Electronic Colloquium on Computational Complexity, Report No. 74 (2018)

# Generalized comparison trees for point-location problems

Daniel M. Kane* Shachar Lovett† Shay Moran‡

April 22, 2018

### Abstract

Let $H$ be an arbitrary family of hyper-planes in $d$-dimensions. We show that the point-location problem for $H$ can be solved by a linear decision tree that only uses a special type of queries called *generalized comparison queries*. These queries correspond to hyperplanes that can be written as a linear combination of two hyperplanes from $H$; in particular, if all hyperplanes in $H$ are $k$-sparse then generalized comparisons are $2k$-sparse. The depth of the obtained linear decision tree is polynomial in $d$ and logarithmic in $|H|$, which is comparable to previous results in the literature that use general linear queries.

This extends the study of comparison trees from a previous work by the authors [Kane et al., FOCS 2017]. The main benefit is that using generalized comparison queries allows to overcome limitations that apply for the more restricted type of comparison queries.

Our analysis combines a seminal result of Forster regarding sets in isotropic position [Forster, JCSS 2002], the margin-based inference dimension analysis for comparison queries from [Kane et al., FOCS 2017], and compactness arguments.

*Department of Computer Science and Engineering/Department of Mathematics, University of California, San Diego. dakane@ucsd.edu Supported by NSF CAREER Award ID 1553288 and a Sloan fellowship.

†Department of Computer Science and Engineering, University of California, San Diego. slovett@cs.ucsd.edu. Research supported by NSF CAREER award 1350481, CCF award 1614023 and a Sloan fellowship.

‡Institute for Advanced Study, Princeton. shaymoran@ias.edu. Research supported by the National Science Foundation under agreement No. CCF-1412958 and by the Simons Foundations.

ISSN 1433-8092

# 1 Introduction

Let $H \subset \mathbb{R}^d$ be a family of $|H| = n$ hyper-planes. $H$ partitions $\mathbb{R}^d$ into $O(n^d)$ cells. The *point-location* problem is to decide, given an input point $x \in \mathbb{R}^d$, to which cell it belongs. That is, to compute the function

$$\mathcal{A}_H(x) := (\text{sign}(\langle x, h\rangle) : h \in H) \in \{-1, 0, 1\}^n.$$

A well-studied computation model for this problem is a *linear decision tree* (LDT): this is a ternary decision tree whose input is $x \in \mathbb{R}^d$ and its internal nodes $v$ make linear/threshold queries of the form $\text{sign}(\langle x, q\rangle)$ for some $q = q(v) \in \mathbb{R}^d$. The three children of $v$ correspond to the three possible outputs of the query : "$-$","0","+". The leaves of the tree are labeled with $\{-1, 0, 1\}^n$ with correspondence to the cell in the arrangement that contains $x$. The complexity of a linear decision tree is its depth, which corresponds to the maximal number of linear queries made on any input.

**Comparison queries.** A *comparison decision tree* is a special type of an LDT, in which all queries are of one of two types:

- Label query: "$\text{sign}(\langle x, h\rangle) = ?$" for $h \in H$.
- Comparison query: "$\text{sign}(\langle x, h' - h''\rangle) = ?$" for $h', h'' \in H$.

In [KLMZ17] it is shown that when $H$ is "nice" then there exist comparison decision trees that computed $\mathcal{A}_H(\cdot)$ and has nearly optimal depth (up to logarithmic factors). For example, for any $H \subset \{-1, 0, 1\}^d$ there is a comparison decision tree with depth $O(d \log d \log|H|)$. This is off by a $\log d$ factor from the basic information theoretical lower bound of $\Omega(d \log|H|)$. Moreover, it is shown there that certain niceness conditions are necessary. Concretely, they give an example of $H \subset \mathbb{R}^3$ such that any comparison decision tree that computes $\mathcal{A}_H(\cdot)$ requires depth $\Omega(|H|)$. This raises the following natural problem: can comparison decision trees be generalized in a way that allows to handle arbitrary point-location problems?

**Generalized comparisons.** This paper addresses the above question by considering generalized comparison queries. A generalized comparison query allows to re-weight its terms: namely, it is query of the form

$$\text{"sign}(\langle x, \alpha h' - \beta h''\rangle) =?\text{"}$$

for $h', h'' \in H$ and some $\alpha, \beta \in \mathbb{R}$. Note that it may be assumed without loss of generality that $|\alpha|+|\beta| = 1$. A generalized comparison decision tree, naturally, is a linear decision tree whose internal linear queries are restricted to be generalized comparisons. Note that generalized comparison queries include as special cases both label queries (setting $\alpha = 1, \beta = 0$) and comparison queries (setting $\alpha = \beta = 1/2$).

Geometrically, generalized comparisons are 1-dimensional in the following sense: let $q = \alpha h' - \beta h''$, with $\alpha, \beta \geq 0$ then $q$ lies on the interval connecting $h'$ and $-h''$. If $\alpha$ and $\beta$ have different signs, $q$ lies on an interval between some other $\pm h'$ and $\pm h''$. So comparison queries are linear queries that lies on the projective lines intervals spanned by $\{\pm h : h \in H\}$. In particular, if each $h \in H$ has sparsity at most $k$ (namely, at most $k$ nonzero coordinates) then each generalized comparison has sparsity at most $2k$.

Our main result is:

**Theorem 1.1** (Main theorem)**.** *Let $H \subset \mathbb{R}^d$. Then there exists a generalized comparison decision tree of depth $O(d^4 \log d \log |H|)$ that computes $\mathcal{A}_H(x)$ for every input $x \in \mathbb{R}^d$.*

**Why consider generalized comparisons?** We consider generalized comparisons for a number of reasons:

- The lower bound against comparison queries in [KLMZ17] was achieved by essentially scaling different elements of $H \subset \mathbb{R}^3$ with exponentially different scales. Allowing for re-scaling (which is what generalized comparisons allow to do) solves this problem.
- Generalized comparisons may be natural from a machine learning perspective, in particular in the context of active learning. A common type of queries used in practice it to give a score to an example (say 1-10), and not just label it as positive (+) or negative (-). Comparing the scores for different examples can be viewed as a "coarse" type of generalized comparisons.
- If the set of original hyperplanes $H$ was "nice", then generalized comparisons maintain some aspects of niceness in the queries performed. As an example that was already mentioned, if all hyperplanes in $H$ are $k$-sparse then generalized comparisons are $2k$-sparse. This is part of a more general line of research, studying what types of "simple queries" are sufficient to obtain efficient active learning algorithms, or equivalently efficient linear decision trees for point-location problems.

## 1.1 Proof outline

Our proof consists of two parts. First, we focus on the case when $H \subset \mathbb{R}^d$ is in *general position*, namely, every $d$ vectors in it are linearly independent. Then, we extend the construction to arbitrary $H$.

The second part is fairly abstract and is derived via compactness arguments. The technical crux lies in the first part: let $H \subseteq \mathbb{R}^d$ be in general position; we first construct a randomized generalized comparison decision tree for $H$, and then derandomize it. The randomized tree is simple to describe: it proceeds by steps, where in each step about $d^2$ elements from $H$ are drawn, labelled, and sorted using generalized comparisons. Then, it is shown that the labels of some $1/d$-fraction of the remaining elements in $H$ are inferred, on average. The inferred vectors are then removed from $H$ and this step is repeated until all labels in $H$ are inferred.

A central technical challenge lies in the analysis of a single step. It hinges on a result by Forster [For02] that transforms a general-positioned $H$ to an *isotropic-positioned* $H'$ (see formal definition below) in a way that comparison queries on $H'$ correspond to generalized comparison queries on $H$. Then, since $H'$ is in isotropic position, it follows that a significant fraction of $H'$ has a large margin with respect to the input $x$. This allows us to employ a variant of the margin-based inference analysis by [KLMZ17] on $H'$ to derive the desired inference of some $\Omega(\frac{1}{d})$-fraction of the remaining labels in each step.

The derandomization of the above randomized LDT is achieved by a double-sampling argument due to [VC71]. A similar argument was used in [KLMZ17], however here several new technical challenges arise, as in each iteration in the above randomized algorithm, we only label a small fraction of the elements on average.

## 1.2 Related work

The point-location problem has been studied since the 1980s, starting from the pioneering work of Meyer auf der Heide [MadH84], Meiser [Mei93], Cardinal et al. [CIO15] and most recently Ezra and Sharir [ES17]. This last work, although not formally stated as such, solves the point-location problem for an arbitrary $H \subset \mathbb{R}^d$ by a linear decision tree whose depth is $O(d^2 \log d \log |H|)$. However, in order to do so, the linear queries used by the linear decision tree could be arbitrary, even when the original family $H$ is very simple (say 3-sparse). This is true for all previous works, as they are all based on various geometric partitioning ideas, which may require the use of quite generic hyperplanes. This should be compared with our results (Theorem 1.1). We obtain a linear decision tree of a bigger depth (by a factor of $d^2$), however the type of linear queries we use remain relatively simple; e.g., as discussed earlier, they are 1-dimensional and preserve sparseness.

## 1.3 Open problems

Our work addresses a problem raised in [KLM17], of whether "simple queries" can be sufficient to solve the point-location problem for general hyperplanes $H$, without making any "niceness" assumptions on $H$. The solution explored here is to allow for generalized comparisons, which are a 1-dimensional set of allowed queries. An intriguing question is whether this is necessary, or whether there are some 0-dimensional gadgets that would be sufficient.

In order to formally define the problem, we need the notion of *gadgets*. A $t$-ary gadget in $\mathbb{R}^d$ is a function $g : (\mathbb{R}^d)^t \to \mathbb{R}^d$. Let $G = \{g_1, \ldots, g_r\}$ be a finite collection of gadgets in $\mathbb{R}^d$. Given a set of hyperplanes $H \subset \mathbb{R}^d$, a $G$-LDT that solves $\mathcal{A}_H(\cdot)$ is a LDT where any linear query is of the form $\operatorname{sign}(\langle q, \cdot \rangle)$ for $q = g(h_1, \ldots, h_t)$ for some $g \in G$ and $h_1, \ldots, h_t \in H$. For example, a comparison decision tree corresponds to the gadgets $g_1(h) = h$ (label queries) and $g_2(h_1, h_2) = h_1 - h_2$ (comparison queries). A generalized comparison decision tree corresponds to the 1-dimensional (infinite) family of gadgets $\{g_\alpha(h_1, h_2) = \alpha h_1 - (1-\alpha) h_2 : \alpha \in [0, 1]\}$. It was shown in [KLMZ17] that comparison decision trees are sufficient to efficiently solve the point-location problem in 2 dimensions, but not in 3 dimensions. So, the problem is already open in $\mathbb{R}^3$.

**Open problem 1.** *Fix $d \geq 3$. Is there a finite set of gadgets $G$ in $\mathbb{R}^d$, such that for every $H \subset \mathbb{R}^d$ there exists a G-LDT which computes $\mathcal{A}_H(\cdot)$, whose depth is logarithmic in $|H|$? Can one hope to get to the information theoretic lower bound, namely to $O(d \log |H|)$?*

Another open problem is whether randomized LDT can always be derandomized, without losing too much in the depth. To recall, a randomized (zero-error) LDT is a distribution over (deterministic) LDTs which each computes $\mathcal{A}_H(\cdot)$. The measure of complexity for a randomized LDT is the expected number of queries performed, for the worst-case input $x$. The derandomization technique we apply in this work (see Lemma 3.9 and its proof for details) loses a factor of $d$, but it is not clear whether this loss is necessary.

**Open problem 2.** *Let $H \subset \mathbb{R}^d$. Assume that there exists a randomized LDT which computes $\mathcal{A}_H(\cdot)$, whose expected query complexity is at most $D$ for any input. Does there always exist a (deterministic) LDT which computes $\mathcal{A}_H(\cdot)$, whose depth is $O(D)$?*

# 2 Preliminaries and some basic technical lemmas

## 2.1 Linear decision trees

Let $T$ be a linear decision tree defined on input points $x \in \mathbb{R}^d$. For a vertex $u$ of $T$ let $C(u)$ denote the set of inputs $x$ whose computation path contains $u$. Let $Z(u)$ denote the queries "sign($\langle x, q\rangle$) =?" on the path from the root to $u$ that are replied by "0", and let $V(u) \leq \mathbb{R}^d$ denote the subspace $\{x : \langle x, q\rangle = 0, \forall q \in Z(u)\}$. We say that $u$ is full dimensional if $\dim(V(u)) = d$ (i.e. no query on the path towards $u$ is replied by a 0).

**Observation 2.1.** *$C(u)$ is convex (as an intersection of open halfspaces and hyperplanes).*

**Observation 2.2.** *$C(u) \subseteq V(u)$ and is open with respect to $V(u)$ (that is, it is the intersection of an open set in $\mathbb{R}^d$ with $V(u)$).*

We say that $T$ *computes* $\text{sign}(\langle h, \cdot\rangle)$ if for every leaf $\ell$ of $T$, the restriction of the function $x \mapsto \text{sign}(\langle h, x\rangle)$ to $C(\ell)$ is constant. Thus, $T$ computes $\mathcal{A}_H(\cdot)$ if and only if it computes $\text{sign}(\langle h, \cdot\rangle)$ for all $h \in H$. We say that $T$ *computes* $\text{sign}(\langle h, \cdot\rangle)$ *almost everywhere* if the restriction of $x \mapsto \text{sign}(\langle h, x\rangle)$ to $C(\ell)$ is constant, for every full dimensional leaf $\ell$.

We will use the following corollary of Observations 2.1 and 2.2. It shows that if $\text{sign}(h, \cdot)$ is not constant in $C(u)$ then it must take all three possible values. In Section 3.4, we show that a linear decision tree that computes $\mathcal{A}_H(\cdot)$ almost everywhere can be "exteneded" to a LDT that computes $\mathcal{A}_H(\cdot)$ everywhere, without increasing the depth or introducing new queries. It relies on the following lemma.

**Lemma 2.3.** *Let $u$ be a vertex in $T$, and assume that the restriction of $x \mapsto sign(\langle h, x\rangle)$ to $C(u)$ is not constant. Then there exist $x_{-1}, x_0, x_{+1} \in C(u)$ such that $sign(\langle h, x_i\rangle) = i$ for every $i \in \{-1, 0, +1\}$.*

*Proof.* Let $x', x'' \in C(u)$ with $\text{sign}(\langle h, x'\rangle) \neq \text{sign}(\langle h, x''\rangle)$. If $\{\text{sign}(\langle h, x'\rangle), \text{sign}(\langle h, x''\rangle)\} = \{\pm 1\}$ then by continuity of $x \mapsto \text{sign}(\langle h, x\rangle)$ there exists some $x_0$ on the interval between $x', x''$ such that $\text{sign}(\langle h, x_0\rangle) = 0$, and $x_0 \in C(u)$ by convexity.

Else, without loss of generality, $\text{sign}(\langle h, x'\rangle) = 0$ and $\text{sign}(\langle h, x''\rangle) = +1$. Therefore, since $C(u)$ is open relative to $V(u)$:

$$x' - \varepsilon \cdot x'' \in C(u)$$

for some small $\varepsilon > 0$. This finishes the proof since $\text{sign}(\langle h, x' - \varepsilon \cdot x''\rangle) = -1$. □

## 2.2 Inferring from comparisons

Let $x, h \in \mathbb{R}^d$ and let $S \subseteq \mathbb{R}^d$.

**Definition 2.4** (Inference)**.** *We say that $S$ infers $h$ at $x$ if sign$(\langle h, x\rangle)$ is determined by the linear queries sign$(\langle h', x\rangle)$ for $h' \in S$. That is, if for any point $y$ in the set*

$$\left\{y \in \mathbb{R}^d : \textit{sign}(\langle h', y\rangle) = \textit{sign}(\langle h', x\rangle) \quad \forall h' \in S\right\}$$

*it holds that sign$(\langle h, y\rangle)$ = sign$(\langle h, x\rangle)$. Define*

$$\mathsf{infer}(S; x) := \{h \in \mathbb{R}^d : h \textit{ is inferred from } S \textit{ at } x\}.$$

The notion of inference has a natural geometric perspective. Consider the partition of $\mathbb{R}^d$ induced by $S$. Then, $S$ infers $h$ at $x$ if the cell in this partition that contains $x$ is either disjoint from $h$ or otherwise is contained in $h$ (so in either case, the value of $\text{sign}(\langle h, \cdot\rangle)$ is constant on the cell).

Our algorithms and analysis are based on inferences from comparisons. Let $S - S$ denote the set $\{h' - h'' : h', h'' \in S\}$.

**Definition 2.5** (Inference by comparisons)**.** *We say that* comparisons on $S$ *infer $h$ at $x$ if $S \cup (S - S)$ infers $h$ at $x$. Define*

$$\mathsf{InferComp}(S; x) := \mathsf{infer}\big(S \cup (S - S); x\big).$$

Thus, $\mathsf{InferComp}(S; x)$ is determined by querying $\text{sign}(\langle h', x\rangle)$ and $\text{sign}(\langle h' - h'', x\rangle)$ for all $h', h'' \in S$. Naively, this requires some $O(|S|^2)$ linear queries. However, using efficient sorting algorithm (e.g. merge-sort) achieves it with just $O(|S| \log |S|)$ comparison queries. A further improvement, when $|S| > d$, is obtained by Fredman's sorting algorithm that uses just $O(|S| + d \log |S|)$ comparison queries [Fre76].

## 2.3 Vectors in isotropic position

Vectors $h_1, \ldots, h_m \in \mathbb{R}^d$ are said to be in *general position* if any $d$ of them are linearly independent. They are said to be in *isotropic position* if for any unit vectors $v \in S^d$,

$$\frac{1}{m} \sum_{i=1}^{m} \langle h_i, v\rangle^2 = \frac{1}{d}.$$

Equivalently, if $\frac{1}{m}\sum h_i h_i^T$ is $\frac{1}{d}$ times the $d \times d$ identity matrix. An important theorem of Forster [For02] (see also Barthe [Bar98] for a more general statement) states that any set of vectors in general position can be scaled to be in isotropic position.

**Theorem 2.6** ([For02])**.** *Let $H \subset \mathbb{R}^d$ be a finite set in general position. Then there exists an invertible linear transformation $T$ such that the set*

$$H' := \left\{ \frac{Th}{\|Th\|_2} : h \in H \right\}$$

*is in isotropic position. We refer to such a $T$ as a* Forster transformation *for $H$.*

We will also need a relaxed notion of isotropic position. Given vectors $h_1, \ldots, h_m \in \mathbb{R}^d$ and some $0 < c < 1$, we say that the vectors are in *$c$-approximate isotropic position*, if for all unit vectors $v \in S^d$ it holds that

$$\frac{1}{m}\sum_{i=1}^{m} \langle h_i, v\rangle^2 \geq \frac{c}{d}.$$

We note that this condition is easy to test algorithmically, as it is equivalent to the statement that the smallest eigenvalue of the positive semi-definite $d \times d$ matrix $\frac{1}{m}\sum_{i=1}^{m} h_i h_i^T$ is at least $\frac{c}{d}$.

We summarize it in the following lemma, which follows from basic real linear algebra.

**Claim 2.7.** *Let $h_1, \ldots, h_m \in \mathbb{R}^d$ be unit vectors. Then the following are equivalent.*

- *$h_1, \ldots, h_m$ are in $c$-approximate isotropic position.*
- *$\lambda_1\left(\frac{1}{m}\sum_{i=1}^{m} h_i h_i^T\right) \geq c/d$,*

*where $\lambda_1(M)$ denotes the minimal eigenvalue of a positive semidefinite matrix $M$.*

We will need the following basic claims. The first claim shows that a set of unit vectors in an approximate isotropic position has many vectors with non-negligible inner product with any unit vector.

**Claim 2.8.** *Let $h_1, \ldots, h_m \in \mathbb{R}^d$ be unit vectors in a $c$-approximate isotropic position, and let $x \in \mathbb{R}^d$ be a unit vector. Then, at least a $\frac{c}{2d}$-fraction of the $h_i$'s satisfy $|\langle h_i, x\rangle| > \sqrt{\frac{c}{2d}}$.*

*Proof.* Assume otherwise. It follows that

$$\frac{1}{m}\sum_{i=1}^{m} |\langle h, x_i\rangle|^2 \leq \frac{c}{2d}\cdot 1 + \left(1 - \frac{c}{2d}\right)\frac{c}{2d} < \frac{c}{2d} + \frac{c}{2d} = \frac{c}{d}.$$

This contradicts the assumption that the $h_i$'s are in $c$-approximate isotropic position. □

The second claim shows that a random subset of a set of unit vectors in an approximate isotropic position is also in approximate isotropic position, with good probability.

**Claim 2.9.** *Let $h_1, \dots, h_m$ be unit vectors in $c$-approximate isotropic position. Let $i_1, \dots, i_k \in [m]$ be independently and uniformly sampled. Then for any $\delta > 0$, the vectors $h_{i_1}, \dots, h_{i_k}$ are in $((1-\delta)c)$-approximate isotropic position with probability at least*

$$1 - d \cdot \left[ \frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right]^{ck/d}.$$

*Proof.* This is an immediate corollary of Matrix Chernoff bounds [Tro12]. By Claim 2.7 the above event is equivalent to that $\lambda_1\left(\frac{1}{k}\sum_{i=1}^{k} h_i h_i^T\right) \geq (1-\delta)\frac{c}{d}$. By assumption, $\lambda_1\left(\frac{1}{m}\sum_{i=1}^{m} h_i h_i^T\right) \geq \frac{c}{d}$. Now, by the Matrix Chernoff bound, for any $\delta \in [0,1]$ it holds that

$$\Pr\left[\lambda_1\left(\frac{1}{k}\sum_{i=1}^{k} h_i h_i^T\right) \leq (1-\delta) \cdot \frac{c}{d}\right] \leq d \cdot \left[ \frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right]^{ck/d}.$$

□

We will use two instantiations of Claim 2.9: (i) $c \geq 3/4$, and $(1-\delta)c = 1/2$, and (ii) $c = 1$ and $(1-\delta)c = 3/4$. In both cases the bound simplifies to

$$1 - d \cdot \left(\frac{99}{100}\right)^{k/d}. \tag{1}$$

# 3 Proof of main theorem

Let $H \subset \mathbb{R}^d$. We prove Theorem 1.1 in four steps:

1. First, we assume that $H$ is in general position. In this case, we construct a randomized generalized comparison LDT which computes $\mathcal{A}_H(\cdot)$, whose expected depth is $O(d^3 \log d \log |H|)$ for any input. This is achieved in Section 3.1, see Lemma 3.1.

2. Next, we derandomize the construction. This gives for any $H$ in general position a (deterministic) generalized comparison LDT which computes $\mathcal{A}_H(\cdot)$, whose depth is $O(d^4 \log d \log |H|)$. This is achieved in Section 3.2, see Lemma 3.9.

3. In the next step, we handle an arbitrary $H$ (not necessarily in general position), and construct by a compactness argument a generalized comparisons LDT of depth $O(d^4 \log d \log |H|)$ which computes it almost everywhere. This is achieved in Section 3.3, see Lemma 3.14.

4. Finally, we show that any LDT which computes $\mathcal{A}_H(\cdot)$ almost everywhere can be "fixed" to a LDT which computes $\mathcal{A}_H(\cdot)$ everywhere. This fixing procedure maintains both the depth of the LDT, as well as the set of queries performed by it. This is achieved in Section 3.4, see Lemma 3.15.

## 3.1 A randomized LDT for $H$ in general position

In this section we construct a randomized generalized comparison LDT for $H$ in general position. Here, by a randomized LDT we mean a distribution over (deterministic) LDT which compute $\mathcal{A}_H(\cdot)$. The corresponding complexity measure is the expected number of queries it makes, for the worst-case input $x$.

**Lemma 3.1.** *Let $H \subseteq \mathbb{R}^d$ be a finite set in general position. Then there exists a randomized LDT that computes $\mathcal{A}_H(\cdot)$, which makes $O\left(d^3 \log d \log|H|\right)$ generalized comparison queries on expectation, for any input.*

The proof of Lemma 3.1 is based on a variant of the margin-based analysis of the inference dimension with respect to comparison queries as in [KLMZ17] (The analysis in [KLMZ17] assumed that all vectors have large margin, where here we need to work under the weaker assumption that only a noticeable fraction of the vectors have large margin). The crux of the proof relies on scaling every $h \in H$ by a carefully chosen scalar $\alpha_h$ such that drawing a sufficiently large random subset of $H$, and sorting the values $\langle \alpha_h h, x \rangle$ using comparison queries (which correspond to generalized comparisons on the $h$'s) allows to infer, on average, at least $\Omega(1/d)$ of the labels of $H$. The scalars $\alpha_h$ are derived via Forster's theorem (Theorem 2.6). More specifically, $\alpha_h = \frac{1}{\|Th\|_2}$, where $T$ is a Forster transformation for $H$.

```
Randomized generalized-comparisons tree for $H$ in general position

Let $H \subseteq \mathbb{R}^d$ in general position.

Input: $x \in \mathbb{R}^d$, given by oracle access for $\mathrm{sign}(\langle \cdot, x\rangle)$
Output: $\mathcal{A}_H(x) = (\mathrm{sign}(\langle h, x\rangle))_{h\in H}$

  (1) Initialize: $H_0 = H$, $i = 0$, $v(h) =?$ for all $h \in H$. Set $k = \Theta(d^2 \log(d))$.
  (2) Repeat while $|H_i| \geq k$:
      (2.1) Let $T_i$ be the Forster transformation for $H_i$. Define $H'_i = \left\{\frac{h}{\|T_i h\|_2} : h \in H_i\right\}$.
      (2.2) Sample uniformly $S_i \subset H'_i$ of size $|S_i| = k$.
      (2.3) Query $\mathrm{sign}(\langle h, x\rangle)$ for $h \in S_i$ (using label queries).
      (2.4) Sort $\langle h, x\rangle$ and $\langle -h, x\rangle$ for $h \in S_i$ (using generalized comparison queries).
      (2.5) For all $h \in H_i$, check if $h \in \mathsf{InferComp}(\pm S_i; x)$, and in case it is, set $v(h) \in$
            $\{-, 0, +\}$ to be the inferred value of $h$.
      (2.6) Remove all $h \in H_i$ for which $\mathrm{sign}\left(\langle h, x\rangle\right)$ was inferred, set $H_{i+1}$ to be the
            resulting set and go to step (2).
  (3) Query $\mathrm{sign}(\langle h, x\rangle)$ for all $h \in H_i$, and set $v(h)$ accordingly.
  (4) Return $v$ as the value of $\mathcal{A}_H(x)$.
```

In order to understand the intuition behind the main iteration (2) of the algorithm, define $x' = (T_i^{-1})^{\mathrm{T}} x$ and for each $h \in H_i$ let $h' = \frac{T_i h}{\|T_i h\|}$. Then $\mathrm{sign}(\langle h, x\rangle) = \mathrm{sign}(\langle h', x'\rangle)$, and so it suffices to infer the sign for many $h' \in H_i$ with respect to $x'$. The main benefit is that we may assume in the analysis that the set of vectors $H'_i$ is in isotropic position; and reduce the analysis to that of using (standard) comparisons on $H'_i$ and $x'$. These then translate to performing generalized comparison queries on $H_i$ and the original input $x$. The following lemma captures the analysis of the main iteration of the algorithm. Below, we denote by $\pm S := S \cup (-S)$.

**Lemma 3.2.** *Let $x \in \mathbb{R}^d$, let $H \subseteq \mathbb{R}^d$ be a finite set of unit vectors in c-approximate isotropic position with $c \geq 3/4$, and let $S \subset H$ be a uniformly chosen subset of size $k = \Omega\left(d^2 \log d\right)$. Then*

$$\mathbb{E}_S\left[|\mathsf{InferComp}(\pm S; x) \cap H|\right] \geq \frac{|H|}{40d}.$$

Note that this proves a stronger statement than needed for Lemma 3.1. Indeed, it would suffice to consider only $H$ that is in (a complete) isotropic position. This stronger version will be used in the next section for derandomizing the above algorithm. Let us first argue how Lemma 3.1 follows from Lemma 3.2, and then proceed to prove Lemma 3.2.

*Proof of Lemma 3.1 given Lemma 3.2.* By Lemma 3.2, in each iteration (2) of the algorithm, we infer on expectation at least $\Omega(1/d)$ fraction of the $h \in H'_i$ with respect to $x' = T_i^{-1}x$. By the discussion above, this is the same as inferring an $\Omega(1/d)$ fraction of the $h_i \in H_i$ with respect to $x$. So, the total expected number of iterations needed is $O(d \log |H|)$. Next, we calculate the number of linear queries performed at each iteration. The number of label queries is $O(k)$ and the number of comparison queries on $H'_i$ (which translate to generalized comparison queries on $H_i$) is $O(k \log k)$ if we use merge-sort, and can be improved to $O(k + d \log k)$ by using Fredman's sorting algorithm [Fre76]. So, in each iteration we perform $O(d^2 \log d)$ queries, and the expected number of iterations is $O(d \log |H|)$. So the expected total number of queries by the algorithm is $O(d^3 \log d \log |H|)$. □

From now on, we focus on proving Lemma 3.2. To this end, we assume from now that $H \subset \mathbb{R}^d$ is in $c$-isotropic position for $c \geq 3/4$. Note that $h$ is inferred from comparisons on $\pm S$ if and only if $-h$ is, and that replacing an element of $S$ with its negation does not affect $\pm S$. Therefore, negating elements of $H$ does not change the expected number of elements inferred from comparisons on $\pm S$. Therefore, we may assume in the analysis that $\langle h, x \rangle \geq 0$ for all $h \in H$. Under this assumption, we will show that

$$\mathbb{E}_S \left[ |\mathsf{InferComp}(S; x) \cap H| \right] \geq \frac{|H|}{40d}.$$

It is convenient to analyze the following procedure for sampling $S$:

- Sample $h_1, \ldots h_{k+1}$ random points in $H$, and $r \in [k+1]$ uniformly at random.
- Set $S = \{h_j : j \in [k+1] \setminus \{r\}\}$.

We will analyze the probability that comparisons on $S$ infer $h_r$ at $x$. Our proof relies on the following observation.

**Observation 3.3.** *The probability, according to the above process, that $h_r \in \mathsf{InferComp}(S; x)$ is equal to the expected fraction of $h \in H$ whose label is inferred. That is,*

$$\Pr \left[ h_r \in \mathsf{InferComp}(S; x) \right] = \mathbb{E} \left[ \frac{|\mathsf{InferComp}(S; x) \cap H|}{|H|} \right].$$

Thus, it suffices to show that $\Pr [h_r \in \mathsf{InferComp}(S; x)] \geq 1/40d$. This is achieved by the next two propositions as follows. Proposition 3.4 shows that $S$ is in a (1/2)-approximate isotropic position with probability at least 1/2, and Proposition 3.5 shows that whenever $S$ is in (1/2)-approximate isotropic position then $h_r \in \mathsf{InferComp}(S; x)$ with probability at least $1/20d$. Combining these two propositions together yields that $\Pr [h_r \in \mathsf{InferComp}(S; x)] \geq 1/40d$ and finishes the proof of Lemma 3.2.

**Proposition 3.4.** *Let $H \subset \mathbb{R}^d$ be a set of unit vectors in c-approximate isotropic position for $c \geq 3/4$. Let $S \subset H$ be a uniformly sampled subset of size $|S| \geq \Omega(d \log d)$. Then $S$ is in $(1/2)$-approximate isotropic position with probability at least $1/2$.*

*Proof.* The proof follows from Claim 2.9 by plugging $k = \Omega(d \log d)$ in Equation (1) and calculating that the bound on the right hand side becomes at least 1/2. □

**Proposition 3.5.** *Let $x \in \mathbb{R}^d$, $S \subset \mathbb{R}^d$ be in (1/2)-approximate isotropic position, where $|S| \geq \Omega\left(d^2 \log d\right)$. Let $h \in S$ be sampled uniformly. Then*

$$\Pr_{h \in S}\left[h \in \mathsf{InferComp}\left(S \setminus \{h\}; x\right)\right] \geq \frac{1}{20d}.$$

*Proof.* We may assume that $x$ is a unit vector, namely $\|x\|_2 = 1$. Let $s = |S|$ and assume that $S = \{h_1, \ldots, h_s\}$ with

$$\langle h_1, x \rangle \geq \langle h_2, x \rangle \geq \ldots \geq \langle h_s, x \rangle \geq 0.$$

Set $\varepsilon = \frac{1}{2\sqrt{d}}$. As $S$ is in (1/2)-approximate isotropic position, Claim 2.8 gives that $\langle h_i, x \rangle \geq \varepsilon$ for at least $|S|/4d$ many $h_i \in S$. Set $t = |S|/8d$ and define

$$T = \{h_1, \ldots, h_t\},$$

where by out assumption $\langle h_t, x \rangle \geq \varepsilon$. Note that in this case, we can compute $T$ from comparison queries on $S$. We will show that

$$\Pr_{h \in T}\left[h \in \mathsf{InferComp}\left(S \setminus \{h\}; x\right)\right] \geq \frac{1}{2},$$

from which the proposition follows. This in turn follows by the following two claims, whose proof we present shortly.

**Claim 3.6.** *Let $h_a \in T$. Assume that there exists a non-negative linear combination $v$ of $\{h_i - h_{i+1} : i = 1, \ldots, a-2\}$ such that*

$$\|h_a - (h_1 + v)\|_2 \leq \varepsilon/4.$$

*Then $h_a \in \mathsf{InferComp}\left(S \setminus \{h_a\}; x\right)$.*

**Claim 3.7.** *The assumption of Claim 3.6 holds for at least half the vectors in $T$.*

Clearly, Claim 3.6 and Claim 3.7 together imply that for at least half of $h_a \in T$, it holds that $h_a \in \mathsf{InferComp}\left(S \setminus \{h_a\}; x\right)$. This concludes the proof of the proposition. □

Next we prove Claim 3.6 and Claim 3.7.

*Proof of Claim 3.6.* Let $S' = S \setminus \{h_a\}$ and $T' = T \setminus \{h_a\}$. As $S$ is in (1/2)-approximate isotropic position then $S'$ is in $c$-approximate isotropic position for $c = 1/2 - d/|S|$. In particular, as $|S| \geq 4d$ we have $c \geq 1/4$. By applying comparison queries to $S'$ we can sort $\{\langle h_i, x \rangle : h_i \in S'\}$. Then $T'$ can be computed as the set of the $t-1$ elements with the largest inner product. Claim 2.8 applied to $S'$ then implies that $\langle h_i, x \rangle \geq \varepsilon/2$ for all $h_i \in T'$. Crucially, we can deduce this just from the comparison queries on $S'$, together with our initial assumption that $S$ is in (1/2)-approximate isotropic position. Thus we deduced from our queries that:

- $\langle h_1, x\rangle \geq \varepsilon/2$.
- $\langle v, x\rangle \geq 0$.

In addition, from our assumption it follows that $|\langle h_a - (h_1 + v), x\rangle| \leq \varepsilon/4$. These together infer that $\langle h_a, x\rangle > 0$. □

The proof of Claim 3.7 follows from the applying the following claim iteratively. We note that this claim appears in [KLMZ17] implicitly, but we repeat it here for clarity.

**Claim 3.8.** *Let $h_1, \ldots, h_t \in \mathbb{R}^d$ be unit vectors. For any $\varepsilon > 0$, if $t \geq 16d \ln(2d/\varepsilon)$ then there exist $a \in [t]$ and $\alpha_1, \ldots, \alpha_{a-2} \in \{0, 1, 2\}$ such that*

$$h_a = h_1 + \sum_{j=1}^{i-2} \alpha_j (h_{j+1} - h_j) + e,$$

*where $\|e\|_2 \leq \varepsilon$.*

In order to derive Claim 3.7 from Claim 3.8, we assume that $|T| \geq 32d \ln((2d)/(\varepsilon/4)) = \Omega(d \log d)$. Then we can apply Claim 3.8 iteratively $|T|/2$ times with parameter $\varepsilon/4$, at each step identify the required $h_a$, remove it from $T$ and continue. Next we prove Claim 3.8.

*Proof of Claim 3.8.* Let $\mathbb{B} := \{h \in \mathbb{R}^d : \|h\|_2 \leq 1\}$ denote the Euclidean ball of radius 1, and let $C$ denote the convex hull of $\{h_2 - h_1, \ldots, h_t - h_{t-1}\}$. Observe that $C \subset 2\mathbb{B}$, as each $h_i$ is a unit vector. For $\beta \in \{0, 1\}^{t-1}$ define

$$h_\beta = \sum \beta_j (h_{j+1} - h_j).$$

We claim that having $t \geq 16d \ln(2d/\varepsilon)$ guarantees that there exist distinct $\beta', \beta''$ for which

$$h_{\beta'} - h_{\beta''} \in \frac{\varepsilon}{4}(C - C).$$

This follows by a packing argument: if not, then the sets $h_\beta + \frac{\varepsilon}{4}C$ for $\beta \in \{0, 1\}^{t-1}$ are mutually disjoint. Each has volume $(\varepsilon/4)^d \mathrm{vol}(C)$, and they are all contained in $tC$ which has volume $t^d \mathrm{vol}(C)$. As the number of distinct $\beta$ is $2^{t-1}$ we obtain that $2^{t-1}(\varepsilon/4)^d \leq t^d$, which contradicts our assumption on $t$.

Let $i \in [t]$ be maximal such that $\beta'_{i-1} \neq \beta''_{i-1}$. We may assume without loss of generality that $\beta'_{i-1} = 0, \beta''_{i-1} = 1$, as otherwise we can swap the roles of $\beta'$ and $\beta''$. Thus we have

$$\sum_{j=1}^{i-1} (\beta'_j - \beta''_j)(h_{j+1} - h_j) \in \frac{\varepsilon}{4}(C - C) \subset \varepsilon\mathbb{B}.$$

Adding $h_i - h_1 = \sum_{j=1}^{i-1} (h_{j+1} - h_j)$ to both sides gives

$$\sum_{j=1}^{i-1} (\beta'_j - \beta''_j + 1)(h_{j+1} - h_j) \in h_i - h_1 + \varepsilon\mathbb{B},$$

which is equivalent to

$$h_i - h_1 \in \sum_{j=1}^{i-1} (\beta'_j - \beta''_j + 1)(h_{j+1} - h_j) + \varepsilon \mathbb{B}.$$

The claim follows by setting $\alpha_j = \beta'_j - \beta''_j + 1$ and noting that by our construction $\alpha_{i-1} = 0$, and hence the sum terminates at $i - 2$. □

## 3.2 A deterministic LDT for $H$ in general position

In this section, we derandomize the algorithm from the previous section. We still assume that $H$ is in general position, this assumption will be removed in the next sections.

**Lemma 3.9.** *Let $H \subseteq \mathbb{R}^d$ be a finite set in general position. Then there exists an LDT that computes $\mathcal{A}_H(\cdot)$ with $O\left(d^4 \log d \log|H|\right)$ generalized comparison queries.*

Note that the this bound is worse by a factor of $d$ than the one in Lemma 3.1. In Open problem 2 we ask whether this loss is necessary, or whether it can be avoided by a different derandomization technique.

Lemma 3.9 follows by derandomizing the algorithm from Lemma 3.1. Recall that Lemma 3.1 boils down to showing that $h \in \mathsf{InferComp}(S_i; x)$ for an $\Omega(1/d)$ fraction of $h \in H_i$ on average. In other words, for every input vector $x$, *most of the subsets $S_i \subseteq H'_i$* of size $\Omega(d^2 \log d)$ allow to infer from comparisons the labels of some $\Omega(1/d)$-fraction of the points in $H_i$. We derandomize this step by showing that *there exists a universal set $S_i \subseteq H'_i$* of size $O(d^3 \log d)$ that allows to infer the labels of some $\Omega(1/d)$-fraction of the points in $H_i$, with respect to *any* $x$. This is achieved by the next lemma.

**Lemma 3.10.** *Let $H \subseteq \mathbb{R}^d$ be a set of unit vectors in isotropic position. Then there exists $S \subseteq H$ of size $O(d^3 \log d)$ such that*

$$\left(\forall x \in \mathbb{R}^d\right) : |\mathsf{InferComp}(S; x) \cap H| \geq \frac{|H|}{100d}.$$

*Proof.* We use a variant of the double-sampling argument due to [VC71] to show that a random $S \subseteq H$ of size $s = O(d^3 \log d)$ satisfies the requirements. Let $S = \{h_1, \ldots, h_s\}$ be a random (multi-)subset of size $s$, and let $E = E(S)$ denote the event

$$E(S) := \left[\exists x \in \mathbb{R}^d : |\mathsf{InferComp}(S; x) \cap H| < |H|/100d\right].$$

Our goal is showing that $\Pr[E] < 1$. To this end we introduce an auxiliary event $F$. Let $t = \Theta(d^2 \log d)$, and let $T = \{h_1, \ldots, h_t\} \subseteq S$ be a subsample of $S$, where each $h_i$ is drawn uniformly from $S$ and independently of the others. Define $F = F(S, T)$ to be the event

$$F(S,T) := \Big[\exists x \in \mathbb{R}^d : |\mathsf{InferComp}(T; x) \cap H| < |H|/100d \text{ and } |\mathsf{InferComp}(T; x) \cap S| \geq |S|/50d\Big].$$

The following claims conclude the proof of Lemma 3.10.

**Claim 3.11.** *If* $\Pr[E] \geq 9/10$ *then* $\Pr[F] \geq 1/200d$.

**Claim 3.12.** $\Pr[F] \leq 1/250d$.

This concludes the proof, as it shows that $\Pr[E] < 9/10$. We next move to prove Claim 3.11 and Claim 3.12.

*Proof of Claim 3.11.* Assume that $\Pr[E] \geq 9/10$. Define another auxiliary event $G = G(S)$ as

$$G(S) := [S \text{ is in } (3/4)\text{-approximate isotropic position}] .$$

Applying Claim 2.9 by plugging $m \geq 100d \ln(10d)$ in Equation (1) gives that $\Pr[G] \geq 9/10$, which implies that $\Pr[E \wedge G] \geq 8/10$. Next, we analyze $\Pr[F | E \wedge G]$.

To this end, fix $S$ such that both $E(S)$ and $G(S)$ hold. That is: $S$ is in (3/4)-approximate isotropic position, and there exists $x = x(S) \in \mathbb{R}^d$ such that $|\mathsf{InferComp}(S; x) \cap H| < |H|/100d$. If we now sample $T \subset S$, in order for $F(S, T)$ to hold, we need that (i) $|\mathsf{InferComp}(T; x) \cap H| < |H|/100d$ , which holds with probability one, as $|\mathsf{InferComp}(S; x) \cap H| < |H|/100d$; and (ii) that $|\mathsf{InferComp}(T; x) \cap S| \geq |S|/50d$. So, we analyze this event next.

Applying Lemma 3.2 to the subsample $T$ with respect to $S$ gives that

$$\mathbb{E}_T \left[ |\mathsf{InferComp}(T; x) \cap S| \right] \geq |S|/40d.$$

This then implies that

$$\Pr \left[ |\mathsf{InferComp}(T; x) \cap S| \geq |S|/100d \right] \geq 1/100d.$$

To conclude: we proved under the assumptions of the lemma that $\Pr_S[E(S) \wedge G(S)] \geq 8/10$; and that for every $S$ which satisfies $E(S) \wedge G(S)$ it holds that $\Pr_T[F(S,T)|S] \geq 1/100d$. Together these give that $\Pr[F(S, T)] \geq 1/200d$. □

*Proof of Claim 3.12.* We can model the choice of $(S, T)$ as first sampling $T \subset H$ of size $t$, and then sampling $S \setminus T \subset H$ of size $s - t$. We will prove the following (stronger) statement: for any choice of $T$,

$$\Pr \left[ F(S, T) | T \right] < 1/250d.$$

So from now on, fix $T$ and consider the random choice of $T' = S \setminus T$. We want to show that:

$$\Pr_{T'} \Big[ (\exists x \in \mathbb{R}^d) : |\mathsf{InferComp}(T; x) \cap H| < |H|/100d \text{ and } |\mathsf{InferComp}(T; x) \cap S| \geq |S|/50d. \Big] \leq 1/250d.$$

We would like to prove this statement by applying a union bound over all $x \in \mathbb{R}^d$. However, $\mathbb{R}^d$ is an infinite set and therefore a naive union seems problematic. To this end we introduce a suitable equivalence relation that is based on the following observation.

**Observation 3.13.** $\mathsf{InferComp}(T; x)$ *is determined by sign*$(\langle h, x \rangle)$ *for* $h \in T \cup (T - T)$.

We thus define an equivalence relation on $\mathbb{R}^d$ where $x \sim y$ if and only if $\mathrm{sign}(\langle h, x\rangle) = \mathrm{sign}(\langle h, y\rangle)$ for all $h \in T \cup (T - T)$. Let $C$ be a set of representatives for this relation. Thus, it suffices to show that

$$\Pr_{T'}\Big[(\exists x \in C) : |\mathsf{InferComp}(T; x) \cap H| < |H|/100d \text{ and}\\ |\mathsf{InferComp}(T; x) \cap S| \geq |S|/50d.\Big] \leq 1/250d.$$

Since $C$ is finite, a union bound is now applicable. Sepcifically, it is enough to show that

$$(\forall x \in C) : \Pr_{T'}\Big[\, |\mathsf{InferComp}(T; x) \cap H| < |H|/100d \text{ and}\\ |\mathsf{InferComp}(T; x) \cap S| \geq |S|/50d.\Big] \leq \frac{1}{250d|C|}.$$

Now, (a variant of) Sauer's Lemma (see e.g. Lemma 2.1 in [KLM17]) implies that

$$|C| \leq (2e \cdot |T \cup (T - T)|)^d \leq \left(2e \cdot t^2\right)^d \leq (20t)^{2d}. \tag{2}$$

Fix $x \in C$. If $|\mathsf{InferComp}(T; x) \cap H| \geq \frac{|H|}{100d}$ then we are done (note that $\mathsf{InferComp}(T; x)$ is fixed since it depends only on $T$ and $x$ and not on $T'$). So, we may assume that $|\mathsf{InferComp}(T; x) \cap H| < \frac{|H|}{100d}$. Then we need to bound

$$\Pr\left[|\mathsf{InferComp}(T; x) \cap S| \geq \frac{|S|}{50d}\right] \leq \Pr\left[|\mathsf{InferComp}(T; x) \cap T'| \geq \frac{|T'|}{75d}\right],$$

where the inequality follows if $t \leq \frac{s}{150d}$, which can be satisfied since $t = \Theta(d^2 \log d)$ and $s = \Theta(d^3 \log d)$. To bound this probability we use the Chernoff bound: let $p = \frac{|\mathsf{InferComp}(T;x) \cap H|}{|H|}$; note that $|\mathsf{InferComp}(T; x) \cap T'|$ is distributed like $\mathsf{Bin}(s - t, p)$. By assumption, $p \leq \frac{1}{100d}$, and therefore:

$$\Pr\left[|\mathsf{InferComp}(T; x) \cap T'| \geq \frac{|T'|}{75d}\right] \leq \exp\left(-\frac{(1/3)^2 \cdot (t/100d)}{3}\right) \leq \frac{1}{250d \cdot (20t)^{2d}} \leq \frac{1}{250d \cdot |C|},$$

where the second inequality follows because $t = \Theta(d^2 \ln(d))$ with a large enough constant, and the last inequality follows by Section 3.2. □

□

## 3.3 An LDT for every $H$ that is correct almost everywhere

In the next two sections we extend the generalized comparison decision tree to arbitrary sets $H$. Let $H \subseteq \mathbb{R}^d$ be an arbitrary finite set (not necessarily in a general position). The first step is to use a compactness argument to derive a decision tree that computes $\mathcal{A}_H(x)$ for *almost every* $x$ in the following sense. Recall that a linear decision tree $T$ computes the function $x \to \mathrm{sign}\left(\langle h, x\rangle\right)$ almost everywhere if this function is constant on every full dimensional leaf of $T$.

**Lemma 3.14.** *Let $H \subseteq \mathbb{R}^d$ be a finite set. Then there exists a generalized comparison LDT of depth $O\left(d^4 \log d \log|H|\right)$ that computes $\mathcal{A}_H(\cdot)$ almost everywhere.*

*Proof.* If $H$ is in general position then this follows from Lemma 3.9. So, assume that $H$ is not in general position. For every $n \in \mathbb{N}$, pick $H_n \subset \mathbb{R}^d$ with $|H_n| = |H|$ in general position such that for every $h \in H$ there exists $h_n = h_n(h) \in H_n$ with $\|h - h_n\|_2 \le 1/n$. By Lemma 3.9 each $H_n$ has a generalized comparisons tree $T_n$ of depth $D = O\left(d^4 \log d \log|H|\right)$ that computes $\mathcal{A}_{H_n}(\cdot)$. A standard compactness[1] argument imply the existence of a sequence of isomorphic trees $\{T_{n_k}\}_{k=1}^{\infty}$, such that for every vertex $v$, the sequence of the $H_{n_k}$-generalized comparisons queries corresponding to $v$ converges to an $H$-generalized comparison query. Let $T_\infty$ denote the limit tree. One can verify that $T_\infty$ satisfies the following property:

$$C_\infty(\ell) \subseteq \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} C_{n_k}(\ell), \text{ for every full dimensional leaf } \ell \text{ of } T_\infty. \tag{3}$$

In words, every $x \in C_\infty(\ell)$ belongs to all except finitely many of the $C_{n_k}(\ell)$. We claim that Equation (3) implies that $T_\infty$ computes $\mathrm{sign}(\langle h, \cdot\rangle)$ almost everywhere, for every $h \in H$. Indeed, let $\ell$ be a full dimensional leaf of $T_\infty$, and let $x', x'' \in C_\infty(\ell)$. Assume towards contradiction that $\mathrm{sign}(\langle h, x'\rangle) \neq \mathrm{sign}(\langle h, x''\rangle)$ for some $h \in H$. By Corollary 2.3 we may assume that $\mathrm{sign}(\langle h, x'\rangle) = -1$ and $\mathrm{sign}(\langle h, x''\rangle) = +1$. By Equation (3), both $x', x''$ belong to all but finitely many of the $C_{n_k}(\ell)$. Moreover, since $\mathrm{sign}(\langle h, x'\rangle), \mathrm{sign}(\langle h, x''\rangle) \neq 0$ and $h_{n_k}(h) \to_{k\to\infty} h$ it follows that $\mathrm{sign}(\langle h, x'\rangle) = \mathrm{sign}(\langle h_{n_k}(h), x'\rangle)$, and $\mathrm{sign}(\langle h, x''\rangle) = \mathrm{sign}(\langle h_{n_k}(h), x''\rangle)$ for all but finitely many $k$'s. Thus, for such $k$'s the function $\mathrm{sign}(\langle h_{n_k}, \cdot\rangle)$ is not constant on $C_{n_k}(\ell)$, which contradicts the assumption that $T_{n_k}$ computes $\mathrm{sign}(\langle h_{n_k}, \cdot\rangle)$. □

## 3.4 An LDT for every $H$

In this section we derive the final generalized comparison decision tree for arbitrary $H$, which implies Theorem 1.1. This is achieved by the next lemma that derives the final tree from the one in Lemma 3.14.

**Lemma 3.15.** *For every LDT $T$ there exists an LDT $T'$ such that*

- *$T'$ uses the same queries as $T$ and has the same depth as $T$.*
- *For every $h \in \mathbb{R}^d$, if $T$ computes sign$(\langle h, \cdot\rangle)$ almost everywhere then $T'$ computes sign$(\langle h, \cdot\rangle)$ everywhere.*

*Proof.* Without loss of generality, we may assume that $T$ is not redundant, in the sense that each query in it is informative. Namely, that $C(u) \neq \emptyset$ for every vertex $u \in T$.

[1] For the compactness argument to carry, each generalized comparison query is normalized so that its coefficient $\alpha, \beta$ are bounded, e.g. $\alpha + \beta = 1$

**Derivation of $T'$.** Given an input point $x$, follow the corresponding computation path in $T$ with the following modification: once a vertex $v$ whose query $q = q(v)$ satisfies $\mathrm{sign}(\langle q, x\rangle) = 0$ is reached, define in $T'$ a new child of $v$ that corresponds to this case, and continue following the same queries like in the subtree of $T$ that corresponds to the case "$\mathrm{sign}(\langle q, x\rangle) = +$".

**Correctness.** We prove that $T'$ computes $\mathrm{sign}(\langle h, \cdot\rangle)$ everywhere. Consider a leaf $\ell$ in $T'$, and let $x', x'' \in C(\ell)$. Assume toward contradiction that $\mathrm{sign}(\langle h, x'\rangle) \neq \mathrm{sign}(\langle h, x''\rangle)$. By Corollary 2.3 we may assume that $\mathrm{sign}(\langle h, x'\rangle) = -1$ and $\mathrm{sign}(\langle h, x''\rangle) = +1$. Let $q_1, \ldots, q_r$ denote the queries on the path towards $\ell$ whose query is replied by 0. Since $T$ is not redundant, it follows that the $q_i$'s are linearly independent. Thus, there is a solution $z$ to the system $\langle q_i, z\rangle = +1$ for $1 \leq i \leq r$. Let $y' = x' + \varepsilon \cdot z$ and $y'' = x'' + \varepsilon \cdot z$ where $\varepsilon > 0$ is sufficiently small such that

(i) $\mathrm{sign}(\langle h, x'\rangle) = -1$ and $\mathrm{sign}(\langle h, x''\rangle) = +1$, and

(ii) $\mathrm{sign}(\langle q, y'\rangle) = \mathrm{sign}(\langle q, x'\rangle) = \mathrm{sign}(\langle q, x''\rangle) = \mathrm{sign}(\langle q, y''\rangle)$ for every query $q$ on the path towards $\ell$ whose sign query is replied by a $\pm 1$.

Thus, by (ii) above and since $\varepsilon > 0$ it follows that $y', y''$ belong to the same full dimensional leaf of $T$. However, (i) above implies that the function $x \mapsto \mathrm{sign}(\langle h, x\rangle)$ is not constant on this leaf, which contradicts the assumption on $T$. □

# References

[Bar98] Franck Barthe. On a reverse form of the Brascamp-Lieb inequality. *Inventiones mathematicae*, 134(2):335–361, 1998.

[CIO15] Jean Cardinal, John Iacono, and Aurélien Ooms. Solving $k$-sum using few linear queries. *arXiv preprint arXiv:1512.06678*, 2015.

[ES17] Esther Ezra and Micha Sharir. A nearly quadratic bound for the decision tree complexity of k-sum. In *33rd International Symposium on Computational Geometry, SoCG 2017, July 4-7, 2017, Brisbane, Australia*, pages 41:1–41:15, 2017.

[For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

[Fre76] Michael L Fredman. How good is the information theory bound in sorting? *Theoretical Computer Science*, 1(4):355–361, 1976.

[KLM17] Daniel M. Kane, Shachar Lovett, and Shay Moran. Near-optimal linear decision trees for k-sum and related problems. *CoRR*, abs/1705.01720, 2017.

[KLMZ17] Daniel Kane, Shachar Lovett, Shay Moran, and Jiapeng Zhang. Active classification with comparison queries. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on.* IEEE, 2017.

[MadH84] Friedhelm Meyer auf der Heide. A polynomial linear search algorithm for the n-dimensional knapsack problem. *Journal of the ACM (JACM)*, 31(3):668–676, 1984.

[Mei93] Stefan Meiser. Point location in arrangements of hyperplanes. *Information and Computation*, 106(2):286–303, 1993.

[Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[VC71] VN Vapnik and A Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.