

A pessimist's approach to one-sided matching

Tom Demeulemeester^{a,b,*}, Dries Goossens^{b,c}, Ben Hermans^a, Roel Leus^a

^a*Research Center for Operations Research & Statistics, KU Leuven, Belgium*

^b*Department of Business Informatics and Operations Management, Ghent University, Belgium*

^c*Corelab CVAMO, FlandersMake@UGent, Belgium*

Abstract

Inspired by real-world applications such as the assignment of pupils to schools or the allocation of social housing, the one-sided matching problem studies how a set of agents can be assigned to a set of objects when the agents have preferences over the objects, but not vice versa. For fairness reasons, most mechanisms use randomness, and therefore result in a probabilistic assignment. We study the problem of decomposing these probabilistic assignments into a weighted sum of ex-post (Pareto-)efficient matchings, while maximizing the worst-case number of assigned agents. This decomposition preserves all the assignments' desirable properties, most notably strategy-proofness. Next to discussing the complexity of the problem, we obtain tight lower and upper bounds on the optimal worst-case number of assigned agents. Moreover, we propose two alternative column generation frameworks for the introduced problem, which prove to be capable of finding decompositions with the theoretically best possible worst-case number of assigned agents, both for randomly generated data, and for real-world school choice data from the Belgian cities Antwerp and Ghent. Lastly, the proposed column generation frameworks are inherently flexible, and can therefore also be applied to settings where other ex-post criteria are desirable, or to find decompositions that satisfy other worst-case measures.

Keywords: Assignment, Ex-post Pareto-efficiency, Probabilistic assignment, Random Serial Dictatorship, Probabilistic Serial mechanism

*Corresponding author

Email address: tom.demeulemeester@kuleuven.be (Tom Demeulemeester)

1. Introduction

We study assignment problems in which a set of *agents* have preferences over a set of indivisible *objects*, but not vice versa, while the agents can be assigned to at most one object. This problem models a wide range of applications, ranging from the assignment of pupils to schools to the allocation of social housing (see Biró (2017) and Elster (1992) for an overview of applications). In this type of problem, which is generally referred to as bipartite matching with one-sided preferences or (Capacitated) House Allocation, fairness plays a crucial role, and monetary transfers are generally not allowed. Most mechanisms used in practice adopt randomness to ensure that all agents are treated equally. As a result, we obtain a *probabilistic assignment* that determines for each agent-object pair the probability of being assigned to each other.

In order to implement this probabilistic assignment, however, we need to decompose it into a set of corresponding *matchings*, which state for each agent-object pair whether they are assigned to each other or not. We will focus on two aspects of finding such a decomposition that have a great practical appeal to decision makers, but that have not yet been studied extensively.

First of all, we make the realistic assumption that decision makers are risk-averse. This implies that they pessimistically fear the worst-case outcome, and want to avoid ending up with an “unlucky” random draw that results, for example, in a matching that assigns exceptionally few agents to an object. We therefore analyze the problem of maximizing the worst-case number of assigned agents by all matchings in the decomposition of a probabilistic assignment. A second natural requirement is that all matchings in the decomposition of a probabilistic assignment be desirable on their own. A crucial property in the context of one-sided matching is *ex-post (Pareto-)efficiency*: a matching is ex-post efficient if there is no other matching that is at least as preferred by all agents and strictly more preferred by at least one agent. If agents want to exchange their assigned objects, for example, this matching is not ex-post efficient.

Importantly, we do not want to harm the desirable properties, such as *strategy-proofness*, that are related to the probabilistic assignment that is decomposed. We therefore focus on decomposing the original probabilistic assignment exactly, rather than to approximate it (contrary to, e.g., Akbarpour & Nikzad (2020), Bronfman et al. (2018)).

Given a probabilistic assignment that satisfies certain fairness criteria, the well-known *Birkhoff-von Neumann theorem* (Birkhoff, 1946; von Neumann, 1953)

guarantees that a decomposition over matchings exists. This decomposition, however, is generally not unique, and not all decompositions are equally desirable. To the best of our knowledge, the problem of selecting the most desirable among all feasible decompositions was only studied explicitly in two papers. First, Budish et al. (2013) proposed a method to minimize the variance in the utility that is experienced by the agents between the different matchings in the decomposition. Nevertheless, their method is not capable of imposing certain desirable criteria, such as ex-post efficiency, upon the matchings in the decomposition. Second, Kesten et al. (2017) require the matchings to be ex-post efficient, but their algorithm does not allow the decision-maker to choose among the possibly many decompositions based on some worst-case measure or on alternative ex-post properties. The frameworks that we propose overcome both limitations.

Lastly, designated algorithms have been proposed to obtain decompositions with attractive ex-post properties for specific probabilistic assignments. Most notably, the decomposition of the assignment by the *Probabilistic Serial* (PS) mechanism (Bogomolnaia & Moulin, 2001) has been studied by Freeman et al. (2020) and by Aziz (2020). Whereas their algorithms can only be applied to find a decomposition for one specific probabilistic assignment, our solution methods can be applied to any probabilistic assignment, as discussed in more detail in Section 5.4.

We introduce two new problems: (1) the problem of finding a decomposition of a probabilistic assignment X that maximizes the worst-case number of assigned agents, which we refer to as $\text{MD}(X)$, and (2) the same problem with the additional requirement that all matchings in the decomposition be ex-post efficient, which we refer to as $\text{MD-SD}(X)$.

Our contributions are both theoretical and practical. From a theoretical perspective, we propose an algorithm to solve $\text{MD}(X)$ in polynomial time, and prove that the optimal value of $\text{MD}(X)$ will always be equal to the expected number of assigned agents by X , rounded down. When solving $\text{MD-SD}(X)$, however, we find that the same algorithm can only be used for a specific class of probabilistic assignments, which includes the well-studied Probabilistic Serial (Bogomolnaia & Moulin, 2001) mechanism, but not the often used *Random Serial Dictatorship* (RSD) (Abdulkadiroğlu & Sönmez, 1998). For general probabilistic assignments, $\text{MD-SD}(X)$ becomes \mathcal{NP} -hard. Specifically for the decomposition of the probabilistic assignment by the RSD mechanism, we provide tight lower and upper bounds on the optimal increase in the worst-case number of assigned agents. Lastly, we propose two column generation frame-

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

works to find an optimal decomposition for MD-SD(X), and we discuss how they can be adapted for finding decompositions that optimize alternative worst-case criteria, or that contain matchings satisfying different ex-post properties.

From a practical perspective, we have conducted computational experiments to assess the performance of the proposed methods, and we show that the realized gain in the worst-case number of assigned agents, in comparison to RSD, supports the adoption of our methods in practical applications. Because, to the best of our knowledge, no set of benchmark instances exists for one-sided matching, we have developed a parameterized data generation tool. Next to evaluating the proposed methods on generated instances of different sizes, we also assessed their performance on two real-world school choice instances from the Belgian cities Antwerp and Ghent.

The remainder of this paper is structured as follows. Section 2 formally defines one-sided matching. Next, Section 3 introduces the problems MD(X) and MD-SD(X), whereas the complexity and the optimal values of both problems are discussed in Section 4. Specifically for MD-SD(X), Section 5 describes the column generation frameworks, while Section 6 analyzes the performance of the proposed methods on existing and randomly generated one-sided matching instances. Lastly, Section 7 concludes.

2. Preliminaries

A *one-sided matching problem* is defined by a four-tuple $(N, O, >, q)$, where N is a finite set of agents, and O is a finite set of objects to which the agents in N want to be assigned. The preference profile $> = (>_1, \dots, >_{|N|})$ contains for each agent $i \in N$ a strict ordering $>_i$ over the objects in $O \cup \emptyset$, where \emptyset refers to the outside option of not being assigned to any object. We write $j >_i k$ if agent $i \in N$ prefers object j to object k , and $\emptyset >_i l$ if she prefers the outside option over object l , with $j, k, l \in O$. Moreover, the capacities $q = (q_1, \dots, q_{|O|})$ determine for each object $j \in O$ the maximum number of agents $q_j \in \mathbb{N}$ that can be assigned to it. Denote the set of all one-sided matching instances by \mathcal{I} .

A *matching* $M = [m_{ij}]$ is a binary matrix, indexed by all agents and objects, in which $m_{ij} = 1$ if M assigns agent $i \in N$ to object $j \in O$, and $m_{ij} = 0$ otherwise. An alternative way to indicate that agent $i \in N$ is assigned to object $j \in O$ is to state that $M(i) = j$. Similarly, a *probabilistic assignment* $X = [x_{ij}]$, with $x_{ij} \in [0, 1]$, can be interpreted as the probabilities with which the agents in N are assigned to the objects in O . A probabilistic assignment is *feasible*

if the sum of the allocation probabilities of each agent is at most one, and the sum of the allocation probabilities of each object does not exceed the object's capacity, i.e.,

- (i) $\sum_{j \in O} x_{ij} \leq 1, \quad \forall i \in N,$
- (ii) $\sum_{i \in N} x_{ij} \leq q_j, \quad \forall j \in O.$

Denote the set of all feasible matchings by \mathcal{M} , and the set of all feasible probabilistic assignments by $\Delta\mathcal{M}$, with $\mathcal{M} \subseteq \Delta\mathcal{M}$. Moreover, denote the expected number of assigned agents in a probabilistic assignment X by $\mu(X) = \sum_{(i,j) \in N \times O} x_{ij}$. Note that the number of agents that is assigned to an object in matching $M \in \mathcal{M}$ equals $\mu(M)$.

2.1. Implementing probabilistic assignments

In most practical applications, one does not make a decision by using the probabilistic assignment directly. Instead, the assignment is typically implemented by a lottery over matchings, in which the weights are chosen such that the expected result is equivalent to the initial probabilistic assignment. A natural question that arises is over which sets of feasible matchings a probabilistic assignment can be implemented in this way. Budish et al. (2013) formalized this using the following definition:

Definition 1. A feasible probabilistic assignment $X \in \Delta\mathcal{M}$ is *implementable over* \mathcal{M}' , where $\mathcal{M}' = \{M^t\}_{t=1}^T \subseteq \mathcal{M}$ is a subset of T feasible matchings, if there exist non-negative numbers $\lambda = \{\lambda^t\}_{t=1}^T$, with $\sum_{t=1}^T \lambda^t = 1$, such that

$$X = \sum_{t=1}^T \lambda^t M^t.$$

Given an arbitrary feasible probabilistic assignment $X \in \Delta\mathcal{M}$, the well-known *Birkhoff-von Neumann theorem* (Birkhoff, 1946; Budish et al., 2013; von Neumann, 1953) ensures that there will always exist a subset of feasible matchings over which X can be implemented. This means that each feasible probabilistic assignment X can be written as a convex combination of a set of feasible matchings $\{M^t\}_{t=1}^T \subseteq \mathcal{M}$. In order to implement X , it suffices to randomize over the matchings in this convex combination in such a way that matching M^t is selected as the final matching with a probability equal to its weight λ^t in the convex combination. The following definition formalizes the difference between implementing and decomposing a probabilistic assignment:

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

Definition 2. Given a probabilistic assignment $X \in \Delta\mathcal{M}$ that is implementable over $\mathcal{M}' = \{M^t\}_{t=1}^T \subseteq \mathcal{M}$ with weights $\lambda = \{\lambda^t\}_{t=1}^T$,

- (i) an *implementation* of X is an algorithm that randomly selects a single matching $M^t \in \mathcal{M}'$ according to the probability distribution defined by the weights λ ;
- (ii) a *decomposition* of X is the tuple (\mathcal{M}', λ) .

In other words, an implementation of a probabilistic assignment is an algorithm that implicitly describes an underlying decomposition by randomly selecting each of the matchings with a probability that is equal to the corresponding weight in the decomposition. In most practical applications, therefore, only an implementation of a probabilistic assignment is needed to make a decision, and not its entire decomposition. Nevertheless, having the entire decomposition increases the transparency of the procedure. To illustrate the introduced terminology, consider the following example.

Example 1. Let $N = \{1, 2, 3, 4\}$ be a set of agents and $O = \{a, b, c\}$ be a set of objects. The capacities of the objects are equal to $q = (2, 1, 1)$. Consider the following preference lists \succ_i of the agents $i \in N$ over the objects $j \in O$, and consider the feasible probabilistic assignment X^1 (in which the agents correspond to the rows and the objects to the columns):

$$\begin{array}{l} \text{agents 1, 2: } a > b > c > \emptyset \\ \text{agents 3, 4: } a > \emptyset > b > c \end{array} \quad X^1 = \begin{pmatrix} 1/2 & 5/12 & 1/12 \\ 1/2 & 5/12 & 1/12 \\ 1/2 & 0 & 0 \\ 1/2 & 0 & 0 \end{pmatrix}$$

Agents 1 and 2 prefer object a to object b , and object b to object c , while agents 3 and 4 only prefer object a to the outside option of not being assigned to any object. Moreover, agent 1 has a probability of $\frac{5}{12}$ of being assigned to object b in X^1 . We can rewrite X^1 in the following way:

$$X^1 = \frac{5}{12} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{5}{12} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \frac{1}{12} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{1}{12} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Denote the four matchings in the decomposition of X^1 by M^1, M^2, M^3 and M^4 . Thus, probabilistic assignment X^1 is implementable over $\{M^t\}_{t=1}^4$. Observe

that $\mu(X^1) = 3$ and that every matching M^t also assigns $\mu(M^t) = 3$ agents to an object, for $t = 1, \dots, 4$.

2.2. Mechanisms and their properties

In one-sided matching, objects are indifferent about which agents will be assigned to them. Therefore, there may exist ties between agents that have to be broken in order to obtain a matching. Define a *tie-breaking rule* $\sigma = [\sigma_j]$ for $j \in O$, where $\sigma_j = (\sigma_j(1), \dots, \sigma_j(|N|))$ is a strict ordering over the agents in N . When object j 's capacity is insufficient, its tie-breaking rule σ_j is used to decide which agent should be assigned to object j , among all agents who prefer that object equally. Denote the set of all tie-breaking rules by Σ .

A *deterministic mechanism* $\pi : \mathcal{I} \times \Sigma \mapsto \mathcal{M}$ is a function that returns a feasible matching $M^{\pi(I, \sigma)} \in \mathcal{M}$ for each one-sided matching instance $I \in \mathcal{I}$, and for each tie-breaking rule $\sigma \in \Sigma$. Similarly, a *probabilistic mechanism* $\psi : \mathcal{I} \mapsto \Delta\mathcal{M}$ returns a feasible probabilistic assignment $X^{\psi(I)} \in \Delta\mathcal{M}$ for each one-sided matching instance $I \in \mathcal{I}$. Note that a probabilistic mechanism only returns a probabilistic assignment, and not necessarily its implementation or decomposition. Unless stated differently, the term “mechanism” will refer to probabilistic mechanisms in the remainder of this paper. Moreover, we will simply refer to $X^{\psi(I)}$ by X^ψ when the instance I is clear from the context.

When designing a mechanism, there are several desirable properties one wants to satisfy. An optimality concept that has received broad attention in the context of one-sided matching (e.g., Abdulkadiroğlu & Sönmez (1998), Abraham et al. (2004)) is *ex-post (Pareto-)efficiency*. A matching $M \in \mathcal{M}$ is ex-post efficient if there is no other matching $M' \in \mathcal{M}$ that is at least as preferred as M by all agents and strictly more preferred than M by at least one agent. The *Serial Dictatorship* (SD) mechanism is a surprisingly simple deterministic mechanism to obtain an ex-post efficient matching. Denote the set of all tie-breaking rules with identical orderings for all objects by $\Sigma' \subseteq \Sigma$. Given a one-sided matching instance $I \in \mathcal{I}$ and a strict ordering $\sigma \in \Sigma'$ of the agents, the SD mechanism will first assign the first-ranked agent $\sigma(1)$ to her most preferred object, then the second-ranked agent $\sigma(2)$ to her most preferred object among the objects with remaining capacity, etc. The resulting matching will be $\text{SD}(I, \sigma) = M^{\text{SD}(I, \sigma)}$. The SD mechanism possesses several desirable properties beside ex-post efficiency (Saban & Sethuraman, 2015): it is *neutral* (invariant to relabeling the objects), *nonbossy* (no agent can change another agent's allocation without changing her own), *strategy-proof* (submitting true preferences

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

is optimal for all agents), and easy to compute. Moreover, given a one-sided matching instance $I \in \mathcal{I}$, the SD mechanism can generate any ex-post efficient matching in I (Abdulkadiroğlu & Sönmez, 1998; Abraham et al., 2004). We will therefore denote the set of all ex-post efficient matchings by $\mathcal{M}^{\text{SD}} \subseteq \mathcal{M}$.

Despite its desirable properties, the SD mechanism is rarely used in real-world applications, because it is not *anonymous*: two agents with the same preference list might be treated differently, simply because one of them appears before the other one in the ordering of the agents that serves as an input for SD. In order to overcome this issue, Abdulkadiroğlu & Sönmez (1998) introduced the *Random Serial Dictatorship* (RSD) mechanism, which will randomly select an ordering to which the SD mechanism is applied. Applying the RSD mechanism to a one-sided matching instance $I = (N, O, >, q) \in \mathcal{I}$ will result in a probabilistic assignment $X^{\text{RSD}(I)}$, which is the equally weighted sum of the matchings obtained by the SD mechanism for all different orderings $\sigma \in \Sigma'$:

$$X^{\text{RSD}(I)} = \text{RSD}(I) = \frac{1}{|N'|!} \sum_{\sigma \in \Sigma'} \text{SD}(I, \sigma). \quad (1)$$

In the remainder of the paper, we will simply refer to $X^{\text{RSD}(I)}$ as X^{RSD} if the instance is clear from the context. Note that the definition of the RSD mechanism is directly linked to an implementation of X^{RSD} in the sense of Definition 2. In the remainder of this paper, we will use the term *RSD mechanism* to refer to the function that maps an instance I to the probabilistic assignment $X^{\text{RSD}(I)}$, and the term *RSD algorithm* to refer to the implementation of $X^{\text{RSD}(I)}$ which results in the matching $\text{SD}(I, \sigma)$ by randomly selecting a tie-breaking rule $\sigma \in \Sigma'$.

Although the RSD algorithm is widely used for real-world one-sided matching problems, Bogomolnaia & Moulin (2001) showed that it fails to satisfy the following notion of efficiency that is stronger than ex-post efficiency.

Definition 3. A probabilistic assignment $X \in \Delta\mathcal{M}$ is *ordinally efficient* if and only if there is no probabilistic assignment $X' \in \Delta\mathcal{M}$ that is at least as preferred as X by all agents and strictly more preferred than X by at least one agent.

Contrary to the RSD mechanism, the *Probabilistic Serial* (PS) mechanism, that was developed by Bogomolnaia & Moulin (2001), will always result in an ordinally efficient assignment. To obtain this probabilistic assignment, assume that time t runs continuously from 0 to 1. At each point in time, each agent “eats” with a uniform eating speed of one from her most preferred object that has remaining capacity. At time $t = 1$, we obtain the ordinally efficient probabilistic

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

assignment X^{PS} . Note that an implementation of X^{PS} is not specified by its construction, in contrast to X^{RSD} .

Next to resulting in an ordinally efficient probabilistic assignment, the PS mechanism satisfies several other desirable properties (Bogomolnaia & Moulin, 2001). Similarly to the RSD mechanism, PS will be anonymous. Moreover, PS will be *envy-free* (no agent will prefer the assignment probabilities of another agent, for any compatible utility function), while RSD is not. In contrast to the RSD mechanism, however, PS will not be strategy-proof. In fact, Bogomolnaia & Moulin (2001) showed that no mechanism can satisfy ordinal efficiency, strategy-proofness and anonymity at the same time. More generally, the extent to which probabilistic assignments can satisfy certain combinations of properties has been widely studied in the literature, leading to an interesting series of impossibility results (Athanasoglou & Sethuraman, 2011; Bogomolnaia & Moulin, 2001; Martini, 2016; Nesterov, 2017; Ramezani & Feizi, 2022; Zhou, 1990).

One crucial observation in the context of this paper is that many relevant properties such as ordinal efficiency, anonymity, and strategy-proofness are inherent to the probabilistic assignment $X \in \Delta\mathcal{M}$. Hence, all such properties are retained regardless of the decomposition of X . Other properties, such as ex-post efficiency, are defined on the level of the matchings in the decomposition of X . To ensure that a decomposition of X satisfies a property of the latter category, we thus have to explicitly enforce this property on all matchings in the decomposition.

3. Maximin decomposition

Although the Birkhoff-von Neumann theorem is an important result, it does not guarantee desirable outcomes in practical applications. From the perspective of an individual agent, a decomposition might be undesirable because one of the matchings in the decomposition is undesirable in itself (Example 2). Decision makers might also prefer one decomposition to another because they are risk-averse (Example 3).

Example 2. Reconsider the decomposition of probabilistic assignment X^1 in Example 1. The third matching in the decomposition, M^3 , is clearly sub-optimal as it is not ex-post efficient. We know that agent 2 prefers object b to object c . Nevertheless, she is assigned to object c while object b has unused capacity. A similar argument holds for agent 1 in M^4 .

Example 3. Consider an instance with four agents and two objects. The capacities of the objects are equal to $q = (2, 2)$, and the agents' preferences are as indicated below. X^2 is a feasible probabilistic assignment for this instance.

$$\begin{array}{l} \text{agents 1, 2 : } a > b > \emptyset \\ \text{agents 3, 4 : } a > \emptyset > b \end{array} \quad X^2 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & 0 \\ 1/2 & 0 \end{pmatrix}$$

It can be easily verified that the following two decompositions of X^2 are both feasible:

$$X^2 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Although both decompositions result in the same probabilistic assignment X^2 , a risk-averse decision maker will probably prefer the second decomposition over the first one: the second decomposition of X^2 is guaranteed to always assign three agents to an object, whereas there is a risk that only two agents will be assigned to an object by the first decomposition of X^2 .

Inspired by Examples 2 and 3, we identify two criteria that impact the desirability of a decomposition. First of all, the agents generally have a set of properties that, depending on the context, they want to see satisfied by all matchings in the decomposition (Example 2). These properties could include, for example, ex-post efficiency, or group-specific quota. Secondly, we argue that risk-averse decision makers want to minimize the risk of ending up with an undesirable matching when implementing a probabilistic assignment. What exactly is considered to be an undesirable matching depends on the application at hand, but examples include matchings with a low average utility of the agents, a high average preference for the received object, a high total travel distance (e.g., Biró & Gudmundsson, 2021), a low *popularity* (e.g., Kondratev & Nesterov, 2022; McCutchen, 2008), or a low total number of assigned agents (Example 3). In this paper, we focus on maximizing the worst-case number of assigned agents. As will be discussed in Section 5.4, however, the column generation framework that we introduce can be easily adapted to find a decomposition that optimizes other worst-case criteria, or can even be adapted to different settings in which other criteria than ex-post efficiency should be satisfied by the matchings in the

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

decomposition.

More specifically, we study the two following, closely related problems. First, consider the problem of maximizing the worst-case number of assigned agents in the decomposition of a probabilistic assignment $X \in \Delta\mathcal{M}$, while allowing all feasible matchings in \mathcal{M} to be used in this decomposition. We will refer to this problem as MAXIMIN DECOMPOSITION OF X , or simply MD(X). Denoting all matchings in \mathcal{M} that assign at least $k \in \mathbb{N}$ agents to an object by $\mathcal{M}_k \subseteq \mathcal{M}$, we formally define MD(X) as follows:

Definition 4. MAXIMIN DECOMPOSITION OF X (MD(X)). Given a one-sided matching instance $I = (N, O, >, q) \in \mathcal{I}$, and a probabilistic assignment $X \in \Delta\mathcal{M}$, find a decomposition of X over \mathcal{M}_k that maximizes k .

Second, we define a problem similar to MD(X), but with the additional requirement that all matchings in the decomposition of X be ex-post efficient. Denote the set of all ex-post efficient matchings that assign at least $k \in \mathbb{N}$ agents to an object by $\mathcal{M}_k^{\text{SD}} \subseteq \mathcal{M}^{\text{SD}}$.

Definition 5. MAXIMIN DECOMPOSITION OF X OVER \mathcal{M}^{SD} (MD-SD(X)). Given a one-sided matching instance $I = (N, O, >, q) \in \mathcal{I}$, and a probabilistic assignment $X \in \Delta\mathcal{M}$, find a decomposition of X over $\mathcal{M}_k^{\text{SD}}$ that maximizes k .

In the remainder of this paper, we will refer to the largest value of k for which X is implementable over \mathcal{M}_k , resp. $\mathcal{M}_k^{\text{SD}}$, as the *optimal value* of MD(X), resp. MD-SD(X), and we will denote the optimal value of MD-SD(X) by $z(X)$.

Denote the floor-operator by $\lfloor \cdot \rfloor$, and the ceiling-operator by $\lceil \cdot \rceil$. We can identify the following intuitive upper bound on the optimal values of MD(X) and MD-SD(X), which is based on the observation that it is not possible to decompose a probabilistic assignment X by only using matchings that assign strictly more agents to an object than the expected number of assigned agents in X .

Proposition 1. *The optimal values of MD(X) and MD-SD(X) are upper bounded by $\lfloor \mu(X) \rfloor$, for all $X \in \Delta\mathcal{M}$.*

4. Obtaining a maximin decomposition: theoretical results

In this section, we examine the complexity and the optimal values of MD(X) and MD-SD(X). We show in Section 4.1 that the optimal value of MD(X) is $\lfloor \mu(X) \rfloor$, for all $X \in \Delta\mathcal{M}$, and we propose a polynomial-time algorithm to find a

corresponding decomposition. In Section 4.2, we establish that for MD-SD(X) the same result only holds for a specific class of probabilistic matchings, whereas MD-SD(X) is \mathcal{NP} -hard for general probabilistic assignments. Specifically for MD-SD(X^{RSD}), in Section 4.3 we provide a lower and upper bound on the optimal gain in the worst-case number of assigned agents compared to the RSD algorithm.

4.1. Complexity of MD(X)

The following theorem states that obtaining a decomposition of X that attains the upper bound $\lfloor \mu(X) \rfloor$ of Proposition 1 can be achieved in polynomial time. The existence of such a decomposition follows from a result by Budish et al. (2013) for a more general constraint structure known as a *bihierarchy*. Budish et al. further describe an implementation in the sense of Definition 2 that runs in polynomial time. The main novelty of our result is that it generalizes the method of Budish et al. in the spirit of the Birkhoff-von Neumann algorithm such that it returns a complete decomposition in polynomial time, and not simply one matching in this decomposition.

Theorem 1. *For every probabilistic assignment $X \in \Delta\mathcal{M}$, we can find a decomposition (\mathcal{M}', λ) of X in polynomial time in which each matching $M \in \mathcal{M}'$ assigns either $\lfloor \mu(X) \rfloor$ or $\lceil \mu(X) \rceil$ agents.*

Proof. Consider a probabilistic assignment $X \in \Delta\mathcal{M}$ and assume that $\mu(X) \in \mathbb{N}$. We will describe an iterative approach to obtain a decomposition of X in which all matchings assign exactly $\mu(X)$ agents. The case where $\mu(X)$ is fractional can be dealt with by adding an additional row and column to X , corresponding to a dummy agent a and a unit-capacity object o , where the unique non-zero element (a, o) has an assignment probability equal to $\lceil \mu(X) \rceil - \mu(X)$. A decomposition that always assigns $\lceil \mu(X) \rceil$ agents in this modified instance, then guarantees to assign either $\lfloor \mu(X) \rfloor$ or $\lceil \mu(X) \rceil$ non-dummy agents in the original instance.

Following Budish et al. (2013), we can rewrite the conditions for X to be a feasible probabilistic assignment using a *constraint structure* $\mathcal{H} \subset 2^{N \times O}$, where \mathcal{H} contains a *constraint set* $S \subseteq N \times O$ with associated *quota* q_S for each agent, for each object, and for each agent-object pair. In particular, define

$$\mathcal{H} = \bigcup_{(i,j) \in N \times O} \{(i, j)\} \cup \bigcup_{i \in N} (\{i\} \times O) \cup \bigcup_{j \in O} (N \times \{j\}),$$

and let $q_S = q_j$ if $S = N \times \{j\}$ for some $j \in O$, and $q_S = 1$ for all other $S \in \mathcal{H}$. A probabilistic assignment X is then feasible if and only if $0 \leq \sum_{(i,j) \in S} x_{ij} \leq q_S$ for every $S \in \mathcal{H}$. Observe that $|\mathcal{H}| = |N| \cdot |O| + |N| + |O| = \mathcal{O}(|N| \cdot |O|)$.

Each iteration t of our iterative approach starts from a given probabilistic assignment X^t , where $X^t = X$ if $t = 1$, and X^t is defined in the previous iteration otherwise. If X^t is a matching, i.e., if $X^t \in \mathcal{M}$, then the procedure terminates and we define $M^t = X^t$. Otherwise, we apply to X^t a modified version of the algorithm by Budish et al. (2013), which we describe in Appendix A, in order to obtain a matching $M^t \in \mathcal{M}$ with the following two properties:

- (i) $\mu(M^t) = \mu(X^t)$;
- (ii) if $\sum_{(i,j) \in S} x_{ij}^t$ is integer, then $\sum_{(i,j) \in S} m_{ij}^t = \sum_{(i,j) \in S} x_{ij}^t$, for each $S \in \mathcal{H}$.

Next, let λ^t be the largest value for $\lambda \in \mathbb{R}$ such that $X^t + \lambda(X^t - M^t) \in \Delta\mathcal{M}$, where λ^t can be determined by checking which of the constraint sets $S \in \mathcal{H}$ becomes binding first. The procedure then defines $X^{t+1} = X^t + \lambda^t(X^t - M^t)$, and proceeds to the next iteration.

Now assume that our procedure terminates in T iterations. It follows by expanding the above recursion that the matchings $\{M^t\}_{t=1}^T$ and weights $\hat{\lambda} = \{\hat{\lambda}^t\}_{t=1}^T$, with

$$\hat{\lambda}^t = \begin{cases} \frac{\lambda^t}{1+\lambda^t} & \text{if } t = 1, \\ \frac{\lambda^t}{1+\lambda^t} \prod_{u=1}^{t-1} \frac{1}{1+\lambda^u} & \text{if } t \in \{2, \dots, T-1\}, \\ \prod_{u=1}^{t-1} \frac{1}{1+\lambda^u} & \text{if } t = T, \end{cases}$$

yield a decomposition of X in which all matchings assign $\mu(X)$ agents. This latter guarantee follows from Property (i) of M^t and by definition of X^{t+1} for each $t = 1, \dots, T-1$.

We claim that our procedure terminates in $T \leq |\mathcal{H}|$ iterations. For each probabilistic assignment $X' \in \Delta\mathcal{M}$, denote by

$$\tau(X') = |\{S \in \mathcal{H} : \sum_{(i,j) \in S} x'_{ij} \in \mathbb{N}\}|$$

the number of constraint sets $S \in \mathcal{H}$ that are integer in X' . Observe that $\tau(X') \leq |\mathcal{H}|$, where equality is reached if and only if $X' \in \mathcal{M}$. To prove our claim, it thus suffices to show that, for every iteration t with $X^t \notin \mathcal{M}$, it holds that $\tau(X^{t+1}) > \tau(X^t)$.

Consider an arbitrary iteration t in which $X^t \notin \mathcal{M}$, and denote $R^t = X^t - M^t$. Observe that $R^t \neq 0$ since $X^t \neq M^t$. By Property (ii) of M^t , it holds for every $S \in \mathcal{H}$ for which $\sum_{(i,j) \in S} x_{ij}^t$ is integer that $\sum_{(i,j) \in S} r_{ij}^t = 0$ and thus, by definition of X^{t+1} , that $\sum_{(i,j) \in S} x_{ij}^{t+1} = \sum_{(i,j) \in S} x_{ij}^t$. Property (ii) of M^t also implies that $\sum_{(i,j) \in S} x_{ij}^t$ is fractional for every $S \in \mathcal{H}$ with $\sum_{(i,j) \in S} r_{ij}^t \neq 0$, and thus there exists a $\lambda > 0$ for which $X^t + \lambda R^t \in \Delta\mathcal{M}$. The maximal choice λ^t for λ is then such that there is a $S' \in \mathcal{H}$ with $\sum_{(i,j) \in S'} x_{ij}^t$ fractional and $\sum_{(i,j) \in S'} x_{ij}^{t+1}$ integer (because a new constraint becomes binding). This yields $\tau(X^{t+1}) > \tau(X^t)$, which proves our claim.

Since the modified algorithm of Budish et al. (2013), described in Appendix A, runs in time $\mathcal{O}(|\mathcal{H}| \min\{|N|, |O|\})$ and since finding λ^t in a given iteration $t < T$ can be done in time $\mathcal{O}(|\mathcal{H}|)$, we can obtain our desired decomposition of X in time $\mathcal{O}(|\mathcal{H}|^2 \min\{|N|, |O|\})$. \square

Theorem 1 implies that the optimal value of MD(X) is equal to $\lfloor \mu(X) \rfloor$, for any $X \in \Delta\mathcal{M}$, and that the corresponding decomposition can be found in polynomial time. In other words, if the decision maker does not require specific properties to be satisfied by all matchings in the decomposition of a probabilistic assignment X , it is always possible to decompose X by only using matchings that assign at least $\lfloor \mu(X) \rfloor$ agents to an object.

4.2. Complexity of MD-SD(X)

In contrast to MD(X), the complexity of MD-SD(X) depends on the properties of the probabilistic assignment X . Consider the following property.

Definition 6. (Aziz et al., 2015) A probabilistic assignment $X \in \Delta\mathcal{M}$ is *robust ex-post efficient* if and only if every possible decomposition of X only contains ex-post efficient matchings.

Denote the set of robust ex-post efficient probabilistic assignments by $\Delta\mathcal{M}^R \subseteq \Delta\mathcal{M}$. Define a probabilistic assignment $X \in \Delta\mathcal{M}$ to be *ex-post efficient* if a decomposition of X exists that only contains ex-post efficient matchings. The following proposition shows the relationship between the three efficiency concepts for probabilistic assignments that have been mentioned in this paper.

Proposition 2. (Aziz et al., 2015) *Ordinal efficiency of a probabilistic assignment implies robust ex-post efficiency, which in turn implies ex-post efficiency.*

Although the above result was originally established for a setting with an equal number of agents and unit-capacity objects (Aziz et al., 2014, Theorem 5),

the proof directly carries over to our setting. Proposition 2 has the following implications for the PS mechanism.

Corollary 1. *The PS mechanism will always return a probabilistic assignment that is robust ex-post efficient.*

Moreover, with respect to the RSD mechanism, Aziz et al. (2015) obtained the following result (see Example 2, where $X^1 = X^{\text{RSD}}$, and $M^3, M^4 \notin \mathcal{M}^{\text{SD}}$).

Proposition 3. *(Aziz et al., 2015) The RSD mechanism may return a probabilistic assignment that is not robust ex-post efficient.*

Using these results, we can now identify an important class of probabilistic assignments for which we obtain the same results as in Theorem 1, i.e., for which the optimal value of MD-SD(X) is guaranteed to be equal to its upper bound $\lfloor \mu(X) \rfloor$, and for which we can find a decomposition in polynomial time.

Theorem 2. *For every robust ex-post efficient probabilistic assignment $X \in \Delta\mathcal{M}^{\text{R}}$, we can find a decomposition (\mathcal{M}', λ) of X in polynomial time in which each matching $M \in \mathcal{M}'$ assigns either $\lfloor \mu(X) \rfloor$ or $\lceil \mu(X) \rceil$ agents.*

Proof. Consider an arbitrary robust ex-post efficient probabilistic assignment $X \in \Delta\mathcal{M}^{\text{R}}$. By using the polynomial-time algorithm from the proof of Theorem 1, we can find a decomposition of X in which all matchings assign either $\lfloor \mu(X) \rfloor$ or $\lceil \mu(X) \rceil$ agents to an object. By definition of robust ex-post efficiency, all matchings in this decomposition will be ex-post efficient. \square

Combining this result with Corollary 1, we can conclude that we can decompose the probabilistic assignment X^{PS} in polynomial time over a set of ex-post efficient matchings that assign at least $\lfloor \mu(X^{\text{PS}}) \rfloor$ agents to an object. From Proposition 3, however, we know that we cannot guarantee the ex-post efficiency of the matchings in the decomposition if we apply the same approach to X^{RSD} .

Although Theorem 2 showed that MD-SD(X) is polynomially solvable when X is robust ex-post efficient, the following result holds for general probabilistic assignments.

Theorem 3. *MD-SD(X) is \mathcal{NP} -hard in general.*

Proof. The result immediately follows from the fact that it is already \mathcal{NP} -complete to determine whether an arbitrary probabilistic assignment can be

implemented over ex-post efficient matchings when objects have unit capacities (Aziz et al., 2015). □

Note that the proof of Theorem 3 does not apply for X^{RSD} , because X^{RSD} is guaranteed to have at least one decomposition over ex-post efficient matchings, given by Equation (1). Hence, the complexity status of MD-SD(X^{RSD}) is open.

4.3. Bounds on the optimal value of MD-SD(X^{RSD})

By Theorem 2, we know that the optimal value $z(X)$ of MD-SD(X) equals its upper bound $\lfloor \mu(X) \rfloor$ when X is robust ex-post efficient. In this section, we analyze the possible range of values of $z(X)$ when $X \notin \Delta\mathcal{M}^R$. More specifically, we focus on the RSD mechanism, given its practical and theoretical importance: we establish that $\frac{1}{2}\lfloor \mu(X^{RSD}) \rfloor$ is an asymptotically tight lower bound on $z(X^{RSD})$, whereas twice the worst-case number of assigned agents by any ex-post efficient matching is an asymptotically tight upper bound on $z(X^{RSD})$.

Denoting the cardinality of a minimum-cardinality ex-post efficient matching in a one-sided matching instance $I \in \mathcal{I}$ by $p^-(I)$, which is equal to the worst-case number of assigned agents by the RSD algorithm, we can identify the following bounds on $z(X^{RSD(I)})$.

Proposition 4. $\frac{1}{2}\lfloor \mu(X^{RSD(I)}) \rfloor < z(X^{RSD(I)}) < 2p^-(I)$, for all $I \in \mathcal{I}$.

Proof. By definition of MD-SD($X^{RSD(I)}$), and given the upper bound on $z(X^{RSD(I)})$ from Proposition 1, we know that $p^-(I) \leq z(X^{RSD(I)}) \leq \lfloor \mu(X^{RSD(I)}) \rfloor$. Hence, what remains to be shown is that $\frac{1}{2}\lfloor \mu(X^{RSD(I)}) \rfloor < p^-(I)$.

Denote by $p^+(I)$ the cardinality of a maximum-cardinality ex-post efficient matching in a one-sided matching instance $I \in \mathcal{I}$, and note that $p^-(I) \leq p^+(I)$. We will consider two cases. First, if $p^-(I) = p^+(I)$, clearly $\frac{1}{2}\lfloor \mu(X^{RSD(I)}) \rfloor < p^-(I)$. Second, if $p^-(I) < p^+(I)$, we know that $\lfloor \mu(X^{RSD(I)}) \rfloor < p^+(I) \leq 2p^-(I)$, where the first inequality follows from Equation (1), and the second inequality is a result by Abraham et al. (2004). □

In the following two theorems, the proofs of which are included in Appendices B and C, we show that for each of the two strict inequalities in Proposition 4, a family of instances exists for which both sides of the inequality converge. First of all, the following result states that there exists a family of instances for which $z(X^{RSD})$ cannot achieve any improvement in the worst-case number of assigned agents compared to the RSD algorithm, causing the ratio of $z(X^{RSD})$ over the upper bound $\lfloor \mu(X^{RSD}) \rfloor$ to converge to $\frac{1}{2}$.

Theorem 4. *There exists a family of instances $I_k \in \mathcal{I}$, with $k \geq 2$ an integer, for which*

$$\lim_{k \rightarrow \infty} \frac{z(X^{\text{RSD}(I_k)})}{\lfloor \mu(X^{\text{RSD}(I_k)}) \rfloor} = \frac{1}{2}.$$

Secondly, we show that there exist instances where $z(X^{\text{RSD}})$ realizes the maximum possible improvement in the worst-case number of assigned agents with respect to the RSD algorithm, as the ratio of $z(X^{\text{RSD}})$ over $p^-(I)$ converges to 2.

Theorem 5. *There exists a family of instances $I_\ell \in \mathcal{I}$, with $\ell \geq 2$ an integer, for which*

$$\lim_{\ell \rightarrow \infty} \frac{z(X^{\text{RSD}(I_\ell)})}{p^-(I_\ell)} = 2.$$

Lastly, we provide some intuition about the relation between the expected number of assigned agents by the RSD mechanism and the maximum number of assigned agents by any ex-post efficient matching in an instance $I \in \mathcal{I}$, which we denote by $p^+(I)$. Following a similar reasoning as in the proof of Proposition 4, we know that when $p^-(I) < p^+(I)$ it must hold that $1 < \frac{p^+(I)}{\mu(X^{\text{RSD}(I)})} < 2$. First, the family of instances I_k as defined in the proof of Theorem 4 shows that this intuitive lower bound is asymptotically tight. Second, this intuitive upper bound can be strengthened for larger instances, as discussed by Bhalgat et al. (2011) and Krysta et al. (2019) for RSD, and by Bogomolnaia & Moulin (2015) for PS.

5. Column generation

In order to be able to solve MD-SD(X) to optimality for reasonably large instances when X is not robust ex-post efficient, we introduce two alternative *column generation* frameworks. Recall that the total number of tie-breaking rules $|\Sigma'|$ that can be used as an input for the SD mechanism to obtain an ex-post efficient matching is equal to $|N|!$, which implies that the number of ex-post efficient matchings $|\mathcal{M}^{\text{SD}}|$ can be exponential in the number of agents. A linear programming model that determines the weights of each matching in the decomposition would therefore contain an exponential number of decision variables. A common approach to tackle this issue is *column generation* (e.g., Bertsimas & Tsitsiklis (1997)). The idea is to first solve the model with a subset of the variables in the *restricted master problem*. Subsequently, we check whether this leads to an optimal solution over all variables. If this is not the case,

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

a new variable will be added to the model, which is then solved again until the restricted master problem returns an optimal solution over all variables. New variables are generated by a separate subproblem called the *pricing problem*, which identifies a variable with a negative *reduced cost* or shows that no such variable exists.

In the remainder of this section, we propose two alternative column generation frameworks that will tackle the problem of finding an optimal solution to MD-SD(X) from different angles. In both frameworks, we solve MD-SD(X) by means of a binary search for $z(X)$, in which each step consists of solving a linear program to check whether a probabilistic assignment X is implementable over $\mathcal{M}_k^{\text{SD}}$ for some integer k . In Sections 5.1 and 5.2, by only using matchings in $\mathcal{M}_k^{\text{SD}}$, we will minimize the largest difference between X and the resulting probabilistic assignment. Alternatively, in Section 5.3, we enforce the resulting decomposition to be exactly equal to X while only using ex-post efficient matchings, and we will maximize the cumulative weights of the matchings in $\mathcal{M}_k^{\text{SD}}$ in this decomposition. Although both frameworks are similar, their optimal solutions differ when X is not implementable over $\mathcal{M}_k^{\text{SD}}$ for some integer k . The performance and the interpretations of both frameworks will be discussed in Section 6.3.

5.1. Minimizing the deviation from X : restricted master problem

In this section, we formulate the restricted master problem [RMP] for a column generation framework that aims to find a decomposition of a probabilistic assignment X in which all matchings are ex-post efficient and assign at least k agents to an object, for a given value of k in the binary search. To guide this search, we will minimize the largest difference over all agent-object pairs between X and the resulting probabilistic assignment. Let $X = [x_{ij}]$, with $(i, j) \in N \times O$, be the probabilistic assignment we want to decompose, and let $\tilde{\mathcal{M}}_k^{\text{SD}} \subseteq \mathcal{M}_k^{\text{SD}}$ be the subset of the variables that are considered in the restricted master problem. Moreover, define decision variables λ^t to be the weight of matching $M^t \in \tilde{\mathcal{M}}_k^{\text{SD}}$ in the decomposition of X , and let s be an auxiliary variable.

$$\begin{aligned}
 \text{[RMP]} \quad & \min && s && (2a) \\
 \text{s.t.} && \sum_{t: M^t \in \tilde{\mathcal{M}}_k^{\text{SD}}} \lambda^t m_{ij}^t - x_{ij} \geq 0 && \forall (i, j) \in N \times O, && (2b)
 \end{aligned}$$

$$\sum_{t: M^t \in \tilde{\mathcal{M}}_k^{\text{SD}}} \lambda^t m_{ij}^t - x_{ij} \leq s \quad \forall (i, j) \in N \times O, \quad (2c)$$

$$\sum_{t: M^t \in \tilde{\mathcal{M}}_k^{\text{SD}}} \lambda^t = 1, \quad (2d)$$

$$\lambda^t \geq 0 \quad \forall t : M^t \in \tilde{\mathcal{M}}_k^{\text{SD}}. \quad (2e)$$

Constraints (2b) impose that the probability of agent i being assigned to object j is at least as high in the solution as in the original probabilistic assignment X . Constraints (2c) ensure, together with the objective function, that s will be equal to the maximum difference between the constructed probabilistic assignment and X , over all agent-object pairs. As a result, a feasible decomposition of X over $\tilde{\mathcal{M}}_k^{\text{SD}}$ has been found if the objective value of [RMP] equals zero. Furthermore, constraints (2d) and (2e) state that the weights in the decomposition are non-negative and sum to one. To ensure the feasibility of [RMP] for any subset $\tilde{\mathcal{M}}_k^{\text{SD}}$, we can include an artificial *super-column* M^0 in which $m_{ij}^0 = 1$ for all agent-object pairs $(i, j) \in N \times O$. Moreover, to reduce the number of constraints, we can remove redundant constraints (2b) when $x_{ij} = 0$, and constraints (2c) when $x_{ij} = 1$.

Denote the dual variables related to the constraints (2b), (2c), and (2d) by $u_{ij} \geq 0$, $v_{ij} \leq 0$, and w , respectively, for each $(i, j) \in N \times O$. Then a solution of [RMP] with dual values $u^* = [u_{ij}^*]$, $v^* = [v_{ij}^*]$, and w^* is optimal over all variables in $\mathcal{M}_k^{\text{SD}}$ if no matching $M \in \mathcal{M}_k^{\text{SD}}$ has a negative reduced cost, i.e., if for all matchings $M \in \mathcal{M}_k^{\text{SD}}$

$$\sum_{(i,j) \in N \times O} -m_{ij}^t (u_{ij}^* + v_{ij}^*) - w^* \geq 0. \quad (3)$$

5.2. Minimizing the deviation from X : pricing problem

In order to determine the existence of a matching that violates (3), we formulate a pricing problem to find a feasible ex-post efficient matching $M \in \mathcal{M}_k^{\text{SD}}$ that minimizes the left-hand side of inequality (3) for given dual values u^* , v^* , and w^* . This pricing problem contains two sets of constraints: constraints (4b)-(4e) ensure that the constructed matching is feasible and assigns at least k agents to an object, while constraints (4f)-(4n) guarantee that the matching is ex-post efficient.

First, define binary decision variables m_{ij} , where $m_{ij} = 1$ if agent $i \in N$ is assigned to object $j \in O$, and $m_{ij} = 0$ otherwise. Then, the following integer

linear program [PP₀] will find a minimum-weight matching that assigns at least k agents to an object.

$$[\text{PP}_0] \quad \min \quad \sum_{i \in N} \sum_{j \in O} -m_{ij}(u_{ij}^* + v_{ij}^*) - w^* \quad (4a)$$

$$\text{s.t.} \quad \sum_{j \in O} m_{ij} \leq 1 \quad \forall i \in N, \quad (4b)$$

$$\sum_{i \in N} m_{ij} \leq q_j \quad \forall j \in O, \quad (4c)$$

$$\sum_{i \in N} \sum_{j \in O} m_{ij} \geq k, \quad (4d)$$

$$m_{ij} \in \{0, 1\} \quad \forall (i, j) \in N \times O. \quad (4e)$$

The objective function of [PP₀] minimizes the left-hand side of inequality (3), and will find a matching with a negative reduced cost if and only if the objective value is strictly negative. In order to reduce the number of variables in [PP₀] we can fix m_{ij} to one (zero) if agent i is always (never) assigned to object j by the probabilistic assignment X that we want to decompose. Note that [PP₀] is, except for the inequality in constraint (4d), identical to the formulation by Dell'Amico & Martello (1997) for the k -CARDINALITY ASSIGNMENT PROBLEM (k -AP), which finds a minimum-weight matching that assigns exactly k agents to an object. They showed that k -AP is polynomially solvable because of the total unimodularity of the constraint matrix. The same holds for [PP₀] if we do not require the final matching to be ex-post efficient.

Additionally, we enforce the matching to be ex-post efficient by appending to [PP₀] the constraints that were recently formulated by Biró & Gudmundsson (2021), slightly simplified to the setting with strict agent preferences. They showed that a matching $M \in \mathcal{M}$ is ex-post efficient if and only if there exists a vector of prices $p \in \{0, \dots, |O|\}^{|O|}$ such that (M, p) is a *competitive equilibrium*. We say that (M, p) is a competitive equilibrium if it satisfies the following two conditions:

- (i) each object $j \in O$ with remaining capacity has a price equal to zero: $q_j - \sum_{i \in N} m_{ij} > 0 \Rightarrow p_j = 0$;
- (ii) each object $k \in O$ that is preferred by at least one agent to its assigned object $j \in O$ has a strictly higher price than j : $\exists i \in N : k \succ_i j = M(i) \Rightarrow p_k > p_j$.

Note that we do not require the third set of conditions by Biró & Gudmundsson (2021), because we assume agent preferences to be strict. Alternatively, we have also tried to enforce ex-post efficiency by translating the three necessary and sufficient conditions for ex-post efficiency by Abraham et al. (2004) into linear constraints, but the recent formulation by Biró and Gudmundsson outperformed this alternative formulation.

In order to implement the conditions for a competitive equilibrium in a linear model, we introduce some additional decision variables. Using the notation by Biró and Gudmundsson, define integer variables $p_j \in \{0, \dots, |O|\}$ which represent the price for each object $j \in O$. Moreover, define an auxiliary variable $f_j \in \{0, 1\}$ for each object $j \in O$, and auxiliary variables $s_{jk} \in \mathbb{N}$ and $\tilde{s}_{jk} \in \{0, 1\}$ for each object pair $\{j, k\} \subseteq O$. Denoting the set of objects that agent i prefers to the outside option by $O_i \subseteq O$, we append the following constraints to [PP₀].

$$[\text{PP}_{\text{PE}}] \quad \sum_{i \in N: k >_i j} m_{ij} = s_{jk} \quad \forall \{j, k\} \subseteq O, \quad (4f)$$

$$\tilde{s}_{jk} \leq s_{jk} \quad \forall \{j, k\} \subseteq O, \quad (4g)$$

$$\tilde{s}_{jk} \cdot |N| \geq s_{jk} \quad \forall \{j, k\} \subseteq O, \quad (4h)$$

$$f_j q_j \leq \sum_{i \in N} m_{ij} \quad \forall j \in O, \quad (4i)$$

$$f_j + q_j \geq \sum_{i \in N} m_{ij} + 1 \quad \forall j \in O, \quad (4j)$$

$$\sum_{l \in O_i} m_{il} + f_j \geq 1 \quad \forall (i, j) \in N \times O_i, \quad (4k)$$

$$f_j \cdot |O| \geq p_j \quad \forall j \in O, \quad (4l)$$

$$(1 - \tilde{s}_{jk})(|O| + 1) + p_k \geq p_j + 1 \quad \forall \{j, k\} \subseteq O, \quad (4m)$$

$$p_j \in \mathbb{N} \quad \forall j \in O. \quad (4n)$$

In summary, constraints (4f)-(4j) define the auxiliary variables, constraints (4k) enforce the resulting matching to be maximal, and constraints (4l) and (4m) ensure that (M, p) is a competitive equilibrium. More specifically, the first set of constraints (4f) sets s_{jk} equal to the number of agents that prefer object k to their assigned object j . Secondly, constraints (4g) and (4h) ensure that \tilde{s}_{jk} equals one if s_{jk} is strictly positive, and that \tilde{s}_{jk} equals zero otherwise. Thirdly, constraints (4i) and (4j) ensure that f_j equals zero when object $j \in O$ has

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in European Journal of Operational Research, Volume 305, Issue 3, Pages 1087-1099 (2023).

unused capacity, and that f_j equals one otherwise. Constraints (4k) enforce the matching to be maximal: if agent $i \in N$ is not assigned to any object, then none of the objects that she prefers to the outside option can have unused capacity. Lastly, constraints (4l) force the price of object $j \in O$ to be zero when it has unused capacity, while constraints (4m) impose $p_k \geq p_j + 1$ whenever $\tilde{s}_{jk} = 1$. Note that adding the integrality conditions (4n) for p_j is not required for the correctness of the formulation, but we found that doing so results in lower computation times.

While the basic pricing problem [PP₀] can be solved in polynomial time, the addition of constraints (4f)-(4n) to [PP₀] causes the problem to become computationally hard. This follows directly from the \mathcal{NP} -hardness of finding an ex-post efficient matching of maximum weight (Biró & Gudmundsson, 2021).

5.3. Maximizing the fraction of large matchings

In this section, we propose an alternative procedure that obtains an exact decomposition of the probabilistic assignment X over ex-post efficient matchings, while minimizing the cumulative weight of the matchings in the decomposition that assign less than k agents to an object.

Consider the same notation as in Section 5.1, and let $k \in \mathbb{N}$ be the the number of assigned agents for which we are checking whether a given probabilistic assignment $X \in \Delta\mathcal{M}$ can be decomposed over $\mathcal{M}_k^{\text{SD}}$. Denote the cumulative weight of the matchings in the decomposition that assign at least k agents to an object by decision variable $\alpha \in [0, 1]$. We can slightly adapt [RMP] to model the restricted master problem for the alternative formulation as follows.

$$[\alpha\text{RMP}] \quad \max \quad \alpha \tag{5a}$$

$$\text{s.t.} \quad \sum_{t: M^t \in \tilde{\mathcal{M}}^{\text{SD}}} \lambda^t m_{ij}^t = x_{ij} \quad \forall (i, j) \in N \times O, \tag{5b}$$

$$\sum_{t: M^t \in \tilde{\mathcal{M}}_k^{\text{SD}}} \lambda^t - \alpha \geq 0, \tag{5c}$$

$$\sum_{t: M^t \in \tilde{\mathcal{M}}^{\text{SD}}} \lambda^t = 1, \tag{5d}$$

$$\lambda^t \geq 0 \quad \forall t: M^t \in \tilde{\mathcal{M}}^{\text{SD}}. \tag{5e}$$

Note that constraints (5b), (5d), and (5e) are equivalent to their counterparts in [RMP] where $\tilde{\mathcal{M}}_k^{\text{SD}}$ is replaced by $\tilde{\mathcal{M}}^{\text{SD}}$. Denote the dual variables of constraints (5b), (5c), and (5d) by u_{ij} , $v \leq -1$, and w respectively, for each

$(i, j) \in N \times O$. Given a solution of $[\alpha\text{RMP}]$ with dual variables $u^* = [u_{ij}^*]$, v^* and w^* , this solution is optimal if no matching $M \in \mathcal{M}^{\text{SD}}$ has a negative reduced cost, i.e., if

$$\begin{cases} \sum_{(i,j) \in N \times O} m_{ij} u_{ij}^* + w < 0 & \forall M \in \mathcal{M}^{\text{SD}} \setminus \mathcal{M}_k^{\text{SD}}, \\ \sum_{(i,j) \in N \times O} m_{ij} u_{ij}^* + v + w < 0 & \forall M \in \mathcal{M}_k^{\text{SD}}. \end{cases} \quad (6)$$

Hence, two pricing problems should be solved to identify whether a solution for $[\alpha\text{RMP}]$ is optimal. The first pricing problem minimizes the first expression in (6) under constraints (4b)-(4n), but without cardinality constraint (4d). The second pricing problem simply minimizes the second expression in (6) under constraints (4b)-(4n). If any of the pricing problems finds a matching with a negative objective value, it is added to $\tilde{\mathcal{M}}^{\text{SD}}$ and to $\tilde{\mathcal{M}}_k^{\text{SD}}$ if the matching assigns at least k agents, and $[\alpha\text{RMP}]$ is solved again. If the objective value for both pricing problems is non-negative, the found solution for $[\alpha\text{RMP}]$ is optimal.

5.4. Flexibility of the column generation framework

In this subsection, we discuss how the inherent flexibility of linear (integer) programming methods can be exploited to adapt the proposed column generation frameworks to find decompositions of probabilistic assignments in different problem settings. As such, our column generation framework can be readily adapted to alternative assignment settings where ties have to be broken, or where probabilistic solution concepts are preferred. Next to RSD and PS, interesting probabilistic assignments to be decomposed are proposed by Ashlagi & Shi (2016), Ashlagi et al. (2020), Kavitha et al. (2011), Kesten et al. (2017), and Shi (2022).

Changes in the problem setting can be classified into two categories. First, the decision-maker could require the matchings in the decomposition to satisfy other ex-post properties than ex-post efficiency. Examples of possible alternative properties are quota, stability notions, or enforcing marriage partners to be matched together (e.g., Bronfman et al., 2018). These alternative properties can be easily enforced by including the relevant constraints in the pricing problem. Note, however, that it might be possible that a given probabilistic assignment is not decomposable over a set of matchings with the desired properties (e.g., Bronfman et al., 2018, Theorem 4.1). In that case, the column generation framework $[\text{RMP}]$ will output a decomposition in which the largest difference

between the resulting allocation probabilities and the initial allocation probability is minimized over all agent-object pairs. By changing the objective function of [RMP], other ways of approximating the initial allocation probabilities can also be modeled.

Second, the decision-maker could be interested in finding a decomposition that maximizes an alternative worst-case measure. Examples of alternative measures include a minimization of the worst-case total travel distance to the schools (e.g., Biró & Gudmundsson, 2021), a minimization of some worst-case *unpopularity* measure (e.g., Kondratev & Nesterov, 2022; McCutchen, 2008), or an optimization of the worst-case average assigned preference or utility of the agents. These alternative worst-case measures can be easily applied by modifying the binary search procedure. More specifically, we replace constraint (4d) by a set of linear constraints to bound the matchings with respect to the evaluated value of the alternative worst-case measure in the binary search.

While obtaining such a set of linear constraints is rather straightforward for most of the discussed worst-case measures, we illustrate how to obtain a matching with a bounded *unpopularity margin*. Denote by $\phi(M, M')$ the number of agents who prefer matching M over matching M' . We say that matching M is *more popular* than matching M' if $\phi(M, M') > \phi(M', M)$. The unpopularity margin of a matching M is then defined as $g(M) = \max_{M' \in \mathcal{M}} (\phi(M', M) - \phi(M, M'))$. The interpretation of a matching M having unpopularity margin $g(M) = \omega$ is that for all matchings M' , the number of agents who prefer M' over M is at most ω units larger than the number of agents who prefer M over M' .

To impose the matchings in the decomposition to have an unpopularity margin of at most ω , we adopt a similar approach as in the recent work by Kavitha et al. (2022). Consider the additional binary variable $m_{i\emptyset}$ which equals one if agent $i \in N$ is assigned to the outside option, and zero otherwise, and let $q_\emptyset = |N|$. Moreover, given a matching $M \in \mathcal{M}$, let weights $\nu_M(i, j)$ be equal to one if agent i prefers object j to $M(i)$, and let $\nu_M(i, j)$ equal minus one if agent i prefers $M(i)$ to j . Following Kavitha et al. (2022), the optimal objective function of formulation [LP₁], in which m'_{ij} are the decision variables, will be equal to the unpopularity margin of a given matching M .

$$[\text{LP}_1] \quad \max \quad \sum_{(i,j) \in N \times (OU\emptyset)} \nu_M(i, j) \cdot m'_{ij} \tag{7a}$$

$$\text{s.t.} \quad \sum_{j \in OU\emptyset} m'_{ij} = 1 \quad \forall i \in N, \tag{7b}$$

$$\sum_{i \in N} m'_{ij} \leq q_j \quad \forall j \in O \cup \emptyset, \quad (7c)$$

$$m'_{ij} \geq 0 \quad \forall (i, j) \in N \times (O \cup \emptyset). \quad (7d)$$

However, when matching M is no longer given, but determined upon by the pricing problem, weights ν_M become decision variables as well. Simply adding a constraint to the pricing problem that upper bounds the objective function of [LP₁] by ω would therefore result in a non-linear integer program. To overcome this issue, we can use strong duality and include the dual constraints of [LP₁] while lower bounding the dual objective by ω . As a result, the unpopularity margin of the found matching will be at most ω if and only if there exists a vector $\alpha \in \mathbb{R}^{|N|+|O|+1}$ and a matching $M = [m_{ij}] \in \mathcal{M}$ that satisfy constraints (8b)-(8d), where α represents the dual variables of constraints (7b) and (7c). Constraint (8a) corresponds to the bounded dual objective of [LP₁], and constraints (8b) and (8d) to the dual constraints. Constraints (8c) determine the values of weights ν_M as defined.

$$\sum_{i \in N} \alpha_i + \sum_{j \in O \cup \emptyset} q_j \alpha_j \geq \omega, \quad (8a)$$

$$\alpha_i + \alpha_j \geq \nu_M(i, j) \quad \forall (i, j) \in N \times (O \cup \emptyset), \quad (8b)$$

$$\sum_{k \in O \cup \emptyset: j >_i k} m_{ik} - \sum_{k \in O \cup \emptyset: k >_i j} m_{ik} = \nu_M(i, j) \quad \forall (i, j) \in N \times (O \cup \emptyset), \quad (8c)$$

$$\alpha_j \geq 0 \quad \forall j \in O \cup \emptyset. \quad (8d)$$

6. Computational experiments

In this section, we discuss the computational performance of the introduced methods for decomposing X^{RSD} , which we test both on existing and on newly generated instances. The choice to apply our methods on the RSD mechanism is motivated by its widespread use in practical applications. In particular, these experiments allow us to gain the two following main insights. Firstly, despite the worst-case result of Theorem 4, the optimal value of MD-SD(X^{RSD}) is equal to its upper bound $\lfloor \mu(X^{\text{RSD}}) \rfloor$ for all generated instances that were solved to optimality. Secondly, we observe that even when the column generation frameworks from Section 5 do not obtain a guaranteed optimal solution within the imposed runtime limit, their results can still be highly valuable in real-world applications, and we discuss two real-world school choice cases from the Belgian

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

cities of Antwerp and Ghent as examples.

6.1. Implementation details

All experiments were implemented with C++, compiled with Microsoft Visual C++ 2019, and run on an Intel Core i7-7700 processor running at 3.60 GHz, with 16GB of RAM memory on a Windows 10 64-bit OS. All linear and integer programs are solved using IBM ILOG CPLEX 12.9, implemented in C++ with Concert Technology, with default parameter settings, and with a precision of 10^{-4} to avoid numerical issues.

In order to decompose X^{RSD} , this matrix first needs to be computed. Although the RSD algorithm can be easily executed by randomly selecting an ordering $\sigma \in \Sigma'$, Aziz et al. (2013) showed that computing X^{RSD} is $\#\mathcal{P}$ -complete, and thus intractable. Moreover, Saban & Sethuraman (2015) found that X^{RSD} cannot even be approximated efficiently. We therefore estimate X^{RSD} in this section by computing Equation (1) over a random sample $\hat{\Sigma}' \subseteq \Sigma'$ of the orderings, and we opt for sample size $|\hat{\Sigma}'| = 10,000$. Because $\hat{\Sigma}'$ is a random sample, both the resulting estimate of X^{RSD} and the solution of the column generation framework are unbiased.

Moreover, we determine $p^-(I)$, i.e., the cardinality of the minimum-cardinality ex-post efficient matching in a one-sided matching instance $I \in \mathcal{I}$, by formulating an IP model with constraints $[\text{PP}_0]$ and $[\text{PP}_{\text{PE}}]$ from the pricing problem while minimizing the number of assigned agents. Note that despite the \mathcal{NP} -hardness of this problem (Abraham et al., 2004), the running times are acceptable for the considered instances.

In all experiments, we include an initial subset of matchings as variables in the initial restricted master problem $[\text{RMP}]$ or $[\alpha\text{RMP}]$. Here, a trade-off emerges: whereas a larger sample size is likely to decrease the number of times the pricing problem will be called, it also tends to increase the solution time of the restricted master problem. To balance these two effects, we adopt a sample size of 10,000 random orderings. For column generation framework $[\text{RMP}]$, described in Sections 5.1 and 5.2, we only retain the matchings that assign at least k agents to an object, where $k \in \mathbb{N}$ is the value to be checked in the binary search. When solving $[\text{RMP}]$ for $\lfloor \mu(X^{\text{RSD}}) \rfloor$, for example, generally slightly more than half of the sampled matchings are retained. Moreover, we start the binary search for both column generation frameworks by setting $k = \lfloor \mu(X^{\text{RSD}}) \rfloor$.

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

6.2. Data generation

We have developed a parameterized data generation tool that is based on the properties of two real-world data sets from the school choice problem in the Belgian cities Antwerp and Ghent (see Section 6.4 for more details). Next to basic parameters, such as the objects' capacities or the length of the preference lists, we also consider the popularity of the objects in our data generation tool. More specifically, the user can specify the desired correlation between an object's capacity and its popularity (ρ , see Table 1), to what extent agents who submit a longer preference list include more popular objects on average (Δ_1), and to what extent the popularity of the objects in an agent's preference list is influenced by the popularity of their first choice (Δ_2). Table 1 contains all parameters of the data generation and their default values, which are based on the data of Ghent, and which we used to generate the instances used in the computational experiments. The data generation proceeds as follows.

Denote by $l_i \in \mathbb{N}$ the preference list length of agent $i \in N$, i.e., the number of objects that are preferred to the outside option by agent i , and denote $l = (l_1, \dots, l_{|N|})$. Recall that we denote the capacities of the objects by $q = (q_1, \dots, q_{|O|})$. Moreover, define the *popularity* $\eta_j \in \mathbb{R}^+$ of an object $j \in O$ as the ratio of the number of times j is preferred to the outside option over the capacity of that object, and denote $\eta = (\eta_1, \dots, \eta_{|O|})$. The fraction $\xi \in [0, 1]$ of the objects with the highest popularity are called *popular*. The data generation itself consists of the following four steps.

- (i) Generate the preference list lengths l_i from a truncated normal distribution $\mathcal{N}(\bar{l}, s^2)$ such that $l_i \geq 1$, and round to the closest integer, for all agents $i \in N$.
- (ii) Generate η and q from truncated standard normal distributions, and ensure that their correlation equals ρ by multiplying them with the upper triangle matrix in the Cholesky decomposition of the covariance matrix corresponding to ρ .
- (iii) Rescale η and q to satisfy the desired means and standard deviations, and round the elements of η and q to the closest integers.
- (iv) Fill in the preference profile $>$. First, determine for each agent $i \in N$ and for each position $t \leq l_i$ in $>_i$ the probability π_{it} with which a popular object is selected, in function of Δ_1 and Δ_2 (as defined in Table 1). Second, each object $j \in O$ in the group of popular, resp. unpopular, objects is

This is an author-version of

<https://doi.org/10.1016/j.ejor.2022.07.013>, published in *European Journal of Operational Research*, Volume 305, Issue 3, Pages 1087-1099 (2023).

6.3 Observations for generated data COMPUTATIONAL EXPERIMENTS

selected with probability $\pi_{it} \frac{\eta_j q_j}{Q}$, resp. $(1 - \pi_{it}) \frac{\eta_j q_j}{Q}$, where Q is the sum of $\eta_k q_k$ for the (un)popular objects $k \in O$ that are not yet in $>_i$. If $Q = 0$ for the popular (unpopular) objects for some values of i and t , then π_{it} is set to zero (one).

Table 1: Data generation parameters and their default values.

Parameter	Description	Default value
C	$\sum_{j \in O} q_j / N $	1.20
\bar{l}	Mean preference list length l	2.42
s	Standard deviation of preference list length l	1.05
ξ	Fraction of the objects that are popular	0.10
ρ	Correlation between q and η	0.21
CV_c	The coefficient of variation of the capacities q	0.80
CV_η	The coefficient of variation of the popularity η	0.60
Δ_1	Difference in average popularity of the objects in a preference list of length $\frac{1}{2}(\max_{i \in N} l_i + 1)$, compared to in a preference list of length one	0.14
Δ_2	Difference in the probability of submitting a popular object between when the first choice is popular and unpopular	0.01

A detailed description of the data generation, its repositories and the generated instances are available online (<https://github.com/DemeulemeesterT/GOSMI.git>).

6.3. Observations for generated data

Table 2 summarizes the main findings of our computational experiments for generated data. We can conclude that both column generation frameworks that were introduced in Section 5 manage to solve relatively large instances to optimality in acceptable runtimes. Column generation framework [RMP] slightly outperforms [α RMP], because it requires less time, on average, and manages to solve more instances to optimality. In general, we observe that both an increase in the number of agents, and an increase in the number of objects per agent cause the runtimes to rise.

Furthermore, for all instances that were solved to optimality, the difference between $z(X^{\text{RSD}})$ and $p^-(I)$, the worst-case number of assigned agents by the RSD algorithm, is substantial, with an average increase of 5.42% of the total number of agents. In fact, this is the maximal possible increase, as we observe that for all optimally solved instances, $z(X^{\text{RSD}})$ is equal to its upper bound

6.3 Observations for generated data 6 COMPUTATIONAL EXPERIMENTS

Table 2: Computational results of column generation frameworks [RMP] and [α RMP] for MD-SD(X^{RSD}). The column CPU displays the average computation time (in s) when optimality was obtained within one hour, and the column \checkmark contains the number of instances solved to optimality, out of 25 instances.

N	N / O = 1				N / O = 10				N / O = 50			
	[RMP]		[α RMP]		[RMP]		[α RMP]		[RMP]		[α RMP]	
	CPU	\checkmark	CPU	\checkmark	CPU	\checkmark	CPU	\checkmark	CPU	\checkmark	CPU	\checkmark
10	0.08	25	0.05	25								
50	0.70	25	0.22	25	0.23	25	0.16	25				
100	2.15	25	0.53	25	0.59	25	0.42	25	0.36	25	0.31	25
500	74.17	25	161.45	22	26.82	25	20.71	22	6.57	25	35.36	24
1,000	1,136.75	24	946.44	17	868.26	22	147.55	8	163.10	25	468.65	13
1,500	–	0	–	0	–	0	–	0	1,311.87	16	1,136.59	5

Table 3: Average objective values for the MD-SD(X^{RSD}) instances that could not be solved to optimality within one hour, for column generation frameworks [RMP] and [α RMP].

N	N / O = 1		N / O = 10		N / O = 50	
	[RMP]	[α RMP]	[RMP]	[α RMP]	[RMP]	[α RMP]
500	–	0.99977	–	0.99978	–	0.9998
1,000	$2.3 \cdot 10^{-4}$	0.99895	$3.1 \cdot 10^{-4}$	0.99174	–	0.99618
1,500	$8.6 \cdot 10^{-4}$	0.97458	$8.1 \cdot 10^{-4}$	0.96016	$2.6 \cdot 10^{-4}$	0.98266

$[\mu(X^{\text{RSD}})]$, despite the worst-case result of Theorem 4. This observation can be explained in two different ways. First, Che & Kojima (2010) showed that X^{PS} and X^{RSD} become equivalent when the market becomes large, and we know from Theorem 2 that $z(X^{\text{PS}}) = [\mu(X^{\text{PS}})]$. Second, an instance similar to the instance I_k that was given in Theorem 4 to prove the lower bound on $z(X^{\text{RSD}})$ is unlikely to occur in practice, because it requires a fraction $\frac{k^2-k}{k^2}$ of the agents to all submit the same object as their unique preference, and this fraction goes to one for large values of k .

Table 3 shows that for the instances that could not be solved to optimality within the runtime limit of one hour, the objective values of [RMP] and [α RMP] are close to their best possible values. Recall that the objective value s^* of [RMP] represents the maximum difference for any agent-object pair between the generated probabilistic assignment by [RMP] and X^{RSD} , while the objective value α^* of [α RMP] represents the cumulative weights of the matchings that assign at least $[\mu(X^{\text{RSD}})]$ agents to an object in the decomposition of X^{RSD} . For [RMP], the worst value of s^* over all unsolved instances is $2.20 \cdot 10^{-3}$, for example, whereas the largest value of s^* over all unsolved instances with

1,000 agents is only $3.4 \cdot 10^{-4}$. Similarly, for $[\alpha\text{RMP}]$, the worst value of α^* over all unsolved instances is 0.92543, whereas the smallest value of α^* over all unsolved instances with 1,000 agents is only 0.97272. We argue that, even when optimality cannot be obtained, the resulting decompositions are still valuable for real-world applications. Both methods will return decompositions in which all matchings are ex-post efficient, but the choice of the method depends on the application. If assigning at least $\lfloor \mu(X^{\text{RSD}}) \rfloor$ agents to an object is more important than decomposing X^{RSD} exactly, then the solution to $[\text{RMP}]$ provides a decomposition in which the implemented probability for each agent-object pair is at most s^* higher than in X^{RSD} , because of constraints (2b). Alternatively, if decomposing X^{RSD} exactly is of higher importance, then $[\alpha\text{RMP}]$ obtains a decomposition in which a fraction α^* of the matchings assign at least $\lfloor \mu(X^{\text{RSD}}) \rfloor$ agents to an object.

Lastly, we have also evaluated the algorithm we introduced in the proof of Theorem 1 to decompose X^{RSD} such that all matchings in the decomposition assign at least $z(X^{\text{RSD}})$ agents to an object. For the instances of Table 2, the algorithm only resulted in an ex-post efficient decomposition for some instances with a small number of objects, namely for 20 among the 75 instances where $(|N|, |O|) \in \{(10, 10), (50, 5), (100, 2)\}$. The algorithm from Theorem 1 is therefore not suited to find an optimal decomposition for $\text{MD-SD}(X)$ when X is not robust ex-post efficient.

6.4. Antwerp and Ghent

Next to evaluating our methods on generated data, we also consider two real-world school choice data sets from the cities of Antwerp and Ghent. The data from Antwerp corresponds to the primary school enrollment in the scholastic year of 2014-2015, and contains 4,236 students and 186 schools. The data of Ghent corresponds to the secondary school enrollment in 2018-2019, and contains 3,081 students and 64 schools. For confidentiality reasons, the data sets cannot be made publicly available.

In both data sets, the possible increase that we can realize in the worst-case number of assigned agents is substantial, compared to the RSD algorithm. For Antwerp, this increase is equal to 295 students, which is an increase by 6.96% of the total number of students, while the increase in Ghent equals 162 students, or 5.26%.

When applying column generation framework $[\text{RMP}]$ of Sections 5.1 and 5.2 to both instances, we notice that a larger initial number of included matchings

leads to better overall solutions. For Ghent, for example, the best found objective value for [RMP] is equal to 0.00161 after twenty-four hours of computing when starting from an initial sample size of 10,000 matchings, among which the matchings that assigned more than $\lfloor \mu(X^{\text{RSD}}) \rfloor$ students were retained. When starting from 50,000 matchings, however, an optimal decomposition is found after almost five hours. Similarly, for Antwerp the objective value decreases from 0.00579 after twenty-four hours when sampling 10,000 initial matchings to $8.7 \cdot 10^{-4}$ after only two hours for an initial sample of 50,000 matchings.

We observe similar behaviour for the alternative column generation framework [α RMP]. For Ghent, when initially including 10,000 matchings, the best found objective value is 0.85873 after twenty-four hours, while an objective value of 0.99900 is found after two hours, in the first iteration, when including 50,000 matchings. Performing the same test for Antwerp causes the objective value to rise from 0.64919 to 0.93339.

7. Conclusion

In this paper, we have studied the problem of decomposing a probabilistic assignment for one-sided matching over ex-post efficient matchings while maximizing the worst-case number of assigned agents. With respect to the two most studied mechanisms for one-sided matching, PS and RSD, we have obtained the following insights.

We have established that it is always possible to decompose the probabilistic assignment by the PS mechanism in polynomial time over ex-post efficient matchings that all assign the expected number of assigned agents by PS, either rounded up or down. The same result does not hold for the RSD mechanism, however. On the one hand, it is possible for the RSD algorithm to assign only half of the optimal worst-case number of agents to an object. On the other hand, instances exist for which we cannot realize any improvement in the worst-case number of assigned agents when decomposing the probabilistic assignment by RSD over ex-post efficient matchings.

In our computational experiments, we have found that for all generated instances that were solved to optimality, the optimal worst-case number of assigned agents when decomposing the probabilistic assignment by RSD over ex-post efficient matchings is equal to the expected number of assigned agents by RSD, rounded down. This promising result encourages the adoption of our solution methods in practical applications. By applying the column generation

frameworks that we have introduced to real-world school choice data sets, we have illustrated that even if optimality cannot be obtained in the desired runtime, the found solution can still be highly valuable in practice.

Acknowledgements Tom Demeulemeester is funded by PhD fellowship 11J8721N of Research Foundation - Flanders. Ben Hermans is funded by post-doc fellowship 12ZZI21N of Research Foundation - Flanders. We are grateful to Steven Penneman from the city of Antwerp, and Pieter De Wilde from the city of Ghent for their collaboration and for making their data sets available to us. Moreover, we would like to thank Markus Brill, Ágnes Cseh, and the members of the ALGO research group at TU Berlin for their feedback. Lastly, we would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

- Abdulkadiroğlu, A., & Sönmez, T. (1998). Random Serial Dictatorship and the core from random endowments in House Allocation problems. *Econometrica*, *66*, 689–701.
- Abraham, D. J., Cechlárová, K., Manlove, D. F., & Mehlhorn, K. (2004). Pareto optimality in house allocation problems. In *International Symposium on Algorithms and Computation* (pp. 3–15). Springer.
- Akbarpour, M., & Nikzad, A. (2020). Approximate random allocation mechanisms. *The Review of Economic Studies*, *87*, 2473–2510.
- Ashlagi, I., Saberi, A., & Shamel, A. (2020). Assignment mechanisms under distributional constraints. *Operations Research*, *68*, 467–479.
- Ashlagi, I., & Shi, P. (2016). Optimal allocation without money: An engineering approach. *Management Science*, *62*, 1078–1097.
- Athanassoglou, S., & Sethuraman, J. (2011). House allocation with fractional endowments. *International Journal of Game Theory*, *40*, 481–513.
- Aziz, H. (2020). Simultaneously achieving ex-ante and ex-post fairness. In *International Conference on Web and Internet Economics* (pp. 341–355). Springer.
- Aziz, H., Brandt, F., & Brill, M. (2013). The computational complexity of Random Serial Dictatorship. *Economics Letters*, *121*, 341–345.
- Aziz, H., Mackenzie, S., Xia, L., & Ye, C. (2014). Structure and complexity of ex post efficient random assignments. *arXiv:1409.6076 [cs.GT]*, .

- Aziz, H., Mackenzie, S., Xia, L., & Ye, C. (2015). Ex post efficiency of random assignments. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1639–1640).
- Bertsimas, D., & Tsitsiklis, J. N. (1997). *Introduction to Linear Optimization* volume 6. Athena Scientific Belmont, MA.
- Bhalgat, A., Chakrabarty, D., & Khanna, S. (2011). Social welfare in one-sided matching markets without money. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques* (pp. 87–98). Springer.
- Birkhoff, G. (1946). Three observations on linear algebra. *Universidad Nacional de Tucuman Revista, Serie A*, 5, 147–151.
- Biró, P. (2017). Applications of matching models under preferences. In U. Endriss (Ed.), *Trends in Computational Social Choice* chapter 18. (pp. 345–373). AI Access.
- Biró, P., & Gudmundsson, J. (2021). Complexity of finding Pareto-efficient allocations of highest welfare. *European Journal of Operational Research*, 291, 614–628.
- Bogomolnaia, A., & Moulin, H. (2001). A new solution to the random assignment problem. *Journal of Economic Theory*, 100, 295–328.
- Bogomolnaia, A., & Moulin, H. (2015). Size versus fairness in the assignment problem. *Games and Economic Behavior*, 90, 119–127.
- Bronfman, S., Alon, N., Hassidim, A., & Romm, A. (2018). Redesigning the Israeli medical internship match. *ACM Transactions on Economics and Computation*, 6, 1–18.
- Budish, E., Che, Y.-K., Kojima, F., & Milgrom, P. (2013). Designing random allocation mechanisms: Theory and applications. *American Economic Review*, 103, 585–623.
- Che, Y.-K., & Kojima, F. (2010). Asymptotic equivalence of probabilistic serial and random priority mechanisms. *Econometrica*, 78, 1625–1672.
- Dell’Amico, M., & Martello, S. (1997). The k -cardinality assignment problem. *Discrete Applied Mathematics*, 76, 103–121.

- Elster, J. (1992). *Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens*. Russell Sage Foundation.
- Freeman, R., Shah, N., & Vaish, R. (2020). Best of both worlds: Ex-ante and ex-post fairness in resource allocation. In *Proceedings of the 21st ACM Conference on Economics and Computation* (pp. 21–22).
- Kavitha, T., Király, T., Matuschke, J., Schlotter, I., & Schmidt-Kraepelin, U. (2022). The popular assignment problem: when cardinality is more important than popularity. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (pp. 103–123). SIAM.
- Kavitha, T., Mestre, J., & Nasre, M. (2011). Popular mixed matchings. *Theoretical Computer Science*, 412, 2679–2690.
- Kesten, O., Kurino, M., & Nesterov, A. S. (2017). Efficient lottery design. *Social Choice and Welfare*, 48, 31–57.
- Kondratev, A. Y., & Nesterov, A. S. (2022). Minimal envy and popular matchings. *European Journal of Operational Research*, 296, 776–787.
- Krysta, P., Manlove, D., Rastegari, B., & Zhang, J. (2019). Size versus truthfulness in the house allocation problem. *Algorithmica*, 81, 3422–3463.
- Martini, G. (2016). Strategy-proof and fair assignment is wasteful. *Games and Economic Behavior*, 98, 172–179.
- McCutchen, R. M. (2008). The least-unpopularity-factor and least-unpopularity-margin criteria for matching problems with one-sided preferences. In *Latin American Symposium on Theoretical Informatics* (pp. 593–604). Springer.
- Nesterov, A. S. (2017). Fairness and efficiency in strategy-proof object allocation mechanisms. *Journal of Economic Theory*, 170, 145–168.
- Ramezani, R., & Feizi, M. (2022). Robust ex-post pareto efficiency and fairness in random assignments: Two impossibility results. *Available at SSRN 4037462*, .
- Saban, D., & Sethuraman, J. (2015). The complexity of computing the Random Priority allocation matrix. *Mathematics of Operations Research*, 40, 1005–1014.

Shi, P. (2022). Optimal priority-based allocation mechanisms. *Management Science*, 68, 171–188.

von Neumann, J. (1953). A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games*, 2. Edited by W. Kuhn and A.W. Tucker. Princeton: Princeton University Press, 1997.

Zhou, L. (1990). On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory*, 52, 123–135.

Appendix A. Modified algorithm of Budish et al. (2013)

This appendix describes a modified version of the algorithm of Budish et al. (2013) that, given a probabilistic assignment $X \in \Delta\mathcal{M}$ with $\mu(X) \in \mathbb{N}$, constructs a matching $M \in \mathcal{M}$ that satisfies Properties (i)-(ii) as stated in the proof of Theorem 1. Our approach simplifies the one of Budish et al. (2013) as it considers only the specific constraint structure \mathcal{H} with quotas $(q_S)_{S \in \mathcal{H}}$ corresponding to $\Delta\mathcal{M}$, as introduced in the proof of Theorem 1, rather than the more general bihierarchical structure considered by Budish et al. (2013). This restriction enables us to reduce the $\mathcal{O}(|\mathcal{H}|^2)$ running time of the algorithm by Budish et al. (2013) to $\mathcal{O}(|\mathcal{H}| \min\{|N|, |O|\})$.

Given an assignment $X \in \Delta\mathcal{M}$, consider a bipartite graph $B(X) = (V, E(X))$ with vertices $V = N \cup O \cup \{s, t\}$ and a set of edges $E(X)$ consisting of

- (i) an edge $\{i, j\}$ for every $(i, j) \in N \times O$ for which x_{ij} is fractional,
- (ii) an edge $\{s, i\}$ for every $i \in N$ for which $\sum_{j \in O} x_{ij}$ is fractional,
- (iii) and an edge $\{j, t\}$ for every $j \in O$ for which $\sum_{i \in N} x_{ij}$ is fractional.

Every iteration u of the algorithm starts from a given probabilistic assignment X^u and a graph $B(X^u)$, where $X^u = X$ if $u = 1$, and X^u is defined in the previous iteration otherwise. If X^u is a matching, i.e., if $X^u \in \mathcal{M}$, then the procedure outputs $M = X^u$ and terminates. Otherwise, we determine a cycle $C = \langle i_1, j_1, i_2, j_2, \dots, i_r, j_r \rangle$ in $B(X^u)$ with $\{i_k, j_k\}, \{j_k, i_{k+1}\} \in E(X^u)$ for every $k = 1, \dots, r$ and $i_{r+1} = i_1$, where the indices are such that $i_k \in N \cup \{t\}$ and $j_k \in O \cup \{s\}$ for every $k = 1, \dots, r$. Next, we let α be the largest real

number such that the probabilistic assignment X^{u+1} with

$$x_{ij}^{u+1} = \begin{cases} x_{ij}^u + \alpha & \text{if } (i, j) = (i_k, j_k) \text{ for some } k = 1, \dots, r, \\ x_{ij}^u - \alpha & \text{if } (i, j) = (i_{k+1}, j_k) \text{ for some } k = 1, \dots, r, \\ x_{ij}^u & \text{otherwise,} \end{cases}$$

is feasible, update the graph $B(X^{u+1})$, and proceed to the next iteration.

To see that the algorithm yields a matching with the desired properties in time $\mathcal{O}(|\mathcal{H}| \min\{|N|, |O|\})$, observe first that in every iteration u the constructed assignment $X^u \in \Delta\mathcal{M}$ satisfies Properties (i)-(ii) as stated in the proof of Theorem 1. Moreover, it follows from the construction of $B(X^u)$ that the degree of each vertex in $B(X^u)$ is either zero or at least two and, since $B(X^u)$ is bipartite, that each cycle contains at most $2 \min\{|N \cup \{t\}|, |O \cup \{s\}|\}$ edges. Hence, if there is at least one $(i, j) \in N \times O$ for which x_{ij}^u is fractional, then we can obtain a cycle of the desired form, compute α , and determine X^{u+1} and $B(X^{u+1})$ in time $\mathcal{O}(\min\{|N|, |O|\})$. Finally, since α is chosen such that $\tau(X^{u+1}) > \tau(X^u)$, where τ is defined as in the proof of Theorem 1, the algorithm terminates after at most $|\mathcal{H}|$ iterations.

Appendix B. Proof of Theorem 4

Consider a family of one-sided matching instances $I_k = (N, O, >, q)$, with $k \geq 2$ an integer. Let $N = \{1, 2, \dots, k^2\}$ be a set of k^2 agents, and let $O = \{o_1, o_2, \dots, o_{k+1}\}$ be a set of $k + 1$ objects. Let the capacities of the objects in O be equal to $q = (k, 1, 1, \dots, 1)$. Moreover, let the preference list $>_i$ of each agent $i \in N$ be equal to

$$>_i = \begin{cases} o_1 > o_2 > \dots > o_{k+1} > \emptyset & \text{if } i \leq k, \\ o_1 > \emptyset > o_2 > \dots > o_{k+1} & \text{if } i > k. \end{cases}$$

We now show that $\mu(X^{\text{RSD}(I_k)}) = 2k - 1$ and $z(X^{\text{RSD}(I_k)}) = k$ for every $k \geq 2$. First, consider an arbitrary strict ordering $\sigma \in \Sigma'$ of the agents that is used by the SD mechanism to construct an ex-post efficient matching $\text{SD}(I_k, \sigma) \in \mathcal{M}^{\text{SD}}$. The preference lists are such that each agent $i \leq k$ always receives an object and such that an agent $i > k$ only receives an object if it is among the first k agents in the ordering σ . Hence, if we denote by τ_i the probability that agent i receives an object by applying the SD mechanism, then we obtain that $\tau_i = 1$

if $i \leq k$, and, since there are k^2 agents, that $\tau_i = k/k^2 = 1/k$ if $i > k$. It follows that

$$\mu(X^{\text{RSD}(I_k)}) = \sum_{i=1}^{k^2} \tau_i = \sum_{i=1}^k \tau_i + \sum_{i=k+1}^{k^2} \tau_i = k + \frac{k^2 - k}{k} = 2k - 1.$$

Second, to prove that $z(X^{\text{RSD}(I_k)}) = k$, recall that the set of all ex-post efficient matchings that assign at least $r \in \mathbb{N}$ agents to an object is denoted by $\mathcal{M}_r^{\text{SD}} \subseteq \mathcal{M}^{\text{SD}}$. In instance I_k , denote the unique ex-post efficient matching that assigns k agents to an object by $M^0 \in \mathcal{M}^{\text{SD}}$, and note that all other matchings in \mathcal{M}^{SD} assign strictly more than k agents to an object. Moreover, note that matching $\text{SD}(I_k, \sigma)$ will only be equal to M^0 if all agents $i \leq k$ are ranked before all $k^2 - k$ other agents $i > k$ in ordering σ . Denote the probability that the RSD algorithm will obtain matching M^0 in instance I_k by $\theta_k = \binom{k^2}{k}^{-1} > 0$. Denoting the weight of matching $M^t \in \mathcal{M}^{\text{SD}}$ by $\lambda^t \geq 0$, every feasible decomposition of $X^{\text{RSD}(I_k)}$ should satisfy

$$\sum_{t: M^t \in \mathcal{M}_{k+1}^{\text{SD}}} \lambda^t = \sum_{i=1}^k x_{i2}^{\text{RSD}(I_k)} = 1 - \theta_k < 1, \quad (\text{B.1})$$

because we know that every matching in $\mathcal{M}_{k+1}^{\text{SD}}$ assigns exactly one agent $i \leq k$ to object o_2 . As $\mathcal{M}^{\text{SD}} = \mathcal{M}_{k+1}^{\text{SD}} \cup M^0$ and $M^0 \notin \mathcal{M}_{k+1}^{\text{SD}}$, Equation (B.1) implies that $\lambda^0 > 0$. This, in turn, implies that $z(X^{\text{RSD}(I_k)}) = k$, because every feasible decomposition of $X^{\text{RSD}(I_k)}$ over ex-post efficient matchings has a strictly positive weight for the matching M^0 that assigns exactly k agents to an object.

Combining equalities $\mu(X^{\text{RSD}(I_k)}) = 2k - 1$ and $z(X^{\text{RSD}(I_k)}) = k$, we obtain that

$$\lim_{k \rightarrow \infty} \frac{z(X^{\text{RSD}(I_k)})}{[\mu(X^{\text{RSD}(I_k)})]} = \lim_{k \rightarrow \infty} \frac{k}{2k - 1} = \frac{1}{2},$$

which proves the theorem.

Appendix C. Proof of Theorem 5

Consider a family of one-sided matching instances $I_\ell = (N, O, >, q)$, with $\ell \geq 2$ an integer. Let $N = \{1, 2, \dots, \ell^2\}$ be a set of ℓ^2 agents, and let $O = \{o_1, o_2\}$ be a set of two objects. Let the capacities of the objects in O be equal to $q = (\ell, \ell)$. Moreover, let the preference list $>_i$ of each agent $i \in N$ be equal

to

$$>_i = \begin{cases} o_1 > o_2 > \emptyset & \text{if } i \leq \ell, \\ o_1 > \emptyset > o_2 & \text{if } i > \ell. \end{cases}$$

First, the minimum-cardinality ex-post efficient matching in this instance will be equal to $\text{SD}(I_\ell, \sigma)$ for every strict ordering $\sigma \in \Sigma'$ that ranks all agents $i \leq \ell$ before agents $i > \ell$. Hence, $p^-(I_\ell) = \ell$.

Second, we show that we can decompose $X^{\text{RSD}(I_\ell)}$ over ex-post efficient matchings that all assign $2\ell - 1$ agents to an object, i.e., $z(X^{\text{RSD}(I_\ell)}) = 2\ell - 1$. To construct a decomposition of $X^{\text{RSD}(I_\ell)}$ over $\mathcal{M}_{2\ell-1}^{\text{SD}}$, consider the following set of ℓ matchings $\{M^t\}_{t=1}^\ell$, where

$$M^t(i) = \begin{cases} o_1 & \text{if } i = t, \\ o_2 & \text{if } i \leq \ell \text{ and } i \neq t, \\ o_1 & \text{if } i \in \cup_{\gamma=2}^\ell \{(\ell-1)t + \gamma\}, \\ \emptyset & \text{otherwise.} \end{cases}$$

Clearly, all matchings in $\{M^t\}_{t=1}^\ell$ are ex-post efficient and assign exactly $2\ell - 1$ agents to an object. Note that each agent is assigned to o_1 by exactly one matching in $\{M^t\}_{t=1}^\ell$, and that each agent $i \leq \ell$ will always be assigned to an object in these matchings. Moreover, note that

$$x_{ij}^{\text{RSD}(I_\ell)} = \begin{cases} 1/\ell & \text{if } j = 1, \\ 1 - 1/\ell & \text{if } j = 2 \text{ and } i \leq \ell, \\ 0 & \text{if } j = 2 \text{ and } i > \ell. \end{cases}$$

Therefore, a feasible decomposition of $X^{\text{RSD}(I_\ell)}$ over $\mathcal{M}_{2\ell-1}^{\text{SD}}$ can be constructed by giving matching M^t a weight of $\lambda^t = 1/\ell$, for each $t = 1, \dots, \ell$.

Combining both results, we obtain that

$$\lim_{\ell \rightarrow \infty} \frac{z(X^{\text{RSD}(I_\ell)})}{p^-(I_\ell)} = \lim_{\ell \rightarrow \infty} \frac{2\ell - 1}{\ell} = 2,$$

which proves the theorem.