Approximation Algorithms for k-Line Center *

Pankaj K. Agarwal[†] Cecilia M. Procopiuc[‡] Kasturi R. Varadarajan[§]

April 15, 2002

Abstract

Given a set P of n points in $\mathbb{R}^d$ and an integer $k \geq 1$, let w^* denote the minimum value so that P can be covered by k cylinders of width at most w^*. We describe an algorithm that, given P and an $\varepsilon > 0$, computes k cylinders of width at most $(1+\varepsilon)w^*$ that cover P. The running time of the algorithm is $O(n \log n)$, with the constant of proportionality depending on k, d, and ε. The running times of the fastest algorithms that compute w^* exactly are of the order of $n^{O(dk)}$. An approximation algorithm with near-linear dependence on n for $k > 1$ was only known for the planar 2-line center problem, i.e., the case $k = 2$, $d = 2$.

We believe that the techniques used in showing this result are quite useful in themselves. We first show that there exists a small "certificate" $Q \subseteq P$, whose size does not depend on n, such that for any k-cylinders that cover Q, an enlargement of these cylinders by a factor of $(1+\varepsilon)$ covers P. We only establish the existence of a small certificate and our proof does not give us an efficient way of constructing one. We then observe that a well-known scheme based on sampling and iterated re-weighting gives us an efficient algorithm for solving the problem. Only the existence of a small certificate is used to establish the correctness of the algorithm. This technique is quite general and can be used in other contexts as well.

1 Introduction

Problem statement and motivation. Given a set P of n points in $\mathbb{R}^d$, an integer $k \geq 1$, and a real $\varepsilon > 0$, we want to compute k cylinders of radius at most $(1+\varepsilon)w^*$ that cover P (that is, the union of the cylinders contains P), where w^* denotes the minimum value so that P can be covered by k cylinders of width at most w^*.

This problem is a special instance of projective clustering. In a more general formulation, a *projective clustering* problem can be defined as follows. Given a set S of n points in $\mathbb{R}^d$ and two integers $k < n$ and $q \leq d$, find k q-dimensional flats $h_1, \ldots, h_k$ and partition S into

*Research by the first author is supported by NSF under grants CCR-00-86013 EIA-98-70724, EIA-01-31905, and CCR-97-32787, and by a grant from the U.S.–Israel Binational Science Foundation.

† Department of Computer Science, Duke University, pankaj@cs.duke.edu

‡ AT&T Shannon Labs, magda@research.att.com

§ Department of Computer Science, University of Iowa, kvaradar@cs.uiowa.edu

k subsets $S_1, \ldots, S_k$ so that $\max_{1 \le i \le k} \max_{p \in Si} d(p, h_i)$ is minimized. That is, we partition S into k clusters and each cluster S_i is projected onto a q-dimensional linear subspace so that the maximum distance between a point p and its projection p^* is minimized. In this paper we study the special case in which $q = 1$, i.e., we wish to cover S by k congruent cylinders of smallest minimum radius.

Clustering is a widely used technique for data mining, indexing, and classification [31, 33]. Most of the methods — both theoretical and practical — proposed in the last few years [2, 31, 17, 18, 22] are "full-dimensional," in the sense that they give equal importance to all the dimensions in computing the distance between two points. While such approaches have been successful for low-dimensional datasets, their accuracy and efficiency decrease significantly in higher dimensional spaces (see [28] for an excellent analysis and discussion). The reason for this performance deterioration is the so-called dimensionality curse. A full-dimensional distance for moderate-to-high dimensional spaces is often irrelevant, as the farthest neighbor of a point is expected to be almost as close as its nearest neighbor [27]. Methods such as principal component analysis, singular value decomposition, and randomized projection reduce the dimensionality of the data by projecting all points on a subspace so that the information loss is minimized. A full-dimensional clustering method is then used in this subspace. However, these methods do not handle well those situations in which different subsets of the points lie on different lower-dimensional subspaces. Motivated by the need for increased flexibility in reducing the data dimensionality, recently a number of methods have been proposed for *projective clustering*, in which points that are closely correlated in some subspace are clustered together [6, 7, 12, 34]. Instead of projecting the entire dataset on a single subspace, these methods project each cluster on its associated subspace, which is generally different from the subspace associated with another cluster.

Previous results. Meggido and Tamir [35] showed that it is NP-complete to decide whether a set of n points in the plane can be covered by k lines. This immediately implies that projective clustering is NP-Complete even in the planar case. In fact, it also implies that approximating the minimum width within a constant factor is NP-Complete. Approximation algorithms for hitting compact sets by minimum number of lines are presented in [25]. Fitting a $(d-1)$-hyperplane through S is the classical *width problem*. The width of a point set can be computed in $\Theta(n \log n)$ time for $d = 2$ [29], and in $O(n^{3/2+\varepsilon})$ expected time for $d = 3$ [1]. Duncan et al. gave an algorithm for computing the width approximately in higher dimensions [16]. Several algorithms with near-quadratic running time are known for covering a set of n points in the plane by two strips of minimum width; see [32] and references therein. Har-Peled and Varadarajan [24] have recently given a polynomial-time approximation scheme for the projective clustering problem in high dimensions for any fixed k and q.

Besides these results, very little is known about the projective clustering problem, even in the plane. A few Monte Carlo algorithms have been developed for projecting S onto a single subspace [30]. An exact solution to the projective clustering problem can be solved in $n^{O(dk)}$ time. We can also use the greedy algorithm [13] to cover points by congruent

q-dimensional hyper-cylinders. More precisely, if S can be covered by k hyper-cylinders of radius r, then the greedy algorithm covers S by $O(k \log n)$ hyper-cylinders of radius r in time $n^{O(d)}$. The approximation factor can be improved to $O(k \log k)$ using the technique by Brönimann and Goodrich [10]. For example, this approach computes a cover of $S \subseteq \mathbb{R}^d$ by $O(k \log k)$ hyper-cylinders of a given radius r in time $O(n^{O(d)} k \log k)$, assuming that S can be covered by k hyper-cylinders of width r each. Agarwal and Procopiuc [4] give a significantly faster scheme to cover S with $O(k \log k)$ hyper-cylinders of radius at most r in $O(dnk^2 \log^2 n)$ time. Combining this result with parametric search, they also give an algorithm that computes, in time $O(dnk^3 \log^4 n)$, a cover of S by $O(k \log k)$ hyper-cylinders of radius at most w^*, where w^* is the minimum radius of a cover by k hyper-cylinders.

The problem that we consider can also be thought of as an instance of a shape fitting problem, where the quality of the fit is determined by the maximum distance of a point from the shape. In these problems one generally wants to fit a shape, for example a line, hyperplane, sphere, or cylinder, through a given set of points P. Approximation algorithms with near-linear dependence on the size of P are now known for most of these shapes [9, 11, 23]. Trying to generalize these techniques to more complicated shapes seems to be the next natural step. The problem that we consider, that of trying to fit a point set with $k > 1$ lines, is an important step in this direction. The only previous result giving a $(1+\varepsilon)$-approximation in near-linear time in n, for $k > 1$, is the algorithm of Agarwal et al. [5] for $k = 2$. However, their techniques do not generalize to higher dimensions or to the case of $k > 2$.

Our results and techniques We present an $(1 + \varepsilon)$-approximation algorithm, with $O(n \log n)$ running time, for the k-line center problem in any fixed dimension; the constant of proportionality depends on k, ε, and d. We believe that the techniques used in showing this result are quite useful in themselves. We first show that there exists a small "certificate" $Q \subseteq P$, whose size does not depend on n, such that for any k-cylinders that cover Q, an enlargement of these cylinders by a factor of $(1+\varepsilon)$ covers P. The proof of this result is non-constructive in the sense that it does not give us an efficient way of constructing a certificate. The ideas used in this proof offer some hope for simplifying some known results as well as for proving the existence of a small certificate for other, more difficult problems.

We then observe that a well-known scheme based on sampling and iterated re-weighting [15] gives us an efficient algorithm for solving the problem. Only the existence of a small certificate is used to establish the correctness of the algorithm. This technique is quite general and can be used in other contexts as well. Thus it allows us to focus our attention on trying to prove the existence of small certificates, which seems to be the right thing to do for more complex shapes.

We present some definitions in Section 2, a proof of the existence of small certificates in Section 3, and our algorithm in Section 4

2 Preliminaries

A cylinder in $\mathbb{R}^d$, defined by specifying a line ℓ in $\mathbb{R}^d$ and a non-negative real number $r \geq 0$, is the set of all points within a distance of r from the line ℓ. We refer to ℓ as the *axis* and r as the *radius* of the cylinder. For $\varepsilon \geq 0$, an ε-enlargement of a cylinder with axis ℓ and radius r is the cylinder with axis ℓ and radius $(1+\varepsilon)r$. We define an *ε-enlargement* of a finite set of cylinders to be the set obtained by replacing each cylinder with its ε-enlargement.

Definition 2.1 Let P be a set of points in $\mathbb{R}^d$, $\varepsilon > 0$, and $k \geq 1$. We say that a subset $Q \subseteq P$ is an (ε, k)-certificate for P if for any k cylinders of *equal* radius that cover Q, an ε-expansion of the k cylinders results in a cover of P. We stress that the cylinders must have equal radius.

Let $I = [a, b]$ be an interval on the real line. For $\varepsilon \geq 0$, we define its ε-expansion or ε-enlargement to be the interval $I = [a - \varepsilon(b-a), b + \varepsilon(b-a)]$. We define an ε-expansion of a set of intervals on the real line analogously.

Definition 2.2 Let P be a set of points on the real line, $\varepsilon > 0$ a real number, and $k \geq 1$ an integer. We say that a subset $Q \subseteq P$ is an (ε, k)-certificate for P if for any set of k intervals that cover Q, an ε-expansion of the k intervals results in a cover for P. We emphasize that the k intervals are allowed to have different lengths.

Though we are using the term (ε, k)-certificate to mean two different notions, the context will clarify which one we are referring to.

3 Existence of Small Certificates

We show in this section that for any point set P in $\mathbb{R}^d$, $\varepsilon > 0$, and integer $k \geq 1$, there is an (ε, k)-certificate for P whose size is independent of the size of P. In order to do this, we first have to show the existence of such certificates for points on the real line.

Lemma 3.1 *Let P be any set of points on the real line, $\varepsilon > 0$ be a real, and $k \geq 1$ be an integer. There is an (ε, k)-certificate $Q \subseteq P$ for P with $g(\varepsilon, k)$ points, where $g(\varepsilon, k) = (k/\varepsilon)^{O(k)}$.*

Proof: The proof is by induction on k. If $k = 1$, we let Q be the two extreme points of P.

If $k > 1$, Q is picked as follows. Let Δ be the length of the interval I spanned by P. We divide I into k intervals of length Δ/k. Let $a_0 < a_1 \cdots < a_k$ denote the endpoints of these intervals (thus, $I = [a_0, a_k]$). For each a_i, $1 \leq i \leq k-1$, and $1 \leq j \leq k-1$, we compute an (ε, j)-certificate for $P \cap [a_0, a_i]$ and an $(\varepsilon, k-j)$-certificate for $P \cap [a_i, a_k]$. Let Q_1 denote the union of all these certificates.

We divide I into k/ε intervals of length $\varepsilon\Delta/k$; we call these *basic* intervals. Let us call an interval of the real line *canonical* if it is a union of basic intervals and has length at least

Δ/k. There are $\Theta(k^2/\varepsilon^2)$ canonical intervals. For each canonical interval I', we compute an $(\varepsilon, k-1)$-certificate for the points in P lying *outside* I'. Let Q_2 denote the union of these certificates.

We let $Q = Q_1 \cup Q_2$. We can easily verify that the size of Q is as claimed. We now argue that Q is an (ε, k)-certificate. Let $s_1, s_2, \ldots, s_k$ be k intervals that cover Q. We first consider the case where all these segments have length smaller than Δ/k. In this case there exists an a_i, $1 \leq i \leq k-1$, that is not contained in any of these segments. Without loss of generality, we may assume that the segments $s_1, \ldots, s_j$ lie to the left of a_i and the segments $s_{j+1}, \ldots, s_k$ lie to the right of a_i, for some $1 \leq j \leq k-1$. Since Q includes an (ε, j)-certificate for the points $P \cap [a_0, a_i]$, and $s_1, \ldots, s_j$ cover this certificate, we conclude that an ε-expansion of $s_1, \ldots, s_j$ contains $P \cap [a_0, a_i]$. By a symmetric arguement, we conclude that an ε-expansion of $s_{j+1}, \ldots, s_k$ covers $P \cap [a_i, a_n]$, and we are done.

We now consider the case when one of the segments, say s_1, has length at least Δ/k. Let I' denote the smallest canonical interval containing s_1; note that I' is covered by an ε-expansion of s_1. Since Q includes an $(\varepsilon, k-1)$-certificate for the points in P outside I', and $s_2, \ldots, s_k$ cover this certificate, we conclude that an ε-expansion of $s_2, \ldots, s_k$ covers the points in P outside I'. Since an ε-expansion of s_1 covers $P \cap I'$, the result follows.

□

We use Lemma 3.1 to prove the existence of a small certificate for any set of points in $\mathbb{R}^d$.

Lemma 3.2 *Let P be any set of points in $\mathbb{R}^d$, $\varepsilon > 0$, and $k \geq 1$. There exists an (ε, k)-certificate for P with $f(\varepsilon, k)$ points, where $f(\varepsilon, k) = k^{O(k)}/\varepsilon^{O(d+k)}$.*

Proof: Consider a cover of P by k congruent cylinders of minimum radius, and let this radius be denoted by w^*. Let $\delta = c\varepsilon$, where $c > 0$ is a sufficiently small constant. Clearly, there is a set L of $O(k/\varepsilon^{d-1})$ lines such that for any point in p, there is a line $\ell \in L$ with $d(p, \ell) \leq \delta w^*$. Indeed, for every cylinder C in the cover, draw a grid of size ε in the $(d-1)$-dimensional ball forming the base of C and draw a line parallel to the axis of C from each of the grid points.

Let P' be the point set obtained by projecting each point in P to the nearest line in L. Let $P'(\ell) \subseteq P'$ denote the points on the line ℓ. Let $Q'(\ell)$ be a (δ, k)-certificate, computed according to Lemma 3.1, for the points $P'(\ell)$ (by treating ℓ as the real line). Let $Q' = \bigcup_{\ell \in L} Q'(\ell)$, and let Q be the original points in P corresponding to Q'.

The bound on the size of Q is easily verified. Consider k cylinders of radius r that cover Q. Expanding each cylinder by an *additive* factor of δw^*, we cover Q'. For each $\ell \in L$, the segments formed by intersecting the cylinders with ℓ cover $Q'(\ell)$. Therefore, the δ-expansion of the segments results in a cover of $P'(\ell)$. It follows that a δ-expansion of the cylinders results in a cover for P' (this step needs a geometric claim that is easily verified). Expanding further by an additive factor of δw^*, we get a cover of P. Since w^* is the width of the optimal cover of P, we have

$$(r + \delta w^*)(1 + \delta) + \delta w^* \geq w^*.$$

Assuming that $\delta < 1/10$, this yields $w^* \leq 2r$. Thus the radius of the cylinders that cover P is at most

$$(r + \delta w^*)(1 + \delta) + \delta w^* \leq r((1 + 2\delta)(1 + \delta) + 2\delta) \leq (1 + \varepsilon)r.$$

□

The ideas used above in proving the existence of a small certificate may prove useful in other contexts also. To illustrate this point, we use these ideas to establish a known result whose earlier proofs relied on powerful tools like the Lowner-John Ellipsoid [3, 9, 21]. Let P be a set of points in $\mathbb{R}^d$, and $\varepsilon > 0$. We sketch a proof of the fact that there exists a subset $Q \subseteq P$ with $O(1/\varepsilon^{d-1})$ points such that for any slab that contains Q an ε-expansion of the slab contains P (That is, Q is a certificate for P.) If $d = 1$, Q consists of the two extreme points. If $d > 1$, consider the minimum width slab that covers P, and let w^* be its width. We find $1/\varepsilon$ slabs such that each point in P is within εw^* of the nearest slab. We move each point in P to the nearest slab. We now have $1/\varepsilon$ $(d-1)$-dimensional point sets. We recursively compute a $(d-1)$-dimensional certificate for each of these point sets, let Q' be their union, and Q the corresponding points in P. We argue as in Lemma 3.2 that Q is a certificate for P.

4 The Algorithm

Let $P \in \mathbb{R}^d$ be a set of n points in $\mathbb{R}^d$, and $\varepsilon > 0$ be a given parameter. Let $w^* \geq 0$ denote the smallest number such that there are k cylinders of radius w^* that cover P. We describe an efficient algorithm to compute a cover of P using k cylinders of radius at most $(1+\varepsilon)w^*$. The algorithm relies on the fact P has a small (ε, k)- certificate. Let $\mu = f(\varepsilon, k)$ denote the size of this certificate. The algorithm is an adaptation of the iterated re-weighting algorithm of Clarkson [15] for linear programming.

The algorithm maintains an integer weight $\mathrm{Wt}(p)$ for each point $p \in P$, which is intialized to 1. For any subset $Q \subseteq P$, let $\mathrm{Wt}(Q)$ denote the sum of the weights $\sum_{q \in Q} \mathrm{Wt}(q)$. Our algorithm is as follows.

1. We repeat the following sampling process until we meet with success.
   (a) Let $r = (2\log_2 e)\mu$ and R be a random sample from P of size $c(d, k)r \log r$, where $c(d, k)$ is a sufficiently large constant. R is constructed by $|R|$ independent repititions of the sampling process in which a point $p \in P$ is picked with probability $\mathrm{Wt}(p)/\mathrm{Wt}(P)$.
   (b) We compute $\mathcal{C}(R)$, the optimal k-cylinder cover for R, using the brute-force exact algorithm in $O(|R|^{2dk})$ time. Let S be the set of points in P not covered by $\mathcal{C}(R)$. If $\mathrm{Wt}(S) > \mathrm{Wt}(P)/r$ we go back to 1(a). Otherwise, we proceed to Step (2).
2. We check if the ε-expansion of $\mathcal{C}(R)$ covers P. If it does, our algorithm returns this ε-expansion as the approximate k-cylinder cover of P and halts. If it does not, we double the weight $\mathrm{Wt}(s)$ of each point $s \in S$ and go back to Step (1).

It is clear that if the algorithm does halt, it does so with an approximate k-cylinder cover of P. This is because $\mathcal{C}(R)$ is an *optimal* k-cylinder cover of R, and an ε-expansion of $\mathcal{C}(R)$ covers P. We now argue that the algorithm does indeed halt.

Lemma 4.1 *Let Q be an (ε, k)-certificate for P with size μ. Step (2) of the algorithm is executed at most $2\mu \log n$ times.*

Proof: Let us assume that Step (2) is executed ℓ times and each time the ε-expansion of $\mathcal{C}(R)$ does not cover P. This means that each time, $\mathcal{C}(R)$ did not cover some point from Q. Let $q \in Q$ be the point that was not covered the most number of times. Then q was not covered at least ℓ/μ times, and so its weight after ℓ executions of Step (2) is at least $2^{\ell/\mu}$.

Note that whenever Step (2) is executed, we have $\mathrm{Wt}(S) \leq \mathrm{Wt}(P)/r$. Since we double the weights of only the points in S, it follows that the weight of P increases by a factor of at most $(1 + 1/r)$ each time Step (2) is executed. After ℓ executions of Step (2), we have that $\mathrm{Wt}(P) \leq n(1 + 1/r)^{\ell}$. Since we must have $\mathrm{Wt}(q) \leq \mathrm{Wt}(P)$, we have

$$2^{\ell/\mu} \leq n(1 + 1/r)^{\ell}.$$

Taking logarithms, and rearranging the terms, we get

$$\ell \leq \frac{\log n}{(1/\mu - \log_2(1 + 1/r))}.$$

Using the fact that $\log_2(1 + a) \leq (\log_2 e)a$ for any $a \geq 0$, and substituting the value of r, it follows that $\ell \leq 2\mu \log n$. $\square$

Lemma 4.2 *Let R be a random sample of P as constructed in Step 1(a) of the algorithm, and let S be the set of points in P not covered by $\mathcal{C}(R)$. Then the probability that $Wt(S) > Wt(P)/r$ is at most $1/2$ if the constant $c(d, k)$ is chosen large enough.*

Proof: This follows from the theory of ε-nets [14, 26, 36]. Let $\mathcal{C}$ be the set of all cylinders in $\mathbb{R}^d$, and let $\mathcal{C}^k$ be the family of k-tuples in $\mathcal{C}$. It can be shown that VC-dimension of the range space $(\mathbb{R}^d, \mathcal{C}^k)$ is finite and depends only on k and d. Assuming that the constant $c(d, k)$ is larger than the VC-dimension of the range space, the lemma follows from a result by Haussler and Welzl [26]. $\square$

Lemma 4.2 implies that the expected number of times we have to iterate Steps 1(a) and 1(b) before we find a sample R such that $\mathrm{Wt}(S) \leq \mathrm{Wt}(P)/r$ is at most 2. Combining this with Lemma 4.1, we see that the expected running time of the algorithm is $O(\ell * n + \ell * (\mu \log \mu)^{2dk})$, where $\ell = \mu \log n$. We have thus obtained the main result of this paper.

Theorem 4.3 *Let P be a set of n points in $\mathbb{R}^d$, $w^* \geq 0$ denote the smallest number such that there are k cylinders of radius w^* that cover P, and $\varepsilon > 0$ be a parameter. We can compute k cylinders of width at most $(1 + \varepsilon)w^*$ that cover P in $O(n \log n)$ time, with the constant of proportionality depending on k, ε, and d.*

It is easy to see that the algorithm is fairly general and would work in related contexts, provided we can demonstrate the existence of a small certificate. One disadvantage is the large dependence of the running time on d and k. We have not tried to optimize this dependence. Some simple techniques, like computing a constant factor approximation first and then refining it to a factor of $(1+\varepsilon)$, may help improve the dependence on some of the parameters.

References

[1] P.K. Agarwal and M. Sharir. Efficient randomized algorithms for some geometric optimization problems. *Discrete Comput. Geom.*, 16:317–337, 1996.

[2] P. K. Agarwal and M. Sharir, Efficient algorithms for geometric optimization, *ACM Comput. Surv.*, 30 (1998), 412–458.

[3] P.K. Agarwal and S. Har-Peled. Maintaining the approximate extent measures of moving points. In *Proc. 12th ACM-SIAM Sympos. Discrete Algorithms*, pages 148–157, 2001.

[4] P.K. Agarwal and C.M. Procopiuc. Approximation algorithms for projective clustering. In *Proc. 11th ACM-SIAM Sympos. Discrete Algorithms*, pages 538–547, 2000.

[5] P. K. Agarwal, C. M. Procopiuc, and K. R. Varadarajan. A $(1+\varepsilon)$-approximation algorithm for 2-line-center. Submitted for publication.

[6] C. C. Aggarwal, C. M. Procopiuc, J. L. Wolf, P. S. Yu, and J. S. Park. Fast algorithms for projected clustering. In *Proc. of ACM SIGMOD Intl. Conf. Management of Data*, pages 61–72, 1999.

[7] C. C. Aggarwal, and P. S. Yu. Finding generalized projected clusters in high dimensional spaces. In *Proc. of ACM SIGMOD Intl. Conf. Management of Data*, pages 70–81, 2000.

[8] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data for data mining applications. In *Proc. ACM SIGMOD Conf. on Management of Data*, pages 94–105, 1998.

[9] G. Barequet and S. Har-Peled. Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. In *Proc. 10th ACM-SIAM Sympos. Discrete Algorithms*, pages 82–91, 1999.

[10] H. Brönnimann and M. T. Goodrich, Almost optimal set covers in finite VC-dimension, *Discrete Comput. Geom.*, 14 (1995), 263–279.

[11] T.M. Chan. Approximating the diameter, width, smallest enclosing cylinder and minimum-width annulus. In *Proc. 16th Annu. ACM Sympos. Comput. Geom.*, pages 300- -309, 2000.

[12] K. Chakrabarti and S. Mehrotra. Local dimensionality reduction: A new approach to indexing high dimensional spaces. In *Proc. 26th Intl. Conf. Very Large Data Bases*, pages 89–100, 2000.

[13] V. Chvátal. A greedy heuristic for the set-covering problem. *Math. Oper. Res.*, 4:233–235, 1979.

[14] K. L. Clarkson. New applications of random sampling in computational geometry. *Discrete Comput. Geom.*, 2:195–222, 1987.

[15] K. L. Clarkson. Las Vegas algorithms for linear and integer programming. *J. ACM*, 42:488–499, 1995.

[16] C. A. Duncan, M. T. Goodrich, and E. A. Ramos, Efficient approximation and optimization algorithms for computational metrology, *Proc. 8th ACM-SIAM Sympos. Discrete Algorithms*, 1997, pp. 121–130.

[17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. 2nd Intl. Conf. Knowledge Discovery and Data Mining*, pages 226–231, 1996.

[18] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. Density-connected setsand their application for trend detection in spatial databases. In *Proc. 3rd Intl. Conf. Knowledge Discovery and Data Mining*, 1997.

[19] C. Faloutsos and K.-I. Lin, FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia databases, *Proc. ACM SIGMOD Conf. on Management of Data*, 1995, pp. 163–173.

[20] T. Feder and D. H. Greene, Optimal algorithms for approximate clustering, *Proc. 20th Annu. ACM Sympos. Theory Comput.*, 1988, pp. 434–444.

[21] P. M. Gruber. Aspects of approximation of convex bodies. In Peter M. Gruber and J. M. Wills, editors, *Handbook of C onvex Geometry*, volume A, chapter 1.10, pages 319–345. North-Holland, Amste rdam, Netherlands, 1993.

[22] S. Guha, R. Rastogi, and K. Shim. CURE: An efficient clustering algorithm for large databases. In *Proc. ACM SIGMOD Intl. Conf. Management of Data*, pages 73–84, 1998.

[23] S. Har-Peled and K. Varadarajan. Approximate Shape Fitting via Linearization. In *Proc. IEEE Symp. Foundations of Comp. Sci.*, pages 66–73, 2001.

[24] S. Har-Peled and K. Varadarajan. Projective clustering in high dimensions using core-sets. To appear in *ACM Symp. Comput. Geom.*, 2002.

[25] R. Hassin and N. Megiddo, Approximation algorithms for hitting objects by straight lines, *Discrete Appl. Math.*, 30 (1991), 29–42.

[26] D. Haussler and Emo Welzl. Epsilon-nets and simplex range queries. *Discrete Comput. Geom.*, 2:127–151, 1987.

[27] A. Hinneburg, C. C. Aggarwal, and D. A. Keim. What is the nearest neighbor in high dimensional spaces? In *Proc. 26th Intl. Conf. Very Large Data Bases*, pages 506–515, 2000.

[28] A. Hinneburg and D. A. Keim. Optimal grid-clustering: Towards breaking the curse of dimensionality in high-dimensional clustering. In *Proc. 25th Intl. Conf. Very Large Data Bases*, pages 506–517, 1999.

[29] M. E. Houle and G. T. Toussaint, Computing the width of a set, *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-10 (1988), 761–765.

[30] P. Indyk, R. Motwani, P. Raghavan, and S. Vempala, Locality-preserving hashing in multidimensional spaces, *Proc. 29th Annu. ACM Sympos. Theory Comput.*, 1997, pp. 618–625.

[31] A. K. Jain, M. N. Murty, and P. J. Flynn, Data clustering: A review, *ACM Comput. Surv.*, 31 (1999), 264–323.

[32] J. W. Jaromczyk and M. Kowaluk, The two-line center problem from a polar view: A new algorithm and data structure, *Proc. 4th Workshop Algorithms Data Struct.*, *Lecture Notes Comput. Sci.*, Vol. 955, Springer-Verlag, 1995, pp. 13–25.

[33] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, 1990.

[34] C. M. Procopiuc, M. Jones, P. K. Agarwal, and T. M. Murali. A Monte Carlo algorithm for fast projective clustering. In *Proc. ACM SIGMOD Intl. Conf. Management of Data*, 2002. to appear.

[35] N. Megiddo and A. Tamir, On the complexity of locating linear facilities in the plane, *Oper. Res. Lett.*, 1 (1982), 194–197.

[36] K. Mulmuley. *Computational Geometry: An Introduction Through Randomized Algorithms.* Prentice Hall, Englewood Cliffs, NJ, 1993.