Near-Optimal ε-Kernel Construction and Related Problems

Sunil Arya* Guilherme D. da Fonseca† David M. Mount‡

Abstract

The computation of (i) ε-kernels, (ii) approximate diameter, and (iii) approximate bichromatic closest pair are fundamental problems in geometric approximation. In each case the input is a set of points in $\mathbb{R}^d$ for a constant dimension d and an approximation parameter $\varepsilon > 0$. In this paper, we describe new algorithms for these problems, achieving significant improvements to the exponent of the ε-dependency in their running times, from roughly d to $d/2$ for the first two problems and from roughly $d/3$ to $d/4$ for problem (iii).

These results are all based on an efficient decomposition of a convex body using a hierarchy of Macbeath regions, and contrast to previous solutions that decompose space using quadtrees and grids. By further application of these techniques, we also show that it is possible to obtain near-optimal preprocessing time for the most efficient data structures to approximately answer queries for (iv) nearest-neighbor searching, (v) directional width, and (vi) polytope membership.

1 Introduction

In this paper we present new faster algorithms to several fundamental geometric approximation problems involving point sets in d-dimensional space. In particular, we present approximation algorithms for ε-kernels, diameter, and bichromatic closest pair. Our results arise from a recently developed shape-sensitive approach to approximating convex bodies, which is based on the classical concept of Macbeath regions. This approach has been applied to computing area-sensitive bounds for polytope approximation [6], polytope approximations with low combinatorial complexity [7], answering approximate polytope-membership queries [8], and approximate nearest-neighbor searching [8]. The results of [8] demonstrated the existence of data structures for these query problems but did not discuss preprocessing in detail. We complete the story by presenting efficient algorithms for building data structures for three related queries: approximate polytope membership, approximate directional width, and approximate nearest-neighbors.

Throughout, we assume that the dimension d is a constant. Our running times will often involve expressions of the form $1/\varepsilon^{\alpha}$. In such cases, $\alpha > 0$ is constant that can be made arbitrarily small. The approximation parameter ε is treated as an asymptotic variable that approaches 0.

We have learned recently of independent results by Timothy Chan for many of the above problems in which the complexity bounds are very similar to ours [17]. Remarkably, the computational techniques seem to be very different, based on Chebyshev polynomials.

The results presented here are based on the upcoming paper on the 33rd International Symposium on Computational Geometry (SoCG 2017).

2 Static Results

Kernel. Given a set S of n points in $\mathbb{R}^d$ and an approximation parameter $\varepsilon > 0$, an *ε-coreset* is an (ideally small) subset of S that approximates some measure over S (see [2] for a survey). Given a nonzero vector $v \in \mathbb{R}^d$, the *directional width* of S in direction v, $\text{width}_v(S)$ is the minimum distance between two hyperplanes that enclose S and are orthogonal to v. A *coreset for the directional width* (also known as an *ε-kernel* and as a *coreset for the extent measure*) is a subset $Q \subseteq S$ such that $\text{width}_v(Q) \geq (1-\varepsilon)\,\text{width}_v(S)$, for all $v \in \mathbb{R}^d$. Kernels are among the most fundamental constructions in geometric approximation, playing a role similar to that of convex hulls in exact computations. Kernels have been used to obtain approximation algorithms to several problems such as diameter, minimum width, convex hull volume, minimum enclosing cylinder, minimum enclosing annulus, and minimum-width cylindrical shell [1, 2].

The concept of ε-kernels was introduced by Agarwal et al. [1]. The existence of ε-kernels with $O(1/\varepsilon^{(d-1)/2})$ points is implied in the works of Dudley [18] and Bronshteyn and Ivanov [15], and this is known to be optimal in the worst case. Agarwal et al. [1] demonstrated how to compute such a kernel in $O(n + 1/\varepsilon^{3(d-1)/2})$ time, which reduces to $O(n)$ when $n = \Omega(1/\varepsilon^{3(d-1)/2})$. While less succinct ε-kernels with

*Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, arya@cse.ust.hk. Research supported by the Research Grants Council of Hong Kong, China under project number 610012.

†Université d'Auvergne and LIMOS, Clermont-Ferrand, France, fonseca@isima.fr.

‡Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, mount@cs.umd.edu. Research supported by NSF grant CCF–1618866.

$O(1/\varepsilon^{d-1})$ points can be constructed in time $O(n)$ time for all n [1, 13], no linear-time algorithm is known to build an ε-kernel of optimal size. Hereafter, we use the term ε-kernel to refer exclusively to an ε-kernel of size $O(1/\varepsilon^{(d-1)/2})$.

Chan [16] showed that an ε-kernel can be constructed in $O((n+1/\varepsilon^{d-2})\log\frac{1}{\varepsilon})$ time, which is nearly linear when $n = \Omega(1/\varepsilon^{d-2})$. He posed the open problem of obtaining a faster algorithm. A decade later, Arya and Chan [4] showed how to build an ε-kernel in roughly $O(n+\sqrt{n}/\varepsilon^{d/2})$ time using discrete Voronoi diagrams. In this paper, we attain the following near-optimal construction time.

Theorem 1 *Give n points in $\mathbb{R}^d$ and an approximation parameter $\varepsilon > 0$, it is possible to construct an ε-kernel of S with $O(1/\varepsilon^{(d-1)/2})$ points in $O(n\log\frac{1}{\varepsilon} + 1/\varepsilon^{(d-1)/2+\alpha})$ time.*

We note that when $n = o(1/\varepsilon^{(d-1)/2})$, the input S is already an ε-kernel and therefore an $O(n)$ time algorithm is trivial. Because the worst-case output size is $O(1/\varepsilon^{(d-1)/2})$, we may assume that n is at least this large, for otherwise we can simply take S itself to be the kernel. Since $1/\varepsilon^{\alpha}$ dominates $\log\frac{1}{\varepsilon}$, the above running time can be expressed as $O(n/\varepsilon^{\alpha})$, which is nearly linear given that α is arbitrarily small.

Diameter. An important application of ε-kernels is to approximate the diameter of a point set. Given n data points, the *diameter* is defined to be the maximum distance between any two data points. An *ε-approximation* of the diameter is a pair of points whose distance is at least $(1-\varepsilon)$ times the exact diameter. There are multiple algorithms to approximate the diameter [1, 3, 4, 12, 16]. The fastest running times are $O((n+1/\varepsilon^{d-2})\log\frac{1}{\varepsilon})$ [16] and roughly $O(n+\sqrt{n}/\varepsilon^{d/2})$ [4]. The algorithm from [16] essentially computes an ε-kernel Q and then determines the maximum value of $\text{width}_v(Q)$ among a set of $k = O(1/\varepsilon^{(d-1)/2})$ directions v by brute force [1]. Discrete Voronoi diagrams [4] permit this computation in roughly $O(n+\sqrt{n}/\varepsilon^{d/2})$ time. Therefore, combining the kernel construction of Theorem 1 with discrete Voronoi diagrams [4], we reduce n to $O(1/\varepsilon^{(d-1)/2})$ and obtain an algorithm to ε-approximate the diameter in roughly $O(n+1/\varepsilon^{3d/4})$ time. However, we show that it is possible to obtain a much faster algorithm, as presented in the following theorem.

Theorem 2 *Given n points in $\mathbb{R}^d$ and an approximation parameter $\varepsilon > 0$, it is possible to compute an ε-approximation to the diameter of S in $O(n\log\frac{1}{\varepsilon} + 1/\varepsilon^{(d-1)/2+\alpha})$ time.*

Bichromatic Closest Pair. In the *bichromatic closest pair* (BCP) problem, we are given n points from two sets, designated red and blue, and we want to find the closest red-blue pair. In the *ε-approximate* version, the goal is to find a red-blue pair of points whose distance is at most $(1+\varepsilon)$ times the exact BCP distance. Approximations to the BCP problem were introduced in [19], and the most efficient randomized approximation algorithm runs in roughly $O(n/\varepsilon^{d/3})$ expected time [4]. We present the following result.

Theorem 3 *Given n red and blue points in $\mathbb{R}^d$ and an approximation parameter $\varepsilon > 0$, there is a randomized algorithm that computes an ε-approximation to the BCP in $O(n/\varepsilon^{d/4+\alpha})$ expected time.*

3 Data Structure Results

Polytope membership. Let P denote a convex polytope in $\mathbb{R}^d$, represented as the bounded intersection of n halfspaces. The *polytope membership problem* consists of preprocessing P so that it is possible to determine efficiently whether a given query point $q \in \mathbb{R}^d$ lies within P. In the ε-approximate version, we consider an expanded convex body $K \supset P$. A natural way to define this expansion would be to consider the set of points that lie within distance $\varepsilon\cdot\text{diam}(P)$ of P, thus defining a body whose Hausdorff distance from P is $\varepsilon\cdot\text{diam}(P)$. However, this definition has the shortcoming that it is not sensitive to the directional width of P. Instead, we define K as follows. For any nonzero vector $v \in \mathbb{R}^d$, consider the two supporting hyperplanes for P that are normal to v. Translate each of these hyperplanes outward by a distance of $\varepsilon\cdot\text{width}_v(P)$, and consider the closed slab-like region lying between them. Define K to be the intersection of this (infinite) set of slabs. This is clearly a stronger approximation than the Hausdorff-based definition. An *ε-approximate polytope membership query* (ε-APM query) returns a positive result if the query point q is inside P, a negative result if q is outside K, and may return either result otherwise.[1]

We recently proposed an optimal data structure to answer approximate polytope membership queries, but efficient preprocessing remained an open problem [8]. In this paper, we present a similar data structure that not only attains optimal storage and query time, but can also be preprocessed in near-optimal time.

Theorem 4 *Given a convex polytope P in $\mathbb{R}^d$ represented as the intersection of n halfspaces and an approximation parameter $\varepsilon > 0$, there is a data structure that can answer ε-APM queries with query time $O(\log\frac{1}{\varepsilon})$, space $O(1/\varepsilon^{(d-1)/2})$, and preprocessing time $O(n\log\frac{1}{\varepsilon} + 1/\varepsilon^{(d-1)/2+\alpha})$.*

[1] Our earlier works on ε-APM queries [5, 8] use the weaker Hausdorff form to define the problem, but the solutions presented there actually achieve the stronger direction-sensitive form.

Directional width. Applying the previous data structure in the dual space, we obtain a data structure for the following *ε-approximate directional width* problem, which is closely related to ε-kernels. Given a set S of n points in $\mathbb{R}^d$ and an approximation parameter $\varepsilon > 0$, the goal is to preprocess S to efficiently ε-approximate $\text{width}_v(S)$, for a nonzero query vector v. We present the following result.

Theorem 5 *Given n points in $\mathbb{R}^d$ and an approximation parameter $\varepsilon > 0$, there is a data structure that can answer ε-approximate directional width queries with query time $O(\log^2 \frac{1}{\varepsilon})$, space $O(1/\varepsilon^{(d-1)/2})$, and preprocessing time $O(n \log \frac{1}{\varepsilon} + 1/\varepsilon^{(d-1)/2+\alpha})$.*

Nearest Neighbor. Let S be a set of n points in $\mathbb{R}^d$. Given any $q \in \mathbb{R}^d$, an *ε-approximate nearest neighbor* (ANN) of q is any point of S whose distance from q is at most $(1+\varepsilon)$ times the distance to q's closest point in S. The objective is to preprocess S in order to answer such queries efficiently. Data structures for approximate nearest neighbor searching (in fixed dimensions) have been proposed by several authors, offering space-time tradeoffs (see [8] for an overview of the tradeoffs). Applying the reduction from approximate nearest neighbor to approximate polytope membership established in [5] together with Theorem 4, we obtain the following result, which matches the best bound [8] up to an $O(\log \frac{1}{\varepsilon})$ factor in the query time, but offers faster preprocessing time.

Theorem 6 *Given n points in $\mathbb{R}^d$, an approximation parameter $\varepsilon > 0$, and m such that $\log \frac{1}{\varepsilon} \leq m \leq 1/(\varepsilon^{d/2} \log \frac{1}{\varepsilon})$, there is a data structure that can answer ε-ANN queries with query time $O(\log n + (\log \frac{1}{\varepsilon})/(m \cdot \varepsilon^{d/2}))$ space $O(n\,m)$, and preprocessing time $O(n \log n \log \frac{1}{\varepsilon} + n\,m/\varepsilon^{\alpha})$.*

4 Techniques

In contrast to previous kernel constructions, which are based on grids and the execution of Bronshteyn and Ivanov's algorithm, our construction employs a classical structure from the theory of convexity, called *Macbeath regions* [20]. Macbeath regions have found numerous uses in the theory of convex sets and the geometry of numbers (see Bárány [11] for an excellent survey). They have also been applied to several problems in the field of computational geometry. However, most previous results were either in the form of lower bounds [9, 10, 14] or focused on existential results [6, 7, 21].

In [8] the authors introduced a data structure based on a hierarchy of ellipsoids based on Macbeath regions to answer approximate polytope membership queries, but the efficient computation of the hierarchy was not considered. In this paper, we show how to efficiently construct the Macbeath regions that form the basis of this hierarchy.

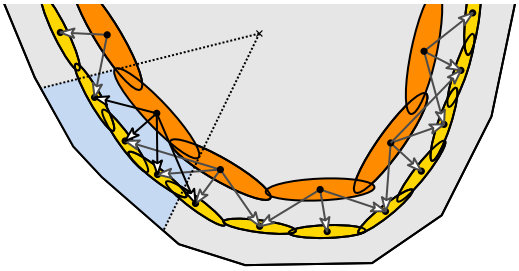

Figure 1: Two levels of the hierarchy of ellipsoids based on Macbeath regions.

Let P denote a convex polytope in $\mathbb{R}^d$. Each level i in the hierarchy corresponds to a δ_i-approximation of the boundary of P by a set of $O(1/\delta_i^{(d-1)/2})$ ellipsoids, where $\delta_i = \Theta(1/2^i)$. Each ellipsoid has $O(1)$ children, which correspond to the ellipsoids of the following level that approximate the same portion of the boundary (see Figure 1). The hierarchy starts with $\delta_0 = \Theta(1)$ and stops after $O(\log \frac{1}{\delta})$ levels when $\delta_i = \delta$, for a desired approximation δ. We present a simple algorithm to construct the hierarchy in $O(n + 1/\delta^{3(d-1)/2})$ time. The polytope P can be presented as either the intersection of n halfspaces or the convex hull of n points.

Our algorithm to compute an ε-kernel in time $O(n \log \frac{1}{\varepsilon} + 1/\varepsilon^{(d-1)/2+\alpha})$ (Theorem 1) is conceptually quite simple. Since the time to build the ε-approximation hierarchy for the convex hull is prohibitively high, we use an approximation parameter $\delta = \varepsilon^{1/3}$ to build a δ-approximation hierarchy in $O(n + 1/\varepsilon^{(d-1)/2})$ time. By navigating through this hierarchy, we partition the n points among the leaf Macbeath ellipsoids in $O(n \log \frac{1}{\varepsilon})$ time, discarding points that are too far from the boundary. We then compute an (ε/δ)-kernel for the set of points in each leaf ellipsoid and return the union of the kernels computed.

Given an algorithm to compute an ε-kernel in $O(n \log \frac{1}{\varepsilon} + 1/\varepsilon^{t(d-1)})$ time, the previous procedure produces an ε-kernel in $O(n \log \frac{1}{\varepsilon} + 1/\varepsilon^{t'(d-1)})$ time where $t' = (4t+1)/6$. Bootstrapping the construction a constant number of times, the value of t goes down from 1 to a value that is arbitrarily close to $1/2$. This discrepancy accounts for the $O(1/\varepsilon^{\alpha})$ factors in our running times.

To prove Theorem 4, we use our kernel construction in the dual space to efficiently build a polytope membership data structure. The key idea is to compute multiple kernels in order to avoid examining the whole polytope. We build a δ-approximate hierarchy (for a proper value of δ) in $O(n + 1/\delta^{3(d-1)/2})$ time. Each leaf node of the data structure is associated with a certain portion of the polytope, called a *shadow*.

We then build an (ε/δ)-kernel (in the dual space) for the shadow of each leaf node, followed by an (ε/δ)-approximate polytope membership data structure for each kernel. Given a query point q, the δ-approximate hierarchy is able to either correctly answer the query (to the desired approximation $\varepsilon \leq \delta$) or to locate a leaf shadow that contains the query point. In the latter case, we transfer the query to the data structure associated with that leaf node. The aforementioned construction reduces the preprocessing time for an approximate polytope membership data structure. Again, we use bootstrapping to obtain a near-optimal preprocessing time.

The remaining theorems follow from Theorems 1 and 4, together with several known reductions.

References

[1] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Approximating extent measures of points. *J. Assoc. Comput. Mach.*, 51:606–635, 2004.

[2] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Geometric approximation via coresets. In J. E. Goodman, J. Pach, and E. Welzl, editors, *Combinatorial and Computational Geometry*. MSRI Publications, 2005.

[3] P. K. Agarwal, J. Matoušek, and S. Suri. Farthest neighbors, maximum spanning trees and related problems in higher dimensions. *Computational Geometry*, 1(4):189–201, 1992.

[4] S. Arya and T. M. Chan. Better ε-dependencies for offline approximate nearest neighbor search, Euclidean minimum spanning trees, and ε-kernels. In *Proc. 30th Annu. Sympos. Comput. Geom.*, pages 416–425, 2014.

[5] S. Arya, G. D. da Fonseca, and D. M. Mount. Approximate polytope membership queries. In *Proc. 43rd Annu. ACM Sympos. Theory Comput.*, pages 579–586, 2011. (Full version available from `http://arxiv.org/abs/1604.01183`, to appear on SIAM J. Computing).

[6] S. Arya, G. D. da Fonseca, and D. M. Mount. Optimal area-sensitive bounds for polytope approximation. In *Proc. 28th Annu. Sympos. Comput. Geom.*, pages 363–372, 2012.

[7] S. Arya, G. D. da Fonseca, and D. M. Mount. On the combinatorial complexity of approximating polytopes. In *Proc. 32nd Internat. Sympos. Comput. Geom.*, pages 11:1–11:15, 2016. (Full version `http://arxiv.org/abs/1604.01175`, to appear on Discrete Comput. Geom.).

[8] S. Arya, G. D. da Fonseca, and D. M. Mount. Optimal approximate polytope membership. In *Proc. 28th Annu. ACM-SIAM Sympos. Discrete Algorithms*, pages 270–288, 2017.

[9] S. Arya, T. Malamatos, and D. M. Mount. The effect of corners on the complexity of approximate range searching. *Discrete Comput. Geom.*, 41:398–443, 2009.

[10] S. Arya, D. M. Mount, and J. Xia. Tight lower bounds for halfspace range searching. *Discrete Comput. Geom.*, 47:711–730, 2012.

[11] I. Bárány. The technique of M-regions and cap-coverings: A survey. *Rend. Circ. Mat. Palermo*, 65:21–38, 2000.

[12] G. Barequet and S. Har-Peled. Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. *Journal of Algorithms*, 38(1):91–109, 2001.

[13] J. L. Bentley, M. G. Faust, and F. P. Preparata. Approximation algorithms for convex hulls. *Commun. ACM*, 25(1):64–68, 1982.

[14] H. Brönnimann, B. Chazelle, and J. Pach. How hard is halfspace range searching. *Discrete Comput. Geom.*, 10:143–155, 1993.

[15] E. M. Bronshteyn and L. D. Ivanov. The approximation of convex sets by polyhedra. *Siberian Math. J.*, 16:852–853, 1976.

[16] T. M. Chan. Faster core-set constructions and data-stream algorithms in fixed dimensions. *Comput. Geom. Theory Appl.*, 35(1):20–35, 2006.

[17] T. M. Chan. Applications of Chebyshev polynomials to low-dimensional computational geometry. In *Proc. 33rd Internat. Sympos. Comput. Geom.*, 2017.

[18] R. M. Dudley. Metric entropy of some classes of sets with differentiable boundaries. *J. Approx. Theory*, 10(3):227–236, 1974.

[19] S. Khuller and Y. Matias. A simple randomized sieve algorithm for the closest-pair problem. *Information and Computation*, 118(1):34–37, 1995.

[20] A. M. Macbeath. A theorem on non-homogeneous lattices. *Ann. of Math.*, 56:269–293, 1952.

[21] N. H. Mustafa and S. Ray. Near-optimal generalisations of a theorem of Macbeath. In *Proc. 31st Internat. Sympos. on Theoret. Aspects of Comp. Sci.*, pages 578–589, 2014.