Shape Dimension and Approximation from Samples *

Tamal K. Dey   Joachim Giesen   Samrat Goswami   Wulue Zhao [†]

Abstract

There are many scientific and engineering applications where an automatic detection of shape dimension from sample data is necessary. Topological dimensions of shapes constitute an important global feature of them. We present a Voronoi based dimension detection algorithm that assigns a dimension to a sample point which is the topological dimension of the manifold it belongs to. Based on this dimension detection, we also present an algorithm to approximate shapes of arbitrary dimension from their samples. Our empirical results with data sets in three dimensions support our theory.

1 Introduction.

Interpretation of shapes from its samples is needed in many scientific and engineering applications. As a result, defining and detecting features that contain useful information about the shape are subjects of active research in shape modeling [4, 13, 21, 22]. A global feature of a shape is its dimension which has been defined in various ways to accommodate intricacies and varieties in shapes. Topological and Hausdorff dimensions are two such definitions that capture the global feature, the first one stresses on the space connectivity and the second one on the space filling property [14, 16]. In this work we focus on the topological dimension of the shapes that are smooth manifolds embedded in an Euclidean space. Data collected for scientific analysis through natural phenomena or simulations lie on such manifolds and can reveal important information if the underlying dimension is detected automatically.

Automatic dimension detection is a major challenge in the fields of learning theory, pattern recognition [5] and artificial intelligence in general [20, 23]. In these applications samples can be generated from an otherwise unknown manifold. The dimension of the manifold, if can be detected from the samples, provides useful feedback to the learning process and sometimes can be used for various matching purposes in pattern recognition.

In two and three dimensions, the topological dimension of a shape becomes obvious with a visual inspection of the data points. However, often a shape needs to be computed from the samples for further processing. For example, in the study of dynamical systems, the quasiperiodic and chaotic orbits can be automatically recognized from sample points in the phase portraits if automatic shape recognition can be performed; see [23]. To this end one may try reconstruction algorithms for curves [2, 11, 17] and for surfaces [1, 3, 6, 8, 12, 18]. Unfortunately, these reconstruction algorithms are of no use if the dimension of the shape is not determined a priori. For example, a surface reconstruction algorithm cannot produce a curve out of a sample that has been derived from a curve. The situation becomes worse when data have samples from shapes of various dimensions such as the data from a phase portrait of a sampled dynamical system. In this case, first, one needs to separate the samples according to their dimensions and then apply respective reconstruction steps.

In this paper we present an algorithm that determines the dimension of the shape from which samples are derived. Specifically, it assigns a dimension k to a sample point if it belongs to a manifold of dimension k. We also present a shape approximation algorithm which filters simplices from a Delaunay complex approximating the sampled shape subsequent to the dimension detection step. The assumption we make on the sampling is that it is feature dependent, i.e., it is dense wherever small features exist and is relatively sparse wherever intricate features are absent. This is similar to the ε-sampling condition introduced in [1, 2]. But, as argued later, we need a stronger assumption on sampling for automatic dimension detection and this assumption cannot be avoided if one wants to guarantee correctness.

Our approach is based on the Voronoi diagrams of the samples. The dimension of a small neighborhood of a point in a manifold depends on the dimension of the manifold itself. For example, a point on a curve has a neighborhood homeomorphic to a 1-disk and a point on a surface has neighborhood homeomorphic to a 2-disk. Under an appropriate sampling condition, the structure of the Voronoi cells of the sample points contains information about the dimension and shape of these neighborhoods. We analyze these structures carefully and extract this information. We test our algorithms on a number of data sets in three dimensions. Empirical results support the theory based on which we design the dimension detection algorithm DIMENSION and the shape approximation algorithm COCONESHAPE.

2 Sampling and Voronoi geometry.

Let $\mathcal{M} = \{M_1, M_2, \ldots, M_\ell\}$ be a collection of smooth compact manifolds embedded in $\mathbb{R}^d$ and $M = |\mathcal{M}|$ be its underlying space. In general, $\mathcal{M}$ can have manifolds of different dimensions, i.e., M_i, $i \in \{1, 2, \ldots, \ell\}$ may be a k-manifold for any k, $1 \le k \le d$. In order to determine the dimensions of the manifolds from their samples, we must assume some density condition on the samples, otherwise they may be too sparse to provide any meaningful information about M.

2.1 Sample density.

We adapt the density assumption of [1] which is based on the *local feature size* of a shape which measures the distance to the *medial axis*. The medial axis of M is the closure of the set of points in $\mathbb{R}^d$ that have two or more closest points on M. The local feature size is a function $f : M \to \mathbb{R}$ where $f(x)$ is the Euclidean distance to the medial axis of a point $x \in M$. Intuitively, $f(x)$ is a measure of how complicated M is in the neighborhood of x. A nice property of $f(\cdot)$ is that it is Lipschitz continuous, i.e., $f(x) \le f(y) + \|x - y\|$ for any two points x and y in M.

A sample $P \subset M$ is called an ε-sample of M if each point $x \in M$ has a sample point $p \in P$ within $\varepsilon f(x)$ distance. This means that each point in M has a sample within ε factor of its local feature size. We argue that this ε-sampling assumption is not suitable for dimension detection even for small values of ε since it cannot prevent ambiguity.
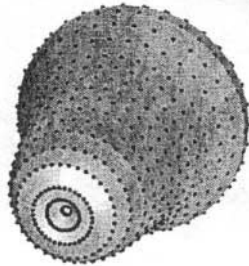

Figure 1: An ambiguous sample.

Consider the sample shown in Figure 1. The sample is taken from the surface of a pear and satisfies the ε-sampling condition for $\varepsilon < 0.4$. However, if it is not known that the sample is from a surface, the points in circular arrangements near the top may as well be taken as a sample of the respective circles. Any dimension detection algorithm would be confused in this case as to conclude if the samples are generated from a surface or a set of curves. This example shows that we need a stricter sampling condition to disambiguate the dimension of the underlying shapes.

DEFINITION 2.1. *A sample P of a collection of smooth manifolds $\mathcal{M}$ is called (ε, δ)-sample of M for $\frac{\varepsilon}{2} \le \delta < \varepsilon < 1$ if each point $x \in M$ has a sample point within $\varepsilon f(x)$ distance and each sample point p has all other sample points at least*

$\delta f(p)$ *away from it.*

Remark. The above definition requires that the sample be dense with respect to the local feature size, but at the same time samples cannot be arbitrarily close to form an arbitrary pattern such as curves on surfaces. Notice that this requirement is much less strict than the condition where the sampling has to be uniform everywhere. We chose the lower bound on δ as $\frac{\varepsilon}{2}$ for making further calculations precise though any other constant greater than 1 will be equally valid for our theoretical analysis. Such sampling conditions have been studied in the context of surface reconstruction in [9, 15].

2.2 Tangent and normal spaces.

Since we are dealing with a collection $\mathcal{M}$ of smooth manifolds in $\mathbb{R}^d$, results from differential geometry ensure that a tangent space at each point $p \in M$ is well defined [7]. The dimension of the tangent space at p coincides with the topological dimension of the manifold containing p. We also refer to this dimension as the *dimension of p*. Our goal is to determine the dimension of the sample points from their Voronoi cells.

Let T_p denote the *tangent space* at p, i.e., the affine subspace of $\mathbb{R}^d$ spanned by tangent vectors at p. The affine subspace spanned by normals to T_p constitute the *normal space N_p at p*. The dimensions of N_p and T_p add up to d.

The main tool we use for detecting dimension is the geometric structure of the Voronoi cells. Let V_P and D_P denote the Voronoi diagram and its dual, the Delaunay triangulation for a sample P. A Voronoi cell V_p for a sample point p is the set of all points that are no farther away from p than any other sample point in P. Let $\tilde{T}_p$ denote the set of all points in T_p that are no farther away from p than from any other sample point, i.e., $\tilde{T}_p = T_p \cap V_p$. Similarly, let $\tilde{N}_p = N_p \cap V_p$ denote the set of points in N_p that are no farther away from p than from any other sample point. We call $\tilde{T}_p$ and $\tilde{N}_p$ the *tangent* and *normal polytopes* of p respectively. The main observation based on which the dimension detection proceeds is that the Voronoi cell V_p of $p \in P$ *approximates* the Minkowski sum of $\tilde{T}_p$ and $\tilde{N}_p$. Figure 2 illustrates the above fact. This figure shows the Voronoi cells of points of various dimensions in an (ε, δ)-sample of M in three dimensions. In the left picture the tangent polytope is the segment going through p and the normal polytope is shaded. This is the case for a sample point on a curve. In the middle picture, the tangent polytope is shaded and the normal polytope is the line segment going through p. This is the case for a sample point on a surface. In the right picture, the tangent polytope is the entire Voronoi cell and the normal polytope is the single point p. This is the case for an interior point in a solid.

2.3 Voronoi subpolytopes.

The Voronoi cell V_p contains information about the dimension of $\tilde{T}_p$ and hence T_p. Our
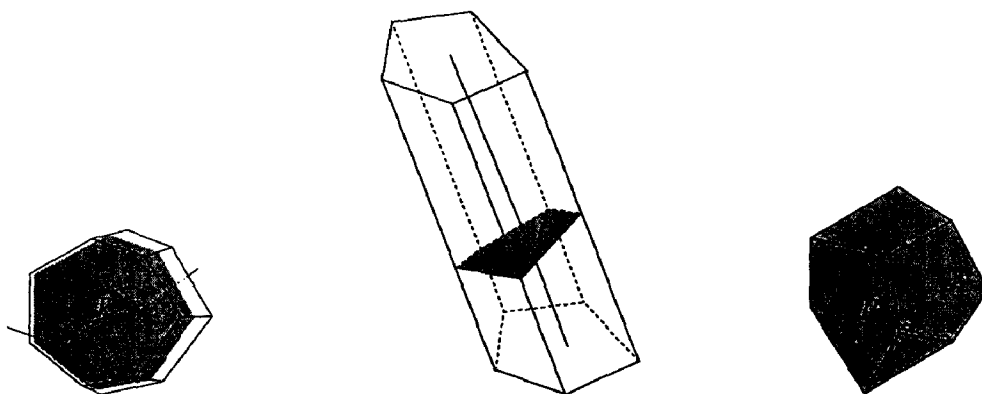
Figure 2: Tangent and normal polytopes of a sample on a curve (left), surface (middle) and solid (right).

task is to separate $\tilde{T}_p$ from $\tilde{N}_p$ from the Voronoi cell V_p. We achieve this by examining the structure of a sequence of *Voronoi subpolytopes* as defined recursively below. The definition of Voronoi subpolytopes is based on the concept of poles for Voronoi cells which was introduced in [1]. We generalize it here for Voronoi subpolytopes. For convenience we use the notation $\angle(\mathbf{v}, \mathbf{w})$ to denote the *acute* angle between the lines supporting two vectors $\mathbf{v}$ and $\mathbf{w}$.

DEFINITION 2.2. *The Voronoi subpolytopes for a sample point* $p \in M$ *are special subsets* $V_p^i \subseteq V_p, i = 1, \ldots, d$ *of the Voronoi cell* V_p. *Let* $V_p^d = V_p$. *Assume that* V_p^i *is already defined. Let* v_p^i *be the farthest point in* V_p^i *from* p. *We call* v_p^i *the pole of* V_p^i *and the vector* $\mathbf{v}_p^i = v_p^i - p$ *its pole vector. If* V_p^i *is unbounded,* v_p^i *is taken at infinity, and the direction of* $\mathbf{v}_p^i$ *is taken as the average of all directions given by unbounded edges. The Voronoi subpolytope* V_p^{i-1} *is the minimal polytope containing all points* $\{x : \angle((x-p), \mathbf{v}_p^i) = \frac{\pi}{2}\}$.

Clearly, $V_p^{i-1} \subset V_p^i$ is a polytope orthogonal to the pole vector of V_p^i. In Figure 3 we show the Voronoi subpolytopes for a sample point on a curve and on a surface, respectively in three dimensions.

We introduce the definition of *height* to measure the structure of Voronoi subpolytopes.

DEFINITION 2.3. *The height* H_p^i *of a Voronoi subpolytope* V_p^i *is the length* $\|\mathbf{v}_p^i\| = \|v_p^i - p\|$.

The height of V_p^i measures its elongation, but we also need a measure for the 'fatness' of V_p^i which is measured by the height of V_p^{i-1}.
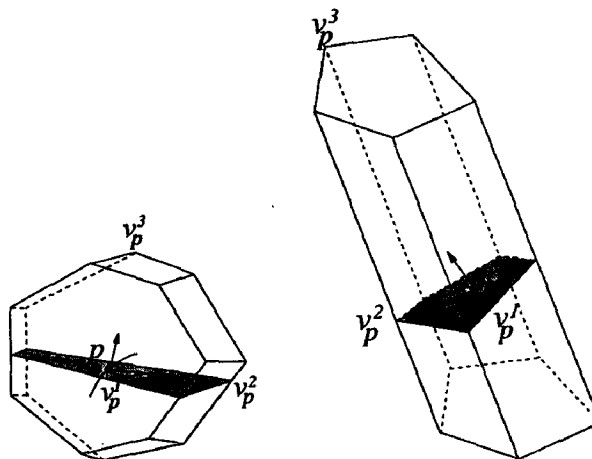

Figure 3: Voronoi subpolytopes for a point on a curve (left) and a surface (right). V_p^3 are the entire Voronoi cells, V_p^2 are the shaded polygons, V_p^1 are the line segments going through p.

3 Geometric guarantees.

Let $p \in P$ be an interior point in a manifold of dimension k, $1 \leq k \leq d$, where P is an (ε, δ)-sample of M, the underlying space of a collection of manifolds $\mathcal{M}$ in $\mathbb{R}^d$. We assume that p satisfies the following condition.

ASSUMPTION 3.1. V_p *does not contain any boundary point of* M.

One way to satisfy the above assumption for all sample points is to consider only manifolds without boundary. Of course, this excludes d-manifolds since they necessarily have boundaries in $\mathbb{R}^d$.

We claim that V_p^k approximates the tangent polytope $\tilde{T}_p$ and thus cannot have large heights. All other higher dimensional Voronoi subpolytopes are long and thin along some normal direction and they have large height. The claims are based on the following two observations:

(i) All Voronoi subpolytopes V_p^i, $k < i \leq d$, contain a boundary point of $\tilde{N}_p$.

(ii) The pole vectors of Voronoi subpolytopes V_p^i, $k < i \leq d$, approximate some vector in the normal space N_p.

Fact (i) implies that the height H_p^i is at least as large as $f(p)$, the local feature size at p, for $k < i \leq d$. Fact (ii) implies that the affine space of V_p^k is normal to k orthogonal vectors each of which approximates a vector in the normal space N_p. This in turn means that the affine space of V_p^k lies close to T_p and V_p^k approximates the tangent polytope $\tilde{T}_p$. We extend a result of [3] to show that the diameter of the tangent polytope is small, $O(\varepsilon)f(p)$ in specific. It follows that the height H_p^i is $O(\varepsilon)f(p)$ for $1 \leq i \leq k$.

Fact (i) and (ii) hold vacuously for $k = d$ since $k < d$ is not satisfied. In this case $V_p^k = V_p$ is the tangent polytope and its height is small due to ε-sampling condition.

LEMMA 3.1. *Let $v \in V_p^i$ be any point in the Voronoi subpolytope V_p^i where $k < i \leq d$ and $\|v - p\| > \mu f(p)$. Then there is a normal vector $\mathrm{n}_p \in N_p$ so that*

$$\angle((v - p), \mathrm{n}_p) \leq \sin^{-1} \frac{\varepsilon}{\mu(1 - \varepsilon)} + \sin^{-1} \frac{\varepsilon}{1 - \varepsilon}.$$

Proof. If $v \in \tilde{N}_p$, then the condition is trivially satisfied. So, assume $v \notin \tilde{N}_p$. First we establish that there is a point $z \in M$ so that the line going through v and z intersects the boundary of the normal polytope $\tilde{N}_p$. Consider the cone $X_v = \{x : x = tv + (1 - t)y\}$ for all $t \in [0, 1]$ and $y \in \tilde{N}_p$. Since $\tilde{N}_p$ has dimension $d - k$, the dimension of X_v is $d - k + 1$. The manifold M intersects X_v at p transversally and $L = M \cap X_v$ must be a manifold of dimension $(d - k + 1) + k - d = 1$. We argue that L must intersect the boundary of X_v at a point $z \notin \tilde{N}_p$. If $z \in \tilde{N}_p$ then there is a medial axis point violating the sampling condition. To see this consider a ball tangential to M at p growing towards z, i.e., its center moving along pz from p towards z. It must touch M at another point before or when it reaches z implying that there is a medial axis point within the segment connecting z and p. This means z has the nearest sample point p at a distance more than $f(z)$, a contradiction to our sampling condition. Thus, z can lie only on $bd(X_v) \setminus \tilde{N}_p$ establishing our claim. See Figure 4 for an illustration.

Let the line going through v and z meet the boundary of $\tilde{N}_p$ at m. Clearly, $m \in bd(V_p)$. Consider the triangle pvm. Take n_p as the vector $r - p$ where r is the other end point of the line segment in $\tilde{N}_p$ going through p and m. We are interested in the angle $\angle rpv$. We have

$$\angle rpv = \angle pmv + \angle pvm.$$

Since $z \in M$ has p as the nearest sample point we have $\|z - p\| \leq \varepsilon f(z)$. Applying Lipschitz condition between $f(z)$
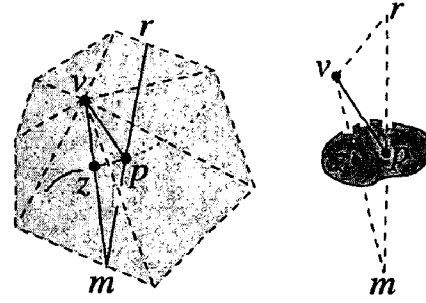

Figure 4: The polytope X_v in Lemma 3.1 is three dimensional for a curve point (left) and is two dimensional for a surface point p (right) in three dimensions. The edges of this polytope are dashed. $\tilde{N}_p$ is the lightly shaded polygon for the curve point, and it is the segment between m and r for the surface point.

and $f(p)$ we get $f(z) \leq \frac{\varepsilon}{1-\varepsilon} f(p)$. Therfore,

$$\|z - p\| \leq \frac{\varepsilon}{1 - \varepsilon} f(p).$$

Also, $\|m - p\| \geq f(p)$ since there is an empty ball touching M at p with center on the medial axis and on the segment pm. To see this, again consider growing a ball touching M at p and moving its center from p towards m until it touches another point on M. The center of such a ball must be in V_p since the ball is empty. It follows that

$$\angle pmv \leq \sin^{-1} \frac{\|z - p\|}{\|m - p\|}$$

$$\leq \sin^{-1} \frac{\varepsilon}{1 - \varepsilon}.$$

Considering the triangle pvm we have

$$\angle pvm \leq \sin^{-1} \frac{\|z - p\|}{\|v - p\|}$$

$$\leq \sin^{-1} \frac{\varepsilon}{\mu(1 - \varepsilon)}.$$

This establishes the claim of the lemma.

LEMMA 3.2. $H_p^i \geq f(p)$ *for $k < i \leq d$.*

Proof. The Voronoi subpolytope V_p^i intersects the normal polytope $\tilde{N}_p$ in a polytope of dimension $i + (d - k) - d = i - k$ which is greater than 0 since $i > k$. It means that V_p^i has a point, say x, where $x \in bd(V_p^i) \cap \tilde{N}_p \subset bd(V_p) \cap \tilde{N}_p$. Consider growing a ball touching M at p always keeping the center on px until it touches M at another point. The center of this grown ball is a medial axis point that must lie in V_p since the ball is empty. Thus, $\|x - p\| \geq f(p)$. In particular, the length of the pole vector $H_p^i = \|v_p^i - p\|$ must be at least $f(p)$.

COROLLARY 3.1. *There exists a normal vector* $\mathbf{n}_p \in N_p$ *for each pole vector* $\mathbf{v}_p^i$, $k < i \leq d$ *so that* $\angle(\mathbf{v}_p^i, \mathbf{n}_p) \leq \alpha = 2\sin^{-1}\frac{\varepsilon}{1-\varepsilon}$.

Proof. From Lemma 3.2 we get $\|\mathbf{v}_p^i\| \geq f(p)$ for $k < i \leq d$. So, plugging $\mu = 1$ into Lemma 3.1 we obtain that $\angle(\mathbf{v}_p^i, \mathbf{n}_p) \leq 2\sin^{-1}\frac{\varepsilon}{1-\varepsilon} = \alpha$ for some vector $\mathbf{n}_p \in N_p$.

LEMMA 3.3. *Let* $x \in V_p^i$ *be any point on the boundary of the Voronoi subpolytope* V_p^i, *where* $i \leq k$. *We must have*

$$\frac{\delta}{2}f(p) \leq \|x - p\| \leq \frac{\varepsilon}{1-\varepsilon} \sec\left(\frac{\alpha}{2}(1 + 4\sqrt{d-k})\right)$$

Proof. The lower bound on $\|x - p\|$ follows from the δ-condition in the sampling density. Since any sample point must be at least $\delta f(p)$ away from p, any point on the boundary of V_p is at least $\frac{\delta}{2}f(p)$ distance away from p. The point x lies on the boundary of V_p^i and hence on the boundary of V_p.

For the upper bound we first exclude the case $k = d$. In this case $V_p^k = V_p$ is the tangent polytope with height $H_p^d \leq \varepsilon f(p)$ due to sampling condition. Thus, the upper bound holds trivially in this case. For the other cases when $k < d$ we first show that the vector $\mathbf{x}_p = x - p$ makes a large angle with any vector in the normal space N_p. Let $\mathbf{n}_p \in N_p$ be a vector so that $\angle(\mathbf{x}_p, \mathbf{n}_p)$ is the smallest.

From Lemma 3.1 it follows that for each pole vector $\mathbf{v}_p^i$, $k < i \leq d$, there exists a vector $\mathbf{n}_p^i$ in the normal space so that $\angle(\mathbf{v}_p^i, \mathbf{n}_p^i) \leq \alpha$, where $\alpha = 2\sin^{-1}\frac{\varepsilon}{1-\varepsilon}$ is small. This result with the fact that the vectors $\mathbf{v}_p^i$ are orthogonal to each other imply that the vectors $\mathbf{n}_p^i$ are linearly independent and thus span N_p. Consider the vector $\mathbf{v}_p = \Sigma_{i=k+1}^d a_i \mathbf{v}_p^i$ if $\mathbf{n}_p = \Sigma_{i=k+1}^d a_i \mathbf{n}_p^i$. Without loss of generality assume that the vectors $\mathbf{v}_p^i$, $\mathbf{n}_p^i$ for $i = k+1, \ldots, d$ are unit. We can also choose $\Sigma_{i=k+1}^d a_i^2 = 1$ which only scales $\mathbf{n}_p$. With these choices $\mathbf{v}_p$ becomes an unit vector. Let $\angle(\mathbf{v}_p, \mathbf{n}_p) = \beta$. We show that β is small.

We can write $\mathbf{n}_p^i = \mathbf{v}_p^i + \mathbf{w}^i$ where $\mathbf{w}^i$ is a small vector with $\|\mathbf{w}^i\| \leq \alpha = 2\sin^{-1}\frac{\varepsilon}{1-\varepsilon}$ since $\mathbf{v}_p^i$ is a unit vector and $\angle(\mathbf{v}_p^i, \mathbf{n}_p^i) \leq \alpha$. Therefore,

$$\mathbf{n}_p = \Sigma_{i=k+1}^d a_i \mathbf{n}_p^i$$
$$= \Sigma_{i=k+1}^d a_i \mathbf{v}_p^i + \Sigma_{i=k+1}^d a_i \mathbf{w}^i$$
$$= \mathbf{v}_p + \Sigma_{i=k+1}^d a_i \mathbf{w}^i$$

It follows that

$$\sin\beta \leq \|\Sigma_{i=k+1}^d a_i \mathbf{w}^i\|$$
$$\leq \Sigma_{i=k+1}^d |a_i| \cdot \|\mathbf{w}^i\|$$
$$\leq \alpha \Sigma_{i=k+1}^d |a_i|$$

The term $\Sigma_{i=k+1}^d |a_i|$ is at most $\sqrt{d-k}$ by Cauchy-Schwartz's inequality with the condition $\Sigma_{i=k+1}^d a_i^2 = 1$. Therefore, for small β

$$\frac{\beta}{2} \leq \sin\beta \leq \alpha\sqrt{d-k}$$

Now consider the angle between $\mathbf{x}_p$ and $\mathbf{n}_p$. By triangular inequality

$$\angle(\mathbf{x}_p, \mathbf{n}_p) \geq \angle(\mathbf{x}_p, \mathbf{v}_p) - \angle(\mathbf{v}_p, \mathbf{n}_p)$$
$$= \frac{\pi}{2} - \beta.$$

Using the contrapositive of Lemma 3.1 we obtain that $\|x - p\| = \mu f(p)$ where

$$\mu \leq \left(\frac{\varepsilon}{1-\varepsilon}\right) \frac{1}{\sin(\frac{\pi}{2} - \beta - \sin^{-1}\frac{\varepsilon}{1-\varepsilon})}$$
$$\leq \left(\frac{\varepsilon}{1-\varepsilon}\right) \frac{1}{\cos(\beta + \frac{\alpha}{2})}$$
$$\leq \left(\frac{\varepsilon}{1-\varepsilon}\right) \sec\left(\frac{\alpha}{2}(1 + 4\sqrt{d-k})\right)$$

completing the proof.

Combining Lemma 3.2 and 3.3 we obtain the following theorem.

THEOREM 3.1. *Let* $p \in P$ *be any point derived from a manifold of dimension* k *embedded in* $\mathbb{R}^d$ *where* P *is an* (ε, δ)-*sample and* V_p *does not contain any boundary point. Then following conditions hold:*

1. $H_p^i \geq f(p)$ *for* $k < i \leq d$

2. $\frac{\delta}{2}f(p) \leq H_p^i \leq \frac{\varepsilon}{1-\varepsilon} \sec(\frac{\alpha}{2}(1 + 4\sqrt{d-k}))f(p)$ *for* $1 \leq i \leq k$.

4 Dimension detection.

Theorem 3.1 is the basis of our algorithm for dimension detection. From Theorem 3.1 we have $H_p^1 = O(\varepsilon)f(p)$ if we assume that the dimension of p is at least one. Certainly, H_p^1 is small if ε is sufficiently small. On the other hand, $H_p^i \geq f(p)$ are large for $k < i \leq d$. Then, one possible algorithm to determine k may proceed as follows. Compare the heights H_p^i with H_p^1 in decreasing sequence of i starting with $i = d$. Continue the sequence of comparisons as long as the ratio $\frac{H_p^1}{H_p^i}$ is less than an user chosen parameter ρ. The comparison stops when the ratio goes above ρ and we note the value of the corresponding i. This strategy has the potential problem that the heights H_p^i for $1 \leq i \leq k$ are small
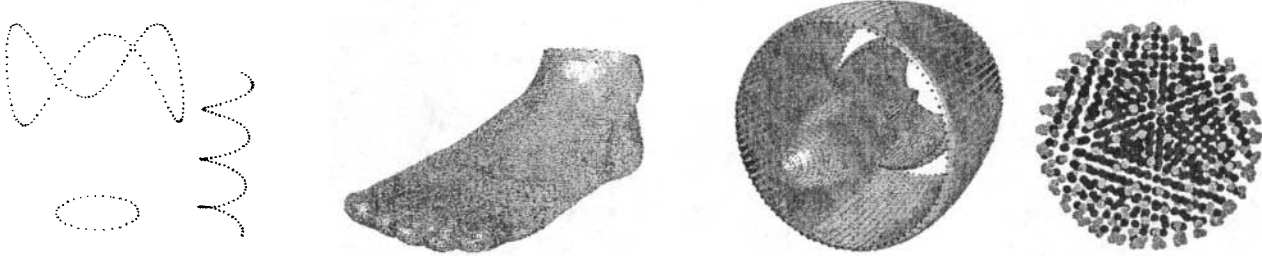
Figure 5: Output of the algorithm DIMENSION on the datasets CURVES, FOOT, ENGINE and BALL. Points classified one-dimensional by the algorithm are colored red (leftmost), points classified two-dimensional are colored green (middle two) and points classified three-dimensional are colored blue (rightmost), respectively.

compared to $f(p)$, but they may not be small with respect to H_p^1. The remedy is obtained by the δ-condition on the (ϵ, δ)-sampling. We note that,

$$\frac{H_p^1}{H_p^i} \geq \frac{\delta(1 - \epsilon)}{2\epsilon \sec\left(\frac{\alpha}{2}(1 + 4\sqrt{d - k})\right)}$$

$$> \frac{(1 - \epsilon)}{4 \sec\left(\frac{\alpha}{2}(1 + 4\sqrt{d - k})\right)}$$

$$= \Omega(1) \text{ for } 1 \leq i \leq k,$$

and for $k < i \leq d$,

$$\frac{H_p^1}{H_p^i} \leq \left(\frac{\epsilon}{1 - \epsilon}\right) \sec\left(\frac{\alpha}{2}(1 + 4\sqrt{d - k})\right)$$

$$= O(\epsilon)$$

Thus, a choice of ρ in the range

$$\left[\left(\frac{\epsilon}{1 - \epsilon}\right) \sec\left(\frac{\alpha}{2}(1 + 4\sqrt{d - k})\right), \frac{(1 - \epsilon)}{4 \sec\left(\frac{\alpha}{2}(1 + 4\sqrt{d - k})\right)}\right]$$

suffices for our algorithm DIMENSION. So, for example, in dimension 5, with ϵ less than 0.03, $\rho \in [0.04, 0.2]$ will suffice though in practice we can deal with larger values of ϵ with larger ρ.

DIMENSION (P, ρ, d)
1   Compute V_P
2   for all $p \in P$
3   compute $H_p^1, \ldots, H_p^d$
4     $i := d$
5   while $H_p^1/H_p^i < \rho$
6       $i := i - 1$
7   endwhile
8     $dim(p) := i$
9   endfor

5   Shape approximation.

We use DIMENSION to design a shape approximation algorithm that can approximate M with a piecewise linear com-

plex interpolating the sample P. This algorithm is dimension independent in that it does not need to know a priori the dimension of the shape from where the sample is derived. The algorithm can be seen as a generalization of our CO-CONE algorithm that reconstructs surfaces in three dimensions [3, 10]. We need the following definition of *cocone* for a sample point p.

DEFINITION 5.1. *Let* $p \in P$ *be a sample point from a manifold of dimension* k. *The cocone for* p, C_p, *is defined as the set of all points* $x \in V_p$ *so that the segment connecting* x *and* p *makes an acute angle less than* $\frac{\pi}{8}$ *with* V_p^k. *See Figure 6 for examples in two and three dimensions.*

Figure 6: Cocone for a curve point in two dimensions (left) and a surface point in three dimensions (right).

The approximation algorithm COCONESHAPE filters a subset of k simplices incident to p from the Delaunay triangulation of P. In the COCONE algorithm for surface reconstruction we selected a set of Delaunay triangles incident to p that are dual to the Voronoi edges intersected by C_p. Generalizing this idea, we compute the set of k-simplices incident to p that are dual to the $d - k$ dimensional Voronoi faces intersected by C_p. All such simplices computed over all sample

points constitute what we call the set of *candidate simplices*. We claim that all candidate simplices lie close to the sampled manifold M.

CoconeShape (P, ρ)

1   Compute V_P
2   Dimension(P, ρ, d)
3   $T := \emptyset$
4   **for all** $p \in P$
5   Let $k := dim(p)$
6   Compute V_p^k and C_p
7   Compute F, the $d - k$ dimensional Voronoi faces intersected by C_p
8     $T := T \cup dual(F)$
9   **endfor**
10   Output T

LEMMA 5.1. *Let t be a candidate simplex computed by* CoconeShape. *There is a ball of radius $O(\varepsilon)f(p)$ where p is any vertex of t.*

Proof. It follows from the definition of cocone C_p and the proof of Lemma 3.3 that any point $x \in C_p$ satisfies the condition $\angle((x - p), \mathbf{n}_p) \geq \frac{3\pi}{8} - O(\varepsilon)$ where $\mathbf{n}_p$ is an anglewise closest vector to $x - p$ in N_p. Using the contrapositive of Lemma 3.1, this implies that $\|x - p\| = O(\varepsilon)f(p)$. Since t is dual to a Voronoi face F that intersects C_p, there exists a ball centered at a point $c = C_p \cap F$ with radius $O(\varepsilon)f(p)$ which circumscribes t.

The above lemma implies that each point on an output simplex has a point p on the manifold M within a small distance of $O(\varepsilon)f(p)$. Also, each point in M has a vertex p on the output complex within $\varepsilon f(p)$ distance. These two facts imply that the Hausdorff distance between M and T is small relative to the feature size.

6 Experimental results.

We experimented with DIMENSION on three dimensional data sets. DIMENSION is implemented in C++ using the computational geometry algorithms library CGAL [25]. We found that robust Delaunay triangulation/Voronoi diagram computations in presence of degeneracies and finite precision arithmetic are absolutely necessary for valid output. To this end we used the filtered floating point arithmetic of CGAL. This simulates exact arithmetic only on a demand basis. Thus, it provides the advantage of exact arithmetic with a nominal increase in running time which is observed to be a factor of two in our case. The reported running times are due to experiments performed on a PC with 933 Mhz Pentium III processor and 512 MB main memory. We tested our implementation of DIMENSION on various datasets. See Figures 5, 8 and Table 1. It turned out that a value of $\rho = 0.3$

gives good results in practice. The results we report here are all obtained using this value.

We developed the algorithm DIMENSION for manifolds without boundaries, but tested it also on manifolds with boundaries. In three dimensions, we can have one and two manifolds without boundaries. However, three manifolds embedded in three dimensions necessarily have boundaries. The boundary of a k-manifold is a $k - 1$-manifold, i.e. the boundary of a volume is a surface. The samples on this surface have Voronoi cells elongated along the normal to the surface. Thus, DIMENSION detects these points as lying on a 2-manifold. Similarly, we observe that points on the boundary of a surface are detected as lying on a 1-manifold. See the datasets FOOT, ENGINE and BALL in Figure 5.

The shapes are approximated with our COCONESHAPE algorithm once the dimensions are detected by DIMENSION. In three dimensions, we can extract a manifold out of the candidate simplices which can be proved to be homeomorphic to the original curve or surface. Our COCONE software works on this principle to reconstruct surfaces in three dimensions [24]. With the manifold extraction step COCONESHAPE is exactly what COCONE does for surface reconstruction. Thus, COCONESHAPE acts like COCONE for samples that are assigned dimension two. It includes samples from the surfaces of the three manifolds. Figure 7 shows an example of a sample from a three manifold in three dimensions and the reconstruction of its boundary with COCONESHAPE after the manifold extraction step.

Figure 7: Output of the algorithm DIMENSION on the dataset TORUS (top) and the reconstruction of the boundary of the same dataset (bottom). This picture needs to be seen in color.

The sample points on the boundaries of the surfaces have dimension one. While computing candidate triangles COCONESHAPE chooses triangles incident to these bound-
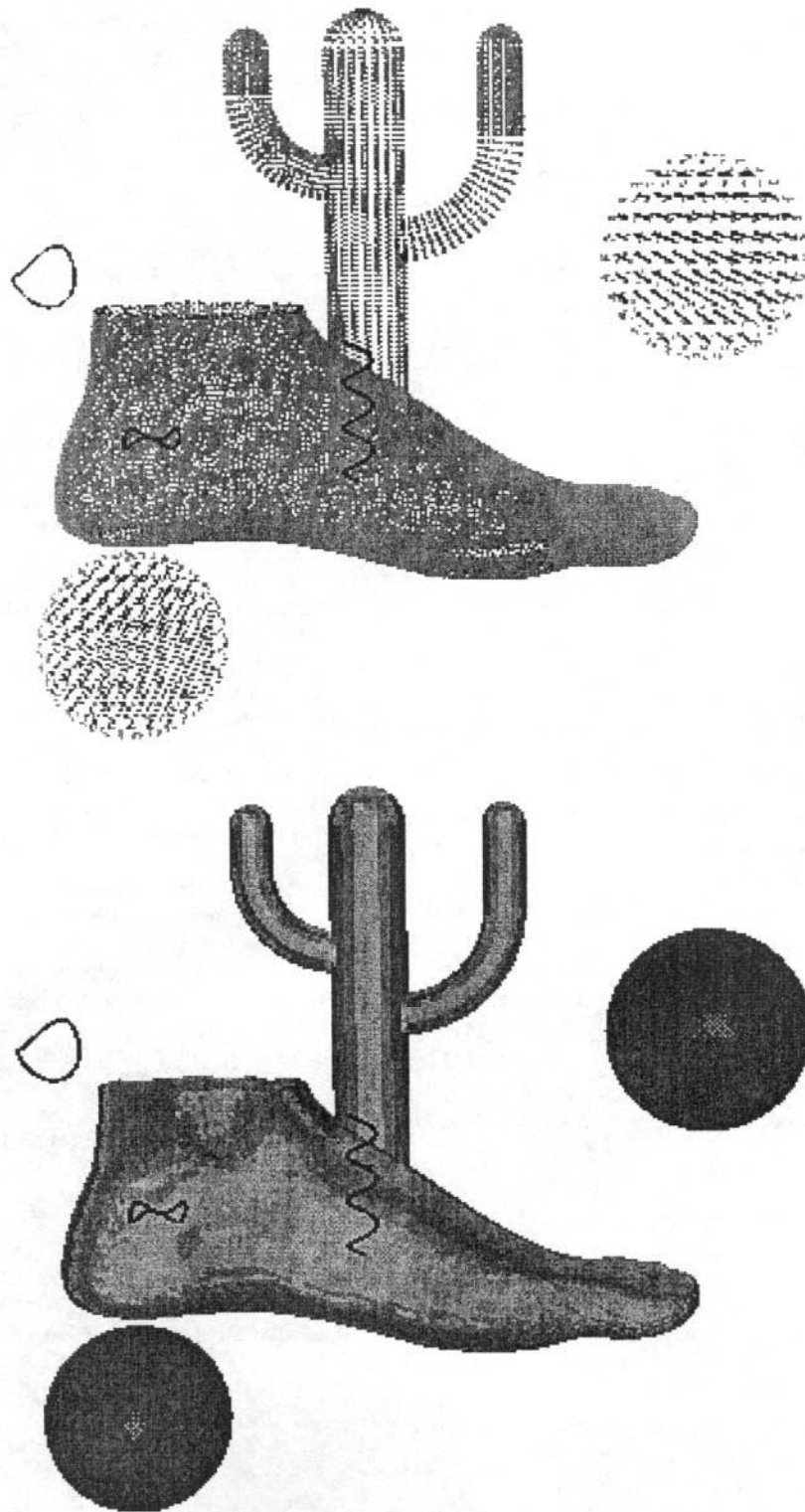
Figure 8: Output of the algorithm COCONESHAPE on the dataset SCENE (top) which has objects of different dimensions and the reconstruction of the same dataset (bottom). This picture needs to be seen in color.