

On Obtaining Stable Rankings

Abolfazl Asudeh[†], H. V. Jagadish[†], Gerome Miklau^{††}, Julia Stoyanovich[‡]

[†]University of Michigan, ^{††}University of Massachusetts Amherst, [‡]Drexel University

[†]{asudeh, jag}@umich.edu, ^{††}miklau@cs.umass.edu, [‡]stoyanovich@drexel.edu

ABSTRACT

We often have to rank items with multiple attributes in a dataset. A typical method to achieve this is to compute a goodness score for each item as a weighted sum of its attribute values, and then to rank by sorting on this score. Clearly, the ranking obtained depends on the weights used for this summation. Ideally, we would want the ranked order not to change if the weights are changed slightly. We call this property *stability* of the ranking. A consumer of a ranked list may trust the ranking more if it has high stability. A producer of a ranked list prefers to choose weights that result in a stable ranking, both to earn the trust of potential consumers and because a stable ranking is intrinsically likely to be more meaningful.

In this paper, we develop a framework that can be used to assess the stability of a provided ranking and to obtain a stable ranking within an “acceptable” range of weight values (called “the region of interest”). We address the case where the user cares about the rank order of the entire set of items, and also the case where the user cares only about the top- k items.

Using a geometric interpretation, we propose the algorithms for producing the stable rankings. We also propose a randomized algorithm that uses Monte-Carlo estimation. To do so, we first propose an unbiased sampler for sampling the rankings (or top- k results) uniformly at random from the region of interest. In addition to the theoretical analyses, we conduct extensive experiments on real datasets that validate our proposal.

1. INTRODUCTION

It is often useful to rank items in a dataset. It is straightforward to sort on a single attribute, but that is often not enough. When the items have more than one attribute based on which they can be compared, it is challenging to place them in ranked order. Consider, for example, ranking of Computer Science departments. Various entities, such as US News and World Report, Times Higher Education, and the National Research Council, produce rankings. These rankings are impactful, yet heavily criticized. Several of these rankings have deficiencies in the attributes they choose to measure and in their data collection methodology, not of relevance to our paper now. Our concern is that even if these deficiencies were addressed,

we have a fundamental difficulty in obtaining a single score/rank for a department by combining multiple objective measures, such as publications, citations, funding, awards, and so on. Different ways of combining values for these attributes can lead to very different ranks. We have similar problems when we want to rank/seed sports teams, rank order cars or other products, and so on. Malcolm Gladwell very nicely describes these problems in a New Yorker article [1].

Differences in rank order can have significant consequences. For example, a company may promote high-ranked employees and fire low-ranked employees. Even in less critical scenarios, such as university rankings, it is well-documented that the ranking formula has a significant effect on policies adopted by universities [2, 3]. In other words, it matters how we choose to combine values of multiple attributes into a scoring formula. Even when there is lack of consensus on a specific way to combine attributes, we should make sure that the method we use is robust: it should not be the case that small perturbations, such as small changes in parameter values, can change the rank order.

In this paper we address the following problem: Assume that a weighted combination of the attribute values is used for assigning a score to each item; then items are sorted to produce a final rank order on items. We want this ranking to be *stable* with respect to changes in the weights used in scoring. Given a particular ranked list of items, one question a consumer of this ranking will ask is how robust the ranking is. If small changes in weights can change the ranked order, then there cannot be much confidence in the correctness of the ranking.

A given ranking of a set of items can be generated by many weight functions. Because this set of functions is continuous, we can think of them as forming a region in the space of all possible weight functions. We call a ranking of items *stable* if the ranking is generated by a weight function in large region of this space.

Note that if some items are very close together in score, it is also possible that small changes to attribute values can change the ordering of these items. Such effects tend to be local, and merely indicate that the affected items are effectively “tied” so that the change in ranking is merely a breaking of the tie. Past work, such as [4], has considered the implications of data uncertainty and sensitivity of rankings to imprecision; it is not our focus here. Instead, we address a much bigger problem: that of changes in the ranking even without any change to the attribute values, but merely due to a small change in the weighting function used to compute the score used for ranking. Such changes can be global, and can drastically change the ranking order, with far-reaching economic and societal effects, as eloquently described in [1].

Stability is a natural concern for consumers of a ranked list (i.e. those who use the ranking to prioritize items and make decisions),

who should be able to assess the magnitude of the region in weight space that produces the observed ranking. If this region is large, then the same ranked order would be obtained for many choices of weights, and the ranking is stable. But if this region is small, then we know that only a few weight choices can produce the observed ranking. This may suggest that the ranking was engineered or “cherry-picked” by the producer to obtain a specific outcome.

Data scientists often act as the producers of ranked lists (i.e. they design weight functions that result in ranked lists), and desire to produce stable results. We argued in [5] that stability in a ranked output is an important aspect of algorithmic transparency, because it allows the producer to justify their ranking methodology, and to gain the trust of consumers. Of course, stability cannot be the only criterion in the choice of ranking function: the result may be weights that seem “unreasonable” to the ranking producer. To support the producer, we allow them to specify a range of reasonable weights, or an *acceptable region* in the space of functions, and help the producer find stable rankings within this region.

Our work is motivated by the lack of formal understanding of ranking stability and the consequent lack of tools for consumers and producers to assess this critical property of rankings. We will show that rank stability hinges on complex geometric properties of rankings and weight functions. We provide novel techniques to identify stable rankings efficiently.

Furthermore, our technique does not stop at proposing just the single most stable choice, or even the k most stable choices for some pre-determined fixed value of k . Rather, it will continue to propose weight choices, and the corresponding rank ordering of items, beginning with the most stable in the specified region of interest, and continuing on in decreasing order of stability, until the user finds one that is satisfactory.

Alternatively, our techniques could provide an overview of all the rankings that occupy a large portion in the acceptable region, and hence are stable, along with an indication of the fraction of the acceptable region occupied by each. Thereby, the user can see at a glance what the stable options are, and also how dominant these are within the acceptable region.

We now motivate our techniques with an example.

EXAMPLE 1. *CSMetrics [6] ranks Computer Science research institutions based on the measured (M) and predicted (P) number of citations. These values are appropriately scaled and used in a weighted scoring formula, with parameter $\alpha \in [0, 1]$ that sets their relative importance (see § 8.1 for details). CSMetrics includes a handful of companies in its ranking, but we focus on academic departments in this example.*

As α ranges from 0 to 1, CSMetrics generates 336 distinct rankings of the top-100 departments. Assuming (as a baseline) that each ranking is equally likely, we would expect an arbitrarily chosen ranking to occur 0.3% of the time, which we take to mean that it occupies 0.3% of the volume in the space of attributes and weights. We formalize this in § 3 and call it stability of a ranking.

Suppose that the ranking with the default $\alpha = 0.3$ is released, placing Cornell (a consumer) at rank 11, just missing the top-10. Cornell then checks the stability of the published ranking (see § 3.2), and learns that its value is 0.3%, matching that of the uniform baseline. With this finding, Cornell asks CSMetrics to justify its choice of α .

CSMetrics (the producer) can respond to Cornell by further interrogating, and potentially revising the published ranking. It first enumerates stable regions (see § 3.4) and finds that the most stable ranking indeed places Cornell at rank 10 (switching with the University of Toronto), and represents 2% of the volume — an order of magnitude more than the reference ranking. However, this

stable ranking is very far from the default, placing more emphasis on measured citations with $\alpha = 0.608$. If this is unsatisfactory, CSRankings can propose another ranking closer to the reference ranking, but with better stability (see § 3.3). Interestingly, Cornell also appears at the top-10 in the most stable ranking that is within 0.998 cosine similarity from the original scoring function.

Contributions:

- We formalize a novel notion of the *stability* of a ranking, for rankings that result from a linear weighting of item attribute values. Stability captures the tolerance to changes in the weights.
- We propose algorithms that enable the efficient testing of ranking stability as well as the enumeration of the most-stable rankings, optionally constrained by a set of acceptable weighting functions. Our algorithms include those that compute exact solutions, as well as approximation algorithms that are based on novel sampling methods.
- Our empirical evaluation demonstrates the efficiency of our techniques and investigates the stability of real published rankings of computer science departments, soccer teams, and diamond retailers. We show that existing rankings in these domains are often unstable and that favoring stability can sometimes have a significant impact on the rank of some items. For instance, our findings question the validity of FIFA rankings that are claimed to be reliable, and are used in making important decisions such as seeding the World Cups final draws.

2. PRELIMINARIES

In this section we describe our data and rankings models, and the geometric interpretations that are the foundation for our algorithms.

2.1 Data model and rankings

We consider a fixed database \mathcal{D} consisting of n items, each with d scalar scoring attributes. In addition to the scoring attributes, the dataset may contain non-scoring attributes that are used for filtering, but they are not our concern here. Thus we represent an item $t \in \mathcal{D}$ as a d -length vector of scoring attributes, $\langle t[1], t[2], \dots, t[d] \rangle$. Without loss of generality, we assume that the scoring attributes have been appropriately transformed: normalized to non-negative values between 0 and 1, standardized to have equivalent variance, and adjusted so that larger values are preferred.

We consider rankings of items that are induced by first applying a linear weight function to each item, then sorting the items by the resulting score to form a ranking.

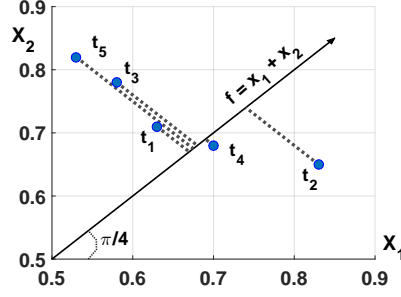
DEFINITION 1 (SCORING FUNCTION). A scoring function $f_{\vec{w}} : \mathbb{R}^d \rightarrow \mathbb{R}$, with weight vector $\vec{w} = \langle w_1, w_2, \dots, w_d \rangle$, assigns the score $f_{\vec{w}}(t) = \sum_{j=1}^d w_j t[j]$ to any item $t \in \mathcal{D}$.

Without loss of generality, we assume that $w_j \in \vec{w} \geq 0$ and when \vec{w} is clear, we denote $f_{\vec{w}}(t)$ by $f(t)$. (Although we restrict our attention to linear scoring functions, our techniques can be used with more general scoring functions by applying non-linear transformations to \mathcal{D} as a preprocessing step; see § 8 for an example.) We use \mathcal{U} to refer to the set of all possible scoring functions.

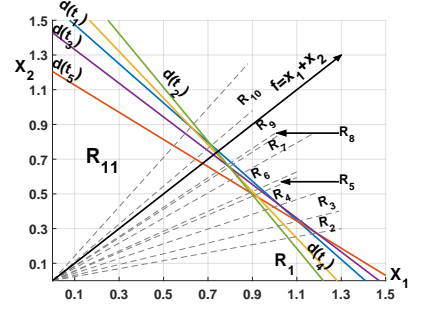
Given a score for each item, the ranking of items induced by f is the permutation of items in \mathcal{D} defined by sorting them by their scores under f , in descending order, and breaking ties consistently by an item id field. We use the notation $\nabla_f(\mathcal{D})$ to denote the ranking of items in \mathcal{D} based on f .

\mathcal{D}			f
id	x_1	x_2	$x_1 + x_2$
t_1	0.63	0.71	1.34
t_2	0.83	0.65	1.48
t_3	0.58	0.78	1.36
t_4	0.7	0.68	1.38
t_5	0.53	0.82	1.35

(a) A sample database, \mathcal{D} , of items with scoring attributes x_1 and x_2 ; and the result of scoring function $f = x_1 + x_2$.



(b) The *original space*: each item is a point. A scoring function is a ray ($f = x_1 + x_2$ is shown) which induces a ranking of the items by their projection.



(c) The *dual space*: items are the hyperplanes (lines here). Each scoring functions is a ray; within a region bounded by the intersections of the item hyperplanes, all scoring functions induce the same ranking.

Figure 1: A sample database and its geometric interpretation in the original space and dual space.

EXAMPLE 2. *The human resources (HR) department of a company wants to prioritize hiring candidates based on two criteria: an aptitude measure x_1 (e.g. a score on a qualifying exam) and an experience measure x_2 (e.g. the number of years of relevant experience). Figure 1a shows the candidates as well as their (normalized) values for x_1 and x_2 . The score for each candidate is also shown, for weight vector $\vec{w} = \langle 1, 1 \rangle$, computed as $f_{\vec{w}}(t) = x_1 + x_2$.*

2.2 Geometry of ranked items

Our algorithms are based on a geometric interpretation of scored items and induced rankings. We now present two geometric views of the database which we respectively refer to as (i) the *original space*, where every item corresponds to a point, and (ii) the *dual space*, where every item corresponds to a hyperplane.

The original space. The *original space* consists of \mathbb{R}^d with each item in \mathcal{D} represented as a point, and a linear scoring function $f_{\vec{w}}$ viewed as a ray starting from the origin and passing through the point $\vec{w} = (w_1, w_2, \dots, w_d)$. The ranking $\nabla_{f_{\vec{w}}}(\mathcal{D})$ corresponds to the ordering of their projections onto this ray.

Continuing our example, Figure 1b shows the items of our sample database in the original space, as points in \mathbb{R}^2 . The function $f = x_1 + x_2$ is shown as a ray passing through $(1, 1)$. The projection of the points onto the vector of f specifies the ranking: the further a point is from the origin, the higher its rank. Looking at Figure 1b, the candidates in Example 2 are ranked as $\langle t_2, t_4, t_3, t_5, t_1 \rangle$ based on f . One can also easily imagine the ranking of items that would result from an extreme scoring function that ranks only by attribute x_1 (i.e. $f = x_1$) by considering the projections onto the x_1 -axis (or respectively for the x_2 -axis).

Viewing the items from \mathcal{D} in the original space provides clarity about the range of rankings that could be induced by scoring functions: since all scoring functions are defined by a ray in the first quadrant of \mathbb{R}^d that is determined by the weight vector. In fact, in the sequel it is sometimes helpful to use polar coordinates to represent a scoring function. Any ray in \mathbb{R}^d starting from the origin (corresponding to a scoring function $f_{\vec{w}}$) can be identified by $(d - 1)$ angles $\langle \theta_1, \theta_2, \dots, \theta_{d-1} \rangle$, each in the range $[0, \pi/2]$. Thus, given a function $f_{\vec{w}}$, its angle vector can be computed using the polar coordinates of \vec{w} . For example, the function f with the weights $\langle 1, 1 \rangle$ in Figure 1b is identified by a single angle $\langle \pi/4 \rangle$. There is a one to one mapping between these rays and the points on the surface of the unit d -sphere (the unit hypersphere in \mathbb{R}^d). Therefore, the first

quadrant of the unit d -sphere represents the universe of functions \mathcal{U} .

The dual space. We are particularly interested in reasoning about the transition points of the weight vector, where we move from one ranking to a different ranking. The *dual space* [7] consists of \mathbb{R}^d but we now represent each item t by a hyperplane $d(t)$ given by the following equation of the d variables $x_1 \dots x_d$:

$$d(t) : t[1]x_1 + \dots + t[d]x_d = 1 \quad (1)$$

Continuing our example, Figure 1c shows the items in the dual space. In \mathbb{R}^2 , every item t is a 2-dimensional hyperplane (i.e. simply a line) given by $d(t) : t[1]x_1 + t[2]x_2 = 1$. In the dual space, functions are represented by the same ray as in the original space. The ordering of the items based on a function f is determined by the ordering of the intersection of the hyperplanes with the vector of f . The closer an intersection is to the origin, the higher its rank. For example, in Figure 1c, the line of t_2 has the intersection with the x_1 -axis that is closest to the origin, so that t_2 has the highest rank for the function $f = x_1$. In addition, we will show in the next section that the intersections of hyperplanes in the dual space define *regions*, within which rankings *do not* change under small changes of the weight vector.

3. STABILITY OF A RANKING

In this section we present our definition of stability and then identify key algorithmic problems that support consumers and producers of rankings.

3.1 Definition of stability

Every scoring function in the universe \mathcal{U} induces a single ranking of the items. But each ranking is generated by many functions. For database \mathcal{D} , let $\mathcal{R}_{\mathcal{D}}$ be the set of rankings over the items in \mathcal{D} that are generated by at least one scoring function $f \in \mathcal{U}$, that is, by at least one choice of weight vector. For a ranking $\tau \in \mathcal{R}_{\mathcal{D}}$, we define its *region*, $R_{\mathcal{D}}(\tau)$, as the set of functions that generate τ :

$$R_{\mathcal{D}}(\tau) = \{f \mid \nabla_f(\mathcal{D}) = \tau\} \quad (2)$$

Figure 1c shows the boundaries (as dotted lines) of the regions for our sample database, each corresponding one of the 11 feasible rankings of the items.

We use the region associated with a ranking to define the ranking’s stability. The intuition is that a ranking is stable if it is supported by a large set of functions that induce it. Recall from Section 2 that each function is a ray, in both the original and the dual space. If the region of a ranking is large, then small changes in the weight vector are not likely to cross the boundary of a region and therefore the ranked order will not change.

DEFINITION 2 (STABILITY OF τ AT \mathcal{D}). *Given a ranking $\tau \in \mathfrak{R}_{\mathcal{D}}$, the stability of τ is the proportion of ranking functions in \mathcal{U} that generate τ . That is, stability is the ratio of the volume of the ranking region of τ to the volume of \mathcal{U} . Formally:*

$$S_{\mathcal{D}}(\tau) = \frac{\text{vol}(R_{\mathcal{D}}(\tau))}{\text{vol}(\mathcal{U})}$$

Here the volume (vol) of a region measures the area of the space carved out in the unit d -sphere by the set of functions in the region.

We emphasize that stability is a property of a ranking (not of a scoring function) and it holds for a particular database, as indicated by the notation $S_{\mathcal{D}}(\tau)$.

Next we develop three alternative problems that build on the notion of stability in Definition 2 and correspond to the views of two different stakeholders: consumers and producers of rankings.

3.2 Consumer’s stability problem

A basic problem for the consumer is *stability verification* in which the consumer seeks to validate the stability of the ranking they have been given. A ranking with higher stability will be more robust and is less likely to be the result of an engineered scoring function.

PROBLEM 1 (STABILITY VERIFICATION). *For dataset \mathcal{D} with n items over d scoring attributes, and a ranking $\tau \in \mathfrak{R}$ of the items in \mathcal{D} , compute the ranking region $R_{\mathcal{D}}(\tau)$ and its stability $S_{\mathcal{D}}(\tau)$.*

Sometimes, the consumer may find relative stability more meaningful than absolute stability. The techniques we describe in this paper can also be configured to score a ranking relative to the one that is most stable, or most stable within some acceptable region, described next.

3.3 Acceptable scoring functions

The producer of a ranking may also care about stability because they prefer to present robust rankings rather than those that might change significantly under small changes of the scoring function. Furthermore, to the extent that consumers trust more stable rankings, producers must try to earn this trust by creating stable rankings. Of course, stability is not the only concern for the producer, and will often be traded off with the choice of a scoring function that is acceptable to the producer a priori. We return to our running example to motivate this point.

EXAMPLE 3. *In producing a ranking of employees, an HR officer believes that aptitude (x_1) should be twice as important as experience (x_2), but this is only a rough guideline. Any weight with a ratio within 20% of 2 is acceptable. By testing different weights within this acceptable range, the officer observes different rankings of candidates and will select one that maximizes stability.*

We allow producers to constrain the scoring function by specifying an acceptable region, denoted $\mathcal{U}^* \subseteq \mathcal{U}$, in one of two ways:

- A *vector and angle distance*¹: the acceptable region is identified by a hypercone around the central ray defined by the

weight vector. For example, a user may equally prefer any function that has at most $\pi/10^\circ$ angle distance (at least 95.1% cosine similarity) with the function f with weight vector $\langle 1, 1 \rangle$.

- A *set of constraints*: the acceptable region is a convex region identified by a set of inequalities. For example, if the user is interested in the functions that weigh x_2 no greater than x_1 , then the acceptable region is constrained by $w_2 \leq w_1$.

We incorporate the notion of an acceptable region into the definition of stability in a natural way. Let \mathfrak{R}^* be the set of rankings that are generated by at least one function $f \in \mathcal{U}^*$. The ranking region in \mathcal{U}^* of a ranking $\tau \in \mathfrak{R}^*$ is: $R^*(\tau) = \{f \in \mathcal{U}^* \mid \nabla_f(\mathcal{D}) = \tau\}$. Accordingly, we modify the definition of stability of a ranking $\tau \in \mathfrak{R}^*$ to be:

$$S(\tau) = \frac{\text{vol}(R^*(\tau))}{\text{vol}(\mathcal{U}^*)}$$

3.4 Producer’s stability problems

With all of the above machinery in place, we can return to helping the producer of a ranking choose one that is stable. To do so, we develop two related methods for exploring stable rankings. We state these problems with respect to an acceptable region \mathcal{U}^* and set $\mathcal{U}^* = \mathcal{U}$ when all scoring functions are acceptable.

First, the producer may wish to enumerate rankings, prioritizing those that are more stable. Below we consider an enumeration of rankings in order of stability, with stopping criteria based either on a stability threshold or a bound on the number of desired rankings.

PROBLEM 2 (BATCH STABLE-REGION ENUMERATION). *For a dataset \mathcal{D} with n items over d scoring attributes, a region of interest \mathcal{U}^* (specified either by a set of constraints or a vector-angle), and a stability threshold s (resp. a value h), find all rankings $\tau \in \mathfrak{R}^*$ such that $S(\tau) \geq s$ (resp. the top- h stable rankings).*

In many scenarios, rather than enumerating rankings, the producer may wish to incrementally generate the stable regions, in order of their stability:

PROBLEM 3 (ITERATIVE STABLE-REGION ENUMERATION). *For a dataset \mathcal{D} with n items over d scoring attributes, a region of interest \mathcal{U}^* , specified either by a set of constraints or a vector-angle, and the top- h stable rankings in \mathcal{U}^* , find the $(h+1)$ -th stable ranking $\tau \in \mathfrak{R}$. That is, find:*

$$\underset{\tau \in \mathfrak{R} \setminus \text{top-}h}{\text{argmax}} (S(\tau))$$

Of course, the two enumeration problems are closely related. In fact, given an algorithm for iterative ranking enumeration, it can be used directly for batch ranking enumeration by making multiple calls to the iterative algorithm. The core of our algorithmic contributions focus on efficiently evaluating an operator we call GETNEXT, which can be used to solve both enumeration problems.

3.5 Stability of partial rankings

So far we focused on complete rankings of n items in \mathcal{D} . However, when n is large, one may be interested in only the highest-ranked k items, for $k \ll n$. This motivates us to reformulate the problems stated above, focusing on the top- k portion of the rankings. One challenge is that, while ranking regions are guaranteed to be convex for complete rankings, they may not be convex for top- k partial rankings. We present sampling-based randomized algorithms that support top- k partial rankings in § 6.1.

¹Note that this is the same as cosine similarity.

4. TWO DIMENSIONAL (2D) RANKING

To develop our intuition, we start by studying the case of $d = 2$ scoring attributes. Using the geometric interpretation of items provided in § 2 while considering the dual representation of the items, we propose two exact algorithms for both stability verification and enumeration.

Consider the items in Example 2, as shown in Figure 1a. While the number of ranking functions, rays between 0° and $\pi/2^\circ$, is infinite, the number of possible orderings between the items is (combinatorially) limited to $n!$. The number of possible orderings based on linear ranking functions is even less.

Consider a pair of items t_i and t_j presented in the dual space in \mathbb{R}^2 . Recall that in 2D, every item t is transformed to the line:

$$d(t) : t[1]x_1 + t[2]x_2 = 1 \quad (3)$$

Also, recall that every function f with the weight vector w is represented with the origin-starting ray passing through the point w , and consider points t_i and t_j . f ranks t_i higher than t_j if the intersection of $d(t_i)$ with f is closer to the origin than the intersection of $d(t_j)$ with f .

Consider f whose origin-starting ray passes through the intersection of $d(t_i)$ and $d(t_j)$. Since both lines intersect with the ray of f at the same point, f assigns an equal score to t_i and t_j . We refer to this function (and its ray) as the *ordering exchange* between t_i and t_j , and denote it \times_{t_i, t_j} [8]. The ordering between t_i and t_j changes on two sides of \times_{t_i, t_j} : t_i is ranked higher than t_j one side of the ray, and t_j is ranked higher than t_i on the other side. For example, consider t_1 and t_4 in Example 2, shown in Figure 1c in the dual space: the blue line represents $d(t_1)$, and the yellow line represents $d(t_4)$. The left-most intersection in the figure is between $d(t_1)$ and $d(t_4)$. The top-left dashed line that starts from the origin and passes through this intersection shows \times_{t_1, t_4} : t_1 is preferred over t_4 on the left of \times_{t_1, t_4} , and t_4 is preferred over t_1 on the right.

Consider two items t and t' that do not dominate each other². Equation 3 can be used for finding the intersection between the lines $d(t)$ and $d(t')$. Considering the polar coordinates of the intersection, $\times_{t, t'}$ is specified by the angle $\theta_{t, t'}$ (between the ordering exchange and the x-axis) as follows:

$$\theta_{t, t'} = \arctan \frac{t'[1] - t[1]}{t[2] - t'[2]} \quad (4)$$

The ordering exchanges between pairs of items of a database partition the space of scoring functions into a set of regions. Each region is identified by the two ordering exchanges that form its borders. Since there are no ordering exchanges within a region, all scoring functions inside a region induce the same ranking of the items. Thus, the number of regions is equal to $|\mathfrak{R}|$, as \mathfrak{R} is the collection of rankings defined by these regions. For instance, Figure 1c shows regions R_1 through R_{11} that define the set of possible rankings of Example 2 for \mathcal{U} .

4.1 Stability Verification

The ordering exchanges are the key to figuring out the stability of a ranking. Consider a ranking τ . For a value of $1 \leq i < n$, let t and t' be the i -th and $(i+1)$ -th items in τ . Following Equation 4, $\theta_{t, t'}$ specifies the ordering exchange between t and t' . If $t[1] < t'[1]$ (resp. $t[1] > t'[1]$), all functions with angles $\theta < \theta_{t, t'}$ (resp. $\theta > \theta_{t, t'}$) rank t higher than t' .

Algorithm 1 uses this idea for computing the stability (and the region) of a given ranking τ . The stability verification algorithm

²An item t dominates an item t' , if $\nexists x_i$ s.t. $t'[i] > t[i]$ and $\exists x_i$ s.t. $t[i] > t'[i]$ [9, 10]. If t dominates t' , then these items do not exchange order.

Algorithm 1 SV_{2D}

Input: Two dimensional dataset \mathcal{D} with n items and the ranking τ
Output: The stability and the region of τ

```

1:  $(\theta_1, \theta_2) = (0, \pi/2)$ 
2: for  $i = 1$  to  $n - 1$  do
3:    $t = \tau[i]; t' = \tau[i + 1]$ 
4:   if  $t$  dominates  $t'$  then continue
5:   if  $t'$  dominates  $t$  then return null
6:    $\theta = \arctan \frac{t'[1] - t[1]}{t[2] - t'[2]}$ 
7:   if  $(t[1] < t'[1] \text{ and } \theta > \theta_1)$  then  $\theta_1 = \theta$ 
8:   if  $(t[1] > t'[1] \text{ and } \theta < \theta_2)$  then  $\theta_2 = \theta$ 
9:   if  $\theta_1 > \theta_2$  then return null
10: end for
11: return  $\frac{\theta_2 - \theta_1}{\pi/2}, (\theta_1, \theta_2)$ 
```

uses the angle range (θ_1, θ_2) for specifying the region of τ . For each value of i in range $[1, n]$, the algorithm considers the items t and t' to be the i -th and $(i+1)$ -th items in τ , respectively. If t' dominates t , the ranking is not valid. Otherwise, if t does not dominate t' , the algorithm computes the ordering exchange $\times_{t, t'}$ and, based on the values of $t[1]$ and $t'[1]$, decides to use it for setting the upper bound or the lower bound of the ranking region. After traversing the ranked list τ , the algorithm returns $\frac{\theta_2 - \theta_1}{\pi/2}$ as the stability value and (θ_1, θ_2) as the region of τ . Since the algorithm scans the ranked list only once, stability verification in 2D is in $O(n)$.

4.2 Stability Enumeration

In 2D, \mathcal{U}^* is identified by two angles demarcating the edges of the pie-slice. For example, let \mathcal{U}_1^* be defined by the set of constraints $\{w_1 \leq w_2, 2w_1 \geq w_2\}$. This defines the set of functions above the line $w_1 = w_2$ and below the line $2w_1 = w_2$, limiting the region of interest to the angles in the range $[\pi/4^\circ, \pi/3^\circ]$. Similarly, let \mathcal{U}_2^* be defined around the function $f = x_1 + x_2$ with the maximum angle $\pi/10^\circ$ (i.e., minimum cosine similarity of 95.1% with f). The $\pi/10^\circ$ angle around f identifies the range $[3\pi/20^\circ, 7\pi/20^\circ]$ as the region of interest. In what follows, we use $[\mathcal{U}^*[1], \mathcal{U}^*[2]]$ to denote the borders of \mathcal{U}^* . Based on Definition 2, the stability of a ranking $\tau \in \mathfrak{R}$ in 2D is the span of its region – the distance between its two borders. For example, in Figure 1c, the regions R_{11} and R_1 are wide and provide stable rankings, while the rankings provided by R_5 and R_8 are hazy and may change by small changes in the weight vector.

We propose the algorithm RAYSWEEPING (Algorithm 2) that starts from the angle $\mathcal{U}^*[1]$ and, while sweeping a ray toward $\mathcal{U}^*[2]$, uses the dual representation of the items for computing the ordering exchanges and finding the ranking regions. The algorithm stores the regions, along with the stability of their rankings, in a heap data structure that is later used by the GET-NEXT_{2D} operator.

Algorithm 2 starts by ordering the items based on $\mathcal{U}^*[1]$. It uses the fact that at any moment, an adjacent pair in the ordered list of items exchange ordering, and, therefore, computes the ordering exchanges between the adjacent items in the ordered list. The intersections that fall into the region of interest are added to the sweeper's min-heap. Until there are intersections over which to sweep, the algorithm pops the intersection with the smallest angle, marks the region between it and the previous intersection in the output max-heap, and updates the ordered list accordingly. Upon updating the ordered list, the algorithm adds the intersections between the new adjacent items to the sweeper. Since the total number of intersections between the items is bounded by $O(n^2)$, and the heap operation is in $O(\log n)$, RAYSWEEPING is in $O(n^2 \log n)$.

Algorithm 2 RAYSWEEPING

Input: Two dimensional dataset \mathcal{D} with n items and the region of interest in the form of an angle range $[\mathcal{U}^*[1], \mathcal{U}^*[2]]$

Output: A heap of ranking regions and their stability

```

1: sweeper = new min-heap( $[\mathcal{U}^*[2]]$ );
2:  $w = (\cos \mathcal{U}^*[1], \sin \mathcal{U}^*[1])$ 
3:  $L = \nabla_f(\mathcal{D})$ 
4: for  $i = 1$  to  $n - 1$  do
5:    $\theta = \arctan(L_{i+1}[1] - L_i[1]) / (L_i[2] - L_{i+1}[2])$ 
6:   if  $\mathcal{U}^*[1] < \theta < \mathcal{U}^*[2]$  then sweeper.push( $(\theta, L_i, L_{i+1})$ )
7: end for
8: heap = new max-heap();  $\theta_p = \mathcal{U}^*[1]$ 
9: while sweeper is not empty do
10:   $(\theta, t, t') = \text{sweeper.pop}()$ 
11:   $i, j = \text{index of } t, t' \text{ in } L$ 
12:  heap.push( $(\frac{\theta - \theta_p}{\mathcal{U}^*[2] - \mathcal{U}^*[1]}, (\theta_p, \theta))$ )
13:  swap  $L_i$  and  $L_j$ 
14:  if  $i > 1$  then
15:     $\theta' = \arctan(t'[1] - L_{i-1}[1]) / (L_{i-1}[2] - t'[2])$ 
16:    if  $\theta < \theta' < \mathcal{U}^*[2]$  then sweeper.push( $(\theta', L_{i-1})$ )
17:  end if
18:  if  $j < n$  then
19:     $\theta' = \arctan(t[1] - L_{j+1}[1]) / (L_{j+1}[2] - t[2])$ 
20:    if  $\theta < \theta' < \mathcal{U}^*[2]$  then sweeper.push( $(\theta, L_{j+1})$ )
21:  end if
22:   $\theta_p = \theta$ 
23: end while
24: return heap

```

After finding the ranking regions and their spans, every call to GET-NEXT_{2D} (Algorithm 3) pops the most stable region from the heap and chooses a scoring function f inside the region. The algorithm returns the ranking $\nabla_f(\mathcal{D})$, along with the width of the region (the stability of the ranking), to the user.

Algorithm 3 GET-NEXT_{2D}

Input: The heap of regions, with the top- h regions being removed from it

Output: the top- $h + 1$ ranking, along with its stability

```

1:  $S, (\theta_1, \theta_2) = \text{heap.pop}()$ 
2:  $w = (\cos \frac{\theta_1 + \theta_2}{2}, \sin \frac{\theta_1 + \theta_2}{2})$ 
3:  $L = \nabla_f(\mathcal{D})$ 
4: return  $L, S, (\theta_1, \theta_2)$ 

```

Since there are no more than $O(n^2)$ regions in the heap, GET-NEXT_{2D} needs $O(\log n)$ to find the $(h + 1)$ -th stable region. Then, it takes $O(n \log n)$ to compute the ranking for the region. As a result, the first call to GET-NEXT_{2D} that creates the heap of regions takes $O(n^2 \log n)$, while subsequent calls take $O(n \log n)$. Note that subsequent GET-NEXT_{2D} calls can be done in the order of $O(\log n)$, with memory cost of $O(n^3)$, by storing the ordered list L for every region in RAYSWEEPING algorithm.

5. MULTI DIMENSIONAL (MD) RANKING

Building upon the intuitions developed from the 2D case, we now turn to the general setting where $d > 2$. Again, we consider the items in dual space and use the ordering exchanges for specifying the borders of the ranking regions. Recall from Equation 1 that an item t is presented as the hyperplane $d(t) : \sum_{i=1}^d t[i] \cdot x_i = 1$.

For a pair of items t_i and t_j the ordering exchange $h = \times_{t_i, t_j}$ is a hyperplane that contains the functions that assign the same score to both items. Therefore:

$$\times_{t_i, t_j} = \sum_{k=1}^d (t_i[k] - t_j[k])x_k = 0 \quad (5)$$

Every hyperplane $h = \times_{t_i, t_j}$ partitions the function space \mathcal{U} in two “half-spaces” [7]:

- $h^- : \sum_{k=1}^d (t_i[k] - t_j[k])x_k < 0$: for the functions in h^- , t_j outranks t_i .
- $h^+ : \sum_{k=1}^d (t_i[k] - t_j[k])x_k > 0$: for the functions in h^+ , t_i outranks t_j .

Similar to § 4, we first show how ordering exchanges can be used for verifying the stability of a ranking and then focus on designing the GET-NEXT operator.

5.1 Stability Verification

Identifying the half-spaces defined by the ordering exchanges between adjacent items in a ranking τ is the key to figuring out its stability. For each value of i in range $[1, n]$, let t and t' be the i -th and $(i + 1)$ -th items in τ . Using Equation 5, every function in the positive half-space $h^+ : \sum_{k=1}^d (t[k] - t'[k])x_k > 0$ ranks t higher than t' . The intersection of these half-spaces specifies an open-ended d -dimensional cone (referred to as d -cone) whose base is a $(d - 1)$ dimensional convex polyhedron. Every function in this cone generates the ranking τ , while no function outside it generates τ . In other words, this d -cone is the ranking region of τ . Algorithm 4 finds the region of a ranking τ as the set of positive half-spaces defined by the ordering exchanges of the adjacent items in τ .

Algorithm 4 SV

Input: Dataset \mathcal{D} with n items and d attributes, and the ranking τ

Output: The stability and the region of τ

```

1:  $R = \{\}$ 
2: for  $i = 1$  to  $n - 1$  do
3:    $t = \tau[i]; t' = \tau[i + 1]$ 
4:   if  $t$  dominates  $t'$  then continue
5:   if  $t'$  dominates  $t$  then return null // infeasible
6:    $R.add(\sum_{k=1}^d (t[k] - t'[k])x_k > 0)$ 
7: end for
8: return  $S(R, \mathcal{U}) / *C.f.f. \quad \S \ 6.3*, R$ 

```

Based on Definition 2, the volume ratio of the region of τ to the one of \mathcal{U} (or, more generally, to \mathcal{U}^*) is its stability. However, since τ is a polyhedron, computing its volume is #P-hard [11]. Therefore, we use numeric methods for estimating this quantity. Throughout this section, we assume the existence of a stability oracle $S(R, \mathcal{U}^*)$ that, given a convex region R in the form of an intersection of half-spaces and a region of interest \mathcal{U}^* , returns the stability of R in \mathcal{U}^* . We will describe in § 6.3 how to construct such an oracle. Stability verification in MD is in $O(n + X_S)$ where X_S is the complexity of constructing the stability oracle.

5.2 Stability Enumeration

Similar to verifying the stability of a ranking, ordering exchanges can be used for finding possible rankings in a region of interest \mathcal{U}^* . The set of ordering exchanges intersecting a region of interest define a dissection of \mathcal{U}^* into connected convex d -cones, each showing a ranking region as the intersection of ordering exchange half-spaces and \mathcal{U}^* . This dissection is named the *arrangement* of

ordering exchange hyperplanes [7]. Algorithm 5 shows the pseudocode of the function \times_{hps} that finds the set of ordering exchange hyperplanes passing through the region of interest \mathcal{U}^* .

Algorithm 5 \times_{hps}

Input: database \mathcal{D} , n , d , region of interest \mathcal{U}^*

Output: the set of ordering exchange hyperplanes intersecting \mathcal{U}^*

```

1:  $H = \{\}$ 
2: for  $t_i, t_j$  in  $\mathcal{D}$  do
3:   if  $t_i$  and  $t_j$  do not dominate each other then
4:     Let  $\times_{t_i, t_j}$  be the hyperplane  $\sum_{k=1}^d (t_i[k] - t_j[k])x_k = 0$ 
5:     if  $\times_{t_i, t_j}$  intersect with  $\mathcal{U}^*$  then  $H.add(\times_{t_i, t_j})$ 
6:   end if
7: end for
8: return  $H$ 

```

THEOREM 1. Every ranking $\tau \in \mathfrak{R}^*$ is provided by the functions in exactly one convex region in the arrangement of ordering exchange hyperplanes in \mathcal{U}^* .

PROOF. First, every ranking $\tau \in \mathfrak{R}$ is generated by at least one function in \mathcal{U}^* . Since the arrangement partitions the \mathcal{U}^* into disjoint regions, this function belongs to one of the regions (d -cones) of the arrangement. Also, because there is no ordering exchange in a region, every function in it will generate the same ranking, i.e., for every ranking there exists at least one region that generates it. Now, to prove the one to one mapping by contradiction, assume the existence of two regions that generate the same ranking. Since these two regions are disjoint, there should exist at least an ordering exchange hyperplane $h = \times_{t_i, t_j}$ between them. Hence, one of them belongs to the half-space h^+ while the other falls into h^- . Therefore, in one of the regions, t_i is preferred over t_j while in the other t_j is ranked higher. This contradicts with the assumption that both regions generate the same ranking. \square

Theorem 1 shows the one-to-one mapping between the rankings $\tau \in \mathfrak{R}^*$ and the regions of the arrangement. Following Theorem 1, the baseline for finding the stable regions in \mathcal{U}^* is to first construct the arrangement and then, similarly to § 4, create a heap of regions and their stabilities. Then, each GET-NEXT operator is as simple as removing the most stable ranking from the heap. The construction of arrangements is extensively studied in the literature [7, 8, 12–15]. The problem with this baseline is that it first needs to construct the complete arrangement of ordering exchange hyperplanes, and to compute the stability of each. The number of ordering exchanges intersecting \mathcal{U}^* is bounded by $O(n^2)$. Therefore the arrangement can contain as many as $O(n^{2d})$ regions [7]. Yet, the baseline needs to compute the stability of each ranking associated with every region. Considering that our objective is to find stable rankings, rather than to discover all possible rankings, and that the user will likely be satisfied after observing a few rankings, this construction is wasteful.

Instead, we propose an algorithm that targets the construction of only the $(h + 1)$ -th stable ranking and delays the construction of other rankings. Arrangement construction is an iterative process that starts by partitioning the space into two half-spaces by adding the first hyperplane $H[1]$ (it partitions the space into $H[1]^-$ and $H[1]^+$). The algorithm then iteratively adds the other hyperplanes; for adding the hyperplane $H[i]$, it first identifies the set of regions in the arrangement of $H[1]$ to $H[i - 1]$ that $H[i]$ intersects with, and then splits each such region into two regions (one including $H[i]^-$ and one $H[i]^+$).

```

struct Region {
  C; // region half-spaces
  S; // the stability of the region
  pending; // the first pending hyperplane
  sb; // samples beginning index (c.f. § 6.4)
  se; // samples end index (c.f. § 6.4)
}

```

Figure 2: The region data structure

The GET-NEXT _{md} algorithm, however, only breaks down the largest region at every iteration, delaying the construction of the arrangement in all other regions. The algorithm uses the “region” data structure (Figure 2) for maintaining a record of each region in the arrangement of ordering exchanges. This data structure contains the following fields: (a) C : the set of half-spaces defining the region; (b) S : the stability of the region, and (c) $pending$: the index of the next hyperplane to be added to the region. In addition, every region contains two extra fields sb and se that are used for determining the intersection of next hyperplanes with it (as well as computing its stability). We defer further details about these two fields to § 6.4.

Algorithm 6 GET-NEXT _{md}

Input: database \mathcal{D} , n , d , region of interest \mathcal{U}^* , top- h stable rankings \mathfrak{R}_h

Output: $(h + 1)$ -th stable ranking and its stability measures

```

1: if  $\mathfrak{R}_h$  is  $\emptyset$  then
2:    $heap = new \text{ max-heap}()$ 
3:    $H = \times_{hps}(\mathcal{D}, n, d, \mathcal{U}^*)$ 
4:    $heap.push(1, new \text{ Region}(C = \{\}, pending=1, S = 1))$ 
5: end if
6: while  $heap$  is not empty do
7:    $s, r = heap.pop()$ 
8:   while true do
9:     if  $r.pending > |H|$  then
10:       $w = \text{a point in } r$ 
11:      return  $\nabla_f(\mathcal{D}), r.S$ 
12:     end if
13:      $k = \text{passThrough}(H[r.pending], r)$  // c.f. § 6.4
14:     if  $k > 0$  /* intersects  $r^*$  */ then break
15:      $r.pending += 1$ 
16:   end while
17:    $C_1 = C_2 = r.C$ ;  $p = r.pending + 1$ 
18:   add  $H[r.pending]^-$  to  $C_1$  and  $H[r.pending]^+$  to  $C_2$ 
19:    $r_1 = new \text{ Region}(C = C_1, pending=p, S = S(C_1))$ 
20:    $r_2 = new \text{ Region}(C = C_2, pending=p, S = S(C_2))$ 
21:    $heap.push(r_1.S, r_1)$ ;  $heap.push(r_2.S, r_2)$ 
22: end while

```

Algorithm 6 shows the pseudo-code of the multi-dimensional GET-NEXT operator. At the high level, while constructing the arrangement of hyperplanes, the algorithm keeps track of the stability of the regions, as it adds hyperplanes to largest one. It uses a max-heap for this purpose. For the first GET-NEXT operation, the algorithm calls Algorithm 5 to find the set of ordering exchanges H that intersect with \mathcal{U}^* . It also creates the root region that contains all functions in \mathcal{U}^* and adds it to the heap. While the heap is not empty, the algorithm pops the most stable region r from it. It then iterates over the pending hyperplanes that can be added to r , attempting to find one that intersects with that region. Testing

whether a hyperplane intersects with a region is done by solving a linear program. Specifically, we solve a quadratic program that looks for a function in \mathcal{U}^* that satisfies both the inequality constraints defined by the half-spaces of the region, and the equality constraints defined by the hyperplane. (Alternatively, sampling can be used for this purpose, as we will show in § 6.4.)

If no more hyperplanes can be added to region r , the algorithm returns r as the next stable region. Otherwise, if a hyperplane is found that intersects with r , then the algorithm uses it to break r into two regions, and adds them to the heap.

Still, in the worst case (where all regions are equally stable) the algorithm may need to construct the arrangement before returning even the first stable region. Therefore, the worst case cost of the algorithm is still $O(n^{2d})$.

Throughout this section, we assumed the existence of an oracle that, given a region in the form of a set of half-spaces, returns its stability. In next section, we discuss unbiased sampling from the function space that plays a key role in the design of the oracle. Such sampling will also enable the design of the randomized algorithm in § 7 that does not depend on the arrangement construction (and therefore, does not suffer its high complexity).

6. UNBIASED FUNCTION SAMPLING

A uniform sampler from the function space is a key component for devising Monte-Carlo methods [16], both for estimating the stability of rankings and for designing randomized algorithms for the problem. While every ranking is generated by continuous ranges of functions, every function f generates only one ranking of items. Moreover, the larger the volume of a ranking region is (i.e. the more stable it is), the higher is the chance of choosing a random function from it. Therefore, *uniform sampling of the function space provides the sampling of rankings based on their stability distribution*. This suggests using such samples from the function space to design a Monte-Carlo method. But first, we need a sampler that generates functions uniformly at random from a region of interest \mathcal{U}^* . In the following, we first discuss sampling from the complete function space and then propose an efficient sampler for \mathcal{U}^* .

6.1 Sampling from the function space

In this subsection we discuss how to generate unbiased samples from the complete function space. Since every function is represented as a vector of $d - 1$ angles, each in range $[0^\circ, \pi/2^\circ]$, one way for generating random functions is by generating angle vectors uniformly at random. This, however, does not provide uniform random functions sampled from the function space, except for 2D.

To demonstrate this in 3D, we generated 1000 two dimensional random vector with values in range $[0, \pi/2]$. Then, for every vector Θ , we plotted the point with the polar cardinalities $(1, \Theta)$ in Figure 3. These represent the end points for the unit vectors with angle Θ . Looking at the figure, it is clear that the distribution is not uniform, as the density of the end points reduces moving from the top of the figure to the bottom. Also, looking at Figure 3, one may notice that the set of points in the first quadrant of the unit d -sphere represent the function space \mathcal{U} . This is because of the one-to-one mapping between the points on the surface of the unit d -sphere and the unit origin-starting rays, each representing a function f . Hence, the problem of choosing functions uniformly at random from \mathcal{U} is equivalent with choosing random points from the surface of a d -sphere. As also suggested in [17], we adopt a method for uniform sampling of the points on the surface of the unit d -sphere [18, 19]. Rather than sampling the angles, this method samples the weights using the *normal distribution*, and normalizes them. This method

works because the normal distribution function has a constant probability on the surfaces of d -spheres with common centers [19, 20]. Algorithm 7, adopt this method to generate random functions from \mathcal{U} .

Algorithm 7 Sample \mathcal{U}

```

1: for  $i = 1$  to  $d$  do
2:    $w_i = |\mathcal{N}(0, 1)|$  //  $\mathcal{N}(0, 1)$  generates a random number
      from the standard normal distribution
3: end for
4: return  $w$ 
```

In order to verify the uniformity of the output of Algorithm 7, we used it to draw 1000 sample functions. Similar to Figure 3, we plotted the endpoints of the unit vectors of sampled functions in Figure 4 – which shows the points are distributed uniformly.

6.2 Sampling from the region of interest

Drawing unbiased samples from a region of interest \mathcal{U}^* is critical for finding stable regions. Given an unbiased sampler for the function space \mathcal{U} , an acceptance-rejection method [21] can be used for drawing samples from \mathcal{U}^* . The idea is simple: (i) draw a sample from the function space; (ii) test if the drawn sample is in the region of interest and, if so, accept it; otherwise reject the sample and try again. Testing if the drawn sample is in the region of interest can be done by: (a) computing its angle distance from the reference ρ and comparing it with the reference angle θ , if \mathcal{U}^* is specified by a ray and angle, or by (b) checking if the sampled point satisfies the constraint, if \mathcal{U}^* is specified by a set of constraints.

The efficiency of this method, however, depends on the acceptance probability p , defined by the volume ratio of \mathcal{U}^* to \mathcal{U} . The expected number of trials for drawing a sample for such probability is $1/p$. Hence, it is efficient if the volume of \mathcal{U}^* is not small.

Therefore, in the following, we alternatively propose an inverse CDF (cumulative distribution function) method [22] for generating random uniform functions from a region of interest. This method is preferred over the acceptance-rejection method when \mathcal{U}^* is small. It generates functions with the maximum angular distance of θ from the reference ray ρ . For a region of interest specified by a set of constraints, the bounding sphere [23] for the base of its d -cone identifies the ray and angle distance that include \mathcal{U}^* . For such regions of interest, the inverse CDF method enables an acceptance-rejection method with higher acceptance rate, leading to better performance.

Consider \mathcal{U}^* as the set of functions with the maximum angle θ around the ray of some function f . We model \mathcal{U}^* by the surface unit d -spherical cap with angle θ around the d -th axis in \mathbb{R}^d (Figure 5). This is similar to the mapping of \mathcal{U} to the surface of the unit hyperspherical, and is due to the one-to-one mapping between the rays in \mathcal{U}^* and the points on the surface unit hyperspherical cap. We use a transformation that maps the ray of f to the d -D axis. After drawing a function we transform it around the ray of f .

For an angle θ , the d -dimension orthogonal plane $x_d = \cos \theta$ partitions the cap from the rest of the d -sphere. Hence, the intersection of the set of the $(d$ -th axis orthogonal) planes $\cos \theta \leq x_d \leq 1$ with the d -sphere define the cap.

The intersection of each such plane with the d -sphere is a $(d-1)$ -sphere. For example, in Figure 5 the intersection of a plane that is orthogonal to the z -axis with the unit sphere is a circle (2-sphere). An unbiased sampler should sample points from the surface of such $(d-1)$ -spheres proportionally to their areas. In the following, we show how this can be done.

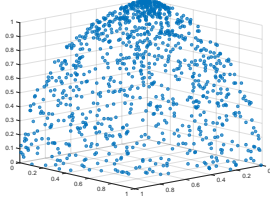


Figure 3: Distribution of 1000 random functions in \mathbb{R}^3 , generated by taking uniform samples of the angles

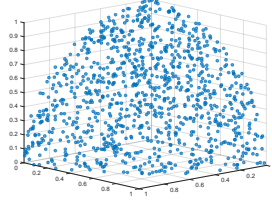


Figure 4: Illustration of 1000 random uniform functions taken in \mathbb{R}^3 , using Algorithm 7

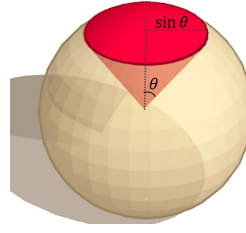


Figure 5: Modeling \mathcal{U}^* as a unit d -spherical cap around the d -th axis

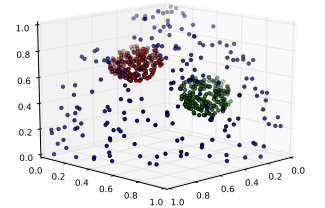


Figure 6: Samples generated using (i) blue: Algorithm 7, (ii) green: Algorithm 9 and Algorithm 8, (iii) red: Algorithm 9 and Equation 9

The surface area of a δ -sphere with the radius r is [24]:

$$A_\delta(r) = \frac{2\pi^{\delta/2}}{\Gamma(\delta/2)} r^{\delta-1} \quad (6)$$

Γ , in the above equation, is the gamma function.

Using this equation, the area of the unit d -spherical cap can be stated as the integral over the surface areas of the $(d-1)$ -spheres, defined by the intersection of the planes $\cos \theta \leq x_d \leq 1$ with the d -sphere, as follows [24]:

$$\begin{aligned} A_d^{cap}(1) &= \int_0^\theta A_{d-1} \sin \phi d\phi \\ &= \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^\theta \sin^{d-2}(\phi) d\phi \end{aligned} \quad (7)$$

Therefore, considering the random angle $0 \leq x \leq \theta$, the cumulative density function (cdf) for x is given by:

$$F(x) = \frac{\int_0^x \sin^{d-2}(\phi) d\phi}{\int_0^\theta \sin^{d-2}(\phi) d\phi} \quad (8)$$

For a specific value of d , one can solve Equation 8, find the inverse of F and use it for sampling. For instance, for $d = 3$:

$$F(x) = \frac{1 - \cos x}{1 - \cos \theta} \Rightarrow F^{-1}(x) = \arccos(1 - (1 - \cos \theta)x) \quad (9)$$

For a general d , we can use the representation of $\int_0^\theta \sin^{d-2}(\phi) d\phi$ in the form of beta function and regularized incomplete beta function [24] and rewrite Equation 8 as³:

$$F(x) = \frac{I_{\sin^2(x)}\left(\frac{d-1}{2}, \frac{1}{2}\right)}{I_{\sin^2(\theta)}\left(\frac{d-1}{2}, \frac{1}{2}\right)} \quad (10)$$

However, since numeric methods are applied for finding the inverse of the regularized incomplete beta function [25], we consider a numeric solution for Equation 8. Consider a regular partition of the interval $[0, \theta]$. The integral $\int_0^\theta \sin^{d-2}(\phi) d\phi$ can be computed as the sequence of Riemann sums over the partitions of the interval. We apply this for computing both the denominator and the numerator of Equation 8. Given the partition of the interval, we start from the angle 0, and for each partition, x' compute the value of $\int_0^{x'} \sin^{d-2}(\phi) d\phi$ as the aggregate over the previous summations and store it in a sorted list. As a result, in addition to the value of

³ $I_z(\alpha, \beta)$ is the regularized incomplete beta function.

the denominator, we have the value of F for each of the partitions. We will later apply binary search on this list, in order to find the angle x that has the area $F(x)$. Algorithm 8 shows the pseudocode of the function *RiemannSums* that computes the denominator and returns the list of partial integrals divided by the denominator. In addition to the angle θ , the function takes the number of partitions as the input.

Algorithm 8 RiemannSums

Input: The angle θ and number of partitions γ

```

1:  $\epsilon = \theta/\gamma$ 
2:  $L = [0]; A = 0; \alpha = \epsilon$ 
3: for  $i = 1$  to  $\gamma$  do
4:    $A = A + \sin^{d-2}(\alpha)$ 
5:    $L.append(A)$ 
6:    $\alpha = \alpha + \epsilon$ 
7: end for
8: for  $i = 1$  to  $\gamma$  do  $L[i] = L[i]/A$ 
9: return  $L$ 
```

Algorithm 9 shows the pseudocode of the inverse CDF sampler. It takes the list L (computed using the function *RiemannSums*) as the input and draws a random function from \mathcal{U}^* . To do so, it first draws a random uniform number y in the range $[0, 1]$. Next, it applies a binary search on the list of partial integrals to find the range to which y belongs. Considering a fine granularity of the partitions, we assume that the areas of all $(d-1)$ -spheres inside each partition are equal. The algorithm, therefore, returns a random angle (drawn uniformly at random) from the selected partition.

Algorithm 9 Sample \mathcal{U}^*

Input: The ray ρ , angle θ , number of partitions γ , and L

```

1:  $y = U[0, 1]$  // draw a uniform sample in range  $[0, 1]$ 
2:  $i = \text{binarySearch}(y, L)$ 
3:  $\hat{\epsilon} = U[0, \epsilon]$ 
4:  $x = i\epsilon + \hat{\epsilon}$ 
5: for  $i = 1$  to  $d-1$  do  $\hat{w}_i = \mathcal{N}(0, 1)$ 
6:  $\langle \theta_1, \dots, \theta_{d-2} \rangle$  = the angles in polar representation of  $\hat{w}$ 
7:  $w = \text{toCartesian}(1, \langle \theta_1, \dots, \theta_{d-2}, x \rangle)$ 
8: return  $\text{Rotate}(w, \rho)$ 
```

Recall that the angle x specifies the intersection of a plane with the d -spherical cap, which is a $(d-1)$ -sphere. For example, in \mathbb{R}^3 , the intersection is a circle (2-sphere). Hence, after finding the angle x , we need to sample from the surface of a $(d-1)$ -sphere,

uniformly at random. Also, recall from § 6.1 that the normalized set of $d - 1$ random numbers drawn from the normal distribution provide a random sample point on the surface of $(d - 1)$ -sphere. The algorithm $\text{Sample}\mathcal{U}^*$ uses this for generating such a random point. It uses the angle combination of the drawn random point from the surface of a $(d - 1)$ -sphere and combines them with the angle x (with the d -th axis). After this step, the point specified by the polar coordinates $(1, \langle \theta_1, \dots, \theta_{d-2}, x \rangle)$ is the random uniform point from the surface of d -spherical cap around the d -th axis. As the final step, the algorithm needs to rotate the coordinate system such that the center of the cap (currently on d -th axis) falls on the ray ρ . We rely on the existence of the function *Rotate* for this, presented in Appendix A.

The final discussion is when to apply the acceptance-rejection method and when the inverse CDF. For a fix d that the inverse CDF is calculated (e.g. Equation 9), the sampling cost is constant and, therefore, it is always preferred over the acceptance/reject method. However, in general where the numeric method is applied for the inverse CDF method (Algorithm 6.1), the cost $O(\log |L|)$. Specifically, in order to make Algorithm 9 run in $O(\log n)$, one can consider $L = O(n)$. On the other hand, the (expected) cost of the acceptance/reject method depends on the ratio of the volume of \mathcal{U} to \mathcal{U}^* . Based on Equations 6 and 7, $1/\int_0^\theta \sin^{d-2}(\phi)d\phi$ shows this ratio. Therefore, if $\log(|L|) > (1/\int_0^\theta \sin^{d-2}(\phi)d\phi)$, the acceptance/reject method is expected to outperform the inverse CDF method; otherwise, the inverse CDF method is preferred.

Figure 6 shows three cases of 200 samples in \mathbb{R}^3 where (i) the blue points are sampled from function space using Algorithm 7, (ii) green points are generated around the ray $(\pi/3, \pi/3)$ with angle $\pi/20$ using Algorithm 9 and Algorithm 8, and (iii) red points are generated around the ray $(\pi/6, \pi/4)$ with angle $\pi/20$ using Algorithm 9 while using Equation 9 for the inverse CDF.

6.3 The stability oracle

In § 5, we relied on the existence of a stability oracle, that given a region, in the form of the intersection a set of halfspaces, computes its stability. Here, we explain the details of that oracle. The unbiased sampling from \mathcal{U}^* enables the Monte-Carlo methods [16] for estimating the stability of the ranking regions. We will also use Monte-Carlo methods for devising randomized algorithms in § 7.

Based on Definition 2, the stability of a region R is equal the ratio of its volume to the one of the region of interest \mathcal{U}^* . Therefore, using the Monte-Carlo estimation, given a set of samples \mathcal{S} , drawn uniformly at random from \mathcal{U}^* , the stability of R can be estimated as the ratio of samples fall inside it.

Consider a region R as the intersection of a set of halfspaces. Recall from § 5 that every halfspace can be considered as an inequality in the forms of $\sum h[i]x_i < 0$ for a negative halfspace h^- and as $\sum h[i]x_i > 0$ for a positive one. Checking if a sample function with the weight vector w falls inside a region R can be done by computing the values of $\sum h[i]w_i$ for the halfspaces defining R . The function falls inside R , if it satisfies all of the constraints enforces by the halfspaces defining it.

Considering $|R|$ as the number of half-spaces defining the region R , the stability oracle is in $O(|R| |\mathcal{S}|)$. Algorithm 10 shows the pseudocode of the stability oracle.

6.4 Partitioning samples for an arrangement

The MD algorithm GET-NEXT_{md} in § 5.2 uses the function *PassThrough* to check if a hyperplane h intersects with a region R . One can use linear programming and develop the function. Instead, here we propose a method that uses the set of samples \mathcal{S} that are used for stability estimation for this purpose. The idea is

Algorithm 10 S // the stability oracle

Input: A set R of halfspaces defining a region and a set \mathcal{S} of unbiased samples taken from \mathcal{U}^*

Output: The stability of R in \mathcal{U}^*

```

1: count = 0
2: for sample  $w$  in  $\mathcal{S}$  do
3:   flag = true
4:   for halfspace  $(h, \text{sign})$  in  $R$  do
5:     sum =  $\sum_{i=1}^d h[i]w_i$ 
6:     if (sign is + and sum < 0) or (sign is - and sum > 0) then
7:       flag = false; break
8:     end if
9:   end for
10:  if flag=true then count = count + 1
11: end for
12: return  $\frac{\text{count}}{|\mathcal{S}|}$ 

```

that h intersects R , if there exists two points w and w' in R such that w belongs to h^- and w' is in h^+ , i.e. $\sum_{i=1}^d h[i]w_i < 0$ and $\sum_{i=1}^d h[i]w'_i > 0$. Since the sample points are used for estimating the stability of the regions, while we are interested in finding the stable regions, \mathcal{S} can also be used for checking if h intersects R . To do so, we partition the samples \mathcal{S} between the current regions of the arrangement, and for a hyperplane h , we check the existence of two points in the given region R that fall in two sides of h .

Interestingly, despite the complexity of the arrangement and independent from the number of dimensions, a one-dimensional array can be used for partitioning the points. Recall that the arrangement construction is iterative, and that every time a hyperplane h is added to a region R , the algorithm partitions it into two sides: the intersection of R with h^- and its intersection with h^+ .

Consider the set \mathcal{S} of samples, drawn from \mathcal{U}^* . Adding the first hyperplane $H[1]$, partitions \mathcal{S} to $\mathcal{S} \cap H[1]^-$ and $\mathcal{S} \cap H[1]^+$. We can apply the famous *partition* algorithm, used in quick-sort [26], and partition the samples in two sets such that all the samples in the range $[\mathcal{S}[1], \mathcal{S}[i]]$ belong to $R_1 = \{H[1]^-\}$ and all the ones with indices $i + 1$ to $|\mathcal{S}|$ belong to $R_2 = \{H[1]^+\}$. We use the fields *sb* and *se* in the Region data structure (Figure 2) to identify the ranges of samples that fall into the region. Interestingly, for every region R , applying the partition algorithm on the points between $R.sb$ and $R.se$ for a hyperplane h both checks if h intersects with R and adds it to R in the case of intersection. If the output index i from the partition algorithm is $R.sb - 1$ or $R.se$, h does not intersect R , otherwise the samples are already partitioned to $R_l = R \cap \{H[1]^-\}$ and $R_r = R \cap \{H[1]^+\}$, while $(R_l.sb, R_l.se) = (R.sb, i)$ and $(R_r.sb, R_r.se) = (i + 1, R.se)$.

Partitioning the samples also enables the constant time computation of the stability estimation as:

$$S(R) = \frac{R.se - R.sb + 1}{|\mathcal{S}|}$$

7. RANDOMIZED GET-NEXT

In a large setting with many items the users are usually not interested in the complete ranking between all of the items. The partial ranking model [27, 28] is the natural fit for such settings, and therefore, is used as the de-facto data retrieval model in the web [29, 30]. In this model, the user is interested in the head (the top- k) of the ranked list, rather than the complete ordering. In the following, we propose a randomized algorithm that, in addition to being scalable for large settings, is applicable for the partial rankings.

As discussed earlier in § 6, the uniform sampling of the functions of a region of interest \mathcal{U}^* provides a sampling of the rankings based on the distribution of their volumes. Therefore, in this section, we use the unbiased sampler provided in § 6 and use the Monte-Carlo methods for Bernoulli random variables [16, 31–34] for designing the randomized GET-NEXT_r operator. This operator works both for finding the stable rankings in a region of interest, as well as the top- k results. In the following, we use ranking for the explanations but all the algorithms and explanations are also valid for top- k .

Monte-Carlo methods work based on repeated sampling and central limit theorem in order to solve the deterministic problems. We consider using these algorithms for numeric integration. Based on the law of large numbers [35], the mean of independent random variables can be used for approximating the integrals. That is possible, as the expected number of occurrence of each observation is proportional to its probability. At high-level the Monte-Carlo methods work as follows: first, they generate a set of large enough inputs based on a probability distribution over a domain; then they use these inputs to do the estimation and aggregate the results.

We use sampling both for discovering the rankings as well as for estimating their stability. We design the GET-NEXT_r operator to allow the user to either (i) specify the sampling budget, or (ii) the confidence interval. Each of these two approaches has their pros and cons. The running time in (i) is fixed but the error is variable. In (ii), on the other hand, the operator guarantees the output quality while the running time is not deterministic. In the following, we first explain the details for (i). Then, we show how this can be adopted for (ii).

7.1 Fixed budget

The sampler explained in § 6.2 draws functions uniformly at random from the function space. Each function is associated with a ranking. The uniform samples on the function space provide ranking samples based on their stabilities (the portion of functions in \mathcal{U}^* generating them). For each ranking $\tau \in \mathfrak{R}$, consider the distribution of drawing a function that generate it. The probability mass function of such distribution is:

$$p(\Theta; S(\tau)) = \begin{cases} S(\tau) & \nabla_{f(\Theta)}(\mathcal{D}) = \tau \\ 1 - S(\tau) & \nabla_{f(\Theta)}(\mathcal{D}) \neq \tau \end{cases} \quad (11)$$

Let the random Bernoulli variable x_τ be 1 if $\nabla_{f(\Theta)}(\mathcal{D}) = \tau$ and 0 otherwise. Recall that the mean and standard deviation of a Bernoulli distribution with the success probability of $S(\tau)$ are $\mu_\tau = S(\tau)$ and $\sigma_\tau = S(\tau)(1 - S(\tau))$. Let m_τ be the average of a set of N samples of the random variable x_τ . Then, $E[m_\tau] = S(\tau)$ and the standard deviation of samples are $s_\tau = m_\tau(1 - m_\tau)$. Based on the central limit theorem, we also know that the distribution of the sample average is $N(\mu_\tau, \frac{\sigma_\tau}{\sqrt{N}})$ – the Normal distribution with the mean μ_τ and standard deviation $\frac{\sigma_\tau}{\sqrt{N}}$. For a large value of N , we can estimate σ_τ by s_τ .

For a confidence level α , the confidence error e identifies the range $[m_\tau - e, m_\tau + e]$ such that:

$$p(m_\tau - e \leq \mu_\tau \leq m_\tau + e) = 1 - \alpha$$

Using the Z-table:

$$\begin{aligned} e &= Z(1 - \frac{\alpha}{2}) \frac{s_\tau}{\sqrt{N}} \\ &= Z(1 - \frac{\alpha}{2}) \sqrt{\frac{m_\tau(1 - m_\tau)}{N}} \end{aligned} \quad (12)$$

Using this argument, we use a set of N samples of the function space for the design of GET-NEXT_r with a fixed budget. Every time

that a GET-NEXT_r operation is called, we collect a set of N samples and use them for finding the next stable ranking and estimating its stability. In order to provide a more accurate estimation, in addition to the N new samples, it uses the aggregates of its previous runs. Algorithm 11 shows the pseudocode of GET-NEXT_r with a fixed budget.

Algorithm 11 GET-NEXT_r

Input: database \mathcal{D} , region of interest \mathcal{U}^* , top- h stable rankings \mathfrak{R}_h , hash of previous aggregates *counts*, total number of previous samples N' , confidence level α , and sampling budget N
Output: $(h + 1)$ -th stable ranking and its stability measures

```

1: for  $k = 1$  to  $N$  do
2:    $w = \text{Sample}\mathcal{U}^*(\mathcal{U}^*)$ 
3:    $\tau = \nabla_f(\mathcal{D})$ 
4:   if  $\tau \in \text{counts.keys}$  then
5:      $\text{counts}[\tau] += 1$ 
6:   else
7:      $\text{counts}[\tau] = 1$ 
8:   end if
9: end for
10: if  $\text{counts.keys} \setminus \mathfrak{R}_h = \emptyset$  then return null
11:  $\tau_{h+1} = \underset{\tau \in \text{counts.keys} \setminus \mathfrak{R}_h}{\text{argmax}} (\text{counts}[\tau])$ 
12:  $S(\tau_{h+1}) = \text{counts}[\tau_{h+1}] / (N + N')$ 
13:  $e = Z(1 - \frac{\alpha}{2}) \sqrt{\frac{S(\tau_{h+1})(1 - S(\tau_{h+1}))}{N + N'}}$ 
14: return  $\tau_{h+1}, S(\tau_{h+1}), e$ 

```

Algorithm 11 uses a hash data structure that contains the aggregates of the rankings it has observed so far. Upon calling the algorithm, it first draws N sample functions from the region of interest \mathcal{U}^* . For each sample function, the algorithm finds the corresponding ranking and checks if it has previously been discovered. If not, it adds the ranking to the hash and sets its count as 1; otherwise, it increments the count of the ranking. If the number of discovered rankings is at most h , the algorithm fails to find a new ranking and returns null. The algorithm then chooses the ranking that does not belong to top- h and has the maximum count. It computes the stability and confidence error of the ranking and returns it.

Considering a budget of N samples while finding the ranking for each sample, the running time of GET-NEXT_r is $O(Nn \log n)$.

The GET-NEXT_r with fixed budget selects the rankings from the ones it has observed. While finding the stable rankings that a relatively large portion of function space generate them is very likely, the rankings that are generated by a small portion of the region is unlikely. First, we would like to remind here that the objective is finding the stable rankings, not the rankings that are very unlikely to generate. Still, if the user is not satisfied with the discovered rankings, she may continue calling the operator to collect more samples and increase the chance of discovering the rare rankings. Next, we explain the other version of GET-NEXT_r that guarantees the confidence error. We will also explain the expected number of samples that are required for the discovery of the rankings with a specific stability.

7.2 Fixed confidence error

While the GET-NEXT_r with fixed budget consumes constant number of samples at every iteration, it does not provide a bound on the confidence error. Here, we fix the error while accepting a non-deterministic number of samples per operation. The algorithm is similar to the previous one and works based the central limit theorem. However, instead of sampling a fixed number of functions,

it continues the sampling until the confidence error for its estimate of the next stable ranking is not more than the required error e . In high-level, the algorithm first finds its estimate of the next stable ranking and computes its confidence error, if it is less than e returns the ranking, otherwise continue the sampling. Algorithm 12 shows the pseudocode of this operation.

Algorithm 12 GET-NEXT_r

Input: database \mathcal{D} , region of interest \mathcal{U}^* , top- h stable rankings \mathfrak{R}_h , hash of previous aggregates $counts$, total number of previous samples N' , confidence level α , confidence error e
Output: $(h + 1)$ -th stable ranking and its stability

```

1:  $N = 0$ 
2:  $\tau_{h+1} = \underset{\tau \in counts.keys \setminus \mathfrak{R}_h}{\operatorname{argmax}} (counts[\tau])$ 
3: while true do
4:   if  $\tau_{h+1}$  is not null then
5:      $S(\tau_{h+1}) = \frac{counts[\tau_{h+1}]}{N+N'}$ 
6:      $e' = Z(1 - \frac{\alpha}{2}) \sqrt{\frac{S(\tau_{h+1})(1-S(\tau_{h+1}))}{N+N'}}$ 
7:     if  $e' \leq e$  then
8:        $N' += N$ 
9:       return  $\tau_{h+1}, S(\tau_{h+1}), N$ 
10:    end if
11:  end if
12:   $w = \operatorname{Sample}\mathcal{U}^*(\mathcal{U}^*); N += 1$ 
13:   $\tau = \nabla_{fw}(\mathcal{D})$ 
14:  if  $\tau$  is in  $counts.keys$  then
15:     $counts[\tau] += 1$ 
16:  else
17:     $counts[\tau] = 1$ 
18:  end if
19:  if  $\tau \notin \mathfrak{R}_h$  and  $(\tau_{h+1}$  is null or  $counts[\tau] > counts[\tau_{h+1}])$  then
20:     $\tau_{h+1} = \tau$ 
21:  end if
22: end while

```

At every iteration, Algorithm 12 looks into the set of rankings it has observed so far. Therefore, it first needs to observe a ranking at least once to be able to estimate its stability. Theorem 2 provides the expected cost for observing a ranking τ with the stability of $S(\tau)$. Following this theorem, GET-NEXT_r needs to spend the expected cost of $1/S(\tau_{h+1})$ for discovering τ_{h+1} .

THEOREM 2. *Given a ranking $\tau \in \mathfrak{R}$ with the stability of $S(\tau)$, the expected number of samples required for observing it is $\frac{1}{S(\tau)}$ with the variance of $\frac{1-S(\tau)}{S(\tau)^2}$.*

PROOF. Consider the probability distribution provided in Equation 11 and the random Bernoulli variable x_τ that is 1 if $\nabla_{fw}(\mathcal{D}) = \tau$ and 0 otherwise. We want to find the likelihood of the event x_τ with the success probability of $S(\tau)$ happening first at exactly the i -th trial. This is modeled by the Geometric distribution with the following probability mass function:

$$p(i; S(\tau)) = (1 - S(\tau))^{i-1} S(\tau)$$

The mean and variance of such distribution are $\mu = \frac{1}{S(\tau)}$ and $\sigma = \frac{1-S(\tau)}{S(\tau)^2}$. These provide the expected number of samples, as well as the variance, required for observing τ for the first time. \square

Theorem 2 shows the reverse relation between the stability of a ranking and the cost for its discovery, which indicates that at the

beginning that the rankings are stable, the cost of discovering them is low; on the other hand, the cost for discovering the unstable rankings that rarely happen is high.

Algorithm 12 guarantees the confidence error e . Using the Equation 12, the expected number of samples for such guarantee for τ_{h+1} is:

$$N_{\tau_{h+1}} = S(\tau_{h+1})(1 - S(\tau_{h+1})) \left(\frac{Z(1 - \frac{\alpha}{2})}{e} \right)^2 \quad (13)$$

7.3 Partial Ranking

In the partial ranking model, instead of the complete ranking, the user is interested in the top- k items. The top- k items may either be treated as a set or a ranked list. A company that considers its top- k candidates for the on-site interview is an example of the top- k set model, whereas for a student that wants to apply for the top colleges, the ranking between the top- k colleges is important.

The challenge for finding the stable partial rankings is that their regions are not necessarily convex. That is because multiple rankings can have equal partial rankings. Let \mathfrak{R}_p be the set of rankings in \mathfrak{R} that have the same partial ranking p . Every function in the region of a ranking $\tau \in \mathfrak{R}_p$ produces the same partial ranking p . Hence, the region of p is $\bigcup_{\tau \in \mathfrak{R}_p} R(\tau)$ and is not convex. As a result, unfortunately the MD algorithm GET-NEXT_{md} is not applicable here.

Fortunately, the randomized algorithm GET-NEXT_r can be applied, as is, for the partial rankings. Instead of maintaining the counts of the complete rankings, it is enough to maintain the counts for the occurrences of partial rankings and all the subsequent results hold. In § 8, we shall show how applying the randomized GET-NEXT_r is both effective and efficient for partial rankings.

8. EXPERIMENTS

Here we validate our stability measure and evaluate the efficiency of our algorithms on three real datasets used for ranking. In particular, we study the stability of two of our datasets in § 8.2, showing that the proposed reference rankings are not stable. In § 8.3, we study the running times of our algorithms, including stability verification, the GET-NEXT problems in 2D and MD, as well as the randomized algorithm and partial ranking.

8.1 Experimental setup

Hardware and platform. The experiments were conducted on a Linux machine with a 3.8 GHz Intel Xeon processor and 64 GB memory. The algorithms were implemented using Python 2.7.

Datasets Our three datasets are CSMetrics ($d = 2$), FIFA ($d = 4$), and Blue Nile ($d = 5$), as described below:

CSMetrics [6]: CSMetrics ranks computer science research institutions based on publication metrics. For each institution, a combination of measured (M) citations and an attribute intended to capture future citations, called predicted (P), is used for ranking, according to the score function: $(M)^\alpha (P)^{1-\alpha}$, for parameter α . This score function is not linear, but under a transformation of the data in which $x_1 = \log(M)$ and $x_2 = \log(P)$ we can write an equivalent score function linearly as: $\alpha x_1 + (1 - \alpha)x_2$. The CSMetrics website uses $\alpha = .3$ as the default value, but allows other values to be selected. We use $\alpha = .3$ and restrict our attention to the top-100 institutions according to this ranking.

FIFA Rankings [36]: The FIFA World Ranking of men's national football teams is based on measures of recent performance. Specifically, the score of a team t depends on team performance

values for A_1 (current year), A_2 (past year), A_3 (two years ago), and A_4 (three years ago). The given score function, from which the reference ranking is derived, is: $t[1] + 0.5t[2] + 0.3t[3] + 0.2t[4]$. FIFA relies on these rankings for modeling the progress and abilities of the national-A soccer teams [37] and to seed important competitions in different tournaments, including the 2018 FIFA World Cup. We consider the top 100 teams in our experiments.

Blue Nile [38]: Blue Nile is the largest online diamond retailer in the world. We collected its catalog that contained 116,300 diamonds at the time of our experiments. We consider the scalar attributes Price, Carat, Depth, LengthWidthRatio, and Table for ranking. For all attributes, except Price, higher values are preferred. We normalize each value v of a higher-preferred attribute A as $(v - \min(A))/(\max(A) - \min(A))$; for a lower-preferred attribute A , we use $(\max(A) - v)/(\max(A) - \min(A))$.

8.2 Stability investigation of real datasets

For the two datasets which provide a reference ranking (CSMetrics and FIFA) we assess these rankings below, using our stability measure and our algorithmic tools.

CSMetrics: In this dataset, items are scored according to two attributes (i.e. $d = 2$). We can therefore use the GET-NEXT operator repeatedly to enumerate all feasible rankings and their stability values. While an upper bound on the number of rankings for $n = 100$ and $d = 2$ is around 5,000, the actual number of feasible rankings for this dataset is 336. Figure 7 shows the distribution of rank stability across all rankings, showing a few rankings with high stability, after which stability rapidly drops. The reference ranking is highlighted in Figure 7; using SV_{2D} , we calculated the stability of the reference ranking to be 0.0032. Notably, the reference ranking did not appear in the top-100 stable rankings (it is the 108th most stable ranking).

Maximizing stability would cause a number of changes compared with the reference ranking. For example, Cornell University is not in the top-10 universities in the reference ranking, but replaces the University of Toronto in the top-10 in the most stable ranking. One of the bigger changes in rank position is Northeastern University which improves from 40 in the reference ranking to 35 in the most stable ranking.

We also study stability for an acceptable region close to the reference ranking. We choose 0.998 cosine similarity ($\theta = \pi/50$) around the weight vector of the reference score function. There are 22 feasible rankings in this acceptable region; their stability distribution is shown in Figure 8. Even in this narrow region of interest, the reference ranking is far below the maximum stability.

FIFA Rankings: Next, we evaluate the higher-dimensional FIFA rankings that are used for important decisions such as seeding different tournaments, including the 2018 FIFA World Cup. We focus on an acceptable region defined by 0.999 cosine similarity ($\theta = \pi/100$) around the reference weights used by FIFA, i.e. $w = \langle 1, 0.5, 0.3, 0.2 \rangle$. Using the MD algorithm GET-NEXT_{md}, we conducted 100 operations to get the distribution of the top-100 stable rankings around the reference weight vector. We considered 10,000 samples drawn using Algorithm 9 for estimations. Figure 9 shows the distribution of stable rankings. First, since $d = 4$, there are many feasible rankings, even in such a narrow region of interest, with a significant drop in stability after the most stable rankings, as was the case for CSMetrics.

Perhaps the most interesting observation is that *the reference ranking did not appear in the top-100 stable rankings* (as a result it is not highlighted in Figure 9). While FIFA advertises this ranking as “a reliable measure for comparing the national A-teams” [37],

our finding questions the FIFA’s trust to such an unstable ranking for making important decisions such as seeding the world cups. To highlight an example, while Tunisia holds a higher rank than Mexico in the reference ranking, Mexico is ranked higher in the most stable ranking. This supports the many critics that have questioned the validity of FIFA rankings in the recent past. Examples of controversial rankings include Brazil at 22 in 2014, the U.S. at 4 in 2006, and Belgium at 1 in 2015.

8.3 Algorithm performance

To evaluate the efficiency of our algorithms, we use our largest dataset, *Blue Nile*, which consists of 116,300 over 5 ranking attributes. To vary the number of items, we take random samples; to vary the number of dimensions to $d = k < 5$ we project the first k attributes. We equally weight the attributes as the default function.

2D: First we study the impact of n , the database size, on the efficiency of SV_{2D} to compute the stability of the default ranking (i.e. $w = \langle 1, 1 \rangle$). We vary n from 100 to 100,000, measuring both time and the stability of the default ranking (Figure 10). As stated in § 4 computing the stability of a ranking in 2D is in $O(n)$. We find that the running time increased linearly and was only 0.12 seconds for the largest data set. We observe that the stability quickly drops from the order of 10^{-2} for $n = 100$ to less than 10^{-6} for $n = 100K$. This is because the larger the dataset is the more the number of ordering exchanges between items grows quickly as n grows, leading to many small regions and low stability measures.

Next, we study the performance of the GET-NEXT_{2D} operator under different database sizes. The first GET-NEXT call needs to perform ray sweeping in order to construct the heap of ranking regions, while the subsequent calls just remove the next most stable ranking from the heap. Therefore, in Figure 11, we separate the first call from subsequent calls. As expected, as the number of items increases, the number of ordering exchanges increase and therefore, the efficiency of the operator drops. Also, the subsequent GET-NEXT calls are significantly faster than the first one. Still, even for the largest setting (i.e., $n = 100K$), the first call of the operator took less than 10 seconds.

MD: Next we study the performance of the stability verification algorithm, SV, and the MD algorithm, GET-NEXT_{md}, for producing the stable rankings in MD. We vary the number of items (n), number of attributes (d), and width of the region of interest (θ).

First we evaluate stability verification (Figure 12). Choosing the default weight vector $w = \langle 1, 1, 1 \rangle$, while setting $d = 3$, we initiate the stability oracle with a set of 1M samples drawn from the entire function space \mathcal{U} and vary the number of items from 100 to 10K. The stability verification algorithm in MD needs to iterate over the sample points, counting those falling inside the ranking region, described as a set of $O(n)$ constraints. This took less than a minute for $n = 10K$. On the other hand, the stability of the default ranking immediately drops to near zero, even for 100 items. Compared to 2D, this is due to the increase in the complexity of function space.

Next, we evaluate the performance of the GET-NEXT_{md} operator under varying n , d , and θ (the width of \mathcal{U}^*). The default values are $n = 100$, $d = 3$ and $\theta = \pi/100$. We use a set of 100K samples from the region of interest for the measurements. Figures 13, 14, and 15 show the performance of the GET-NEXT_{md} algorithm for the top 10 stable rankings for varying (i) number of items from $n = 10$ to $n = 10K$, (ii) number of attributes from $d = 3$ to $d = 5$, and (iii) width of the region of interest from $\theta = \pi/10$ to $\pi/100$, respectively.

Overall, the running time decreases for subsequent calls. This is because the algorithm initially builds part of the arrangement (at

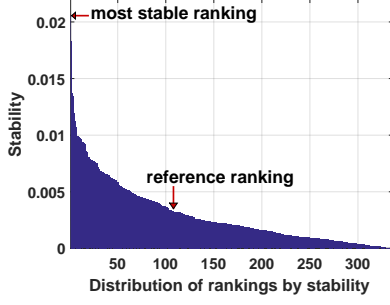


Figure 7: CSMetrics: overall distribution of rankings by Stability

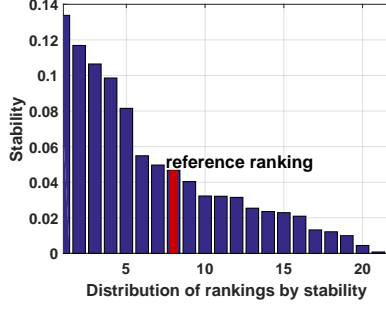


Figure 8: CSMetrics: distribution of rankings by Stability around reference weight vector $\langle 0.3, 0.7 \rangle$ with 0.998 cosine similarity.

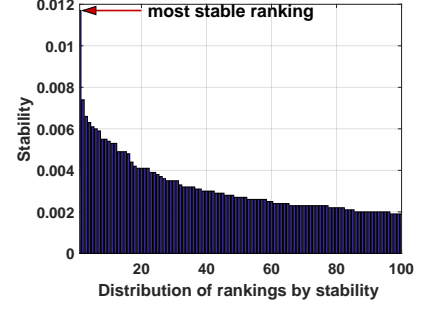


Figure 9: FIFA: distribution of rankings by Stability around reference weight vector with 0.999 cosine similarity.

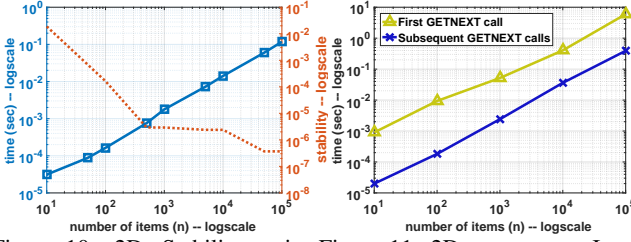


Figure 10: 2D: Stability verification, Impact of dataset size (n)

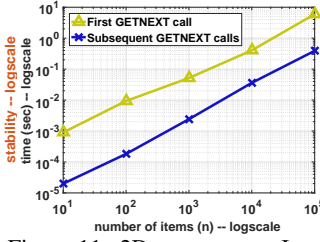


Figure 11: 2D: GET-NEXT, Impact of dataset size (n)

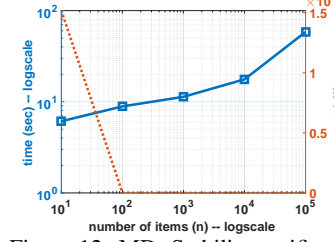


Figure 12: MD: Stability verification, Impact of dataset size

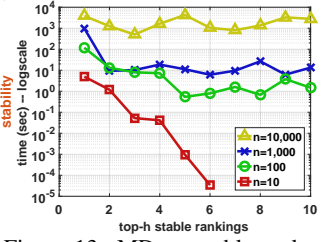


Figure 13: MD: to stable rankings, Impact of dataset size (n)

least for the stable rankings) before it finds the most stable ranking. As the arrangement gets more complete, less time is required for finishing it for the next stable region. As shown in Figure 13, every GET-NEXT call took up to several thousand seconds for the large setting of 10K items. That is because of the complexity of the arrangement of $O(n^2)$ ordering exchanges which makes even the focus on the most stable region inefficient. Note that in such complex situations all the regions are very small and unstable, as too many ordering exchanges happen in very narrow regions of interest. Nevertheless, we note that in a large setting with tens of thousands of items, it is more reasonable to consider smaller partial rankings. Our proposal for such settings is the randomized operator.

The next observation is that the running times are similar for different values of d and θ . While the complexity of the space changes for the $O(n^2)$ ordering exchanges, the search is still done using the set of 100K samples and, using the partitioning algorithm, only the subset of points that fall into a region are used for constructing the arrangement inside it. As a result, the lines in Figures 14 and 15 show similar behaviors for different settings.

Randomized algorithm: As the complexity of arrangement increases, GET-NEXT_{md} becomes less efficient. On the other hand, when the number of items is large, users may be more interested in partial rankings: that is they may focus on the top of the ranked list. In § 7, we proposed a Monte-Carlo-based randomized algorithm to handle these cases. As the last set of experiments, we evaluate the performance of the randomized algorithm under different settings. We look at two sets of partial rankings, (i) top- k partial rankings and (ii) top- k sets. In (i) the user is interested in the orderings among the top- k items, whereas in (ii) the user's interest is on the top- k sets in the ranking lists. We consider a budget of 5,000 samples (from the region of interest) for the first GET-NEXT_r call and 1,000 for subsequent calls. The default values are number of items

$n = 10,000$, number of attributes $d = 3$, the width of the region of interest $\theta = \pi/50$, and $k = 10$.

Figures 16 and 18 show the running time of the first GET-NEXT_r call and the stability of most stable ranking for varying the number of items from 1K to 100K, and the number of attributes from 3 to 5, while considering the top- k partial rankings (the running times are similar for top- k sets). The plots verify the scalability of the randomized algorithm for large settings, as it took a few minutes for 100K items while the running times for $d = 3, 4$, and 5 are similar. Looking at right-y-axis in Figure 16, despite the increase in the number of items from 1K to 100K, the stability of the most stable partial ranking did not noticeably decrease. This confirms the feasibility of using the partial rankings for the large settings.

Figures 17 and 19 shows the stability of the top-10 stable rankings for both top- k partial rankings as well as top- k sets. In both figures, the top- k sets are more stable than the top- k rankings. The reason is that the top- k sets do not consider the ordering between the items, and thus reducing the variety of possible outcomes compared to top- k rankings. An observation in Figure 17 is the similarity of the stability distributions for different numbers of items, which, again, confirms the feasibility of using partial rankings for large settings. In Figure 19, as expected, the number of attributes have a negative correlation with the stability of the partial rankings.

9. RELATED WORK

Efficient processing of the ranking queries [29, 39–43], its variations such as top- k queries [27], and finding representatives [44–46] has been extensively studied in the literature. This includes the ranking for the noisy datasets with uncertainty on the existence of items or on the attribute values [47, 48] or even using the human computation for the absence of such values [10]. However, to the

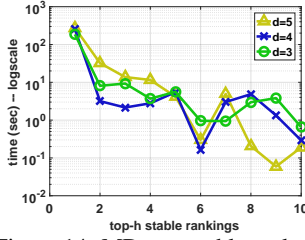


Figure 14: MD: top stable rankings, Impact of number of attributes (d)

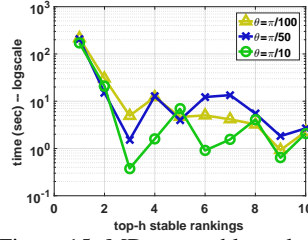


Figure 15: MD: top stable rankings, Impact of width of region of interest (θ)

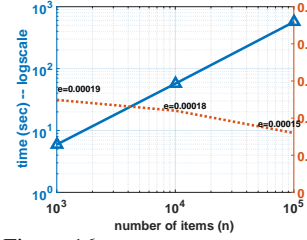


Figure 16: GET-NEXT_r: top stable partial rankings, Impact of dataset size (n)

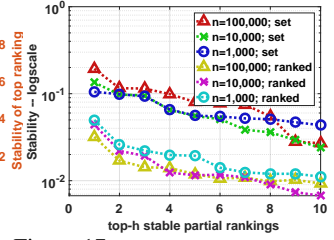


Figure 17: GET-NEXT_r: top stable partial rankings, Impact of dataset size (n) on stability

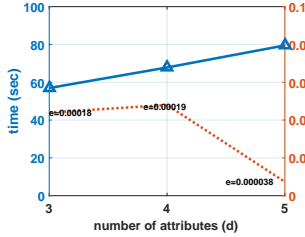


Figure 18: GET-NEXT_r: top stable partial rankings, Impact of number of attributes (d)

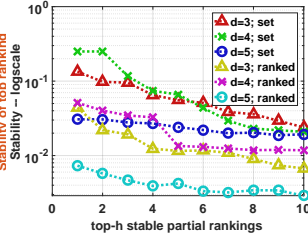


Figure 19: GET-NEXT_r: top stable partial rankings, Impact of number of attributes (d)

best of our knowledge, none of existing work considers a range of acceptable functions and discovers its stable rankings.

The work for finding outstanding items of a database with multiple attributes has two main directions: (i) when there exists a specific ranking function reflecting the user's preference, and (ii) in the absence of such preferences. Our proposal stands in the middle of these two as it is not limited to a specific function for reflecting user's preference, yet the preference is specified as a region of interest. The span of the region of interest can be as narrow as a single function or as wide as the whole function space.

The first direction includes ranking and its varieties [39–43]. Ranking in spatial databases [49] and web databases [29, 30] are a few of such varieties. There also has been extensive effort for efficient processing the top- k queries [27]. While threshold-based algorithms [28] consider parsing presorted lists along each attribute, view-based approaches [50, 51] utilize presorted lists that are built on various angles of the function space, and indexing-based methods [52] create layers of extreme points for efficient processing of queries. Ranking on probabilistic databases and noisy environments has also been the focus of works such as [47, 48, 53]. The assumption in the probabilistic databases is uncertainty on the existence of items or in the attribute values. While probabilistic databases consider the uncertainty on the data we consider it on the function reflecting the user's preference. There also has been recent attentions to ranking in the ethical data processing [8, 54, 55]. For instance, given a weight vector for ranking, [8] looks for a similar vector that provides fairness in its output.

In the absence of user preferences, the effort is on finding representatives such as skyline [9, 44, 56], also known as pareto-optimal [10], that is the set of non-dominated items. Since the size of such representatives can be large [46], works such as [17, 45, 46, 57] look for small subsets of data that minimize a measure of regret for using these sets for finding the outstanding items.

In this paper, we used notions such as half-space, duality, and arrangement from combinatorial geometry that are explained in de-

tails in [7, 58]. Arrangement of hyperplanes, its complexity, construction, and applications are studied in [7, 12–15].

10. FINAL REMARKS

In this paper, we studied the problems of obtaining stable rankings for databases with multiple attributes when the rankings are produced through a goodness score for each item as a weighted sum of its attribute values. A stable ranking is more meaningful than one susceptible to small changes in scoring weights, and hence engenders greater trust. We developed a framework that gives consumers the facility to assess the stability of a ranking and enables producers to discover stable rankings. We devised an unbiased function sampler that enables Monte-Carlo methods. We designed a randomized algorithm for the problem that works both for the complete ranking of items, as well as the top- k partial rankings. The experiments on three real datasets demonstrated the validity of our proposal.

11. REFERENCES

- [1] Malcolm Gladwell. The order of things: What college rankings really tell us. *The New Yorker Magazine*, Feb 14, 2011.
- [2] Nicholas A Bowman and Michael N Bastedo. Anchoring effects in world university rankings: exploring biases in reputation scores. *Higher Education*, 61(4):431–444, 2011.
- [3] James Monks and Ronald G Ehrenberg. The impact of us news and world report college rankings on admission outcomes and pricing decisions at selective private institutions. Technical report, National Bureau of Economic Research, 1999.
- [4] Amy Langville and Carl Meyer. *Who's #1? The Science of Rating and Ranking*. Princeton University Press, 2012.
- [5] Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, HV Jagadish, and Jerome Miklau. A nutritional label for rankings. In *ACM SIGMOD*, 2018.
- [6] CSMetrics. www.csmetrics.org/. [Online; accessed April 2018].
- [7] Herbert Edelsbrunner. *Algorithms in combinatorial geometry*, volume 10. Springer Science & Business Media, 2012.
- [8] Abolfazl Asudeh, H.V. Jagadish, Julia Stoyanovich, and Gautam Das. Designing fair ranking schemes. *CoRR*, abs/1712.09752, 2017.
- [9] Abolfazl Asudeh, Saravanan Thirumuruganathan, Nan Zhang, and Gautam Das. Discovering the skyline of web databases. *VLDB*, 2016.
- [10] Abolfazl Asudeh, Gensheng Zhang, Naeemul Hassan, Chengkai Li, and Gergely V. Zaruba. Crowdsourcing

- pareto-optimal object finding by pairwise comparisons. In *CIKM*, 2015.
- [11] Martin E. Dyer and Alan M. Frieze. On the complexity of computing the volume of a polyhedron. *SIAM Journal on Computing*, 17(5):967–974, 1988.
 - [12] Peter Orlik and Hiroaki Terao. *Arrangements of hyperplanes*, volume 300. Springer, 2013.
 - [13] Branko Grünbaum. Arrangements of hyperplanes. In *Convex Polytopes*. Springer, 2003.
 - [14] Vadim V Schechtman and Alexander N Varchenko. Arrangements of hyperplanes and lie algebra homology. *Inventiones mathematicae*, 106(1), 1991.
 - [15] Pankaj K Agarwal and Micha Sharir. Arrangements and their applications. *Handbook of computational geometry*, 2000.
 - [16] Christian P Robert. *Monte carlo methods*. Wiley Online Library, 2004.
 - [17] Abolfazl Asudeh, Azade Nazi, Nan Zhang, Gautam Das, and H.V. Jagadish. Rrr: Rank-regret representative. *CoRR*, abs/1802.10303, 2018.
 - [18] Mervin E Muller. A note on a method for generating points uniformly on n-dimensional spheres. *Communications of the ACM*, 2(4), 1959.
 - [19] George Marsaglia et al. Choosing a point from the surface of a sphere. *The Annals of Mathematical Statistics*, 43(2), 1972.
 - [20] Harald Cramér. *Mathematical methods of statistics (PMS-9)*, volume 9. Princeton university press, 2016.
 - [21] S Lucidl and M Piccioni. Random tunneling by means of acceptance-rejection sampling for global optimization. *Journal of optimization theory and applications*, 62(2):255–277, 1989.
 - [22] Luc Devroye. Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pages 260–265. ACM, 1986.
 - [23] Kaspar Fischer, Bernd Gärtner, and Martin Kutz. Fast smallest-enclosing-ball computation in high dimensions. In *European Symposium on Algorithms*, pages 630–641. Springer, 2003.
 - [24] Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, 4(1):66–70, 2011.
 - [25] GW Cran, KJ Martin, and GE Thomas. Remark as r19 and algorithm as 109: A remark on algorithms: As 63: The incomplete beta integral as 64: Inverse of the incomplete beta function ratio. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26(1):111–114, 1977.
 - [26] Thomas H Cormen. *Introduction to algorithms*. MIT press, 2009.
 - [27] Ihab F Ilyas, George Beskales, and Mohamed A Soliman. A survey of top-k query processing techniques in relational database systems. *CSUR*, 40(4), 2008.
 - [28] Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. *Journal of Computer and System Sciences*, 66(4), 2003.
 - [29] Abolfazl Asudeh, Nan Zhang, and Gautam Das. Query reranking as a service. *VLDB*, 9(11), 2016.
 - [30] Yeshwanth D. Gunasekaran, Abolfazl Asudeh, Sona Hasani, Nan Zhang, Ali Jaoua, and Gautam Das. Qr2: A third-party query reranking service over web databases. In *ICDE Demo*, 2018.
 - [31] Helge Blaker. Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics*, 28(4):783–798, 2000.
 - [32] Fred J Hickernell, Lan Jiang, Yuewei Liu, and Art B Owen. Guaranteed conservative fixed width confidence intervals via monte carlo sampling. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*, pages 105–128. Springer, 2013.
 - [33] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.
 - [34] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.
 - [35] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.
 - [36] The Fédération Internationale de Football Association (FIFA). Fifa rankings. www.fifa.com/fifa-world-ranking/ranking-table/men/index.html. [Online; accessed April 2018].
 - [37] FIFA. Fifa/coca-cola world ranking procedure. <http://www.fifa.com/fifa-world-ranking/procedure/men.html>, 28 March 2008.
 - [38] Blue Nile. www.bluenile.com/diamond-search? [Online; accessed Feb. 2018].
 - [39] Floris Geerts, Heikki Mannila, and Evimaria Terzi. Relational link-based ranking. In *VLDB*, 2004.
 - [40] Chengkai Li, Mohamed A Soliman, Kevin Chen-Chuan Chang, and Ihab F Ilyas. Ranksql: supporting ranking queries in relational database management systems. In *VLDB*, 2005.
 - [41] Surajit Chaudhuri and Gautam Das. Keyword querying and ranking in databases. *VLDB*, 2(2), 2009.
 - [42] Rakesh Agrawal, Ralf Rantza, and Evimaria Terzi. Context-sensitive ranking. In *SIGMOD*. ACM, 2006.
 - [43] Pankaj K Agarwal, Lars Arge, Jeff Erickson, Paulo G Franciosa, and Jeffrey Scott Vitter. Efficient searching with linear constraints. *JCSS*, 61(2), 2000.
 - [44] S Borzsony, Donald Kossmann, and Konrad Stocker. The skyline operator. In *ICDE*, 2001.
 - [45] Danupon Nanongkai, Atish Das Sarma, Ashwin Lall, Richard J Lipton, and Jun Xu. Regret-minimizing representative databases. *VLDB*, 2010.
 - [46] Abolfazl Asudeh, Azade Nazi, Nan Zhang, and Gautam Das. Efficient computation of regret-ratio minimizing set: A compact maxima representative. In *SIGMOD*. ACM, 2017.
 - [47] Surajit Chaudhuri, Gautam Das, Vagelis Hristidis, and Gerhard Weikum. Probabilistic ranking of database query results. In *VLDB*. VLDB Endowment, 2004.
 - [48] Jian Li, Barna Saha, and Amol Deshpande. A unified approach to ranking in probabilistic databases. *VLDB*, 2(1), 2009.
 - [49] Gisli R Hjaltason and Hanan Samet. Ranking in spatial databases. In *SSTD*. Springer, 1995.
 - [50] Vagelis Hristidis and Yannis Papakonstantinou. Algorithms and applications for answering ranked queries using ranked views. *VLDBJ*, 2004.
 - [51] Gautam Das, Dimitrios Gunopulos, Nick Koudas, and Dimitris Tsirgiannis. Answering top-k queries using views. In *VLDB*, 2006.
 - [52] Yuan-Chi Chang, Lawrence Bergman, Vittorio Castelli, Chung-Sheng Li, Ming-Ling Lo, and John R Smith. The onion technique: indexing for linear optimization queries. In

SIGMOD, 2000.

- [53] Jian Li and Amol Deshpande. Ranking continuous probabilistic datasets. *VLDB*, 3(1-2), 2010.
- [54] L. Elisa Celis, Damian Straszak, and Nisheeth K. Vishnoi. Ranking with fairness constraints. *CoRR*, abs/1704.06840, 2017.
- [55] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo A. Baeza-Yates. FA*IR: A fair top-k ranking algorithm. In *CIKM*, 2017.
- [56] Md Farhadur Rahman, Abolfazl Asudeh, Nick Koudas, and Gautam Das. Efficient computation of subspace skyline over categorical domains. In *CIKM*. ACM, 2017.
- [57] Sean Chester, Alex Thomo, S Venkatesh, and Sue Whitesides. Computing k-regret minimizing sets. *VLDB*, 7(5), 2014.
- [58] Mark De Berg, Otfried Cheong, Marc Van Kreveld, and Mark Overmars. *Computational Geometry: Introduction*. Springer, 2008.
- [59] Andrew Baker. *Matrix groups: An introduction to Lie group theory*. Springer Science & Business Media, 2012.

APPENDIX

A. COORDINATE SYSTEM ROTATION

In order to generate random functions in \mathcal{U}^* , Algorithm 9 models the region of interest as a d -spherical cap around the d -th axis. Therefore, after picking a random vector w , it needs to *rotate* the space such that the d -th axis falls on the input vector ρ . This moves the drawn sample to the region of interest. Such rotation is done via a so-called “transformation matrix” [59]. Having such a $d \times d$ rotation matrix M , the result of rotation on a vector w is a vector w' , generated as $w' = Mw$. For example in \mathbb{R}^2 , the following matrix rotates the coordinate system counterclockwise to an angle of θ :

$$M = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

We use this matrix for deriving the rotation matrix we are looking for. The idea is that instead of applying the the rotation at once, we

can do the rotation on axes separately. For example, for \mathbb{R}^3 , we first can fix the z -axis and do the rotation on the x - y plane and then fix the y -axis and do the rotation on the x - z plane.

The d by d matrix M_i , specified in Equation 14, rotates the coordinate system on the x_1 - x_{i+1} plane counterclockwise to an angle of ρ_i . All the values in M except the diameter, $M[1, i+1]$, and $M[i+1, 1]$ are zero. Also, all the values on the diameter, except $M[1, 1]$ and $M[i+1, i+1]$ are one.

$$M_i = \begin{pmatrix} 1 & 2 & \cdots & i+1 & \cdots & d \\ \cos \rho_i & 0 & \cdots & -\sin \rho_i & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \sin \rho_i & 0 & \cdots & \cos \rho_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \begin{matrix} 1 \\ 2 \\ \vdots \\ i+1 \\ \vdots \\ d \end{matrix} \quad (14)$$

The remaining point is that in order to rotate the d -th axis to ρ , on all x_1 - x_{i+1} planes except the last one the rotations are counterclockwise, while for the x_1 - x_d plane the rotation is clockwise. We change ρ_{d-1} to $(\pi/2 - \rho_{d-1})$ to make the last rotation also counterclockwise. Algorithm 13 shows the pseudocode of the function *Rotate*, used in Algorithm 9. Algorithm 13 is in $O(d^3)$, which is negligible for small values of d .

Algorithm 13 Rotate

Input: vector w and ray ρ (in form of $d-1$ angles)

Output: vector w'

- 1: $w' = w$
 - 2: $\rho_{d-1} = \pi/2 - \rho_{d-1}$
 - 3: **for** $i = d-1$ down to 1 **do**
 - 4: $M_i = \text{Equation 14}$
 - 5: $w' = M_i \times w'$
 - 6: **end for**
 - 7: **return** w'
-