THE GEOMETRY OF RANDOM PAIRED COMPARISONS

Andrew K. Massimino and Mark A. Davenport

School of Electrical and Computer Engineering,
Georgia Institute of Technology, Atlanta, GA

ABSTRACT

Suppose that we are able to obtain binary paired comparisons of the form "$\mathbf{x}$ is closer to $\mathbf{p}$ than to $\mathbf{q}$" for various choices of vectors $\mathbf{p}$ and $\mathbf{q}$. Such observations arise in a variety of contexts, including nonmetric multidimensional scaling, unfolding, and ranking problems, often because they provide a powerful and flexible model of preference. In this paper we give a theoretical bound for how well we can expect to estimate $\mathbf{x}$ under a randomized model for $\mathbf{p}$ and $\mathbf{q}$. We also show that we can achieve significant gains by adaptively changing the distribution for choosing $\mathbf{p}$ and $\mathbf{q}$.

***Index Terms*—** paired comparisons, ideal point models, recommender systems, 1-bit compressive sensing, adaptivity

1. INTRODUCTION

1.1. The localization problem

In this paper we consider the problem of determining the location of a point in Euclidean space based on distance comparisons to a set of known points, where our observations are nonmetric. Let $\mathbf{x} \in \mathbb{R}^n$ be the true position of the point that we are trying to learn, and let $\{(\mathbf{p}_i, \mathbf{q}_i) \in \mathbb{R}^n \times \mathbb{R}^n : i \in [m]\}$ be pairs of known positions. Rather than directly observing the raw distances from $\mathbf{x}$, i.e., $\|\mathbf{x} - \mathbf{p}_i\|$ and $\|\mathbf{x} - \mathbf{q}_i\|$, we instead obtain only paired comparisons of the form $\|\mathbf{x}-\mathbf{p}_i\| < \|\mathbf{x}-\mathbf{q}_i\|$. Our goal is to estimate $\mathbf{x}$ from a set of such inequalities. Nonmetric observations of this type arise in applications of multidimensional scaling and triangulation [1]. These methods are often applied in situations where we have a collection of items and hypothesize that it is possible to embed the items in $\mathbb{R}^n$ in such a way that the Euclidean distance between points corresponds to their "dissimilarity," with small distances corresponding to similar items. Similarity might be derived from human judgements and may be difficult to quantify.

As a motivating example, we consider the problem of estimating a user's preferences from limited response data. This is, for instance, useful in a recommendation system or psychological study. A common and intuitively appealing way to model preference is via the *ideal point model*, which supposes preference for a particular item varies inversely with distance in Euclidean space [2]. We assume that the items to be rated are represented by $\mathbf{p}_i$ and $\mathbf{q}_i$ and a user's preference is modeled as $\mathbf{x}$, called the individual's "ideal point". This represents a hypothetical "perfect" item satisfying all of the user's criteria for evaluating items. Using response data consisting of paired comparisons between items (e.g., "user $\mathbf{x}$ prefers item $\mathbf{p}_i$ to item $\mathbf{q}_i$") is a natural approach when dealing with human subjects since it avoids requiring people to assign precise numerical scores to different items (which is generally a difficult task, especially when multiple factors impact preference [3]). In contrast, human subjects find pairwise judgements much easier to make [4]. Data consisting of paired comparisons is also generated implicitly in contexts where the user has the option to act on two (or more) alternatives; for instance they may choose to watch a particular movie, or click a particular advertisement, out of those displayed to them [5]. In such contexts, the "true distances" in the ideal point model's preference space are generally inaccessible directly, but it is nevertheless possible to obtain an estimate of a user's ideal point.

1.2. Main results

The fundamental question of interest in this paper is how many paired comparisons we need (and how to choose them) to add $\mathbf{x}$ to an existing embedding up to a desired degree of accuracy. The item embedding could be generated using various methods, such as multidimensional scaling applied to a set of item features, or even using the results of previous paired comparisons via an approach like that in [6]. Given an embedding of ℓ items, there are a total of $\binom{\ell}{2} = \Theta(\ell^2)$ possible paired comparisons. In a system with thousands (or more) items, it will be prohibitive to acquire this many comparisons as a typical user will likely only provide comparisons for a handful of items. Fortunately, in general we can expect that many, if not most, of the possible comparisons are redundant.

Any precise answer to this question will depend on the underlying geometry of the item embedding. Each comparison essentially divides $\mathbb{R}^n$ in half, indicating on which side of a hyperplane $\mathbf{x}$ lies, and some arrangements of hyperplanes will yield better tessellations of the preference space than will others. Thus, to gain some intuition on this problem without reference to the geometry of a particular embedding, we instead consider a probabilistic model where the items are

This work was supported by grants NRL N00173-14-2-C001, AFOSR FA9550-14-1-0342, NSF CCF-1350616, CCF-1409406, and CMMI-1537261.
email: {massimino,mdav}@gatech.edu.

generated at random from a particular distribution. In this case we show that under certain natural assumptions on the distribution, it is possible to estimate the location of any $\mathbf{x}$ to within an error of ϵ using a number of comparisons which, up to log factors, is proportional to n/ϵ. This is essentially optimal, so that no set of comparisons can provide a uniform guarantee with significantly fewer comparisons. We then describe a simple extension to an *adaptive* scheme where we actively select the comparisons (manifested here in adaptively altering the mean and variance of the distribution generating the items) to substantially reduce the required number of comparisons.

1.3. Related work

It is important to note that the ideal point model, while similar, is distinct from the low-rank model used in *matrix completion* [7, 8]. Although both models suppose user choices are guided by a number of attributes, the ideal point model leads to preferences that are *non-monotonic* functions of those attributes. The ideal point model suggests that each feature has an ideal level; too much of a feature can be just as undesirable as too little. It is not possible to obtain this kind of performance with a traditional low-rank model, though if points are limited to the sphere, then the ideal point model can duplicate the performance of a low-rank factorization. There is also empirical evidence that the ideal point model captures behavior more accurately than factorization based approaches do [9, 10].

There is a large body of work that studies the problem of learning to rank items from various sources of data, including paired comparisons of the sort we consider in this paper [11, 12, 13]. We first note that in most work on rankings, the central focus is on learning a correct rank-ordered list for a particular user, without providing any guarantees on recovering a correct parameterization for the user's preferences as we do here. While these two problems are related, there are natural settings where it might be desirable to guarantee an accurate recovery of the underlying parameterization ($\mathbf{x}$ in our model). For example, one could exploit these guarantees in the context of an iterative algorithm for nonmetric multidimensional scaling which aims to refine the underlying embedding by updating each user and item one at a time (e.g., see [14]), in which case an understanding of the error in the estimate of $\mathbf{x}$ is crucial. Moreover, we believe that our approach provides an interesting alternative perspective as it yields natural robustness guarantees and suggests simple adaptive schemes.

Also closely related is the work in [15, 16, 17] which consider paired comparisons and more general ordinal measurements in the similar context of low-rank factorizations. Finally, while seemingly unrelated, we note that our work builds on the growing body of literature of 1-bit compressive sensing. In particular, our results are largely inspired by those in [18, 19], and borrow techniques from [20] in the proofs of some of our main results. We emphasize that in our model, both the direction and length of the preference vector are recoverable from the paired comparisons; multiplying a user's point by a scalar will likely cause many comparisons to change.

2. A RANDOMIZED OBSERVATION MODEL

We will consider the "noise-free" setting where a user *always* prefers the item closest to the user's ideal point $\mathbf{x}$ with probability 1. In this case we can represent the observed comparisons mathematically by letting $\mathcal{A}_i(\mathbf{x})$ denote the i^{th} observation, which consists of comparisons between $\mathbf{p}_i$ and $\mathbf{q}_i$, and setting

$$\mathcal{A}_i(\mathbf{x}) := \operatorname{sign}\left(\|\mathbf{x}-\mathbf{q}_i\|^2 - \|\mathbf{x}-\mathbf{p}_i\|^2\right).$$

We will also use $\mathcal{A}(x) := [\mathcal{A}_1(\mathbf{x}), \cdots, \mathcal{A}_m(\mathbf{x})]^T$ to denote the vector of all observations resulting from m comparisons. If we set $\bar{\mathbf{a}}_i := (\mathbf{p}_i - \mathbf{q}_i)$ and $\bar{\tau}_i := \frac{1}{2}(\|\mathbf{p}_i\|^2 - \|\mathbf{q}_i\|^2)$, then we can re-write our observation model as

$$\mathcal{A}_i(\mathbf{x}) = \operatorname{sign}\left(2\bar{\mathbf{a}}_i^T\mathbf{x} - 2\bar{\tau}_i\right) = \operatorname{sign}\left(\bar{\mathbf{a}}_i^T\mathbf{x} - \bar{\tau}_i\right). \quad (1)$$

This is reminiscent of the standard setup in one-bit compressive sensing (with dithers) [18, 19] with the important differences that: *(i)* we do not make any kind of sparsity or other structural assumption on $\mathbf{x}$ and, *(ii)* the "dithers" $\bar{\tau}_i$, at least in this formulation, are dependent on the $\bar{\mathbf{a}}_i$, which results in difficulty applying standard results from this theory to this setting. However, many of the techniques from this literature will be helpful in analyzing this problem. We consider a randomized observation model where the pairs $(\mathbf{p}_i, \mathbf{q}_i)$ are chosen independently with i.i.d. entries drawn according to a normal distribution, i.e., $\mathbf{p}_i, \mathbf{q}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. In this case, we have that the entries of our sensing vectors are i.i.d. with $\bar{\mathbf{a}}_i(j) \sim \mathcal{N}(0, 2\sigma^2)$. Moreover, if we define $\mathbf{b}_i = \mathbf{p}_i + \mathbf{q}_i$, then we also have that $\mathbf{b}_i \sim \mathcal{N}(0, 2\sigma^2\mathbf{I})$, and $\frac{1}{2}\bar{\mathbf{a}}_i^T\mathbf{b}_i = \frac{1}{2}(\|\mathbf{p}_i\|^2 - \|\mathbf{q}_i\|^2) = \bar{\tau}_i$. To simplify, we re-normalize by dividing by $\|\bar{\mathbf{a}}_i\|$, i.e., setting $\mathbf{a}_i := \bar{\mathbf{a}}_i / \|\bar{\mathbf{a}}_i\|$, $\tau_i := \bar{\tau}_i / \|\bar{\mathbf{a}}_i\|$, and

$$\mathcal{A}_i(\mathbf{x}) = \operatorname{sign}\left(\mathbf{a}_i^T\mathbf{x} - \tau_i\right).$$

It is easy to see that $\mathbf{a}_i$ is distributed uniformly on the sphere $\mathbb{S}^{n-1} = \{\mathbf{a} \in \mathbb{R}^n : \|\mathbf{a}\| = 1\}$. Since $\bar{\mathbf{a}}_i$ and $\mathbf{b}_i$ are independent, $\mathbf{a}_i$ and $\mathbf{b}_i$ are also independent. Moreover, for any unit-vector $\mathbf{a}_i$, if $\mathbf{b}_i \sim \mathcal{N}(0, 2\sigma^2\mathbf{I})$ then $\mathbf{a}_i^T\mathbf{b}_i \sim \mathcal{N}(0, 2\sigma^2)$. Thus, we must have $\tau_i \sim \mathcal{N}(0, \sigma^2/2)$, independent of $\mathbf{a}_i$, which is the key insight that enables the analysis below.

3. GUARANTEES UNDER THE RANDOM MODEL

We now state our main result concerning localization under the noise-free random model from Section 2.

Theorem 1 (Performance with Gaussian items)**.** *Suppose m item point pairs $\{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^m$ are generated by drawing each $\mathbf{p}_i$ and $\mathbf{q}_i$ independently from $\mathcal{N}(0, \sigma^2 I)$ where $\sigma^2 = 2R^2/n$. There exist constants c and C such that if $\epsilon > 0$, $\eta > 0$, and*

$$m \geq \frac{R}{C\epsilon(1 - c/\sqrt{n})}\left(n \log\frac{3R\sqrt{n}}{c\epsilon} + \log\frac{1}{\eta}\right), \quad (2)$$

then with probability at least $1-\eta$*, for all pairs of signals* $\mathbf{x},\mathbf{y} \in B_R := \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\| \le R\}$ *such that* $\mathcal{A}(\mathbf{x}) = \mathcal{A}(\mathbf{y})$*,*

$$\|\mathbf{x}-\mathbf{y}\| \le \epsilon.$$

The key message of this theorem is that if one chooses the variance σ^2 of the distribution generating the items appropriately, then it is possible to estimate $\mathbf{x}$ to within ϵ using a number of comparisons that is nearly linear in n/ϵ. A natural question is what would happen with a different choice of σ^2. In fact, this assumption is critical—if σ^2 is substantially smaller the bound quickly becomes vacuous, and as σ^2 grows much past $2R^2/n$ the bound begins to become steadily worse. It should also be somewhat intuitive: if σ^2 is too small, then nearly all the hyperplanes induced by the comparisons will pass very close to the origin, so that accurate estimation of even $\|\mathbf{x}\|$ becomes impossible. On the other hand, if σ^2 is too large, then an increasing number of these hyperplanes will not even intersect the ball of radius R in which $\mathbf{x}$ is presumed to lie, thus yielding no new information.

Lemma 2 ([21])**.** *Let* $\mathbf{w},\mathbf{z} \in \mathbb{B}_R^n$ *be distinct and non-zero. Fix* $\delta > 0$ *and let* $B_\delta(\mathbf{w})$ *be the points with Euclidean distance at most* δ *from* $\mathbf{w}$*:* $B_\delta(\mathbf{w}) := \{\mathbf{u} \in B_R : \|\mathbf{u}-\mathbf{w}\| \le \delta\}$*. Denote by* P_1*,*

$$P_1 := \mathbb{P}\{\forall \mathbf{u} \in B_\delta(\mathbf{w}), \forall \mathbf{v} \in B_\delta(\mathbf{z}) : \mathcal{A}_i(\mathbf{u}) \ne \mathcal{A}_i(\mathbf{v})\},$$

the probability that all points $\mathbf{u}$ *and* $\mathbf{v}$*, which are within* δ *of* $\mathbf{w}$ *and* $\mathbf{z}$ *respectively, differ by a random observation denoted by* $\mathcal{A}_i$ *(i.e., the two* δ*-balls are separated by hyperplane* i*). Set* $\epsilon_0 \le \|\mathbf{w}-\mathbf{z}\|$*. Then,*

$$P_1 \ge \frac{\epsilon_0 - 8\delta\sqrt{2n}}{16\sqrt{12\pi e}R}.$$

Proof of Theorem 1. Let U be a δ-covering set for B_R with $|U| \le (3R/\delta)^n$. For any pair $\mathbf{x},\mathbf{y} \in B_R$ there exist $\mathbf{w},\mathbf{z} \in U$ such that $\|\mathbf{x}-\mathbf{w}\| \le \delta$ and $\|\mathbf{y}-\mathbf{z}\| \le \delta$. Note that $\|\mathbf{w}-\mathbf{z}\| \ge \|\mathbf{x}-\mathbf{y}\| - 2\delta$. Assume m satisfies (2) as in Theorem 1. Suppose that $\|\mathbf{x}-\mathbf{y}\| > \epsilon$ for some $\epsilon > 0$. Let P_1 be defined as in Lemma 2; setting $\epsilon_0 = \|\mathbf{w}-\mathbf{z}\|$, it follows that $P_1 \ge (C_1\epsilon_0 + C_2\delta\sqrt{n})/R$. We have

$$P_1 \ge (C_1(\|\mathbf{x}-\mathbf{y}\| - 2\delta) + C_2\delta\sqrt{n})/R.$$

Let P_1' be the conditioning of P_1 on vectors $\mathbf{x},\mathbf{y}$ farther than ϵ apart. We consider the opposite of this event and set $\delta = c_0\epsilon/\sqrt{n}$ for some $c_0 > 0$,

$$1 - P_1' \le 1 - (C_1(\epsilon - 2c_0\epsilon/\sqrt{n}) + C_2c_0\epsilon_0)/R,$$

This controls the probability that a single comparison fails to distinguish points $\mathbf{x}$ and $\mathbf{y}$. Let P_m be the probability that m such comparisons are identical. Since the comparisons are independent, we extend the previous inequality over m events,

$$P_m \le \left(1 - ((C_1 + C_2c_0)\epsilon - 2C_1c_0\epsilon/\sqrt{n})/R\right)^m.$$

Now by a union bound over pairs $(\mathbf{w},\mathbf{z})$ in the covering set with $|U \times U| \le (3R/\delta)^{2n}$,

$$\begin{aligned} P_t &\le \left(\frac{3R}{\delta}\right)^{2n} \left(1 - C\epsilon(1-c/\sqrt{n})/R\right)^m \\ &\le \exp\left(2n\log\frac{3R}{\delta}\right)\exp\left(-C\epsilon m(1-c/\sqrt{n})/R\right) \\ &= \exp\left(2n\log\frac{3R\sqrt{n}}{c\epsilon} - C\epsilon m(1-c/\sqrt{n})/R\right). \end{aligned}$$

Upper bounding this by η,

$$\begin{aligned} &n\log\frac{3R\sqrt{n}}{c\epsilon} - C\epsilon m(1-c/\sqrt{n})/R \le \log\eta \\ &\implies m \ge \frac{R}{C\epsilon(1-c/\sqrt{n})}\left(n\log\frac{3R\sqrt{n}}{c\epsilon} + \log\frac{1}{\eta}\right) \end{aligned}$$

C and c are constants that are linked by Lemma 2 but do not depend on n. □

4. ADAPTIVE LOCALIZATION

Here we describe a simple extension to our previous (noiseless) theory and show that if we modify the mean and variance of the sampling distribution of items over a number of stages, we can localize *adaptively* and produce an estimate with many fewer comparisons than possible in a non-adaptive strategy. We assume t stages ($t = 1$ for the non-adaptive approach). At each stage $\ell \in [t]$ we will attempt to produce an estimate $\widehat{\mathbf{x}}^\ell$ such that $\|\mathbf{x}-\widehat{\mathbf{x}}^\ell\| \le \epsilon_\ell$ where $\epsilon_\ell = R_\ell/2 = R2^{-\ell}$, then recentering to our previous estimate and dividing the problem radius in half. In stage ℓ, each $\mathbf{p}_i,\mathbf{q}_i \sim \mathcal{N}(\widehat{\mathbf{x}}, 2R_\ell^2/n\mathbf{I})$. After t stages we will have $\|\mathbf{x}-\widehat{\mathbf{x}}^t\| \le R2^{-t} =: e_t$ with probability at least $1-t\eta$.

Theorem 3. *Let* $\mathbf{x} \in \mathbb{R}^n$*,* $\|\mathbf{x}\| \le R$ *and* $\eta > 0$*. There are constants* c *and* C *such that if* $\epsilon_t > 0$ *is the target final accuracy and* m *total comparisons are taken following the adaptive scheme where*

$$m \ge \frac{2\log_2(2R/\epsilon_t)}{C(1-c/\sqrt{n})}\left(n\log\frac{6\sqrt{n}}{c} + \log\frac{1}{\eta}\right),$$

then with probability at least $1-\log_2(2R/\epsilon_t)\eta$*, for any estimate* $\widehat{\mathbf{x}}$ *satisfying* $\mathcal{A}(\widehat{\mathbf{x}}) = \mathcal{A}(\mathbf{x})$*,*

$$\|\mathbf{x}-\widehat{\mathbf{x}}\| \le \epsilon_t.$$

Proof. The adaptive scheme uses $t = \lceil\log_2(R/\epsilon_t)\rceil \le \log_2(2R/\epsilon_t)$ stages. Assume each stage is allocated m_ℓ comparisons. By Theorem 1, localization at each stage ℓ can be accomplished with high probability when

$$\begin{aligned} m_\ell &\ge \frac{R_\ell}{C\epsilon_\ell(1-c/\sqrt{n})}\left(n\log\frac{3R_\ell\sqrt{n}}{c\epsilon_\ell} + \log\frac{1}{\eta}\right) \\ &= \frac{2}{C(1-c/\sqrt{n})}\left(n\log\frac{6\sqrt{n}}{c} + \log\frac{1}{\eta}\right). \end{aligned}$$

This condition is met by giving an equal number of comparisons to each stage, $m_\ell = \lfloor m/t \rfloor$. Each stage fails with probability η. By a union bond, the target localization fails with probability at most $t\eta$. Hence, localization succeeds with probability at least $1 - t\eta$. □

Theorem 3 implies $m_{\text{adapt}} \asymp (n \log n) \log_2(R/\epsilon_t)$ comparisons suffice to estimate $\mathbf{x}$ to within ϵ_t. This represents an exponential improvement in terms of number of total comparisons as a function of the target accuracy, ϵ_t, as compared to a lower bound on the number of required comparisons, $m_{\text{lower}} := 2nR/(e\epsilon_t)$, which can be shown to hold for *any* non-adaptive strategy through a simple volumetric argument.

5. SIMULATIONS

Given a set of comparisons $\mathcal{A}(\mathbf{x})$, we may produce an estimate $\widehat{\mathbf{x}}$ by finding *any* feasible point satisfying all the paired comparisons. A simple approach is the following convex program:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{z}} \|\mathbf{z}\|^2 \text{ subject to } \mathcal{A}_i(\mathbf{x})(\mathbf{a}_i^T\mathbf{z} - \tau_i) \geq 0 \ \forall i.$$

This is easy to solve since the constraints are simple linear inequalities and the feasible region is convex. We do not consider noise in these simulations, but in situations where comparison inconsistencies may exist, this optimization program could be made more robust by introducing slack variables [1].

5.1. Adaptive generation

In Fig. 1, we show the effect of varying levels of adaptivity, starting with the completely non-adaptive approach up to using 10 stages where we progressively re-center and re-scale the hyperplane offsets. In each case, we generate $x \in \mathbb{R}^3$ where $\|\mathbf{x}\| = 0.75$ and choosing the direction randomly. The total comparisons are held fixed and are split as equally as possible among the number of stages (preferring earlier stages when rounding). We set $\sigma^2 = R = 1$ and plot the average over 700 independent trials. As the number of stages increases, performance worsens if the number of comparisons are kept small due to bad localization in the earlier stages. However, if the number of total comparisons is sufficiently large, an exponential improvement over non-adaptivity is possible.

5.2. Adaptive selection with a fixed non-Gaussian dataset

In Fig. 2, we demonstrate the effect of adaptively choosing item pairs from a fixed synthetic dataset over four stages versus choosing items non-adaptively.We first generated 10,000 items uniformly distributed inside the 3-dimensional unit ball and a signal $\mathbf{x} \in \mathbb{R}^3$ where $\|\mathbf{x}\| = 0.4$. In both cases, we generate pairs of Gaussian points and choose the items from the fixed dataset which lie closest to them. In the adaptive case over four stages, we progressively re-center and re-scale the generated points; the initial σ^2 is set to the variance of the dataset and

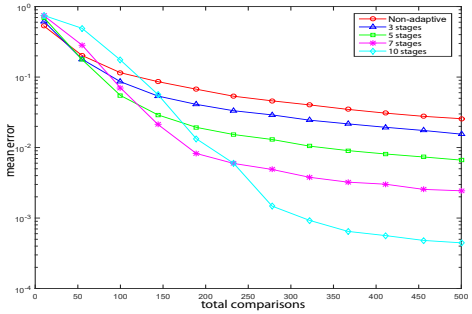

Fig. 1: Error norm $\|\mathbf{x} - \widehat{\mathbf{x}}\|$ versus total comparisons for a sequence of experiments with varying number of adaptive stages.

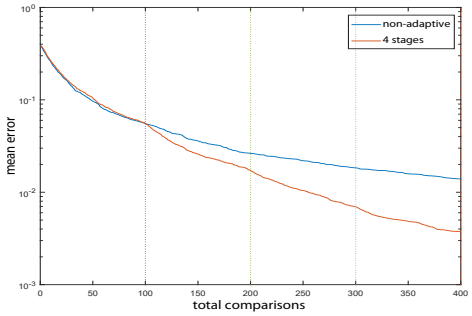

Fig. 2: Mean error norm $\|\mathbf{x} - \widehat{\mathbf{x}}\|$ versus total comparisons for non-adaptive and adaptive selection. Dotted lines denote stage boundaries.

is reduced dyadically after each stage. The total number of comparisons is held fixed and is split as equally as possible among the number of stages (preferring later stages when rounding). We plot the mean error over 200 independent-dataset trials.

6. DISCUSSION

We have shown that given the ability to generate item pairs according to a Gaussian distribution with a particular variance, it is possible to localize a point $\mathbf{x}$ to within ϵ with roughly n/ϵ paired comparisons (ignoring log factors). If one is able to shift the distribution of the items drawn, adaptive localization gives a substantial improvement over a non-adaptive strategy. To directly implement such a scheme, one would require the ability to generate items arbitrarily in $\mathbb{R}^n$. While there may be some cases where this is possible (e.g., in market testing of items where the features correspond to known quantities that can be manually manipulated, such as the amount of various ingredients in a beverage) in most settings considered by recommendation systems the only items which can be compared belong to a fixed set of points. While our theory would still provide rough guidance as to how accurate of a localization is possible, the algorithm must be adapted, as done in Section 5.2. There are many other ways that the adaptive scheme could be modified to account for this restriction. We leave the exploration of additional techniques for future work.

7. REFERENCES

[1] M. Davenport, "Lost without a compass: Nonmetric triangulation and landmark multidimensional scaling," in *Proc. IEEE Int. Work. Comput. Adv. Multi-Sensor Adaptive Processing (CAMSAP)*, Saint Martin, Dec 2013.

[2] C. Coombs, "Psychological scaling without a unit of measurement.," *Psych. Rev.*, vol. 57, no. 3, pp. 145–158, 1950.

[3] G. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information.," *Psych. review*, vol. 63, no. 2, pp. 81, 1956.

[4] H. David, *The method of paired comparisons*, London, UK: Charles Griffin & Company Limited, 1963.

[5] F. Radlinski and T. Joachims, "Active exploration for learning rankings from clickthrough data," in *Proc. ACM SIGKDD Int. Conf. on Knowledge, Discovery, and Data Mining (KDD)*, San Jose, CA, August 2007, ACM.

[6] S. Agarwal, J. Wills, L. Cayton, G. Lanckriet, D. Kriegman, and S. Belongie, "Generalized non-metric multidimensional scaling," in *Proc. Int. Conf. Art. Intell. Stat. (AISTATS)*, San Juan, Puerto Rico, Mar. 2007.

[7] J. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proc. Int. Conf. Machine Learning (ICML)*, Bonn, Germany, Aug. 2005.

[8] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[9] B. Dubois, "Ideal point versus attribute models of brand preference: a comparison of predictive validity.," *Adv. Consumer Research*, vol. 2, no. 1, pp. 321–333, 1975.

[10] A. Maydeu-Olivares and U. Böckenholt, "Modeling preference data," *The SAGE handbook of quantitative methods in psychology*, pp. 264–282, 2009.

[11] K. Jamieson and R. Nowak, "Active ranking using pairwise comparisons," in *Proc. Adv. in Neural Processing Systems (NIPS)*, Granada, Spain, Dec. 2011.

[12] K. Jamieson and R. Nowak, "Low-dimensional embedding using adaptively selected ordinal data," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, 2011.

[13] F. Wauthier, M. Jordan, and N. Jojic, "Efficient ranking from pairwise comparisons," in *Proc. Int. Conf. Machine Learning (ICML)*, Atlanta, GA, June 2013.

[14] M. O'Shaughnessy and M. Davenport, "Localizing users and items from paired comparisons," in *Proc. IEEE Int. Work. on Machine Learning for Signal Processing (MLSP)*, Vietri sul Mare, Salerno, Italy, 2016, to appear.

[15] Y. Lu and S. Negahban, "Individualized rank aggregation using nuclear norm regularization," *arXiv:1410.0860*, 2014.

[16] D. Park, J. Neeman, J. Zhang, S. Sanghavi, and I. Dhillon, "Preference completion: Large-scale collaborative ranking form pairwise comparisons," in *Proc. Int. Conf. Machine Learning (ICML)*, Lille, France, July 2015.

[17] S. Oh, K. Thekumparampil, and J. Xu, "Collaboratively learning preferences from ordinal data," in *Proc. Adv. in Neural Processing Systems (NIPS)*, Montréal, Québec, Dec. 2014.

[18] K. Knudson, R. Saab, and R. Ward, "One-bit compressive sensing with norm estimation," *IEEE Trans. Inform. Theory*, vol. 62, no. 5, pp. 2748–2758, 2016.

[19] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, "Exponential decay of reconstruction error from binary measurements of sparse signals," *arXiv:1407.8246*, 2014.

[20] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inform. Theory*, vol. 59, no. 4, April 2013.

[21] A. Massimino and M. Davenport, "Binary stable embedding via paired comparisons," in *Proc. IEEE Work. on Statistical Signal Processing (SSP)*, Palma de Mallorca, Spain, 2016.