On the Communication Complexity of Stable Matching under Metric Space Market Embeddings

Naveen Durvasula

under the direction of

Prof. Scott Duke Kominers
Harvard Business School

Research Science Institute
July 31, 2018

Abstract

In this paper, we analyze the communication complexity of the stable marriage problem under a model in which agents' preferences are implicitly described by an embedding into a metric space. In our model, we define preference by the "closeness" of two agents given by the pairwise distance operator of the metric space. We answer an open question posed by Gonczarowski et al.: *is it possible to decrease the communication complexity of finding a stable matching when preference is implicitly indicated by input (for example, when agents are represented as points in a Euclidean space)?* We show that under such a model, there is no improvement in the communication complexity from the existing $\Theta(n^2)$ result shown for Boolean queries. We extend this result by using Matoušek's variant of the Johnson-Lindenstrauss transform to show an upper bound on the amount of miscommunication that can occur while preserving a preference profile.

Summary

In the stable marriage problem, there are n "men" and n "women", each with a preference ordering over members of the "opposite sex". We aim to find a *stable marriage*, or a one-to-one matching between the men and women such that no pair containing a man and a woman mutually prefer each other over their assigned match. In real-world markets, stability has proven to be an important condition for matching mechanisms to prevent market failures. As such, minimizing the amount of communication among agents, or the communication complexity, necessary to find a stable matching is of practical importance, as in doing so we reduce market congestion. Gonczarowski et al. show that in any market where communication can only proceed in the form of yes/no questions, on the order of n^2 queries are necessary to find a stable matching or verify that given matching is stable. We study the communication complexity of the stable marriage problem when the "men" and "women" are represented as points in space, and agents prefer others that are spatially "closer" to themselves. We study this model due to its connection to real-world markets: potential matching partners are defined by some number of properties, and preferences are formed entirely by these properties. Queries under this model are of the form, "What is an agent's kth property?". We show that even under this model, on the order of n^2 queries are required to find or verify a stable matching. We extend these results to show bounds on the amount of miscommunication that can occur among agents while still maintaining the original preference ordering.

1 Introduction

In the *stable marriage problem*, there are n men and n women, each with preference ordering over members of the opposite sex. We call the full set of preference orderings for all men and women the men's and women's *preference profiles* respectively. A *marriage* is stable iff there are no *blocking pairs*, that is, a man and a woman that both prefer each other over their assigned partners. In their seminal paper, Gale and Shapley introduced the *deferred acceptance* (DA) algorithm to (in $\Theta(n^2)$ time) find a stable marriage efficiently for any matching market [1]. Experimental evidence has demonstrated the importance of stability in real-world markets as a means of preventing market failures. Hiring markets with stable matching mechanisms did not undergo the process of *unraveling*, or the phenomenon in which contracts are formed far before the employment date due to competition among those hiring to match with the best candidates [2].

As real-world markets increasingly adopt stable matching mechanisms, minimizing the amount of information that relevant parties need to communicate (and hence the amount of congestion in the market), is of practical significance. We analyze the *communication complexity*, or the minimum amount of information in bits that must be passed among agents to solve the stable marriage problem. This definition of communication complexity leads to a natural method of analysis: given that agents answer Boolean queries such as, "Does woman w prefer man m over man m'?", or vice-versa, what is the minimum number of queries necessary? We motivate this type of analysis by interpreting it in the context of real-world markets. In practice, especially in large markets, it is unlikely and/or unreasonable to assume that each agent has formed preferences over all potential match candidates. Furthermore, in decentralized markets, agents learn preferences through pairwise queries. Finding answers to these queries can be costly as resource intensive methods such as interviews may be necessary. Therefore, a protocol that minimizes the number of queries necessary can reduce market congestion.

Under a specified Boolean query scheme, Ng and Hirschberg show that $\Theta(n^2)$ queries are required to construct a stable matching or to verify that a given matching is stable [3]. In [4], Segal expanded this query model to allow for any set of Boolean queries, and proved a weaker bound of $\Omega(n^2)$. In [5], Chou and Lu show a bound of $\Theta(n^2 \log(n))$ for all deterministic communication protocols in which at most a constant fraction of the population forms blocking pairs. In [6], Gonczarowski et al. proved by various reductions to the disjointness problem [7, 8] that any Boolean query system requires $\Omega(n^2)$ queries in the worst case to find a stable or *approximately*[1] stable marriage. The $\Omega(n^2)$ bound applies to both *deterministic communication protocols* and *randomized communication protocols*. While a deterministic communication protocol must always output the correct answer, a randomized communication protocol must for any fixed input output the correct answer with probability $p \geq \frac{1}{2}$ (it is standard to let $p = \frac{2}{3}$). In [9], Ashlagi et al. give a more optimistic result than the $\Omega(n^2)$ bound by constructing a communication protocol that requires $O^*(\sqrt{n})$ bits[2] of communication per agent and finds a stable matching with high probability under natural assumptions of the agent's prior knowledge.

1.1 Our model

We consider a general formulation intended to more realistically model matching processes that occur in the real-world. The applied significance of communication complexity analysis is in understanding the minimum amount of market congestion required to *naturally* arise at a stable matching. If instead a stable matching mechanism is employed, the communication complexity becomes irrelevant, as there is no practical cost to having agents

[1]This definition is based on the minimum *divorce distance* to a stable marriage, or the minimum number of pairs that must be divorced to reach the most similar stable marriage. We direct the reader to [6] for a more formal definition.

[2]$f(n) = O^*(g(n))$ if there exists n_0 and $C > 0$ such that $f(n) \leq (\log n)^C g(n) \ \forall n \geq n_0$.

share information once all of the preference lists have been submitted to a central clearinghouse. In the natural setting, it is unrealistic to assume that agents have defined preferences over match candidates before communicating with them. Our formulation aims to incorporate *preference learning* into the communication process. As agents aim to find a stable matching, they are simultaneously learning their preferences, and the preferences of others.

In our formulation, men and women are represented as elements of *metric spaces*, or sets of elements combined with a pairwise distance operator. We use two metric spaces in our model: the *man space* representing the set of possible men and the *woman space* representing the set of possible women. The sets and distance operators are known publicly to all agents. We let each agent have an *ideal* and an *identity*, both of which are known privately to the agent. An agent's ideal is its most desirable match partner. Agents seek candidates on the other side of the market who have identities that are most similar to the agent's ideal. This measure of "similarity" is represented as an arbitrary distance metric. We note that a man's ideal is a woman, and therefore belongs to the woman space. Consequently, the woman space must capture all of the men's preferences.

We note that under this formulation, the information that each agent has access to differs from that in the various formulations used to generate previous results. In these standard formulations, each agent knows privately their preferences over agents on the other side of the market. However, no agent knows how they rank in the preference lists of others. In this formulation, agents effectively have private access to two *comparing operators*. One of which takes as input the identities of two potential match partners, and by computing the distance between the given agent's ideal and the two identities, the operator can determine which candidate is preferred by the agent. The other operator that the agent has access to takes as input the ideals of two potential match candidates, and by computing the distance between the given agent's identity and the two ideals, the operator can determine which candidate *prefers* the agent more. Thus, agents, to some degree, have access to the preferences of *both* sides of the market.

As an example, consider the scenario where job candidates are being matched to firms. The job candidate's ideal is their ideal firm, and their identity refers to themselves. Likewise, a firm's ideal is their ideal job candidate, and their identity is themselves. To arrive at a stable matching, both sides must learn each others identities, so that they may rank matching candidates, and they must learn each other's ideals to arrive at a stable match. It is in this way that we capture the process of preference learning in our model. This allows the communication complexity to refer to the practical amount of correspondence required to arise at a stable match.

Our query model also differs from that employed in prior analysis. In standard formulations, one might ask either a series of Boolean queries or a series of integer-response queries (the latter can be reduced to the former by allowing each bit of the integer response to be queried separately). In our model, agents may be queried for the information they know privately: their ideal and identity. However, if we wish for one query to correspond to the natural communication cost of extracting some information from some agent, it is unrealistic for the agent to disclose the entirety of their identity or the entirety of their ideal in one query. In context of our running example, it is unreasonable for a firm to learn the entirety of the job candidate's identity within a single correspondence. Instead, the applicant will probably submit a resume and go through a number of interviews, where in each of the steps, the firm learns more about the candidate until (in the ideal case), the candidate's identity is fully discovered. Thus, we let the information defining the agents' ideals and identities be arbitrarily partitioned, and attribute a unit query cost to revealing each block of the partition. We enforce that all ideals and identities belonging to the man space be partitioned in the same way. We impose the same restriction for women.

While we formulate our model mathematically in Section 2, we describe it more formally here. Agents initially know their own ideal and identity as elements of the two metric spaces.

The structures of the metric spaces (the sets and paired distance operators) are known globally among the participants. Each of the agents' ideals and identities are then partitioned into query-able blocks such that all elements belonging to the same metric space are partitioned in the same way. While the amount of information contained within each block need not be the same, the practical cost of querying the information within each block should be the same.

The communication process occurs in two phases: the *sketching phase* and the *query phase*. In the sketching phase, no communication is counted. However, with two globally known transformations, men/women may transform their identities/ideals to a subspace of the man space such that the same distance operators continue to indicate preference. Likewise, the same process is done with elements of the woman space. We call these transformations the *men's sketch* and *women's sketch* respectively. While the transformations are globally known, they may not depend on the agents' identities and ideals, as this would effectively allow agents to freely communicate. The idea behind the sketching phase is to reduce the amount of information necessary to communicate an agent's identity or ideal. Suppose that we have partitioned an agent's identity or ideal into k blocks. If we are able to transform every man's/woman's identity/ideal into a subset of man space that can be fully represented by $l < k$ blocks, then we can reduce the total number of queries necessary. In the query phase, a server queries the agents for their identities and ideals to find a stable matching. The server may ask any agent for the value of the ith block of their identity or ideal. We attribute a unit communication cost for each such query. We choose a server based model as it establishes a lower bound on the amount of communication required should the $2n$ agents communicate with each other.

Our use of the terms *server* and *sketch* coincide with those used by Chou and Lu in [5]. Indeed, we show that for specific metric space formulations that are of practical significance, the results of Chou and Lu can be directly applied.

2 Preliminaries

2.1 Stability and Matching Markets

Let M and W be disjoint sets of men and women respectively where $|M|=|W|=n$. Let a *matching* μ be a one-to-one mapping between M and W. For each agent $i \in M \cup W$ we define a preference operator $\succ^i$, where $a \succ^i b$ iff i prefers a over b. We consider only *strict* preferences; that is, we do not allow indifferences between matching candidates. For any agent $i \in M \cup W$, we let $\mu(i)$ be the agent that i is matched to under μ. A matching μ has a *blocking pair* if there exists a tuple (m, w) where $m \in M$ and $w \in W$ such that $w \succ^m \mu(m)$ and $m \succ^w \mu(w)$. A *stable matching* is a matching with no blocking pairs. We let the set $P_M = (\succ^m)_{m \in M} \subset \mathcal{P}_M(W)$ be the *preference profile* of the men, where $\mathcal{P}_M(W)$ is the set of all possible preference operators over the set W. We analogously define P_W and $\mathcal{P}_W(M)$ for the women. A preference profile is *complete* if every agent prefers being matched over being unmatched. We consider only complete preference profiles. A *matching market* is given by the tuple (M, W, P_M, P_W). We denote by n the *size* of the market (M, W, P_M, P_W). In the stable marriage problem, we aim to find a stable matching μ given some matching market (M, W, P_M, P_W).

2.2 Our formulation

We now formulate mathematically our model described in Section 1.1.

Definition 2.1 (Metric Space)**.** A metric space (S, d) consists of a set S endowed with a distance function $d : S \times S \mapsto \mathbb{R}_{\geq 0}$ such that the following properties hold for any $a, b, c \in S$

$$d(a,b) \geq 0 \quad (1)$$

$$d(a,b) = 0 \iff a = b \quad (2)$$

$$d(a,b) = d(b,a) \quad (3)$$

$$d(a,c) \leq d(a,b) + d(b,c) \quad (4)$$

Property (1) guarantees that the distance between two elements is non-negative. Property (2) states that if two elements have distance 0, they must be the same element. Property (3) states that distances are symmetric. Property (4) states that the triangle inequality must hold.

We define the *man space* $(\mathbb{M}, d_M)$ and the *woman space* $(\mathbb{W}, d_W)$ as *finite* cardinality metric spaces with arbitrary distance operators. For each agent we let $\theta : M \cup W \mapsto \{M, W\}$ take as input an agent i and returns the set M if $i \in M$ and W otherwise. Likewise, we define $\phi : M \cup W \mapsto \{M, W\}$ which takes as input an agent i and returns the set of the *opposite* gender (i.e. if $i \in M$, $\phi(i) = W$). Occasionally (as in the instance directly below), we will abuse notation and allow $\phi(i)$ and $\theta(i)$ to return the metric spaces $\mathbb{M}$ and $\mathbb{W}$ instead of the sets M and W.

Definition 2.2 (Ideals and Identities)**.** Let identity : $M \cup W \mapsto \mathbb{M} \cup \mathbb{W}$ map each agent i to an element of $\theta(i)$. Let ideal : $M \cup W \mapsto \mathbb{M} \cup \mathbb{W}$ map each agent i to an element of $\phi(i)$.

In our model, each agent initially knows privately the values of identity(i) and ideal(i). We now mathematically formalize the partitioning system used in our query system as specified in Section 1.1. Depending on the size of $\mathbb{M}$ or $\mathbb{W}$, the amount of information necessary to represent the identities and ideals of the agents can vary. We wish to partition each identity and ideal of each agent into query-able blocks of information such that the practical cost of querying for each block is the same. We do this by defining *index sets* for the metric spaces $\mathbb{M}$ and $\mathbb{W}$. We first define the *indexing function* $I(k, \mathbf{v})$ for $k \in \mathbb{N}$ and $\mathbf{v} \in \mathbb{N}^k$ which takes in a vector $\mathbf{v}$ of *indices* and returns the set given by

$$I(k, \mathbf{v}) = \prod_{i=1}^{k} \{1, \ldots, v_i\}$$

Definition 2.3 (Index Set)**.** If for some $k, \mathbf{v}$, $I(k, \mathbf{v}) \simeq \mathbb{S}$, we say that $I(k, \mathbf{v})$ *indexes* $\mathbb{S}$.

The set $I(k, \mathbf{v})$ indexes a set $\mathbb{S}$ if and only if $|I(k, \mathbf{v})| = \prod_{i=1}^{k} v_i = |\mathbb{S}|$ by construction. Should this be the case, the $\mathbf{v}$ vector induces a *partition structure* on $\mathbb{S}$, partitioning the set into k blocks. The ith block is given by the ith term in the Cartesian product $\{1, \ldots, v_i\}$. We can reference some element $s \in \mathbb{S}$, by selecting exactly one element from each block. Note that as we have required $\mathbb{M}$ and $\mathbb{W}$ to be finite spaces, we can always construct index sets isomorphic to $\mathbb{M}$ and $\mathbb{W}$ using the indexing function I.

We use the partition structure induced by an index set in our query model. We require as input to our model index sets $I(k_M, \mathbf{v}_M) \simeq \mathbb{M}$ and $I(k_W, \mathbf{v}_W) \simeq \mathbb{W}$. We overload our notation, and let identity(i, j) for $i \in M \cup W$ and $j \in \left\{1, \ldots, k_{\theta(i)}\right\}$ return the value of the jth block of agent i's identity. Likewise, we let let ideal(i, j) for $i \in M \cup W$ and $j \in \left\{1, \ldots, k_{\phi(i)}\right\}$ return the value of the jth block of agent i's ideal. The only queries that the server in our model may make are identity(i, j) and ideal(i, j). Note that as the value of identity(i, j) or ideal(i, j) can take on one of $v_{\theta(i)j}$ or $v_{\phi(i)j}$ values respectively, the number of bits required to send this response is given by $\log(v_{\theta(i)j})$ or $\log(v_{\phi(i)j})$. However, as the practical cost of obtaining information from each block is assumed to be the same, we attribute unit cost to making either of these queries.

As stated, our model is quite general, admitting any partition structures over the metric spaces $(\mathbb{M}, d_M)$ and $(\mathbb{W}, d_W)$. However, there are a certain family of index sets which are of practical importance. We call these sets *k-unique sets*. In a k-unique index set, we let v_i be a function of $n = |M| = |W|$. The purpose of this is to allow each agent to have, roughly speaking, *unique* values in each of the blocks. In our running example of matching job candidates to firms, suppose that a job candidate's identity is partitioned by an index set into their personality and their ability. In other words, firms may query candidates for their personality and ability, and once both are known, the firm knows the candidate's identity in its entirety. Formally, this would be an index set of the form $I(2, \mathbf{v})$. As both of these attributes are quite abstract, we would like to allow each job candidate to have unique representations of their personality and ability. Thus, v_1 and v_2 should be at least equal n, the number of job candidates. We formalize this definition below

Definition 2.4 (k-unique set)**.** An index set $I(k, \mathbf{v})$ is a k-unique set if for all $v_i \in \mathbf{v}$, $v_i = \text{poly}(n)$.

We let v_i be a polynomial of n as this makes the total amount of information contained within each block $\log(v_i) = \log(\text{poly}(n)) = \Theta(\log(n))$ easy to compute. A k-unique set which we will continue to refer to is the *computable real numbers* $\mathbb{R}_c$, and its k-fold Cartesian product $\mathbb{R}_c^k$. The computable real numbers are a finite-cardinality representation of the real numbers. For example, *long variables* are 64 bit representations of the real numbers (the uncountable set is reduced to one of size 2^{64}). Suppose we were to uniquely embed n agents into $\mathbb{R}_c$. We cannot let $|\mathbb{R}_c|$ be a constant, as if n exceeds this constant, we cannot allow each agent to have a unique value. Thus, we let the size of $|\mathbb{R}_c|$ grow as an arbitrary polynomial of n. We likewise define $\mathbb{R}_c^k$ as $\prod_{i=1}^{k} \mathbb{R}_c$. We partition $\mathbb{R}_c^k$ in the natural way with the index set $I(k, \mathbf{v})$ where $v_i = |\mathbb{R}_c|$.

2.3 Sketches and Embeddings

We define the *space profile* as the aggregation of the metric spaces along with the parameters that define them.

Definition 2.5 (Space Profile)**.** A *space profile* $\mathcal{S} = [M, W, (\mathbb{M}, d_M), (\mathbb{W}, d_W), I(k_M, \mathbf{v}_M), I(k_W, \mathbf{v}_W)]$ consists of the sets of agents M and W, and the metric spaces $(\mathbb{M}, d_M)$ and

$(\mathbb{W}, d_W)$ which are indexed by $I(k_M, \mathbf{v}_M)$ and $I(k_W, \mathbf{v}_W)$ respectively. If $I(k_M, \mathbf{v}_M)$ is k_M-unique and $I(k_W, \mathbf{v}_W)$ is k_W-unique, we call $\mathcal{S}$ a *unique space profile*.

We define two equivalence relations among space profiles: *isometry* and *transformability*. Two space profiles are *isometric* if their index sets are identical and their metric spaces are effectively the same.

Definition 2.6 (Isometry)**.** A space profile $\mathcal{S} = [M,\ W,\ (\mathbb{M}, d_M),\ (\mathbb{W}, d_W),\ I(k_M, \mathbf{v}_M),$ $I(k_W, \mathbf{v}_W)]$ is *isometric* to a space profile $\mathcal{S}' = [M',\ W',\ (\mathbb{M}', d'_M),\ (\mathbb{W}', d'_W),\ I(k'_M, \mathbf{v}'_M),$ $I(k'_W, \mathbf{v}'_W)]$ if and only if

1. $M = M'$ and $W = W'$
2. $\mathbf{v}_M = \mathbf{v}'_M$ and $\mathbf{v}_W = \mathbf{v}'_W$
3. there exist bijective transformations $\varphi_M : \mathbb{M} \mapsto \mathbb{M}'$ and $\varphi_W : \mathbb{W} \mapsto \mathbb{W}'$ such that for all elements $x, y \in \mathbb{M}$, $d_M(x, y) = d'_M(\varphi_M(x), \varphi_M(y))$ and for all elements $x, y \in \mathbb{W}$, $d_W(x, y) = d'_W(\varphi_W(x), \varphi_W(y))$

We say that two space profiles are *transformable* if each of the distance metrics can be written as an invertible function of the other.

Definition 2.7 (Transformability)**.** A space profile $\mathcal{S} = [M, W, (\mathbb{M}, d_M), (\mathbb{W}, d_W), I(k_M, \mathbf{v}_M),$ $I(k_W, \mathbf{v}_W)]$ is *transformable* to a space profile $\mathcal{S}' = [M', W', (\mathbb{M}', d'_M), (\mathbb{W}', d'_W), I(k'_M, \mathbf{v}'_M),$ $I(k'_W, \mathbf{v}'_W)]$ if and only if

1. $M = M'$ and $W = W'$
2. $\mathbf{v}_M = \mathbf{v}'_M$ and $\mathbf{v}_W = \mathbf{v}'_W$
3. there exist isomorphisms $\varphi_M : \mathbb{M} \mapsto \mathbb{M}'$ and $\varphi_W : \mathbb{W} \mapsto \mathbb{W}'$

4. there exist invertible functions $\tau_M : \mathbb{R} \mapsto \mathbb{R}$ and $\tau_W : \mathbb{R} \mapsto \mathbb{R}$ such that for all elements $x, y \in \mathbb{M}$, $d_M(x, y) = \tau_M(d'_M(\varphi_M(x), \varphi_M(y)))$ and for all elements $x, y \in \mathbb{W}$, $d_W(x, y) = \tau_W(d'_W(\varphi_W(x), \varphi_W(y)))$

Transformability is a strictly more general relation than isometry. Two transformable space profiles are isometric if and only if τ_M and τ_W are the identity maps. We now tie together everything defined in Section 2.2 in the definition of an *instance*.

Definition 2.8 (Instance)**.** An *instance* ($\mathcal{S}$, identity, ideal) consists of a space profile $\mathcal{S}$ and the maps ideal and identity.

We wish for the distance between an agent's ideal and matching candidate's identity to indicate the degree to which the agent prefers that candidate. We formalize this by associating a *preference function* and corresponding preference profiles with an instance.

Definition 2.9 (Preference Functions and Profiles)**.** Given an instance X, its corresponding preference function

$$P_X(i, j) = d_{\phi(i)}(\text{ideal}(i), \text{identity}(j))$$

returns the distance between the ideal matching partner of an agent $i \in M \cup W$ and the identity of some candidate $j \in \phi(i)$. We use this function to associate preference profiles $P_M^X = (\succ_X^m)_{m \in M}$ and $P_W^X = (\succ_X^w)_{w \in W}$ with an instance X. For all $i \in M \cup W$ and $j, k \in \phi(i)$ we let

$$j \succ_X^i k \iff P(i, j) < P(i, k)$$

We now define the *market embedding* which relates the stable marriage problem as defined in 2.1 with our metric space formulation.

Definition 2.10 (Embedding)**.** An embedding E of a matching market (M, W, P_M, P_W) is an instance that satisfies the conditions $P_M^E = P_M$ and $P_W^E = P_W$.

We say that a space profile $\mathcal{S}$ *admits* a matching market (M, W, P_M, P_W) iff there exists some maps identity and ideal such that instance ($\mathcal{S}$, identity, ideal) is an embedding of that market. We call space profiles that admit *all* size-n matching markets n*-complete* space profiles.

In our communication model, we have two phases: the sketching phase and the query phase. In Section 2.2, we specified the nature of the nature of the queries that the server may make. The sketching phase occurs prior to the query phase. In this phase, agents transform their identities and ideals into subspaces of $\mathbb{M}$ and $\mathbb{W}$ so as to compress the information they have into fewer query blocks. The preference profile may change under the transformation, but the set of stable matchings given the altered preference profile must remain the same. We formalize this mathematically below.

Definition 2.11 (Sketch)**.** Let $\mathcal{S}$ be a n-complete space profile. Let $\ell_M \leq k_M$ and $\ell_W \leq k_W$. Let $\mathbf{u}_M = \left(v_{Mf(i)}\right)_{i=1}^{\ell_M}$ and $\mathbf{u}_W = \left(v_{Wg(i)}\right)_{i=1}^{\ell_W}$ for $f, g : \mathbb{N} \mapsto \mathbb{N}$ be *subsequences* of $\mathbf{v}_M$ and $\mathbf{v}_W$ of lengths ℓ_M and ℓ_W respectively. A *sketch* $(T_M, T_W)_{\mathcal{S}}$ of $\mathcal{S}$ is a tuple of transformations $T_M : \mathbb{M} \mapsto S_M$ and $T_W : \mathbb{W} \mapsto S_W$ such that the following hold:

1. $S_M \simeq I(\ell_M, \mathbf{u}_M)$ and $S_W \simeq I(\ell_W, \mathbf{u}_W)$

2. Let $\mathcal{S}'$ be the space profile given by $[M, W, (S_M, d_M), (S_W, d_W), I(\ell_M, \mathbf{u}_M), I(\ell_W, \mathbf{u}_W)]$. Let $c = (\text{identity}, \text{ideal})$ be an arbitrary tuple of identity and ideal maps. We denote by
$$O_c = (\mathcal{S}, \text{identity}, \text{ideal})$$
the *original instance* given by c and the original space profile $\mathcal{S}$. Similarly, we denote by X_c the *transformed instance* given by
$$X_c = (\mathcal{S}', T_\theta \circ \text{identity}, T_\phi \circ \text{ideal})$$
generated by transforming the ideals and identities of the agents by T_M and T_W. We enforce that for all c, the set of stable matchings given the market $(M, W, P_M^{X_c}, P_W^{X_c})$

must be identical to the set of stable matchings given the market $(M, W, P_M^{O_c}, P_W^{O_c})$.

In the above definition, we compress the identities of the men and the ideals of the women into $\ell_M \leq k_M$ query blocks. Likewise, we compress the identities of the women and the ideals of the men into $\ell_W \leq k_W$ query blocks. We select which ℓ_M and ℓ_W query blocks we compress to using the subsequences $\mathbf{u}_M$ and $\mathbf{u}_W$. Sketches are weaker constructions than embeddings as we do not require the induced preference profiles $(P_M^{X_c}, P_W^{X_c})$ to be the same as $(P_M^{O_c}, P_W^{O_c})$. Furthermore, the transformed space profile $\mathcal{S}'$ is not required to be n-complete. We let the set of all preference profiles $P_M^{X_c}$ induced by the set of transformed instances X_c be $\mathcal{P}_M^{(T_M,T_W)\mathcal{S}}$. Likewise, we let the set of all $P_W^{X_c}$ be $\mathcal{P}_W^{(T_M,T_W)\mathcal{S}}$. These two sets denote the condensed sets of preference profiles admitted by $\mathcal{S}'$ after transforming the identities and ideals of each of the agents in accordance with the sketching transformations.

In a real-world context, sketches represent systems like application processes. The objective of an application process is to encode an applicant's identity into a condensed form. This compression may alter the preference profiles of the agents. However, in the ideal case, the set of stable matchings remains the same.

3 Computing the Communication Complexity

In this section, we show bounds on the communication complexity of stable matching in our formulation (both in the sketching and query phases). We also prove facts about space profiles that relate to real-world matching mechanisms.

3.1 Restrictions on n-complete Space Profiles

We characterize the nature of n-complete space profiles. Recall that these space profiles must admit *all* matching markets (M, W, P_M, P_W) of size n. We accomplish this by showing bounds on the parameters of the index sets $I(k_M, \mathbf{v}_M)$ and $I(k_W, \mathbf{v}_W)$. We then extend our results to the realistic case where $\mathcal{S}$ is a unique space profile.

We begin by characterizing the index vectors $\mathbf{v}_M$ and $\mathbf{v}_W$.

Lemma 3.1. *If $\mathcal{S}$ is a n-complete space profile, then $\sum_{i=1}^{k_M} \log(v_{Mi}) = \Omega(n \log(n))$ and $\sum_{i=1}^{k_W} \log(v_{Wi}) = \Omega(n \log(n))$.*

Proof. Let $\mathcal{P}_n$ denote the set of possible preference profiles P_M (or identically P_W) in a matching market of size n. As each preference operator induces an ordering among n matching candidates, there are $n!$ such preference operators. It follows that $|\mathcal{P}_n| = (n!)^n$. Thus, the number of bits contained within a given preference profile is given by $\log((n!)^n) = \Theta(n^2 \log(n))$.

We further note by construction that for any embedding $E = (\mathcal{S}, \text{identity}, \text{ideal})$, elements of the metric space $\mathbb{W}$ fully determine the preference profile P_M^E, and identically, elements of the metric space $\mathbb{M}$ fully determine the preference profile P_W^E. Thus, the total number of bits H necessary to store the n ideals of agents on one side of the market and the n identities of agents on the other side of the market must be equal to $\Omega(n^2 \log(n))$.

We show the computation of H for elements of the metric space $\mathbb{W}$. The argument is symmetric for elements of $\mathbb{M}$. As $|\mathbb{W}| = \prod_{i=1}^{k_W} v_{Wi}$, the number of bits contained within an element of $\mathbb{W}$ is given by $\log\left(\prod_{i=1}^{k_W} v_{Wi}\right) = \sum_{i=1}^{k_W} \log(v_{Wi})$. As we are given n male ideals and n female identities, we are given $2n$ elements of $\mathbb{W}$. Thus, we have that $H = 2n \sum_{i=1}^{k_W} \log(v_{Wi}) = \Omega(n^2 \log(n))$. The desired result $\sum_{i=1}^{k_W} \log(v_{Wi}) = \Omega(n \log(n))$ follows. □

We extend this result to unique n-complete space profiles.

Corollary 1. *If $\mathcal{S}$ is a unique n-complete space profile, then $k_M = \Omega(n)$ and $k_W = \Omega(n)$.*

Proof. Given k-unique index sets $I(k_M, \mathbf{v}_M)$ and $I(k_W, \mathbf{v}_W)$,

$$\sum_{i=1}^{k_M} \log(v_{Mi}) = \sum_{i=1}^{k_M} \Theta(\log(n)) = k_M \Theta(\log(n))$$

Likewise, $\sum_{i=1}^{k_W} \log(v_{Wi}) = k_W \Theta(\log(n))$. Lemma 3.1 gives us that $k_M \Theta(\log(n)) = \Omega(n \log(n))$. Dividing by $\log(n)$, we get our desired result $k_M = \Omega(n)$, and identically, $k_W = \Omega(n)$. □

Corollary 1 has tangible real-world significance. Suppose agents determine their preferences by analyzing k different attributes about each potential match candidate. Furthermore, suppose that, to some degree, candidates may have unique values for each of the attributes. Corollary 1 tells us that if we wish for every preference profile to be realizable in this system, the number of attributes k to be considered must be asymptotically at least equal to the number of match candidates. This is a completely unrealistic condition to impose from a practical perspective. Thus, it is likely that many preference profiles cannot be realized in real-world markets.

3.2 The Sketching Phase

We now utilize the results from Chou and Lu in [5] to show bounds on the effectiveness of *sketches* (as defined in Section 2.3) in reducing the communication complexity of stable matching.

We restate Theorem 1 from [5] in the language of Section 2.

Theorem 3.2 (Chou and Lu). *Let $(T_M, T_W)_{\mathcal{S}}$ be a sketch of a n-complete space profile $\mathcal{S}$. The number of bits that a preference profile $P_M^{X_c} \in \mathcal{P}_M^{(T_M,T_W)_\mathcal{S}}$ (or identically, a profile $P_W^{X_c} \in \mathcal{P}_W^{(T_M,T_W)_\mathcal{S}}$) contains must be at least $n \log(n!)$.*

Theorem 3.2 allows us to extend our results from Section 3.1 to *sketches*. Recall that in a sketch, we transform the identities and ideals of the agents to subsets of the spaces $\mathbb{M}$ and $\mathbb{W}$ that can be indexed by fewer query blocks. Furthermore, we don't require the induced preference profiles under the transformation to be the same as the original one. Instead, we only require that the set of stable matchings remain invariant under the transformation.

We first prove the analog of Lemma 3.1 for sketches.

Lemma 3.3. *Let $(T_M, T_W)_{\mathcal{S}}$ be a sketch of a n-complete space profile $\mathcal{S}$. Let $\mathcal{S}'$ be the corresponding transformed space profile. $\sum_{i=1}^{\ell_M} \log(u_{Mi}) = \Omega(n\log(n))$ and $\sum_{i=1}^{\ell_W} \log(u_{Wi}) = \Omega(n\log(n))$.*

Proof. By Theorem 3.2, we have that the total number of bits H necessary to store the n transformed ideals of agents on one side of the market and the n transformed identities of agents on the other side of the market must be equal to $\Omega(n^2\log(n))$. We show the computation of H for elements of the metric space S_W. The argument is symmetric for elements of S_M. As $|S_W| = \prod_{i=1}^{\ell_W} u_{Wi}$, the number of bits contained within an element of S_W is given by $\log\left(\prod_{i=1}^{\ell_W} u_{Wi}\right) = \sum_{i=1}^{\ell_W} \log(u_{Wi})$. As we are given n transformed male ideals and n transformed female identities, we are given $2n$ elements of S_W. Thus, we have that $H = 2n\sum_{i=1}^{\ell_W} \log(u_{Wi}) = \Omega(n^2\log(n))$. The desired result $\sum_{i=1}^{\ell_W} \log(u_{Wi}) = \Omega(n\log(n))$ follows. □

We now show the corresponding corollary for unique space profiles.

Corollary 2. *Let $(T_M, T_W)_{\mathcal{S}}$ be a sketch of a unique n-complete space profile $\mathcal{S}$. Let $\mathcal{S}'$ be the corresponding transformed space profile. $\ell_M = \Omega(n)$ and $\ell_W = \Omega(n)$.*

Proof. As $\mathcal{S}$ is a unique space profile and $\mathbf{u}_M$ and $\mathbf{u}_W$ are subsequences of $\mathbf{v}_M$ and $\mathbf{v}_W$ respectively, $\mathcal{S}'$ must also be a unique space profile. Given k-unique index sets $I(\ell_M, \mathbf{u}_M)$ and $I(\ell_W, \mathbf{u}_W)$,

$$\sum_{i=1}^{\ell_M} \log(u_{Mi}) = \sum_{i=1}^{\ell_M} \Theta(\log(n)) = \ell_M\Theta(\log(n))$$

Likewise, $\sum_{i=1}^{\ell_W} \log(v_{Wi}) = \ell_W\Theta(\log(n))$. Lemma 3.3 gives us that $\ell_M\Theta(\log(n)) = \Omega(n\log(n))$. Dividing by $\log(n)$, we get our desired result $\ell_M = \Omega(n)$, and identically, $\ell_W = \Omega(n)$. □

Corollary 2 leads to a profound but unfortunate conclusion about real-world markets. Suppose again that agents' preferences are determined by k different attributes about each potential match candidate. Furthermore, suppose that, to some degree, candidates may have unique values for each of the attributes. Corollary 2 tells us that if agents estimate their preferences by considering $\ell < k$ attributes where ℓ is not asymptotically at least equal to the number of potential match candidates, then a match that is stable according to their estimated preferences may not be stable according to their true preferences. As in practice, even in stable matching mechanisms such as the National Residency Matching Program, agents likely do not consider $\Omega(n)$ attributes about each candidate, it is quite possible that their estimated preferences lead to an unstable match.

3.3 The Query Phase

We now show that in unique space profiles, a server that can actively query for $\text{ideal}(i, j)$ and $\text{identity}(i, j)$ must make $\Theta(n^2)$ queries in order to find a stable matching. Again, we adapt a result from Chou and Lu in [5] to prove this statement.

We restate Theorem 3 from [5].

Theorem 3.4 (Chou and Lu)**.** *Any server which outputs a stable matching must receive* $\Theta(n^2 \log(n))$ *bits of information.*

We build on this to get the desired result.

Corollary 3. *Given an embedding* $(\mathcal{S}, \text{identity}, \text{ideal})$ *where* $\mathcal{S}$ *is a unique space profile, a server must make* $\Theta(n^2)$ *queries of the type* $\text{ideal}(i, j)$ *and* $\text{identity}(i, j)$ *to output a stable matching.*

Proof. As $\mathcal{S}$ is a unique space profile, the server receives $\Theta(\log(n))$ bits from any query of the type $\text{ideal}(i, j)$ or $\text{identity}(i, j)$. If Q is the total number of queries the server must make, we have by Theorem 3.4 that $Q\Theta(\log(n)) = \Theta(n^2 \log(n))$. Dividing by $\log(n)$, we get that $Q = \Theta(n^2)$. □

Corollary 3 tells us that even under this more realistic, and seemingly more relaxed model, $\Theta(n^2)$ queries are still required to compute a stable matching. Thus, even in practice, finding a stable matching requires a lot of communication.

4 Converting Stretches to Sketches

One advantage of the metric space formulation is that it allows us to analyze the process of preference learning in detail. In this section, we analyze the effects of *miscommunication* on stable matching. In real-world markets, when one agent asks another (perhaps through an interview or an application) for the value of identity(i, j) or ideal(i, j), it is likely that the response will be misreported or misinterpreted in some way. The purpose of this analysis is to quantify the amount of miscommunication that may occur while still ensuring that the set of resultant stable matchings remains the same. We first compute this bound for unique space profiles where $(\mathbb{M}, d_M)$ and $(\mathbb{W}, d_W)$ are *isometric* to computable Euclidean space $(\mathbb{R}_c^k, \|\cdot\|_2)$ (the computable real numbers are defined at the end of Section 2.2). Next, we extend these results to more general families of metric spaces. Lastly, we analyze the limit behavior of these bounds in the context of real-world markets.

We begin by formally defining miscommunication. Intuitively, a miscommunication is much like a sketch: we transform each of the identities and ideals while ensuring that the resultant set of stable matchings is invariant under the transformation. Thus, we use the sketch framework to define miscommunication.

Definition 4.1 (Miscommunication Sketch)**.** Let $(T_M, T_W)_{\mathcal{S}}$ be a sketch of a n-complete space profile $\mathcal{S}$. Let $c = (\text{identity}, \text{ideal})$ be an arbitrary tuple of identity and ideal maps. Let X_c be the transformed instance and O_c be the original instance. If for all tuples c, agents $i \in M \cup W$, and match candidates $j \in \phi(i)$,

$$(1-\epsilon)P_{O_c}(i, j) \leq P_{X_c}(i, j) \leq (1+\epsilon)P_{O_c}(i, j)$$

then we say that $(T_M, T_W)_{\mathcal{S}}$ is a *miscommunication sketch*. We call ϵ the *stretch factor* of the miscommunication sketch.

In the miscommunication sketch, T_M and T_W denote the transformations that offset the agents' identities and ideals. We quantify the *amount* of miscommunication by introducing the stretch factor ϵ, which measures the maximum distortion in the distance between an agent's ideal and a matching candidate's identity.

We first bound the stretch factor ϵ in the case where the unique n-complete space profile $\mathcal{S}$ is isometric to the space profile $\mathcal{S}_R = [M, W, (\mathbb{R}_c^{k_M}, \|\cdot\|_2), (\mathbb{R}_c^{k_W}, \|\cdot\|_2), I(k_M, \mathbf{v}_M), I(k_W, \mathbf{v}_W)]$ (isometry between space profiles is defined in Section 2.3). We generate this bound by showing that if the stretch factor exceeds the bound, we can construct a miscommunication sketch with that stretch factor that contradicts Corollary 2. In other words, if the stretch factor becomes too large, we can devise a representation of the agents' identities and ideals that can be stored in fewer bits than the bound given by Chou and Lu in Theorem 3.2. Thus, the resulting set of stable matchings may not be invariant under the set of transformations.

The sketch transformations that we use to reach the contradiction are given by a variant of the *Johnson-Lindenstrauss transform* (defined in [10]). Building off the JL transform, Matoušek proved the existence of a linear transformation that maps points to a subspace of lower dimension while ensuring that no pairwise distance is distorted beyond a set stretch factor ϵ. We define it below.

Definition 4.2 (JL Transform Variant)**.** Suppose there are n points in $\mathbb{R}^k$. In [11], Matoušek shows the existence of a transformation $T_{JL} : \mathbb{R}^k \mapsto \mathbb{R}^{\frac{C}{\epsilon^2}\log(n)}$ such that for all $\binom{n}{2}$ pairs (i, j) of points,

$$(1-\epsilon)\|i-j\|_2 \leq \|T(i)-T(j)\|_2 \leq (1+\epsilon)\|i-j\|_2$$

We use the transformation T_{JL} to construct a sketch for space profiles $\mathcal{S}$ isometric to $\mathcal{S}_R$.

Definition 4.3 (Isometric JL Sketch)**.** Let $\mathcal{S}$ be isometric to $\mathcal{S}_R$. The *isometric JL sketch*

$(T_M^{IJL}, T_W^{IJL})_{\mathcal{S}}$ is given by

- $T_M^{IJL} = \varphi_M^{-1} \circ T_{JL} \circ \varphi_M$
- $T_W^{IJL} = \varphi_W^{-1} \circ T_{JL} \circ \varphi_W$

where φ_M and φ_W are the bijective maps satisfying the isometry condition.

We use the isometric JL sketch to show our first result.

Theorem 4.1. *Let $\mathcal{S}$ be a space profile isometric to $\mathcal{S}_R$. If pairwise distances are distorted by a miscommunication sketch $(T_M, T_W)_{\mathcal{S}}$ with a stretch factor ϵ, then $\epsilon = O\left(\sqrt{\frac{\log(n)}{n}}\right)$*

Proof. We generate our bound by considering the sketch $(T_M^{IJL}, T_W^{IJL})_{\mathcal{S}}$. By Definition 4.2, $\ell_M = \ell_W = \frac{C}{\epsilon^2}\log(n)$. By Corollary 2, $\ell_M = \Omega(n) = \ell_W$. Thus, we have that $\frac{\log(n)}{\epsilon^2} = \Omega(n)$. The desired result $\epsilon = O\left(\sqrt{\frac{\log(n)}{n}}\right)$ follows. $\square$

Corollary 4. *Let $\mathcal{S}$ be a space profile isometric to $\mathcal{S}_R$. If pairwise distances are distorted by a miscommunication sketch $(T_M, T_W)_{\mathcal{S}}$ with a stretch factor ϵ, then* $\lim_{n\to\infty} \epsilon = 0$

Theorem 4.1 and Corollary 4 again lead to a profound but unfortunate conclusion about real-world markets. Suppose that agents' preferences are determined by computing some analog of Euclidean distance between their ideal match partner and a match candidate's identity. As the number of market participants increase, the extent to which identities and ideals are misreported or misinterpreted must decrease, falling to 0 in the limit case. As this is not a trend that holds true in practice, miscommunication can lead to unstable matchings, even when stable matching mechanisms are used.

We extend these results to the more general case where the n-complete unique space profile $\mathcal{S}$ is *transformable* to $\mathcal{S}_R$ (we define transformability in Section 2.3). Just as we did in the isometric case, we define the *transformable JL sketch*.

Definition 4.4 (Transformable JL Sketch)**.** Let $\mathcal{S}$ be transformable to $\mathcal{S}_R$. The *transformable JL sketch* $(T_M^{TJL}, T_W^{TJL})_{\mathcal{S}}$ is given by

- $T_M^{TJL} = \varphi_M^{-1} \circ T_{JL} \circ \varphi_M$
- $T_W^{TJL} = \varphi_W^{-1} \circ T_{JL} \circ \varphi_W$

where φ_M and φ_W are the isomorphisms between the two spaces.

Theorem 4.2. *Let $\mathcal{S}$ be a space profile transformable to $\mathcal{S}_R$. If pairwise distances are distorted by a miscommunication sketch $(T_M, T_W)_{\mathcal{S}}$ with a stretch factor ϵ, then* $\lim_{n\to\infty} \epsilon = 0$.

Proof. As in the isometric case, we consider the sketch $(T_M^{TJL}, T_W^{TJL})_{\mathcal{S}}$. However, as the distance metrics are now transformed by the τ_M and τ_W maps, the stretch factor given by the JL transform variant is *not* equal to the stretch factor of the sketch $(T_M^{TJL}, T_W^{TJL})_{\mathcal{S}}$. We compute the transformed stretch factor for elements of the metric space $\mathbb{W}$. The argument is symmetric for elements of the metric space $\mathbb{M}$. Let ϵ' be the stretch factor of the JL transform variant. From Definitions 4.1, 4.2 and 4.4, we have that

$$(1-\epsilon')\tau_W^{-1}(P_{O_c}(m,w)) \leq \tau_W^{-1}(P_{X_c}(m,w)) \leq (1+\epsilon')\tau_W^{-1}(P_{O_c}(m,w))$$

which we can rewrite as

$$\tau_W((1-\epsilon')\tau_W^{-1}(P_{O_c}(m,w))) \leq P_{X_c}(m,w) \leq \tau_W((1+\epsilon')\tau_W^{-1}(P_{O_c}(m,w)))$$

Compare this with the definition of the stretch factor from Definition 4.1

$$(1-\epsilon)P_{O_c}(m,w) \leq P_{X_c}(m,w) \leq (1+\epsilon)P_{O_c}(m,w)$$

Setting the left sides of both statements equal, we have that

$$(1-\epsilon)P_{O_c}(m,w) = \tau_W((1-\epsilon')\tau_W^{-1}(P_{O_c}(m,w)))$$

Taking τ_W^{-1} on both sides we get

$$\tau_W^{-1}((1-\epsilon)P_{O_c}(m,w)) = (1-\epsilon')\tau_W^{-1}(P_{O_c}(m,w))$$

which we can rearrange as

$$\epsilon' = 1 - \frac{\tau_W^{-1}((1-\epsilon)P_{O_c}(m,w))}{\tau_W^{-1}(P_{O_c}(m,w))}$$

Invoking Corollary 4, we have that $\lim_{n\to\infty} \epsilon' = 0$. Thus, we have that

$$\frac{\tau_W^{-1}((1-\lim_{n\to\infty}\epsilon)P_{O_c}(m,w))}{\tau_W^{-1}(P_{O_c}(m,w))} = 1$$

by taking the limit of both sides and moving the 1 over to the other side of the equation. The above is only true when $1 - \lim_{n\to\infty}\epsilon = 1$. The desired result $\lim_{n\to\infty}\epsilon = 0$ follows. The argument is symmetric in the case that we set the right hand terms to be equal, or if we considered elements of $\mathbb{M}$ instead of elements of $\mathbb{W}$. □

As transformability is a very general relation between space profiles, Theorem 4.2 motivates the idea that in any practical scenario, shifts in the agents' preference profiles due to miscommunication among the agents can cause match instabilities.

5 Infinite Set Embeddings

In Sections 3 and 4, we dealt exclusively with finite metric spaces. In this section, we extend our discussion to metric spaces with infinite elements. We show that in space profiles with metric spaces of infinite cardinality, we can construct a sketch $(T_M, T_W)_{\mathcal{S}}$ such that $\ell_M = \Theta(1) = \ell_W$. We subsequently show that only $\Theta(n)$ queries are required to find a stable matching in this case.

At the moment, these are still conjectures, since I haven't had the time to finish the proofs. The general idea behind these statements is that the finite Cartesian product of an infinite set is isomorphic to the set itself (e.g. $\mathbb{N}^k \simeq \mathbb{N}$ and $\mathbb{R}^k \simeq \mathbb{R}$), so we can use these isomorphisms to construct the sketch.

6 Acknowledgments

I would like to thank my mentor Prof. Scott Duke Kominers for being an inspiring RSI mentor and bearing with my constant topic changes over the course of the program. I would also like to thank Franklyn Wang for preventing me from veering too far off the path of mathematical validity. Dr. Jenny Sendova gave me Great advice on how to proceed with my research. I would like to thank MIT, CEE, and RSI for making this research possible. Lastly, I would like to thank my sponsors Ms. Rachel Guerra, Ms. Susan Wildstrom, Ms. Cara Leonetti Esposito, and Professors Joseph and Nell Sedransk for so generously making this opportunity possible for me.

References

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

[2] A. E. Roth. A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the united kingdom. *American Economic Review*, pages 415–440, 1991.

[3] C. Ng and D. S. Hirschberg. Lower bounds for the stable marriage problem and its variants. *SIAM Journal on Computing*, 19(1):71–77, 1990.

[4] I. Segal. The communication requirements of social choice rules and supporting budget sets. *Journal of Economic Theory*, 136(1):341–378, 2007.

[5] J.-H. Chou and C.-J. Lu. Communication requirements for stable marriages. In *International Conference on Algorithms and Complexity*, pages 371–382. Springer, 2010.

[6] Y. A. Gonczarowski, N. Nisan, R. Ostrovsky, and W. Rosenbaum. A stable marriage requires communication. In *Proceedings of the twenty-sixth annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1003–1017. SIAM, 2014.

[7] A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[8] B. Kalyanasundaram and G. Schintger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992.

[9] I. Ashlagi, M. Braverman, Y. Kanoria, and P. Shi. Communication requirements and informative signaling in matching markets. Working Paper. 2017.

[10] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics*, 26(189-206):1, 1984.

[11] J. Matoušek. On variants of the johnson–lindenstrauss lemma. *Random Structures & Algorithms*, 33(2):142–156, 2008.