Stable Matching with Proportionality Constraints

Thành Nguyen* and Rakesh Vohra†

January 5, 2017

Abstract

The problem of finding stable matches that meet distributional concerns is usually formulated by imposing various side constraints. Prior work has focused on constraints whose “right hand sides” are absolute numbers specified before the preferences or number of agents on the “proposing” side are known. In many cases it is more natural to express the relevant constraints as proportions. We treat such constraints as soft, but provide ex-post guarantees on how well the constraints are satisfied while preserving stability. Our technique requires an extension of Scarf’s lemma, which is of independent interest.

Keywords: stable matching, diversity, Scarf’s lemma

JEL classification: C78, D47

1 Introduction

The elegance and simplicity of Gale and Shapley’s deferred acceptance algorithm (DA), Gale and Shapley (1962), has made it the algorithm of choice for determining stable matchings in a variety of settings: doctors to hospitals, students to school seats. Each setting has imposed new demands on the algorithm beyond returning a stable matching that satisfies

*Krannert School of Management, Purdue University, nguye161@purdue.edu
†University of Pennsylvania, rvohra@seas.upenn.edu
capacity constraints. There is a desire to satisfy side constraints motivated by equity and
distributional considerations.

It is usual to express the relevant constraints as proportions. As an example, in 1989, the
city of White Plains, New York required each school to have the same proportions of Blacks,
Hispanics, and “others”, a term that includes Whites and Asians. The plan allowed for a
discrepancy among schools of only 5 percent. Similarly, the 2003 Cambridge, Massachusetts
Public School District’s goal for a matching was for each grade in each school to be within
a range of plus or minus 15 percentage points of the district-wide percentage of low-SES
students.

In prior work, these constraints are translated into absolute numbers. For example, a
requirement that at least 10 percent of students in a school with capacity 100 belong to
a particular SES becomes a constraint that at least 10 students in the school belong to
the relevant SES. This assumes that each school will be fully allocated. If the number of
students is less than the number of slots, this clearly cannot be true. The Chicago Public
Schools (CPS), the third largest in the US, for example, saw a drop in enrollment from
426,215 in 2000 to about 350,535 in 2013. It classifies almost 50 percent of Chicago’s public
schools as half-empty Even if the number of students exceeds the number of slots, whether
a school is fully allocated or not depends on the actual matching and the outside options of
the students.

In this paper, we consider both lower and upper bound constraints on the proportions
of students from each category. Constraints on proportions have the advantage of not
committing to an absolute number as a target. Insofar as we are aware, propotionality
constraints have not previously been incorporated into stable matching problems.

The first contribution of this paper is a definition of stability adapted to the case of

---

1See Case Studies of School Choice and Open Enrollment in Four Cities, (Cowen Institute, 2011).
2Other school districts facing steep enrollment declines are Buffalo, Philadelphia, Columbus (Ohio),
Pittsburgh, Cleveland, Detroit and Kansas City.
proportionality constraints. Adaptation is required because proportionality constraints are implicit capacity constraints. To see why, suppose the number of minority students applying to a given school is 10. Suppose also that the school must ensure that at least 20 percent of admitted students be minority. To satisfy this proportionality constraint, the school cannot admit more than $\frac{10}{0.20} = 50$ students. This could be below the actual capacity of the school. Furthermore, the “effective capacity” implied by the proportionality constraint depends endogenously on the matching. For example, suppose that among the 10 minority students applying to this school, 4 are accepted to schools they prefer more and which will accept them. The given school’s effective effective capacity will be reduced to $\frac{6}{0.20} = 30$. The effect on capacity is compounded with the introduction of upper bound proportionality constraints as well.

Satisfying both stability and proportionality constraints in general is impossible. So, we treat the proportionality constraints as soft, but provide ex-post guarantees on how well they are satisfied while preserving stability. Our algorithm relies on the technique introduced in [Nguyen and Vohra (2016)], which is based on Scarf’s lemma. Adapting it to proportionality constraints requires a significant generalization of the lemma, which is of independent interest. This generalization establishes the existence of a “fractional” stable matching that satisfies the proportionality constraints. Subsequently, this fractional matching is rounded into an integral stable matching that only violates the proportionality constraints (but not the capacity constraints) in a limited way.

Below we briefly summarize prior work emphasizing its differences with this paper. Section 2 defines lower bound proportionality constraints and the notion of stability. Section 3 introduces Scarf’s lemma and extends it to problems with lower bound proportional constraints. Section 4 describes the algorithm. Section 5 analyzes the stability of the rounded solution. Section 6 extends the result to upper bound proportionality constraints and gives an explicit algorithm. We conclude in Section 7
1.1 Related Work

There is now a large collection of papers devoted to finding stable matchings under distributional constraints. No attempt is made to provide a comprehensive survey here. However, all prior work can be be placed into one of four categories that we list below and give illustrative examples of papers in each category.

• Ceilings

Distributional concerns are modeled as ceilings on the number of agents of each type from the ‘proposing’ side that can be accepted. Ceiling constraints are generally considered easy to accommodate (see Abdulkadiroğlu and Sönmez (2003)). Regional capacity constraints are ceilings that apply to subsets on the “accepting” side rather than just individual members. As long as the subsets have a laminarity property, satisfying these constraints as well as the stability is not difficult. See Fleiner and Kamiyama (2012) and Kojima et al. (2014) for examples. Ceiling constraints, however, can disadvantage minorities. This is discussed in Hafalir et al. (2013).

• Floors

Instead of imposing ceilings, one imposes floors on the number of proposers of a particular type. Satisfying floors and stability is generally difficult to do. This is discussed in Biró et al. (2010) and Huang (2010) which also describe some solvable cases.

• Set Asides

Instead of ceilings and floors, one sets aside capacity for each subgroup and then run a separate matching process for each subgroup. This approach generally produces inefficiencies and other perverse effects in the resulting matching (see Kojima (2012) and Ellison and Pathak (2016)), which are subsequently addressed by adjusting the set asides either dynamically or ex-post (see Fragiadakis and Troyan (2016) and Aygun and Turhan (2016) for examples).
• Modifying Priorities

Instead of focusing on floors and ceilings, one modifies the choice function on the “accepting” side so as to favor various groups. If the modified choice function is specified in the right way, the DA algorithm (or some variant) will find a stable matching. However, there is no ex-post guarantee on realized distribution. An example of this approach can be found in Ehlers et al. (2014). In lieu of an ex-post guarantee, some authors focus on priorities that will produce distributions that are closest to a target distribution, see Erdil and Kumano (2012) for an example.

In the first three cases the relevant “right-hand sides” are quantities specified before agents on the “proposing” side make their participation decisions. This may over-constrain the problem because the number of “proposers” who will be matched is not known in advance but is rather endogenous. In the remaining cases, targeted groups are ‘favored’ but no guarantee is provided on the realized distribution.

2 Matching with Proportionality Constraints

To describe the stable matching problem we use the language of doctors and hospitals to distinguish between the two sides of the market. Denote by \( H \) the set of hospitals and \( D \) the set of doctors. Each doctor \( d \in D \) has a strict preference ordering \( >_d \) over \( H \cup \{\emptyset\} \), where \( \emptyset \) denotes the outside option for each doctor. If \( \emptyset >_d h \), we say that hospital \( h \) is not acceptable for \( d \). Each hospital \( h \in H \) has capacity \( k_h > 0 \) and a strict priority ordering \( >_h \) over elements of \( D \cup \{\emptyset\} \). If \( \emptyset >_h d \), we say \( d \) is not acceptable for \( h \).

A matching is an assignment of each doctor to a hospital or his/her outside option; each hospital is assigned an acceptable set of doctors that does not exceed its capacity. Given a matching \( \mu \), let \( \mu(h) \) denote the subset of doctors matched to \( h \) and \( \mu(d) \) denote the

\footnote{If one is not careful, there is also a “circularity” problem, in that stability is defined with respect to the modified choice function.}
position that $d$ obtains in the matching. Thus $\mu$ satisfies:

\begin{align*}
  i) \quad & \mu(d) >_d \emptyset \\
  ii) \quad & \text{if } d \in \mu(h) \text{ then } d >_h \emptyset \\
  iii) \quad & |\mu(h)| \leq k_h 
\end{align*}

(1)

Next, we introduce the proportionality constraints for hospitals. For each hospital $h$, let $D^h := \{ d : d >_h \emptyset, h >_d \emptyset \}$ be the set of doctors acceptable to $h$ who find $h$ acceptable. Each $D^h$ is partitioned into $T_h$ sets:

$$D^h = D^h_1 \cup D^h_2 \cup \ldots \cup D^h_{T_h}.$$ 

Different hospitals can have different partitions. A doctor $d \in D^h_t$ is said to be of type $t$ for hospital $h$. The type of a doctor may represent a characteristic like intended specialty. In the school choice context, where hospitals correspond to schools and doctors to students, a type can represent an SES category. Allowing different schools to have different partitions allows schools the flexibility to use categories depending on the proximity of the student's residence to the school.\(^4\)

The lower bound proportionality constraint at each hospital $h \in H$ is

$$\alpha^h_t \cdot |\mu(h)| \leq |\mu(h) \cap D^h_t| \quad \forall t = 1, \ldots, T_h,$$

where $0 \leq \alpha^h_t \leq 1$, $\sum_t \alpha^h_t \leq 1$. \(2\)

A matching satisfying (1) and (2) is called feasible.

Constraint (2) ensures that the proportion of doctors of each type in $D^h$ who are matched to hospital $h$ is above some threshold. These constraints don’t need to hold for each hospital-

\(^4\)Here we assume the categories are disjoints. Our model can be extended to capture overlapping categories.
type pair. This can be captured by setting $\alpha_t^h = 0$. Unlike floor constraints, the left-hand side of (2) are endogenous.

Our goal is to find a feasible stable matching. Stability in the presence of (2) is defined below. We denote an instance of the problem by $(\{d\}_{d \in D}, \{h\}_{h \in H}, \{\alpha^h\}_{h \in H}, \{k^h\}_{h \in H})$. As a feasible stable matching need not exist, we derive a feasible stable matching, $\bar{\mu}$ for a nearby instance $(\{d\}_{d \in D}, \{h\}_{h \in H}, \{\bar{\alpha}^h\}_{h \in H}, \{k^h\}_{h \in H})$ where $|\alpha_t^h - \bar{\alpha}_t^h| \leq \frac{2}{|\bar{\mu}(h)|}$ for all $h$ and $t$. The precise form can be found in Section 4.3. The inverse dependence on $\bar{\mu}(h)$ is unavoidable. Small changes in the number of doctors of type $t$ matched to $h$ can have big effects on the proportion of such doctors if $|\bar{\mu}(h)|$ is small.

Our results extend to the case with both upper and lower bounds on the proportions of each type to be matched as well. This is described in Section 6. That section also describes an explicit algorithm for determining the matching.

2.1 Stability

A feasible matching is blocked if there is a doctor-hospital pair that would prefer to be matched to each other rather than the partners assigned to them in the matching. In the presence of (2) we must modify the usual notion of blocking to rule out blocking pairs that violate (2).

Example 2.1 shows why such a modification is required.

Example 2.1. Consider a single hospital $h$ with capacity 3 and priority order

$$d_1 >_h d_3 >_h d_4 >_h d_2.$$ 

Let $D = \{d_1, d_2, d_3, d_4\}$ with $D_1^h = \{d_1, d_2\}$ and $D_2^h = \{d_3, d_4\}$. All doctors strictly prefer to be

\footnote{This is standard, see Kamada and Kojima (2014) for example.}
matched to \( h \) than remain unmatched. We impose the following side constraint.

\[
\frac{2}{3} |\mu(h)| \leq |\mu(h) \cap D_1^h|.
\]  

(3)

Consider the matching \( \mu \) that assigns \( d_1, d_2, d_3 \) to \( h \). Observe that (3) binds at this matching. In the absence of (3), \((d_4, h)\) would be a blocking pair because \( h \) prefers \( d_4 \) to \( d_2 \). However, \( h \) cannot accept \( d_4 \) without violating (3). Thus, the proportionality constraints can conflict with the priority order of hospitals.

The idea behind the modification is to consider only blocking pairs that violate neither the capacity nor the proportionality constraints. To this end we define for each feasible matching a set of protected doctors and for each hospital its effective capacity. A protected doctor can never be rejected by any hospital they are matched to in favor of another doctor.

We start with the notion of wait-listed doctors.

**Definition 2.1 (Wait-listed Doctor).** Given a feasible matching \( \mu \), a doctor \( d \) is wait-listed at \( h \) if either \( d \) is unmatched and \( d \succeq_h \emptyset \), or \( h \succeq_d \mu(d) \).

Thus, the wait-listed doctors of a hospital \( h \) are those who prefer to be matched with \( h \) over their current outcome. In other words, each of these doctors would like to form a blocking coalition with \( h \).

Fix a hospital \( h \). Suppose the set of doctors of type \( t \) at this hospital, \( D_t^h \), does not contain any wait-listed doctors. Then, \( h \) cannot increase the number of admitted doctors of type \( t \) because all such doctors are already matched to a more preferred hospital. In this case, the proportionality constraint corresponding to type \( t \), \( \alpha_t^h |\mu(h)| \leq |\mu(h) \cap D_t^h| \), implies that the total number of doctors that hospital \( h \) can accept is at most

\[
\frac{1}{\alpha_t^h}|\mu(h) \cap D_t^h|.
\]
This motivates the following definition of a hospital’s effective capacity.

**Definition 2.2 (Effective Capacity).** Consider a feasible matching $\mu$ and a hospital $h$. Let $T_0$ be the set of types $t$, such that $D^h_t$ contains no wait-listed doctor. Denote hospital $h$’s effective capacity with respect to $\mu$ by $k^\mu_h$, where

$$k^\mu_h := \min\{k_h, \min_{t \in T_0} \frac{1}{\alpha^h_t} |\mu(h) \cap D^h_t|\}.$$  

If $T_0 = \emptyset$, $k^\mu_h = k_h$.

**Remark 1.** Given $\mu$, $k^\mu_h$ is an upper bound on the number of positions that $h$ can fill by accepting more wait-listed doctor without violating any proportionality constraints. Because $\mu$ is feasible, it satisfies both the capacity and the side constraints. Thus, it is clear that $|\mu(h)| \leq k^\mu_h$.

From the definition above, if hospital $h$ is not at its effective capacity with respect to $\mu$, $|\mu(h)| < k^\mu_h$, then

- $|\mu(h)| < k_h$, and
- there is no $t \in T_0$ such that the proportionality constraint corresponding to $D^h_t$ binds, that is, $|\mu(h)| = \frac{1}{\alpha^h_t} |\mu(h) \cap D^h_t|$.

In what follows, when $\mu$ is clear from context we will omit the qualifier ‘with respect to $\mu$’ when referring to a hospitals effective capacity. Next we define the types of doctors who are protected.

**Definition 2.3 (Protected Type of Doctors).** Given a feasible matching $\mu$, the set of type $t$ doctors at hospital $h$, is protected with respect to $\mu$ if (2) binds with respect to the effective capacity, that is

$$|\mu(h) \cap D^h_t| = \alpha^h_t \cdot k^\mu_h.$$  

(4)
In what follows, when $\mu$ is clear from context we will omit the qualifier ‘with respect to $\mu$’ when referring to protected doctors.

**Example 2.2.** We illustrate the definition of effective capacity and protected doctor using Figure 1.

Consider the group $D_{2}^{h_{1}}$. Doctor $d_{7}$ is the only member of $D_{2}^{h_{1}}$ not matched to $h_{1}$ and she prefers $h_{2}$ to $h_{1}$. Thus, the group $D_{2}^{h_{1}}$ does not contain any wait-listed doctor. This means that $h_{1}$ cannot admit more doctors from $D_{2}^{h_{1}}$. Together with the proportionality constraint for group $D_{2}^{h_{1}}$, it implies that hospital $h_{1}$ cannot admit more than $2/(\alpha_{2}^{h_{1}}) = 6$ doctors. The effective capacity of $h_{1}$ becomes 6 instead of its original capacity of 8.

As the effective capacity of $h_{1}$ is 6, the proportionality constraints corresponding to $D_{1}^{h_{1}}$ and $D_{2}^{h_{1}}$ bind, thus $D_{1}^{h_{1}}$ and $D_{2}^{h_{1}}$ are protected. If a type is protected, it means that the hospital cannot decrease the number of doctors of this type who are matched to it. Why? If the hospital decreases this number, it will also need to decrease the total number of doctors to satisfy the proportionality constraints.

To motivate the definition of stability below, consider $d_{3}$, who prefers to be matched with $h_{1}$ rather than his current match, $h_{2}$. As $h_{1}$ is at its effective capacity, $h_{1}$ must reject a doctor currently matched to $h_{1}$ in order to accept $d_{3}$. To satisfy the proportionality constraint, $h_{1}$
can only reject either a doctor of the same type as \( d_3 \) (\( d_1 \) or \( d_2 \)), or an unprotected doctor (\( d_9 \) or \( d_{10} \)). The definition below requires this matching to be stable if \( h_1 \) has no incentive to replace \( d_3 \) with any of these doctors. That is, \( h_1 \) prefers each of \( d_1, d_2, d_9 \) and \( d_{10} \) to \( d_3 \).

Next, we introduce the notion of stability motivated by example 2.2.

**Definition 2.4 (Bilateral Stability).** A feasible matching \( \mu \) is (bilaterally) **stable** if it satisfies the following two conditions:

1. Each hospital is at its effective capacity, that is \( |\mu(h)| = k^\mu_h \).
2. If \( d_a \) is on the wait list of \( h \), \( d_r \in \mu(h) \) and \( d_a >_h d_r \), then \( d_a \) is protected and \( d_a \) and \( d_r \) are not of the same type.

The first condition ensures that no hospital can increase the number of doctors it accepts. The second condition ensures that if \( h \) tries to reject \( d_r \) to accept a better \( d_a \), then it will violate the side constraint. This is because \( d_r \) is protected and \( d_a \) is not of the same type as \( d_r \), thus by rejecting \( d_r \) hospital \( h \) will violate the side constraint of the group containing \( d_r \).

We show that (bilateral) stability implies coalitional stability. To do so we must specify each hospital’s preferences over subsets of \( D \). This is summarized by \( h \)’s choice function \( \text{Choice}_h(\cdot) : 2^D \rightarrow 2^D \).

**Definition 2.5.** The choice function of \( h \) on a subset of acceptable doctors \( D^* \), denoted \( \text{Choice}_h(D^*) \), is the subset of \( D^* \) with largest cardinality that satisfies the capacity constraints of \( h \) and the proportionality constraints. If there are multiple such subsets, then \( \text{Choice}_h(D^*) \) is the best one in the lexicographical order according to \( >_h \).

If \( \alpha^h_t = 0 \) for all \( h \) and \( t \), this choice function reduces to being responsive: for any set \( D^* \subset D \), hospital \( h \)’s choice from \( D^* \), consists of the (up to) \( k_h \) highest priority doctors among the feasible doctors in \( D^* \). Responsiveness is a standard assumption in the literature.
With this we have the following theorem.

**Theorem 2.1.** Let \( \mu \) be a stable matching, then for any group of doctors \( D^* \) on the wait list of \( h \), \( \text{Choice}_h(\mu(h) \cup D^*) = \mu(h) \).

**Proof.** First notice that because \( h \) is at its effective capacity, \( h \) cannot increase the number of doctors without violating the proportionality constraints. Thus,

\[
|\text{Choice}_h(\mu(h) \cup D^*)| = |\mu(h)|.
\]

Let \( D_A := \text{Choice}_h(\mu(h) \cup D^*) \setminus \mu(d) \) be the set of accepted doctors. Let \( D_R := \mu(d) \setminus \text{Choice}_h(\mu(h) \cup D^*) \) be the set of rejected doctors. Assume \( D_A \) and \( D_R \) are not empty. Let \( d_{\text{min}} \) be the lowest ranked doctor among \( D_R \) according to \( >_h \). Because \( h \) breaks ties according to the lexicographical order, all doctors in \( D_A \) must be more preferred than \( d_{\text{min}} \).

If \( d_{\text{min}} \) is unprotected, then it is a contradiction to the definition of bilateral stability because any \( d_a \in D_A \) can replace \( d_{\text{min}} \) at \( h \). If \( d_{\text{min}} \) is protected then in order to keep the side constraint of this type, \( h \) needs to accept a doctor of the same type. Thus, there should be a \( d_a \in D_A \) that is of the same type as \( d_{\text{min}} \). In this case we can replace \( d_{\text{min}} \) with a better doctor, \( d_a \), which is a contradiction to the definition of bilateral stability. Hence \( D_A \) and \( D_R \) are empty and thus, \( \text{Choice}_h(\mu(h) \cup D^*) = \mu(h) \). \( \square \)

### 3 Fractional Stable Matching

This section is the technical heart of the paper. We use Scarf’s lemma to obtain a fractional stable matching. A direct application of the lemma will not accommodate (2). We derive what we call a ‘conic representation’ of the lemma. This is both new and general enough to apply to other types of side constraints, but in this paper we confine ourselves to proportionality constraints.
We first describe Scarf’s lemma and show how it can be applied to matching.

### 3.1 Scarf’s Lemma

**Definition 3.1.** Let $A$ be an $m \times n$ nonnegative matrix with at least one positive entry in each row and column and $b \in \mathbb{R}^m$ be a positive vector. Associated with each row $i$ of $A$ is a strict ranking $>_i$ over the columns in $\{1 \leq j \leq n : A_{ij} > 0\}$. Let $P = \{x : x \geq 0, Ax \leq b\}$. We say $x \in P$ **dominates** column $j$ if there exists a row $i$ such that:

- $A_{ij} > 0$ and the constraint $i$ binds, that is, $(Ax)_i = b_i$ and
- $k >_i j$ for all columns $k \neq j$ such that $A_{ik}x_k > 0$.

In this case we say that $x$ dominates column $j$ via row $i$.

**Theorem 3.1 (Scarf (1967)).** There exists an extreme point of $P$ that dominates every column of $A$.

To understand the connection of domination to stability, it is helpful to consider the matching problem without side constraints. For each $d \in D \cup \{\emptyset\}$ and $h \in H \cup \{\emptyset\}$ let $x(d,h) = 1$ if we assign $d$ to $h$ and zero otherwise. Now, each doctor $d \in D$ can be assigned to at most one hospital:

$$\sum_{h \in H \cup \{\emptyset\}} x(d, h) \leq 1 \forall d \in D. \quad (5)$$

Each hospital $h$ can be assigned at most $k_h$ doctors:

$$\sum_{d \in D \cup \{\emptyset\}} x(d, h) \leq k_h \forall h \in H. \quad (6)$$

Each inequality (5) inherits the order that doctor $d$, i.e., $>_d$, has over $H \cup \{\emptyset\}$. Each inequality (6) inherits the priority ordering that hospital $h$, i.e., $>_h$, has over $D \cup \{\emptyset\}$. As
follows, system (5-6), along with a non-negativity restriction on the $x$ variables, satisfies the conditions of Scarf’s lemma.

Now, as is well known, every non-negative extreme point of (5-6) corresponds to a matching. By Scarf’s lemma, one of these extreme points, $x^*$, say, is dominating. To show that the matching implied by $x^*$ is stable, suppose a pair $(d^*, h^*)$ such that $x^*(d^*, h^*) = 0$. We show that $(d^*, h^*)$ cannot be a blocking pair. By domination, there must exist a binding constraint from (5-6). Either it is indexed by $d^*$ or $h^*$. Say, $d^*$. Then,

$$\sum_{h \in H \cup \{\varnothing\}} x^*(d^*, h) = 1.$$ 

As $x^*(d^*, h^*) = 0$ it follows that exactly one $h' \in H \cup \{\varnothing\}$ exists such that $x^*(d^*, h') = 1$. Further, by domination, $h' > d$, $h^*$, which means $(d^*, h^*)$ cannot be a blocking pair.

The side constraints in (2) can be written as

$$\alpha^h_t \cdot \sum_{d \in D} x(h, d) \leq \sum_{d \in D^h_t} x(h, d) \quad \forall t = 1, \ldots, T_h, \forall h \in H. \quad (7)$$

It is tempting to append inequality (7) to (5-6) and invoke Scarf’s lemma. However, it is not possible. If we rewrite (7) in the form $Ax \leq b$, the relevant inequalities have negative coefficients and the corresponding coordinates of $b$ are 0. Therefore, Scarf’s lemma does not apply. Also, the condition of stability in our setting is now endogenous and depends on the effective capacity of a hospital. It is not clear how one can apply Scarf’s lemma directly as in Nguyen and Vohra (2016).

### 3.2 Conic Representation

We need another approach to determine a dominating solution of (5-6) that satisfies (7). We exploit the fact that the constraints in (7) form a polyhedral cone. Therefore, any point
in the cone can be expressed as a non-negative linear combination of its generators. We give a high level description first.

Consider the problem of finding a dominating solution satisfying resource constraints $Ax \leq b$ and side constraints $Mx \geq 0$. The set $\{x \in \mathbb{R}_n^+ | Mx \geq 0\}$ is a polyhedral cone and can be rewritten as $\{Vz | z \geq 0\}$, where $V$ is a finite non-negative matrix. The columns of $V$ correspond to the generators of the cone $\{x \in \mathbb{R}_n^+ | Mx \geq 0\}$. The ‘trick’ is to apply Scarf’s lemma to $Q = \{z \geq 0 : AVz \leq b\}$. To do so, we need to endow each row of $AV$ with an ordering so that domination with respect to this system will correspond to stability.

Figure 2: Geometric presentation

Figure 2 gives a geometric illustration of (5, 6, 7). The polytope in Figure 2 corresponds to the matching polytope (5, 6). The inequalities (7) are represented by the cone. The conic version of Scarf’s lemma gives us a fractional dominating solution, $x^*$, say, that is inside the cone but on the boundary of the matching polytope. In particular, there is no other point in the matching polyhedron that vector dominates $x^*$. Our rounding algorithm, described in the next section, will round $x^*$ into an integral solution on the boundary of the polytope, but possibly outside the cone.

Next, we show how to determine the generators of the cone associated with (7).
Generators of a Cone

The following is standard (see Nemhauser and Wolsey (1988)).

**Lemma 1.** For any matrix $\mathcal{M}$, if the set $\{x \in \mathbb{R}^n | \mathcal{M}x \geq 0\}$ contains a non-zero vector, there exists a finite set of non-negative vectors $\mathcal{V}$ such that this set can be expressed as

$$\{ \sum_{v_i \in \mathcal{V}} z_i v_i | z_i \geq 0 \}.$$

The set of vectors, $\mathcal{V}$, are called the **generators** of $\{x \in \mathbb{R}^n | \mathcal{M}x \geq 0\}$.

![Cone and its generators](image)

Figure 3: Cone and its generators

The proportionality constraints are of the form $\mathcal{M}x \geq 0$. To determine the generators of (7), fix a hospital $h \in H$ and focus on

$$\alpha_t^h \cdot \sum_{d \in D} x(h, d) \leq \sum_{d \in D_t^h} x(h, d) \quad t = 1, \ldots, T_h. \quad (8)$$

It is straightforward to see that the generators can be described in this way:

1. Select one doctor from each $D_t^h$ and call it $d_t$.

2. Select an extreme point of the system

$$\sum_{t=1}^{T_h} \alpha_t^h = 1, \quad \alpha_t^h \leq v(d_t, h) \quad \forall t = 1, \ldots, T_h.$$
An extreme point can be determined using the following two-step procedure.

(a) Choose an index $t^* \in \{1, \ldots, T_h\}$ and set $v(d_{t^*}, h) = 1 - \sum_{t \neq t^*} \alpha^h_t \geq \alpha^h_{t^*}$.

(b) For all $t \neq t^*$, set $v(d_t, h) = \alpha^h_t$.

As there are $T_h$ types of doctors, and each type contains $|D^h|$ doctors, the number of generators associated with hospital $h$ can be as large as $T_h \cdot \prod_t |D^h_t|$.

Let $\mathcal{V}_h$ be the set of generators associated with hospital $h$. Each $v \in \mathcal{V}_h$ has $T_h$ non-zero coordinates and can be interpreted as a probability vector. Thus, from the point of view of each $h \in H$, each $v \in \mathcal{V}_h$ can be seen as a lottery over doctors in $D$.

We illustrate with an example.

**Example 3.1.** Suppose $H = \{h_1, h_2\}$ and 2 doctors $d_1 \in D^{h_1}_1$ and $d_2 \in D^{h_1}_2$. Hospital $h_2$ considers all doctors to be the same type, i.e., $D^{h_2}_1 = \{d_1, d_2\}$.

The only proportionality constraints are imposed on $h_1$: the number of type 1 doctors should be at least $1/3$ of the total number of doctors assigned to $h_1$. This constraint can be written as

$$\frac{1}{3} [x(d_1, h_1) + x(d_2, h_1)] \leq x(d_1, h_1)$$

The set of generators for this constraint are

$$\mathcal{V}_{h_1} = \{(1/3, 2/3); (1/3, 2/3)\} := \{v_1, v_2\}.$$ 

We can interpret $v_1 = (1/3, 2/3)$ to mean assigning $d_1$ and $d_2$ to $h_1$ with probability $1/3$ and $2/3$ respectively. All solutions satisfying the proportionality constraints at $h_1$ can be expressed as a linear combination of $v_1$ and $v_2$.

There are no proportionality constraints imposed on $h_2$. This is the same as setting $T_h$.

Footnote: The number of types, $T_h$, will typically be a small constant. Hence, the number of generators is polynomial in the number of doctors.
\( \alpha_{h_2} = 0. \) The set of generators for \( h_2 \) are

\[ V_{h_2} = \{(1,0), (0,1)\} = \{v_3, v_4\}. \]

We interpret \( v_3 = (1,0) \) to mean assigning \( d_1 \) to \( h_2 \) with probability 1, and \( d_2 \) to \( h_2 \) with probability zero.

### 3.3 Conic version of Scarf’s lemma

Associated with each hospital \( h \in H \) is a set \( V_h \) of generators. Let \( V \) be the matrix whose columns correspond to the generators in \( \bigcup_{h \in H} V_h \). Let \( A \) be the constraint matrix associated with (5-6). Each row of the matrix \( A \cdot V \) corresponds to an element of either \( D \) or \( H \). The columns of \( A \cdot V \) correspond to the set of generators. For each \( h \in H \), a column in \( A \cdot V \) that corresponds to \( v \in V_h \), will have a ‘1’ in the \( h^{th} \) row and \( v(d,h) \) in the \( d^{th} \) row. All other entries in that column will be zero. Let \( z \in \mathbb{R}_{+}^{\bigcup_{h \in H} V_h} \) be a non-negative weight vector on the set of generators. The constraints \( A \cdot V \cdot z \leq b \) can be interpreted as follows:

- For each hospital \( h \), the total weight of generators in \( V_h \) is at most \( k_h \).

- For each doctor \( d \), the weight of generators that assigns \( d \) to a hospital is at most 1.

**Example 3.2.** Consider example 3.1. Suppose \( k_h = 2 \). The polyhedron \( Q \) is displayed below.
To invoke Scarf’s lemma we need each row of \( \mathcal{A} \mathcal{V} \) to have a strict ordering over the columns, i.e., generators, in its support. The support of each generator corresponds to one hospital and a coalition of doctors, at most one of each type.

1. For each \( h \in H \) we order the generators in \( \mathcal{V}_h \) lexicographically. Given \( v, v' \in \mathcal{V}_h \), among the doctors who are assigned by \( v, v' \) with positive probability to \( h \), let \( d_1 \) and \( d'_1 \) be the lowest ranked doctors (according to \( >_h \)). If \( d_1 >_h d'_1 \), then \( h \) ranks \( v \) over \( v' \). We will write this as \( v >_h v' \). If \( d_1 = d'_1 \), we compare \( v(d_1, h) \) and \( v'(d_1, h) \). If \( v(d_1, h) > v'(d'_1, h) \), then \( h \) ranks \( v \) over \( v' \), i.e., \( v >_h v' \). If it is the reverse, then, \( v' >_h v \). If \( v(d_1, h) = v'(d'_1, h) \), move to the next lowest ranked doctor in each generator and so on. Because \( v \neq v' \), this procedure must terminate in an unambiguous ordering.

2. For each \( d \in D \) and any \( v, v' \in \bigcup_{h \in H} \mathcal{V}_h \), we rank \( v \) above \( v' \) if \( v \) assigns \( d \) to a higher ranked hospital (according to \( >_d \)) than \( v' \) does. We write this as \( v >_d v' \). If \( v, v' \in \mathcal{V}_h \) for some \( h \in H \), then, \( v >_d v' \) if \( v(d, h) > v'(d, h) \) and the reverse otherwise. If \( v(d, h) = v'(d, h) \) we order \( v \) and \( v' \) in the same way that \( h \) would.

**Example 3.3.** Continuing example (3.1), let \( h_1 >_d h_2 \) for all \( d \in D \) and \( d_1 >_h d_2 \) for all \( h \in H \). The order of each element of \( D \cup H \) over the set of generators is displayed below.

\[
\begin{array}{cccc}
  & v_1 & v_2 & v_3 & v_4 \\
  d_1 & 1/3 & 2/3 & 1 & 0 \\
  d_2 & 2/3 & 1/3 & 0 & 1 \\
  h_1 & 1 & 1 & 0 & 0 \\
  h_2 & 0 & 0 & 1 & 1
\end{array}
\]

\[
\begin{array}{cccc}
  & 1 & 2 & 3 & 4 \\
  v_1 & v_2 & v_3 & v_4 \end{array}
\]

Consider the order for \( h_1 \) on \( v_1, v_2 \). Because they both assign \( d_1 \) and \( d_2 \) to \( h_1 \), we need to compare the probability of assigning \( d_2 \), which is the worst doctor for \( h_1 \). Because \( v_1 \) assigns \( d_2 \) with higher probability, \( v_2 >_{h_1} v_1 \).
Consider the order for $d_1$ on $v_1, v_2$ and $v_3$. Because $v_1, v_2$ assigns $d_1$ to $h_1$, and $v_3$ assigns $d_1$ to $h_2$, thus $d_1$ prefers both $v_1$ and $v_2$ to $v_3$. Between $v_1$ and $v_2$, the one that assigns with a lower probability is better, and thus $v_1 >_{d_1} v_2$.

**Remark 2.** By Theorem 3.1, there exists a dominating solution of $Q$, call it $z^*$. Let

$$V^* = \{v \in V : z_v^* > 0\} \quad (9)$$

be the set of generators with positive $z^*$ weight. Denote by $V_h^*$ the subset of generators in $V^*$ that assign doctors to $h$ and similarly denote by $V_d^*$ the subset of generators in $V^*$ that assign $d$ to a doctor. Because $z^*$ is a dominating solution, for every generator $v$ that assigns a group of doctors $d_1, \ldots, d_T$ to $h$, one of the following must be true:

- **The constraint** $Q$ **corresponding to** $h$ **binds. That is, $h$ is fully allocated and $h$ ranks all the generators in** $V_h^*$ **over** $v$.

- **There is a** $d_i \in \{d_1, \ldots, d_T\}$ **such that the constraint in** $Q$ **corresponding to** $d_i$ **binds and** $d_i$ **ranks all the generators in** $V_{d_i}^*$ **over** $v$.

In example 3.3, the following is a dominating solution (that we will interpret later):

$$z_{v_1}^* = 1; z_{v_2}^* = 1; z_{v_3}^* = 0; z_{v_4}^* = 0.$$

We can recover the corresponding matching $x^*$ by setting $x^* := Vz^*$:

$$x^*(d_1, h_1) = 1; x^*(d_2, h_1) = 1; x^*(d_1, h_2) = 0; x^*(d_2, h_2) = 0.$$

Notice, in this case, the matching is integral. If $x^*$ is integral, it will correspond to a stable matching. However, in general, $x^*$ is fractional. In the next section, we provide an algorithm to convert $x^*$ to an integral solution.
4 Algorithm

In the previous section, we showed how to obtain a fractional matching from a dominating solution. In particular, we let $z^\ast$ be a dominating solution and set $x^\ast = \forall z^\ast$. We show below how to round $x^\ast$ into an integer $\bar{x}$ that satisfies (5-6) and almost satisfies (7).

4.1 Rounding

**Lemma 2.** Given $x^\ast$, there exists integral $\bar{x}$ such that

- $x^\ast(d,h) = 0 \Rightarrow \bar{x}(d,h) = 0$.
- $\left\lceil \sum_{h\in H} x^\ast(d,h) \right\rceil \leq \sum_{h\in H} \bar{x}(d,h) \leq 1 \forall d \in D$
- $\left\lceil \sum_{d\in D} x^\ast(d,h) \right\rceil \leq \sum_{d\in D} \bar{x}(d,h) \leq k_h \forall h \in H$
- $\left\lceil \sum_{d\in D^h_t} x^\ast(d,h) \right\rceil \leq \sum_{d\in D^h_t} \bar{x}(d,h) \leq \left\lfloor \sum_{d\in D^h_t} x^\ast(d,h) \right\rfloor \forall t = 1, \ldots, T_h, \forall h \in H$

Furthermore, $\bar{x}$ can be found by a polynomial time algorithm.

Lemma 2 shows that we can always round a matching $x^\ast$ to $\bar{x}$ such that capacities at the hospitals are not violated and the number of doctors for each type is rounded either up or down to the closest integral number. This is essentially the best integer solution that can be hoped for.

**Proof.** We show that the problem of finding $\bar{x}$ can be formulated as the problem of finding a feasible flow in a network, all of whose arc capacities are integral. Integrability of $\bar{x}$ follows immediately.

Introduce a source node $\sigma$, a sink node $\tau$, one node for each $d \in D$, $h \in H$ and $D^h_t$. For each $d \in D$ there is an arc directed from $\sigma$ to $d$ with upper bound arc capacity of 1. For each $d$ there is an arc directed to $D^h_t$ if $d \in D^h_t$ and $x^\ast(d,h) > 0$ with upper bound arc capacity of

---

This is similar to Theorem 3 in [Budish et al. 2013](#).
1 and lower bound of \( \left[ \sum_{h \in H} x^*(d, h) \right] \). For each \( D^h_t \) there is an arc directed to \( h \) with upper bound arc capacity of \( \left[ \sum_{d \in D^h_t} x^*(d, h) \right] \) and lower bound arc capacity of \( \left[ \sum_{d \in D^h_t} x^*(d, h) \right] \). For each \( h \in H \) there is an arc directed from \( h \) to \( \tau \) with upper bound arc capacity of \( k_h \) and lower bound arc capacity of \( \left[ \sum_{d \in D^h_t} x^*(d, h) \right] \).

Note that \( x^* \) is a feasible flow in this network, so we know that a feasible integer flow exists.

\[ \square \]

### 4.2 Modifying \( \alpha \)

Denote by \( \bar{\mu} \) the matching associated with \( \bar{x} \). Due to rounding, the proportionality constraints might be violated. We will need to change \( \alpha \) to make \( \bar{\mu} \) feasible. In particular, consider a group \( D^h_t \).

- If in the fractional solution \( \sum_{d \in D^h_t} x^*(d, h) = \alpha_t^h \sum_{d \in D^h_t} x^*(d, h) \), then let

\[
\bar{\alpha}_t^h = \frac{\sum_{d \in D^h_t} \bar{x}(d, h)}{\sum_{d \in D^h_t} \bar{x}(d, h)}.
\]

- If \( \sum_{d \in D^h_t} x^*(d, h) > \alpha_t^h \sum_{d \in D^h_t} x^*(d, h) \) but \( \sum_{d \in D^h_t} \bar{x}(d, h) < \alpha_t^h \sum_{d \in D^h_t} \bar{x}(d, h) \), then also let \( \bar{\alpha}_t^h \) be as above. Otherwise \( \bar{\alpha}_t^h = \alpha_t^h \).

With this, our main result is the following.

**Theorem 4.1.** Given the fractional stable matching \( x^* \), let \( \bar{\mu} \) be the matching obtained from \( x^* \) via Lemma 2. Then \( \bar{\mu} \) is feasible and stable for the instance \( (\{d\}_{d \in D^t}, \{h\}_{h \in H}, \{\bar{\alpha}_h^h\}_{h \in H}) \).

The proof is given in Section 5. In the following, we show proximity bounds for the new matching.

22
4.3 Proximity Bounds

We can use Lemma 2 to quantify the closeness of $\bar{\mu}$ to $x^*$. By Lemma 2,

$$|\bar{\mu}(h)| \in \left\{ \left[ \sum_{d \in D} x^*(d, h) \right], \left[ \sum_{d \in D} x^*(d, h) \right] \right\}.$$ 

Thus, we never violate the capacity constraint of $h$. Furthermore, the rounding bound also implies that

$$\|\bar{\mu}(h)| - \sum_{d \in D} x^*(d, h)\| \leq 1 \forall h \in H.$$ 

By Lemma 2, $|\bar{\mu}(h) \cap D^h_t| \in \left\{ \left[ \sum_{d \in D^h_t} x^*(d, h) \right], \left[ \sum_{d \in D^h_t} x^*(d, h) \right] \right\}$.

Hence $\|\bar{\mu}(h) \cap D^h_t| - \sum_{d \in D^h_t} x^*(d, h)\| \leq 1 \forall D^h_t$.

In fact, when the proportionality constraint associated with $D^h_t$ binds, we can say something more:

$$\|\bar{\mu}(h) \cap D^h_t| - \alpha^h_t \sum_{d \in D} x^*(d, h)\| \leq 1$$

because

$$|\bar{\mu}(h) \cap D^h_t| \in \left\{ \left[ \alpha^h_t \sum_{d \in D} x^*(d, h) \right], \left[ \alpha^h_t \sum_{d \in D} x^*(d, h) \right] \right\}.$$ 

Notice that the total number of doctors assigned to any $h$ differs by at most 1 from the original fractional quantity. The same is true for the number of doctors of a particular type. The proportions, however, can behave quite differently. Using these proximity bounds it is straightforward to argue that

$$|\alpha^h_t - \bar{\alpha}^h_t| = |\alpha^h_t - \frac{|\bar{\mu}(h) \cap D^h_t|}{|\bar{\mu}(h)|}| \leq \frac{2}{1 + \sum_{d \in D} x^*(d, h)} \forall D^h_t.$$ 

Of course, if $h$ were fully allocated under $x^*$, this bound would reduce to $\frac{2}{\sqrt{k_h}}$. If the
proportionality constraint for $D_h^t$ binds, this bound improves to $\frac{1 + \alpha_h^t}{1 + \sum_{d \in D} x^*(d, h)}$. In all cases, the closeness of the realized proportions, $\frac{|\hat{\mu}(h)^c \cap D_h^t|}{|\hat{\mu}(h)|}$ to $\alpha_h^t$, depend upon the size of $|\hat{\mu}(h)|$, the number of doctors matched to $h$. If $|\hat{\mu}(h)|$ is small, even small changes in the number of doctors assigned to $h$ can have a large effect on the relevant proportions. If large, then a change in one doctor more or less will have a negligible effect on the relevant proportions.

5 Stability of $\bar{\mu}$

Recall that our algorithm starts from a dominating solution, $z^*$, which is a weight vector of the generators. The algorithm converts $z^*$ to $x^* \coloneqq Vz^*$, which is a fractional matching between doctors and hospitals. The solution $x^*$ is then rounded to an integral solution $\bar{x}$. We denote $\bar{\mu}$ to be the corresponding matching.

We now show that $\bar{\mu}$ is stable with respect to $\bar{\alpha}$. We prove this by contradiction. Assume that it is not stable. We will construct a generator that is not dominated by $z^*$. Below we use the notation of $V^*, V_h^*$ defined in Remark 2.

A group $D_h^t$ reaches its lower bound in $x^*$ if the corresponding proportionality constraint binds, that is,

$$\sum_{d \in D_h^t} x^*(d, h) = \alpha_h^t \sum_{d \in D} x^*(d, h).$$

The following observations will be helpful in the proof.

**Observation 1.** If a group $D_h^t$ reaches its lower bound in $x^*$, then it also reaches its lower bound with respect to $\bar{x}$ and the modified $\bar{\alpha}$ defined in equation (10).

**Proof.** Observation comes directly from the definition of $\bar{\alpha}$.

**Observation 2.** If a group $D_h^t$ reaches its lower bound in $x^*$, then for every $v \in V_h^*$ and $d \in D_h^t$ such that $v(h, d) > 0$, $v(h, d) = \alpha_h^t$.
Proof. Observation 2 comes from the way we construct the generators. In particular, for all generator \( v \in \mathcal{V}_h^* \), if \( v(h,d) > 0 \), then \( v(h,d) \geq \alpha_h^h \). Thus, if \( D^h \) reaches its lower bound in \( x^* \), then there cannot be a \( v \in \mathcal{V}_h^* \) such that \( v(h,d) \) is strictly greater than \( \alpha_h^h \).

**Observation 3.** If \( d \) is a wait-listed doctor at \( h \) under \( \bar{\mu} \), then, any generator in \( \mathcal{V}_h \) that assigns \( d \) to \( h \) cannot be dominated via \( d \).

Proof. If \( d \) is assigned by \( \bar{\mu} \) to a hospital \( h' \) such that \( h \succ_d h' \), this means that there is a generator \( v \in \mathcal{V}_{h'}^* \). This implies that any generator assigning \( d \) to \( h \) will be ranked above \( v \). If \( d \) is unassigned under \( \bar{\mu} \), it means that the constraint of \( d \) does not bind, that is, \( \sum_{h \in H} x^*(d,h) < 1 \). Hence the constraint corresponding to \( d \) in \( Q \) also does not bind. Thus, no generator can be dominated at this constraint.

**Proof of Theorem 4.1**

Suppose \( \bar{\mu} \) is not stable. This means that either \( h \) does not reach its effective capacity or there exists \( d_r \) currently matched with \( h \) that can be exchanged by a better doctor \( d_a \) who is on the wait list of \( h \). There are 2 cases in which \( d_r \) can be replaced by \( d_a \). They are either of the same type, or of a different type but \( d_r \) is not protected. We will construct a generator that is not dominated by \( z^* \), which leads to a contradiction.

**Case 1:** \( h \) is not at its effective capacity under \( \bar{\mu} \).

From Remark 1, this means that \( \sum_{d \in D} \bar{x}(d,h) < k_h \). However, because of the rounding procedure, this implies that \( \sum_{d \in D} x^*(d,h) < k_h \). Thus, no generator can be dominated at \( h \). It remains to create a generator \( v \in \mathcal{V}_h \) that is not dominated via any doctor. This will lead to a contradiction because \( z^* \) is a dominating solution.

Let \( \{i_1, \ldots, i_k\} \) be the set of types that contain wait-listed doctors. Choose one wait-listed doctor from each type to be part of the generator \( v \). Let they be \( d_{i_1}, \ldots, d_{i_k} \). Also let \( v(d_{i_1}, h) = 1 - \sum_{t \in i_1} \alpha_t^h \) and \( v(d_{i_2}, h) = \alpha_{i_2}^h, \ldots, v(d_{i_k}, h) = \alpha_{i_k}^h \). By Observation 3, the generator
that we are constructing cannot be dominated at \( d_i, \ldots, d_k \).

Denote the remaining set of types by \( \{i_{k+1}, \ldots, i_{T_h}\} \). These types do not contain any wait-listed doctor, because \( h \) is not at its effective capacity, according to Remark 1, their side constraints does not bind in the matching \( \bar{\mu} \). According to Observation 1, this means that the side constraints of these types do not bind in the fractional solution \( x^* \). Because of Observation 2, this means that for each type \( t \in \{i_{k+1}, \ldots, i_{T_h}\} \), there exists \( v_t \in \mathcal{V}_h^* \) that assigns a doctor \( d_t \in D_h^t \) to \( h \) with probability higher than \( \alpha_t^h \), i.e., \( v_t(d_t, h) > \alpha_t^h \). Let \( d_t \) be part of the generator \( v \), and let \( v(d_t, h) = \alpha_t^h \). By the way the preference order is defined for doctors, \( d_t \) prefers this new generator to \( v_t \). Thus, \( v \) cannot be dominated at any doctor. This contradicts the fact that \( z^* \) is a dominating solution.

**Case 2a:** \( d_r \) and \( d_a \) are of the same type.

Let \( v_r \in \mathcal{V}_h^* \) be a generator that assigns \( d_r \) to \( h \). Let \( v_a \) be the generator obtained from \( v_r \) by assigning \( d_a \) to \( h \) instead of assigning \( d_r \) to \( h \) with the same probability. Clearly, because \( d_a > d_r \), \( v_a \) is ranked above \( v_r \) by \( h \) and all doctors of different types than \( d_a \). Because of Observation 3, \( v_a \) is not dominated at \( d_a \). Thus \( z^* \) does not dominate \( v_a \), a contradiction.

**Case 2b:** \( d_a \in D_h^a \) and \( d_r \in D_h^r \) are not of the same type and \( d_r \) is not protected under \( \bar{\mu} \).

Because \( d_r \) is not protected under \( \bar{\mu} \), in the fractional solution the side constraint of \( D_h^r \) does not bind (does not reach its lower bound). Among all the doctors \( d \) whose side constraints does not bind and \( x^*(d, h) > 0 \), let \( d_{\text{min}} \) be the least preferred doctor according to \( h \). Assume \( d_{\text{min}} \in D_{\text{min}}^h \). Clearly, \( d_a \geq h \geq d_{\text{min}} \). If \( d_a \) and \( d_{\text{min}} \) are of the same type, we return to Case 2a. Assume therefore that they are of different types.

Let \( v_{\text{min}} \in \mathcal{V}_h^* \) be a generator that assigns \( d_{\text{min}} \) to \( h \). Because \( x^*(d_{\text{min}}, h) > 0 \), such a \( v_{\text{min}} \) exists. There might be several such generators; if so, choose one with the highest probability of assigning \( d_{\text{min}} \) to \( h \), i.e, the highest \( v_{\text{min}}(d_{\text{min}}, h) \). We will construct \( v_a \in \mathcal{V}_h \) by modifying \( v_{\text{min}} \) such that \( v_a \) is dominated by neither \( h \) nor any doctor.
a) $v_a$ assigns $d_a$ to $h$ with probability $1 - \sum_{t \neq a} \alpha_t^h$. By Observation 3, $v_a$ cannot be dominated via $d_a$.

b) Because the side constraint $D_{min}^h$ does not bind in the fractional solution, there exists $v'_{min} \in \mathcal{V}_h^*$ that assigns a doctor $d'_{min} \in D_{min}^h$ to $h$ with probability higher than $\alpha_{min}^h$ ($d'_{min} \neq d_{min}$ and $d_{min}$ can coincide). Let $v_a$ assign $d'_{min}$ to $h$ with probability $\alpha_{min}^h$. Thus, $v_a$ will not be dominated at $d'_{min}$.

c) $v_a$ assigns the same doctor in the remaining groups as in $v_{min}$, with probability $\alpha_t^h$ for group $D_t^h$. With this choice the sum of the components of $v$ added up over the doctors will be 1.

The set of doctors assigned by $v_a$ and $v_{min}$ are different only in $D_a^h$ and $D_{min}^h$. To compare $v_a$ and $v_{min}$, we only need to compare these doctors. First, notice that $d_a > d_{min}$. Now, if $d'_{min} \neq d_{min}$, then $d'_{min} >_h d_{min}$ because of the choice of $d_{min}$. Thus, that $h$ ranks the generators according to the lexicographical order; therefore, $v_a$ is better than $v_{min}$ because it replaces $d_{min}$ with a better doctor. For the case $d'_{min} = d_{min}$, notice that $v_a$ assigns $d_{min}$ with probability $\alpha_{min}^h$, which is less than the probability of $v_{min}$. Therefore $h$ also prefers $v_a$ to $v_{min}$. Hence $v_a$ cannot be dominated via $h$.

Furthermore, the doctors breaks ties among generators according to $h$’s lexicographical order $v_a$ cannot be dominated via any doctor in the remaining groups.

Hence, we conclude that $v_a >_h v_{min}$ and cannot be dominated by $z^*$, which is a contradiction.
Lower Bounds and Upper Bound

In some cases, the proportion of individuals of a particular type matched to a school or hospital is constrained to fall within some interval. To accommodate this, we extend the earlier analysis to include both lower and upper bound proportionality constraints. Using the notation from prior sections, we consider the following constraints.

\[ \alpha^h_t \cdot |\mu(h)| \leq |\mu(h) \cap D^h_t| \leq \beta^h_t \cdot |\mu(h)| \quad \forall t = 1, \ldots, T_h, \]

where \( 0 \leq \alpha^h_t \leq \beta^h_t \leq 1, \sum_t \alpha^h_t \leq 1 \leq \sum_t \beta^h_t. \)

(11)

Call a matching that satisfies (11) feasible. If we choose \( \beta^h_t = 1 \) for all \( h \) and \( t \), we recover (2). We maintain the same notation as before, and departures are noted as they arise.

We develop an algorithm to find a stable matching that slightly violates the proportionality constraints (11). The main result is in Theorem 6.2. We first define the notion of stability in Section 6.1. Section 6.2 describes the set of cone generators used in the algorithm presented in Section 6.3. The main proof to show that the matching we obtain by this algorithm is stable is given in Section 6.4.

6.1 Stability

The presence of upper and lower bounds on the relevant proportions requires a modification of the definition of a hospital’s effective capacity with respect to \( \mu \). To see why, fix a hospital \( h \) and a subset \( S \subset \{1, \ldots, T_h\} \) of the types at that hospital. The upper bounds for all types in \( \{1, \ldots, T_h\} \setminus S \) induce a lower bound on the percentage of doctors whose type is in \( S \). Specifically, the number of doctors with types in \( S \) needs to be at least a \( 1 - \sum_{t \in S} \beta^h_t \) fraction of all the doctors assigned to \( h \). That is,

\[ |\mu(h) \cap (\bigcup_{s \in S} D^h_s)| \geq (1 - \sum_{t \in S} \beta^h_t) \cdot |\mu(h)|. \]

(12)
Given a matching \( \mu \), if among the doctors in \( \bigcup_{s \in S} D_s^h \), there are no wait-listed doctors, then, \( h \) cannot hope to increase the number of admitted doctors with types in \( S \). Therefore, from (12) the effective capacity of \( h \) will be at most

\[
\frac{|\mu(h) \cap (\bigcup_{s \in S} D_s^h)|}{1 - \sum_{t \in S} \beta_t^h}.
\]

This motivates the following extension of definition 2.2.

**Definition 6.1 (Effective Capacity).** Consider a feasible matching \( \mu \) and a hospital \( h \). Let \( T_0 \) be the set of types \( t \), such that \( D_h^t \) contains no wait-listed doctor. Let

\[
\text{bound}_1 := \min_{t \in T_0} \frac{1}{\alpha_t^h} |\mu(h) \cap D_h^t|, \quad \text{bound}_2 := \min_{S \subseteq T_0} \frac{1}{1 - \sum_{t \in S} \beta_t^h} |\mu(h) \cap \bigcup_{s \in S} D_s^h|.
\]

The effective capacity of hospital \( h \) with respect to \( \mu \), denoted by \( k_{\mu}^h \), is \( \min \{ k_h, \text{bound}_1, \text{bound}_2 \} \). If \( T_0 = \emptyset \), \( \text{bound}_1 \) and \( \text{bound}_2 \) are set to infinity.

As before, when \( \mu \) is clear from context we omit its mention when referring to a hospital’s effective capacity.

**Remark 3.** \( k_{\mu}^h \) is an upper bound on the number of slots that \( h \) can fill by accepting more wait-listed doctor without violating the side constraints. Because \( \mu \) is feasible, it satisfies the capacity and the side constraints. Thus, it is clear that \( |\mu(h)| \leq k_{\mu}^h \).

From definition 6.1, if \( h \) is not at its effective capacity, \( |\mu(h)| < k_{\mu}^h \), the following three statements hold.

1. \( |\mu(h)| < k_h \).

2. There is no \( t \in T_0 \) such that the lower bound proportionality constraint corresponding to \( D_t^h \) binds, that is, \( |\mu(h)| = \frac{1}{\alpha_t^h} |\mu(h) \cap D_t^h| \).
3. There is no \( S \subset T_0 \), such that all upper bound proportionality constraints for types \( t \notin S \) bind. That is, \( |\mu(h)| = \frac{1}{\beta_t} |\mu(h) \cap D_t^h| \), for all \( t \notin S \).

We must also extend Definition 2.3:

**Definition 6.2 (Protected and Surplus Doctors).** Given a feasible matching \( \mu \), a doctor of type \( t \) is protected at \( h \in H \) with respect to \( \mu \) if the lower bound proportionality constraint associated with \( D_t^h \) binds with respect to its effective capacity.

A doctor of type \( t \) is surplus at \( h \) with respect to \( \mu \) if the upper bound proportionality constraint associated with type \( D_t^h \) binds with respect to the hospital’s effective capacity.

As before when \( \mu \) is clear from context we omit the qualifier ‘with respect to \( \mu \)’.

As in Definition 2.3, if a type is protected (surplus) at \( h \) it means that \( h \) cannot be matched to fewer (more) doctors of this type without reducing the number of positions at \( h \).

Now, consider a hospital \( h \) and two doctors \( d_a >_h d_r \). Assume that \( d_r \) is currently matched with \( h \), while \( d_a \) is wait-listed at \( h \). This means that \( h \) has an incentive to exchange \( d_r \) for \( d_a \).

The definition of bilateral stability below allows for such a blocking coalition, but requires that if \( h \) does so, it will violate the side constraints. This means that \( h \) must either decrease the number of protected doctors or increase the number of surplus doctor. Specifically, we have the following definition:

**Definition 6.3 (Bilateral Stability).** A feasible matching \( \mu \) is called bilaterally stable if the following two conditions hold:

1. Every hospital is at its effective capacity, that is \( |\mu(h)| = k^D_h \forall h \in H \).

2. For any \( d_a, d_r \in D \) for which \( d_a \) is wait-listed for \( h \), \( \mu(d') = h \), and \( d_a >_h d_r \), then \( d_a, d_r \) are of different types and either \( d_a \) is surplus or \( d_r \) is protected.
The first condition does not permit a hospital to increase its intake. The second says permitting \( h \) to replace \( d_r \) with \( d_a \) will violate (11).

As in Section [2.1] we show that (bilateral) stability implies coalitional stability. A hospital’s choice function in the presence of side constraints is defined next.

**Definition 6.4.** The choice function of \( h \) on a subset of acceptable doctors \( D^* \), denoted \( \text{Choice}_h(D^*) \), is the subset of \( D^* \) with largest cardinality that satisfies the capacity constraints of \( h \) and the proportionality constraints. If there are multiple such subsets, then \( \text{Choice}_h(D^*) \) is the best one in the lexicographical order according to \( \succ_h \).

With this we have the following theorem.

**Theorem 6.1.** Let \( \mu \) be a stable matching, then for any group of doctors \( D^* \) on the wait list of \( h \), \( \text{Choice}_h(\mu(h) \cup D^*) = \mu(h) \).

The proof of Theorem 6.1 is analogous to that of Theorem 2.1 and is omitted.

### 6.2 Cone Generators

We take the same steps as before. The first is to determine the generators of (11). Fix a hospital \( h \in H \) and focus on:

\[
\alpha_t^h \cdot \sum_{d \in D} x(h,d) \leq \sum_{d \in D_t^h} x(h,d) \leq \beta_t^h \cdot \sum_{d \in D} x(h,d) \text{ for } t = 1, \ldots, T_h.
\]  

The generators are the extreme points of the system

\[
\sum_{t=1}^{T_h} v(d_t,h) = 1, \quad \alpha_t^h \leq v(d_t,h) \leq \beta_t^h \forall t = 1, \ldots, T_h.
\]  

It is easy to see that an extreme point of (14) can be determined using the following algorithm.
1. Select one doctor from each $D^h_t$ and call it $d_t$.

2. Choose an ordering of the selected doctors, and call it $\sigma$.

3. Set $v(d_t, h) = \alpha_t^h$ for $i = 1, \ldots, T_h$.

4. In the order selected, increase the value of each $v(d_t, h)$ as much as possible (up to $\beta_t^h$) until the remaining mass of $1 - \sum_{t=1}^{T_h} \alpha_t^h$ is exhausted.

With this algorithm, consider a generator for hospital $h$. If we order the $T_h$ non-zero components in the order that they are selected by the algorithm, they will be of the form

$$\beta_{i_1}^h, \ldots, \beta_{i_k}^h, \gamma, \alpha_{i_{k+2}}^h, \ldots, \alpha_{T_h}^h,$$

where $\gamma = 1 - \beta_{i_1}^h - \ldots - \beta_{i_k}^h - \alpha_{i_{k+2}}^h - \ldots - \alpha_{T_h}^h$.

Denote the resulting extreme point by $\{v^\sigma(d, h)\}_{d \in D, h \in H}$. Keep in mind that it is possible for two distinct orderings to give rise to the same extreme point. For subsequent arguments it is useful to distinguish between the two and hence the need to record the order.

**Definition 6.5.** A generator $v^\sigma \in V_h$ **contains** doctor $d \in D$ if $v^\sigma(d, h) > 0$. The **order** of a doctor $d$ contained in generator $v^\sigma \in V_h$ is the order of $d$ in $\sigma$ and is denoted $\sigma(d)$. The order of $d$ is undefined if the generator does not contain $d$.

**Example 6.1.** Consider the example in Figure [1]. Assume $\beta_{1}^{h_1} = \beta_{2}^{h_1} = \beta_{3}^{h_1} = .45$.

We describe one generator of this system.

1. Select $d_2 \in D_{1}^{h_1}$; $d_6 \in D_{2}^{h_1}$; $d_8 \in D_{3}^{h_1}$

2. Let $\sigma$ be the order $(d_6, d_2, d_8)$

3. Set $v^\sigma(d_2, h_1) = 1/3$, $v^\sigma(d_6, h_1) = 1/3$; $v^\sigma(d_8, h_1) = 1/5$. The remaining coordinates are set to zero.
4. The remaining mass is $1 - 1/3 - 1/3 - 1/5 = 2/15$ is distributed in the order of $\sigma$. This gives $v^\sigma(d_6, h_1) = .45, v^\sigma(d_2, h_1) = .35, v^\sigma(d_8, h_1) = 1/5 = .2$

We say the generator $v^\sigma$ contains $d_2, d_6, d_8$. The order of the doctors in this generator are $\sigma(d_2) = 2; \sigma(d_6) = 1$ and $\sigma(d_8) = 3$.

6.3 Algorithm

Ranking of Columns in $A V$

We consider the conic version of Scarf’s lemma as in Section 3.3 The system $A \cdot V \cdot z \leq b$ is constructed in the same way as in Section 3.3 Each column of $A \cdot V$ corresponds to a generator $v^\sigma$, each row of $A \cdot V$ corresponds to a constraint for either a doctor or a hospital.

We now describe how each agent in $D \cup H$ ranks the columns of $A V$. (We use the word “rank” to distinguish between the ordering over the columns of $A V$ and $A$.)

- $h$ ranks two generators $v^\sigma, \bar{v}^\sigma' \in V_h$ according to the lowest ranked doctor (according to $>_d$) contained in each of them. If the lowest ranked doctor of both $v^\sigma$ and $\bar{v}^\sigma'$ are the same, say $d_{\text{min}}$, then break ties by comparing $\sigma(d_{\text{min}})$ and $\sigma'(d_{\text{min}})$. The lower the order, the less preferred. If they are equal, move to the second worst doctor contained in each and so on.

- $d$ compares two generators $v^\sigma, \bar{v}^\sigma'$ that contain $d$ according to the hospital that each assigns $d$ to using $>_d$. If $v^\sigma, \bar{v}^\sigma'$ both assign $d$ to the same hospital $h$, break ties by comparing $\sigma(d)$ and $\sigma'(d)$. Specifically, if $\sigma(d) > \sigma'(d)$, then $v^\sigma$ is ranked above $\bar{v}^\sigma'$. If $\sigma(d) = \sigma'(d)$, then $d$ uses $h$’s ordering over the generators to break the tie.

Scarf’s algorithm and rounding

We use the algorithm in Scarf (1967) to derive a dominating $z^* \in Q$. Set $x^* = Vz^*$, and use Lemma 2 to round $x^*$ into an integer $\bar{x}$ that satisfies (5) and almost satisfies (7). Let $\bar{\mu$
be the corresponding matching.

**Define** \( \bar{\alpha} \) and \( \bar{\beta} \).

- If \( \sum_{d \in D^h_t} x^*(d, h) = \alpha^h_t \sum_{d \in D^h_t} x^*(d, h) \), then, let
  \[
  \bar{\alpha}^h_t = \frac{\sum_{d \in D^h_t} \bar{x}(d, h)}{\sum_{d \in D^h_t} \bar{x}(d, h)}
  \]  \hspace{1cm} (15)

- If \( \sum_{d \in D^h_t} x^*(d, h) > \alpha^h_t \sum_{d \in D^h_t} x^*(d, h) \) but \( \sum_{d \in D^h_t} \bar{x}(d, h) < \alpha^h_t \sum_{d \in D^h_t} \bar{x}(d, h) \), then let \( \bar{\alpha}^h_t \) be as in (15). Otherwise \( \bar{\alpha}^h_t = \alpha^h_t \).

- Similarly, if \( \sum_{d \in D^h_t} x^*(d, h) = \beta^h_t \sum_{d \in D^h_t} x^*(d, h) \), then, let
  \[
  \bar{\beta}^h_t = \frac{\sum_{d \in D^h_t} \bar{x}(d, h)}{\sum_{d \in D^h_t} \bar{x}(d, h)}
  \]  \hspace{1cm} (16)

- If \( \sum_{d \in D^h_t} x^*(d, h) < \beta^h_t \sum_{d \in D^h_t} x^*(d, h) \) but \( \sum_{d \in D^h_t} \bar{x}(d, h) > \beta^h_t \sum_{d \in D^h_t} \bar{x}(d, h) \), then also let \( \bar{\beta}^h_t \) be as in (16). Otherwise \( \bar{\beta}^h_t = \beta^h_t \).

An argument similar that in Section 4.3 yields the following proximity bounds for \( \bar{\alpha} \) and \( \bar{\beta} \):

\[
|\alpha^h_t - \bar{\alpha}^h_t| = |\alpha^h_t - \frac{\bar{\mu}(h) \cap D^h_t}{|\bar{\mu}(h)|}| \leq \frac{2}{1 + \sum_{d \in D_t} x^*(d, h)} \forall D^h_t,
\]

and

\[
|\beta^h_t - \bar{\beta}^h_t| = |\beta^h_t - \frac{\bar{\mu}(h) \cap D^h_t}{|\bar{\mu}(h)|}| \leq \frac{2}{1 + \sum_{d \in D_t} x^*(d, h)} \forall D^h_t.
\]

Our main result is

**Theorem 6.2.** \( \bar{\mu} \) is feasible and stable for the instance \( \{\gamma_d\}_{d \in D}, \{\gamma_h\}_{h \in H}, \{\bar{\alpha}^h_t\}_{h \in H}, \{\bar{\beta}^h_t\}_{h \in H} \).
6.4 Stability of $\bar{\mu}$

Recall that $\mathcal{V}^* = \{ v^\sigma \in \mathcal{V} : z^*_{\omega} > 0 \}$. For each hospital $h$, the set of generators in $\mathcal{V}^*$ associated with hospital $h$ is denoted $\mathcal{V}^*_h$.

A group $D^h_i$ reaches its lower bound in $x^*$ if

$$\sum_{d \in D^h_i} x^*(d, h) = \alpha^h_i \sum_{d \in D} x^*(d, h).$$

Similarly, $D^h_i$ reaches its upper bound in $x^*$ if

$$\sum_{d \in D^h_i} x^*(d, h) = \beta^h_i \sum_{d \in D} x^*(d, h).$$

We use Observation 3 and the one below in the proof.

**Observation 4.** Fix a hospital $h$ and consider a partition of the set of types into 2 groups such that Group 1 contains $D^h_{i_1}, \ldots, D^h_{i_k}$ and group 2 contains $D^h_{i_{k+1}}, \ldots, D^h_{i_{T_h}}$ if for all doctors $d$ in group 2 and all $v^\sigma \in \mathcal{V}^*_h$, $\sigma(d) \geq k + 1$. Then, either each member of group 1 reaches its upper bound or each member in group 2 reaches its lower bound.

**Proof.** Notice that, $D^h_i$ reaches its lower bound in $x^*$ if and only if for all generator $v^\sigma \in \mathcal{V}^*_h$ that contains $d \in D^h_i$, $v^\sigma(h, d) = \alpha^h_i$. Similarly, $D^h_i$ reaches its upper bound in $x^*$ if and only if for all generator $v^\sigma \in \mathcal{V}^*_h$ containing $d \in D^h_i$, $v^\sigma(h, d) = \beta^h_i$.

According to the algorithm producing the generators, if we order the $T_h$ non-zero components in the order that they are selected by the algorithm, they will be of the form

$$\beta^h_{i_1}, \ldots, \beta^h_{i_k}, \gamma, \alpha^h_{i_{k+2}}, \ldots, \alpha^h_{T_h},$$

where $\gamma = 1 - \beta^h_{i_1} - \ldots - \beta^h_{i_k} - \alpha^h_{i_{k+2}} - \ldots - \alpha^h_{T_h}$.

Because of the assumption that for all doctors $d$ in group 2 and all $v^\sigma \in \mathcal{V}^*_h$, $\sigma(d) \geq k + 1$,
all doctors in group 2 is always selected after all doctors in group 1. Thus, either all the
types in group 1 reach their upper bound, and if not, it is because the remaining mass of
$1 - \sum_t \alpha^h_t$ is exhausted, and therefore all types in group 2 reach their lower bound.

Proof of Theorem 6.2

We are now ready to prove Theorem 6.2. Suppose, for a contradiction, that $\bar{\mu}$ is not stable.
Let $d_a, d_r \in D$ be such that $\bar{\mu}(d_r) = h,\ \bar{\mu}(d_a) \neq h$ and $d_a >_h d_r$, i.e., $d_a$ is wait-listed at $h$.
There are two cases to consider. In the first, $h$ does not reach to its effective capacity. In
the second, we can exchange $d_r$ for $d_a$ without violating (11).

Our goal is to construct a generator that is not dominated by $z^*$, which is a contradiction
to the domination of $z^*$.

Case 1: $h$ does not reach its effective capacity.

From Remark 3, this means that $|\bar{\mu}(h)| < k_h$. Because of the rounding does not violate
hospital capacity, the capacity constraint at $h$ does not bind in the fractional solution, that
is $\sum_{d \in D} x^*(d,h) < k_h$. Thus, no generator can be dominated at $h$. It remains to create a
generator $w^{\sigma'} \in \mathcal{V}_h$ that is not dominated via any doctor. This will lead to a contradiction
because $z^*$ is a dominating solution.

The first step is to choose $\sigma'$. Order the types so that all the types that contain a wait-
listed doctor come first, and choose one wait-listed doctor from each of the types to be part
of the generator. Let $\{i_1, \ldots, i_k\}$ be the set of these types. By observation 3 the generator
that we are constructing cannot be dominated by the the constraints at these doctors.

Denote the remaining set of types by $\{i_{k+1}, \ldots, i_{T_h}\}$. If there is a doctor $d \in D^h_{i_{k+1}} \cup\ldots\cup D^h_{i_{T_h}}$
and a $v^\sigma \in \mathcal{V}_h^*$ such that $\sigma(d) \leq k$, then take this doctor to be the next in the order $\sigma'$. Hence,$\sigma'(d) > \sigma(d)$. Therefore, $w^{\sigma'}$ cannot be dominated via $d$. Repeat, until we cannot find such

\footnote{$d_a, d_r$ denote for the doctor to accept and the doctor to reject, respectively.}
a doctor. Without loss of generality suppose this happens at the first instance. According to Observation 4, either $D^h_{h_1}, \ldots , D^h_{i_k}$ reaches its upper bound in $z^*$ or $D^h_{i_{k+1}} \cup \ldots \cup D^h_{r_{th}}$ reaches its lower bound in $z^*$. Because there is no wait-listed doctor in $D^h_{i_{k+1}} \cup \ldots \cup D^h_{r_{th}}$. Because of the rounding and modifying of $\alpha$, this means that the corresponding constraints in the rounded matching $\mu$ also bind. However, these types define the effective capacity at $h$, which contradicts the fact that $h$ is not at its effective capacity.

By this argument we create a generator $w^{\sigma'}$ not dominated via any doctor. This contradicts the fact that $z^*$ is a dominating solution.

**Case 2: $d_a$ can be exchanged for $d_r$.**

- We argue that $d_a$ and $d_r$ are not of the same type. Suppose they are of the same type. Let $v^\sigma \in V^*_h$ be a generator containing $d_r$. Let $w^{\sigma} \in V^*_h$ be obtained from $v^\sigma$ by shifting the probability weight from $d_r$ to $d_a$ but keeping the same order. Generator $w^{\sigma}$ is ranked above $v^\sigma$ by $h$ and all doctors of types that differ from $d_a$ and $d_r$. Moreover, because $d_a$ is a wait-listed doctor, this generator cannot be dominated via $d_a$. Thus, this new generator is not dominated by $z^*$, a contradiction.

- Given that $d_a \in D^h_{a}, d_r \in D^h_{r}$ are not of the same type, we argue that either $D^h_{a}$ reaches its upper bound, or $D^h_{r}$ reaches its lower bound in $z^*$. Suppose, for a contradiction, otherwise. Then, $D^h_{a}$ has not reached its upper bound, and $D^h_{r}$ has not reached its lower bound in $z^*$.

Among all doctors $d$ that have a type that has not reached its lower bound in $z^*$, and $x^*(d, h) > 0$ (doctor $d_r$ is a member of this set), let $d_{min}$ be the least preferred according to $>_h$. Let $D^h_{min}$ be the set of doctors of the same type as $d_{min}$. Let $v^\sigma \in V^*_h$ be a generator containing $d_{min}$ such that $d_{min}$’s order, $\sigma(d_{min})$, is as small as possible.  

---

9Because of the rounding procedure, and the modifying of $\alpha$-s, this implies that either $d_a$ is at surplus or $d_r$ is protected at the rounded matching $\bar{\mu}$, with the modified $\bar{\alpha}$-s, $\bar{\beta}$-s. This is what we need to prove.
Such a generator exists because $x^*(d,h) > 0$.

**Claim 1.** $d_{\min}, d_a$ are of different types; furthermore, if $d'_a \in D^h_a$ is contained in a generator $v^{\sigma'}$ ($d'_a$ may be the same as $d_a$), then, $\sigma(d_{\min}) > \sigma(d'_a)$.

**Proof.** If $d_{\min}$ and $d_a$ are of the same type, we could, in $v^{\sigma}$, shift the probability weight from $d_{\min}$ to $d_a$. This produces a generator that is not dominated via $h$ because $d_a >_h d_r \succeq_h d_{\min}$. It is clearly not dominated via $d_a$. Finally, it is not dominated via any doctor other than $\{d_a, d_{\min}\}$ in the two generators, who will face a tie and break it in $h$'s favor.

If $\sigma$ orders $d_{\min}$ before the type of $d'_a$, i.e., $\sigma(d_{\min}) < \sigma(d'_a)$, then switch the order of these two types, and if $d'_a \neq d_a$, shift the probability weight from $d'_a$ to $d_a$. This new generator is not dominated via $d_a$ because of Observation 3. Second, because we have switched the order of $d_{\min}$ and $d'_a$, this new generator is not dominated via $d_{\min}$. Third, because $d_a >_h d_r$ for all other doctors and for $h$, the new generator is ranked above $v^{\sigma}$. Therefore, it cannot be dominated. □

![Figure 4: The generator $v$](image)

Let $D^*$ be the set of doctors who belong to types that have not reached their lower bounds in $z^*$, and their order in $v^{\sigma}$ is at least the order of $\sigma(d_{\min})$. See Figure 4.
Because of Claim 1, $D^h_a \cap D^* = \emptyset$.

Suppose, among the doctors in $D^*$, there is a $d_t \in D^h_t$ whose order under a different generator $\tilde{\sigma}' \in V^*_h$ is no larger than $\sigma(d_{\min})$. Notice that $d_t$ could be of the same type as $d_{\min}$, but $d_t \neq d_{\min}$ because $d_{\min}$ is the least preferred under $>_h$.

Create a new generator from $\tilde{\sigma}'$ by switching the order of $D^h_t$ and $D^h_{\min}$, and let $d_t$ be part of this new generator (if $d_t, d_{\min}$ are the same type, replace $d_{\min}$ with $d_t$). The new generator is not dominated via $d_t$ because $d_t$ ranks it above $\tilde{\sigma}'$. It is also ranked above $v^{\sigma}$ by $h$, because $d_{\min}$ has a higher order, and thus is also ranked higher by all other doctors because they break ties according to $>_h$. Hence, this new generator is not dominated.

We are left with the case that the order of each doctor in $D^*$ in every generator in $V^*_h$ is at least $\sigma(d_{\min})$. This means that for all generators in $V^*_h$, doctors in $D^*$ are ordered after $\bar{D}$, where $\bar{D}$ is the set of doctors whose types are ordered before $D^h_{\min}$ under $\sigma$.

Similar to the argument in Observation 4, this means that either $\bar{D}$ reaches its upper bound, or $D^*$ reaches its lower bound. However, this is impossible because $D^h_{\min} \in D^*$ was chosen such that it does not reach its lower bound, and $D^h_a \in \bar{D}$ is assumed to not reach its upper bound.

This concludes the proof.

7 Conclusion

It is common to require that a matching satisfy a variety of distributional goals. These are sometimes expressed as lower or upper bounds on the proportion of agents of a particular type being matched. This paper is the first that we are aware of to address this problem. It uses a novel extension of Scarf’s lemma to identify a stable matching that approximately
satisfies such proportionality constraints. In addition, ex-post bounds on the deviation between the realized and desired proportions are provided.

References


