Distance-sensitive hashing*

Martin Aumüller[1], Tobias Christiani[1], Rasmus Pagh[1], and Francesco Silvestri[2]

1 **IT University of Copenhagen, Denmark, {maau, tobc, pagh}@itu.dk**
2 **University of Padova, Italy, silvestri@dei.unipd.it**

Abstract

We initiate the study of *distance-sensitive hashing*, a generalization of locality-sensitive hashing that seeks a family of hash functions such that the probability of two points having the same hash value is a given function of the distance between them. More precisely, given a distance space (X, dist) and a "collision probability function" (CPF) $f\colon \mathbb{R} \to [0, 1]$ we seek a distribution over pairs of functions (h, g) such that for every pair of points $\mathbf{x}, \mathbf{y} \in X$ the collision probability is $\Pr[h(\mathbf{x}) = g(\mathbf{y})] = f(\mathrm{dist}(\mathbf{x}, \mathbf{y}))$. Locality-sensitive hashing is the study of how fast a CPF can *decrease* as the distance grows. For many spaces f can be made exponentially decreasing even if we restrict attention to the symmetric case where $g = h$. In this paper we study how *asymmetry* makes it possible to achieve CPFs that are, for example, increasing or unimodal. Our original motivation comes from *annulus queries* where we are interested in searching for points at distance approximately r from a query point, but we believe that distance-sensitive hashing is of interest beyond this application.

1998 ACM Subject Classification H.3.3 Information Search and Retrieval

Keywords and phrases locality-sensitive hashing, annulus queries, recommender systems.

Digital Object Identifier 10.4230/LIPIcs...

1 Introduction

High-dimensional nearest neighbor search in a point set P is a building block in a variety of applications. A classical application is recommender systems: Suppose you have shown interest in a particular item, for example a news article $\mathbf{x}$. The semantic meaning of a piece of text can be represented as a high-dimensional *feature vector*, for example computed using latent semantic indexing [16]. In order to recommend other news articles we might search the set P of article feature vectors for articles that are similar to $\mathbf{x}$. But in general it is not clear that it is desirable to recommend the most similar articles. Indeed, it might be desirable to recommend articles that are on the same topic but are not *too* aligned with $\mathbf{x}$, and may provide a different perspective.

For many applications of nearest neighbor search it is acceptable to approximate distances such that the points reported are only approximately as close to $\mathbf{x}$ as the true set of closest points. Locality-sensitive hashing (LSH), first defined by Indyk and Motwani [19], is a powerful framework for approximate nearest neighbor search (ANN) in high dimensions that achieves sublinear query time. However, existing LSH techniques do not allow us to search for points that are "close, but not too close". In a nutshell: LSH provides a sequence of hash

* The research leading to these results has received funding from the European Research Council under the European Union's 7th Framework Programme (FP7/2007-2013) / ERC grant agreement no. 614331. Silvestri has also been supported by MIUR of Italy under project AMANDA.

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

functions $h_1, h_2, \ldots$ such that if $\mathbf{x}$ and $\mathbf{y}$ are close we have $h_i(\mathbf{x}) = h_i(\mathbf{y})$ for some i with high probability, while if $\mathbf{x}$ and $\mathbf{y}$ are distant we have that $h_i(\mathbf{x}) \neq h_i(\mathbf{y})$ for all i with high probability. In this paper we seek techniques that allow us to refine the first requirement: If $\mathbf{x}$ and $\mathbf{y}$ are "too close" we would like collisions to occur only with very small probability. At first sight this seems impossible because we will, by definition, have a collision when $\mathbf{x} = \mathbf{y}$. However, this objection is overcome by switching to an *asymmetric* setting where we work with pairs of functions (h_i, g_i) and are concerned with collisions of the form $h_i(\mathbf{x}) = g_i(\mathbf{y})$. More generally, we initiate the study of the following question: In the asymmetric setting, what is the class of functions f for which it is possible to achieve $\Pr[h(\mathbf{x}) = g(\mathbf{y})] = f(\mathrm{dist}(\mathbf{x}, \mathbf{y}))$, where the probability is over the choice of (h, g) and $\mathrm{dist}(\mathbf{x}, \mathbf{y})$ is the distance between $\mathbf{x}$ and $\mathbf{y}$. We refer to such a function as a *collision probability function* (CPF). More formally:

▶ **Definition 1.** A distribution $\mathcal{D}$ over pairs of functions $h, g \colon X \to \mathbb{R}$ is called *distance-sensitive* for the space (X, dist) with collision probability function (CPF) $f \colon \mathbb{R} \to [0, 1]$ if for each pair $\mathbf{x}, \mathbf{y} \in X$ and (h, g) sampled according to $\mathcal{D}$ we have $\Pr[h(\mathbf{x}) = g(\mathbf{y})] = f(\mathrm{dist}(\mathbf{x}, \mathbf{y}))$.

1.1 Our results

On a high level our results go into two different directions. First, we show that distance-sensitive hash families with certain CPFs allow us to reuse the standard LSH data structure [19] to solve problems where standard LSH families do not yield satisfactory solutions. Second, we describe constructions of distance-sensitive hash families that achieve certain CPFs and study lower bounds on distance-sensitive hash families with monotonically increasing CPFs.

We consider a standard RAM model of computation with word size $\Theta(\log n)$ bits where $n = |P|$ is the size of the set of points. For simplicity we also assume that a point in (X, dist) can be stored using d words and that the time complexity is $O(d)$ for performing distance computations, as well as sampling and evaluating functions from a distance-sensitive family (if this is not the case, the space and time bounds can be adjusted accordingly).

Applications. *Approximate annulus search* is the problem of finding a point in the set P of data points with distance in an interval $[r_-, r_+]$ from a query point. Having access to a distance-sensitive hash family with a CPF that peaks inside $[r_-, r_+]$ and is significantly smaller at the ends of the interval gives an LSH-like solution to this problem.

▶ **Theorem 2.** *Suppose we have a set P of n points, an interval $[r_-, r_+]$, a distance $r \in [r_-, r_+]$, and assume we are given a distance-sensitive family with CPF f such that $f(r') \leq 1/n$ for all $r' \notin [r_-, r_+]$. Then there exists a data structure that, given a query $\mathbf{q}$ for which there exists $x \in P$ with $\mathrm{dist}(\mathbf{q}, \mathbf{x}) = r$, returns $\mathbf{x}' \in P$ with $\mathrm{dist}(\mathbf{q}, \mathbf{x}') \in [r_-, r_+]$ with probability at least $1/2$. The data structure uses space $O(n^{1+\rho^*}/f(r) + dn)$ and has query time $O(dn^{\rho^*})$, where $\rho^* = \log(1/f(r))/\log n$.*

Obtaining a CPF that peaks inside of $[r_-, r_+]$ can be achieved by combining a standard locality-sensitive hash family with a distance-sensitive family that has an increasing CPF. On the d-dimensional unit sphere under inner product similarity, our strongest construction for solving the annulus search problem, described in section 2.2, allows us to search a point set P of unit vectors for a vector approximately orthogonal to a query vector $\mathbf{q}$ in time $dn^{\rho^*+o(1)}$ for $\rho^* = \frac{1-\alpha^2}{1+\alpha^2}$, where we guarantee to return a vector $\mathbf{x}$ with $\langle \mathbf{x}, \mathbf{q} \rangle \in [-\alpha, \alpha]$ if an orthogonal vector exists (a special case of Theorem 28).

Approximate spherical range reporting [1] aims to report all points in P within distance r

from a query point. CPFs that have a (roughly) fixed value in $[0, r]$ and then decrease rapidly to zero yield data structures with good *output sensitivity*.

▶ **Theorem 3.** *Suppose we have a set P of n points and two distances $r < r_+$. Assume we are given a distance-sensitive family with CPF f where $f(r') \leq 1/n$ for all $r' \geq r_+$, and let $f_{\min} = \inf_{t\in[0,r]} f(t)$, $f_{\max} = \sup_{t\in[0,r]} f(t)$. Then there exists a data structure that, given a query $\mathbf{q}$, returns $S \subseteq \{\mathbf{x} \in P \mid \operatorname{dist}(\mathbf{q}, \mathbf{x}) \leq r_+\}$ such that for each $\mathbf{x} \in P$ with $\operatorname{dist}(\mathbf{q}, \mathbf{x}) \leq r$, $\Pr[\mathbf{x} \in S] > 1/2$. The data structure uses space $O(n^{1+\rho^*} + dn)$ and the query has expected running time $O(dn^{\rho^*} + d|S|f_{\max}/f_{\min})$, where $\rho^* = \log(1/f_{\min})/\log(1/f(r_+))$.*

In particular, if we have a constant bound on $f_{\max}/f_{\min}$ the output sensitivity is optimal in the sense that the time to report an additional close point is $O(d)$ which is the time it takes to verify its distance to the query point. CPFs with this property are implicit in the linear space extremes of the space-time tradeoff techniques for similarity search [4, 13], but a better value of ρ^* could possibly be obtained by allowing a higher space usage.

We note that the assumption $f(r_+) \leq 1/n$ in both theorems is not critical: the standard technique of *powering* (see Lemma 6) allows us to work with the CPF $f(x)^k$ for integer k, where k is the smallest integer such that $f(x)^k \leq 1/n$.

The proofs of the theorems, which follow strictly along the lines of proofs for the standard LSH data structure in [19], are sketched in Appendix A for completeness.

Constructions and lower bound. Section 2 presents our constructions of distance-sensitive hash families. As a warm-up we consider a simple construction of a distance-sensitive hash family with an increasing CPF for Hamming space building upon the well-known bit-sampling approach from [19]. While bit-sampling is in a certain sense optimal [26] as a locality-sensitive hash family with decreasing CPF w.r.t. the gap of collision probabilities at distance r and cr, it turns out that it is possible to find distance-sensitive hash families with CPFs that have a larger gap between the collision probabilities at distances r and r/c.

We describe two such families. The central tool in both constructions is the projection of vectors $\mathbf{x} \in \mathbb{R}^d$ to $\mathbb{R}$ by taking the inner product $\langle \mathbf{x}, \mathbf{z} \rangle$ where $\mathbf{z} \sim \mathcal{N}^d(0, 1)$. This is a well-known technique in the locality-sensitive hashing literature and it has been used in many constructions of locality-sensitive families [15, 4, 13]. In our first construction, we consider an asymmetric version of the classical E2LSH family [15] for Euclidean space, namely sampling pairs (h, g) with

$$h\colon \mathbf{x} \mapsto \left\lfloor \tfrac{\langle \mathbf{a}, \mathbf{x}\rangle + b}{w} \right\rfloor, \qquad g\colon \mathbf{x} \mapsto \left\lfloor \tfrac{\langle \mathbf{a}, \mathbf{x}\rangle + b}{w} \right\rfloor + k, \tag{1}$$

where $b \in [0, w]$ is uniformly random and $\mathbf{a} \sim \mathcal{N}^d(0, 1)$ is a d-dimensional random Gaussian vector. We show that this method, for suitable choice of parameters $w \in \mathbb{R}$ and $k \in \mathbb{N}$, provide a near-optimal gap of $1/c^2 + o(1)$ in the ratio of the logarithms of collision probabilities between close points at distance r and very close points at distance r/c. This is surprising, since the classical E2LSH is not optimal as an LSH for Euclidean space [3].

In order to find a lower bound for the gap in the collision probabilities, we consider vectors $\mathbf{x}, \mathbf{y} \in \{0, 1\}^d$ that are random and α-correlated, i.e., for each $i \in \{1, \ldots, d\}$ we have $\Pr[\mathbf{x}_i = \mathbf{y}_i] = \frac{1+\alpha}{2}$ independently. The expected Euclidean squared distance is $\mathrm{E}[||\mathbf{x} - \mathbf{y}||_2^2] = (1 - \alpha)d/2$, and by applying a Chernoff bound we have that for large d the distance is within a factor $1 + o_d(1)$ of the expectation with high probability. We show the following lower bound:

▶ **Theorem 4.** *Let $\mathcal{D}$ be a distribution over pairs of functions $h, g\colon \{0, 1\}^d \to R$, and define $\hat{f}\colon [-1, 1] \to [0, 1]$ by $\hat{f}(\alpha) = \Pr[h(\mathbf{x}) = g(\mathbf{y})]$ where $\mathbf{x}, \mathbf{y}$ are randomly α-correlated and (h, g) is sampled according to $\mathcal{D}$. Then for every $0 \leq \alpha < 1$ we have that $\hat{f}(\alpha) \geq \hat{f}(0)^{\frac{1+\alpha}{1-\alpha}}$.*

That is, the collision probability for α-correlated vectors cannot be too much smaller than the collision probability for random (0-correlated) vectors, a statement *dual* to standard (symmetric) LSH lower bounds [22, 6, 26]. Since correlation 0 corresponds to Euclidean distance $r = \sqrt{d/2}$ and correlation α to Euclidean distance $r/c = \sqrt{(1-\alpha)d/2}$ it follows that the lower bound on the collision probability can also be stated in terms of Euclidean distance with approximation factor $c = 1/\sqrt{1-\alpha}$. Then the exponent of the bound is $\frac{1+\alpha}{1-\alpha} = \frac{2-1/c^2}{1/c^2} = 2c^2 - 1$. This matches the exponent shown for (1) in section 2.1 up to a constant factor, but a gap remains. Using our second construction that is based on the recently-discovered concept of *locality-sensitive filters* [7] and takes ideas from [4] and [13], it turns out that the lower bound can be matched up to a *lower-order term* in the exponent on the unit sphere.

▶ **Theorem 5.** *Let $\varepsilon > 0$ be constant. For every $t > 0$ there exists a family $\mathcal{D}_-$ of distance-sensitive functions for $(\mathbb{S}^{d-1}, \langle\cdot,\cdot\rangle)$ with CPF f such that for $\alpha \in [-1+\varepsilon, 1-\varepsilon]$ we have that*

$$\ln(1/f(\alpha)) = \tfrac{1+\alpha}{1-\alpha}\tfrac{t^2}{2} + \Theta(\log t) \tag{2}$$

The complexity of sampling, storing, and evaluating $(h, g) \in \mathcal{D}$ is $O(dt^4 e^{t^2/2})$.

Note that this shows the exponent in Theorem 4 is tight up to an additive $o_t(1)$ term.

Finally, in section 2.3 we consider the following natural question: Let $\mathcal{P}(t)$ be a polynomial. Does there exist a distance-sensitive hash family with CPF $f(t) = \mathcal{P}(t)$? We present two general approaches of constructing CPFs for the unit sphere and Hamming space that cover a wide range of such polynomials.

1.2 Related work

A substantial literature has been devoted to the study of locality-sensitive hashing (LSH). Here we review only selected results, and refer to [36] for a comprehensive survey. For simplicity we consider only LSH constructions that are *isometric* in the sense that the probability of a hash collision depends only on the distance $\mathrm{dist}(\mathbf{x}, \mathbf{y})$. In other words, there exists a *collision probability function* (CPF) $f\colon \mathbb{R} \to [0,1]$ such that $\Pr[h(\mathbf{x}) = h(\mathbf{y})] = f(\mathrm{dist}(\mathbf{x}, \mathbf{y}))$. Almost all LSH constructions whose collision probability has been rigorously analyzed are isometric. Notable exceptions are recent *data dependent* LSH methods such as [5] where the LSH distribution, and thus the collision probabilities, depends on the structure of data.

ρ-values. Much attention has been given to optimal ρ-values of locality-sensitive hash functions, where we consider *non-increasing* CPFs. Suppose we are interested in hash collisions when $\mathrm{dist}(\mathbf{x}, \mathbf{y}) = r_1$ but want to avoid hash collisions when $\mathrm{dist}(\mathbf{x}, \mathbf{y}) \geq r_2$, for some $r_2 > r_1$. The ρ-value of this setting is the real number in $[0,1]$ such that $f(r_1) = f(r_2)^\rho$; it measures the gap between collision probabilities $f(r_1)$ and $f(r_2)$. The ρ-value determines the performance of LSH-based data structures for the (r_1, r_2)-*approximate near neighbor* problem: Assume that it takes $O(d)$ time to sample and evaluate a locality-sensitive hash function and compute a distance between two points. Then we can preprocess a point set P of n points in time $O(d\, n^{1+\rho})$ such that for a query point $\mathbf{q}$ from which there exists a point in P within distance r_1, you can return $\mathbf{x} \in P$ within distance r_2 of q in time $O(dn^\rho)$. In many spaces a good upper bound on ρ can be given in terms of the ratio $c = r_2/r_1$, but in general the smallest possible ρ can depend on r_1, r_2, $f(r_1)$, as well as the number of dimensions d. In this paper we consider collision probabilities of the form $\Pr[h(\mathbf{x}) = g(\mathbf{y})]$; as stated in Theorem 2 and Theorem 3 it remains relevant to compare collision probabilities using

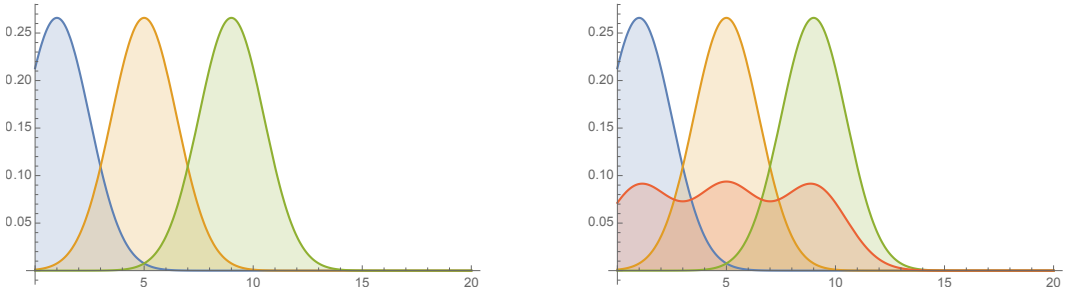

Figure 1 Composing several unimodal CPFs (left) to form a plateau CPF (red curve on the right) using Lemma 6. Such a CPF is particularly interesting when applying Theorem 3.

ρ-values, but we are not limited to non-increasing CPFs so the design space is significantly larger.

LSHable functions. Charikar [10] gave a necessary condition that all CPFs in the symmetric setting must fulfill, namely, $\text{dist}(\mathbf{x}, \mathbf{y}) = 1 - \Pr[h(\mathbf{x}) = h(\mathbf{y})]$ must be the distance measure of a metric, and more specifically this metric must be isometrically embeddable in ℓ_1. In the asymmetric setting this condition no longer holds as can be seen, for example, by noting that we can obtain $\text{dist}(\mathbf{x}, \mathbf{x}) = 1 - \Pr[h(\mathbf{x}) = g(\mathbf{x})] > 0$.

Chierichetti and Kumar [11, Lemma 7] considered transformations that can be used to create new CPFs. Though they are considered in a symmetric setting, the same constructions apply in an asymmetric setting and give the following result:

▶ **Lemma 6.** *Let $\{\mathcal{D}_i\}_{i=1}^n$ be a collection of n distance-sensitive families with CPFs $\{f_i\}_{i=1}^n$.*

(a) *There exists a distance-sensitive family $\mathcal{D}_{concat}$ with CPF $f(x) = \prod_{i=1}^n f_i(x)$.*

(b) *Given a probability distribution $\{p_i\}_{i=1}^n$ over $\{\mathcal{D}_i\}$, there exists a distance-sensitive family $\mathcal{D}^p$ with CPF $f(x) = \sum_{i=1}^n p_i f_i(x)$.*

Figure 1 shows an example application of Lemma 6. For completeness we present a proof of Lemma 6 in Appendix B.1. Interestingly, at least in the symmetric setting, the application of this lemma to a single CPF yields *all* transformations that are guaranteed to map a CPF to a CPF. Chierichetti et al. [12] recently extended the study of CPFs in the symmetric setting to allow *approximation*, i.e., allowing the collision probability to differ from a target function by a given approximation factor.

Asymmetric locality-sensitive hashing. Motivated by applications in machine learning, Vijayanarasimhan et al. [35] presented asymmetric LSH methods for Euclidean space where the collision probability is a decreasing function of $|\langle \mathbf{x}, \mathbf{y} \rangle|$. Shrivastava and Li [31] also explored how asymmetry can be used to achieve new CPFs (increasing), in settings where the inner product of vectors is used to measure closeness. Neyshabur and Srebro [24] extended this study by showing that the extra power obtained by asymmetry hinges on restrictions on the vector pairs for which we consider collisions: If vectors are not restricted to a bounded region of $\mathbb{R}^d$, no nontrivial CPF (as a function of inner product) is possible. On the other hand, if one vector is normalized (e.g. a query vector), the performance of known asymmetric LSH schemes can be matched with a symmetric method. But in the case where vectors are bounded but not normalized, asymmetric LSH is able to obtain CPFs that are impossible for symmetric LSH. Ahle et al. [2] showed further impossibility results for asymmetric LSH applied to inner products, and that symmetric LSH is possible in a bounded domain even without normalization if we just allow collision probability 1 when vectors coincide.

In section 3.2 we show that asymmetry does not help us when attempting to distinguish between random and positively correlated points in the Hamming cube using distance-sensitive hashing. Stated in terms of the ρ-value we get that $\rho \geq 1/(2c-1) - o_d(1)$ for distance-sensitive hashing, matching tight lower bounds from the symmetric LSH setting [22, 6]. We note that the asymmetric lower bound also follows implicitly from recent space-time tradeoff lower bounds [4, 13].

Indyk [17] showed how asymmetry can be used to enable new types of embeddings. More recently asymmetry has been used in the context of locality-sensitive *filters* [4, 13] and *maps* [14]. The idea is to map each point $\mathbf{x}$ to a pair of sets $(h(\mathbf{x}), g(\mathbf{x}))$ such that $\Pr[h(\mathbf{x}) \cap g(\mathbf{y}) \neq \emptyset]$ is constant if $\mathbf{x}$ and $\mathbf{y}$ are close, and very small if $\mathbf{x}$ and $\mathbf{y}$ are far from each other. This yields a similarity search data structure that adds for each vector $\mathbf{x} \in P$ the elements of $h(\mathbf{x})$ to a hash table; a query for a vector $\mathbf{q}$ proceeds by looking up each key in $g(\mathbf{q})$ in the hash table. One can transform such methods into asymmetric LSH methods by using min-wise hashing [8, 9] to sample a single element from each of the sets $h(\mathbf{x})$ and $g(\mathbf{x})$ (see [13, Theorem 1.4]).

Recommender systems. Returning to our motivating example we are not the first to address the topic of getting "interesting" recommendations using similarity search methods. Indyk et al. [18] build a similarity search data structure on a *core-set* of P to guarantee diverse query results. However, this method effectively discards much of the data set, so may not be suitable in all settings. Pagh et al. [27] consider the type of annulus queries that is interesting for recommendation, but their solution does not use the LSH framework and is limited to Euclidean space.

2 Constructions

Bit sampling [19] is one of the simplest LSH families for Hamming space, yet gives optimal ρ-values in terms of the approximation factor [26]. Its CPF is $f(t) = 1 - t$, where t is the relative Hamming distance. By using a function pair $(\mathbf{x} \mapsto x_i, \mathbf{x} \mapsto 1 - x_i)$ where $i \in \{1, \ldots, d\}$ is random, we get a simple asymmetric distance-sensitive family for Hamming space whose CPF $f(t) = t$, is monotonically increasing in the relative Hamming distance. We refer to increasing CPFs as *anti-LSH*, and the specific family as *anti bit-sampling* (because it gives a collision exactly when bit-sampling would not). Formally we have the family $\mathcal{H}_{ab} = \{(h_i, g_i) \mid 1 \leq i \leq d, h_i, g_i : \{0,1\}^d \to \{0,1\}, h_i : \mathbf{x} \mapsto x_i, g_i : \mathbf{x} \mapsto 1 - x_i\}$, which has CPF $f(t) = t$.

Anti-LSH is relevant since by concatenating an anti-LSH with a standard LSH, multiplying the CPFs (cf. Lemma 6(b)), we get unimodal CPFs that can be used to answer annulus queries. Let us set $r_- = r/c$ and $r_+ = cr$ for some $r > 0$ and $c > 1$. Let f^+ and f^- denote the CPFs of the LSH and anti-LSH families. Then, by Theorem 2, the annulus problem can be solved with $\rho^* \leq \rho^+ + \rho^-$, where $\rho^+ = \log(f^+(r))/\log(f^+(cr))$ and $\rho^- = \log(f^-(r))/\log(f^-(r/c))$. For anti bit-sampling, we get that $\rho^- = \Theta(1/\log c)$ as soon as r (normalized in $[0,1]$) is a constant factor from 1, and hence $\rho^* = \Theta(1/\log c)$.

Perhaps surprisingly, this anti-LSH approach is not optimal and a better result, with $\rho^* = O(1/c)$, follows by using an anti-LSH construction for Euclidean space proposed in section 2.1 and an anti-LSH based on filters for the unit sphere proposed in section 2.2, both yielding $\rho^- = O(1/c^2)$.

It is natural to wonder if more advanced CPFs can be obtained. We provide some results in this direction by describing in Section 2.3 two constructions yielding a wide class of CPFs.

2.1 An Anti-LSH construction in Euclidean Space

A simple and elegant distance-sensitive hash family in Euclidean space is given by a natural extension of the locality-sensitive hash family introduced by Datar et al. [15], where we project a point onto a line and split this line up into buckets. Let k and w be two suitable parameters to be chosen below. Consider the family of pairs of functions (h, g) with

$$h\colon \mathbf{x} \mapsto \left\lfloor \frac{\langle \mathbf{a}, \mathbf{x}\rangle + b}{w} \right\rfloor, g\colon \mathbf{y} \mapsto \left\lfloor \frac{\langle \mathbf{a}, \mathbf{y}\rangle + b}{w} \right\rfloor + k,$$

indexed by a uniform real number $b \in [0, w]$ and a d-dimensional random Gaussian vector $\mathbf{a} \sim \mathcal{N}^d(0, 1)$. We have the following result:

▶ **Theorem 7.** *Let r_- and r be two real values such that $0 < r_- < r$, and let $c = r/r_-$. We have that*

$$\rho^- = \frac{\log(1/f(r))}{\log(1/f(r_-))} = \frac{1}{c^2} + o_{c,k}(1).$$

Proof. For the sake of simplicity we assume $r = 1$ in the analysis (otherwise it is enough to scale down vectors accordingly). Let $\mathbf{x}$ and $\mathbf{y}$ be two points in $\mathbb{R}^d$ with distance Δ. We know that for $\mathbf{a} \sim \mathcal{N}^d(0, 1)$ the inner product $\langle \mathbf{a}, (\mathbf{x} - \mathbf{y})\rangle$ is distributed as $\mathcal{N}(0, \Delta)$. A necessary but not sufficient condition to have a collision between $\mathbf{x}$ and $\mathbf{y}$ is that $\langle \mathbf{a}, (\mathbf{x} - \mathbf{y})\rangle$ lies in the interval $[(k-1)w, (k+1)w]$. Now, if $t := \langle \mathbf{a}, (\mathbf{x} - \mathbf{y})\rangle \in [(k-1)w, kw]$, then the random offset b must lie in an interval of length $t - (k-1)w$, putting $\langle \mathbf{a}, \mathbf{x}\rangle$ and $\langle \mathbf{a}, \mathbf{y}\rangle - (k-1)w$ into different buckets. For the interval $[kw, (k+1)w]$ similar observations show that b has to be chosen in an interval of length $(k+1)w - t$. Let $\phi(t) = 1/\sqrt{2\pi}e^{-t^2/2}$ be the density function of a standard normal random variable. Similarly to the calculations in [15], the collision probability at distance Δ can be calculated as follows:

$$\begin{aligned} f(\Delta) &= \Pr\left(\left\lfloor \frac{\langle \mathbf{a}, \mathbf{x}\rangle + b}{w} \right\rfloor - \left\lfloor \frac{\langle \mathbf{a}, \mathbf{y}\rangle + b}{w} \right\rfloor = k\right) \\ &= \int_{(k-1)w}^{kw} \frac{\phi(t/\Delta)}{\Delta}\left(\frac{t}{w} - (k-1)\right) dt + \int_{kw}^{(k+1)w} \frac{\phi(t/\Delta)}{\Delta}\left(k + 1 - \frac{t}{w}\right) dt - \frac{\phi(kw/\Delta)}{\Delta} \\ &= \frac{1}{\sqrt{2\pi}\Delta}\left(\int_{(k-1)w}^{kw} e^{-\frac{t^2}{2\Delta^2}}\left(\frac{t}{w} - (k-1)\right) dt + \int_{kw}^{(k+1)w} e^{-\frac{t^2}{2\Delta^2}}\left(k + 1 - \frac{t}{w}\right) dt - e^{-\frac{(kw)^2}{2\Delta^2}}\right). \end{aligned}$$

We now proceed to upper bound ρ^- by finding an upper bound on $f(1/c)$ and a lower bound on $f(1)$. Simple calculations give an upper bound of

$$f(1/c) \le \frac{2wc}{\sqrt{2\pi}} e^{-(c(k-1)w)^2/2}. \tag{3}$$

For the lower bound, we only look at the interval $t \in [kw, (k+1/2)w]$ and obtain the bound:

$$f(1) \ge \frac{1}{\sqrt{2\pi}} \int_{kw}^{(k+1/2)w} e^{-\frac{t^2}{2}}\left(k + 1 - \frac{t}{w}\right) dt \ge \frac{w}{4\sqrt{2\pi}} e^{-((k+1/2)w)^2/2}. \tag{4}$$

Now we multiply the ratio of the logarithms of the right-hand sides of (3) and (4) with c^2 and look at the limit behavior for $c \to \infty$. We obtain that

$$\lim_{c\to\infty} \frac{\log\left(\frac{w}{4\sqrt{2\pi}} e^{-((k+1/2)w)^2/2}\right)}{\log\left(\frac{2wc}{\sqrt{2\pi}} e^{-(c(k-1)w)^2/2}\right)} c^2 = -\frac{2\log\left(\frac{w}{4\sqrt{2\pi}} e^{-\frac{1}{8}(2kw+w)^2)}\right)}{(k-1)^2 w^2},$$

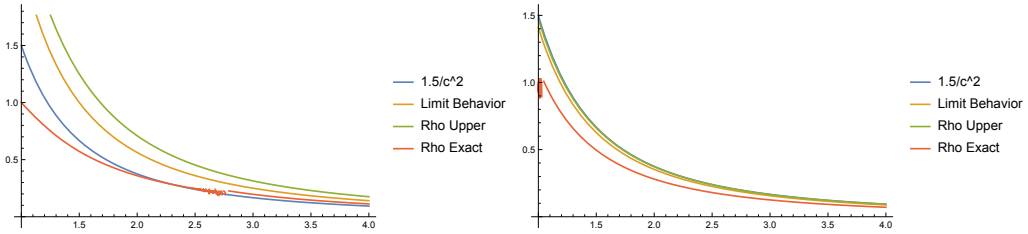

Figure 2 Graph depicting differences between the upper and exact bounds on the ρ^- value of the Euclidean space anti-LSH. The ρ^- value is depicted on the y-axis, the approximation factor c on the x-axis. The graph also shows the behavior of $f(\Delta)$ in the limit when k and w go to ∞, and the function $1.5/c^2$. Left: Parameter setting $k = 4, w = 1$; right: parameter setting $k = 9, w = 1$.

and notice that the right-hand side goes to 1 for $k \to \infty$ and arbitrary $w > 0$. This shows the claimed result. ◀

The result of Theorem 7 only holds for large k. For fixed k, ρ^- behaves asymptotically as $\frac{(2k+1)^2}{4c^2(k-1)^2} + o_{c,w}(1)$. For example, numerical calculations for $k = 9$ and $w = 1$ give an upper bound on the ρ^- value of $\frac{1.5}{c^2}$. Figure 2 compares the exact ρ^- value and our upper bound for two choices of parameters.

2.2 Optimal monotonic distance-sensitive hashing for the unit sphere

In this section we will show how to construct distance-sensitive hash families with monotonically increasing and decreasing CPFs for the unit sphere under inner product similarity that match the lower bounds shown in section 3. In particular we will prove Theorem 5 by showing the existence of a family $\mathcal{D}_-$ with a CPF $f_- : [-1, 1] \to [0, 1]$ that is monotonically decreasing in the similarity between points $\text{sim}(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$. The construction of $\mathcal{D}_-$ follows as a corollary from the construction of a family $\mathcal{D}_+$ with a CPF f_+ that is monotonically increasing in the similarity, and in fact, we have that $f_+(\alpha) = f_-(-\alpha)$ when $\mathcal{D}_+$ and $\mathcal{D}_-$ are parameterized in the same way. As an application of these families we show how they can be combined to yield powerful solutions to the approximate annulus search problem for a large, natural class of annuli. This application is further described in Appendix D.

The main new contribution compared to existing filter approaches [4, 13] is to make use of the asymmetry granted by $(h, g) \sim \mathcal{D}_-$ to show the existence of a family with a monotonically decreasing CPF. Furthermore, our analysis makes use of powerful tail bounds for the bivariate normal distribution [30] that allows us to provide guarantees for $\mathcal{D}_-, \mathcal{D}_+$ that span the entire range of similarities.

The distance-sensitive families. We begin by describing the family $\mathcal{D}_+$. The family takes as parameter a real number $t > 0$ and an integer m that we will later set as a function of t. We sample a pair of functions (h, g) from $\mathcal{D}_+$ by sampling m vectors $z_1, \ldots, z_m$ where $\mathbf{z}_i \sim \mathcal{N}^d(0, 1)$. The functions h, g map a point $\mathbf{x} \in \mathbb{S}^{d-1}$ to the index i of the first projection $\mathbf{z}_i$ where $\langle \mathbf{z}_i, \mathbf{x} \rangle \geq t$. If no such projection is found, then we ensure that $h(\mathbf{x}) \neq g(\mathbf{x})$ by mapping them to different values. Formally, we set

$$h_+(\mathbf{x}) = \min(\{i \mid \langle \mathbf{z}_i, \mathbf{x} \rangle \geq t\} \cup \{m + 1\}),$$
$$g_+(\mathbf{x}) = \min(\{i \mid \langle \mathbf{z}_i, \mathbf{x} \rangle \geq t\} \cup \{m + 2\}).$$

The collision probability for $(h, g) \sim \mathcal{D}_+$ depends only on the similarity $\alpha = \langle \mathbf{x}, \mathbf{y} \rangle$ between

the pair of points being evaluated and is given by

$$f_+(\alpha) = \Pr[h(\mathbf{x}) = g(\mathbf{y})] = \Pr[h(\mathbf{x}) \leq m \wedge g(\mathbf{y}) \leq m] \frac{\Pr[\langle \mathbf{z}, \mathbf{x} \rangle \geq t \wedge \langle \mathbf{z}, \mathbf{y} \rangle \geq t]}{\Pr[\langle \mathbf{z}, \mathbf{x} \rangle \geq t \vee \langle \mathbf{z}, \mathbf{y} \rangle \geq t]}.$$

The only way the family $\mathcal{D}_-$ differs from $\mathcal{D}_+$ is in the definition of $g_-(\mathbf{x})$ where we replace the condition as follows:

$$g_-(\mathbf{x}) = \min(\{i \mid \langle \mathbf{z}_i, \mathbf{x} \rangle \leq -t\} \cup \{m+2\}).$$

The collision probability $f_-(\alpha)$ of $\mathcal{D}_-$ follows analogously. We observe the following connection between $\mathcal{D}_+$ and $\mathcal{D}_-$.

▶ **Lemma 8.** *Given $\mathcal{D}_+$ and $\mathcal{D}_-$ with identical parameters we have that $f_+(\alpha) = f_-(-\alpha)$.*

Proof. A bivariate normally distributed variable with correlation α can be represented as a pair (X, Y) with $X = Z_1$ and $Y = \alpha Z_1 + \sqrt{1-\alpha^2} Z_2$ where Z_1, Z_2 are i.i.d. standard normal. By the symmetry of the standard normal distribution around zero it is straightforward to verify that $\Pr[Z_1 \geq t \wedge \alpha Z_1 + \sqrt{1-\alpha^2} Z_2 \geq t] = \Pr[Z_1 \geq t \wedge -\alpha Z_1 + \sqrt{1-\alpha^2} Z_2 \leq -t]$. ◀

Bounding the CPF. We use tail bounds for the standard normal distribution and the tail bounds by Savage [30] for the bivariate standard normal distribution in order to obtain the following lemma, the details of which are provided in Appendix C. We remark that this lemma provides also bounds for $f_-(\alpha)$ through the observation in Lemma 8.

▶ **Lemma 9.** *For every $t > 0$ and $\alpha \in (-1, 1)$ the family $\mathcal{D}_+$ satisfies*

$$f_+(\alpha) < \bar{f}_+(\alpha) := \frac{1}{\sqrt{2\pi}} \frac{t+1}{t^2} \frac{(1+\alpha)^2}{\sqrt{1-\alpha^2}} \exp\left(-\frac{1-\alpha}{1+\alpha} \frac{t^2}{2}\right),$$
$$f_+(\alpha) > \left(1 - \frac{(2-\alpha)(1+\alpha)}{1-\alpha} \frac{1}{t^2}\right) \frac{t}{t+1} \bar{f}_+(\alpha) - 2\exp(-t^3).$$

The complexity of sampling, storing, and evaluating a pair of functions $(h, g) \in \mathcal{D}_+$ is $O(dt^4 e^{t^2/2})$.

Results. Combining the above ingredients show Theorem 5. We also note that a similar statement holds for $\mathcal{D}_+$:

▶ **Corollary 10.** *Theorem 5 holds for $\mathcal{D}_+$ with the* CPF *bound*

$$\ln(1/f(\alpha)) = \frac{1-\alpha}{1+\alpha} \frac{t^2}{2} + \Theta(\log t).$$

Results on the unit sphere can be extended to ℓ_s-spaces for $0 < s \leq 2$ through the embedding result by Rahimi and Recht [29] as shown in [13]. A more careful analysis of the collision probabilities is required in order to combine the families $\mathcal{D}_-$ and $\mathcal{D}_+$ to form a unimodal family that can be used to solve the annulus search problem, see Theorem 2. These results are stated in Appendix D.

2.3 General constructions

So far we have focused our attention on anti-LSH constructions, which just represent one kind of distance-sensitive functions. We now overview general constructions targeting wider classes of CPFs.

Angular similarity functions

We say that $\mathrm{sim}\colon [-1,1] \to [0,1]$ is an *LSHable angular similarity function* if there exists a distance-sensitive hash family $\mathcal{S}$ with collision probability function $\mathrm{sim}(\langle \mathbf{x}, \mathbf{y} \rangle)$ for each $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{d-1}$. For example, the function $\mathrm{sim}(t) = 1 - \arccos(t)/\pi$ is LSHable using the *SimHash* construction of Charikar [10].

Valiant [34] described a pair of mappings $\varphi_1^{\mathcal{P}}, \varphi_2^{\mathcal{P}} \colon \mathbb{R}^d \to \mathbb{R}^D$, where $D = O(d^k)$, such that $\varphi_1^{\mathcal{P}}(\mathbf{x}) \cdot \varphi_2^{\mathcal{P}}(\mathbf{y}) = \mathcal{P}(\langle \mathbf{x}, \mathbf{y} \rangle)$, for any polynomial $\mathcal{P}(t) = \sum_{i=0}^{k} a_i t^i$. By leveraging this construction, it is possible to derive the following result (see Appendix B.2 for the proof).

▶ **Theorem 11.** *Suppose that sim is an LSHable angular similarity function and that the polynomial $\mathcal{P}(t) = \sum_{i=0}^{k} a_i t^i$ satisfies $\sum_{i=0}^{k} |a_i| = 1$. Then there exists a distribution over pairs (h, g) of functions such that for all $\mathbf{x}, \mathbf{y} \in S^{d-1}$, $\Pr[h(\mathbf{x}) = g(\mathbf{y})] = \mathrm{sim}(\mathcal{P}(\langle \mathbf{x}, \mathbf{y} \rangle))$.*

The computational cost of a naïve implementation of the proposed scheme may be prohibitive when d^k is large. However, by using the so-called *kernel approximation* methods [28], we can in near-linear time compute approximations $\hat{\varphi}_1^{\mathcal{P}}(\mathbf{x})$ and $\hat{\varphi}_2^{\mathcal{P}}(\mathbf{y})$ that satisfy $\hat{\varphi}_1^{\mathcal{P}}(\mathbf{x}) \cdot \hat{\varphi}_2^{\mathcal{P}}(\mathbf{y}) = \mathcal{P}(\langle \mathbf{x}, \mathbf{y} \rangle) \pm \varepsilon$ with high probability for a given approximation error $\varepsilon > 0$.

Hamming distance functions

For Hamming space, it is natural to wonder which CFPs can be expressed as a function of the relative Hamming distance $d_h(\mathbf{x}, \mathbf{y})$. A first positive answer follows by using the anti bit-sampling approach mentioned at the beginning of this section together with Lemma 6. This gives a scheme for matching any polynomial $\mathcal{P}(t) = \sum_{i=0}^{k} a_i t^i$ that satisfies $\sum_{i=0}^{k} a_i = 1$ and $a_i > 0$ for each i.

In this section, we provide another construction that matches, up to a scaling factor Δ, any polynomial $\mathcal{P}(t)$ having no roots with a real part in $(0, 1)$. The scaling factor depends only on the roots of the polynomial. We have the following result that is proven in Appendix B.3:

▶ **Theorem 12.** *Let $\mathcal{P}(t) = \sum_{i=0}^{k} a_i t^i$, Z be the multiset of roots of $\mathcal{P}(t)$, and $\psi \leq k$ be the number of roots with negative real part. Then there exists a distance-sensitive hash family with collision probability $\Pr(h(\mathbf{x}) = g(\mathbf{y})) = \mathcal{P}(d_h(\mathbf{x}, \mathbf{y}))/\Delta$ with $\Delta = a_k 2^{\psi} \prod_{z \in Z, |z| > 1} |z_i|$.*

The construction exploits the factorization $\mathcal{P}(t) = a_k \prod_{z \in Z}(t - z)$ and consists of a combination of $|Z|$ variations of bit-sampling and anti bit-sampling. We refer to Theorem 12 in Appendix B.3 for the construction. Although the proposed scheme may not reach the ρ value given by the polynomial $\mathcal{P}(t)$, it can be used for estimating $\mathcal{P}(d_H(\mathbf{x}, \mathbf{y}))$ since the scaling factor is constant and only depends on the polynomial.

We remark that a scaling factor Δ is unavoidable in the general case. Otherwise, it would be possible to match the CFP $1 - t^2$ for Hamming space, which implies $\rho \leq 1/c^2$ in contradiction with the lower bound $1/c$ in [26]. However, it is an open question to provide better bounds on Δ.

Finally, we observe that our scheme can be used to approximate any function $f(t)$ that can be represented with a Taylor series: indeed, it is sufficient to truncate the series to the term that gives the desired approximation, and then to apply our construction to the resulting truncated polynomial.

3 Lower bound

In this section we will show lower bounds on the CPFs of distance-sensitive families in Hamming space under relative Hamming distance. These results extend to the unit sphere

and Euclidean space through standard embeddings. Our primary focus will be to obtain a lower bound for the case of a CPF that is increasing with the distance, i.e., decreasing in the similarity. As with our upper bounds for the unit sphere, re-applying the same techniques also yields a lower bound for the case of an increasing CPF in the similarity.

The proof combines the (reverse) small-set expansion theorem by O'Donnell [25] with techniques inspired by the LSH lower bound of Motwani et al. [22]. The reverse small-set expansion theorem lower bounds the probability that random α-correlated points $(\mathbf{x}, \mathbf{y})$ end up in a pair of subsets A, B of the Hamming cube, as a function of the size of the subsets. The main contribution here is to extend this lower bound for pairs of subsets of Hamming space to our object of interest: distributions over pairs of functions that partition space. We begin by introducing the required tools from [25].

▶ **Definition 13.** For $0 \leq \alpha \leq 1$ we say that $(\mathbf{x}, \mathbf{y})$ are α-correlated if $\mathbf{x}$ is chosen uniformly at random from $\{0,1\}^d$ and $\mathbf{y}$ is constructed by rerandomizing each bit from $\mathbf{x}$ independently at random with probability $1 - \alpha$.

In the following we refer to the volume of $A \subset \{0,1\}^d$ as $|A|/2^d$.

▶ **Theorem 14** (Reverse Small-Set Expansion)**.** *Let $0 \leq \alpha \leq 1$. Let $A, B \subseteq \{0,1\}^d$ have volumes $\exp(-a^2/2)$, $\exp(-b^2/2)$, respectively, where $a, b \geq 0$. Then we have that*

$$\Pr_{\substack{(\mathbf{x},\mathbf{y}) \\ \alpha\text{-corr.}}}[\mathbf{x} \in A, \mathbf{y} \in B] \geq \exp\left(-\frac{1}{2}\frac{a^2 + 2\alpha ab + b^2}{1 - \alpha^2}\right).$$

We define a probabilistic version of the probability collision function that we will state results for. Later we will apply concentration bounds on the similarity between α-correlated pairs of points in order to make statements about the actual CPF. We will use R to denote the range of a family of functions which, without loss of generality, we can assume to be finite.

▶ **Definition 15** (Probabilistic CPF)**.** Let $\mathcal{D}$ be a distribution over pairs $h, g \colon \{0,1\}^d \to R$ and $0 \leq \alpha < 1$. Then we define the probabilistic CPF $\hat{f} \colon [0,1] \to [0,1]$ by

$$\hat{f}(\alpha) = \Pr_{\substack{(h,g)\sim\mathcal{D} \\ (\mathbf{x},\mathbf{y})\ \alpha\text{-corr.}}}[h(\mathbf{x}) = g(\mathbf{y})].$$

We are now ready to state our main lemma that lower bounds $\hat{f}(\alpha)$ in terms of $\hat{f}(0)$. This immediately implies Theorem 4.

▶ **Lemma 16.** *For every distribution $\mathcal{D}$ over pairs of functions $h, g \colon \{0,1\}^d \to R$ and every $0 \leq \alpha < 1$ we have that $\hat{f}(\alpha) \geq \hat{f}(0)^{\frac{1+\alpha}{1-\alpha}}$.*

Proof. For a function $h \colon \{0,1\}^d \to R$ define its inverse $h^{-1} \colon R \to 2^{\{0,1\}^d}$ by $h^{-1}(i) = \{\mathbf{x} \in \{0,1\}^d \mid h(\mathbf{x}) = i\}$. For a pair of functions $(h, g) \in \mathcal{D}$ and $i \in R$ we define $a_{h,i}, b_{g,i} \geq 0$ such that $|h^{-1}(i)|/2^d = \exp(-a_{h,i}^2/2)$ and $|g^{-1}(i)|/2^d = \exp(-b_{g,i}^2/2)$. For fixed (h, g) define $\hat{f}_{h,g}(\alpha) = \Pr_{(\mathbf{x},\mathbf{y})\ \alpha\text{-corr.}}[h(\mathbf{x}) = g(\mathbf{y})]$. We obtain a lower bound on $\hat{f}(\alpha)$ in the following

way:

$$\begin{aligned}
\hat{f}(\alpha) &= \mathop{\mathrm{E}}_{(h,g)\sim\mathcal{D}}\left[\sum_{i\in R}\Pr_{(\mathbf{x},\mathbf{y})\ \alpha\text{-corr.}}[h(\mathbf{x})=g(\mathbf{y})=i]\right] \\
&\geq \mathop{\mathrm{E}}_{(h,g)\sim\mathcal{D}}\left[\sum_{i\in R}\exp\left(-\frac{1}{2}\frac{a_{h,i}^2+2\alpha a_{h,i}b_{g,i}+b_{g,i}^2}{1-\alpha^2}\right)\right] &(5)\\
&\geq \mathop{\mathrm{E}}_{(h,g)\sim\mathcal{D}}\left[\sum_{i\in R}\exp\left(-\frac{1}{2}\frac{a_{h,i}^2+b_{g,i}^2}{1-\alpha}\right)\right] &(6)\\
&\geq \mathop{\mathrm{E}}_{(h,g)\sim\mathcal{D}}\hat{f}_{h,g}(0)^{\frac{1+\alpha}{1-\alpha}} &(7)\\
&\geq \left(\mathop{\mathrm{E}}_{(h,g)\sim\mathcal{D}}\hat{f}_{h,g}(0)\right)^{\frac{1+\alpha}{1-\alpha}} &(8)\\
&= \hat{f}(0)^{\frac{1+\alpha}{1-\alpha}}
\end{aligned}$$

Inequality (5) is due to Theorem 14. Inequality (6) follows from the simple fact that $a^2+\alpha(a^2+b^2)+b^2 \geq a^2+2\alpha ab+b^2$. Inequality (7) follows from the result of an optimization problem further described in Appendix E.1. Finally, inequality (8) follows from a standard application of Jensen's Inequality. ◀

3.1 Bounding the CPF

We will use Lemma 16 to a show a lower bound for distance-sensitive families that have the *opposite* properties of locality-sensitive hash families. Our lower bound holds for "similarity"-sensitive hash families, where we replace the distance function in the space (X, dist) in Definition 1 by a space (X, sim) equipped with similarity measure $\mathrm{sim}\colon X \times X \to [0,1]$. The following definition covers both standard locality-sensitive hash families and families having the opposite behavior from a similarity perspective.

▶ **Definition 17** (Similarity-[in]sensitive hash families)**.** Let $\mathcal{D}$ be a similarity-sensitive family for (X, sim) with CPF f. We say that $\mathcal{D}$ is $(\alpha_-, \alpha_+, f_-, f_+)$-[in]sensitive if it satisfies:

- For $\alpha \leq \alpha_-$ we have that $f(\alpha) \leq f_-$ $[f(\alpha) \geq f_-]$.
- For $\alpha \geq \alpha_+$ we have that $f(\alpha) \geq f_+$ $[f(\alpha) \leq f_+]$.

We state our results in the natural similarity-version of Hamming space that also corresponds to embedding Hamming space into the unit sphere, namely the space $(\{0,1\}^d, \mathrm{sim}_H)$ where $\mathrm{sim}_H(\mathbf{x},\mathbf{y}) = 1 - 2\left\|\mathbf{x}-\mathbf{y}\right\|_1/d$. In the following theorem we extend the lower bound from Lemma 16 that considers the relation between $\hat{f}(0)$ and $\hat{f}(\alpha)$ to a wider range of parameters $0 < \alpha_- < \alpha_+ < 1$ and we consider the relation between $f(\alpha_-)$ and $f(\alpha_+)$. The proof has been deferred to Appendix E.2.

▶ **Theorem 18.** *Let $0 < \alpha_- < \alpha_+ < 1$ be constants. Then every $(\alpha_-, \alpha_+, f_-, f_+)$-insensitive family $\mathcal{D}$ for $(\{0,1\}^d, \mathrm{sim}_H)$ must satisfy*

$$\frac{\log(1/f_-)}{\log(1/f_+)} \geq \frac{1-\alpha_+}{1+\alpha_+-2\alpha_-} - O(\sqrt{\log(1/f_+)/d}).$$

▶ Remark. In the statement of Theorem 18 we may replace the properties from Definition 17 that hold for every $\alpha \leq \alpha_-$ and every $\alpha \geq \alpha_+$ with less restrictive versions that hold in an ε-interval around α_-, α_+ for some $\varepsilon = o_d(1)$.

▶ Remark. If we rewrite the bound in terms of relative Hamming distances δ and δ/c where δ, c are constants, we obtain a lower bound of $1/(2c-1) - o_d(1)$ — an expression that is familiar from known LSH lower bounds [22, 6].

3.2 The other direction

We can re-apply the techniques behind Lemma 16 and Theorem 18 to state similar results in the other direction where for $\alpha_- < \alpha_+$ we are interested in upper bounding $f(\alpha_+)$ as a function of $f(\alpha_-)$. This is similar to the well-studied problem of constructing LSH lower bounds and our results match known LSH bounds [22, 6], indicating that the asymmetry afforded by $\mathcal{D}$ does not help us when we wish to construct similarity-sensitive families with monotonically increasing CPFs. Implicitly, this result already follows from the space-time tradeoff lower bounds for similarity search shown independently by Andoni et al. [4] and Christiani [13]. As with Lemma 16, the following theorem by O'Donnell [25] is the foundation of our lower bounds.

▶ **Theorem 19** (Generalized Small-Set Expansion). *Let $0 \leq \alpha \leq 1$. Let $A, B \subseteq \{0,1\}^d$ have volumes $\exp(-a^2/2)$, $\exp(-b^2/2)$ and assume $0 \leq \alpha b \leq a \leq b$. Then,*

$$\Pr_{\substack{(\mathbf{x},\mathbf{y}) \\ \alpha\text{-corr.}}}[\mathbf{x} \in A, \mathbf{y} \in B] \geq \exp\left(-\frac{1}{2}\frac{a^2 + 2\alpha ab + b^2}{1-\alpha^2}\right).$$

▶ **Lemma 20.** *For every distribution $\mathcal{D}$ over pairs of functions $h, g\colon \{0,1\}^d \to R$ and every $0 \leq \alpha < 1$ we have that $\hat{f}(\alpha) \leq \hat{f}(0)^{\frac{1-\alpha}{1+\alpha}}$.*

▶ Remark. The restriction from Theorem 19 that $0 \leq \alpha b \leq a \leq b$ can be ignored when attempting to upper bound $\hat{f}$ in the proof of Lemma 20 as further asymmetry does not increase the probability of collision. The solution to the optimization problem underlying the lower bound in Lemma 20 has $a = b$ regardless.

We are now ready to state the corresponding result for similarity-sensitive families.

▶ **Theorem 21.** *Let $0 < \alpha_- < \alpha_+ < 1$ be constants. Then every $(\alpha_-, \alpha_+, f_-, f_+)$-sensitive family $\mathcal{D}$ for $(\{0,1\}^d, \mathrm{sim}_H)$ must satisfy*

$$\frac{\log(1/f_+)}{\log(1/f_-)} \geq \frac{1-\alpha_+}{1+\alpha_+-2\alpha_-} - O(\sqrt{\log(1/f_-)/d}).$$

4 Conclusion

We have initiated the study of *distance-sensitive hashing*, an asymmetric class of LSH methods that considerably extend the capabilities of standard LSH. We proposed some applications and described different constructions of such hash families. Though we settled some basic questions regarding what is possible using distance-sensitive hashing, many questions remain. Ultimately, one would like for a given space a complete characterization of the CPFs that can be achieved, with emphasis on extremal properties. For example: For a CPF that has $f(x) = \Theta(\varepsilon)$ for $x \in [0, r]$, how small a value $\rho(c) = \log(f(r))/\log(f(cr))$ is possible outside of this range? Additionally, our solution to the annulus problem works by combining an LSH and an anti-LSH family to obtain a unimodal family. While we know lower bounds for both, it is not clear whether combining them yields optimal solutions for this problem. Moreover, it is also of interest to consider other applications in approximation algorithms. For example, CPFs appear relevant for efficient kernel density estimation, see, e.g. [20].

Acknowledgement. We thank Thomas D. Ahle for insightful conversations.

References

1. Thomas D. Ahle, Martin Aumüller, and Rasmus Pagh. Parameter-free locality sensitive hashing for spherical range reporting. In *Proceedings of 28th Symposium on Discrete Algorithms (SODA)*, pages 239–256, 2017. doi:10.1137/1.9781611974782.16.
2. Thomas Dybdahl Ahle, Rasmus Pagh, Ilya P. Razenshteyn, and Francesco Silvestri. On the complexity of inner product similarity join. In *Proceedings of 35th ACM Symposium on Principles of Database Systems (PODS)*, pages 151–164, 2016. doi:10.1145/2902251.2902285.
3. Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Proceedings of 47th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 459–468, 2006. doi:10.1109/FOCS.2006.49.
4. Alexandr Andoni, Thijs Laarhoven, Ilya P. Razenshteyn, and Erik Waingarten. Optimal hashing-based time-space trade-offs for approximate near neighbors. In *Proceedings of 28th Annual Symposium on Discrete Algorithms (SODA)*, pages 47–66, 2017. doi:10.1137/1.9781611974782.4.
5. Alexandr Andoni and Ilya P. Razenshteyn. Optimal data-dependent hashing for approximate near neighbors. In *Proceedings of 47th Annual Symposium on Theory of Computing (STOC)*, pages 793–801, 2015. doi:10.1145/2746539.2746553.
6. Alexandr Andoni and Ilya P. Razenshteyn. Tight lower bounds for data-dependent locality-sensitive hashing. In *Proceedings of 32nd International Symposium on Computational Geometry (SoCG)*, pages 9:1–9:11, 2016. doi:10.4230/LIPIcs.SoCG.2016.9.
7. Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In *Proceedings of the 27th Symposium on Discrete Algorithms (SODA)*, pages 10–24, 2016. doi:10.1137/1.9781611974331.ch2.
8. Andrei Z Broder. On the resemblance and containment of documents. In *Proceedings of Compression and Complexity of Sequences*, pages 21–29. IEEE, 1997.
9. Andrei Z Broder, Steven C Glassman, Mark S Manasse, and Geoffrey Zweig. Syntactic clustering of the web. *Computer Networks and ISDN Systems*, 29(8-13):1157–1166, 1997.
10. Moses Charikar. Similarity estimation techniques from rounding algorithms. In *Proceedings of 34th ACM Symposium on Theory of Computing (STOC)*, pages 380–388, 2002.
11. Flavio Chierichetti and Ravi Kumar. Lsh-preserving functions and their applications. *Journal of the ACM (JACM)*, 62(5):33, 2015.
12. Flavio Chierichetti, Alessandro Panconesi, Ravi Kumar, and Erisa Terolli. The distortion of locality sensitive hashing. In *Proceedings of ACM Conference on Innovations in Theoretical Computer Science (ITCS)*, 2017.
13. Tobias Christiani. A framework for similarity search with space-time tradeoffs using locality-sensitive filtering. In *Proceedings of 28th Symposium on Discrete Algorithms (SODA)*, pages 31–46, 2017.
14. Tobias Christiani and Rasmus Pagh. Set similarity search beyond minhash. In *Proceedings of 49th Annual Symposium on Theory of Computing (STOC), to appear*, 2017.
15. Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the 20th Annual Symposium on Computational Geometry (SoCG)*, pages 253–262. ACM, 2004.
16. Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391, 1990.
17. Piotr Indyk. Better algorithms for high-dimensional proximity problems via asymmetric embeddings. In *Proceedings of 14th Symposium on Discrete Algorithms (SODA)*, pages 539–545, 2003.

18 Piotr Indyk, Sepideh Mahabadi, Mohammad Mahdian, and Vahab S Mirrokni. Composable core-sets for diversity and coverage maximization. In *Proceedings of 33rd Symposium on Principles of Database Systems (PODS)*, pages 100–108. ACM, 2014.

19 Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of 30th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 604–613, 1998.

20 C. G. Lambert, S. E. Harrington, C. R. Harvey, and A. Glodjo. Efficient on-line nonparametric kernel density estimation. *Algorithmica*, 25(1):37–57, 1999. doi:10.1007/PL00009282.

21 M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, New York, NY, 2005.

22 R. Motwani, A. Naor, and R. Panigrahy. Lower bounds on locality sensitive hashing. *SIAM J. Discrete Math.*, 21(4):930–935, 2007.

23 Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Chapman & Hall/CRC, 2010.

24 Behnam Neyshabur and Nathan Srebro. On symmetric and asymmetric lshs for inner product search. In *Proceedings 32nd Conference on Machine Learning (ICML)*, pages 1926–1934, 2015.

25 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

26 Ryan O'Donnell, Yi Wu, and Yuan Zhou. Optimal lower bounds for locality-sensitive hashing (except when q is tiny). *ACM Transactions on Computation Theory (TOCT)*, 6(1):5, 2014.

27 Rasmus Pagh, Francesco Silvestri, Johan Sivertsen, and Matthew Skala. Approximate furthest neighbor with application to annulus query. *Information Systems*, 64:152–162, 2017.

28 Ninh Pham and Rasmus Pagh. Fast and scalable polynomial kernels via explicit feature maps. In *Proceedings of 19th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 239–247. ACM, 2013.

29 Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Proceedings of 21st Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1177–1184, 2007. URL: http://papers.nips.cc/paper/3182-random-features-for-large-scale-kernel-machines.

30 I. R. Savage. Mill's ratio for multivariate normal distributions. *Jour. Res. NBS Math. Sci.*, 66(3):93–96, 1962.

31 Anshumali Shrivastava and Ping Li. Asymmetric LSH (ALSH) for sublinear time maximum inner product search (MIPS). In *Proceedings of 27th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2321–2329, 2014.

32 S. J. Szarek and E. Werner. A nonsymmetric correlation inequality for gaussian measure. *Journal of Multivariate Analysis*, 68(2):193–211, 1999.

33 F. Topsœ. *Some Bounds for the Logarithmic Function*, volume 4, pages 137–151. Nova Science, 2007.

34 Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and the closest pair problem. *Journal of the ACM (JACM)*, 62(2):13, 2015.

35 Sudheendra Vijayanarasimhan, Prateek Jain, and Kristen Grauman. Hashing hyperplane queries to near points with applications to large-scale active learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(2):276–288, February 2014. doi:10.1109/TPAMI.2013.121.

36 J. Wang, H. T. Shen, J. Song, and J. Ji. Hashing for similarity search: A survey. *CoRR*, abs/1408.2927, 2014. URL: http://arxiv.org/abs/1408.2927.

A Applications

A.1 Unimodal CPFs for annulus queries

Suppose we are given a distance-sensitive hash family with CPF $f(t)$. In this section we prove Theorem 2 by using a simple adaptation of the standard construction of a near neighbor data structure with LSH. We observe that this data structure improves the trivial scanning solution when $\rho^* = \log(1/f(r))/\log(1/n) < 1$, that is when $f(r) > f(r_-)$ and $f(r) > f(r_+)$. This is satisfied by *unimodal* distance-sensitive hash families, that is when the CPF has a single maximum at t^* and is decreasing for both $t \leq t^*$ and $t \geq t^*$: as soon as t^* lies in the interval (r_-, r_+) we obtain a data structure with sublinear query time.

Proof of Theorem 2. The data structure is a straight-forward adaptation of the construction of a near neighbor data structure using LSH. Associate with each data point $\mathbf{x}$ and query point $\mathbf{y}$ the hash values $h(\mathbf{x})$ and $g(\mathbf{y})$, where (h, g) are independently sampled from the distance-sensitive family. Store all points $\mathbf{x} \in S$ according to $h(\mathbf{x})$ in a hash table. Let $\mathbf{y}$ be the query point and let $\mathbf{x}$ be a point at distance r. Compute $g(\mathbf{y})$ and retrieve all the points from S that have the same hash value. If a point within distance $[r_-, r_+]$ is among the points, output one such point. We expect $\max\{f(r_-)n, f(r_+)n\} \leq 1$ collisions with points at distance at most r_- or at least r_+. The probability of finding $\mathbf{x}$ is at least $f(r)$. Thus, $L = 1/f(r) \leq n^{\rho^*}$ repetitions suffice to retrieve $\mathbf{x}$ with constant probability $1/e$. If the algorithm retrieves more than $8L$ points, none of which is in the interval $[r_-, r_+]$, the algorithm terminates. By Markov's inequality, the probability that the algorithm retrieves $8L$ points, none of which is in the interval $[r_-, r_+]$, is at most $1/8$. ◀

A.2 Plateau CPFs for spherical range reporting

A common problem with LSH-based solutions for reporting all close points is that the CPF is monotonically decreasing starting with collision probability very close to 1 for points that are very close to the query point. On the other hand, many repetitions are necessary to find points at the target distance r. This means that the algorithm retrieves many duplicates for solving range reporting problems. The state-of-the-art data structure for range reporting queries [1] requires $O\left((1 + |S^*|)(n/|S^*|)^\rho\right)$, where S^* is the set of points at distance at most r_+. The following Theorem 3 provides a better analysis of the performance of a standard LSH data structure that takes into account the gap between $f_{\min}$ and $f_{\max}$.

Proof of Theorem 3. We assume that we build a standard LSH data structure as in the proof of Theorem 2 above. We use $1/f_{\min}$ repetitions such that each point within distance r is found with constant probability. Each repetition will contribute $O\left(1 + |S^*| f_{\max}\right)$ points in expectation. Thus, the total cost will be $O\left((1 + |S^*| f_{\max})/f_{\min}\right)$ from which the statement follows. ◀

In particular, if we have a constant bound on $f_{\max}/f_{\min}$ the output sensitivity is optimal. A technique for getting a CPF with a small $f_{\max}/f_{\min}$ gap is to average several unimodal functions: given k such families, we randomly select one of them with probability $1/k$. A graphical example is given in Figure 1. For a more concrete example in Hamming space, consider the scheme given by selecting with probability $1/2$ a standard bit-sampling ($f_1(t) = 1 - t$), and with probability $1/2$ a scheme consisting of bit-sampling and anti bit-sampling ($f_2(t) = t(1 - t)$). The resulting CPF is $f(t) = (1 - t^2)/2$, the gap is $f_{\max}/f_{\min} = 1/(1 - t^2)$. CPFs with constant bounds are implicit in the linear space extremes of the time-space

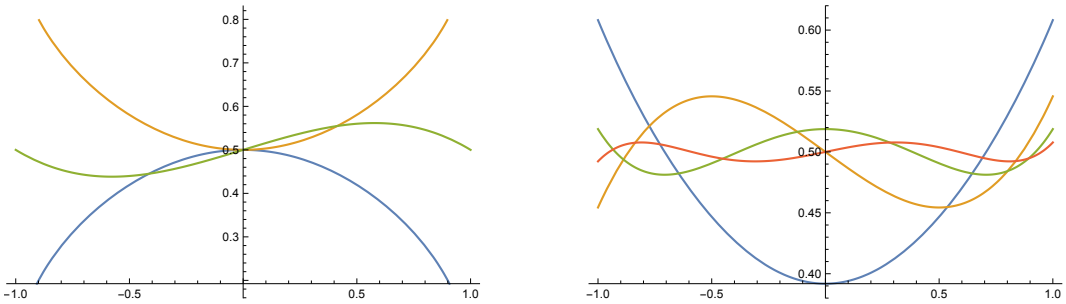

Figure 3 Examples of collision probability functions obtained using Theorem 11. The polynomials used are t^2, $-t^2$, $(-t^3+t^2-t)/3$ (left), and $(2t^2-1)/3$, $(4t^3-3t)/7$, $(8t^4-8t^2+1)/17$, $(16t^5-20t^3+5t)/41$ (right).

trade-off-aware techniques for similarity search [14], but a better value of ρ^* could possibly be obtained by allowing a higher space usage.

Since we use a standard LSH data structure for spherical range reporting, we get the following adaptive variant by using Algorithm 1 from [1].

▶ **Corollary 22.** *Suppose we have a set P of n points and two distances $r < r_+$. Assume we have access to a distance-sensitive hash family with CPF f with $f_{\min} = \inf_{t\in[0,r]} f(t)$. Then Theorem 5.1 from [1] holds for*

$$W_{\text{single}} = \min_{0\le k\le K} \left[f_{\min}^{-k}(1 + \sum_{x\in S} f(\text{dist}(q,x)) \right].$$

B Distance-sensitive constructions

B.1 Proof of Lemma 6

Proof. Let $\mathbf{x}, \mathbf{y}$ be two arbitrary points from X. Part (a): Sample a pair (h_i, g_i) from $\mathcal{D}_i$ for each $i \in \{1,\ldots,n\}$ and set $h(\mathbf{x}) = (h_1(\mathbf{x}),\ldots,h_n(\mathbf{x}))$ and $g(\mathbf{y}) = (g_1(\mathbf{y}),\ldots,g_n(\mathbf{y}))$. We observe that

$$\Pr(h(\mathbf{x}) = g(\mathbf{y})) = \prod_{i=1}^{n} \Pr(h_i(\mathbf{x}) = g_i(\mathbf{y})) = \prod_{i=1}^{n} f_i(\text{dist}(\mathbf{x},\mathbf{y})).$$

Part (b): Pick an integer $i \in \{1,\ldots,n\}$ according to $\{p_i\}$ at random. Then sample a pair (h_i, g_i) from $\mathcal{D}_i$. The hash function pair (h, g) is given by $(i, h_i(\mathbf{x}))$ and $(i, g_i(\mathbf{y}))$. We observe that

$$\Pr(h(\mathbf{x}) = g(\mathbf{y})) = \sum_{i=1}^{n} p_i \Pr_{(h,g)\sim\mathcal{D}_i}(h(\mathbf{x}) = g(\mathbf{y})) = \sum_{i=1}^{n} p_i f_i(\text{dist}(\mathbf{x},\mathbf{y})).$$

◀

B.2 Angular similarity function

This section shows how to derive a distance sensitive scheme with collision probability $\text{sim}(\mathcal{P}(\langle \mathbf{x},\mathbf{y}\rangle))$, when $\sum_{i=0}^{k} |a_i| = 1$. Figure 3 gives some examples of functions that can be obtained from Theorem 11 using SimHash [10].

Proof of Theorem 11. Valiant [34] has shown how, for any real degree-k polynomial p, to construct a pair of mappings $\varphi_1^p, \varphi_2^p : \mathbb{R}^d \to \mathbb{R}^D$, where $D = O(d^k)$, such that $\varphi_1^p(\mathbf{x}) \cdot \varphi_2^p(\mathbf{y}) = \mathcal{P}(\langle \mathbf{x}, \mathbf{y} \rangle)$. For completeness we outline the argument here: First consider the monomial $\mathcal{P}(t) = a_k t^k$. For $\mathbf{x} \in \mathbb{R}^d$, let $\mathbf{x}^{(k)}$ denote the vector of dimension d^k indexed by vectors $\mathbf{i} = (i_1, \ldots, i_k) \in [d]^k$, where $\mathbf{x}_{\mathbf{i}}^{(k)} = \prod_{j=1}^k x_{i_j}$. It is easy to verify that $\langle \mathbf{x}^{(k)}, \mathbf{y}^{(k)} \rangle = (\langle \mathbf{x}, \mathbf{y} \rangle)^k$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. With this notation in place we can define $\varphi_1^p(\mathbf{x}) = \sqrt{|a_i|}\, \mathbf{x}^{(k)}$ and $\varphi_2^p(\mathbf{y}) = (a_i / \sqrt{|a_i|})\, \mathbf{y}^{(k)}$ which satisfy $\varphi_1^p(\mathbf{x}) \cdot \varphi_2^p(\mathbf{x}) = a_i (\langle \mathbf{x}, \mathbf{y}, \rangle)^k$. The asymmetry of the mapping is essential to allow a negative coefficient a_k. To handle an arbitrary real polynomial $\mathcal{P}(t) = \sum_{i=0}^k a_i t^i$ we simply concatenate vectors corresponding to each monomial, obtaining a vector of dimension $\sum_{i=0}^k d^i = O(d^k)$.

Observe that $||\mathbf{x}^{(k)}||_2^2 = \langle \mathbf{x}^{(k)}, \mathbf{x}^{(k)} \rangle = (\langle \mathbf{x}, \mathbf{x}, \rangle)^k = ||\mathbf{x}||_2^{2k}$. This means that for $||\mathbf{x}||_2^2 = 1$ we have $||\varphi_1^p(\mathbf{x})||^2 = \sum_{i=0}^k \sqrt{|a_i|}^2 = 1$, using the assumption $\sum_{i=0}^k |a_i| = 1$. Similarly, for $||\mathbf{y}||_2^2 = 1$ we have $||\varphi_2^p(\mathbf{y})||^2 = \sum_{i=0}^k (a_i / \sqrt{|a_i|})^2 = \sum_{i=0}^k |a_i| = 1$. Thus, φ_1^p and φ_2^p map S^{d-1} to S^{D-1}.

Our family $\mathcal{F}$ samples a function s from the distribution $\mathcal{S}$ corresponding to sim and constructs the function pair (h, g) with $h(\mathbf{x}) = s(\varphi_1^p(\mathbf{x}))$, $g(\mathbf{y}) = s(\varphi_2^p(\mathbf{y}))$. Using the properties of the functions involved we have

$$\Pr[h(\mathbf{x}) = g(\mathbf{y})] = \mathrm{sim}(\langle \varphi_1^p(\mathbf{x}), \varphi_2^p(\mathbf{y}) \rangle) = \mathrm{sim}(\mathcal{P}(\langle \mathbf{x}, \mathbf{y} \rangle)) \ .$$

◀

B.3 Hamming distance functions

Proof of Theorem 12. We initially assume that $a_0 \neq 0$ (i.e., 0 is not a root of $\mathcal{P}(t)$), and then remove this assumption at the end of the proof. We recall that a root of $\mathcal{P}(t)$ can appear with multiplicity larger than 1 and that, by the complex conjugate root theorem, if $z = a + bi$ is a complex root then so is its conjugate $z' = a - bi$. We let Z be the multiset containing the k roots of $\mathcal{P}(t)$, with Z_{r+} and Z_{r-} being the multiset of positive and negative real roots, respectively, and with Z_c being the multiset consisting of pairs of conjugate complex roots. By factoring $\mathcal{P}(t)$, we get:

$$\mathcal{P}(t) = a_k \prod_{z \in Z} (t - z) = |a_k| \prod_{z \in Z_{r+}} (z - t) \prod_{z \in Z_{r-}} (t + |z|) \prod_{z = a+bi \in Z_c} (t^2 - 2at + a^2 + b^2), \quad (9)$$

where in the last step we exploited that $a_k \prod_{z \in Z_{r+}} (z - t) = |a_k| \prod_{z \in Z_{r+}} (t - z) > 0$. Indeed, $\mathcal{P}(t)$ is positive in $(0, 1)$ and the multiplicative terms associated with complex and negative real roots are positive in this range; this implies that the remaining terms are positive as well.

We need to introduce scaled and biased variations of bit-sampling or anti bit-sampling. Anti-bit sampling with scaling factor $\alpha \in [0, 1]$ and bias $\beta \in [0, 1]$ has the CPF $f(t) = \beta/2 + \alpha t/2$ and is given by randomly selecting one of following two schemes: (1) with probability $1/2$, the scheme is a standard hashing that maps data and query points to 0 with probability β, and otherwise to 0 and 1 respectively; (2) with probability $1/2$, the scheme is anti bit-sampling where the sampled bit is set to 0 with probability $1 - \alpha$ on both data and query points, or kept unchanged otherwise. Similarly, bit-sampling with scaling factor $\alpha \in [0, 1]$ has the CPF $f(t) = (1 - \alpha t)$ and is given by using bit-sampling, where the sampled bit is set to 0 with probability $1 - \alpha$ on both data and query points. (We do not need a biased version of bit-sampling.)

We now assign to each multiplicative term of (9) a scaled and biased version of bit-sampling or anti bit-sampling as follows:

- z **is real and** $z < -1$**.** We assign to z an anti bit-sampling with bias 1 and scaling factor $1/|z| \leq 1$: the CPF is $S_1(t,z) = (1/2 + t/(2|z|))$, and we have $(t + |z|) = 2|z|S_1(t,z)$.
- z **is real and** $-1 \leq z < 0$**.** We assign to z an anti bit-sampling with bias $|z| \leq 1$ and scaling factor 1: the CPF is $S_2(t,z) = |z|/2 + t/2$, and we have $(t + |z|) = 2S_2(t,z)$.
- z **is real and** $z \geq 1$**.** We assign to z a bit-sampling with scaling factor $1/z \leq 1$: the CPF is $S_3(t,z) = (1 - t/z)$, and we have $(t - z) = zS_3(t,z)$.
- (z, z') **are conjugate complex roots and** $\text{Real}(z) < -1$**.** Let $z = a + bi$ and $z' = a - bi$. The assigned scheme has CPF $S_4(t,z) = \left(\frac{b^2}{4(a^2+b^2)} + \frac{a^2}{a^2+b^2}\left(\frac{x}{2|a|} + \frac{1}{2}\right)^2\right)$ and is obtained as follows: with probability $b^2/(a^2+b^2)$, the scheme maps data and query points to 0 and 0 with probability 1/4, or to 0 and 1 with probability 3/4; with probability $a^2/(a^2+b^2)$, the schemes consists of the concatenation of two anti bit-sampling with bias 1 and scaling factor $1/|a|$. Note that $t^2 - 2at + a^2 + b^2 = 4(a^2+b^2)S_4(t,z)$.
- (z, z') **are conjugate complex roots and** $\text{Real}(z) \geq 1$**.** The scheme is similar to the previous one where we use two bit-sampling with scaling factor $1/a$ instead of the anti bit-sampling. The CPF is $S_5(t,z) = \left(\frac{b^2}{a^2+b^2} + \frac{a^2}{a^2+b^2}\left(1 - \frac{x}{a}\right)^2\right)$, and we get $t^2 - 2at + a^2 + b^2 = (a^2+b^2)S_5(t,z)$.
- (z, z') **are conjugate complex roots,** $-1 \leq \text{Real}(z) \leq 0$**, and** $|z| = a^2 + b^2 \geq 1$**.** We assign the following scheme with CPF $S_6(t,z) = \left(\frac{x^2}{4(a^2+b^2)} + \frac{|a|x}{2(a^2+b^2)} + \frac{1}{4}\right)$: with probability 1/4 the scheme maps data and query points to 0; with probability 1/2, the scheme consists of anti bit-sampling with bias 0 and scaling factor $|a|/(a^2+b^2) \leq 1$; with probability 1/4 the scheme consists of two anti bit-sampling with bias 0 and scaling factor $\sqrt{a^2+b^2}$ each. We have $t^2 - 2at + a^2 + b^2 = 4(a^2+b^2)S_6(t,z)$.
- (z, z') **are conjugate complex roots,** $-1 \leq \text{Real}(z) \leq 0$**, and** $|z| = a^2 + b^2 < 1$**.** We use the scheme of the previous point with different parameters, giving CPF $S_7(t,z) = \left(\frac{x^2}{4} + \frac{|a|x}{2} + \frac{a^2+b^2}{4}\right)$. The scheme is the following: with probability 1/4, the scheme is a standard hashing scheme where data points are always mapped to 0 and where a query point is mapped to 0 with probability $a^2 + b^2$ and to 1 with probability $1 - a^2 + b^2$; with probability 1/2, the scheme consists of anti bit-sampling with bias 0 and scaling factor $|a| \leq 1$; with probability 1/4, the scheme consists of two anti bit-sampling with bias 0 and scaling factor 1 each. We have $t^2 - 2at + a^2 + b^2 = 4S_7(t,z)$.

Consider the scheme obtained by concatenating the above ones for each real root and each pair of conjugate roots. Its CPF is $S(t) = \prod_{i=1}^{6} \prod_{z \in Z_i} S_i(t,z)$, where Z_i contains root with CPF S_i. Then, by letting ψ denote the number of roots with negative real part, we get from Equation 9:

$$\mathcal{P}(t) = \left(2^{\psi}|a_k| \prod_{z \in Z, |\text{Real}(z)|>1} |z|\right) S(t) = \Delta S(t).$$

Consider now $a_k = 0$ and let ℓ be the largest value such that $\mathcal{P}(t) = t^{\ell}\mathcal{P}'(x)$ with $\mathcal{P}'(0) \neq 0$. We get the claimed result by concatenating ℓ anti bit-sampling, which gives a CPF of x^{ℓ}, and the scheme for $\mathcal{P}'(t)$ obtained by the procedure described above. ◀

C CPF bounds for the unit sphere

Gaussian tail bounds. We will make use the following tail bounds for the univariate and bivariate normal distribution.

▶ **Lemma 23** (Follows Szarek & Werner [32]). *Let Z be a standard normal random variable. Then, for every $t \geq 0$ we have that*

$$\frac{1}{\sqrt{2\pi}}\frac{1}{t+1}e^{-t^2/2} \leq \Pr[Z \geq t] \leq \frac{1}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/2}.$$

▶ **Lemma 24** (Savage [30]). *Let $\alpha \in (-1,1)$ and let $Z_1, Z_2 \sim \mathcal{N}(0,1)$. Define $X_1 = Z_1$ and $X_2 = \alpha Z_1 + (1-\alpha^2)Z_2$. Then, for every $t > 0$ we have that*

$$\Pr[X_1 \geq t \wedge X_2 \geq t] < \frac{1}{2\pi t^2}\frac{(1+\alpha)^2}{\sqrt{1-\alpha^2}}\exp\left(-\frac{t^2}{1+\alpha}\right),$$
$$\Pr[X_1 \geq t \wedge X_2 \geq t] > \left(1 - \frac{(2-\alpha)(1+\alpha)}{1-\alpha}\frac{1}{t^2}\right)\frac{1}{2\pi t^2}\frac{(1+\alpha)^2}{\sqrt{1-\alpha^2}}\exp\left(-\frac{t^2}{1+\alpha}\right).$$

▶ **Corollary 25.** *The above bounds apply to $\Pr[X_1 \geq t \wedge X_2 \leq -t]$ if we replace all occurrences of α with $-\alpha$.*

Bounding the CPF. We proceed by deriving upper and lower bounds on the collision probability.

$$\begin{aligned} f_+(\alpha) &\leq \frac{\Pr[\langle \mathbf{z}, \mathbf{x}\rangle \geq t \wedge \langle \mathbf{z}, \mathbf{y}\rangle \geq t]}{\Pr[\langle \mathbf{z}, \mathbf{x}\rangle \geq t]} \\ &\leq \frac{1}{\sqrt{2\pi}}\frac{t+1}{t^2}\frac{(1+\alpha)^2}{\sqrt{1-\alpha^2}}\exp\left(-\frac{1-\alpha}{1+\alpha}\frac{t^2}{2}\right). \end{aligned}$$

We derive the lower bound in stages.

$$\Pr[h(\mathbf{x}) = g(\mathbf{y})] \geq \frac{\Pr[\langle \mathbf{z}, \mathbf{x}\rangle \geq t \wedge \langle \mathbf{z}, \mathbf{y}\rangle \geq t]}{2\Pr[\langle \mathbf{z}, \mathbf{x}\rangle \geq t]} - \Pr[h(\mathbf{x}) > m \vee g(\mathbf{y}) > m].$$

The first part is lower bounded by

$$\begin{aligned} &\frac{\Pr[\langle \mathbf{z}, \mathbf{x}\rangle \geq t \wedge \langle \mathbf{z}, \mathbf{y}\rangle \geq t]}{2\Pr[\langle \mathbf{z}, \mathbf{x}\rangle \geq t]} \geq \\ &\left(1 - \frac{(2-\alpha)(1+\alpha)}{1-\alpha}\frac{1}{t^2}\right)\frac{1}{2\sqrt{2\pi}}\frac{1}{t}\frac{(1+\alpha)^2}{\sqrt{1-\alpha^2}}\exp\left(-\frac{1-\alpha}{1+\alpha}\frac{t^2}{2}\right). \end{aligned}$$

The probability of not being captured by a projection depends on the number of projections m. In order to make this probability negligible we can set $m = \lceil t^3/p' \rceil$ where p' denotes the lower bound from Lemma 23.

$$\begin{aligned} \Pr[h(\mathbf{x}) > m \vee g(\mathbf{y}) > m] &\leq 2(1 - \Pr[\langle \mathbf{z}, \mathbf{x}\rangle \geq t])^m \\ &\leq 2(1-p')^{t^3/p'} \\ &\leq 2e^{-t^3}. \end{aligned}$$

D Annulus search on the unit sphere

We will construct a distance sensitive family $\mathcal{D}$ for solving the approximate annulus search problem. Let $\mathcal{D}_+$ be parameterized by t_+ and let $\mathcal{D}_-$ be parameterized by t_-. To sample a pair of functions (h, g) from $\mathcal{D}$ we independently sample a pair (h_+, g_+) from $\mathcal{D}_+$ and (h_-, g_-) from $\mathcal{D}_-$ and define (h, g) by $h(\mathbf{x}) = (h_+(\mathbf{x}), h_-(\mathbf{x}))$ and $g(\mathbf{x}) = (g_+(\mathbf{x}), g_-(\mathbf{x}))$.

Let $f(\alpha)$ denote the CPF of $\mathcal{D}$. We would like to be able to parameterize $\mathcal{D}$ such that $f(\alpha)$ is somewhat symmetric around a unique maximum value of α. It can be verified from

the definition of $\mathcal{D}_+$ that $p_+(-1) = 0$ which implies that $f(-1) = f(1) = 0$. If we ignore lower order terms and define $\gamma > 0$ by $t_- = \gamma t_+$, then we can see that

$$\ln(1/f(\alpha)) \approx \frac{1-\alpha}{1+\alpha}\frac{t_+^2}{2} + \frac{1+\alpha}{1-\alpha}\frac{\gamma^2 t_+^2}{2}.$$

For simplicity, temporarily define $a(\alpha) = (1-\alpha)/(1+\alpha) > 0$. Given a fixed γ, the equation $a + \gamma^2/a$ is minimized (corresponding to approximately maximizing $f(\alpha)$) when setting $a = \gamma$. Let $\alpha_{\max} \in (-1, 1)$ and set $\gamma = a_{\max} = (1-\alpha_{\max})/(1+\alpha_{\max})$. To find values $\alpha_- < \alpha_{\max} < \alpha_+$ where $\ln(1/f(\alpha_-)) \approx \ln(1/f(\alpha_+))$ note that this condition holds for every $s > 1$ when we set $a_- = s a_{\max}$ and $a_+ = (1/s)a_{\max}$. We therefore parameterize $\mathcal{D}$ by $\alpha_{\max} \in (-1, 1)$ and $t > 0$ and set $t_+ = t$ and $t_- = (1-\alpha_{\max})/(1+\alpha_{\max})t_+$. By combining our bounds from Lemma 9 with the above observations we are able to obtain the following theorem which immediately yields a solution to the approximate annulus search problem.

▶ **Theorem 26.** *For every choice of $t > 0$ and constant $\alpha_{\max} \in (-1, 1)$ the family $\mathcal{D}$ satisfies the following: For every choice of constant $s > 1$ consider the interval $[\alpha_-, \alpha_+]$ defined to contain every α such that $\frac{1}{s}\frac{1-\alpha_{\max}}{1+\alpha_{\max}} \leq \frac{1-\alpha}{1+\alpha} \leq s\frac{1-\alpha_{\max}}{1+\alpha_{\max}}$, then*

- *For $\alpha \in [\alpha_-, \alpha_+]$ we have that $f(\alpha) = \Omega((1/t^2)\exp(-(s+1/s)\frac{1-\alpha_{\max}}{1+\alpha_{\max}}\frac{t^2}{2}))$.*
- *For $\alpha \notin [\alpha_-, \alpha_+]$ we have that $f(\alpha) = O((1/t^2)\exp(-(s+1/s)\frac{1-\alpha_{\max}}{1+\alpha_{\max}}\frac{t^2}{2}))$.*

The complexity of sampling, storing, and evaluating a pair of functions $(h, g) \in \mathcal{D}$ is $O(dt^4 e^{t^2/2})$.

See Figure 4 for a visual representation of the annulus for given parameters $\alpha_{\max}$ and s.

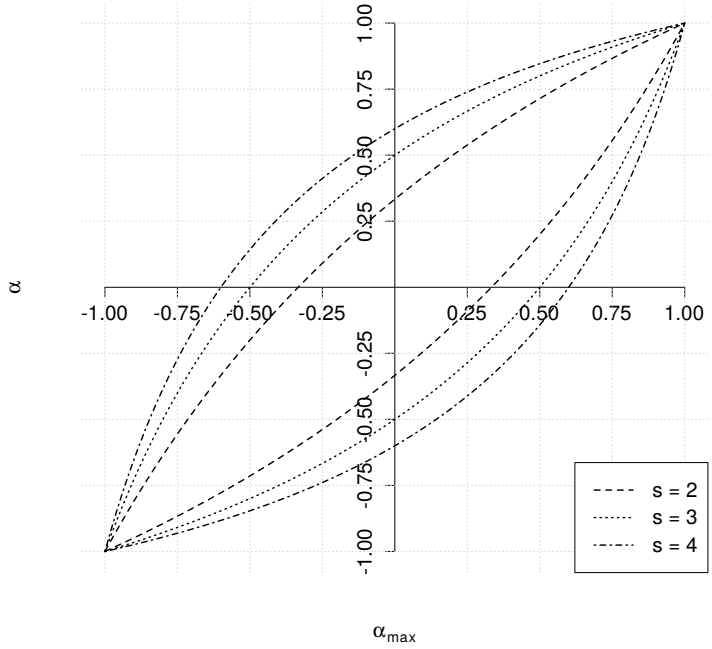

Figure 4 Annuli as defined in Theorem 26 for every value of $\alpha_{\max}$ and $s = 2, 3, 4$.

We define an approximate annulus search problem for similarity spaces and proceed by applying Theorem 26 to provide a solution for the unit sphere, resulting in Theorem 28.

▶ **Definition 27.** Let $\beta_- < \alpha_- \le \alpha_+ < \beta_+$ be given real numbers. For a set P of n points in a similarity space (X, sim) a solution to the $((\alpha_-, \alpha_+), (\beta_-, \beta_+))$-annulus search problem is a data structure that supports a query operation that takes as input a point $\mathbf{x} \in X$ and if there exists a point $\mathbf{y} \in P$ such that $\alpha_- \le \mathrm{sim}(\mathbf{x}, \mathbf{y}) \le \alpha_+$ then it returns a point $\mathbf{y}' \in P$ such that $\beta_- \le \mathrm{sim}(\mathbf{x}, \mathbf{y}') \le \beta_+$.

▶ **Theorem 28.** *For every choice of constants $-1 < \beta_- < \alpha_- < \alpha_+ < \beta_+ < 1$ such that $\frac{1-\alpha_-}{1+\alpha_-}\frac{1-\alpha_+}{1+\alpha_+} = \frac{1-\beta_-}{1+\beta_-}\frac{1-\beta_+}{1+\beta_+}$ we can solve the $((\alpha_+, \alpha_-), (\beta_+, \beta_-))$-annulus problem for $(\mathbb{S}^{d-1}, \langle \cdot, \cdot \rangle)$ with space usage $dn + n^{1+\rho+o(1)}$ words and query time $dn^{\rho+o(1)}$ where*

$$\rho = \frac{c_\alpha + 1/c_\alpha}{c_\beta + 1/c_\beta} \le \frac{2}{c + 1/c}$$

and we define $1 < c_\alpha < c_\beta$ by $c_\alpha = \sqrt{\frac{1-\alpha_-}{1+\alpha_-}/\frac{1-\alpha_+}{1+\alpha_+}}$, $c_\beta = \sqrt{\frac{1-\beta_-}{1+\beta_-}/\frac{1-\beta_+}{1+\beta_+}}$, and $c = c_\beta/c_\alpha$.

E Details of the lower bound

E.1 Optimal partition pair problem

The lower bound in Lemma 16 relies on the following lemma which we prove by showing that the pair of partitions induced by $(h, g) \in \mathcal{D}$ minimizes a convex function of the probability of the parts, under the additional constraint that $\Pr[h(\mathbf{x}) = g(\mathbf{y})] = p$, when all parts are of equal volume and as small as possible.

▶ **Lemma 29.** *For every $0 \le \alpha < 1$, every pair of functions $h, g \colon \{0,1\}^d \to R$ must satisfy*

$$\sum_{i \in R} \left(\Pr_{\substack{(\mathbf{x},\mathbf{y}) \\ 0\text{-}corr.}} [h(\mathbf{x}) = g(\mathbf{y}) = i] \right)^{\frac{1}{1-\alpha}} \ge \left(\sum_{i \in R} \Pr_{\substack{(\mathbf{x},\mathbf{y}) \\ 0\text{-}corr.}} [h(\mathbf{x}) = g(\mathbf{y}) = i] \right)^{\frac{1+\alpha}{1-\alpha}}.$$

Proof. For $i \in R$ define $A_i = |h^{-1}(i)|/2^d$ and $B_i = |g^{-1}(i)|/2^d$. Given some value of $p = \sum_i A_i B_i$ we would like to choose the partition to minimize $\sum_i (A_i B_i)^{\frac{1}{1-\alpha}}$. In order to avoid complications due to integrality constraints on the number of partitions, we define a weighted version of the problem with the property that its solution never exceeds the solution of the original problem.

$$\begin{aligned} \underset{w_i, A_i, B_i}{\text{minimize}} \quad & \sum_i w_i \, (A_i B_i)^{\frac{1}{1-\alpha}} \\ \text{subject to} \quad & \sum_i w_i A_i B_i = p, \\ & \sum_i w_i A_i = 1, \ \sum_i w_i B_i = 1. \end{aligned}$$

To ease notation and due to symmetry we will suppress different values $i, i' \in R$ in what

follows. The Lagrangian for this problem and its first order partial derivatives are given by

$$\begin{aligned}
L(w_i, A_i, B_i, \lambda_A, \lambda_B, \lambda_p) &= \sum_i w_i \left(A_i B_i\right)^{\frac{1}{1-\alpha}} - \lambda_A(\sum_i w_i A_i - 1) \\
&\quad - \lambda_B(\sum_i w_i B_i - 1) - \lambda_p(\sum_i w_i A_i B_i - p), \\
\frac{\partial L}{\partial w_i} &= (A_i B_i)^{\frac{1}{1-\alpha}} - \lambda_A A_i - \lambda_B B_i - \lambda_p A_i B_i, \\
\frac{\partial L}{\partial A_i} &= w_i \left(\frac{1}{1-\alpha} (A_i B_i)^{\frac{\alpha}{1-\alpha}} B_i - \lambda_A - \lambda_p B_i \right), \\
\frac{\partial L}{\partial B_i} &= w_i \left(\frac{1}{1-\alpha} (A_i B_i)^{\frac{\alpha}{1-\alpha}} A_i - \lambda_A - \lambda_p A_i \right), \\
\frac{\partial L}{\partial \lambda_A} &= \sum_i w_i A_i - 1, \\
\frac{\partial L}{\partial \lambda_B} &= \sum_i w_i B_i - 1, \\
\frac{\partial L}{\partial \lambda_p} &= \sum_i w_i A_i B_i - p.
\end{aligned}$$

We will proceed by deriving necessary conditions for a solution by manipulating the first order conditions that all the partial derivatives are equal to zero. Consider the following sum:

$$\sum_i \left(\frac{\partial L}{\partial A_i} \right) A_i = \frac{1}{1-\alpha} \sum_i w_i \left(A_i B_i\right) - \lambda_A \sum_i w_i A_i - \lambda_p \sum_i w_i A_i B_i.$$

Setting this equal to the corresponding sum for B_i allows us to conclude that $\lambda_A = \lambda_B$. Setting $\frac{\partial L}{\partial A_i} A_i = \frac{\partial L}{\partial B_i} B_i$ allows us to conclude that $w_i A_i = w_i B_i$. Consider now an i for which $w_i \neq 0$ which implies that $A_i = B_i$. Further assume that $A_i \neq 0$ since the case of $A_i = B_i = 0$ will not affect the problem. Setting the first order conditions $\frac{\partial L}{\partial w_i} = 0$ and $\frac{\partial L}{\partial A_i} = 0$ equal to each other we get

$$A_i^{\frac{2}{1-\alpha}} - 2\lambda_A A_i - \lambda_p A_i^2 = \frac{1}{1-\alpha} A_i^{\frac{1+\alpha}{1-\alpha}} - \lambda_A - \lambda_p A_i.$$

This allows us to conclude that

$$\lambda_A = -\frac{\alpha}{1-\alpha} A_i^{\frac{1+\alpha}{1-\alpha}}, \quad \lambda_p = \frac{1+\alpha}{1-\alpha} A_i^{\frac{2\alpha}{1-\alpha}}.$$

Because the same derivation holds for every i, for $i \neq j$ and under the assumptions that $w_i, A_i \neq 0$ and $w_j, A_j \neq 0$ it must hold that $A_i = A_j$. We can therefore restrict our attention to the case of a single w_i and $A_i = B_i$ since all other solutions will result in the same value in the optimum. From the first order conditions we have that $w_i A_i = 1$ and $w_i A_i^2 = p$ and it is therefore easy to see that an optimal solution is

$$w_1 = 1/p, \qquad A_1 = B_1 = p$$

with everything else set to zero. In the unweighted, original formulation of the problem, this corresponds to the partitions induced by h, g each consisting of $1/p$ equal parts of volume p. ◀

E.2 Lower bounding the CPF in Hamming space

Here we prove Theorem 18 for Hamming space, using the concept of a (r, c, p, q)-insensitive family.

▶ **Definition 30** (Anti Locality-Sensitive Hashing)**.** A distribution $\mathcal{A}$ over pairs of functions $h, g\colon X \to R$ is (r, c, p, q)-*insensitive* for (X, dist) if for all pairs of points $\mathbf{x}, \mathbf{y}$ and (h, g) sampled randomly from $\mathcal{A}$ we have that:

- If $\mathrm{dist}(\mathbf{x}, \mathbf{y}) \geq r$ then $\Pr[h(\mathbf{x}) = g(\mathbf{y})] \geq p$.
- If $\mathrm{dist}(\mathbf{x}, \mathbf{y}) \leq r/c$ then $\Pr[h(\mathbf{x}) = g(\mathbf{y})] \leq q$.

We prove the following theorem, which can easily be converted to Theorem 18 in the main text.

▶ **Theorem 31.** *For every constant $\varepsilon > 0$, every (r, c, p, q)-insensitive family $\mathcal{A}$ for $\{0,1\}^d$ under Hamming distance with $r \leq (1-\varepsilon)d/2$ must satisfy*

$$\rho(\mathcal{A}) = \frac{\log 1/p}{\log 1/q} \geq \frac{1}{2c-1} - O(\sqrt{(c/r)\log(1/q)}).$$

Proof. Given $\mathcal{A}$ we define a distribution $\hat{\mathcal{A}}$ over pairs of functions $\hat{h}, \hat{g}\colon \{0,1\}^{\hat{d}} \to R$ where $\hat{d} \leq d$ remains to be determined. We sample a pair of functions $(\hat{h}, \hat{g})$ from $\hat{\mathcal{A}}$ by sampling (h, g) from $\mathcal{H}$ and setting $\hat{h}(\mathbf{x}) = h(\mathbf{x} \circ \mathbf{1})$ and similarly $\hat{g}(\mathbf{x}) = g(\mathbf{x} \circ \mathbf{1})$ where $\mathbf{1}$ denotes the $(d - \hat{d})$-dimensional all-ones vector. We will now turn to the process of relating p to $\hat{p} = \hat{f}(0)$ and q to $\hat{q} = \hat{f}(\alpha)$ for $\hat{\mathcal{A}}$.

Let $0 < \varepsilon_p < 1$ and set $\hat{d} = \lceil 2r/(1-\varepsilon_p) \rceil$. Then by applying standard Chernoff bounds we get

$$\Pr_{(\mathbf{x},\mathbf{y})\ 0\text{-corr.}}[\mathrm{dist}(\mathbf{x}, \mathbf{y}) \leq r] \leq \exp\left(-\frac{\varepsilon_p^2}{1-\varepsilon_p}\frac{r}{2}\right).$$

For convenience, define $\delta_p = \exp\left(-\frac{\varepsilon_p^2}{1-\varepsilon_p}\frac{r}{2}\right)$. We now have $\hat{p} \geq (1-\delta_p)p$.

In order to tie $\hat{q}$ to q we consider the probability of α-correlated points having distance greater than r/c. The expected Hamming distance of α-correlated $(\mathbf{x}, \mathbf{y})$ in $\hat{d}$ dimensions is $\hat{d}(1-\alpha)/2$. We would like to set α such that the probability of the distance exceeding r/c is small. Let X denote $\mathrm{dist}(\mathbf{x}, \mathbf{y})$, then the standard Chernoff bound states that:

$$\Pr[X \geq (1+\varepsilon)\mu] \leq e^{-\varepsilon^2\mu/3}.$$

For a parameter $0 < \varepsilon_q < 1$ we set α such that the following is satisfied:

$$(1+\varepsilon_q)\mu \geq (1+\varepsilon_q)\frac{2r}{1-\varepsilon_p}\frac{1-\alpha}{2} \geq r/c.$$

This results in a value of $\alpha = 1 - \frac{1-\varepsilon_p}{1+\varepsilon_q}\frac{1}{c}$ and we observe that

$$\delta_q \leq \exp(-\varepsilon_q^2\mu/3) \leq \exp(-\frac{\varepsilon_q^2}{1+\varepsilon_q}\frac{r}{3c}).$$

It follows that

$$\hat{q} \leq (1-\delta_q)q + \delta_q.$$

Let us summarize what we know so far:

$$\begin{aligned}
&\hat{p} \geq (1-\delta_p)p \\
&\hat{q} \leq (1-\delta_q)q + \delta_q \leq q(1+\delta_q/q) \\
&0 < \varepsilon_p, \varepsilon_q < 1 \\
&\delta_p \leq \exp\left(-\frac{\varepsilon_p^2}{1-\varepsilon_p}\frac{r}{2}\right) \\
&\delta_q \leq \exp\left(-\frac{\varepsilon_q^2}{1+\varepsilon_q}\frac{r}{3c}\right) \\
&\alpha = 1 - \frac{1-\varepsilon_p}{1+\varepsilon_q}\frac{1}{c} \\
&\hat{q} \geq \hat{p}^{\frac{1+\alpha}{1-\alpha}}.
\end{aligned}$$

We assume that $0 < q < p < 1$ and furthermore, without loss of generality we can assume that $q \leq 1/e$ due to the powering technique. In our derivations we also assume that $\delta_p \leq 1/2$ and $\delta_q \leq 1/(2e)$ such that $\delta_q/q \leq 1/2$. This will later be implicit in the statement of the result in big-Oh notation. From our assumptions and standard bounds on the natural logarithm we are able to derive the following:

$$\begin{aligned}
\frac{\ln(1/p)}{\ln(1/q)} &\geq \frac{\ln(1-\delta_p)\ln(1/\hat{p})}{\ln(1/q)} \\
&\geq \frac{\ln(1/\hat{p})}{\ln(1/q)} - 2\delta_p \\
&\geq \frac{\ln(1/\hat{p})}{\ln(1+\delta_q/q)+\ln(1/\hat{q})} - 2\delta_p \\
&\geq \frac{\ln(1/\hat{p})}{\ln(1/\hat{q})}\left(1 - \frac{\ln(1+\delta_q/q)}{\ln(1/\hat{q})}\right) - 2\delta_p \\
&\geq \frac{\ln(1/\hat{p})}{\ln(1/\hat{q})} - \frac{\ln(1+\delta_q/q)}{\ln(1/(1+\delta_q/q)q)} - 2\delta_p \qquad (10) \\
&\geq \frac{\ln(1/\hat{p})}{\ln(1/\hat{q})} - 2\delta_q/q - 2\delta_p.
\end{aligned}$$

In equation (10) we use the statement itself combined with our assumptions on p and q to deduce that

$$1 > \frac{\ln(1/p)}{\ln(1/q)} \geq \frac{\ln(1/\hat{p})}{\ln(1/\hat{q})}.$$

We proceed by lower bounding $\hat{\rho}$. Temporarily define $1-\varepsilon' = \frac{1-\varepsilon_p}{1+\varepsilon_q}$ and observe that

$$\begin{aligned}
\frac{\ln(1/\hat{p})}{\ln(1/\hat{q})} &\geq \frac{1-\alpha}{1+\alpha} \\
&= \frac{(1-\varepsilon')/c}{2-(1-\varepsilon')/c} \\
&\geq \frac{1}{2c-1} - \frac{\varepsilon'}{(2c-1)^2} - \frac{\varepsilon'}{2c-1}.
\end{aligned}$$

We have that

$$\varepsilon' = 1 - \frac{1-\varepsilon_p}{1+\varepsilon_q} = \frac{1+\varepsilon_q-(1-\varepsilon_p)}{1+\varepsilon_q} \leq \varepsilon_q + \varepsilon_p,$$

and combining these bounds results in

$$\frac{\ln(1/p)}{\ln(1/q)} \geq \frac{1}{2c-1} - 2(\varepsilon_q + \varepsilon_p - \delta_q/q - \delta_p).$$

We can now set $\varepsilon_q = \varepsilon_p = K \cdot \sqrt{(c/r)\ln(1/q)}$ for some universal constant K to obtain Theorem 18. ◀

E.3 Tools

For completeness we here state some standard technical lemmas used in our derivation of the lower bound.

▶ **Lemma 32** (Chernoff [21, Theorems 4.4 and 4.5]). *Let $X_1, \ldots, X_n$ be independent Poisson trials and define $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathrm{E}[X]$. Then, for $0 < \varepsilon < 1$ we have*

- $\Pr[X \geq (1+\varepsilon)\mu] \leq e^{-\varepsilon^2\mu/3}$.
- $\Pr[X \leq (1-\varepsilon)\mu] \leq e^{-\varepsilon^2\mu/2}$.

Bounding the natural logarithm and approximating the exponential function:

▶ **Lemma 33** ([33]). *For $x > -1$ we have that $\frac{x}{1+x} \leq \ln(1+x) \leq x$.*

▶ **Lemma 34** ([23, Prop. B.3]). *For all $t, n \in \mathbb{R}$ with $|t| \leq n$ we have that $e^t(1 - \frac{t^2}{n}) \leq (1 + \frac{t}{n})^n \leq e^t$.*