Abstract

An $\varepsilon$-approximate incidence between a point and some geometric object (line, circle, plane, sphere) occurs when the point and the object lie at distance at most $\varepsilon$ from each other. Given a set of points and a set of objects, computing the approximate incidences between them is a major step in many database and web-based applications in computer vision and graphics, including robust model fitting, approximate point pattern matching, and estimating the fundamental matrix in epipolar (stereo) geometry.

In a typical approximate incidence problem of this sort, we are given a set $P$ of $m$ points in two or three dimensions, a set $S$ of $n$ objects (lines, circles, planes, spheres), and an error parameter $\varepsilon > 0$, and our goal is to report all pairs $(p,s) \in P \times S$ that lie at distance at most $\varepsilon$ from one another. We present efficient output-sensitive approximation algorithms for quite a few cases, including points and lines or circles in the plane, and points and planes, spheres, lines, or circles in three dimensions. Several of these cases arise in the applications mentioned above. Our algorithms report all pairs at distance $\leq \varepsilon$, but may also report additional pairs, all of which are guaranteed to be at distance at most $\alpha \varepsilon$, for some problem-dependent constant $\alpha > 1$. Our algorithms are based on simple primal and dual grid decompositions and are easy to implement. We note that (a) the use of duality, which leads to significant improvements in the overhead cost of the algorithms, appears to be novel for this kind of problems; (b) the correct choice of duality in some of these problems is fairly intricate and requires some care; and (c) the correctness and performance analysis of the algorithms (especially in the more advanced versions) is fairly non-trivial. We analyze our algorithms and prove guaranteed upper bounds on their running time and on the “distortion” parameter $\alpha$. 

---

*A preliminary version of this work appeared in Proc. 25th European Sympos. Algorithms (ESA), 2017, 5:1–5:13.

†Google Inc.; email: aigerd@google.com

‡School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel; email: haimk@tau.ac.il. Work by Haim Kaplan has been supported by Grants 1161/2011 and 1367/2017 from the German-Israeli Science Foundation, by Grants 1841-14 and 1595-19 from the Israel Science Foundation and by the Blavatnik Research Fund in Computer Science at Tel Aviv University.

§School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel; email: michas@tau.ac.il. Work by Micha Sharir has been supported by Grant 2012/229 from the U.S.-Israel Binational Science Foundation, by Grant 892/13 from the Israel Science Foundation, by the Blavatnik Research Fund in Computer Science at Tel Aviv University, and by the Hermann Minkowski–MINERVA Center for Geometry at Tel Aviv University. Both Kaplan and Sharir have also been supported by the Israeli Centers for Research Excellence (I-CORE) program (center no. 4/11).
1 Introduction

Approximate incidences. Given a finite point set $S_1$ and a finite set $S_2$ of geometric primitives (e.g., lines, planes, circles, or spheres in $\mathbb{R}^2$ or $\mathbb{R}^3$), and some $\varepsilon > 0$, we define the set of $\varepsilon$-incidences (also referred to as $\varepsilon$-approximate incidences, or just approximate incidences) between $S_1$ and $S_2$ to be

$$I_\varepsilon(S_1, S_2) = \{(s_1, s_2) \mid s_1 \in S_1, s_2 \in S_2, \text{dist}(s_1, s_2) \leq \varepsilon\},$$

where \text{dist}(s_1, s_2) = \inf\{\text{dist}(s_1, y) \mid y \in s_2\}$ is the Euclidean distance between $s_1$ and $s_2$.

We are interested in efficient algorithms for computing $I_\varepsilon(S_1, S_2)$, ideally in time linear in $|S_1| + |S_2| + |I_\varepsilon(S_1, S_2)|$.

Most of the classical work in discrete and computational geometry on this kind of problems is focused on computing exact incidences ($\varepsilon = 0$). The simplest, and perhaps archetypal instance of this task is Hopcroft’s problem, where we want to determine whether there exists at least one incidence between a set $S_1$ of $m$ points and a set $S_2$ of $n$ lines in the plane. Solutions to this problem and its obvious generalizations run in time close to $m^{2/3}n^{2/3} + m + n$; see [2] [16]. The cases of more general families of curves or surfaces have received less attention.

In principle, this problem is a special case of batched range searching, where the data set is $S_1$ and the ranges are the objects in $S_2$. These problems can be solved using standard range searching techniques, as reviewed, e.g., in [2], but the resulting running times, while subquadratic, are sometimes inferior to the best known combinatorial bounds on the number of incidences (unlike the situation with Hopcroft’s problem and its variants, where the running time is similar to the incidence bound). We note that a major difference between approximate incidences and exact incidences is that the number of exact incidences is always asymptotically smaller than $nm$, where $m = |S_1|$ and $n = |S_2|$, whereas the number of approximate incidences could well be $nm$.

In contrast, the notion of approximate incidences, as we define here, has received less attention in the practical consideration, but it has many important applications which we review below. We consider the problem of reporting all pairs in $I_\varepsilon(S_1, S_2)$. Our algorithms, though, can also estimate $|I_\varepsilon(S_1, S_2)|$, rather than report its members, and do it faster when $|I_\varepsilon(S_1, S_2)|$ is small.

The problem of finding approximate incidences can also be viewed as a range searching problem. Specifically, we treat each member $s_2$ of $S_2$ as the range $s_2(\varepsilon) = \{p \in \mathbb{R}^d \mid \text{dist}(p, s_2) \leq \varepsilon\}$. Here $d$ is the dimension of the ambient space, which in this paper is 2 or 3. By definition, $s_2(\varepsilon)$ is the Minkowski sum of $s_2$ with a disk (ball in $\mathbb{R}^3$) of radius $\varepsilon$ (centered at the origin); thus points become disks (in $\mathbb{R}^2$) or balls (in $\mathbb{R}^3$), lines become slabs (in $\mathbb{R}^2$) or cylinders (in $\mathbb{R}^3$), circles become annuli (in $\mathbb{R}^2$) or tori (in $\mathbb{R}^3$), and so on. The goal now is to report all pairs $(s_1, s_2) \in S_1 \times S_2$ such that $s_1 \in s_2(\varepsilon)$. As mentioned, the known algorithms for such tasks have a rather large overhead. For example, when $S_1$ is a set of $m$ points and $S_2$ is a set of $n$ lines in the plane, i.e., the ranges $s_2(\varepsilon)$ are fixed-width slabs, the best known algorithms for solving the problem have an overhead close to $m^{2/3}n^{2/3}$, and there are matching lower bounds in certain models of computation. The overhead is larger when the objects in $S_2$ are of more complex shapes (e.g., arbitrary circles) or when we move to three (or higher) dimensions; see [2]. In addition, these algorithms, while interesting and sophisticated from a theoretical point of view, are a nightmare to implement in practice.

Instead, with the goal of obtaining algorithms that are really simple to implement (and
therefore with good performance in practice), and that run in time that is (nearly) linear in the input and output sizes, we adopt the approach of using approximation schemes, in which we still report all the pairs \((s_1, s_2)\) that satisfy \(\text{dist}(s_1, s_2) \leq \varepsilon\), but are willing to report additional pairs, provided that all pairs that we report satisfy \(\text{dist}(s_1, s_2) \leq \alpha \varepsilon\), for some constant problem-dependent parameter \(\alpha > 1\). To be more precise, assuming that the test whether \(\text{dist}(s_1, s_2) \leq \varepsilon\) is cheap, we can filter the reported pairs by such a test, and actually report only the pairs that pass it. The actual number of pairs that we have to inspect will typically be larger than \(|I_\varepsilon(S_1, S_2)|\), but it will always be at most \(|I_{\alpha \varepsilon}(S_1, S_2)|\) (and in practice considerably less than that), and the hope is that the number of inspected pairs will not be much larger than those that we actually report. (We expect it to be larger by only a constant factor, which depends on \(\alpha\) and on the geometry of the setup under consideration.)

Our results. We present simple and efficient output-sensitive algorithms (in the above sense) for approximate-incidence reporting problems between points and various simple geometric shapes, in two and three dimensions.

To calibrate the merits of our solutions, we first note that these approximate incidence reporting problems can also be solved by naive grid-based algorithms, as follows. Consider, for example, the problem of reporting approximate incidences between a set \(S_1\) of \(m\) points and a set \(S_2\) of \(n\) lines in the plane. We assume that all the incidences that we seek occur in the unit disk (ball in \(\mathbb{R}^2\)). We partition the unit disk by a uniform grid, each of whose cells is a square of side length \(\varepsilon\). We store each point in \(S_1\) in a bucket corresponding to the grid cell that contains it, and, for each line \(\ell \in S_2\), we report all the pairs involving \(\ell\) and the points in the grid cells that \(\ell\) crosses, and in their neighboring cells. The running time is \(O(m + n/\varepsilon + k)\), where \(k\) is the number of reported approximate incidences. Clearly, all pairs \((p, \ell) \in S_1 \times S_2\) with \(\text{dist}(p, \ell) \leq \varepsilon\) are reported, and each reported pair \((p, \ell)\) satisfies \(\text{dist}(p, \ell) \leq 2\sqrt{2} \varepsilon\), as is easily checked. If \(n\) is much larger than \(m\), we can use duality (where some care is needed to preserve point-line distances), to map the points to lines and the lines to points, and thereby reduce the complexity to \(O(n + m + \min\{m, n\}/\varepsilon + k)\). This method can also be applied in three dimensions, and yields the same time bounds as in the preceding primal-only approach (duality is much trickier in these situations), namely, \(O(m + n/\varepsilon + k)\), when \(S_2\) consists of one-dimensional objects (e.g., lines or circles), but the running time deteriorates to \(O(m + n/\varepsilon^2 + k)\) when \(S_2\) consists of surfaces (e.g., planes or spheres). In these latter cases (involving planes or congruent spheres) duality can be applied, to improve the time bound to \(O(n + m + \min\{m, n\}/\varepsilon^2 + k)\).

While superficially these simple solutions might look ideal, as they are linear in \(m, n,\) and \(k\), their dependence on \(\varepsilon\) is too naive and weak, and when \(m\) and \(n\) are large and \(\varepsilon\) small (as is typically the case in practice), the algorithms are rather slow in practice.

In this paper we address this issue, and develop a series of “primal-dual” grid-based algorithms for several approximate incidence reporting problems, that are faster than this naive scheme for suitable ranges of the parameters \(m, n,\) and \(\varepsilon\) (which cover most of the practical instances of these problems). Specifically, we present the following results. In all of them, \(S_1\) is a set of \(m\) points, contained in the unit ball in two or three dimensions, and \(k\) is the number of points that we inspect; the actual output size might be smaller.

(a) In the plane, for a set \(S_2\) of \(n\) lines, all \(k\) approximate incidences can be reported in
time $O(m + n + \sqrt{mn}/\sqrt{\varepsilon} + k)$. (The dependency of the complexity on $\varepsilon$ is improved by a factor of $\sqrt{\varepsilon}$ compared to the naive scheme when $n$ and $m$ are comparable.). See Section 2.

(b) In three dimensions, for a set $S_2$ of $n$ planes, all $k$ approximate incidences can be reported in time $O(m + n + \sqrt{mn}/\varepsilon + k)$. (The dependency of the complexity on $\varepsilon$ is improved by a factor of $\varepsilon$ compared to the naive scheme, when $n$ and $m$ are comparable.). See Section 3.

(c) In the plane, for a set $S_2$ of $n$ congruent circles, all $k$ approximate incidences can be reported in time $O(m + n + \sqrt{mn}/\sqrt{\varepsilon} + k)$. See Section 4.

(d) In the plane, for a set $S_2$ of $n$ arbitrary circles, all $k$ approximate incidences can be reported in time $O(m + n + m^{1/3}n^{2/3}/\varepsilon^{2/3} + k)$. See Section 5.

(e) In three dimensions, for a set $S_2$ of $n$ congruent spheres, all $k$ approximate incidences can be reported in time $O((m + n)/\varepsilon + k)$. See Section 6.

(f) In three dimensions, for a set $S_2$ of $n$ lines, all $k$ approximate incidences can be reported in time $O(m + n + m^{1/3}n^{2/3}/\varepsilon^{2/3} + k)$. See Section 7.

(g) In three dimensions, for a set $S_2$ of $n$ congruent circles, all $k$ approximate incidences can be reported in time

$$O\left((m + n)/\varepsilon^{1/2} + m^{1/3}n^{2/3}/\varepsilon^{7/6} + k\right).$$

See Section 8.

In Section 9, we use the algorithms in (e) and (g), to obtain an efficient algorithm for finding triangles that are nearly congruent to a given triangle in a three-dimensional point set. This is the first step in solving the approximate point pattern matching problem in $\mathbb{R}^3$. The exact version of this problem (which is to report all triangles spanned by a set of $n$ points in 3-space which are congruent to a given triangle) has been solved by Agarwal and Sharir [1], in time close to $n^{5/3}$.

A comparison with the naive solutions sketched above clearly shows the superiority of our technique. For example, for lines or congruent circles in the plane, assuming that $n \leq m$, our algorithms (in (a) and (c), respectively) are asymptotically faster than the naive method when $\sqrt{mn}/\varepsilon \leq n/\varepsilon$, that is, when $\varepsilon \leq n/m$, an assumption that holds in most practical applications.

To recap, one can obtain substantially better bounds than the naive ones. Our methods are based on grids and on duality—they construct much coarser primal grids, and pass each subproblem, consisting of the points in a grid cell and of the objects that pass through or near that cell, to a secondary dual stage, in which another coarse grid is constructed in a suitably defined dual space. The output pairs are obtained from the cells of these secondary grids, and the gain is in the overhead, as each primal or dual object crosses much fewer grid cells than in the naive solutions. Although this primal-dual paradigm is fairly standard, its power in the approximate incidences context, as considered here, has not been demonstrated before (to the best of our knowledge). The analysis (and the particular duality one has to use) for some of the three-dimensional variants is fairly challenging, but the algorithms all remain simple to describe and to implement. We have actually implemented some of the
algorithms and have experimented with them on several data sets. This implementation is reported in Section 10.

Motivation and applications. Approximate incidence reporting and counting problems arise in several basic practical applications, in computer vision, pattern recognition, and related areas. Three major applications of this sort are robust model fitting, approximate point pattern matching under rigid motions, and estimating the fundamental matrix in (stereo) epipolar geometry. All three problems share a common paradigm, which we first explain for model fitting. In this problem, we are given a set \( P \) of \( n \) points, say in \( \mathbb{R}^3 \) (typically, these are so-called interest points, extracted from some image or 3D sensors), and we want to fit objects (called models) from some given family, such as lines, circles, planes, or spheres, so that each model passes near (i.e., is approximately incident to) many points of \( P \); the quality of the model is measured in terms of the number of approximately incident points. The standard approach is to construct (usually, by repeated random sampling) a sufficiently rich collection of candidate models. (For example, for line models, one can simply sample pairs of points of \( P \), and for each pair construct the line passing through its points.) One then counts, for each candidate line, the number of approximately incident points (for some specified error parameter \( \varepsilon > 0 \)), and reports the models that have sufficiently many such points.

Similar reductions arise in the other problems. In approximate point pattern matching, we are given two sets \( A, B \) of points, and want to find rigid motions that map sufficiently large subsets of \( A \) to sets whose (unidirectional) Hausdorff distance to \( B \) is at most \( \varepsilon \). Here too we construct candidate rigid motions, and test the quality of each of them. For example, in the plane, we sample pairs of points from \( A \), and find, for each sampled pair, the pairs of points of \( B \) that are nearly at the same distance. For each such pair of pairs we construct a rigid motion that maps the first pair to near the other pair, and then test the quality of each of these motions, namely, the number of points of \( A \) that lie, after the motion, near points of \( B \). The first step can be reduced to approximate incidence counting involving circles (whose radii correspond to the distance between the pairs of sampled points of \( A \), and which are centered at the points of \( B \)) and the points of \( B \). In three dimensions, we need to sample triples of points of \( A \), and for each triple \( a, b, c \), we need to find those triples of \( B \) that span triangles that are nearly congruent to \( \Delta abc \) (because to determine a rigid motion in \( \mathbb{R}^3 \) we need to specify how it maps three (noncollinear) source points to three respective image points). This step is described in detail in Section 9.

In epipolar geometry, we have two stereo images \( A, B \) of the same scene, and we want to estimate the fundamental matrix \( F \) that best matches \( A \) to \( B \), where a point \( p \in A \) is (exactly) matched to a point \( q \in B \) if \( p^T F q = 0 \). We construct a sample of candidate matrices, by repeatedly sampling \( O(1) \) interest points from both images, and test the quality of each matrix. To do so for a candidate matrix \( F \), we left-multiply each point \( p \in A \) by \( F \), interpret the resulting vectors \( p^T F \), for \( p \in A \), as lines, and count the approximate incidences of each line with the points of \( B \). If sufficiently many lines have sufficiently high counts, we regard \( F \) as a good fit and output it.

To recap, in each of these applications, and in other applications of a similar nature, we generate a random sample of candidate models, motions, or matrices, and need to test the quality of each candidate. Approximate incidence reporting and counting arises either in the generation step, or in the quality testing step, or in both. Improving the efficiency
of these steps is therefore a crucial ingredient of successful solutions for these problems. The standard approach, used “all over” in computer vision in practice, is the RANSAC technique [8, 10], which checks in brute force each model against each point. Replacing it by efficient methods for approximate incidence counting, which is our focus here, can drastically improve the running time of these applications.

To support the claim that this is indeed the case in practice, we have conducted, as already mentioned, preliminary experiments with some of our algorithms, tested them on real and random data, and compared them with other existing methods. Roughly, they demonstrate that our approach is significantly faster than the other approaches. Our experiments also support our feeling that the cost of reporting more pairs than really needed (pairs that might be at most $\alpha \varepsilon$ apart, rather than just $\varepsilon$), is negligible compared to the cost of the other steps (in themselves much more efficient than the competing techniques). We leave the project of conducting a thorough experimental study for future work. While we will present the implementation that we have performed, the focus of this paper will be on developing the algorithms and establishing their worst-case guarantees.

**Related work.** Model fitting and point pattern matching have been the focus of many studies, both theoretical and practical; see for example [3, 4, 5, 6, 9, 11, 12, 13, 15].

We first note that in many of the common approaches used in practice (e.g., RANSAC for model fitting [8, 10]), reporting or counting approximate incidences between models and points is done using brute force, examining every pair of a model and a point. Some heuristic improvements have also been proposed (see, e.g., [6] and the references therein). A similar brute-force technique is commonly used for approximate point pattern matching too (e.g., in the Alignment method [13] and its many variants).

The use of (exact) geometric incidences in algorithms for exact point pattern matching is well established; see, e.g., Brass [5] for details. Similar connections have also been used for the more practical problem of approximate point pattern matching. Gavrilov et al. [11] gave efficient algorithms for approximate pattern matching in two and three dimensions (where the entire sets $A$ and $B$ are to be matched), that use algorithms for reporting approximate incidences. One of the main results in [11] is that in the plane, all pairs of points at distance in $[(1 - \varepsilon)r, (1 + \varepsilon)r]$ can be reported in $O(n\sqrt{r/\varepsilon})$ time, using a grid-based search. (In a way, part of the study in this paper formalizes, extends, and improves this method.)

Aiger et al. [4] proposed a method for point pattern matching in $\mathbb{R}^3$, called 4PCS (4-Points Congruent Sets), which iterates over all pairs of coplanar quadruples of points, one from $A$ and one from $B$, that can be matched via an affine transformation, and then tests the quality of each pair, focusing on pairs where the transformation is rigid. This algorithm does not use approximate incidences, and assumes the existence of coplanar tuples.

In a more recent work, Aiger and Kedem [3] describe another algorithm for computing approximate incidences of points and circles, following a similar approach by Fonseca and Mount [9] for points and lines, which is better than the one of [11] for $n = \Omega(1/\varepsilon^{3/2})$, and use this for approximate point pattern matching. This algorithm has been used in Mellado et al. [15], to reduce the running time of the 4PCS algorithm in [4] to be asymptotically linear in $n$ and in the output size.

The method of [3, 9] provides an alternative approach to approximate incidence reporting, for the cases of points and lines or congruent circles (the analysis in [3] is rather sketchy,
though). This technique runs in $O(m + n + \log(1/\varepsilon)/\varepsilon^2 + k)$ time. For the case of lines in the plane, the scheme exploits the fact that we can approximate (up to an error of $O(\varepsilon)$) all lines in the plane that cross the unit disk, by $O(1/\varepsilon^2)$ representative lines, such that if a point in the unit disk is close to a representative line $\ell$, then it is also close (up to some small negligible additive error) to all the lines in the input that $\ell$ represents (and vice versa). Assuming, for example, that $m$ is constant, this alternative scheme is better than our new algorithm (for these restricted scenarios) when $\sqrt{n}/\sqrt{\varepsilon} \geq 1/\varepsilon^2$, that is, when $n \geq 1/\varepsilon^3$ (we ignore the factor $\log(1/\varepsilon)$ in this calculation). (This technique seems to be extendible to three dimensions, and to surfaces, but the formal details have not yet been worked out, as far as we know.)

2 Approximate incidences in planar point-line configurations

We consider the approximate incidences problem between a set $P$ of $m$ points in the unit disk $B$ in $\mathbb{R}^2$, and a set $L$ of $n$ lines that cross $B$, with a given accuracy parameter $0 < \varepsilon \leq 1/2$.

We approximate the distance $\text{dist}(p, \ell)$ by the vertical distance between $p \in P$ and $\ell \in L$, which we denote by $\text{dist}_v(p, \ell)$. For this approximation to be good, the angle between $\ell$ and the $x$-direction should not be too large. To ensure this, we partition $L$ into two subfamilies, one consisting of the lines with positive slopes, and one of the lines with negative slopes. We fix one subfamily, rotate the plane by 45°, and get the desired property. In what follows we assume that all the lines of $L$ are "nearly horizontal", in this sense.

![Figure 1: The partition of $S$ into subsquares, and the subproblem associated with the middle highlighted subsquare.](https://example.com/figure1.png)

Without loss of generality, we replace the unit disk $B$ by the unit square $S = [0,1]^2$ (scaling down the plane by a factor of 2), and apply the following two-stage partitioning procedure. First we partition $S$ into $1/\delta_1^2$ pairwise openly disjoint smaller squares, each of side length $\delta_1$, where $\delta_1$ is a parameter whose exact value will be set later. See Figure 1. We ignore in what follows rounding issues and assume, for example, that $1/\delta_1^2$ is an integer.

Enumerate these squares as $S_1, S_2, \ldots, S_1/\delta_1^2$. For $i = 1, \ldots, 1/\delta_1^2$, let $P_i$ denote the set of all points of $P$ that lie either in $S_i$ or in one of the two squares that are directly above and below $S_i$ (if they exist), and let $L_i$ be the set of all the lines of $L$ that cross $S_i$. Put $m_i := |P_i|$ and $n_i := |L_i|$. We have $\sum_i m_i \leq 3m$ and $\sum_i n_i \leq 2n/\delta_1$, because each line of $L$ crosses at most $2/\delta_1$ squares $S_i$. 7
We now apply a duality transformation to each small square $S_i$ separately. For notational simplicity, and without loss of generality, we may assume that $S_i = [-\delta_1/2, \delta_1/2]^2$. (Technically, this means that we shift the cells by $\delta_1/2$ in both coordinate directions, so that the grid vertices now represent the centers of the cells.) We map each point $p = (\xi, \eta)$ in $P_i$ to the line $p^* : y = \xi x - \eta$, and each line $\ell : y = cx + d$ in $L_i$ to the point $\ell^* = (c, -d)$. This duality preserves the vertical distance $\text{dist}_v$ between a point $p$ and a line $\ell$; that is, $\text{dist}_v(p, \ell) = \text{dist}_v(\ell^*, p^*)$. Note that the slope condition ensures that $\text{dist}(p, \ell) \leq \text{dist}_v(p, \ell) \leq \sqrt{2}\text{dist}(p, \ell)$. See Figure 2.

![Figure 2: The relation between dist(p, l) and dist_v(p, l).](image)

Let $\ell : y = cx + d$ be a line in $L_i$, that is, $\ell$ crosses $S_i$. By the slope condition we have $-1 \leq c \leq 1$ and $-\delta_1 \leq d \leq \delta_1$, so the dual point $\ell^*$ lies in the rectangle $R := [-1, 1] \times [-\delta_1, \delta_1]$. Each point $p = (\xi, \eta) \in P_i$ satisfies $-\delta_1/2 \leq \xi \leq \delta_1/2$ and $-3\delta_1/2 \leq \eta \leq 3\delta_1/2$ so the coefficients of the dual line $p^* : y = \xi x - \eta$ satisfy these respective inequalities.

We now partition $R$ into $1/\delta_2^2$ small rectangles, each of width $2\delta_2$ and height $2\delta_1\delta_2$, where $\delta_2$ is another parameter that we will shortly specify. Each dual line $p^*$ crosses at most $2/\delta_2$ small rectangles. To facilitate the following analysis, we choose $\delta_1, \delta_2$ so that they satisfy $\delta_1\delta_2 = \varepsilon$; we still have one degree of freedom in choosing them, which we will exploit later.

**Lemma 2.1** For each small rectangle $R'$, if $\ell^*$ is a dual point in $R'$ and $p^*$ is a dual line that crosses either $R'$ or one of the small rectangles directly above or below $R'$ (in the $y$-direction, if they exist), then the vertical distance $\text{dist}_v(\ell^*, p^*)$ (which is the same as $\text{dist}_v(p, \ell)$) is at most $5\delta_1\delta_2 = 5\varepsilon$.

**Proof.** Indeed, if $p^*$ crosses a small rectangle $R''$, which is either $R'$ or one of the two adjacent rectangles, as above, then, since the slope of $p^*$ is in $[-\delta_1/2, \delta_1/2]$, its maximum vertical deviation from $R''$ is at most $2\delta_2 \cdot (\delta_1/2) = \delta_1\delta_2$. Adding the heights $2\delta_1\delta_2$ of $R''$, and of $R'$ when $R'' \neq R'$, the claim follows. $\square$

**Lemma 2.2** (a) Let $(p, \ell) \in P \times L$ be such that $\text{dist}(p, \ell) \leq \varepsilon$. Let $S_i$ be the small square containing $p$. If $\delta_1 \geq \varepsilon\sqrt{2}$, then $\ell$ must cross either $S_i$ or one of the two squares directly above and below $S_i$. In other words, there exists a $j$ such that $(p, \ell) \in P_j \times L_j$.

(b) Continue to assume that $\text{dist}(p, \ell) \leq \varepsilon$, let $i$ be such that $(p, \ell) \in P_i \times L_i$, and let $R'$ be the dual small rectangle (that arises in the dual processing of $S_i$) that contains $\ell^*$. Then the dual line $p^*$ must cross either $R'$ or one of the two small rectangles lying directly above and below $R'$ (in the $y$-direction, if they exist).

**Proof.** Both claims are obvious; in (a) we use the fact that $\text{dist}_v(p, \ell) \leq \varepsilon\sqrt{2}$, and the assumption that $\varepsilon\sqrt{2} \leq \delta_1$; see below how this is enforced. In (b) we use the fact that $\text{dist}_v(p, \ell) = \text{dist}_v(\ell^*, p^*)$ and that the height of a small rectangle is $2\delta_1\delta_2 = 2\varepsilon > \varepsilon\sqrt{2}$. $\square$
The algorithm. We first compute, for each point $p \in P$, the square $S_i$ it belongs to; this can be done in $O(1)$ time, assuming a model of computation in which we can compute the floor function in constant time. Similarly, we find, for each line $\ell \in L$, the squares that it crosses, in $O(1/\delta_1)$ time. This gives us all the sets $P_i, L_i$, in overall $O(m + n/\delta_1)$ time.

We then iterate over the small squares in the partition of $S$. For each such square $S_i$, we construct the dual partitioning of the resulting dual rectangle $R$ into the smaller rectangles $R'$. As above, we find, for each dual point $\ell^*$, for $\ell \in L_i$, the small rectangle that contains it, and, for each dual line $p^*$, for $p \in P_i$, the small rectangles that it crosses. This takes $O(n_i + m_i/\delta_2)$ time.

We now report, for each small rectangle $R'$, all the pairs $(p, \ell) \in P_i \times L_i$ for which $\ell^*$ lies in $R'$ and $p^*$ crosses either $R'$ or one of the small rectangles lying directly above or below $R'$ (if they exist). We repeat this over all small squares $S_i$ and all respective small rectangles $R'$. Note that a pair $(p, \ell)$ may be reported more than once in this procedure, but its multiplicity is at most some small absolute constant. The running time of this algorithm is

$$O \left( m + n/\delta_1 + \sum_{i=1}^{1/\delta_1^2} \left( n_i + m_i/\delta_2 \right) + k \right) = O \left( \frac{n}{\delta_1} + \frac{m}{\delta_2} + k \right),$$

where $k$ is the number of pairs that we report. Lemma 2.1 guarantees that each reported pair is at distance at most $5\varepsilon$ and Lemma 2.2 guarantees that every pair $(p, \ell)$ at distance at most $\varepsilon$ is reported.

We optimize the running time by choosing $\delta_1, \delta_2$ to satisfy $m/\delta_2 = n/\delta_1$ and $\delta_1 \delta_2 = \varepsilon$. That is, we want to choose $\delta_1 = \sqrt{m/n}$ and $\delta_2 = \sqrt{n\varepsilon/m}$. These choices are effective, provided that both $\delta_1, \delta_2$ are at most 1, for otherwise the primal partition or the dual partitions does not exist. If $\delta_2 > 1$, that is, if $n < m \varepsilon$, we simply choose $\delta_1 = \varepsilon$, and run only the primal part of the algorithm, outputting all the pairs in $\bigcup_i P_i \times L_i$. The cost is now $O(m + n/\varepsilon + k) = O(m + k)$. (This is the naive implementation, which is now efficient since $n$ is so small.) If $\delta_1 > 1$, we pass directly to the dual plane, flip the roles of $P$ and $L$, and solve the problem in the naive manner just described, at the cost of $O(n + k)$. Otherwise (when both $\delta_1$ and $\delta_2$ are $\leq 1$), the cost is $O(\sqrt{mn}/\sqrt{\varepsilon} + k)$. The cost of the algorithm is therefore always bounded by $O(n + m + \sqrt{mn}/\sqrt{\varepsilon} + k).

Recall also that in the proof of Lemma 2.2 we needed the inequality $\varepsilon \sqrt{2} \leq \delta_1$. This will hold when $m \leq n$ (and $\varepsilon \leq 1/2$, as we assume). In the complementary case $m > n$, we simply flip the roles of points and lines (that is, we start the analysis in the dual plane).

In conclusion, we have obtained the following main result of this section.

Theorem 2.3 Let $P$ be a set of $m$ points in the unit disk $B$ in the plane, let $L$ be a set of $n$ lines that cross $B$, and let $0 < \varepsilon \leq 1/2$ be a prescribed parameter. We can report all pairs $(p, \ell) \in P \times L$, for which $\text{dist}(p, \ell) \leq \varepsilon,$ in time $O(n + m + \sqrt{mn}/\sqrt{\varepsilon} + k)$, where $k$ is the actual number of pairs that we report; all pairs at distance at most $\varepsilon$ are reported, and every reported pair lies at distance at most $5\varepsilon$.

Another useful feature of the algorithm is that, rather than reporting all the pairs that it produces, it can output a compact representation of them, as a union of complete bipartite graphs $P_\alpha \times L_\alpha$. The number of such graphs is $O \left( \frac{1}{\delta_1^2}, \frac{1}{\delta_2^2} \right) = O(1/\varepsilon^2)$, and the sum of the
cardinalities of their vertex sets is $O(m + n + \sqrt{mn}/\sqrt{\varepsilon})$. A similar feature holds for the algorithms in the forthcoming sections.

3 Near neighbors in point-plane configurations

As a second application of the methodology illustrated in the preceding section, we apply a similar approach in three dimensions. That is, given a set $P$ of $m$ points in the unit ball $B$ in $\mathbb{R}^3$, a set $\Pi$ of $n$ planes crossing $B$, and a prescribed error parameter $0 < \varepsilon \leq 1/2$, we solve the approximate incidences problem for $P$ and $\Pi$ with accuracy $\varepsilon$.

We approximate the distance $\text{dist}(p, \pi)$ by the $z$-vertical distance $\text{dist}_v(p, \pi)$ between $p$ and $\pi$. For this approximation to be good, we partition $\Pi$ into $O(1)$ subfamilies, such that, for each subfamily $\Pi'$ there exists a direction $u'$ such that the angle between $u'$ and the normal of each plane of $\Pi'$ is at most $\pi/4$. We apply the construction to each subset $\Pi'$ separately (with respect to all the points in $P$). When we apply it to a subfamily $\Pi'$, we rotate the space such that $u'$ becomes the $z$-direction. In what follows, we fix one subfamily, continue to denote it as $\Pi$, and assume that $u'$ is indeed the $z$-direction.

As in the two-dimensional case, we assume that all the points of $P$ are contained in the unit cube $S = [0, 1]^3$.

We apply a two-stage partitioning procedure analogous to the one of Section 2. First we partition $S$ into $\frac{1}{\sqrt{3}}$ pairwise openly disjoint smaller cubes, each of side length $\delta_1$, where $\delta_1$ is a parameter whose exact value will be set later.

Consider one such small cube $S_i$, and assume that $S_i = [-\delta_1/2, \delta_1/2]^3$ (translate space by $-\delta_1/2$ in each axis). Let $P_i$ denote the set of all points of $P$ that lie either in $S_i$ or in one of the two cubes that lie directly above and below $S_i$ in the $z$-direction, (if they exist), and let $\Pi_i$ be the set of all the planes of $\Pi$ that cross $S_i$. Put $n_i := |\Pi_i|$ and $m_i := |P_i|$. We have $\sum_i n_i \leq 3m$ and $\sum_i n_i = O(n/\delta_1^2)$, because each plane of $\Pi$ crosses $O(1/\delta_1^2)$ cubes $S_i$.

For each such $S_i$, we pass to the dual space, mapping each point $p = (\xi, \eta, \zeta)$ in $P_i$ to the plane $p^* : z = \xi x + \eta y - \zeta$, and map each plane $\pi : z = ax + by + c$ in $\Pi_i$ to the point $\pi^* = (a, b, -c)$. This duality preserves the vertical distance $\text{dist}_v$ between a point $p$ and a plane $\pi$; that is, $\text{dist}_v(p, \ell) = \text{dist}_v(\ell^*, p^*)$. As in the planar case, the normal direction condition is easily seen to ensure that $\text{dist}(p, \ell) \leq \text{dist}_v(p, \ell) \leq \sqrt{2}\text{dist}(p, \ell)$.

The normal direction condition also implies that, for each plane $\pi : z = ax + by + c$ in the current subproblem,

$$\frac{(a, b, 1) \cdot (0, 0, 1)}{||(a, b, 1)||} = \frac{1}{\sqrt{a^2 + b^2 + 1}} \geq \cos(\pi/4) = \frac{1}{\sqrt{2}},$$

so $a^2 + b^2 \leq 1$, and therefore $|a| \leq 1$ and $|b| \leq 1$.

Let $\pi : z = ax + by + c$ be a plane in $\Pi_i$. We then have $-1 \leq a \leq 1$, $-1 \leq b \leq 1$, and $-3\delta_1/2 \leq c \leq 3\delta_1/2$, so the dual point $\pi^*$ lies in the box $R := [-1, 1]^2 \times [-3\delta_1/2, 3\delta_1/2]$. Each point $p = (\xi, \eta, \zeta) \in P_i$ satisfies $-\delta_1/2 \leq \xi \leq \delta_1/2$, $-\delta_1/2 \leq \eta \leq \delta_1/2$, and $-3\delta_1/2 \leq \zeta \leq 3\delta_1/2$.

1There exists a point $(x_1, y_1, z_1) \in \pi \cap S_i$, and then we have $\delta_1/2 \leq x_1 \leq \delta_1/2$, $-\delta_1/2 \leq y_1 \leq \delta_1/2$, $-\delta_1/2 \leq z_1 \leq \delta_1/2$, and $z_1 = ax_1 + by_1 + c$. Thus $c = z_1 - ax_1 - by_1$, which, with $|a| \leq 1$ and $|b| \leq 1$, implies that $-3\delta_1/2 \leq c \leq 3\delta_1/2$. 

10
\[ \zeta \leq 3\delta_1 / 2, \] so the coefficients of the dual plane \( p^* : z = \xi x + \eta y - \zeta \) satisfy these respective inequalities.

We now partition \( R \) into \( 1/\delta^3 \) small boxes, each of \( x \)-range and \( y \)-range \( 2\delta_2 \), and of \( z \)-range \( 3\delta_1\delta_2 \). Each dual plane \( p^* \) crosses at most \( O(1/\delta^2) \) small boxes. We choose \( \delta_1, \delta_2 \) so that they satisfy \( \delta_1 \geq \varepsilon\sqrt{2} \) and \( \delta_1\delta_2 = \varepsilon \), and prove lemmas analogous to Lemma 2.1 and Lemma 2.2. We omit both the statements and the proofs, which are almost verbatim to those in Section 2. In the analog of Lemma 2.2, the constant 5 has to be replaced by 7, as is easily checked.

**The algorithm.** We map each point \( p \in P \) to the cube \( S_i \) containing it and each plane \( \pi \in \Pi \) to the cubes that it crosses, thereby obtaining all the sets \( P_i, \Pi_i \). This takes \( O(m + n/\delta^2) \) time. We then iterate over the cubes in the partition of \( S \). For each such cube \( S_i \), we construct the dual partitioning of the resulting dual box \( R \) into the smaller boxes \( R' \). As above, we find, for each dual point \( \ell^* \), the small box that contains it, and, for each dual plane \( p^* \), the small boxes that it crosses. This takes \( O(n_i + m_i/\delta^2) \) time.

We now report, for each small box \( R' \), all the pairs \( (p, \ell) \) for which \( \ell^* \) lies in \( R' \) and \( p^* \) crosses either \( R' \) or one of the small boxes lying directly above or below \( R' \) (in the third coordinate, if they exist). The overall running time is

\[
O\left( \frac{n}{\delta^2} + \frac{m}{\delta^2} + k \right),
\]

where \( k \) is the number of pairs that we report.

We optimize the running time by choosing \( \delta_1, \delta_2 \) to satisfy

\[
\frac{m}{\delta^2} = \frac{n}{\delta^2}, \quad \text{and} \quad \delta_1\delta_2 = \varepsilon.
\]

That is, we choose

\[
\delta_1 = \left( \frac{n\varepsilon^2}{m} \right)^{1/4}, \quad \text{and} \quad \delta_2 = \left( \frac{m\varepsilon^2}{n} \right)^{1/4}.
\]

As before, these choices make sense only when both \( \delta_1 \) and \( \delta_2 \) are at most 1. When one of them is larger than 1, we proceed as in the two-dimensional case, performing either only the primal stage or only the dual one, and obtain the cost \( O(m + n + k) \). Thus, the total cost of the algorithm is \( O(n + m + \sqrt{mn}/\varepsilon + k) \). The requirement that \( \delta_1 \geq \varepsilon\sqrt{2} \) can be enforced as in the planar case.

In conclusion, we have obtained the following main result of this section.

**Theorem 3.1** Let \( P \) be a set of \( m \) points in the unit ball \( B \) of \( \mathbb{R}^3 \), let \( \Pi \) be a set of \( n \) planes that cross \( B \), and let \( 0 < \varepsilon \leq 1/2 \) be a prescribed parameter. We can report all pairs \( (p, \pi) \in P \times \Pi \) for which \( \text{dist}(p, \pi) \leq \varepsilon \), in time \( O(n + m + \sqrt{mn}/\varepsilon + k) \), where \( k \) is the actual number of pairs that we report; all pairs at distance at most \( \varepsilon \) will be reported, and every reported pair lies at distance at most \( 7\varepsilon \).

4 **Nearly congruent pairs in the plane**

In this section we consider the following problem. We are given two point sets \( P, Q \) in the plane, of respective sizes \( m \) and \( n \) (which would be the same set in some applications),
and we wish to report all pairs \((p, q) \in P \times Q\) such that \(|pq| \in [r - \varepsilon, r + \varepsilon]|. Here too we consider the approximation version, where we want all such pairs to be reported, and want every reported pair to satisfy \(|pq| \in [r - \alpha \varepsilon, r + \alpha \varepsilon]|, for a suitable absolute constant \(\alpha\).

This problem is equivalent to an approximate incidences problem between \(P\) and the set of congruent circles \(C := \{c_q \mid q \in Q\}\) where \(c_q\) the circle of radius \(r\) centered at a point \(q\).

We assume that \(0 < \nu \leq r \leq 1/2\) for some fixed positive constant \(\nu\).

In the following subsections we present two different solutions to the problem. The first solution, inspired by a similar idea due to Indyk, Motwani, and Venkatasubramanian [14], does not use duality. It is simple and elegant, but its major drawback is that it is not sensitive to cases where \(m\) and \(n\) differ significantly. The second solution does use duality, and is sensitive to such differences; it is closer to the preceding solutions for the point-line and point-plane approximate incidences problems.

### 4.1 Reporting all nearly congruent pairs in the plane I

We take the circle \(c_o\) of radius \(r\) centered at the origin \(o\), and partition it into \(2\pi/\sqrt{\varepsilon}\) equal canonical arcs, each with a central angle \(\sqrt{\varepsilon}\), delimited at the points on \(c_o\) at orientations \(0, \sqrt{\varepsilon}, 2\sqrt{\varepsilon}, \ldots\) (again, we ignore in what follows the routine rounding issues). Consider one such arc \(\tilde{ab}\); see Figure 3. Let \(A_o\) denote the annulus centered at the origin with inner radius \(r - \varepsilon\) and outer radius \(r + \varepsilon\). Let \(A_{ab}\) be the portion of \(A_o\) within the wedge \(W_{ab}\) that defines the central angle of \(\tilde{ab}\); that is, \(W_{ab}\) is the wedge with \(o\) as an apex, bounded by the rays \(\tilde{oa}\) and \(\tilde{ob}\). Denote by \(a_1b_1\) and \(a_2b_2\) the respective inner and outer arcs that bound \(A_{ab}\). Let \(R_{ab}\) be the smallest enclosing rectangle of \(A_{ab}\) whose longer side is parallel to \(ab\) (and to \(a_1b_1, a_2b_2\)); see Figure 3.

The short edge, \(ef\), of \(R_{ab}\) is of length

\[
2\varepsilon + (r - \varepsilon) - (r - \varepsilon) \cos(\sqrt{\varepsilon}/2) \leq r + \varepsilon - (r - \varepsilon)(1 - \varepsilon/8) = \varepsilon + r\varepsilon/8 + \varepsilon - \varepsilon^2/8 \leq 3\varepsilon.
\]

The length of the large edge, \(de\), of \(R_{ab}\) is

\[
2(r + \varepsilon) \sin(\sqrt{\varepsilon}/2) \leq (r + \varepsilon)\sqrt{\varepsilon} \leq \sqrt{\varepsilon}.
\]

In these derivations we use the inequalities \(\cos x > 1 - \frac{1}{2}x^2\) and \(\sin x < x\), for \(x > 0\), and, in the very last inequality, also the fact that \(r \leq 1/2\). Note that these upper bounds on the side lengths of \(R_{ab}\) are tight up to a constant factor.

**Lemma 4.1** (a) Let \(q\) be a point at distance \(\leq \varepsilon\) from \(c_o\), so that the point of \(c_o\) nearest to \(q\) lies on \(\tilde{ab}\). Then \(q \in R_{ab}\).

(b) Let \(R_{ab}^*\) denote the homothetic copy of \(R_{ab}\) scaled by a factor of 3 about its center. Then every point in \(R_{ab}^*\) is at distance \(\leq 5\varepsilon\) from \(c_o\).

**Proof.** (a) is trivial to prove because \(q\) must lie in \(A_{ab}\). For (b), we estimate the smallest and largest distances from \(o\) to points of \(R_{ab}^*\). The smallest distance is attained at the midpoint \(\mu^*\) of the longer edge of \(R_{ab}^*\) that is closer to \(o\).
Figure 3: The rectangle $R_{\widehat{ab}}$ that bounds the sector with central angle $\sqrt{\varepsilon}$ of the annulus around the arc $\widehat{ab}$.

The distance of the midpoint $\mu$ of $gf$ (the longer edge of $R_{\widehat{ab}}$ that is closer to $o$) from $o$ is equal to

$$(r - \varepsilon)\cos(\sqrt{\varepsilon}/2) > (r - \varepsilon)(1 - \varepsilon/8) > r - (1 + r/8)\varepsilon > r - 2\varepsilon.$$ 

Since the width of $R_{\widehat{ab}}$ is at most $3\varepsilon$, the image $\mu^*$ of $\mu$ under this homothetic transformation is closer to $o$ by at most $3\varepsilon$, so the distance of $\mu^*$ from $o$ is at least $r - 5\varepsilon$.

The largest distance from $o$ to a point of $R_{\widehat{ab}}$ is attained at the images $d^*$ and $e^*$ of the respective vertices $d$ and $e$ of the longer edge of $R_{\widehat{ab}}$ that is farther from $o$. To estimate the distance from $o$ to $e^*$, say, we argue as follows. The image of the midpoint of $de$ is at distance at most $(r + \varepsilon) + 3\varepsilon = r + 4\varepsilon$ from $o$, and half the length of the image of $de$ is at most $3\sqrt{\varepsilon}/2$. Hence the distance from $o$ to $e^*$ is at most

$$(r + 4\varepsilon)^2 + 9\varepsilon/4)^{1/2} = (r^2 + (8r + 9/4)\varepsilon + 16\varepsilon^2)^{1/2} \leq r + \alpha\varepsilon,$$

for any constant $\alpha$ satisfying $\alpha \geq 4$ and $\alpha \geq (8r + 9/4)/2$, as is easily checked. Since we assume that $r \leq 1/2$, we can take $\alpha = 4$. This establishes (b). \hfill \square

Let $R_{\widehat{ab}}(q)$ be the rectangle $R_{\widehat{ab}}$ translated by the point (vector) $q$. For each canonical arc $\widehat{ab}$ of $c_o$, we consider all the rectangles $\{R_{\widehat{ab}}(q) \mid q \in Q\}$, and aim to find all pairs $(p, c_q)$, for $p \in P$, $q \in Q$, such that $p$ is contained in $R_{\widehat{ab}}(q)$. This is done as follows.

We rotate the plane such that each rectangle $R_{\widehat{ab}}(q)$ becomes axis-parallel with its long edge parallel to the $x$-axis (as depicted in Figure 3). Clearly, in the rotated coordinate system, we can enclose all rectangles and points in a disk centered at $o$ of radius slightly larger than 1. Proceeding as in the previous sections, we may assume that all our axis-parallel rectangles are contained in the unit square $S = [0, 1]^2$.

We partition $S$ into a grid $G$ of isothetic copies of $R_{\widehat{ab}}$, that is, rectangles of size roughly $\sqrt{\varepsilon} \times 3\varepsilon$. There are $O((1/\sqrt{\varepsilon}) \cdot (1/\varepsilon)) = O(1/\varepsilon^{3/2})$ such rectangles in $G$, and each rectangle $R_{\widehat{ab}}(q)$ intersects (the interiors of) at most four rectangles of $G$. For each $q \in Q$, we report all the points of $P$ that lie in any of the four corresponding rectangles of $G$. 

13
The following lemma, combined with Lemma 4.1, establishes the correctness of our scheme.

**Lemma 4.2** We report all pairs \((p, q) \in P \times Q\) such that \(p \in R_{ab}(q)\). Every pair \((p, q)\) that we report is such that \(p\) is at distance at most \(5\varepsilon\) from \(c_q\).

**Proof.** The first part is obvious. The second part follows from the observation that any grid cell that meets \(R_{ab}(q)\) is fully contained in \(R^*_{ab}(q)\), which, combined with Lemma 4.1(b), establishes the claim. \(\square\)

It takes \(O(m)\) time to assign each point of \(P\) to the cell of \(G\) that contains it. It then takes \(O(n + k_{ab})\) time to find and report all the \(k_{ab}\) pairs \((p, q)\) such that \(p\) lies in one of the four grid cells that \(R_{ab}(q)\) overlaps. Thus the total running time per arc \(\hat{ab}\) is \(O(m + n + k_{ab})\). Adding up these bounds, over all \(O(1/\sqrt{\varepsilon})\) arcs \(\hat{ab}\) we get that the total running time is \(O((m + n)/\sqrt{\varepsilon} + k)\), where \(k\) is the total number of reported pairs. Clearly, every pair \((p, q)\), where \(p\) is at distance \(\leq \varepsilon\) from \(c_q\), is reported, and we report each pair \((p, q)\) only a constant number of times.

**Theorem 4.3** Let \(P\) and \(Q\) be two sets of \(m\) and \(n\) points, respectively, in the unit disk \(B\), and let \(0 < \nu \leq r \leq 1/2\) for some fixed constant \(\nu\). We can report all pairs \((p, q) \in P \times Q\) for which \(\text{dist}(p, q) \in [r - \varepsilon, r + \varepsilon]\), in time

\[O\left(\frac{m + n}{\sqrt{\varepsilon}} + k\right),\]

where \(k\) is the actual number of pairs that we report; all pairs at distance in \([r - \varepsilon, r + \varepsilon]\) will be reported, and every reported pair lies at distance in \([r - 5\varepsilon, r + 5\varepsilon]\).

### 4.2 Reporting all nearly congruent pairs in the plane II

We next present an alternative approach to the problem considered in the preceding subsection. Let \(P, Q, m, n, r,\) and \(\varepsilon\) be as above. Again, we may assume that \(P\) and \(Q\) are bounded in the unit square \(S = [0, 1]^2\).

We apply a two-stage partitioning procedure, similar to the one given for the cases of lines and planes. We fix two real positive parameters \(\delta_1, \delta_2\), whose values will be set later. First we partition \(S\) into \(1/\delta_1^2\) pairwise openly disjoint smaller squares, each of side length \(\delta_1\). Enumerate these squares as \(S_1, \ldots, S_1/\delta_1^2\). Let \(\hat{S}_i\) denote the union of \(S_i\) and the (at most) eight squares adjacent to \(S_i\). Let \(P_i\) denote the set of all points of \(P\) that lie in \(\hat{S}_i\), and let \(C_i\) denote the set of all the circles \(c_q \in C\) that cross \(S_i\). Put \(m_i := |P_i|\) and \(n_i := |C_i|\), for \(i = 1, \ldots, 1/\delta_1^2\). We have \(\sum_i m_i = O(m)\), and \(\sum_i n_i = O(n/\delta_1)\).

Fix a small square \(S_i\). To find all the \(\varepsilon\)-near pairs among points in \(P_i\) and circles in \(C_i\), we pass to the dual plane, where (i) we map each point \(p \in P_i\) to the circle \(c_p\) of radius \(r\) centered at \(p\), and (ii) we map each circle \(c_q \in C_i\) to its center \(q\) (so now the elements of \(Q\) become points and those of \(P\) become circles). The distance between \(q\) and \(c_p\) is the same as the distance between \(p\) and \(c_q\).

Let \(c_q\) be a circle in \(C_i\). Clearly, \(q\) has to lie in the Minkowski sum \(K_i\) of \(S_i\) and the circle of radius \(r\) centered at the origin. As is easily checked, \(K_i\) is contained in the annulus...
that is centered at the center \( o_i \) of \( S_i \) and has radii \( r \pm \delta_1/\sqrt{2} \) (note that \( \delta_1/\sqrt{2} \) is half the diameter of \( S_i \)). (We assume that \( \delta_1 < r \).) To simplify the notation, denote this annulus also as \( K_i \); we will use this annulus instead of the Minkowski sum in what follows.

Passing to polar coordinates \((\rho, \theta)\) about \( o_i \), we get that \( K_i \) becomes the rectangle

\[
R = [r - \delta_1/\sqrt{2}, r + \delta_1/\sqrt{2}] \times [0, 2\pi].
\]

We partition \( R \) into \( 1/\delta_2^2 \) small (polar) rectangles, each of width \((\rho\)-range\() \sqrt{2}\delta_1\delta_2\) and height \((\theta\)-range\) \(2\pi\delta_2\); in the standard coordinate frame, each small rectangle is a sector of some (narrower) annulus centered at \( o_i \), with the above width and angle. Each dual circle \( c_{p} \) crosses at most \( O(1/\delta_2) \) small rectangles (that is, annulus sectors) of this grid. This easily follows from the fact that the circle \( c_{p} \) is the graph of a well-defined function \( r = f_p(\theta) \), of constant complexity, in our polar coordinate frame.

To facilitate the following analysis, we will choose \( \delta_1 \ll r \) (recall that \( v \leq r \) for some fixed constant \( v \)), \( \delta_2 \leq \sqrt{\varepsilon/\rho} \), and \( \delta_1\delta_2 = \sqrt{2}\varepsilon \); see below for the way in which we ensure that these constraints hold. The latter choice makes the \( \rho\)-range of each small polar rectangle in the decomposition of \( R \) equal to \( \sqrt{2}\delta_1\delta_2 = 2\varepsilon \).

**Lemma 4.4** Let \( R' \) be a small polar rectangle in the decomposition in the dual problem of \( S_i \). Let \( q \) be a dual point in \( R' \), and let \( c_{p} \) be a dual circle that crosses \( R' \) or one of its two adjacent rectangles with the same \( \theta\)-range. Then \( r - \alpha\varepsilon \leq |pq| \leq r + \alpha\varepsilon \) for some suitable absolute constant \( \alpha \).

**Proof.** Let \( u \) be a point in the intersection of \( c_{p} \) with \( R' \) or with one of its adjacent rectangles with the same \( \theta\)-range, and let \( o = o_i \) be the center of \( S_i \); see Figure 4. We know that (i) \(|pu| = r\), (ii) \(|op| \leq 3\delta_1/\sqrt{2}\), (iii) \(|ou| - r| \leq \delta_1/\sqrt{2}\), and (iv) \(|oq| - |ou| | \leq 2\sqrt{2}\delta_1\delta_2 = 4\varepsilon \). We want to show that \( r - \alpha\varepsilon \leq |pq| \leq r + \alpha\varepsilon \), for some absolute constant \( \alpha \).

Let \( v \) be the point on \( oq \) satisfying \(|ov| = |ou| \). By (iv), we have \(|qv| \leq 4\varepsilon \). It therefore suffices to show that \(|pu| - r| \leq \varepsilon \varepsilon \), for a suitable constant \( c \).

In the isosceles triangle \( \Delta_{ouv} \), the angle at \( o \) is at most \( 2\pi\delta_2 \), so its base \( uv \) is of length

\[
|uv| \leq 2|ou| \sin \pi\delta_2 \leq |ou| \cdot 2\pi\delta_2 \leq \left( r + \delta_1/\sqrt{2}\right) \cdot 2\pi\delta_2 \leq 10r\delta_2,
\]

which can be assumed in view of (iii) and the assumption that \( \delta_1 \ll r \). Moreover, since \( \Delta_{ouv} \) is isosceles we have:

\[
\xi = \angle ouv = \frac{\pi}{2} - \frac{\angle ouv}{2} 
= \frac{\pi}{2} - \pi\delta_2.
\]

Consider next the triangle \( \Delta_{oup} \) and its angle \( \beta = \angle pou \). By the Law of Sines, we have

\[
\frac{|op|}{\sin \beta} = \frac{|pu|}{\sin \angle pou} \geq |pu| = r.
\]

Hence, by (ii)

\[
\sin \beta \leq \frac{|op|}{r} \leq \frac{3\delta_1}{r\sqrt{2}},
\]

so we may conclude that \( \beta \leq \frac{3\delta_1}{r} \), again under the assumption that \( \delta_1 \ll r \).
Consider now the triangle $\Delta_{puv}$, and let $\gamma$ denote its angle $\angle_{puv}$. Regardless of how the two triangles $\Delta_{puo}$ and $\Delta_{ouv}$ are juxtapositioned, we have

$$\xi - \beta \leq \gamma \leq \xi + \beta.$$ 

Subtracting this inequality from $\pi/2$ we get

$$\frac{\pi}{2} - (\xi - \beta) \geq \frac{\pi}{2} - \gamma \geq \frac{\pi}{2} - (\xi + \beta).$$

Since $\xi \leq \pi/2$ the right hand side is at least $-\beta$, and by Equation (1) the left hand side is at most $\beta + \pi\delta_2/2$.

Combining this with our conclusion above that $\beta \leq \frac{3\delta_1}{r}$ we get that

$$\left|\frac{\pi}{2} - \gamma\right| \leq \beta + \pi\delta_2 \leq \frac{3\delta_1}{r} + \pi\delta_2.$$ 

Hence, by the Law of Cosines,

$$|pv|^2 = |pu|^2 + |uv|^2 - 2|pu||uv| \cos \gamma = r^2 + |uv|^2 - 2r|uv| \sin (\pi/2 - \gamma).$$

Write the right-hand side as $r^2(1 + x)$, where

$$|x| = \frac{1}{r^2} | |uv|^2 - 2r|uv| \sin (\pi/2 - \gamma) |$$

$$\leq \frac{1}{r^2} (|uv|^2 + 2r|uv| (3\delta_1/r + \pi\delta_2)) = O\left(\delta_2^2 + \frac{\delta_1\delta_2}{r}\right).$$

We thus have $|pv| = r(1 + x)^{1/2}$, and $1 - |x| \leq (1 + x)^{1/2} \leq 1 + \frac{1}{2} |x|$ (where the left inequality holds for $|x| < 1$, which we may assume to be the case). In other words,

$$|pv| - r \leq r|x| = O\left(r\delta_2^2 + \delta_1\delta_2\right),$$

which, by the assumptions we have made, is $O(\varepsilon)$, as asserted. $\square$

**Lemma 4.5** (a) Let $(p, q) \in P \times Q$ be such that $r - \varepsilon \leq |pq| \leq r + \varepsilon$. Let $S_i$ be the small square containing $p$. Then $c_q$ must cross either $S_i$ or one of its adjacent squares. So there must be a (unique) index $j$ such that $(p, c_q) \in P_j \times C_j$.

(b) Let $j$ be the index for which $(p, c_q) \in P_j \times C_j$, and let $R'$ be the dual small polar rectangle (that arises in the dual processing of $S_j$) that contains $q$. Then the dual circle $c_p$ must cross either $R'$ or one of the two small rectangles lying directly above and below $R'$ (in the $p$-direction, if they exist).

**Proof.** The proof of part (a) is trivial: Since the distance between $c_q$ and $p$ is at most $\varepsilon = \delta_1\delta_2/\sqrt{2} \leq \delta_1$ (the latter inequality holds since $\delta_2 \leq 1$, by construction), $c_q$ must cross a square $S_j$ adjacent to $S_i$.

For part (b), let $o$ be the center of $S_j$, and let $b$ be the point on the ray through $oq$ such that $|pb| = r$ (for the assumed ranges of $r$ and $\varepsilon$, $b$ is unique). Assume that $b$ lies between $o$ and $q$; the case where $b$ lies beyond $q$ is handled analogously. It suffices to show that $|qb| \leq 2\varepsilon$, which is the $p$-range of a small polar rectangle $R'$. See Figure 5.
Figure 4: An illustration of the proof of Lemma 4.4.

Let $\beta = \angle obp$. Applying the Law of Sines in the triangle $\triangle obp$, we get that

$$\frac{|op|}{\sin \beta} = \frac{|pb|}{\sin \angle pob} \geq r.$$ 

Hence $\sin \beta \leq \frac{|op|}{r} \leq \frac{3\delta_1/\sqrt{2}}{r}$.

Since we assume that $\delta_1 \ll r$, we may also assume, as in the proof of Lemma 4.4, that $\beta < \frac{3\delta_1}{r} \ll 1$. Hence $\angle pbq > \pi/2$, and thus $|pq| > |pb| = r$. Let $a$ be the point on $pq$ for which $|pa| = r$. Applying the Law of Sines in the triangle $\triangle abq$, we get that

$$\frac{|qb|}{\sin \angle qab} = \frac{|qa|}{\sin \angle qba}. \quad (2)$$

By assumption, $|qa| \leq \epsilon$. Also, $\angle qba = \pi - \beta - \angle pba$. In the isosceles triangle $\triangle pab$, we have

$$\frac{\pi}{2} > \angle pba = \frac{\pi}{2} - \frac{\angle bpa}{2} > \frac{\pi}{2} - \frac{\beta}{2},$$

and therefore

$$\frac{\pi}{2} - \beta < \angle qba < \frac{\pi}{2} - \frac{\beta}{2}.$$ 

Substituting these bounds in Equation (2) we get

$$|qb| = \frac{|qa| \sin \angle qab}{\sin \angle qba} < \frac{\epsilon}{\cos(\pi/2 - \angle qba)} < \frac{\epsilon}{\cos \beta},$$

which is smaller than $2\epsilon$ when $\beta$ is small enough, that is, when $r$ is sufficiently larger than $\delta_1$. $\square$

The algorithm. The preceding analysis yields the following straightforward implementation, analogous to the one of Section 2. We first compute, for each point $p \in P$, the square
S_i that contains it and we find, for each circle c_q ∈ C, the squares that it crosses. This gives us all the sets P_i, C_i. We then iterate over the small squares in the partition of S. For each such square S_i, we construct the dual partitioning (in polar coordinates) of the resulting dual rectangle R_i into the smaller rectangles R'_i. As above, we find, for each dual point q, for which c_q ∈ C_i, the small rectangle R'_i that contains it, and, for each dual circle c_p, for p ∈ P_i, the small rectangles that it crosses. We now report, for each small rectangle R'_i, all the pairs (p,q) for which q lies in R'_i and c_p crosses either R'_i or one of the small rectangles lying directly above and below R'_i (in the ρ-direction, if they exist). We repeat this over all small primal squares S_i and all respective small rectangles R'_i. Note that a pair (p,q) may be reported more than once in this procedure, but its multiplicity is at most some small absolute constant.

As in the case of lines, the running time of this algorithm is O(n/δ_1 + m/δ_2 + k), where k is the output size. By Lemmas 4.4 and 4.5, every pair (p,q) at distance in [r − ε, r + ε] will be reported, and every reported pair lies at distance in [r − αε, r + αε], for some small absolute constant α, provided that we enforce the constraints δ_1 ≪ r, δ_2 ≤ √ε/r, and δ_1 δ_2 = √2ε.

As in Section 2, to minimize the running time, while satisfying δ_1 δ_2 = √2ε, we want to pick
\[ \delta_1 = \sqrt{\frac{2n\varepsilon}{m}}, \quad \text{and} \quad \delta_2 = \sqrt{\frac{m\varepsilon}{n}}. \]
The other two constraints amount to requiring that \( \varepsilon n \ll mr^2 \) and \( mr \leq n \). That is,
\[ \frac{\varepsilon}{r^2} \ll \frac{m}{n} \leq \frac{1}{r}. \] (3)
Since the problem is symmetric in P and Q, we may assume that m ≤ n (otherwise we simply flip the roles of P and Q). Hence the right inequality in (3) holds (recall that we assume that r ≤ 1/2). If the other inequality does not hold, say, \( m/n \leq 100\varepsilon/r^2 \), we skip
the primal stage, apply only the dual partitioning, with \( \delta_2 = \Theta(\varepsilon) \), and get the bound
\[
O\left(\frac{m}{\varepsilon} + n + k\right) = O\left(\frac{m}{\varepsilon} + k\right) = O(n + k).
\]
(Recall that we assume that \( r \) is bounded from below by a constant \( v \).)

In the remaining case, \([3]\) holds, and then \( \delta_1, \delta_2 \) are both \( \leq 1 \), as is easily checked, and then the bound is \( O(\sqrt{mn}/\sqrt{\varepsilon} + k) \). Including the symmetric case \( m \geq n \), we get the following theorem, which improves upon Theorem \([4.3]\) when the values of \( m \) and \( n \) are “unbalanced”.

**Theorem 4.6** Let \( P \) and \( Q \) be two sets of \( m \) and \( n \) points, respectively, in the unit disk \( B \), and let \( v \leq r \leq 1/2 \) for a constant \( v \). We can report all pairs \((p,q) \in P \times Q\) for which \(|pq| - r| \leq \varepsilon\), in
\[
O\left(m + n + \frac{\sqrt{mn}}{\sqrt{\varepsilon}} + k\right),
\]
time, where \( k \) is the actual number of pairs that we report; all pairs at distance in \([r - \varepsilon, r + \varepsilon]\) will be reported, and every reported pair lies at distance in \([r - \alpha\varepsilon, r + \alpha\varepsilon]\), for some absolute constant \( \alpha \).

## 5 Near-neighbor point-circle configurations

In this section we study the near-neighbor problem for points and arbitrary circles, extending the results from the previous section. Specifically, we are given a set \( P \) of \( m \) points in the unit disk \( B \) in the plane, and a set \( C \) of \( n \) circles intersecting \( B \), where we assume that the radii of the circles in \( C \) all lie in a fixed interval \([r_1, r_2]\), for \( \varepsilon \leq r_1 \leq 1/2 \leq r_2 \). We want to compute the \( \varepsilon \)-approximate incidences between \( P \) and \( C \).

We solve this problem by a two-stage partition scheme, similar to those used before, except that one stage takes place in the plane, and the other in three dimensions.

The first stage is more or less identical to that used in Section \([4.2]\). That is, we assume that \( P \) is contained in the unit square \( S = [0,1]^2 \). We fix two real positive parameters \( \delta_1 \), \( \delta_2 \), whose values will be set later. We partition \( S \) into \( \frac{1}{\delta_1^2} \) pairwise openly disjoint smaller squares, each of side length \( \delta_1 \). We enumerate these squares as \( S_1, \ldots, S_{1/\delta_1^2} \), and let \( \hat{S}_i \) denote the union of \( S_i \) and the (at most) eight squares adjacent to \( S_i \). Let \( P_i \) denote the set of all points of \( P \) that lie in \( \hat{S}_i \), and let \( C_i \) denote the set of all the circles \( c \in C \) that cross \( S_i \). Put \( m_i := |P_i| \) and \( n_i := |C_i| \), for \( i = 1, \ldots, 1/\delta_1^2 \). We have \( \sum_i m_i = O(m) \), and \( \sum_i n_i = O(n/\delta_1) \).

The second stage is different, because the varying values of \( r \) do not allow us to apply the simple duality that we used in Section \([4.2]\). Instead we first move to a different notion of distance between a point and a circle, which is the power (see, e.g., \([7]\)). The power of a point \( p \) with respect to a circle \( c \) centered at \( q \) with radius \( r \) is \( \Pi(p,c) = |pq|^2 - r^2 \).

The notions of Euclidean distance and power are closely related: Let \( p \) be a point and \( c \) a circle centered at a point \( q \) with radius \( r \). Notice that \( \text{dist}(p,c) = |pq| - r \) and
\[
|\Pi(p,c)| = |pq|^2 - r^2 = \text{dist}(p,c)(|pq| + r) = \text{dist}(p,c)(2r + \text{dist}(p,c)).
\]
Hence, we always have
\[
|\Pi(p,c)| \geq 2r \text{dist}(p,c) \geq 2r_1 \text{dist}(p,c), \quad (4)
\]
and if $\text{dist}(p, c) \leq r$, which certainly holds for all circles $c$ which are approximately incident to $p$, we have
\[ |\Pi(p, c)| \leq 3r \text{dist}(p, c) \leq 3r_2 \text{dist}(p, c). \]  

By Equation (5), for every pair $p, c$ such that $\text{dist}(p, c) \leq \varepsilon$ we have that $|\Pi(p, c)| \leq 3r_2 \varepsilon$. On the other hand, for a pair $p, c$ such that $|\Pi(p, c)| \leq 3r_2 \varepsilon$ we know, by Equation (4), that $\text{dist}(p, c) \leq 3r_2 \varepsilon / 2r_1$. Thus, our task now is to report all pairs $p, c$ such that $|\Pi(p, c)| \leq 3r_2 \varepsilon$.

We use the standard lifting transform where each point $p = (x, y) \in \mathbb{R}^2$ is mapped to the point $(x, y, x^2 + y^2)$ on the paraboloid $z = x^2 + y^2$ in 3-space, and each circle $c$ with center $q = (q_1, q_2)$ and radius $r$ is mapped to the plane $z = 2q_1 x + 2q_2 y + (r^2 - q_1^2 - q_2^2)$. The vertical distance between the images of $p$ and $c$ is
\[ |x^2 + y^2 - 2q_1 x - 2q_2 y + q_1^2 + q_2^2 - r^2| = ||pq||^2 - r^2 = |\Pi(p, c)|. \]

We now dualize 3-space by mapping points to planes and planes to points, in a standard manner that preserves vertical distances between points and planes. We get a set $P_i^*$ of $m_i$ planes and a set $C_i^*$ of $n_i$ points, and want to report all point-plane pairs at vertical distance $\leq 3r_2 \varepsilon$. This is handled exactly as in the second stage in Section 3.

Specifically, since each circle $c \in C_i$ crosses $S_i$, its distance from the center $o_i$ of $S_i$ is at most $\delta_1 / \sqrt{2}$, so the vertical distance between the plane $o_i^*$ and the point $c^*$ is at most $3r_2 \cdot \delta_1 / \sqrt{2} < 3r_2 \delta_1$. Moreover, the $xy$-projection of $c^*$ is $2q$, which lies in a suitable annulus centered at $2o_i$ with radii proportional to $r_1$ and $r_2$; this holds if we require that $\delta_1 < r_1$, say. For simplicity, enclose this annulus by an axis-parallel square $R_0$ of side length $cr_2$, for a suitable constant $c$, and let $R$ denote the parallelepiped bounded between the two planes that are shifts of $o_i^*$ by $\pm 3r_2 \delta_1$ and having $R_0$ as its $xy$-projection.

We now partition $R$ into $O(1/\delta_2^2)$ small homothetic copies, each scaled down by $\delta_2$. Each small region $R'$ has an $xy$-projection of size $cr_2 \delta_2 \times cr_2 \delta_2$, and its vertical width (in the $z$-direction) is $3r_2 \delta_1 \delta_2$.

For each small region $R'$, we report all the pairs $(p, c) \in P_i \times C_i$ for which $c^* \in R'$ and $p^*$ crosses either $R'$ or one of the two regions above and below $R'$ with the same $xy$-projection. Each dual plane $p^*$ crosses $O(1/\delta_2^2)$ small regions.

Applying the arguments used in the case of points and planes in $\mathbb{R}^3$, given in Section 3 and in particular ensuring that $\delta_1 \delta_2 = \varepsilon$, we conclude that the algorithm correctly reports all pairs $(p, c)$ for which $\text{dist}(p, c) \leq \varepsilon$, and that each pair that it reports satisfies $\text{dist}(p, c) \leq \alpha \varepsilon$, for a suitable constant $\alpha$. The overall running time is
\[ O \left( \frac{n}{\delta_1} + \frac{m}{\delta_2} + k \right), \]
where $k$ is the number of pairs that we report. To optimize this bound, we choose $\delta_1$ and $\delta_2$ to satisfy
\[ \frac{n}{\delta_1} = \frac{m}{\delta_2^2}, \quad \text{and} \quad \delta_1 \delta_2 = \varepsilon, \]
that is,
\[ \delta_1 = \left( \frac{\varepsilon^2 n}{m} \right)^{1/3}, \quad \text{and} \quad \delta_2 = \left( \frac{\varepsilon m}{n} \right)^{1/3}. \]

We require that $\delta_1 < r_1$, $\delta_2 \leq 1$. In case $\delta_1 > r_1$, that is, $m < n \varepsilon^2 r_1^3$, we skip the first stage, and run the second stage over the full data, with $\delta_2 = \varepsilon$, resulting in running time
\(O\left(\frac{m^2}{\varepsilon^2} + n + k\right) = O(n + k)\). Similarly, in case \(\delta_2 > 1\), that is, \(n < m\varepsilon\), we perform only the first stage, with \(\delta_1 = \varepsilon\), and the running time is then \(O\left(m + \frac{n}{\varepsilon} + k\right) = O(m + k)\). Otherwise the running time is \(O\left(m^{1/3}n^{2/3}/\varepsilon^{2/3} + k\right)\).

Thus, we have obtained the following theorem.

**Theorem 5.1** Let \(P\) be a set of \(m\) points in the unit disk \(B\) in \(\mathbb{R}^2\), let \(C\) be a set of \(n\) circles of radii in the range \([r_1, r_2]\), for some positive constants \(\varepsilon \leq r_1 \leq 1/2 \leq r_2\), that cross \(B\), when \(\varepsilon > 0\) is a prescribed error parameter. We can report all pairs \((p, q) \in P \times C\) for which \(\text{dist}(p, c) \leq \varepsilon\) in time

\[
O\left( m + n + \frac{m^{1/3}n^{2/3}}{\varepsilon^{2/3}} + k \right),
\]

where \(k\) is the actual number of pairs that we report; all pairs at distance at most \(\varepsilon\) will be reported, and every reported pair lies at distance at most \(\alpha\varepsilon\) for some constant \(\alpha\) (proportional to \(r_2/r_1\)).

### 6 Reporting all nearly congruent pairs in three dimensions

In this section we consider the three-dimensional version of the problem studied in Section 4. That is, we are given sets \(P\) and \(Q\) of \(m\) and \(n\) points, respectively, in the unit ball \(B\) in \(\mathbb{R}^3\), and parameters \(0 < v \leq r \leq 1/2\), for a constant \(v\), and wish to report all pairs \((p, q) \in P^2\) such that \(\text{dist}(p, q) \in [r - \varepsilon, r + \varepsilon]\). As usual, we allow more pairs to be reported, but require that each pair \((p, q)\) that we report satisfies \(\text{dist}(p, q) \in [r - \alpha\varepsilon, r + \alpha\varepsilon]\), for some absolute constant \(\alpha\). This is the approximate incidences problem between \(P\) and spheres of radius \(r\) centered at the points of \(Q\).

As in Section 4, there are two alternative solutions, one using the technique of Indyk et al. [14], and one using duality. In the following we derive Theorem 6.2 using Indyk et al.’s approach. We omit the tedious derivation using duality which would give a result analogous to Theorem 4.6 (with \(\varepsilon\) rather than \(\sqrt{\varepsilon}\) in the denominator).

Let \(\sigma_o\) denote the sphere of radius \(r\) centered at the origin \(o\). We can cover \(\sigma_o\) with \(O(1/\varepsilon)\) congruent caps, each of opening angle \(\sqrt{\varepsilon}\), so that no point of \(\sigma_o\) is covered by more than \(O(1)\) caps. Let \(U\) be the set of directions from \(o\) to the centers of these caps, \(|U| = O(1/\varepsilon)|. In the following we fix one direction \(u \in U\), which, without loss of generality, we assume to be the positive \(z\)-direction.

Let \(\tilde{u}\) denote the cap of \(\sigma_o\) with \(u\) as a central direction. Let \(A_{\tilde{u}}\) be a cap portion of a spherical shell centered at \(o\), with inner radius \(r - \varepsilon\) and outer radius \(r + \varepsilon\), which is the intersection of the entire such shell with the cone with apex \(o\), axis \(u\), and opening angle \(\sqrt{\varepsilon}\). Let \(R_{\tilde{u}}\) denote the smallest enclosing axis-parallel box of \(A_{\tilde{u}}\) (Figure 3 can serve as a schematic two-dimensional illustration of this setup).

Let \(R_{\tilde{u}}(q)\) be \(R_{\tilde{u}}\) translated by a vector (point) \(q \in Q\). Let \(R\) denote the collection of the boxes \(R_{\tilde{u}}(q)\), for \(q \in Q\). Note that the members of \(R\) are translates of one another.

---

2One can do this by packing disjoint caps of opening angle \(\sqrt{\varepsilon}/2\) on \(\sigma_o\), and taking \(U\) as the set of directions to the centers of these caps.
We now construct a grid $G$ whose cells are translates of $R_\bar{u}$, assign each point of $P$ to the cell of $G$ containing it, and assign each point $q \in Q$ to the at most (exactly, in general position) eight cells that $R_\bar{u}^*(q)$ overlaps. We then report, over all grid cells, all the pairs $(p, q) \in P \times Q$ that are assigned to the same cell.

We repeat this procedure for each of the $O(1/\varepsilon)$ orientations in $U$, and the overall output of the algorithm is the union of the outputs for the individual orientations. The overall running time is $O((m + n)/\varepsilon + k)$, where $k$ is the number of distinct pairs that we report. The term $O(k)$ is justified because each pair $(p, q)$ is reported once for each shell-cap of $q$ such that the box $R_\bar{u}^*$ which is a homothetic copy $R_\bar{u}$ scaled by a factor of 3 about its center, contains $p$. It is easy to check that if $p$ lies in $R_\bar{u}^*$ then the angle between $\bar{q}p$ and $u$ is $O(\sqrt{\varepsilon})$ so there could be only $O(1)$ such directions $u$.

As in Section 4.1 the correctness of this algorithm is a consequence of the following lemma.

**Lemma 6.1** (a) Let $q$ be a point at distance $\leq \varepsilon$ from $\sigma_o$, so that the point of $\sigma_o$ nearest to $q$ lies in $\bar{u}$. Then $q \in R_\bar{u}$.

(b) Let $R_\bar{u}^*$ denote the homothetic copy $R_\bar{u}$ scaled by a factor of 3 about its center. Then every point in $R_\bar{u}^*$ is at distance $\leq \alpha \varepsilon$ from $\sigma_o$, for a suitable small absolute constant $\alpha$.

**Proof.** (a) is trivial since in this case $q$ must lie in $A_\bar{u}$ and therefore also in $R_\bar{u}$.

We establish (b) by giving a lower (resp., upper) bound on the shortest (resp., longest) distance of a point in $R_\bar{u}^*$ from $o$.

Clearly, the point of $R_\bar{u}$ closest to $o$ is the center point $\mu$ of its bottom face. This point $\mu$ lies on the cross section of $R_\bar{u}$ with the $yz$-plane. This cross section is congruent to the rectangle $R_{\bar{a}b}$ of Figure 3 bounding the annulus $A_{\bar{a}b}$ around an arc $\bar{a}b$ with opening angle $\sqrt{\varepsilon}$.

Arguing as in the proof of Lemma 4.1 the distance of $\mu$ from $o$ is at least $r - 2\varepsilon$, and therefore the distance of the center $\mu^*$ of the bottom face of $R_\bar{u}^*$ from $o$ is at least $r - 5\varepsilon$.

The points of $R_\bar{u}^*$ farthest from $o$ are the four vertices of its top face $f$. Arguing as in the proof of Lemma 4.1 the center point of $f$ lies at distance at most $r + 4\varepsilon$ from $o$. The side length of $f$ is the same as the side length of its cross section with the $yz$-plane, which is at most $3\sqrt{\varepsilon}$, as in the proof of Lemma 4.1 so the distance of a vertex of $f$ from its center point is at most $3\sqrt{\varepsilon/2}$. By the Pythagorean theorem, we obtain that the distance of $o$ from a vertex of $f$ is at most

$$(r + 4\varepsilon)^2 + 9\varepsilon/2)^{1/2} = (r^2 + (8r + 9/2)\varepsilon + 16\varepsilon^2)^{1/2} < r + \alpha \varepsilon,$$

for a suitable constant $\alpha$ that depends on $r$ (analogously to the analysis in Section 4.1). □

The following theorem summarizes the main result of this section.

**Theorem 6.2** Let $P$ and $Q$ be two sets, of respective sizes $m$ and $n$, in the unit ball $B$ in $\mathbb{R}^3$, and let $0 < v \leq r \leq 1/2$ for some constant $v$. For any small $\varepsilon$, we can report all pairs $(p, q) \in P \times Q$ for which $\text{dist}(p, q) \in [r - \varepsilon, r + \varepsilon]$, in

$$O\left(\frac{m + n}{\varepsilon} + k\right)$$
time, where \( k \) is the actual number of distinct pairs that we report. All pairs at distance in \([r - \varepsilon, r + \varepsilon]\) will be reported, and every reported pair lies at distance in \([r - \alpha \varepsilon, r + \alpha \varepsilon]\), for some constant \( \alpha \) that depends on \( r \).

**Remark.** Note that both techniques work in any dimension, more or less verbatim. Consider for example Indyk et al.’s technique. One major difference is that the size of the set \( U \) of directions in \( d \) dimensions is \( O(1/\varepsilon^{(d-1)/2}) \), so the algorithm runs in time \( O((m + n)/\varepsilon^{(d-1)/2} + k) \); the naive grid-based approach, discussed in the introduction, would take \( O(m + n/\varepsilon^{d-1} + k) \). There is also the issue of applying the Pythagorean theorem, where the factor \( \sqrt{2} \) has to be replaced by \( \sqrt{d - 1} \). The rest of the analysis goes more or less unchanged.

### 7 Reporting all point-line neighbors in three dimensions

Let \( P \) be a set of \( m \) points in the unit ball \( B \) in three dimensions, let \( L \) be a set of \( n \) lines that cross \( B \), and let \( \varepsilon > 0 \) be a given error parameter. In this section we present an algorithm for the approximate incidence reporting problem involving \( P \) and \( L \).

We represent each line in \( \mathbb{R}^3 \) by the pair of equations\(^3\)

\[
y = ax + b, \quad z = cx + d.
\]

Let \( \ell \) be the line \( y = ax + b, \quad z = cx + d \), and let \( p = (\xi, \eta, \zeta) \) be a point in \( \mathbb{R}^3 \). We approximate \( \text{dist}(p, \ell) \) by slicing space by the plane \( \pi_p : x = \xi \), and by computing the distance between the points \( p \) and \( \ell_p := \ell \cap \pi_p = (\xi, a\xi + b, c\xi + d) \).

As in Section 2, for this approximation to be good, the angle between \( \ell \) and the \( x \)-direction should not be too large. To ensure this, similarly to what we did in Section 3, we partition \( L \) into \( O(1) \) subfamilies, such that, for each subfamily \( L' \) there exists a direction \( u' \) such that the angle between \( u' \) and each line of \( L' \) is at most \( \pi/4 \). We apply the construction to each subset \( L' \) separately (with respect to all the points in \( P \)). To apply it for a specific \( L' \), we first rotate \( \mathbb{R}^3 \) so that \( u' \) becomes the \( x \)-direction. In what follows, we fix one subfamily, continue to denote it as \( L \), and assume that \( u' \) is indeed the \( x \)-direction.

We make the following two easy observations (compare with the analysis in Section 3).

**Lemma 7.1** The slopes \( a \) and \( c \) of any line \( y = ax + b, \quad z = cx + d \) in \( L \) satisfy \( a^2 + c^2 \leq 1 \).

**Proof.** The parametric representation of the line \( y = ax + b, \quad z = cx + d \) is \( \{(0, b, d) + t(1, a, c) \mid t \in \mathbb{R}\} \). By our assumption, the angle \( \gamma \) between the vectors \( (1, a, c) \) and \( (1, 0, 0) \) is at most \( \pi/4 \). Hence we have

\[
\cos \gamma = \frac{1}{\sqrt{1 + a^2 + c^2}} \geq \cos(\pi/4) = \frac{1}{\sqrt{2}},
\]

from which the lemma follows. \( \square \)

**Lemma 7.2** Let \( p = (\xi, \eta, \zeta) \) be a point in \( P \), let \( \ell \) be a line in \( L \), and let \( \ell_p := \ell \cap \pi_p \) where \( \pi_p \) is the plane \( x = \xi \). Then \( \text{dist}(p, \ell_p) \leq \sqrt{2}\text{dist}(p, \ell) \).

\(^3\)We assume without loss of generality that no line is orthogonal to the \( x \)-axis.
Proof. Let \(a\) be the point on \(\ell\) closest to \(p\), and consider the triangle \(\triangle a\ell p\). Since the angle between \(\ell\) and the \(x\)-direction is at most \(\pi/4\), the angle \(\theta_0\) between \(\ell\) and its projection on \(\pi_p\) is at least \(\pi/4\). Since \(\theta_0\) is the smallest angle between \(\ell\) and any line in \(\pi_p\), it follows that the angle \(\angle a\ell p\) is also at least \(\pi/4\), and therefore

\[
\frac{\text{dist}(p, \ell)}{\text{dist}(p, \ell_p)} = \frac{\text{dist}(p, a)}{\text{dist}(p, \ell_p)} \geq \sin(\pi/4) = \frac{1}{\sqrt{2}},
\]

as claimed. See Figure 6. \(\Box\)

Figure 6: We approximate the distance between \(p\) and \(\ell\) by the distance between \(p\) and \(\ell_p\).

As in the preceding sections, we replace \(B\) by the unit cube \(S = [0,1]^3\) and we assume \(P \subseteq S\). Then we apply the following two-stage partitioning procedure.

**The primal stage.** We fix two parameters \(\delta_1, \delta_2\), whose values will be set later. We first partition \(S\) into \(\frac{1}{\delta_1}\) pairwise openly disjoint smaller cubes, each of side length \(\delta_1\). Enumerate these cubes as \(S_1, S_2, \ldots, S_{\frac{1}{\delta_1}}\). For \(i = 1, \ldots, \frac{1}{\delta_1}\), let \(P_i\) denote the set of all points of \(P\) that lie in \(S_i\) or in one of the (at most) eight cubes that surround \(S_i\) and have the same \(x\)-projection as \(S_i\). Let \(L_i\) denote the set of all the lines of \(L\) that cross \(S_i\). Put \(m_i := |P_i|\) and \(n_i := |L_i|\), for \(i = 1, \ldots, \frac{1}{\delta_1}\). We have \(\sum_i m_i \leq 9m\), and \(\sum_i n_i \leq 3n/\delta_1\), as is easily checked (to cross from a cube to an adjacent cube, the line has to cross one of the \(3/\delta_1\) planes that define the grid).

**The dual stage.** For each such small cube \(S_i\), we now pass to a parametric dual four-dimensional space (with coordinates \((x, y, z, w)\)), in which we represent each line \(\ell \in L_i\), given by \(y = ax + b, z = cx + d\), by the point \(\ell^* = (a, b, c, d)\), and represent each point \(p = (\xi, \eta, \zeta) \in P_i\) by the 2-plane (in \(\mathbb{R}^4\))

\[
p^* = \{(a, b, c, d) \mid a\xi + b = \eta, c\xi + d = \zeta\};
\]

\(p^*\) is the locus of all points dual to lines that pass through \(p\).

We define the distance in the dual space between a point \(\ell^* = (a, b, c, d)\) and a plane \(p^*\), for a primal point \(p = (\xi, \eta, \zeta)\), to be the distance between \(\ell^*\) and the point \((a, \eta - a\xi, c, \zeta - c\xi)\), which is the intersection of \(p^*\) with the plane defined by \(x = a\) and \(z = c\). In the primal space, the point \((a, \eta - a\xi, c, \zeta - c\xi)\) corresponds to a line parallel to \(\ell\) that passes through \(p\). The following lemma is immediate from this definition.
Lemma 7.3 The distance between $\ell^*$ and $p^*$, as defined above, is equal to $\text{dist}_\ast(p, \ell_p)$ in the primal space.

Fix a small cube $S_i$, and assume without loss of generality that $S_i = [0, \delta_1]^3$. Let $\ell$ be a line in $L_i$, given by $y = ax + b$, $z = cx + d$. Since we assume that the angle between each each line of $L$ and the $x$-axis is at most $\pi/4$, the $y$- and $z$-spans of the intersection of $\ell$ with the slab $0 \leq x \leq \delta_1$ are each at most $\delta_1$. This implies that $-\delta_1 \leq b \leq 2\delta_1$, and $-\delta_1 \leq d \leq 2\delta_1$.

It also follows from Lemma 7.1 that $|a|, |c| \leq 1$. Therefore, in the dual parametric four-dimensional space, $\ell^*$ lies in the box $R$ given by

1. $-1 \leq a, c \leq 1$
2. $-\delta_1 \leq b, d \leq 2\delta_1$.

We now partition $R$ into $1/\delta_2^4$ smaller boxes, each of which is a homothetic copy of $R$ scaled down by $\delta_2$. Concretely, each smaller box $R'$ is congruent to the box $[0, 2\delta_2] \times [0, 3\delta_1\delta_2] \times [0, 2\delta_2] \times [0, 3\delta_1\delta_2]$.

Lemma 7.4 For each small box $R'$, if $\ell^* = (a_\ell, b_\ell, c_\ell, d_\ell)$ is a dual point (of some $\ell \in L_i$) in $R'$ and $p^*$ is a dual plane (of some point $p = (\xi, \eta, \zeta) \in P_i$) that crosses $R'$ or one of its (at most eight) surrounding boxes of the same $xz$-range, then $\text{dist}(p, \ell) \leq 8\sqrt{2}\delta_1\delta_2$.

Proof. Assume without loss of generality that $S_i$ is the cube $[0, \delta_1]^3$ and that $R'$ is the box $[0, 2\delta_2] \times [0, 3\delta_1\delta_2] \times [0, 2\delta_2] \times [0, 3\delta_1\delta_2]$. Since $\ell^*$ is in $R'$ we have

1. $0 \leq a_\ell, c_\ell \leq 2\delta_2$ \hspace{1cm} (6)
2. $0 \leq b_\ell, d_\ell \leq 3\delta_1\delta_2$. \hspace{1cm} (7)

Let $(a, b, c, d)$ be a point in $p^* \cap \hat{R}'$, where $\hat{R}'$ is $R'$ or one of its surrounding boxes of the same $xz$-range. By definition of $p^*$, we have $b = \eta - a\xi$ and $d = \zeta - c\xi$. Since $(a, b, c, d) \in \hat{R}'$, we have

3. $0 \leq a, c \leq 2\delta_2$ \hspace{1cm} (8)
4. $-3\delta_1\delta_2 \leq \eta - a\xi, \zeta - c\xi \leq 6\delta_1\delta_2$. \hspace{1cm} (9)

Finally, since $p \in P_i$, we have

5. $0 \leq \xi \leq \delta_1$ \hspace{1cm} (10)
6. $-\delta_1 \leq \eta \leq 2\delta_1$
7. $-\delta_1 \leq \zeta \leq 2\delta_1$.

We have

$$\text{dist}(p, \ell) \leq \text{dist}(p, \ell_p) = \left((a_\ell\xi + b_\ell - \eta)^2 + (c_\ell\xi + d_\ell - \zeta)^2\right)^{1/2}.$$ 

Let us estimate the first term $a_\ell\xi + b_\ell - \eta$ in the square root; the second term is estimated in a fully analogous manner. We have

$$a_\ell\xi + b_\ell - \eta = (a_\ell - a)\xi + b_\ell + a\xi - \eta.$$ 

Note that we include here adjacent regions that lie outside $R$, because suitable shifts of them would arise when $R'$ is a generic small region.
By Equations (6) and (8), \((a_\ell - a) \in [-2\delta_2, 2\delta_2]\), and by Equation (10), \(\xi \in [0, \delta_1]\), so \((a_\ell - a)\xi \in [-2\delta_1\delta_2, 2\delta_1\delta_2]\). By Equation (7), \(b_\ell \in [0, 3\delta_1\delta_2]\), and by equation (9), \(a\xi - \eta \in [-6\delta_1\delta_2, 3\delta_1\delta_2]\). Adding up these estimates, we get

\[
a_\ell\xi + b_\ell - \eta \in [-8\delta_1\delta_2, 8\delta_1\delta_2],
\]

and, by a fully symmetric argument,

\[
c_\ell\xi + d_\ell - \zeta \in [-8\delta_1\delta_2, 8\delta_1\delta_2].
\]

Hence \(\text{dist}(p, \ell) \leq 8\sqrt{2}\delta_1\delta_2\), as asserted. \(\square\)

For the following lemma, we constrain \(\delta_1\) and \(\delta_2\) to satisfy \(\delta_1 \delta_2 = \varepsilon\), and \(\delta_2 \leq 1/\sqrt{2}\).

**Lemma 7.5**

(a) Let \(p = (\xi, \eta, \zeta)\) be a point in \(P\) and \(\ell\) be a line in \(L\), given by \(y = ax + b, z = cx + d\), such that \(\text{dist}(p, \ell) \leq \varepsilon\). Let \(S_i\) be the small primal cube containing \(p\). Then \(\ell\) must cross either \(S_i\) or one of the at most eight cubes that surround \(S_i\) and have the same \(x\)-projection as \(S_i\). Therefore, there exists a \(j\) such that \((p, \ell) \in P_j \times L_j\).

(b) Let \(p\) and \(\ell\) be as in (a), let \(j\) be such that \((p, \ell) \in P_j \times L_j\), and let \(R'\) be the dual small region in \(\mathbb{R}^4\) (that arises in the dual processing of \(S_j\)) that contains \(\ell^*\). Then the dual plane \(p^*\) must cross either \(R'\) or an adjacent small region with the same \(xz\)-projection.

**Proof.**

(a) The line \(\ell\) crosses the plane \(x = \xi\) at the point \(q = (\xi, a\xi + b, c\xi + d)\) which, by Lemma 7.2, lies at distance at most \(\sqrt{2}\varepsilon\) from \(p\). That is, \(q\) lies in a cube with the same \(x\)-projection as \(S_i\), at distance at most \(\sqrt{2}\varepsilon = \sqrt{2}\delta_1\delta_2\) from \(S_i\), so, for \(\delta_2 \leq 1/\sqrt{2}\), it must lie either in \(S_i\) or in one of the eight surrounding cubes with the same \(x\)-projection, as claimed.

(b) Assume without loss of generality that \(R' = [0, 2\delta_2] \times [0, 3\delta_1\delta_2] \times [0, 2\delta_2] \times [0, 3\delta_1\delta_2]\). The unique intersection point of \(p^*\) with the plane \(\pi: x = a, z = c\) is the point \(\lambda^* = (a, \eta - a\xi, c, \zeta - c\xi)\). Within \(\pi\), the absolute value of the \(y\)-shift (resp., \(w\)-shift) between \(\ell^*\) and \(\lambda^*\) is \(|a\xi + b - \eta|\) (resp., \(|c\xi + d - \zeta|\)). By construction and by Lemma 7.2, each of these quantities is at most \(d(p, \ell_P) \leq \sqrt{2}\varepsilon = \sqrt{2}\delta_1\delta_2\). This implies that the region \(R''\) containing \(\lambda^*\) must be adjacent to \(R'\) and of the same \(xz\)-projection as \(R'\). \(\square\)

**The algorithm.**

The preceding analysis yields the following straightforward implementation. We compute the sets \(P_i, L_i\), for \(i = 1, \ldots, 1/\delta_1^3\), in overall \(O(m + n/\delta_1)\) time. Then, for each small cube \(S_i\), we consider the partitioning of the resulting dual box \(R\) into the smaller boxes \(R'\). As above, we find, for each \(\ell \in L_i\), the small region that contains the dual point \(\ell^*\), and, for each \(p \in P_i\), the small regions that the dual plane \(p^*\) crosses. This takes \(O(n + m/\delta_2^3)\) time.

We now report, for each small region \(R'\), all the pairs \((p, \ell) \in P_i \times L_i\) for which \(\ell^*\) lies in \(R'\) and \(p^*\) crosses either \(R'\) or one of the at most eight small regions that surround \(R'\) and have the same \(xz\)-range. We repeat this over all small cubes \(S_i\) and all respective small regions \(R'\). The running time of the algorithm is

\[
O\left(\frac{m + n}{\delta_1} + \frac{1}{\delta_1^3} \sum_{i=1}^{1/\delta_1^3} \left(\frac{n}{\delta_1} + \frac{m}{\delta_2^3}\right) + k\right) = O\left(\frac{n}{\delta_1^3} + \frac{m}{\delta_2^3} + k\right),
\]
where \( k \) is the number of distinct point-line pairs that we report. By Lemmas 7.4 and 7.5, every pair \((p, \ell)\) at distance at most \( \varepsilon = \delta_1 \delta_2 \) will be reported, and every reported pair lies at distance at most \( 8\sqrt{2} \varepsilon \). Moreover, no pair is reported more than a constant number of times. To minimize the running time overhead, as a function of \( \varepsilon \), we choose \( \delta_1 \) and \( \delta_2 \) to satisfy

\[
\frac{n}{\delta_1} = \frac{m}{\delta_2^2}, \quad \text{and} \quad \delta_1 \delta_2 = \varepsilon,
\]

that is,

\[
\delta_1 = \left( \frac{\varepsilon^2 n}{m} \right)^{1/3}, \quad \text{and} \quad \delta_2 = \left( \frac{\varepsilon m}{n} \right)^{1/3}.
\]

As in all the preceding cases, we need to require that \( \delta_1 \) and \( \delta_2 \) are both at most 1, and in fact we want \( \delta_2 \) to be at most \( 1/\sqrt{2} \). If one of these conditions does not hold, we skip the corresponding primal or dual stage, set the other parameter to \( \varepsilon \), and conclude, as is easily verified, that the running time is \( O(m + n + k) \). Otherwise, the running time is \( O \left( m^{1/3} n^{2/3} / \varepsilon^{2/3} + k \right) \).

In summary, we obtain the following result.

**Theorem 7.6** Let \( P \) be a set of \( m \) points in the unit ball \( B \) in \( \mathbb{R}^3 \), let \( L \) be a set of \( n \) lines in \( \mathbb{R}^3 \) that cross \( B \), and let \( \varepsilon > 0 \) be a prescribed error parameter. We can report all pairs \((p, \ell) \in P \times L \) for which \( \text{dist}(p, \ell) \leq \varepsilon \) in

\[
O \left( m + n + \frac{m^{1/3} n^{2/3}}{\varepsilon^{2/3}} + k \right)
\]

time, where \( k \) is the actual number of distinct pairs that we report. All pairs at distance \( \leq \varepsilon \) will be reported, and every reported pair lies at distance at most \( 8\sqrt{2} \varepsilon \).

8 Reporting all point-circle neighbors in three dimensions

In preparation for our final algorithm, of finding all nearly congruent copies of a given triangle in a set of \( n \) points in \( \mathbb{R}^3 \), we first solve the following problem. Let \( P \) be a set of \( m \) points in the unit ball \( B \) in \( \mathbb{R}^3 \), let \( C \) be a set of \( n \) congruent circles in \( \mathbb{R}^3 \) of radius \( r \leq 1/2 \) that cross \( B \), and let \( \varepsilon \ll r \) be a prescribed error parameter. We present an efficient algorithm for the approximate incidence reporting problem for \( P \) and \( C \).

We relax the problem further, as follows. For each circle \( c \in C \), denote by \( \lambda_c \) the axis of \( c \), which is the line that passes through the center of \( c \) and is orthogonal to the plane \( \pi_c \) containing \( c \). We partition \( C \) into \( O(1) \) subsets, corresponding to some canonical set \( U \) of \( O(1) \) directions in \( \mathbb{R}^3 \), so that we associate with each \( u \in U \) all the circles \( c \in C \) for which the angle between \( u \) and \( \lambda_c \) is at most some small but constant value \( \theta_0 \); in general, this is not a partition, but we turn it into a partition by assigning each circle \( c \) to an arbitrary single set from among those it belongs to. We apply the following procedure separately for each of these subsets, and focus on a single such set, where we assume, without loss of generality, that the corresponding direction \( u \) is the \( z \)-vertical direction. For simplicity of notation, continue to denote the corresponding subset of \( C \) as \( C \).
Now fix a circle \( c \in C \), and denote by \( T_c \) the torus that is the Minkowski sum of \( c \) and the ball \( B_\varepsilon \) of radius \( \varepsilon \) centered at the origin; our goal is to report all points in \( P \cap T_c \), for each circle \( c \in C \).

Let \( c_0 \) be the circle of radius \( r \) in the \( xy \)-plane centered at the origin. We partition \( T_{c_0} \) into sectors by roughly \( \pi/\sqrt{\varepsilon} \) planes through the \( z \)-axis such that the dihedral angle between each pair of consecutive planes is \( \sqrt{\varepsilon} \). We enumerate the sectors as \( S_{c_0}^{(1)}, \ldots, S_{c_0}^{(\kappa)} \), where \( \kappa = \pi/\sqrt{\varepsilon} \). We now focus on one such sector \( S_{c_0}^{(j)} \) and to simplify the notation drop the index \( j \) from \( S_{c_0}^{(j)} \) hereafter. Let \( \gamma_{c_0} \) be the arc \( c_0 \cap S_{c_0} \) of \( c_0 \), and let \( s_{c_0} \) be the chord connecting the endpoints of \( \gamma_{c_0} \). We rotate \( \mathbb{R}^3 \) around the \( z \)-axis so that \( s_{c_0} \) is parallel to the \( x \)-axis. Let \( Q_{c_0} \) be the smallest cylinder enclosing \( S_{c_0} \) whose axis is parallel to \( s_{c_0} \). The cross section of \( Q_{c_0} \) with the \( xy \)-plane is a rectangle \( R_{c_0} \), similar to the one shown in Figure \ref{fig:compare} (where \( \overline{ab} \) is now \( \gamma_{c_0} \)). As the calculations in Section \ref{sec:poincare} show, the width of \( R_{c_0} \) is \( < 3\varepsilon \) and its length is \( < \sqrt{\varepsilon} \) (and these bounds are tight up to an absolute constant factor). In other words, the radius of \( Q_{c_0} \) is at most \( \varepsilon' = 1.5\varepsilon \).

Now consider a circle \( c \in C \). Let \( \pi \) be the translation of the \( xy \)-plane that maps the origin to the center of \( c \). Tilt \( \pi \) by some angle \( \theta \), which is at most \( \theta_0 \), around its intersection line with \( \pi_c \) until it coincides with \( \pi_c \); this also makes the image of \( c_0 \) coincide with \( c \), and that of \( T_{c_0} \) coincide with \( T_c \). Let \( S_c \) be the sector of \( T_c \) that is the corresponding image of \( S_{c_0} \). Let \( \gamma_c, s_c, Q_c \) be the corresponding arc, segment, and cylinder, respectively.

Our approximation goal now is to report all pairs \( (p, c) \), such that \( p \) lies in \( Q_c \), and do so for every \( c \in C \) and for every sector \( S_{c_0}^{(j)} \) of \( S_{c_0} \). By construction, every pair \( (p, c) \) satisfying \( \text{dist}(p, c) \leq \varepsilon \) (such that \( c \) is in our current subset) is such that \( p \) lies in at least one of the \( O(1/\sqrt{\varepsilon}) \) cylinders \( Q_c \) of \( c \), and will therefore be reported. We perform this task for every sector index \( j \). (As is easily checked, \( p \) lies in at most two cylinders \( Q_c \), for the same circle \( c \), so \( (p, c) \) will be reported at most twice.)

This new problem is reminiscent of the problem of reporting all near point-line pairs in three dimensions, as presented in Section \ref{sec:main}, with a major difference that instead of lines we have segments (namely, the bounded axes of the cylinders \( Q_c \)). Our next step further reduces our problem to the point-line scenario.

Let \( \hat{s}_c \) denote the projection of \( s_c \) onto the \( xy \)-plane. We claim that \( |s_{c_0}| \cos \theta_0 \leq |\hat{s}_c| \leq |s_{c_0}| \), and that the angle \( \beta \) between \( \hat{s}_c \) and \( s_{c_0} \) is small (satisfying \( \beta = O(\theta_0^2) \)). Indeed (refer to Figure \ref{fig:main}), assume without loss of generality that \( s_{c_0} = AB \) is such that its endpoint \( A \) lies on \( \ell = \pi \cap \pi_c \). The rotation of \( \pi \) by angle \( \theta \leq \theta_0 \) around \( \ell \) brings \( B \) to a point \( B' \in \pi_c \), and \( s_{c_0} \) to \( s_c = AB' \). Let \( D \) be the projection of \( B' \) on \( \pi \), and let \( C \) be the point on \( \ell \) nearest to \( D \). The projection \( \hat{s}_c \) of \( s_c \) onto \( \pi \) is \( AD \), and the projection of \( CB' \) onto \( \pi \) is \( CD \). As is easily verified, \( D \) lies on the segment \( CB \), \( \angle B'CD = \theta \), and \( |CB'| = |CB| \). Since \( |AB'| = |AB| \), we have \( |\hat{s}_c| \leq |s_{c_0}| \). On the other hand, \( |AB'| \geq |CB'| \), so \( \delta = \angle B'AD \leq \theta \). Hence, \( |\hat{s}_c| = |AB'| \cos \delta \geq |s_{c_0}| \cos \theta_0 \). To estimate \( \beta \), we have \( |BD| = |BC| - |B'C| \cos \theta = |BC| (1 - \cos \theta) = |BC| \cdot O(\theta_0^2) \). Hence, sin \( \beta = |BD| \sin(\angle ADB)/|AB| \leq |BD|/|CB| = O(\theta_0^2) \).

By our assumption that \( s_{c_0} \) is parallel to the \( x \)-axis, we obtain that each of the projections \( \hat{s}_c \), for \( c \in C \), is of length at most \( r\sqrt{\varepsilon} < \sqrt{\varepsilon} \), and is almost parallel to the \( x \)-axis, forming with it an angle which is \( O(\theta_0^2) \).

We partition \( \mathbb{R}^3 \) into vertical slabs that are orthogonal to the \( x \)-axis, and are of width
equal to $|s_{c_0}| \leq \sqrt{\varepsilon}$. Our bounds on the lengths of $\hat{s}_c$ and their angles with $s_{c_0}$ imply that the axis $a_c$ of each cylinder $Q_c$ crosses only $O(1)$ slabs. Furthermore, if we stretch $a_c$ by a factor of 3 about its center then the projection of this larger segment onto the $x$-direction completely covers each of the slabs that $a_c$ intersects, assuming that $\theta_0$ is sufficiently small.

We fix a slab $\sigma$, take the subset of the cylinders that cross $\sigma$, and replace each such cylinder $Q_c$ by the (entire) line $\ell_c$ containing its axis. See Figure 8. Then we apply the algorithm in Section 7 to the set of these lines and to the set of points of $P$ within $\sigma$, with an error parameter $\varepsilon' \leq 1.5\varepsilon$, equal to the common radius of all the cylinders $Q_c$. Let $Q_c^*$ be a cylinder obtained from $Q_c$ by stretching its axis about its center by a factor of 3 and increasing its radius to $8\sqrt{2}\varepsilon$ (see Theorem 7.6). By the discussion above, the output will contain all pairs $(p, \ell_c)$ such that $p \in Q_c$, and every output pair will satisfy $p \in Q_c^*$.

As is easily checked, each pair $(p, c)$ can be reported once for each sector $j$, as $p$ is
contained in only one slab. The same pair can be reported at most twice, in the subproblems associated with a pair of adjacent sectors, when \( p \) is contained in both corresponding (and slightly overlapping) cylinders \( Q^*_c \).

We apply the algorithm in Section 7 to every slab that contains at least one point and is crossed by at least one cylinder. For each slab \( \sigma \), let \( m_{\sigma} \) denote the number of points of \( P \) in \( \sigma \), and let \( n_{\sigma} \) denote the number of cylinders that cross \( \sigma \). As noted, we have \( \sum_{\sigma} m_{\sigma} \leq m \), and \( \sum_{\sigma} n_{\sigma} = O(n) \). The time required by the algorithm in Section 7, for a slab \( \sigma \), is

\[
O \left( m_{\sigma} \right. + n_{\sigma} + m^{1/3}_{\sigma} n^{2/3}_{\sigma} / \varepsilon^{2/3} + k_{\sigma}^{(u,j)} \big) ,
\]

where \( k_{\sigma}^{(u,j)} \) is the number of reported pairs for the subproblem associated with \( u, j \), and \( \sigma \). Summing over all slabs \( \sigma \) (with \( u,j \) still fixed), and using Hölder’s inequality, the total running time is

\[
O \left( m + n + m^{1/3} n^{2/3} / \varepsilon^{2/3} + k^{(u,j)} \right) ,
\]

where \( k^{(u,j)} \) is the overall number of reported pairs for the subproblem associated with \( u \) and \( j \). Summing over the \( O(1) \) values of \( u \), and the \( O(1/\sqrt{\varepsilon}) \) sector indices \( j \), and observing (as already noted) that a pair \( (p,c) \) is reported at most \( O(1) \) times, we get a total running time of

\[
O \left( m + n + \frac{m^{1/3} n^{2/3}}{\varepsilon^{7/6}} + k \right) ,
\]

where \( k \) is the number of (distinct) reported pairs, over all possible choices of all parameters.

**Correctness.** Similar to the previous cases, the correctness is established in the following lemma.

**Lemma 8.1** (a) Each pair \( (p,c) \in P \times C \) satisfying \( \text{dist}(p,c) \leq \varepsilon \) will be reported by the algorithm.

(b) Each pair \( (p,c) \in P \times C \) reported by the algorithm satisfies \( \text{dist}(p,c) \leq \alpha \varepsilon \), for a suitable absolute constant \( \alpha \).

**Proof.** (a) Let \( (p,c) \in P \times C \) be such that \( \text{dist}(p,c) \leq \varepsilon \). Let \( u \in U \) be the direction associated with \( c \) and consider the sectors and slabs associated with \( u \).

Let \( S_c \) be the sector of \( T_c \) that contains \( p \) (there exists at least one, and in general exactly one such sector). Clearly, the enclosing cylinder \( Q_c \) also contains \( p \). Let \( \sigma \) be the slab that contains \( p \) in the subproblem associated with the sector \( S_c \). Then \( Q_c \) too must cross \( \sigma \), and the correctness of the algorithm in Section 7 implies that \( (p,c) \) will be reported.

(b) Let \( (p,c) \in P \times C \) be a pair that we report, at some subproblem with the direction \( u \in U \), associated with \( c \), at some the sector \( S_c \), and at the corresponding slab \( \sigma \) that contains \( p \). Refer to Figure 8. By the discussion above, we have \( p \in Q^*_c \). The proof is completed by arguing that any point \( p \in Q^*_c \) is at distance at most \( \alpha \varepsilon \) from \( c \) for some absolute constant \( \alpha \). This can be done by estimating the distance of the furthest point in \( Q^*_c \) from the center of \( c \) using the Pythagorean theorem as in the proofs of Lemmas 4.1 and 6.1. \( \square \)

In summary, we obtain the following result.
**Theorem 8.2** Let \( P \) be a set of \( m \) points in the unit ball \( B \) of \( \mathbb{R}^3 \), let \( C \) be a set of \( n \) congruent circles in \( \mathbb{R}^3 \), of common radius \( r \leq 1/2 \), that cross \( B \), and let \( \varepsilon \ll r \) be a prescribed error parameter. We can report all pairs \((p,c) \in P \times C\) that satisfy \(\text{dist}(p,c) \leq \varepsilon\), in time
\[
O\left( m + n + \frac{m^{1/3}n^{2/3}}{\varepsilon^{7/6}} + k \right),
\]
where \( k \) is the number of (distinct) pairs that we report. Each pair satisfying \(\text{dist}(p,c) \leq \varepsilon\) will be reported, and each reported pair satisfies \(\text{dist}(p,c) \leq \alpha \varepsilon\), for some absolute constant \( \alpha \). Moreover, each pair is reported at most \( O(1) \) times.

9 Reporting all nearly congruent triangles

In this section we put to work the algorithms in Sections 6 and 8 (see also (e) and (g) in Section 1), to obtain an efficient solution of the first step in solving the approximate point pattern matching problem in \( \mathbb{R}^3 \) (see its review in the introduction), where we are given a sampled “reference” triangle \( \Delta abc \), for a triple of points \( a, b, c \) in the first set \( A \), and a prescribed error parameter \( \varepsilon > 0 \). Our goal is to report all triples \( p, q, o \) in the second set \( B \) that span a triangle “nearly congruent” to \( \Delta \); that is, triples that satisfy
\[
|pq| - |ab| \leq \varepsilon, \quad |po| - |ac| \leq \varepsilon, \quad \text{and} \quad |qo| - |bc| \leq \varepsilon. \tag{11}
\]

We require that all such triples are reported, but we also allow to report triples that satisfy \((11)\) with \( \alpha \varepsilon \) on the right-hand sides rather than \( \varepsilon \), for some fixed absolute constant \( \alpha \). Let \( ab \) be the longest edge of \( \Delta \). We require that \( \beta \leq |ab| \leq 1/2 \) for some fixed constant \( \beta \). We also require that the height \( h \) of \( \Delta \) from \( c \) (perpendicular to \( ab \)) is larger than some fixed constant \( s \). We assume that \( \beta, s \gg \varepsilon \). Our approximation guarantee \( \alpha \) increases as \( \beta \) and \( s \) decrease.

We first report all pairs \((p,q) \in B^2 \) such that \( |pq| - |ab| \leq \varepsilon \), using the algorithm in Section 6 which involves incidences between congruent spheres and points). This takes \( O(n/\varepsilon + N) \) time, where \( N \) is the number of pairs that we report. Let \( \Pi \) denote the set of reported pairs. We know that all the desired pairs are included in \( \Pi \), and that every pair \((p,q) \in \Pi \) satisfies \( |pq| - |ab| \leq \alpha \varepsilon \), for some absolute constant \( \alpha ' \). We prune \( \Pi \), leaving in it only pairs \((p,q) \) satisfying \( |pq| - |ab| \leq \varepsilon \). We continue to denote the resulting set as \( \Pi \), and its size by \( N \).

Let \((p,q)\) be a pair in \( \Pi \). Any point \( o \) that satisfies \( |po| - |ac| \leq \varepsilon \) and \( |qo| - |bc| \leq \varepsilon \) lies in the intersection \( K = K_{p,q} \) of two spherical shells, one centered at \( p \) with radii \( |ac| \pm \varepsilon \), and one centered at \( q \) with radii \( |bc| \pm \varepsilon \). The following lemma allows us to replace \( K \) by a torus that is congruent to a fixed torus that depends only on \( \Delta \). See Figure 9.

**Lemma 9.1** Assume that \( \Delta \) is sufficiently fat, in the sense that \( \beta \leq |ab| \leq 1/2 \) and \( h \geq s \), for some absolute positive constants \( \beta, s \) that satisfy \( \varepsilon \ll \beta, s \). Then there exists a circle \( \gamma_{p,q} \) of radius \( h \) such that \( K \) is contained in the torus \( T_{p,q} \) that is the Minkowski sum of \( \gamma_{p,q} \) and a ball of radius \( \varepsilon' \leq \delta \varepsilon \) around the origin, where the constant \( \delta \) depends on \( \beta \) and \( s \).

**Proof.** Denote the lengths of the edges of the triangle \( \Delta abc \) by \( u = |ab|, v = |ac| \) and \( w = |bc| \). Let \( g \) the point where \( h \) meets \( ab \) and let \( z = |ag| \). We have \( z^2 + h^2 = v^2 \). 

31
and \((u - z)^2 + h^2 = w^2\), from which we obtain that \(z = \frac{u^2 + v^2 - w^2}{2u}\), and we denote this expression as \(z = z(u, v, w)\). Consider an alignment of \(\Delta\) within the plane of \(\Delta pqo\), such that \(a\) coincides with \(p\) and \(ab\) overlaps \(pq\). Let \(g\) now be a point on \(pq\) at distance \(z\) from \(p = a\). Then \(c\) lies on the circle \(\gamma_{p,q}\) of radius \(h\), centered at \(g\), and contained in the plane perpendicular to \(pq\) through \(g\). See Figure 9.

Fix some point \(o \in K\). We claim that \(o\) must be at distance \(\leq \delta \varepsilon\) from \(\gamma_{p,q}\), for some fixed constant \(\delta\) that depends on \(\beta\) and \(s\). Indeed, since \((p, q) \in \Pi\) and \(o \in K\), we can write \(|pq| = u + \varepsilon_1|, |po| = v + \varepsilon_2|, \text{ and } |go| = w + \varepsilon_3|, \text{ where } |\varepsilon_i| \leq \varepsilon\) for \(i = 1, 2, 3\).

Consider the alignment of \(\Delta\) with \(\Delta pqo\), as above, and imagine that we perturb the edges \(ab, ac, \) and \(bc\) of \(\Delta\) by \(\varepsilon_1, \varepsilon_2, \) and \(\varepsilon_3, \) respectively, so that \(\Delta\) is continuously deformed into \(\Delta pqo\). We claim that \(o\) cannot move too far as a result of this deformation so the distance between \(o\) and \(c\) must be small.

To see this, let \(h'\) be the height of \(\Delta pqo\) from \(o\), let \(g'\) be the point at which \(h'\) meets \(pq\), and let \(z' = |pq'|\). We claim that \(|z' - z| \leq \delta \varepsilon\) and \(|h' - h| \leq \delta \varepsilon\) for some absolute constant \(\delta\). To see this, using the function \(z = z(u, v, w)\) defined above, we have \(z' = z(u + \varepsilon_1, v + \varepsilon_2, w + \varepsilon_3)\), and routine calculations show that, for \(\varepsilon\) sufficiently small, we have \(|z' - z| = O(|\nabla z(u, v, w) \cdot (\varepsilon_1, \varepsilon_2, \varepsilon_3)|) \leq \delta' \varepsilon\), where \(\delta'\) depends on \(\beta\).

Similarly, by Heron’s formula, we can think of \(h\) as a function \(h(u, v, w)\), given by

\[
    h(u, v, w) = \frac{2 \text{Area}(\Delta)}{u} = \frac{2\sqrt{\tau(u - \tau)(v - \tau)(w - \tau)}}{u},
\]

where \(\tau = \frac{1}{2}(u + v + w)\). Then \(h' = h(u + \varepsilon_1, v + \varepsilon_2, w + \varepsilon_3)\), and, by another routine calculation, \(|h' - h| = O(|\nabla h(u, v, w) \cdot (\varepsilon_1, \varepsilon_2, \varepsilon_3)|) \leq \delta'' \varepsilon\), for another constant \(\delta''\) that depends on \(\beta\) and \(s\). (Simple calculations show that \(\nabla h\) becomes smaller as \(s\) increases.) Take \(\delta = \sqrt{\delta'(\delta' + \delta'')^2}\), and the lemma follows. □

We have thus reached the following scenario. We have a set \(\mathcal{T}\) of \(N\) congruent tori \(T_{p,q}\), for \((p, q) \in \Pi\), and a set \(B\) (the original one) of \(n\) points. By construction, each triple \((p, q, o)\) that defines a triangle for which \(\square\) holds, satisfies \(o \in T_{p,q}\). Using our algorithm for point-circle near neighbors in \(\mathbb{R}^3\), as reviewed in Section 8, we can report all the triples \((p, q, o)\) such that \(o \in T_{p,q}\), in time \(O\left(n + N/\varepsilon^{1/2} + n^{1/3}N^{2/3}/\varepsilon^{7/6} + k\right)\), where \(k\) is the
number of (distinct) triples that we report; each of the desired triples is reported, and each
triple that we report is such that the distance from $o$ to $\gamma_{p,q}$ is at most $\alpha \varepsilon$
for some other fixed constant $\alpha > \delta$. Therefore each triple which we report satisfies (11)
with $\alpha \varepsilon$ on the right-hand sides, rather than $\varepsilon$. In summary, we have:

**Theorem 9.2** Let $B$ be a set of $n$ points in the unit ball in $\mathbb{R}^3$. Let $\Delta_{abc}$ be a fixed
reference triangle and let $\varepsilon$ an error parameter, so that $\Delta$ and $\varepsilon$ satisfy the constraints
specified in Lemma 9.4. We can then report all triples $(p, q, o) \in B^3$ that span a triangle
nearly congruent to $\Delta$, in the sense of (11), in time $\left(n + N/\varepsilon^{1/2} + n^{1/3}N^{2/3}/\varepsilon^{7/6} + k\right)$,
where $N$ is the number of pairs reported by our algorithm for approximate congruent pairs
in $\mathbb{R}^3$ (presented in Section 6), applied to $P$ with distance $|ab|$, the largest edge length of $\Delta$,
and $k$ is the number of (distinct) triples that the algorithm in this section reports; each of
the desired triples is reported, and each triple that we report satisfies (11) with $\alpha \varepsilon$ replacing
$\varepsilon$, where $\alpha$ is a suitable absolute constant. Each pair is reported at most $O(1)$ times.

### 10 Implementation and experiments

To test the effectiveness of the methodology proposed in this paper, we implemented the
algorithm of Section 2, for incidences of points and lines in the plane, and tested it on real
and random data. We compared its performance to three other approaches that are used in
practice, and were mentioned in the introduction. Specifically we compared the algorithms:

**Naive**: Based on constructing a grid of cell size $\varepsilon$ only in the primal plane. Its running time
is $O(m + n/\varepsilon + k)$.

**Naive-duality**: Use the naive approach when $m > n$. Otherwise apply the naive solution in
the dual plane. The running time is $O(m + n + \min(m, n)/\varepsilon + k)$.

**Large-n (the dense case)**: The alternative solution of Aiger and Kedem [3]. Its running time
is $O(m + n + \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} + k)$.

**Efficient-duality** (Efficient for short in the plots): This is our solution, with running time
$O(m + n + \sqrt{nm}/\sqrt{\varepsilon} + k)$.

The output size $k$, which appears in the four time bounds listed above, is not a fixed
quantity, because it depends on the specific algorithm being used. More precisely, for a
fixed input instance, denote by $k_{\text{true}}$ the real output size, which is the number of pairs at
distance at most $\varepsilon$ apart. Each algorithm encounters its own superset of these pairs, and
its running time degrades linearly with the size of this superset.

As a matter of fact, our Efficient-duality algorithm tends to have a larger value of $k$, becausel each of its primal and dual steps makes some worst-case assumptions that affect
the size of the grid cells that are used, allowing more pairs to be reported. The Efficient-
duality algorithm might report pairs at distance up to $5\sqrt{2}\varepsilon$ (see Section 2), whereas each pair reported by the Naive implementation is only at distance at most $2\sqrt{2}\varepsilon$, as is easily
checked.

Our random data set consisted of $n$ points drawn uniformly at random in the unit square
and $n$ random lines crossing that square, for various values of $n$. For this data the value
of $k$ tends to increase quadratically in the respective factor $5\sqrt{2}\varepsilon$, $2\sqrt{2}\varepsilon$, and the difference could become significant when $\varepsilon$ is large.

Our real data set was extracted from the image depicted in Figure 10(a). That is, we have applied a standard edge detection procedure to this image, resulting in the edges depicted in Figure 10(b), from which we have sampled our points. The lines that we use were obtained by sampling pairs of these points, in the hope that some of the sampled lines will be very close to the actual edges, and will be detected as such by the approximate incidence reporting algorithms. In other words, the experiments that we have conducted on this data were made with the application of robust model fitting in mind; see later in this section.

Figure 10: The input with real data: (a) The image. (b) The detected edges, from which we sample our input points.

In real data, if we use an algorithm that allows pairs at distance up to $\alpha \varepsilon$ to be reported, we expect that the number $k$ of reported (i.e., inspected) pairs will grow only linearly in $\alpha$.

Our results are as follows.

**Random points.** Figures 11 and 12 show the runtime of the three algorithms Naïve, Efficient-duality, and Large-n for various values of $n$ and $\varepsilon$ (since the number of points is the same as the number of lines, there is no need to consider Naïve-duality). Each of the three subfigures (a)–(c), in both figures, is for a different choice of $\varepsilon$, which are, respectively, 0.001, 0.002 and 0.004. The executions reported in Figure 11 only count the number of output (that is, inspected) pairs, essentially making the running time independent of the corresponding value of $k$. In contrast, the executions reported in Figure 12 include the cost of reporting the output pairs, so their running time also depends on $k$.

As can be seen, Efficient-duality always performs considerably better than Naïve, where the difference is substantial for a wide range of $n$ and $\varepsilon$. The difference is less significant when $\varepsilon$ increases (also in the counting versions), but Efficient-duality still outperforms Naïve. Even the quadratic growth of $k$ in the reporting version still leaves our algorithm superior, for the (fairly wide) ranges of $n$ and $\varepsilon$ depicted in the figures. The implementation of the Large-n algorithm is more complex, resulting in a large constant of proportionality in the overhead, which makes it efficient only for very large values of $n$ (for practical values of $\varepsilon$).

While serving as a useful testbed for comparing the algorithms, the random case is not very practical. Moreover, as can be seen in Figure 12, the cost of handling the $k$ output
pairs (collecting, inspecting and outputting) tends to become rather large for larger values of $\varepsilon$, and dominates the runtime.

![Graphs](image)

Figure 11: The runtime of the counting versions of Naive, Efficient-duality, and Large-$n$ vs. the number of points (and lines), for different values of $\varepsilon$: (a): $\varepsilon = 0.001$; (b): $\varepsilon = 0.002$; (c) $\varepsilon = 0.004$.

**Real data.** Our first experiment still used an equal number of points and lines. The results are shown in Figure 13. Our algorithm wins with a substantial margin. Note the rather minor difference between the time for counting and the time for reporting (because a relatively small number of pairs is reported here).

Figure 14 shows the actual number of pairs reported by the Naive and Efficient-duality algorithms, as well as the number of true pairs (those at distance at most $\varepsilon$) for various values of $\varepsilon$. Part (a) shows the actual number of pairs, and part (b) shows the ratio between the numbers of reported and true pairs. As these figures show, (i) the larger $\alpha$ in the Efficient-duality algorithm does indeed causes it to produce more pairs than the naive one; (ii) these numbers grow linearly with $\varepsilon$, as expected; (iii) in fact, the ratio between reported and true pairs is more or less a constant (1.8 for Naive and about 5 for Efficient-duality). Still, in spite of this discrepancy (in favor of Naive), the moderate growth of $k$, combined with the much faster overhead, makes our algorithm a clear winner in these experiments.

**RANSAC line fitting with our method.** We ran a complete RANSAC line fitting algorithm where we used the Naive, Naive-duality, and Efficient-duality methods to count and report the nearby points for each candidate line. The input consists of points sampled

35
Figure 12: The runtime of the reporting versions of Naive, Efficient-duality, and Large-n vs. the number of points (and lines), for different values of $\varepsilon$: (a) $\varepsilon = 0.001$; (b) $\varepsilon = 0.002$; (c) $\varepsilon = 0.004$.

Figure 13: Image points and lines incidences: (a) Counting (b) Reporting.

from the input image in Figure 10. In all experiments we sampled 19955 points, and we randomly generated increasing numbers of lines by sampling pairs from these points (here the number of lines was not equal to the number of points). Duality allows us to exploit the fact that $n$ and $m$ are different, in both the Naive-duality and Efficient-duality methods. The asymptotic theoretical bounds for these two techniques (see the beginning of the section) show that, for a sufficiently large number of lines, namely, larger than $1/\varepsilon$.
times the number of points, Naive-duality will become superior to Efficient-duality. Figure 13 indicates this trend, but shows that the number of lines needed for this to happen (in this example) has indeed to be very large.

Figure 14: True pairs and reported pairs by the Naive and Efficient-duality algorithms. (a) The actual number of pairs. (b) The ratio between the number of reported and true pairs.

Figure 15: Line fitting: (a) Counting time. (b) Reporting time. (c) Fitted lines out of 40000 sampled candidate lines.
References

[1] P. Agarwal and M. Sharir. The number of congruent simplices in a point set. *Discrete Comput. Geom.*, 28(2):123–150, 2002.

[2] P. K. Agarwal and J. Erickson. Geometric range searching and its relatives. In Contemp. Math. 223 (B. Chazelle J. E. Goodman and R. Pollack eds.), editors, *Advances in Discrete and Computational Geometry*, pages 1–56. Amer. Math. Soc. Press, Providence, RI, 1999.

[3] D. Aiger and K. Kedem. Approximate input sensitive algorithms for point pattern matching. *Pattern Recognition*, 43(1):153–159, 2010.

[4] D. Aiger, N. J. Mitra, and D. Cohen-Or. 4-points congruent sets for robust pairwise surface registration. *ACM Trans. Graphics*, 27(3):Article 85, 2008.

[5] P. Brass. Combinatorial geometry problems in pattern recognition. *Discrete Comput. Geom.*, 28(4):495–510, 2002.

[6] O. Chum and J. Matas. Optimal randomized RANSAC. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(8):1472–1482, 2008.

[7] H. S. M. Coxeter. *Introduction to Geometry*. 2nd ed. Wiley Interscience, New York, 1969.

[8] *IEEE Intl Workshop 25 Years of RANSAC in conjunction with CVPR 06 (RANSAC25 06)*. IEEE Computer Society, 2006.

[9] G. D. da Fonseca and D. M. Mount. Approximate range searching: The absolute model. *Comput. Geom. Theory Appl.*, 43(4):434–444, 2010.

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1982.

[11] M. Gavrilov, P. Indyk, R. Motwani, and S. Venkatasubramanian. Combinatorial and experimental methods for approximate point pattern matching. *Algorithmica*, 38(1):59–90, 2003.

[12] P. J. Heffernan and S. Schirra. Approximate decision algorithms for point set congruence. *Comput. Geom. Theory Appl.*, 1(4):137–156, 1994.

[13] D. P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *Internat. J. Computer Vision*, 5(2):195–212, 1990.

[14] P. Indyk, R. Motwani, and S. Venkatasubramanian. Geometric matching under noise: Combinatorial bounds and algorithms. *Proc. 10th ACM–SIAM Sympos. Discrete Algorithms*, pages 457–465, 1994.

[15] N. Mellado, D. Aiger, and N. J. Mitra. Super 4pcs: fast global pointcloud registration via smart indexing. *Comput. Graphics Forum*, 33(5):205–215, 2014.
[16] J. Pach and M. Sharir. Geometric incidences. In ed.) Contemp. Math. 342 (J. Pach, editor, *Towards a Theory of Geometric Graphs*, pages 185–223. Amer. Math. Soc. Press, Providence, RI, 2004.