

However, in general, determining whether all sequences of *BP*-swaps terminate in a Pareto-optimal matching, i.e., checking convergence of *BP*-dynamics to a Pareto-optimal matching, is hard. This is due to the hardness of checking the existence of a cycle in *BP*-dynamics.

Theorem 2. *Determining whether BP-dynamics can cycle in marriage and roommate markets is* NP*-hard.*

The proof is omitted due to space restrictions.

Corollary 3. ∀-BP-ParetoSequence *is* co-NP*-hard in marriage and roommate markets.*

Nevertheless, when preferences are 1-Euclidean, we can always reach a stable matching thanks to *BP*-dynamics in both settings.

Indeed, a marriage market under 1-Euclidean preferences is a particular case of a *correlated two-sided market* [4] where all the possible pairs are *globally ranked* [see, also 3]. In such a correlated market, the preferences of the agents are induced from the global order by taking into account the order over the pairs to which they belong. It has been proved that *BP*-dynamics always converges in correlated marriage markets [4]. Moreover, it is easy to see that from 1-Euclidean preferences, a global ranking over all possible pairs can be extracted by sorting all pairs according to the distance on the real line between the two partners.[3] Therefore, we obtain the following corollary.

Corollary 4. *BP-dynamics always converges in marriage markets for 1-Euclidean preferences.*

In roommate markets, there always exists a unique *BP*-stable matching under 1-Euclidean preferences [6]. We further prove that convergence to this matching is guaranteed using a potential function argument.

Proposition 5. *BP-dynamics always converges in roommate markets for 1-Euclidean preferences.*

[3] The presence of a global ranking over all possible pairs does not imply that preferences are 1-Euclidean. Consider for instance, in roommate markets, the following preference profile: $1 : 2 \succ 3 \succ 4$, $2 : 1 \succ 4 \succ 3$, $3 : 4 \succ 1 \succ 2$, $4 : 3 \succ 2 \succ 1$.

Proof. Denote by $E : N \to \mathbb{R}$ the embedding of the agents on the real line such that their preferences are 1-Euclidean w.r.t. this embedding. Let $d(\mu)$ be the $n/2$-vector of distances in E of all the different pairs in μ, i.e., $d(\mu) = (|E(i) - E(j)|)_{i,j \text{ s.t. } \mu(i)=j}$. Now consider a sequence of BP-swaps given by the following sequence of matchings $(\mu^0, \mu^1, \dots, \mu^r)$. Then, between each pair of matchings μ^t and μ^{t+1} with $0 \leq t < r$, a BP-swap is performed, say w.r.t. blocking pair (i, j) of agents. By definition of a BP-swap, agents i and j prefer to be together than being with their partner in μ^t, so $j = \mu^{t+1}(i) \succ_i \mu^t(i)$ and $i = \mu^{t+1}(j) \succ_j \mu^t(j)$, which implies that $|E(i) - E(j)| < |E(i) - E(\mu^t(i))|$ and $|E(i) - E(j)| < |E(j) - E(\mu^t(j))|$. Therefore, $(|E(i) - E(j)|, |E(\mu^t(i)) - E(\mu^t(j))|)$ is lexicographically smaller than $(|E(i) - E(\mu^t(i))|, |E(j) - E(\mu^t(j))|)$. Since the rest of the pairs remains unchanged between μ^t and μ^{t+1}, $d(\mu^{t+1})$ is lexicographically strictly smaller than $d(\mu^t)$. Because the number of different matchings is finite, we conclude that BP-dynamics always converges. $\square$

Since every BP-stable matching is Pareto-optimal, Corollary 4 and Proposition 5 imply the following corollary.

Corollary 5. *BP-dynamics always converges to a Pareto-optimal matching in marriage and roommate markets for 1-Euclidean preferences.*

6 Fully Rational Swaps

Just as in the case of *ER*-swaps and housing markets, *FR*-swaps always represent Pareto improvements because all involved agents are strictly better off after the swap. Hence, *FR*-stable matchings are guaranteed to exist because every Pareto-optimal matching is *FR*-stable and *FR*-dynamics always converges because the number of agents is finite.

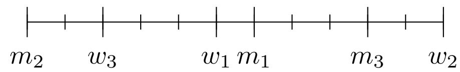
In Section 4, we have shown that *ER*-dynamics always converges to a Pareto-optimal matching when the preferences of the agents are 1-Euclidean. It turns out that this does not hold for *FR*-dynamics.

Proposition 6. *A sequence of FR-swaps may not converge to a Pareto-optimal matching in marriage and roommate markets, even for 1-Euclidean preferences.*

Proof. Consider a marriage market with three women and three men. The preferences are given below, where the initial assignment is marked with frames.

$$w_1 : \textcircled{m_1} \succ \boxed{m_3} \succ m_2 \qquad m_1 : \textcircled{w_1} \succ \boxed{w_3} \succ w_2$$
$$w_2 : \textcircled{m_3} \succ m_1 \succ \boxed{m_2} \qquad m_2 : \textcircled{w_3} \succ w_1 \succ \boxed{w_2}$$
$$w_3 : \textcircled{m_2} \succ \boxed{m_1} \succ m_3 \qquad m_3 : \textcircled{w_2} \succ \boxed{w_1} \succ w_3$$

The initial matching is the only reachable matching, because no *FR*-swap is possible in this matching. However, there is another matching (circled agents) which is not reachable but which Pareto dominates this only reachable matching. The preferences are 1-Euclidean w.r.t. the following embedding on the real line.



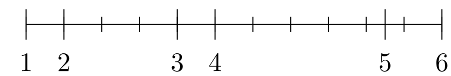
Now, consider a roommate market with six agents. The preferences of the agents are given below, where the initial assignment is marked with frames.

$1 : \textcircled{2} \succ 3 \succ 4 \succ 5 \succ \boxed{6}$ $\qquad$ $4 : \textcircled{3} \succ 2 \succ \boxed{5} \succ 1 \succ 6$

$2 : \textcircled{1} \succ \boxed{3} \succ 4 \succ 5 \succ 6$ $\qquad$ $5 : \textcircled{6} \succ \boxed{4} \succ 3 \succ 2 \succ 1$

$3 : \textcircled{4} \succ \boxed{2} \succ 1 \succ 5 \succ 6$ $\qquad$ $6 : \textcircled{5} \succ 4 \succ 3 \succ 2 \succ \boxed{1}$

The initial matching is the only reachable matching, because there is no *FR*-swap from this matching. However, there is another matching (circled agents) which is not reachable but which Pareto dominates this only reachable matching. The preferences are 1-Euclidean w.r.t. the following embedding on the real line.



□

The proof of Theorem 1 only dealt with instances in which *FR*-swaps are identical to *ER*-swaps. We thus immediately obtain hardness of ∃-FR-ParetoSequence.

Theorem 3. ∃-FR-ParetoSequence *is* NP*-hard in marriage and roommate markets.*

7 Conclusion

We have studied the properties of different dynamics of rational swaps in matching markets with initial assignments and, in particular, the question of convergence to a Pareto-optimal matching. For all considered settings, the dynamics may not terminate in a Pareto-optimal matching because *(i)* there is no stable matching, *(ii)* the dynamics does not converge, or *(iii)* the stable matching that is eventually reached is not Pareto-optimal. An overview of our results is given in Table 1.

Computationally, determining whether there exists a sequence of rational swaps terminating in a Pareto-optimal matching is NP-hard for fully rational swaps and exchange rational swaps in all matching markets (Th. 1 and Th. 3). For swaps based on blocking pairs, this problem can be solved efficiently (Cor. 2). However, the convergence to a Pareto-optimal matching is co-NP-hard to decide (Cor. 3). Since determining the existence of a sequence of fully rational or exchange rational swaps terminating in a Pareto-optimal matching is already hard, we did not investigate the complexity of convergence to a Pareto-optimal matching (which means that all sequences terminate) for these swaps. We leave it as an open problem that we conjecture to be hard.

The convergence to a Pareto-optimal matching in housing markets for exchange rational dynamics and single-peaked preferences [14] does not hold for more general settings where the "objects" are agents who have preferences. However, this convergence is guaranteed under 1-Euclidean preferences in marriage

Table 1. Summary of the results on the existence of a stable matching (Stable), the guarantee of convergence (Conv) and the guarantee of convergence to a Pareto-optimal matching (Pareto) for the three different matching markets under study, according to different types of rational swaps and under different preference domains (General, single-peaked (SP), and 1-Euclidean (1-D)). The only meaningful type of rational swaps in housing markets are exchange-rational swaps; hence, the empty spaces.

Market	Prefs	Exchange Rational Swaps			Blocking Pair Swaps			Fully Rational Swaps		
		Stable	Conv	Pareto	Stable	Conv	Pareto	Stable	Conv	Pareto
Housing	General	✓	✓	– (Prop. 1)						
	SP	✓	✓	✓ [14]						
	1-D	✓	✓	✓						
Marriage	General	–	–	–	✓ [16]	–	–	✓	✓	–
	SP	– [11]	–	–	✓	–	–	✓	✓	–
	1-D	✓	✓	✓ (Prop. 2)	✓	✓	✓ (Cor. 4)	✓	✓	– (Prop. 6)
Roommate	General	–	–	–	– [16]	–	–	✓	✓	–
	SP	– [5]	–	–	–	–	–	✓	✓	–
	1-D	✓	✓	✓ (Prop. 2)	✓	✓	✓ (Prop. 5)	✓	✓	– (Prop. 6)

and roommate markets. Hence, the generalization of this convergence result to more general settings requires more structure in the preferences.

A natural extension of this work would be to study meaningful dynamics for hedonic games, where agents form groups consisting of more than two agents.

References

1. Abdulkadiroğlu, A., Sönmez, T.: House allocation with existing tenants. Journal of Economic Theory **88**(2), 233–260 (1999)
2. Abeledo, H., Rothblum, U.G.: Paths to marriage stability. Discrete Applied Mathematics **63**(1), 1–12 (1995)
3. Abraham, D.J., Leravi, A., Manlove, D.F., O'Malley, G.: The stable roommates problem with globally-ranked pairs. In: Proc. of 3rd WINE. LNCS, vol. 4858, pp. 431–444. Springer (2007)
4. Ackermann, H., Goldberg, P.W., Mirrokni, V.S., Röglin, H., Vöcking, B.: Uncoordinated two-sided matching markets. SIAM Journal on Computing **40**(1), 92–106 (2011)
5. Alcalde, J.: Exchange-proofness or divorce-proofness? Stability in one-sided matching markets. Review of Economic Design **1**(1), 275–287 (1994)
6. Arkin, E.M., Bae, S.W., Efrat, A., Okamoto, K., Mitchell, J.S.B., Polishchuk, V.: Geometric stable roommates. Information Processing Letters **109**(4), 219–224 (2009)
7. Aziz, H.: Algorithms for Pareto optimal exchange with bounded exchange cycles. Operations Research Letters **47**(5), 344–347 (2019)
8. Aziz, H., Brandt, F., Harrenstein, P.: Pareto optimality in coalition formation. Games and Economic Behavior **82**, 562–581 (2013)

9. Aziz, H., Goldwaser, A.: Coalitional exchange stable matchings in marriage and roommate markets (extended abstract). In: Proc. of 16th AAMAS Conference. pp. 1475–1477. IFAAMAS (2017)
10. Black, D.: On the rationale of group decision-making. Journal of Political Economy **56**(1), 23–34 (1948)
11. Cechlárová, K.: On the complexity of exchange-stable roommates. Discrete Applied Mathematics **116**(3), 279–287 (2002)
12. Cechlárová, K., Manlove, D.F.: The exchange-stable marriage problem. Discrete Applied Mathematics **152**(1–3), 109–122 (2005)
13. Coombs, C.H.: Psychological scaling without a unit of measurement. Psychological Review **57**(3), 145–158 (1950)
14. Damamme, A., Beynier, A., Chevaleyre, Y., Maudet, N.: The power of swap deals in distributed resource allocation. In: Proc. of 14th AAMAS Conference. pp. 625–633. IFAAMAS (2015)
15. Diamantoudi, E., Miyagawa, E., Xue, L.: Random paths to stability in the roommate problem. Games and Economic Behavior **48**(1), 18–28 (2004)
16. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. The American Mathematical Monthly **69**(1), 9–15 (1962)
17. Gourvès, L., Lesca, J., Wilczynski, A.: Object allocation via swaps along a social network. In: Proc. of 26th IJCAI. pp. 213–219. IJCAI (2017)
18. Hoefer, M.: Local matching dynamics in social networks. In: Proc. of 38th ICALP. LNCS, vol. 6756, pp. 113–124. Springer (2011)
19. Hoefer, M., Vaz, D., Wagner, L.: Dynamics in matching and coalition formation games with structural constraints. Artificial Intelligence **262**, 222–247 (2018)
20. Huang, S., Xiao, M.: Object reachability via swaps along a line. In: Proc. of 33rd AAAI Conference. pp. 2037–2044. AAAI Press (2019)
21. Irving, R.W.: An efficient algorithm for the “stable roommates” problem. Journal of Algorithms **6**(4), 577–595 (1985)
22. Klaus, B., Manlove, D.F., Rossi, F.: Matching under preferences. In: Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D. (eds.) Handbook of Computational Social Choice, chap. 14. Cambridge University Press (2016)
23. Knuth, D.E.: Mariages stables. Les Presses de l'Université de Montréal (1976)
24. Manlove, D.F.: Algorithmics of Matching Under Preferences. World Scientific Publishing Company (2013)
25. Morrill, T.: The roommates problem revisited. Journal of Economic Theory **145**(5), 1739–1756 (2010)
26. Roth, A.E., Vande Vate, J.H.: Random paths to stability in two-sided matching. Econometrica **58**(6), 1475–1480 (1990)
27. Saffidine, A., Wilczynski, A.: Constrained swap dynamics over a social network in distributed resource reallocation. In: Proc. of 11th SAGT. LNCS, vol. 11059, pp. 213–225. Springer (2018)
28. Shapley, L.S., Scarf, H.: On cores and indivisibility. Journal of Mathematical Economics **1**(1), 23–37 (1974)