CSC_42021_EP PI: Matching under constraints

Patrick Loiseau

patrick.loiseau@inria.fr

Final version – November 29, 2024

In this project, we study many-to-one matching problems (such as the matching of students to schools) where we have non-trivial constraints on the possible matchings. We consider in particular *non-discrimination constraints*, as well as budget constraints. We investigate both the deferred-acceptance algorithm and its extensions, as well as newly proposed algorithms specifically done for the constrained problem.

Project's rules: The preferred language is Python but Java or C++ are allowed. It is encouraged to read through the entire project before starting to code. For every algorithm throughout the project, it is asked to discuss its implementation and choice of data structure, and to derive its complexity. Section 5 is independent from the rest—it is not mandatory and will be treated as bonus.

1 Introduction

We consider a classical matching problem, which we describe following [2] (and using their notation).[1] Throughout the project, we use the terminology of matching *students* to *schools*, but the formalism is the same for many other many-to-one matching problems (e.g., allocation of daycare slots to babies, allocation of professors to schools, etc.)—although the interpretation of some constraints may differ.

Let I be the set of students and S be the set of schools (both non-empty). Each student $i \in I$ has a strict preference relation denoted $\succ_i$ over the schools and being unmatched (denoted by $\emptyset$). Similarly, each school s has a strict preference relation denoted $\succ_s$ over the set of students. School s is said to be acceptable to student i if $s \succ_i \emptyset$. Each school considers all students as acceptable.

Each school is subject to a constraint $\mathcal{F}_s \in 2^I$, which corresponds to a collection of sets of students. A set $I' \subseteq I$ is feasible at school s if and only if $I' \in \mathcal{F}_s$. We define $\mathcal{F}_S = (\mathcal{F}_s)_{s \in S}$. This formalism allows specifying arbitrary constraints. In the project, we will consider special cases of constraints based on the characteristics of admitted students.

The students-to-schools problem is a many-to-one matching problem; that is, each student can be matched to at most one school but each school may admit multiple students. Formally, a matching μ is a mapping that satisfies:

[1] We give here the minimum amount of details needed for the project. We encourage students to go look at the original paper for additional details.

(i) $\mu_i \in S \cup \emptyset$ for all $i \in I$;

(ii) $\mu_s \subseteq I$ for all $s \in S$; and

(iii) $\mu_i = s$ if and only if $\mu_s \ni i$ for any $i \in I$ and $s \in S$.

In words, a matching simply specifies which student is assigned to which school (if any).

There are a number of potentially desirable characteristics of a matching. First, a matching μ is said to be *feasible* if $\mu_s \in \mathcal{F}_s$ for all $s \in S$. Second, a matching μ is *individually rational* if $\mu_i \succ_i \emptyset$ for all $i \in I$ such that $\mu_i \neq \emptyset$. Third, we say that i has a *justified envy* towards i' if there exists s such that $s \succ_i \mu_i$, $i' \in \mu_s$, and $i \succ_s i'$. A matching μ is fair[2] if there exist no students i and i' such that i has a justified envy toward i' (regardless of feasibility of the swap). Fourth, a matching μ is non-wasteful if there is no pair $(i, s) \in I \times S$ such that $s \succ_i \mu_i$ and $\mu_s \cup \{i\}$ is feasible at s. Finally, a matching μ is said to be stable if it is feasible, individual rational, fair, and non-wasteful.

Each student $i \in I$ is endowed with two characteristics: a group $g_i \in \mathcal{G}$ that belongs to a finite set $\mathcal{G}$ of possible groups, and a cost $w_i \in \mathbb{R}_+$. Intuitively, the student's group represents its demographic attributes such as gender, ethnicity, or disability; based on which discrimination is often forbidden (this will later appear as a constraint). The student's cost represents the cost for the school having this student. Different students might bring different costs; for instance disable students yield higher costs due to the need for special accommodations, international students yield higher costs due to the need to offer them courses of the local language, etc.

2 The deferred acceptance algorithm and its variants

2.1 Capacity constraints

We first consider the simplest case, where the only constraint is a *capacity constraint*. Each school s has a given capacity $q_s \in \mathbb{N}$ and can only admit up to q_s students.

Task 1: Show that, in this case, the above notion of stability coincides with the standard notion from Gale and Shapley [1].

Task 2: Devise and implement a variant of the deferred acceptance algorithm for the school admission problem with capacity constraints (i.e., the many-to-one matching problem).

2.2 Maximum quotas

The previous algorithm does not consider student's group. We first consider a simple case of affirmative action, based on maximum quotas. Specifically, in addition to its global quota q_s, each school s now also has a quota per group $q_s^g \in \mathbb{N}$, for all $g \in \mathcal{G}$. That is, school s can only admit up to q_s^g students of group g. We assume that $q_s^g \leq q_s$ for all g, and $\sum_{g \in \mathcal{G}} q_s^g \geq q_s$.

Task 3: Propose and implement a modification of the algorithm of Task 2 that outputs a matching respecting the above maximum quota constraints.

Remark: For the above task, we do not require that the obtained matching satisfies the exact stability notion definition defined in Section 1.

[2] Note that this "fairness" notion has nothing to do with non-discrimination of demographic groups. While non-discrimination is sometimes also called fairness, we stick to the terminology introduced to avoid confusion with the fairness notion in matching.

2.3 Tests

Task 4: Test the algorithms from Tasks 2 and 3 in the following instances:

Instance 1: There are 4 students $I = \{i_1, i_2, i_3, i_4\}$ and 2 schools $S = \{s_1, s_2\}$ with the following preferences:[3]

$$\begin{aligned} &\succ_{i_1}: s_1, s_2 & & \\ &\succ_{i_2}: s_2, s_1 & &\succ_{s_1}: i_4, i_3, i_2, i_1 \\ &\succ_{i_3}: s_1 & &\succ_{s_2}: i_4, i_3, i_2, i_1 \\ &\succ_{i_4}: s_2 & & \end{aligned}$$

The set of groups is $\mathcal{G} = \{A, B\}$. Students $I = \{i_1, i_2, i_3\}$ are of group A, student i_4 is of group B. The costs are $w_1 = w_2 = w_3 = 1$, $w_4 = 10$. Each school has a capacity of 2 slots and a quota of 2 for each group.

For this instance, you will graphically represent the matching obtained.

Instance 2: There are n students $I = \{i_1, \cdots, i_n\}$ and 2 schools $S = \{s_1, s_2\}$. Student i's preference is $\succ_i$: s_1, s_2 or $\succ_i$: s_2, s_1, drawn randomly with uniform probability (independently of the student's group). Each student i has a latent quality W_i, drawn randomly according to a standard normal distribution. Each school observes a noisy version of the student's quality $\hat{W}_i = W_i + \epsilon_i$ where ϵ_i is a random noise that also follows a standard normal distribution. The noises are independent for the 2 schools. Then each school's preference list simply ranks the students in decreasing order of observed quality $\hat{W}$.

The set of groups is $\mathcal{G} = \{A, B\}$. Students $I = \{i_1, \cdots, i_m\}$ are of group A, where $m = \lfloor p_A \cdot n \rfloor$ and $p_A = 9/10$ is the fraction of students of group A; other students are of group B. The costs are $w_A = 1$ for students of group A and $w_B = 10$ for students of group B. Each school has a capacity of $\lfloor n/4 \rfloor$, and a quota of $\lfloor 0.9n/4 \rfloor$ for each group.

For this instance, you will first generate a realization of this random instance and then do the matching on it. We ask that you compute (and show) the number of students that are assigned to their first choice, for each group. You will run multiple realizations and compute average values. You may test multiple values of n and investigate how the results depend on n.

Instance 3: Same as Instance 2 but with 4 groups $\{A, B, C, D\}$ of fractions $p_A = 10/20$, $p_B = 6/20$, $p_C = 3/20$. The costs are $w_A = 1, w_B = 5, w_C = 6, w_D = 10$ for students of groups A, B, C, D respectively. Each school still has a quota of $\lfloor 0.9n/4 \rfloor$ for each group.

Remark: The above instances define all parameters used throughout the project—not all are needed for this particular task (e.g., costs will only be used later).

3 Fair rankings

The maximum quotas constraints are limited in that they do not guarantee a *minimum* number of students from the different groups, which non-discrimination rules typically do. A more stringent

[3] We give preferences in the order most to least preferred. We truncate preference lists of students at $\emptyset$.

notion is known as the 4/5-rule: for any group, the fraction of students of the group at any school cannot be less than 80% or more than 120% of the fraction of students of that group in I.

Guaranteeing that the obtained matching satisfies the 4/5-rule is non-trivial. We consider here a first natural idea: we will modify the schools preference lists such that, for any rank k, the subset of students ranked k or better satisfy the 4/5-rule up to one unit.

Task 5: Propose and implement an algorithm to change schools preference lists so that they satisfy the 4/5-rule up to one unit while minimizing the distance to the initial ranking.

Task 6: Apply the algorithm from Task 2 on the new rankings in the instances from Task 4. Do the final matchings satisfy the 4/5-rule?

4 A fixed-point algorithm for arbitrary constraints

In a recent paper, Kamada and Kojima [2] proposed a new algorithm based on a fixed-point to find a matching with good properties under arbitrary constraints. With arbitrary constraints, finding a stable matching is impossible in general (see [2] for a counter example). Therefore, we will drop the property of non-wastefulness and look for matchings that are feasible, individually rational and fair. Kamada and Kojima [2] showed that such matchings can be characterized as the solutions to a fixed point problem as follows.

Consider the space of cutoff profiles $P = \{1, \cdots, |I|, |I|+1\}^S$ (i.e., the space of functions from S to $\{1, \cdots, |I|, |I|+1\}$), endowed with a partial order $\leq$ such that two profiles $p, p' \in P$ satisfy $p \leq p'$ if and only if p_s is weakly smaller than p'_s for all $s \in S$. For each school $s \in S$, let $i^{(s,l)}$ be the student whose rank is the l^{th} from the bottom according to the preference list of s (e.g., $i^{(s,|I|)}$ is the best student for s). Also consider a hypothetical student $i^* \notin I$ such that $i^* = i^{(s,|I|+1)}$ for all $s \in S$ and expand the domain of $\succ_s$ for each $s \in S$ so that $i^* \succ_s i$ for all $i \in I$. Given a cutoff profile $p \in P$, we define the demand at school s as

$$D_s(p) = \{i \in I | i \succeq_s i^{(s,p_s)} \text{ and } s \succ_i \emptyset; i \succeq_{s'} i^{(s',p_{s'})} \Rightarrow s \succeq_i s'\}.$$

In this definition, the first part "$i \succeq_s i^{(s,p_s)}$ and $s \succ_i \emptyset$" says that student i is as good as the cutoff student $i^{(s,p_s)}$ and finds s acceptable, while the second part "$i \succeq_{s'} i^{(s',p_{s'})} \Rightarrow s \succeq_i s'$" says that s is the most preferred school among the ones at which i passes the cutoff.[4] The demand for s is a collection of students who meet those two criteria.

We consider a mapping $T : P \to P$, called the cutoff adjustment function, defined as follows: for all $s \in S$

$$T_s(p) = \begin{cases} p_s + 1 & \text{if } D_s(p) \notin \mathcal{F}_s, \\ p_s & \text{if } D_s(p) \in \mathcal{F}_s; \end{cases}$$

where we set by convention $(|I|+1)+1 = 1$ to ensure that the range of T is P. That is, for each s and cutoff profile, this mapping raises the cutoff at s by one if the demand for s is infeasible, and leaves the cutoff unchanged if it is feasible.

For each $p \in P$, let μ^p be the matching such that

$$\mu^p_s = D_s(p) \text{ for each } s \in S.$$

That is, μ^p_s simply matches all students who demand s given cutoff profile p.

Then, Kamada and Kojima [2] show the following result:

[4] In particular, if $p_s = |I|+1$, then $i \succeq_s i^{(s,p_s)} = i^*$ (which is well defined as we expanded the domain of s) does not hold for any $i \in I$, so $D_s(p)$ is empty.

Theorem 1. *If a cutoff profile $p \in P$ is a fixed point of the cutoff adjustment function T, then μ^p is feasible, individually rational, and fair. Moreover, if μ is a feasible, individually rational, and fair matching, then there exists a cutoff profile $p \in P$ with $\mu = \mu^p$ that is a fixed point of T.*

There exist sufficient conditions to guarantee the existence of a fixed-point (called "general upper bound" [2]). Here, we admit that these conditions are satisfied for capacity constraints and for maximum quota constraints (as in Section 2.2). They are not satisfied, however, for constraints corresponding to the 4/5-rule.

Task 7: Implement the algorithm from [2] described above with a feasibility function encoding arbitrary feasible sets.

Task 8: Test your algorithm on the Instances 1-3 from Task 4, with only capacity constraints. Compare to the outcome of the algorithm from Task 2.

Task 9: Test your algorithm on the Instances 1-3 from Task 4 with the maximum quota constraint (in addition to capacity constraints). Compare to the outcome of the algorithm from Task 3.

Task 10: What happens if you try to run the algorithm with the 4/5-rule constraint?

Note that in these tasks, you will not use the costs defined in the Instances 1-3 from Task 4.

5 Matching without stability constraints

In this last section, we disregard stability (or fairness) conditions for the matchings and instead look for "optimal" matchings in a centralized sense (i.e., assuming that a central authority can enforce the matching even in the presence of blocking pairs). Instead of a capacity constraint, we now assume that each school s has a budget constraint W_s, that is, the school can admit a subset I' of students only if $\sum_{i \in I'} w_i \leq W_s$.

5.1 The case of a single school

We first consider the case where there is a single school. In this case, there is no longer a preference list for students as each student only has one possible choice of school.

Task 11: Propose and implement an algorithm that maximizes the number of assigned students with a budget constraints and the 4/5-rule constraint.

Task 12: Test your algorithm on the Instances 1-3 from Task 4 with the following modification: s_1 and s_2 are merged—i.e., they are the same school. The budget is $W_s = 11$ for Instance 1 and $W_s = 3n$ for Instances 2-3.

5.2 The case of a multiple schools

In the case of multiple schools, students preferences become important. To account for them, we define a penalty of l_i^2 for matching student i to the school corresponding to their l^{th} choice in their preference list according to $\succ_i$.[5] We assign a penalty $(|S|+1)^2$ for students who remain unmatched.

[5] A similar penalty is used in the algorithm to assign military training for X students.

Task 13: Propose and implement an algorithm that minimizes the total students penalty (i.e., the sum of penalties of all students) with a budget constraints and the 4/5-rule constraint.

Task 14: Test your algorithm on the Instances 1-3 from Task 4. The budget is $W_s = 11$ for Instance 1 and $W_s = 3n$ for Instances 2-3.

Note that in this last section, schools preferences have not been taken into account.

References

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

[2] Yuichiro Kamada and Fuhito Kojima. Fair matching under constraints: Theory and applications. *The Review of Economic Studies*, 91(2):1162–1199, 04 2023.