Higher education admission in Hungary by a "score-limit algorithm"*

Péter Biró

Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK
Email: pbiro@dcs.gla.ac.uk.

Abstract

The admission procedure of higher education institutions is organized by a centralized matching program in Hungary since 1985. We present the implemented algorithm, which is similar to the college-proposing version of the Gale-Shapley algorithm. The difference is that here ties must be handled, since applicants may have equal scores. The presented score-limit algorithm finds a solution that satisfies a special stability condition. We describe the applicant-proposing version of the above algorithm and we prove that the obtained solutions by these algorithms are the maximal and the minimal stable score-limits, respectively.

1 Introduction

The college admission problem was introduced and studied by Gale and Shapley [5]. Later Roth [8] described the history of the National Intern Matching Program, that have used the same type of algorithm since 1952. Further literature about the two-sided matching markets can be found in the book of Roth and Sotomayor [10].

Recently, the student admission problem came again into prominence (detailed description about several applications can be found in the paper of Abdulkadiroğlu and Sönmez [3]). New centralized matching programs have been implemented for public schools in Boston, and for high schools in New York (see [1] and [2]).

However, there are some studies about existing college admissions programs as well (see the papers [7] and [4] about the programs in Spain and in Turkey, respectively), the description of many other important implementations are not available in the literature.

In Hungary, the admission procedure of higher education institutions is organized by a centralized matching program. The Ministry of Education founded the Admission to Higher Education National Institute (OFI) in 1985 in order to create, operate and develop the admission system of the higher education. The number of applicants is around 150000 in each year, about 100000 of them are admitted, and the fees are payed by the state for approximately 60% of the students (exact statistics in Hungarian are available at [6]).

First, we note that instead of colleges, in Hungary the universities have faculties, where the education is organized in different fields of studies quite independently. So here, students apply for fields of studies of particular faculties. For simplicity, these fields are referred as colleges later in order to keep the original terminology of Gale and Shapley.

At the beginning of the procedure, students give their ranking lists that correspond to their preferences over the fields they apply for. There is no limit for the length of the list, however applicants are charged after each item. The students receive scores at each

*This work was supported by EPSRC grant EP/E011993/1 and by OTKA grant K69027.

field they applied for according to their final notes at the high school, and entrance exams. Note, that the score of a student can differ at two fields. These scores are integer numbers, currently limited to 144. Universities can admit a limited number of students to each of their fields, these *quotas* are determined by the Ministry of Education.[1]

The administration is organized by a state-owned center. After collecting the applicants' rankings and their scores, a centralized program computes the score-limits of the fields. An applicant is admitted by the first place on his list, where he is above the score-limit.

Here, we present the currently used basic algorithm that yields a kind of stable assignment. This algorithm is very similar to the Gale–Shapley [5] algorithm, in fact, if the score of the applicants are different at each place then this algorithm is equivalent to the college-proposing algorithm of Gale and Shapley. This explains why it is not suprising that similar statements can be proved for the score-limit algoritms. Here, we show that the score-limits at each field is maximal for the college-proposing version, and minimal for the applicant-proposing version in the set of the stable score-limits.

2 The definition of stable score-limit

Let $A = \{a_1, a_2, \ldots, a_n\}$ be the set of applicants and C be the set of colleges, where q_u denotes the quota of college c_u. Let the ranking of the applicant a_i be given by a preference list P^i, where $c_v >_i c_u$ denotes if c_v preceeds c_u in the list, i.e. if the applicant a_i prefers the college c_v to c_u. Let s^i_u be a_i's final score at the college c_u.

The score-limit l is a nonnegative integer mapping $l : C \to \mathbb{N}$. An applicant a_i is admitted by a college c_u, if he achieves the limit at college c_u, and that is the first such place in his list, i.e. $s^i_u \geq l(c_u)$, and $s^i_v < l(c_v)$ for every college $c_v >_i c_u$. If the score-limit l implies that a college c_u admits applicant a_i, then we set the boolean variable $x^i_u(l) = 1$, and 0 otherwise. Let $x_u(l) = \sum_i x^i_u(l)$ be the number of applicants allocated to c_u. A score-limit l is *feasible* if $x_u(l) \leq q_u$ for every college.

Let $l^{u,t}$ be defined as follows: $l^{u,t}(u) = l(u) - t$ and $l^{u,t}(v) = l(v)$ for every $v \neq u$. That is, we decrease the score limit by t at college c_u, by leaving the other limits unchanged. We say that a score-limit l is *stable* if l is feasible but for each college c_u, $l^{u,1}$ is not feasible. This stability condition means that no college can decrease its limit without violating its quota (assuming that the others do not change their limits). We note that if no ties occur (i.e. two applicants have different scores at each college), then this stability condition is equivalent to the original one by Gale and Shapley.

3 Score-limit algorithms and optimality

First we present the currently used algorithm and verify its correctness, then we describe its applicant-proposing version. Finally, we prove that these algorithms produce the maximal and the minimal stable score-limits, respectively.

The score-limit algorithms

Both score-limit algorithms are very similar to the two versions of the original Gale–Shapley algorithm. The only difference is that now, the colleges cannot select exactly as many best applicants as their quotas are, since the applicants may have equal scores.

[1] We describe some further specialities and requirements in the last subsection, that are not included in the presented basic model.

Here, instead each college sets its score-limits always to be the smallest one, such that its quota is not exceeded. If the scores of the applicants are distinct at each college then these algorithms are equivalent to the original ones.

College-proposing algorithm:

In the first stage of the algorithm, let us set the score-limit at each college independently to be the smallest value such that the number of admitted applicants does not exceed its quota by considering all its applications. Let us denote this limit by l_1. Obviously, there can be some applicants, who are admitted by several places. These applicants keep their best offer, and reject all the less preferred ones, moreover they cancel also their less preferred applications.

In the further stages, the colleges check whether their score-limits can be further decreased, since some of their applications may have been cancelled in the previous stage, hence they look for new students to fill up the empty places. So each college sets its score-limit independently to be the least possible, considering their actual applications. If an applicant is admitted by some new, better place, then he accepts the best offer in suspense, and rejects or cancels his other, worse applications.

Formally, let l_k be the score-limit after the k-th stage. In the next stage, at every college c_u, the largest integer t_u is chosen, such that $x_u(l_k^{u,t_u}) \leq q_u$. That is, by decreasing its score-limit by t_u, the number of admitted applicants by c_u does not exceed its quota, supposing that all other score-limits remained the same. For every college let $l_{k+1}(c_u) := l_k^{u,t_u}(c_u)$ be the new score-limit. If some limits are decreased simultaneously, then some applicants can be admitted by more than one place, so $x_u(l_{k+1}) \leq x_u(l_k^{u,t_u})$. Obviously, the new score-limit remains feasible.

Finally, if no college can decrease its limit, then the algorithm stops. The stability of the final score-limit is obvious by definition.

Example 1. *In this example we consider only 3 colleges, c_{cs}, c_e and c_m (i.e. college of computer science, economics and maths, respectively) and the effect caused by two applicants, a_i and a_j. We suppose that all the other applicants have only one place in their lists. The preferences of a_i and a_j are the following: $P^i = c_e, c_{cs}, c_m, \ldots$ and $P^j = c_{cs}, c_m, c_e, \ldots$. Their scores are: $s^i_{cs} = 112$, $s^i_e = 100$, $s^i_m = 117$, $s^j_{cs} = 110$, $s^j_e = 103$ and $s^j_m = 105$. Let the quotas be $q_{cs} = 500$, $q_e = 500$ and $q_m = 100$. We suppose that the number of applicants having*

- *at least* 110 points at c_{cs} is 510,
- *more than* 110 points at c_{cs} is 483,
- *at least* 100 points at c_e is 501,
- *more than* 100 points at c_e is 460,
- *at least* 105 points at c_m is 101,
- *more than* 105 points at c_m is 87.

In the first stage of the college-proposing algorithm the score-limits are $l_1(c_{cs}) = 111$, $l_1(c_e) = 101$ and $l_1(c_m) = 106$. At these limits a_i is admitted to the college of computer science and to the college of maths too, while a_j is admitted to the college of economics only. Since a_i prefers the computer science, he rejects the latter offer (and he cancels also his other less preferred applications.) Now, in the second stage, the score-limit can be decreased by one at the college of maths, because the number of currect applications having at least 105 points is exactly 100. In this way, a_j becomes admitted to this college, and since he prefers maths to economics, he rejects his offer there. In the third stage, the score-limit can be decreased by one at the college of economics. After this change a_i is

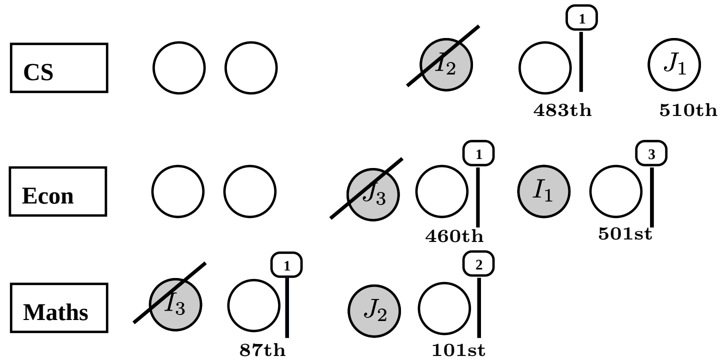

Figure 1: The score-limits in the college-proposing algorithm

admitted to the college of economics, that is his most preferred place, so he cancels all his other applications. In the final stage no score-limit can be decreased, so the algorithm stops.

Applicant-proposing algorithm:
Let each applicant propose to his first choise in his list. If a college receives more applications than its quota, then let this score-limit be the smallest value such that the number of temporary accepted applicants does not exceed its quota. We set the other limits to be 0.

Let the score-limit after the k-th stage be l_k. If an applicant has been rejected in the k-th stage, then let him apply for the next place in his list, say c_u where he achieves the actual score-limit $l_k(c_u)$, (if there remained such a place in his list). Some colleges may receive new proposals, so if the number of admitted applicants exceeds their quota at a college, they set a new, higher score-limit l_{k+1}. At the same time, they reject all those applicants that do not achieve this new limit.

The algorithm stops if there is no new application. The final score-limit is obviously feasible. It is also stable, because after a limit is increased for the last time, then the rejected applicants get worse and worse offers during the algorithm. So if the limit were decreased by one at the final solution in this place, then these applicants would accept the offer, and the quota would have been exceeded.

Theorem 3.1. *Both the score-limit l_C, obtained by the college-proposing algorithm and the score-limit l_A, obtained by the applicant-proposing algorithm are stable.*

Below, we give a simple example to show that not only some applicants can be admitted by preferred places in l_A than in l_C, but the number of admitted applicants can also be larger in l_A.

Example 2. *We consider only two places c_{cs} and c_e with two applicants a_i and a_j. We suppose that all the other applicants have only one single place in their lists. The preference-lists of a_i and a_j are $P^i = c_e, c_{cs}, \dots$ and $P^j = c_{cs}, c_e, \dots$, and their scores are: $s^i_{cs} = 112$, $s^i_e = 100$, $s^j_{cs} = 110$ and $s^j_e = 103$. Both places have quotas 500. We suppose that the number of applicants having*

- *at least* 110 points at c_{cs} is 501,
- *more than* 110 points at c_{cs} is 487,
- *at least* 100 points at c_e is 501,
- *more than* 100 points at c_e is 460.

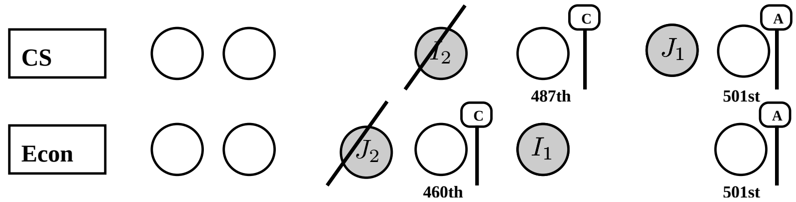

Figure 2: The final score-limits of the college-proposing and the applicants-proposing algorithm

Here, both algorithms stop after one stage. The final score-limit obtained by the college-proposing algorithm is $l_C(c_{cs}) = 111$ and $l_C(c_e) = 101$. The number of admitted applicants are $x_{cs}(l_C) = 487$ and $x_e(l_C) = 460$, respectively. While, the final score-limit obtained by the applicant-proposing algorithm is $l_A(c_{cs}) = 110$ and $l_A(c_e) = 100$. Moreover, the number of admitted applicants are 500 at both places. This extreme example shows that the difference between the solutions can be relevant.

The optimality

We say that a score-limit l is *better* than l_* for the applicants if $l \leq l_*$, (i.e. $l(c_u) \leq l_*(c_u)$ for every college c_u). In this case every applicant is admitted by the same or by a preferred place at score-limit l than at l_*.

Theorem 3.2. *l_C is the worst possible and l_A is the best possible stable score-limit for the applicants, i.e. for any stable score-limit l, $l_A \leq l \leq l_C$ holds.*

Proof. Both proofs are based on indirect arguments, that are similar to the original one of Gale and Shapley's.

Suppose first, that there exists a stable score-limit l_* and a college c_u such that $l_*(c_u) > l_C(c_u)$. During the college-proposing algorithm there must be two consecutive stages with score-limits l_k and l_{k+1}, such that $l_* \leq l_k$ and $l_*(c_u) > l_{k+1}(c_u)$ for some college c_u. Obviously, $l_k^{u,t_u}(c_u) = l_{k+1}(c_u)$ by definition and $x_u(l_k^{u,t_u}) \leq q_u < x_u(l_*^{u,1})$ by the stability of l_*. So, on the one hand, there must be an applicant, say a_i who is admitted by c_u at $l_*^{u,1}$ but not admitted by c_u at l_k^{u,t_u}. On the other hand, the indirect assumption $l_k^{u,t_u}(c_u) = l_{k+1}(c_u) \leq l_*(c_u) - 1 = l_*^{u,1}(c_u)$ implies that a_i must be admitted by another, preferred place than c_u at l_k^{u,t_u} (since a_i has at least $l_k^{u,t_u}(c_u)$ score there), and obviously also at l_k. That is impossible if $l_* \leq l_k$, a contradiction.

To prove the other direction, we suppose that there exist a stable score-limit l_* and a place c_u such that $l_*(c_u) < l_A(c_u)$. During the applicant-proposing algorithm there must be two consecutive stages with score-limits l_k and l_{k+1}, such that $l_* \geq l_k$ and $l_*(c_u) < l_{k+1}(c_u)$ for some college c_u. At this moment, the reason of the incrementation is that more than q_u students are applying for c_u with at least l_* score. This implies that one of these students, say a_i is not admitted by c_u at l_* (however he has at least $l_*(c_u)$ score there). So, by the stability of l_*, he must be admitted by a preferred place, say c_v at l_*. Consequently, a_i must have been rejected by c_v in a previous stage of the algorithm, that is possible only if $l_*(c_v) < l_k(c_v)$, a contradiction. $\square$

4 Further notes

There are many further rules required by the law. Some of them are considered in the present algorithm, some are handled manually afterwards.

At each place there is a minimum score that is generally equal to 60% of the maximum score (that is 144 points usually). If an applicant does not have the minimum score at a place, then this application is simply deleted.

In Hungary, some studies are completely financed by the state, some are partly financed by the students. At most of the places there are two different quotas for both kind of studies. The applicants have to indicate also in their rankings which kind of study they apply for at some field.[2] These are considered in the algorithm as distinct places with distinct quotas. However, there are some requirements on their score-limits: the difference between the score-limits of the state-financed and the privately-financed studies at the same place can not be more than 10%. This rule is tracted by the current algorithm.

Another speciality is that certain pairs of fields can be chosen simultaneously, and some others must come in pairs. These cases are solved manually after the first run of the program, and might cause overflowings.

An actual problem of the program is that the scoring system is not fine enough, that is why huge ties are likely to emerge. As a consequence, the difference between the quota and the number of admitted applicants can be large. Moreover, in an extreme case, if the number of applicants having maximum score is greater than the quota of that place, no student can be admitted. So, it is a good news, that recently a finer scoring system has been proposed in the actual law that will increase the maximum score from 144 to 480.

In our opinion, to change the direction of the algorithm would also be reasonable. Not just because some applicants could be admitted by preferred places, but also because the number of admitted applicants could increase too. We think that the effect of such a change would be more significant than the effect of a similar change in the National Resident Matching Program (see the study of Roth and Peranson [9] about this).

References

[1] A. Abdulkadiroğlu, A. E. P. A. Pathak, Roth, and T. Sönmez. The Boston public school match. *American Economic Review, Papers and Proceedings*, 95(2):368–371, 2005.

[2] A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth. The New York City high school match. *American Economic Review, Papers and Proceedings*, 95(2):364–367, 2005.

[3] Atila Abdulkadiroğlu and Tayfun Sönmez. School Choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

[4] M. Balinski and T. Sönmez. A tale of two mechanisms: student placement. *J. Econom. Theory*, 84(1):73–94, 1999.

[5] D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *Amer. Math. Monthly*, 69(1):9–15, 1962.

[6] Natinal Higher Education Information Centre. *http://www.felvi.hu*.

[7] A. Romero-Medina. Implementation of stable solutions in a restricted matching market. *Review of Economic Design*, 3(2):137–147, 1998.

[2] An applicant may rank first a state-financed study in economics at a university in Budapest, then secondly another state-financed study in economics at another university in Pécs, and thirdly a privately-financed study in economics at the first university in Budapest. So the fees are included in the preferences of the applicants in this way.

[8] A. E. Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy*, 6(4):991–1016, 1984.

[9] A. E. Roth and E. Peranson. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748–780, 1999.

[10] A. E. Roth and M. Sotomayor. *Two-sided matching*, volume 18 of *Econometric Society Monographs*. Cambridge University Press, Cambridge, 1990. A study in game-theoretic modeling and analysis, With a foreword by Robert Aumann.