BINARY STABLE EMBEDDING VIA PAIRED COMPARISONS

Andrew K. Massimino and Mark A. Davenport

Georgia Institute of Technology,
School of Electrical and Computer Engineering
{massimino,mdav}@gatech.edu

ABSTRACT

Suppose that we wish to estimate a vector $\mathbf{x}$ from a set of 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 $\mathbf{q}$. The problem of estimating $\mathbf{x}$ from this type of observation arises in a variety of contexts, including nonmetric multidimensional scaling, "unfolding," and ranking problems, often because it provides a powerful and flexible model of preference. The main contribution of this paper is to show that under a randomized model for $\mathbf{p}$ and $\mathbf{q}$, a suitable number of binary paired comparisons yield a stable embedding of the space of target vectors.

Index Terms— paired comparisons, binary stable embeddings, recommendation systems, 1-bit compressive sensing

1. INTRODUCTION

The central problem we consider in this paper is the estimation of a vector $\mathbf{x} \in \mathbb{R}^n$ where, rather than directly observing $\mathbf{x}$, we assume that we are restricted to m observations of the form "$\mathbf{x}$ is closer to $\mathbf{p}_i$ than $\mathbf{q}_i$," where $\mathbf{p}_i, \mathbf{q}_i \in \mathbb{R}^n$ correspond to points whose locations are known. Each of these comparisons essentially divides $\mathbb{R}^n$ in half and tells us which side of a hyperplane the point $\mathbf{x}$ lies. Observations of this form arise in a variety of contexts, but a particularly important class of applications involve recommendation systems, targeted advertisement, and psychological studies where $\mathbf{x}$ represents an *ideal point* that models a particular user's preferences and the $\mathbf{p}_i$ and $\mathbf{q}_i$ represent items that the user compares [1]. Items which are close to $\mathbf{x}$ are those most preferred by the user. Paired comparisons arise naturally in this context since precise numerical scores quantifying a user's preference are generally much more difficult to assign than comparative judgements [2, 3]. Moreover, data consisting of paired comparisons is often 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 [4]. In such contexts, the "true distances" in the ideal point model's preference space are generally inaccessible in any direct way, but it is nevertheless still possible to obtain a good estimate of a user's ideal point.

The fundamental question which interests us in this paper is how many comparisons suffice in order to guarantee that the number of differing paired comparisons generated by $\mathbf{x}$ and $\mathbf{y}$ is roughly proportional to the Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$. This allows us to understand how well $\mathbf{x}$ can potentially be estimated from highly quantized information and gives an idea of the stability in the presence of labeling errors.

We consider the case where we are given an existing embedding of the items (as in a mature recommender system) and focus on the on-line problem of locating a single new user from their feedback (consisting of binary data generated from paired comparisons). The item embedding could be generated using a variety of 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 [5]. We wish to understand how many comparisons are then required to accurately localize a user's ideal point. Any precise answer to this question would depend on the underlying geometry of the item embedding (since some sets of hyperplanes will yield better tesselations of the preference space than others). Thus, to gain some intuition on this problem without reference to the geometry of a particular embedding, we will instead consider a probabilistic model where the items are generated at random from a particular distribution.

It is important to note that the ideal point model, while similar, is distinct from the low-rank factor or attribute model used in the *matrix completion* approaches which have recently gained much attention as applied to recommendation systems, e.g., [6, 7]. 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. There is also empirical evidence that the ideal point model captures user behavior more accurately than factorization based approaches do [8, 9].

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. See, for example, [10, 11, 12] and references therein. 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. Perhaps most

This work was supported by grants NRL N00173-14-2-C001, AFOSR FA9550-14-1-0342, NSF CCF-1350616, CCF-1409406, and CMMI-1537261.

closely related to our work is that of [10, 11], which examines the problem of learning a rank ordering using the same ideal point model considered in this paper. The message in this work is broadly consistent with ours, in that the number of comparisons required should scale with the dimension of the preference space (not the total number of items). Also closely related is the work in [13, 14, 15] which consider paired comparisons and more general ordinal measurements in the similar (but as discussed above, subtly different) context of low-rank factorizations. Finally, while seemingly unrelated, we note that our work builds on the growing body of literature on 1-bit compressive sensing. In particular, our results are largely inspired by those in [16, 17], and borrow techniques from [18] in the proofs of some of our main results.

2. THE RANDOM OBSERVATION MODEL

For the moment we will consider the "noise-free" setting where a user *always* prefers the item closest to the user's ideal point $\mathbf{x}$. 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

$$\begin{aligned}\mathcal{A}_i(\mathbf{x}) &:= \operatorname{sign}\left(\|\mathbf{x}-\mathbf{q}_i\|^2 - \|\mathbf{x}-\mathbf{p}_i\|^2\right) \\ &= \begin{cases} +1 & \text{if } \mathbf{x} \text{ is closer to } \mathbf{p}_i \\ -1 & \text{if } \mathbf{x} \text{ is closer to } \mathbf{q}_i. \end{cases}\end{aligned}$$

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

$$\mathcal{A}_i(\mathbf{x}) = \operatorname{sign}\left(2\mathbf{a}_i^T\mathbf{x} - 2\tau_i\right) = \operatorname{sign}\left(\mathbf{a}_i^T\mathbf{x} - \tau_i\right). \tag{2.1}$$

This is reminiscent of the one-bit compressive sensing setting (with dithers) [16, 17] with the important differences that: *(i)* we have not yet made any kind of sparsity or other structural assumption on $\mathbf{x}$ and, *(ii)* the "dithers" τ_i, at least in this formulation, are dependent on the $\mathbf{a}_i$, which results in difficulty applying existing results from this theory to this setting.

However, many of the techniques from this literature will nevertheless be helpful in analyzing this problem. To see this, 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., $\mathbf{p}_i, \mathbf{q}_i \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. In this case, we have that the entries of our sensing vectors are i.i.d. with $a_{ij} \sim \mathcal{N}(0, 2\sigma^2)$. 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 we can write $\tau_i = \frac{1}{2}\mathbf{a}_i^T\mathbf{b}_i$. Note that while τ_i is clearly dependent on $\mathbf{a}_i$, we do have that $\mathbf{a}_i$ and $\mathbf{b}_i$ are independent. To further simplify things, let's re-normalize by dividing by $\|\mathbf{a}_i\|$, i.e., setting $\widetilde{\mathbf{a}}_i = \mathbf{a}_i / \|\mathbf{a}_i\|$ and $\widetilde{\tau}_i = \tau_i / \|\mathbf{a}_i\|$ and

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

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

3. MAIN RESULT

Under the model described above, we would like to understand how much the sign patterns of two different signals can differ from their Euclidean distance. Our main result is that given enough comparisons there is an approximate embedding of the preference space into $\{-1, 1\}^m$ via our measurement model. We assume any signal of interest has norm at most R, i.e., we say signals $\mathbf{x}, \mathbf{y} \in \mathbb{B}_R^n$, the n-ball of radius R. As an example application, $\mathbf{x}$ might be considered the "true" user's ideal point and $\mathbf{y}$ some other possible estimate. Theorem 3.1 states that if $\mathbf{x}$ and $\mathbf{y}$ are sufficiently nearby, the respective sign-measurement patterns $\mathcal{A}(\mathbf{x})$ and $\mathcal{A}(\mathbf{y})$ closely match. We denote by d_H the Hamming distance, i.e., d_H counts the number of $\{-1, +1\}$ sign measurement errors;

$$d_H(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) := \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}|\mathcal{A}_i(\mathbf{x}) - \mathcal{A}_i(\mathbf{y})|. \tag{3.1}$$

Note that the following result applies *for all* $\mathbf{x}$ and $\mathbf{y}$, and follows from Lemma 3.2, which in contrast applies only for fixed $\mathbf{x}$ and $\mathbf{y}$.

Theorem 3.1. *Let $\eta > 0$ and $\zeta > 0$. Suppose m total measurements of the form (2.1) may be taken, i.e., each $\mathbf{a}_i$ is drawn uniformly on the unit sphere and $\tau \in \mathbb{R}^m$ is a vector with entries $\tau_i \sim \mathcal{N}(0, \sigma^2/2)$ independent of the $\mathbf{a}_i$, where $\sigma^2 = 2R/n$. If*

$$m \geq \frac{1}{2\zeta^2}\left(2n\log\left(\frac{3\sqrt{n}}{\zeta}\right) + \log\left(\frac{2}{\eta}\right)\right).$$

Then there are constants c_1, c_2, c_3, such that with probability at least $1-\eta$, for all points $\mathbf{x}, \mathbf{y} \in \mathbb{B}_R^n$,

$$\begin{aligned}&\frac{c_1\|\mathbf{x}-\mathbf{y}\|}{R} - \zeta\left(\frac{2c_1}{\sqrt{n}} + c_2 + 1\right) \\ &\leq d_H(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) \\ &\leq \frac{c_3\|\mathbf{x}-\mathbf{y}\|}{R} + \zeta\left(\frac{2c_3}{\sqrt{n}} + \sqrt{\frac{2}{\pi}} + 1\right).\end{aligned}$$

Proof. By Lemma 3.2, for any fixed pair $\mathbf{w}, \mathbf{z} \in \mathbb{B}_R^n$ we have bounds on the Hamming distance with probability at least $1 - 2\exp(-2\zeta^2 m)$, for all $\mathbf{u} \in B_\delta(\mathbf{w})$ and $\mathbf{v} \in B_\delta(\mathbf{z})$. Since the radius R ball can be covered with a set U of radius δ balls with $|U| \leq (3R/\delta)^n$, by a union bound we have with probability at least $1-2(3R/\delta)^{2n}\exp(-2\zeta^2 m)$, for *all* $\mathbf{w}, \mathbf{z} \in U$, for all $\mathbf{u} \in B_\delta(\mathbf{w})$ and $\mathbf{v} \in B_\delta(\mathbf{z})$,

$$\frac{c_1 \|\mathbf{w}-\mathbf{z}\|}{R} - \frac{c_2\delta\sqrt{n}}{R} - \zeta \le d_H(\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v})) \le \frac{c_3 \|\mathbf{w}-\mathbf{z}\|}{R} + \frac{\delta}{R}\sqrt{\frac{2n}{\pi}} + \zeta.$$

Since $\|\mathbf{x}-\mathbf{y}\| - 2\delta \le \|\mathbf{w}-\mathbf{z}\| \le \|\mathbf{x}-\mathbf{y}\| + 2\delta$,

$$\begin{aligned}&\frac{c_1 \|\mathbf{x}-\mathbf{y}\|}{R} - \frac{2c_1\delta}{R} - \frac{c_2\delta\sqrt{n}}{R} - \zeta \\ &\le d_H(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) \\ &\le \frac{c_3 \|\mathbf{x}-\mathbf{y}\|}{R} + \frac{2c_3\delta}{R} + \frac{\delta}{R}\sqrt{\frac{2n}{\pi}} + \zeta.\end{aligned}$$

Letting $\delta = \zeta R/\sqrt{n}$ this becomes

$$\begin{aligned}&\frac{c_1 \|\mathbf{x}-\mathbf{y}\|}{R} - \zeta\left(\frac{2c_1}{\sqrt{n}} + c_2 + 1\right) \\ &\le d_H(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) \\ &\le \frac{c_3 \|\mathbf{x}-\mathbf{y}\|}{R} + \zeta\left(\frac{2c_3}{\sqrt{n}} + \sqrt{\frac{2}{\pi}} + 1\right).\end{aligned}$$

Lower bounding the probability by $1-\eta$,

$$2(3R/\delta)^{2n}\exp(-2\zeta^2 m) \le \eta.$$

Rearranging, we have the desired result. □

We comment that Theorem 3.1 concerns a particular choice of the variance parameter σ^2. A natural question is what would happen with a different choice of σ^2. In fact, this assumption is critical—intuitively, 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 3.2. *Let $\mathbf{w}, \mathbf{z} \in \mathbb{B}^n_R$ be fixed. Let m measurements of the form (2.1) be taken in which each $\mathbf{a}_i$ is drawn uniformly on the unit sphere and let $\tau \in \mathbb{R}^m$ be a vector with entries $\tau_i \sim \mathcal{N}(0, \sigma^2/2)$ independent of the $\mathbf{a}_i$. Fix $\zeta > 0$, $\delta > 0$, and define*

$$B_\delta(\mathbf{w}) := \{\mathbf{u} \in \mathbb{B}^n_R : \|\mathbf{u}-\mathbf{w}\| \le \delta\}.$$

Then there are constants c_1, c_2, and c_3 such that with probability at least $1-\exp(-2\zeta^2 m)$, for all $\mathbf{u} \in B_\delta(\mathbf{w})$ and $\mathbf{v} \in B_\delta(\mathbf{z})$,

$$\frac{c_1 \|\mathbf{w}-\mathbf{z}\|}{R} - \frac{c_2\delta\sqrt{n}}{R} - \zeta \le d_H(\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v})) \le \frac{c_3 \|\mathbf{w}-\mathbf{z}\|}{R} + \frac{\delta}{R}\sqrt{\frac{2n}{\pi}} + \zeta,$$

where d_H is the Hamming distance (3.1).

Proof. Fix $\delta > 0$ and let $\mathbf{u} \in B_\delta(\mathbf{w}), \mathbf{v} \in B_\delta(\mathbf{z})$. Recall that the Hamming distance d_H is a sum of independent identically distributed Bernoulli random variables and we may bound it using Hoeffding's inequality. Since our probabilistic upper and lower bounds must hold for all $\mathbf{u}, \mathbf{v}$ as described above, we introduce quantities L_0 and L_1 which represent two "extreme cases" of the Bernoulli variables:

$$\begin{aligned}L_0 &:= \sup_{\mathbf{u}\in B_\delta(\mathbf{w}),\mathbf{v}\in B_\delta(\mathbf{z})} \frac{1}{2m}\sum_{i=1}^m |\mathcal{A}_i(\mathbf{u}) - \mathcal{A}_i(\mathbf{v})| \\ L_1 &:= \inf_{\mathbf{u}\in B_\delta(\mathbf{w}),\mathbf{v}\in B_\delta(\mathbf{z})} \frac{1}{2m}\sum_{i=1}^m |\mathcal{A}_i(\mathbf{u}) - \mathcal{A}_i(\mathbf{v})|.\end{aligned}$$

Then we have

$$L_1 \le d_H(\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v})) \le L_0$$

Denote $P_0 = 1 - \mathbb{E}\, L_0$ and $P_1 = \mathbb{E}\, L_1$, i.e.,

$$\begin{aligned}P_0 &= \mathbb{P}\{\forall \mathbf{u} \in B_\delta(\mathbf{w}), \forall \mathbf{v} \in B_\delta(\mathbf{z}) : \mathcal{A}_i(\mathbf{u}) = \mathcal{A}_i(\mathbf{v})\} \\ P_1 &= \mathbb{P}\{\forall \mathbf{u} \in B_\delta(\mathbf{w}), \forall \mathbf{v} \in B_\delta(\mathbf{z}) : \mathcal{A}_i(\mathbf{u}) \ne \mathcal{A}_i(\mathbf{v})\}.\end{aligned}$$

We give a lower bound for P_0 in Lemma 3.3 and for P_1 in Lemma 3.4. Now by Hoeffding's inequality,

$$\begin{aligned}\mathbb{P}\{L_0 > (1-P_0) + \zeta\} &\le \exp(-2m\zeta^2) \\ \mathbb{P}\{L_1 < P_1 - \zeta\} &\le \exp(-2m\zeta^2).\end{aligned}$$

Hence, with probability at least $1 - 2\exp(-2m\zeta^2)$,

$$P_1 - \zeta \le d_H(A(\mathbf{u}), A(\mathbf{v})) \le (1-P_0) + \zeta.$$

Since

$$P_1 \ge \frac{\|\mathbf{w}-\mathbf{z}\| - 2\delta\sqrt{n}}{16e\sqrt{12\pi}R},$$

and recalling $\sigma^2 = 2R^2/n$,

$$\begin{aligned}1 - P_0 &\le \frac{2}{\sigma\pi}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})}\|\mathbf{w}-\mathbf{z}\| + \frac{2\delta}{\sigma\sqrt{\pi}} \\ &\le \frac{\sqrt{2n}}{R\pi}\sqrt{\frac{2\pi}{n}}\|\mathbf{w}-\mathbf{z}\| + \frac{2\delta\sqrt{n}}{R\sqrt{2\pi}} \\ &= \frac{2\|\mathbf{w}-\mathbf{z}\|}{R\sqrt{\pi}} + \frac{\delta}{R}\sqrt{\frac{2n}{\pi}},\end{aligned}$$

the lemma is satisfied with appropriate c_1, c_2, c_3. □

3.1. Probability estimates

Lemma 3.3. *Let $\mathbf{w}, \mathbf{z} \in \mathbb{B}^n_R$ be distinct. Fix $\delta > 0$. Denote by P_0 the probability that all points $\mathbf{u}$ and $\mathbf{v}$ that are within δ of $\mathbf{w}$ and $\mathbf{z}$ respectively do not differ in the random measurement denoted by $\mathcal{A}_i$ (i.e., the two δ-balls are separated by hyperplane i). The direction and threshold of hyperplane i are denoted by $\mathbf{a}$ and τ respectively. That is,*

$$P_0 = \mathbb{P}\{\forall \mathbf{u} \in B_\delta(\mathbf{w}), \forall \mathbf{v} \in B_\delta(\mathbf{z}) : \mathcal{A}_i(\mathbf{u}) = \mathcal{A}_i(\mathbf{v})\}.$$

Then

$$1 - P_0 \le \frac{2}{\sigma\pi} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} \|\mathbf{w} - \mathbf{z}\| + \frac{2\delta}{\sigma\sqrt{\pi}}.$$

Proof. This result will require the following integral;

$$\int_{\mathbb{S}^{n-1}} |\mathbf{a}^T(\mathbf{w} - \mathbf{z})| \, \nu(\mathrm{d}\mathbf{a}) = \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} \|\mathbf{w} - \mathbf{z}\|,$$

where ν is the uniform probability measure on the unit sphere. We need a lower bound on P_0 which is equivalent to an *upper* bound on

$$1 - P_0 = \mathbb{P}\{\mathcal{A}_i(\mathbf{u}) \neq \mathcal{A}_i(\mathbf{v}) \text{ for some } \mathbf{u} \in B_\delta(\mathbf{w}), \mathbf{v} \in B_\delta(\mathbf{z})\}.$$

Assume $\mathbf{a}^T\mathbf{w} > \mathbf{a}^T\mathbf{z}$. Then this probability is just

$$\mathbb{P}\{\mathbf{a}^T\mathbf{v} < \tau < \mathbf{a}^T\mathbf{u} \text{ for some } u \in B_\delta(\mathbf{w}), v \in B_\delta(\mathbf{z})\} = \mathbb{P}\{\min_{\mathbf{v} \in B_\delta(\mathbf{z})} \mathbf{a}^T\mathbf{v} < \tau < \max_{\mathbf{u} \in B_\delta(\mathbf{w})} \mathbf{a}^T\mathbf{u}\}$$

But

$$\min_{\mathbf{v} \in B_\delta(\mathbf{z})} \mathbf{a}^T\mathbf{v} \ge \mathbf{a}^T\mathbf{z} - \delta, \quad \max_{\mathbf{u} \in B_\delta(\mathbf{w})} \mathbf{a}^T\mathbf{u} \le \mathbf{a}^T\mathbf{w} + \delta,$$

so we have

$$\begin{aligned}
1 - P_0 &\le \mathbb{P}\{\mathbf{a}^T\mathbf{z} - \delta < \tau < \mathbf{a}^T\mathbf{w} + \delta\}. \\
&= \int_{\mathbb{S}^{n-1}} \left| \Phi\left(\frac{\mathbf{a}^T\mathbf{w} + \delta}{\sigma/\sqrt{2}}\right) - \Phi\left(\frac{\mathbf{a}^T\mathbf{z} - \delta}{\sigma/\sqrt{2}}\right) \right| \nu(\mathrm{d}\mathbf{a}) \\
&\le \int_{\mathbb{S}^{n-1}} \frac{1}{\sigma\sqrt{\pi}} \left| |\mathbf{a}^T(\mathbf{w} - \mathbf{z})| + 2\delta \right| \nu(\mathrm{d}\mathbf{a}) \\
&\le \frac{1}{\sigma\sqrt{\pi}} \int_{\mathbb{S}^{n-1}} |\mathbf{a}^T(\mathbf{w} - \mathbf{z})| \, \nu(\mathrm{d}\mathbf{a}) + \frac{2\delta}{\sigma\sqrt{\pi}} \\
&\le \frac{2}{\sigma\pi} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} \|\mathbf{w} - \mathbf{z}\| + \frac{2\delta}{\sigma\sqrt{\pi}}. \quad \square
\end{aligned}$$

Lemma 3.4 (Probability of separation with a single random measurement)**.** *Let $\mathbf{w}, \mathbf{z} \in \mathbb{B}_R^n$ be distinct. Fix $\delta > 0$. Denote by P_1 the probability that all points $\mathbf{u}$ and $\mathbf{v}$ that are within δ of $\mathbf{w}$ and $\mathbf{z}$ respectively differ by a random measurement denoted by $\mathcal{A}_i$ (i.e., the two δ-balls are separated by hyperplane i). The direction and threshold of hyperplane i are denoted by $\mathbf{a}$ and τ respectively. That is,*

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

Set $\epsilon_0 \le \|\mathbf{w} - \mathbf{z}\|$. Then,

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

Proof. Let $\epsilon = \|\mathbf{w} - \mathbf{z}\|$. Assume without loss of generality that $\|\mathbf{w}\| \ge \|\mathbf{z}\|$, otherwise swap $\mathbf{w}$ and $\mathbf{z}$ in the argument that follows. If $B_\delta(\mathbf{w}) \cap B_\delta(\mathbf{z}) \neq \emptyset$, then the chance that a hyperplane with orientation $\mathbf{a}$ and offset τ splits the two balls is zero. Hence, we may consider a restriction to the portion of the sphere where $|\mathbf{a}^T(\mathbf{w} - \mathbf{z})| \ge 2\delta$. Further, since the distribution of τ is symmetric, we can restrict to $C_\alpha = \{\mathbf{a} : \mathbf{a}^T(\mathbf{w} - \mathbf{z}) \ge \alpha\}$ for some $\alpha \ge 2\delta$ and double the integral. Hence C_α is a hyper-spherical cap of height $1 - \alpha/\epsilon$.

$$\begin{aligned}
P_1 &= \mathbb{P}\{\mathbf{a}^T\mathbf{z} + \delta \le \tau \le \mathbf{a}^T\mathbf{w} - \delta \\
&\quad \vee \mathbf{a}^T\mathbf{w} + \delta \le \tau \le \mathbf{a}^T\mathbf{z} - \delta\} \\
&\ge 2\int_{C_\alpha} \left| \Phi\left(\frac{\mathbf{a}^T\mathbf{w} - \delta}{\sigma/\sqrt{2}}\right) - \Phi\left(\frac{\mathbf{a}^T\mathbf{z} + \delta}{\sigma/\sqrt{2}}\right) \right| \omega(\mathrm{d}\mathbf{a})
\end{aligned}$$

To obtain a lower bound, we consider the area of an arbitrary subset $C'_\alpha \subset C_\alpha$ and multiply this by the minimum value of the integrand over that set. Let $\xi > 0$ and $W = \{\mathbf{a} : \mathbf{a}^T\mathbf{w} \le \xi \|\mathbf{w}\|\}$. For any $a \in C_\alpha$, since $\mathbf{a}^T(\mathbf{w} - \mathbf{z}) \ge 2\delta$, we have $\mathbf{a}^T\mathbf{z} + \delta \le \mathbf{a}^T\mathbf{w} - \delta \le \xi R$. Let $C'_\alpha = C_\alpha \cap W$. One can show, by properties of the normal distribution, for all $\mathbf{a} \in C'_\alpha$,

$$\begin{aligned}
&\left| \Phi\left(\frac{\sqrt{2}(\mathbf{a}^T\mathbf{w} - \delta)}{\sigma}\right) - \Phi\left(\frac{\sqrt{2}(\mathbf{a}^T\mathbf{z} + \delta)}{\sigma}\right) \right| \\
&\ge (\mathbf{a}^T(\mathbf{w} - \mathbf{z}) - 2\delta) \frac{\sqrt{2}}{\sigma} \phi\left(\frac{\sqrt{2}}{\sigma}\xi R\right).
\end{aligned}$$

Since $C'_\alpha = C_\alpha \cap W = C_\alpha \setminus W^c$ is the difference of hyper-spherical caps C_α and W^c, to obtain a lower bond on $\nu(C'_\alpha)$, it suffices to consider the case where $W^c \subset C_\alpha$. Since the area of C'_α does not change by varying the orientation of $\mathbf{w}$, we may assume $\mathbf{w}$ and $\mathbf{z}$ are colinear, or $\mathbf{z} = (1 - \epsilon_0)\mathbf{w}$. We set $\xi = \xi'/\sqrt{n}$, and $\alpha = \alpha'/\sqrt{n}$. Recall by assumption $\sigma = R\sqrt{2/n}$. Then the lower bound above becomes

$$\begin{aligned}
\left(\frac{\alpha'}{\sqrt{n}} - 2\delta\right) \frac{\sqrt{n}}{R} \phi(\sqrt{n}\xi) &= \left(\alpha' - 2\delta\sqrt{n}\right) \frac{\phi(\xi')}{R} \\
&= \left(\alpha' - 2\delta\sqrt{n}\right) \frac{\exp(-(\xi')^2/2)}{\sqrt{2\pi}R}.
\end{aligned}$$

By integrating, it can be shown that the normalized area of C'_α is bounded by

$$\nu(C'_\alpha) \ge \frac{1}{4\sqrt{3}}(\xi' - \alpha'/\epsilon) \exp(-(\xi')^2/2).$$

Combining the two previous formulae, and setting $\alpha' = (2\delta\sqrt{n} + \xi'\epsilon)/2$, we have,

$$\begin{aligned}
P_1 &\ge \left(\alpha' - 2\delta\sqrt{n}\right) \frac{(\xi' - \alpha'/\epsilon)}{4\sqrt{6\pi}R} \exp(-(\xi')^2) \\
&\ge \frac{(2\delta\sqrt{n} - \epsilon\xi')^2}{4\epsilon} \frac{\exp(-\xi'^2)}{4\sqrt{6\pi}R} \\
&\ge (\xi'^2\epsilon - 2\xi'\delta\sqrt{n}) \frac{\exp(-\xi'^2)}{16e\sqrt{6\pi}R}
\end{aligned}$$

We obtain the stated result by setting $\xi' = 1$ and $\epsilon \ge \epsilon_0$. $\square$

4. REFERENCES

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

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

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

[4] 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, Aug. 2007.

[5] 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.

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

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

[8] 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.

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

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

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

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

[13] Y. Lu and S. Negahban, "Individualized rank aggregration using nuclear norm regularization," *arxiv:1410.0860*, 2014.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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, pp. 2082–2102, 2013.