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

Péter Biró*
Department of Computer Science and Information Theory
Budapest University of Technology and Economics
pbiro@cs.bme.hu

March 30, 2007

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.

Keywords: college admissions, mechanism design, stable marriage, two-sided market, many-to-one matching

Introduction

The college admission problem was studied by Gale and Shapley [3]. Later Roth [5] described the history of the National Resident Matching Program, that used the same type of algorithm from 1952. Further literature about the

*The affiliation of the author will change to Department of Computing Science, University of Glasgow from 1 June 2007.

two-sided matching markets can be found in the book of Roth and Sotomayor [7]. Recently, centralized matching programs have been implemented for student admission of public schools in Boston, and of high schools in New York (see [1] and [2]).

In Hungary, the admission procedure of higher education institutions is organized by a centralized matching program. The Ministry of Education has 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 [4]).

First, we note that instead of colleges, in Hungary the universities have faculties and the faculties have departments, where the education is organized quite independently. So here, students apply for departments.

At the beginning of the procedure, students give their ranking lists that correspond to their preferences over the departments they apply for. There is no limit for the length of the list, however the applicant are charged after each item. The students receive scores at each department they applied for according to their final notes at the high school, and the entrance exams. Note, that the scores of a student can differ at two departments. The scores are integer numbers, currently limited to 144. Each department can admit a limited number of students, these *quotas* are determined by the Ministry.[1]

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

Here, we present the used basic algorithm that preserve a kind of stable assignment. This algorithm is very similar to the Gale-Shapley [3] algorithm, in fact, if the score of the applicants are different at each department then this algorithm is equivalent to the college-proposing algorithm of Gale and Shapley. That is 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 department is maximal by the college-proposing version, and minimal by the applicant-proposing version in the set of the stable score-limits.

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

1 The definition of stable score-limit

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

The score-limit l is a nonnegative integer mapping $l : D \to \mathbb{N}$. An applicant a_i is admitted by department d_u, if he achieves the limit at department d_u, and that is the first such department in his list, i.e. $s^i_u \geq l(d_u)$, and $s^i_u < l(d_v)$ for every department $d_v >_i d_u$. If the score-limit l implies that the department d_u admit 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 d_u. A score-limit l is *feasible* if $x_u(l) \leq q_u$ for every department.

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 department d_u, by leaving the other limits unchanged. We say that a score-limit l is *stable* if l is feasible but for each department d_u, $l^{u,1}$ is not feasible. This stability condition means that no department can decrease its limit without violating its quota (assuming that the others do not change their limits).

2 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.

2.1 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 departments cannot select exactly as many best applicants as their quotas are, since the applicants can have equal scores. Here, instead each department sets its score-limit always to be the smallest one, for that its quota is not exceeded. If the scores of the applicant are distict at each department 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 department 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 departments. 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, each department checks whether its score-limit can be decreased, since some of its applications may have been cancelled in the previous stage, hence it looks for new students to fill up the empty places. So each department independently sets its score-limit to be the least possible, considering its actual applications. If an applicant is admitted by some new, better department, 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, each department, say d_u choses the largest integer t_u, 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 d_u does not exceed its quota, supposing that the other departments do not change their score-limits. For every department let $l_{k+1}(d_u) := l_k^{u,t_u}(d_u)$ be the new score-limit. If some departments decrease their limits simultaneously, then some applicants can be admitted by more than one department, so $x_u(l_{k+1}) \leq x_u(l_k^{u,t_u})$. Obviously, the new score-limit remains feasible.

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

Example 1 *In this example we consider only 3 departments, d_{cs}, d_e and d_m (i.e. department 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 department in their lists. The preferences of a_i and a_j are the following: $P^i = d_e, d_{cs}, d_m, \dots$ and $P^j = d_{cs}, d_m, d_e, \dots$. 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 d_{cs} is 510,
- *more than* 110 points at d_{cs} is 483,
- *at least* 100 points at d_e is 501,
- *more than* 100 points at d_e is 460,

- *at least* 105 points at d_m is 101,
- *more than* 105 points at d_m is 87.

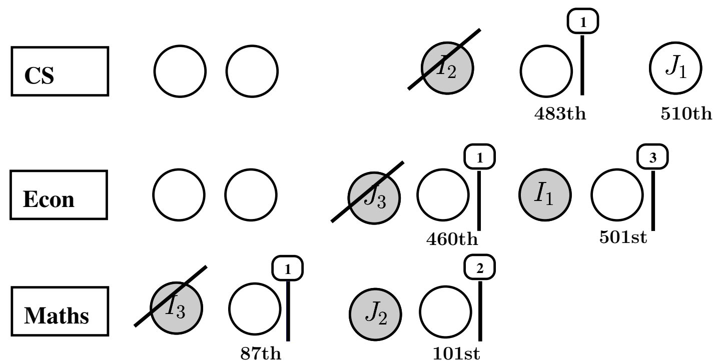

Figure 1: The score-limits in the college-proposing algorithm

In the first stage of the college-proposing algorithm the score-limits are $l_1(d_{cs}) = 111$, $l_1(d_e) = 101$ and $l_1(d_m) = 106$. At this limit a_i is admitted by the department of computer science and by the department of maths too, while a_j is admitted by the department 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 department of maths can decrease its score-limit by one, because the number of currect applications having at least 105 points is exactly 100. In this way, a_j becomes admitted by this department, and since he prefers maths to economics, he rejects his offer there. In the third stage, the department of economics, can decrease its score-limit by one. After this change a_i is admitted by the department of economics, that is his most preferred place, so he cancels all his other applications. In the final stage no department can decrease its limit, so the algorithm stops.

Applicant-proposing algorithm:
Let each applicant propose to his first choise in his list. If a department receives more applications than its quota, then let its 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 department in his list, say d_u where he achieves the actual score-limit $l_k(d_u)$, (if there remained such a department in his list). Some departments can receive new proposals, so if the number of admitted applicants exceed their quotas, 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 department increase its limit for the last time, then the rejected applicants get worse and worse departments during the algorithm. So if this department would decrease its limit by one at the final solution, then these applicants would accept the offer, and the quota would been exceeded.

Theorem 1 *Both the score-limit l_D, 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 departments in l_A than in l_D, but the number of admitted applicants can be also larger in l_A.

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

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

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

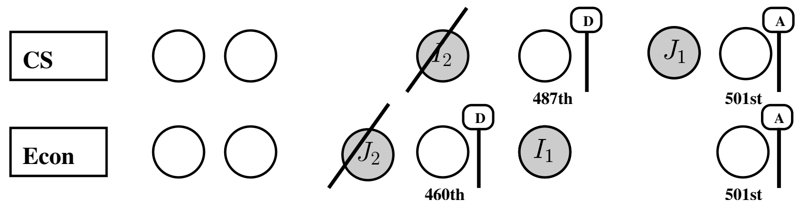

Figure 2: The final score-limits of the college-proposing and the applicants-proposing algorithm

2.2 The optimality

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

Theorem 2 *l_D 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, $l_A \leq l \leq l_D$ 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 department d_u such that $l_*(d_u) > l_D(d_u)$. 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 $l_*(d_u) > l_{k+1}(d_u)$ for some department d_u. Obviously, $l_k^{u,t_u}(d_u) = l_{k+1}(d_u)$ 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 one hand, there must be an applicant, say a_i who is admitted by d_u at $l_*^{u,1}$ but not admitted by d_u at l_k^{u,t_u}. On the other hand, the indirect assumption $l_k^{u,t_u}(d_u) = l_{k+1}(d_u) \leq l_*(d_u) - 1 = l_*^{u,1}(d_u)$ implies that a_i must be admitted by another, preferred department than d_u at l_k^{u,t_u} (since a_i has at least $l_k^{u,t_u}(d_u)$ 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 department d_u such that $l_*(d_u) < l_A(d_u)$. 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 $l_*(d_u) < l_{k+1}(d_u)$ for some department d_u. At this moment, the reason of the incrementation is that more than q_u students are applying for d_u with at least l_* score. This implies that one of these students, say a_i is not admitted by d_u at l_* (however he has at least $l_*(d_u)$ score there). So, by the stability of l_*, he must be admitted by a preferred department, say d_v at l_*. Consequently, a_i must have been rejected by d_v in a previous stage of the algorithm, that is possible only if $l_*(d_v) < l_k(d_v)$, a contradiction. □

3 Further notes

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

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

In Hungary, some studies are completely financed by the state, some are partionally financed by the students. At most of the departments there are two different quotas for both kind of studies. The applicants have to note also in their rankings which kind of study they apply for at some department.[2] These are considered in the algorithm as distict departments with distict 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 department can not be more than 10%. This rule is tracted by the actual algorithm.

Another speciality that some few pairs of departments 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 that the scoring system is not fine enough, that is why huge ties likely emerge. As a conclusion, 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 department, no student can be admitted. So, it is a good news, that recently a more fine scoring system has been proposed in the actual law that will increase the maximum score from 144 to 500.

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

In our opinion, to change the direction of the algorithm would be also reasonable. Not just because some applicants could be admitted by preferred departments, 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 Perantson [6] about this).

References

[1] Abdulkadiroglu A, Pathak PA, Roth AE (2005) The New York city high school match. American Economic Review 95(2):364-367

[2] Abdulkadiroglu A, Pathak PA, Roth AE (2005) The Boston public school match. American Economic Review 95(2):368-371

[3] Gale D, Shapley LS (1962) College admissions and stability of marriage. American Mathematical Monthly 69:9-15

[4] National Higher Education Information Centre, www.felvi.hu

[5] Roth AE (1984) The evolution of the labor market for medical interns and residents: a case study in game theory. Journal of Political Economy 6:991-1016

[6] Roth AE, Peranson E (1999) The redesign of the matching market for american physicians: some engineering aspects of economic design. American Economic Review 89:748-752

[7] Roth AE, Sotomayor M (1990) Two-sided matching: A study in game-theoretic modeling and analysis. Econometric Society Monograph Series, Cambridge University Press