

HOW MANY POINTS IN A POINT CLOUD IS SUFFICIENT FOR ACCURATE ESTIMATION OF THE CURVATURE

R. Mirzaie

Abstract. We introduce an estimator for the curvature of curves and surfaces by using finite sample points drawn from a probability distribution whose support lies on the curve or surface. First, we give a probabilistic procedure for estimating the curvature at a given point of a curve. Then, we extend it to estimate the Gaussian curvature of surfaces. In the proposed probabilistic procedures, we use a relation between the number of selected points in the point cloud and the probability that a given point has a sufficient number of nearby points. This relation allows us to control the required number of points in the point cloud.

1. Introduction

Recently, there has been significant progress in manifold learning, which has attracted much interest because of its useful applications in topological data analysis, geometric measure theory, integral geometry, image processing, and biomathematics. However, a main problem remains open: What is the best algorithm for estimating the curvature of a manifold using a given point cloud?

There are many algorithms suited for Gaussian curvature and mean curvature estimation. Almost all of these algorithms rely on estimating the intrinsic geometric properties of a surface from a polygonal mesh obtained from data. For example, a widely used approximation of the curvature of a surface is based on the fact that a surface around a point can be represented as the graph of a bivariate function, and the curvature can be approximated by the curvature of a hyperbola passing through the neighbouring points of the given point in the selected point cloud (see [12]). The authors of [1] use principal component analysis to estimate the Ricci curvature of a manifold at each point of a point cloud set, selected randomly from the points of a Riemannian submanifold of Euclidean space. The estimator of [9], for the scalar curvature, is based on the fact that the scalar curvature at a point x of a Riemannian manifold characterizes the growth rate of the volume of a geodesic ball $B(x, r)$ as r

2020 Mathematics Subject Classification: 53C21, 53C23, 62D05

Keywords and phrases: Curve; surface; curvature; manifold.

decreases. Thus, the estimation of the scalar curvature reduces to the estimation of volumes. We refer to [14] for many methods to estimate Gaussian curvature from point clouds sampled from surfaces. We also refer the reader to [13] for surveys. The authors of [7] locally fit a set of circles through the neighbouring vertices of the triangulations as a main tool for estimation.

In [2], a suitable curve is constructed at each vertex of the surface triangulation, and then the local differential structure is derived.

In the present article, we give a probabilistic procedure to estimate the curvature of a curve by using the osculating circle at a given point. Then, we use it to estimate the Gaussian curvature of surfaces.

Using the osculating circle is a traditional method for estimating the curvature of plane curves, but using it to estimate the principal curvatures of surfaces from random point clouds is new. Another novel aspect of the present article is the following observation:

Using random point clouds to estimate geometric properties with global characteristics works very well. However, when we use random point clouds to estimate geometric properties with local characteristics (for example, curvature at a point), there may be significant estimation errors. Let a be an arbitrary point on a surface M and $B = \{x_1, \dots, x_m\}$ be a subset of points on M . To estimate the curvature at the point a , the points of B that are close to a play an essential role. However, we are uncertain whether there are enough points in B close to a .

In practice, since we usually have no information about the underlying manifold, the easiest way to solve the problem is to choose a large number of points. However, the large number of points makes the computation process and the runtime of the computer program long.

To overcome these two problems, we use Theorem 2.7 and Theorem 2.13 to provide a method for finding the minimum number m of points in the point cloud. The value of m is suitable for a probability p , where p can be chosen arbitrarily close to one.

2. Results

In what follows, each curve and surface is considered to be differentiable, with finite length and area, and embedded in R^2 and R^3 , respectively. Since we intend to estimate their curvature at a generic point, we suppose that they have no boundary points.

We suppose that each sample set $B = \{x_1, \dots, x_m\}$ of points on the curve or surface is i.i.d. (independent and identically distributed) and uniformly distributed (uniform with respect to the Riemannian volume measure induced from the ambient space).

REMARK 2.1. Let $\eta > 0$ and let $\gamma : [0, \eta] \rightarrow R^2$ be a differentiable function. We call γ a plane curve, and throughout we assume that γ is one-to-one. If the length of γ is l then, after a reparametrization if necessary, we may assume that $l = \eta$.

Let $r = r(a, b, c)$ be radius of the circle generated by three nonlinear points a, b, c

belonging to γ . The curvature of γ at the point a which we denote by $\kappa(a)$ is defined by $\kappa(a)^{-1} = \lim_{(b,c) \rightarrow (a,a)} r(a,b,c)$. A circle with radius $\kappa(a)^{-1}$ tangent to γ at the point a is called the osculating circle at a . Thus, for the points b, c sufficiently close to a , $\frac{1}{r}$ is an estimation for $\kappa(a)$. If $\max\{|b-a|, |c-a|\} < \epsilon$, then we say that $\frac{1}{r}$ is a ϵ -triangle based estimation of $\kappa(a)$. Let d_1, d_2, d_3 be the length of the sides of the triangle abc and $D = \frac{d_1+d_2+d_3}{2}$. Then,

$$r = \frac{d_1 d_2 d_3}{4\sqrt{D(D-d_1)(D-d_2)(D-d_3)}}.$$

Thus, ϵ -triangle based estimation of $\kappa(a)$ is

$$\kappa(a) = \frac{4\sqrt{D(D-d_1)(D-d_2)(D-d_3)}}{d_1 d_2 d_3}.$$

In what follows, we denote by $d(x, y)$ the distance between x, y in R^n .

If S is a connected curve or a surface, then the distance between x, y in S is denoted by $d_S(x, y)$, which is equal to the infimum of the length of the curves in S joining x to y .

We denote by γ and M a differentiable plan curve with the finite length and a differentiable surface with the finite area, respectively. $B(x, r)$ is a ball with the center x and of radius $r > 0$ (in R^n). $B(r)$ will denote a ball of radius r (non-specified center). If M is a surface in R^3 and $x \in M$, we denote by $B_M(x, r)$ the geodesic ball in M with the center x and the radius r , that is $B_M(x, r) = \{y \in M : d_M(x, y) \leq r\}$. We use the symbol $vol(M)$ to denote the volume(area) of M .

REMARK 2.2. $vol(B_M(x, r))$ is a continuous function of r . Thus, for each positive number ϵ (sufficiently small), there is a number $r = r(x, \epsilon)$ such that $vol(B_M(x, r)) = \pi\epsilon^2$ (the area of the usual ball with the radius ϵ in a plane). In this case, we denote $B_M(x, r)$ by $B'(x, \epsilon)$. We use the symbol $B'(\epsilon)$, when the center x is not specified.

The following simple corollary about number of the balls covering a curve or a surface will be useful.

COROLLARY 2.3. (i) Let l be the length of a differentiable plan curve γ and let $\epsilon > 0$. Then, there is a covering of γ with a collection $\{B_1(\epsilon), \dots, B_N(\epsilon)\}$ of balls centered on γ such that $N \leq \frac{l}{\epsilon}$.

(ii) Let M be a compact surface in R^3 with the area $vol(M)$. Then, there is a covering of M with a collection $\{B'_1(\epsilon), \dots, B'_N(\epsilon)\}$ of balls of M such that $N \leq \frac{3vol(M)}{\pi\epsilon^2}$.

Proof. (i) Given a ball $B_i(\epsilon)$ centered on γ . Put $\gamma_i = \gamma \cap B_i(\epsilon)$ and let l_i be the length of γ_i . It is clear that l_i is at least equal to the diagonal of $B_i(\epsilon)$. Then, $l_i \geq 2\epsilon$. We can choose the collection of balls in such a way that for each i , $1 < i < N-1$, γ_i has overlaps at most by γ_{i-1} and γ_{i+1} . Then, $\sum_i l_i \leq 2l$ and $2l \geq \sum_i l_i \geq 2N\epsilon$.

(ii) It is possible to choose a collection $\{B'_1(\epsilon), \dots, B'_N(\epsilon)\}$ of balls of M covering M in such a way that each point $x \in M$ belongs to at most three balls. Then, $\sum_i vol(B'_i(\epsilon)) \leq 3vol(M)$. Thus, $N\pi\epsilon^2 \leq 3vol(M)$. which gives the result. $\square$

LEMMA 2.4. Let γ be a curve of length l , and let $\gamma_1, \dots, \gamma_n$ be sub-curves of γ with lengths l_i . Suppose there exists a constant $a > 0$ such that $l_i > a$ for all $i = 1, \dots, n$.

Let $B = \{x_1, \dots, x_m\}$ be a set of i.i.d. points uniformly distributed on γ , and let $0 < p < 1$. If

$$m > \frac{2(\log(1-p) - \log n)}{\log\left(1 - \left(\frac{a}{l}\right)^2\right)} + 1,$$

then, with probability at least p , each sub-curve γ_i contains at least two distinct points of B , for all $i \in \{1, \dots, n\}$.

Proof. Consider γ as a map $\gamma : L = [0, l] \rightarrow R^2$ (see Remark 2.1). Each point $x = \gamma(t)$ on γ corresponds to a parameter $t \in L$, and each sub-curve γ_i corresponds to a subinterval $L_i \subset L$ of length l_i . Let $y_1, y_2 \in B$ such that $y_1 = \gamma(t_1), y_2 = \gamma(t_2)$. The event that $\{y_1, y_2\} \subset \gamma_i$ is equivalent to the event that $(y_1, y_2) \in \gamma_i \times \gamma_i$, which corresponds to the event that $(t_1, t_2) \in L_i \times L_i$, so we have

$$P(\{y_1, y_2\} \subset \gamma_i) = P(\{(t_1, t_2) \in L_i \times L_i\}) = \frac{l_i^2}{l^2}. \quad (1)$$

In what follows, for two events E and F , we denote by $E \cup F$ the event ' E or F '. Let F_i denote the event that γ_i contains at least a couple of points of B . Let $k = \lfloor \frac{m}{2} \rfloor$ and put $A_1 = \{x_1, x_2\}$, $A_2 = \{x_3, x_4\}, \dots, A_k = \{x_{k-1}, x_{2k}\}$. Denote by $\hat{A}_j$ the event that $A_j \subset \gamma_i$. Then,

$$P(F_i) \geq P(\hat{A}_1 \cup \hat{A}_2 \cup \dots \cup \hat{A}_k). \quad (2)$$

It is clear that the events $\hat{A}_1, \dots, \hat{A}_k$ are independent. Thus, for $1 \leq j_1 < j_2 < \dots < j_s \leq k$, we have

$$P(\hat{A}_{j_1} \cap \hat{A}_{j_2} \cap \dots \cap \hat{A}_{j_s}) = P(\hat{A}_{j_1})P(\hat{A}_{j_2}) \cap \dots \cap P(\hat{A}_{j_s}). \quad (3)$$

Since by (1), $P(\hat{A}_j) = \frac{l_i^2}{l^2}$, then (3) implies

$$P(\hat{A}_{j_1} \cap \hat{A}_{j_2} \cap \dots \cap \hat{A}_{j_s}) = \frac{l_i^{2s}}{l^{2s}}. \quad (4)$$

Also, we have

$$P(\hat{A}_1 \cup \dots \cup \hat{A}_k) = \sum_{j=1}^k P(\hat{A}_j) - \sum_{j_1 < j_2} P(\hat{A}_{j_1} \cap \hat{A}_{j_2}) + \dots + (-1)^{k+1} P(\hat{A}_1 \cap \dots \cap \hat{A}_k). \quad (5)$$

By using (2), (4) and (5), we get

$$P(F_i) \geq k \frac{l_i^2}{l^2} - \binom{k}{2} \frac{l_i^4}{l^4} + \binom{k}{3} \frac{l_i^6}{l^6} - \dots + (-1)^{k+1} \binom{k}{k} \frac{l_i^{2k}}{l^{2k}}. \quad (6)$$

Using the binomial formula $(1-x)^k = \sum_{j=0}^k \binom{k}{j} (-1)^j x^j$, (6) can be written as

$$P(F_i) \geq 1 - \left(1 - \frac{l_i^2}{l^2}\right)^k.$$

Since by assumption, $l_i > a$, then

$$P(F_i) \geq 1 - \left(1 - \frac{a^2}{l^2}\right)^k. \quad (7)$$

Now, denote by E_i the event that no couples of points of B belongs to γ_i . Then,

$$P(E_i) = 1 - P(F_i). \quad (8)$$

The probability that for some i , γ_i contains no couples of B is equal to $P(\bigcup_i E_i)$. What we are looking is $1 - P(\bigcup_i E_i)$. By (7) and (8) we get

$$1 - P(\bigcup_i E_i) \geq 1 - \sum_i P(E_i) \geq 1 - n(1 - (\frac{a}{l})^2)^k.$$

A simple computation shows that $1 - n(1 - (\frac{a}{l})^2)^k > p$ is equivalent to $k > \frac{\log(1-p) - \log n}{\log(1 - (\frac{a}{l})^2)}$. Since $k = \lfloor \frac{m}{2} \rfloor$, then $m > \frac{2(\log(1-p) - \log n)}{\log(1 - (\frac{a}{l})^2)} + 1$ implies $1 - P(\bigcup_i E_i) > p$. $\square$

REMARK 2.5. The criterion of the previous lemma, by a similar proof, can be mentioned in a general form as follows:

Let $A_1, \dots, A_n$ be subsets of a measurable set X and $B = \{x_1, \dots, x_m\} \subset X$. If the probability that a point of X belongs to A_i is at least α and $m > \frac{2(\log(1-p) - \log n)}{\log(1 - \alpha^2)} + 1$, then with the probability at least p , each A_i contains two different points of B .

DEFINITION 2.6. Let γ be a connected curve and x, x', x'' be three different points on γ . We say that x is between x', x'' and we denote by $[x', x, x'']$, if for a parametrization $\gamma(t)$, $t \in I \subset \mathbb{R}$, of γ , $x' = \gamma(t')$, $x = \gamma(t)$, $x'' = \gamma(t'')$, $t' < t < t''$ or $t'' < t < t'$.

Suppose that we have a point cloud $\{x_1, \dots, x_m\}$ of the points on a curve γ , drawn in i.i.d fashion. The following theorem proposes a method to estimate the curvature with a suitable confidence by choosing sufficiently big number of points. If x, y are two points on γ , we denote by $d_\gamma(x, y)$ the length of the γ between x and y which by definition is equal to the length of the shortest path on γ joining x to y . It is clear that $d(x, y) \leq d_\gamma(x, y)$.

THEOREM 2.7. Let $\epsilon > 0$, and γ be a plan curve of length l with $l \geq \epsilon$. Let $B = \{x_1, \dots, x_m\}$ be a set of i.i.d. points samples uniformly from γ . If

$$m > \frac{2(\log(1-p) - \log l + \log(2\epsilon))}{\log(1 - (\frac{\epsilon}{l})^2)} + 1,$$

then, with probability at least p , for each point $x \in \gamma$ there exists x', x'' in B such that $[x', x, x'']$ and $\max\{d(x, x'), d(x, x'')\} < \epsilon$. Consequently, $r^{-1}(x, x', x'')$ is an ϵ -triangle based estimation of $\kappa(x)$.

Proof. Fix $x \in \gamma$, and let $\{A_i = B_\gamma(y_i, \frac{\epsilon}{2})\}_{i=1}^n$ be a collection of γ -balls covering γ , with $y_1 = x$. Without loss of generality, we may assume that for all i , $l_i := l(A_i) = \epsilon$. Indeed, if γ is an open bounded curve and the distance between x and one of its endpoints e is $\delta < \frac{\epsilon}{2}$, we may extend γ beyond e to a curve $\tilde{\gamma}$ such that $l(\tilde{\gamma}) = l(\gamma) + (\epsilon - \delta)$. This modification does not affect the probabilistic estimates below, and thus assuming $l(A_i) = \epsilon$ imposes no restriction.

Since the sampling is uniform along arc length, the probability that a randomly chosen point of γ lies in A_i equals $\alpha = \frac{\epsilon}{l}$.

By Remark 2.5, if

$$m > \frac{2(\log(1-p) - \log n)}{\log\left(1 - \left(\frac{\epsilon}{l}\right)^2\right)} + 1, \quad (9)$$

then, with probability at least p , there exist two sampled points $x_\alpha, x_\beta \in \{x_1, \dots, x_m\}$ that belong to A_1 .

If the points are ordered along γ as $[x_\alpha, x, x_\beta]$, we are done. Otherwise, either $[x, x_\alpha, x_\beta]$ or $[x, x_\beta, x_\alpha]$ holds. Assume $[x, x_\alpha, x_\beta]$; the other case is analogous.

Let $z \in \gamma$ be the point such that $[z, x, x_\alpha]$ and $d_\gamma(z, x) = \epsilon$. Applying the same argument, we conclude that there exist two points $x''', x'''' \in \{x_1, \dots, x_m\}$ lying in the ball $B_\gamma(z, \frac{\epsilon}{2})$. Clearly, $x''', x'''' \in B_\gamma(x, \epsilon)$ and $[x''', x, x_\alpha]$.

Finally, since $\{A_i\}_{i=1}^n$ covers γ , we must have $n \geq \frac{l}{2\epsilon}$. Substituting this bound into (9) yields the desired result. $\square$

Let $e = (x_0, y_0)$ be a unit vector at the origin of R^2 . The following subset of R^2 is called a standard cone with the radius r , angle θ , and direction e .

$$c_0(e, \theta, r) = \{t(x_0 \cos \psi - y_0 \sin \psi, x_0 \sin \psi + y_0 \cos \psi) : t \in [0, r], \psi \in [-\frac{\theta}{2}, \frac{\theta}{2}]\}.$$

We recall that if M is a surface and $x \in M$, then the exponential map $\exp_x : T_x M \rightarrow M$, defines a map from $T_x M$ (the vector space all vectors tangent to M at x) to M . For each vector $v \in T_x M$, if γ_v is the unit speed geodesic in M starting from x , then $\exp_x(v) = \gamma_v(1)$.

DEFINITION 2.8. Let M be a surface, $x \in M$ and e be a unit vector in $T_x M (\simeq R^2)$. Consider the exponential map $\exp_x : T_x M \rightarrow M$. The image of a standard cone under the exponential map is called a cone with the vertex x in M , i.e. $C = \text{cone}(x, e, \theta, r) = \exp_x(c_0(e, \theta, r))$. We denote the $\text{cone}(x, -e, \theta, r)$ by $-C$. We say that two points x_1, x_2 of M are well ordered with respect to $\text{cone}(x, e, \theta, r)$ if there is a differentiable curve $\gamma : I \rightarrow C \cup (-C) \subset M$ such that $[x_1, x, x_2]$ on γ .

REMARK 2.9 ([11]). Let M be a surface and $x \in M$. Consider a vector N_x which is normal to M at x . Planes containing N_x are called normal planes at x . The intersection of a normal plane and M will form a curve γ , called a normal section, and the curvature of this curve is the normal curvature. The maximum and minimum values of the normal curvatures, say κ_1, κ_2 , are called the principal curvatures. The normal sections γ_1 and γ_2 whose curvatures are κ_1 and κ_2 are called the principal curves at x . There are (unit) vectors tangent to M at x such that e_1 is normal to e_2 and γ_1, γ_2 are tangent to e_1, e_2 at x . e_1 and e_2 are called the principal vectors. The (Gaussian) curvature of M at x is the product $\kappa = \kappa_1 \kappa_2$.

DEFINITION 2.10. Let e_1 and e_2 be the principal vectors of M at x . The cones $C = \text{cone}(x, e_1, \theta, r)$ and $D = \text{cone}(x, e_2, \theta, r)$ are called the principal (θ, r) -cones at x . We say that four distinct points x_1, x_2, x_3, x_4 are (θ, r) -principal ordered at x , if two points are well ordered with respect to C and the other two points are well ordered with respect to D .

THEOREM 2.11. *Let $\epsilon, \theta > 0$, and let $M \subset \mathbb{R}^3$ be a surface of area s . Let $B = \{x_1, \dots, x_m\}$ be a set of i.i.d. points sampled uniformly from M . If*

$$m \geq \log\left(\frac{1 - p^{\frac{\pi\epsilon^2}{12s}}}{4}\right) \left(\log\left(1 - \frac{\theta\epsilon^2}{2s}\right)\right)^{-1},$$

then, with probability at least p , for every point $b \in M$ there exist points $x_{11}, x_{12}, x_{21}, x_{22} \in B$ that are (θ, ϵ) -principal ordered at b .

Proof. Consider a covering of M by the $\frac{\epsilon}{2}$ -balls of M , say $\{B'(b_1, \frac{\epsilon}{2}), \dots, B'(b_n, \frac{\epsilon}{2})\}$, $b_1 = b$. By Corollary 2.3, we may choose such a covering so that $n \leq \frac{3s}{\pi(\frac{\epsilon}{2})^2} = \frac{12s}{\pi\epsilon^2}$.

For each i , denote by C_i and D_i the $(\theta, \frac{\epsilon}{2})$ -principal cones at b_i . The area of C_i (and similarly of D_i), denoted by $\mu(C_i)$, is at least the area of a standard $(\theta, \frac{\epsilon}{2})$ -cone in $\mathbb{R}^2$. Hence, $\mu(C_i) \geq \frac{\theta}{2}\epsilon^2$, $\mu(D_i) \geq \frac{\theta}{2}\epsilon^2$. Therefore, the probability that a randomly chosen point of M lies in C_i satisfies $\frac{\mu(C_i)}{s} \geq \frac{\theta\epsilon^2}{2s}$.

Let E_i be the event that C_i contains no points of B . Then $\mathbb{P}(E_i) \leq \left(1 - \frac{\theta\epsilon^2}{2s}\right)^m$. The same estimate holds for the cones $-C_i$, D_i , and $-D_i$. Thus, for a fixed index i , the probability that at least one of the four sets $C_i, -C_i, D_i, -D_i$ contains no point of B is at most $4\left(1 - \frac{\theta\epsilon^2}{2s}\right)^m$. Consequently, the probability that all four sets contain at least one point of B is at least $1 - 4\left(1 - \frac{\theta\epsilon^2}{2s}\right)^m$.

Therefore, the probability P of the event that, for all i , the sets $C_i, -C_i, D_i$, and $-D_i$ contain a point of B satisfies

$$P \geq \left(1 - 4\left(1 - \frac{\theta\epsilon^2}{2s}\right)^m\right)^n.$$

Hence, $P \geq p$ provided that

$$\left(1 - 4\left(1 - \frac{\theta\epsilon^2}{2s}\right)^m\right)^n \geq p,$$

which is equivalent to

$$m \log\left(1 - \frac{\theta\epsilon^2}{2s}\right) \leq \log\left(\frac{1 - p^{1/n}}{4}\right).$$

Since θ and ϵ are small, we obtain

$$m \geq \log\left(\frac{1 - p^{1/n}}{4}\right) \left(\log\left(1 - \frac{\theta\epsilon^2}{2s}\right)\right)^{-1}.$$

Using the bound $n \leq \frac{12s}{\pi\epsilon^2}$, we have

$$\log\left(\frac{1 - p^{\frac{\pi\epsilon^2}{12s}}}{4}\right) \left(\log\left(1 - \frac{\theta\epsilon^2}{2s}\right)\right)^{-1} \geq \log\left(\frac{1 - p^{1/n}}{4}\right) \left(\log\left(1 - \frac{\theta\epsilon^2}{2s}\right)\right)^{-1}.$$

Therefore, $P \geq p$ whenever

$$m \geq \log\left(\frac{1 - p^{\frac{\pi\epsilon^2}{12s}}}{4}\right) \left(\log\left(1 - \frac{\theta\epsilon^2}{2s}\right)\right)^{-1}. \quad \square$$

REMARK 2.12 (curvature estimation on a surface). Let M be a surface and $b \in M$, and let $x_{11}, x_{12}, x_{21}, x_{22}$ be (θ, ϵ) -ordered at p such that x_{11}, x_{12} be well ordered with respect to $\text{cone}(b, e_1, \theta, \epsilon)$ and x_{21}, x_{22} be well ordered with respect to $\text{cone}(b, e_2, \theta, \epsilon)$. Let β_1 be a broken-geodesic in M (probably broken at b) joining x_{11} to b and b to x_{12} (i.e., $[x_{11}, b, x_{12}]$ on β_1). Similarly consider the broken-geodesic β_2 for x_{21}, x_{22} . β_1 and β_2 are estimations for the principal curves at b :

$$\kappa_1(b) \simeq (r(x_{11}, b, x_{12}))^{-1} \quad \text{and} \quad \kappa_2(b) \simeq (r(x_{21}, b, x_{22}))^{-1}.$$

Therefore, $\kappa(b) \simeq (r(x_{11}, b, x_{12}))^{-1}(r(x_{21}, b, x_{22}))^{-1}$.

Now, by Theorem 2.11 and Remark 2.12, we get the following theorem.

THEOREM 2.13. *Let $\epsilon, \theta > 0$ and M be a surface in R^3 with the area s . If we choose a subset $B = \{x_1, \dots, x_m\}$ of points, i.i.d. uniformly distributed on M , in such way that*

$$m \geq \log\left(\frac{1 - p^{\frac{\pi\epsilon^2}{12s}}}{4}\right) (\log(1 - \frac{\theta\epsilon^2}{2s}))^{-1},$$

then with the probability at least p , for each point $b \in M$, there are points $x_{11}, x_{12}, x_{21}, x_{22}$ in B such that $\kappa(b) \simeq (r(x_{11}, b, x_{12}))^{-1}(r(x_{21}, b, x_{22}))^{-1}$.

3. Probabilistic procedures

There are many algorithms for estimation of the length or area of the curves or surfaces from a point cloud, which can be used to estimation of l or s in the following probabilistic procedures.

3.1 Curvature of a curve at a given point

1. Input the point a , the positive (small) number ϵ , and the expected probability p .
2. Input the point cloud $B = \{x_1, \dots, x_m\}$.
3. Compute the length of the curve l . Put $\epsilon_1 = \min\{l, \epsilon\}$.
4. Compute λ : $\lambda = \frac{2(\log(1-p) - \log l + \log(2\epsilon_1))}{\log(1 - (\frac{\epsilon_1}{l})^2)} + 1$.
5. If $m > \lambda$, then go to 7.
6. . Add more points to B , then go to 3.
7. Find y_1, y_2 in B such that $[y_1, a, y_2]$ and $\max\{d(a, y_1), d(a, y_2)\} < \epsilon_1$.
8. Compute the ϵ_1 triangle based estimation of $\kappa(a) = r^{-1}(y_1, a, y_2)$ (since $\epsilon \leq \epsilon_1$, it is also ϵ triangle based estimation).

3.2 Principal curvature of a surface at a point

Following Remark 2.12 and Theorem 2.13, denote the point cloud by C . For estimation of the principal curvatures, first step is to find the best estimate for β_1 and its curvature at the point a . Consider a point q in C , near a . Let q' be a point in C with the property that $d(q, a) + d(q', a) - d(q, q')$ is equal to the minimum of the following set: $\{d(y_1, a) + d(y_2, a) - d(y_1, y_2) : y_1, y_2 \in C\}$. We call q' a C -conjugate of q . If β is a geodesic joining q to a and a to q' , then its curvature is almost equal to $r^{-1}(q, a, q')$. Thus, we have $\kappa_1(a) = \min\{r^{-1}(c, a, c') : c \in C\}$.

For estimation of $\kappa_2(a)$, note that β_2 is included in a plane normal to the plane which contains β_1 . Thus, if q belongs to β_1 , a point b belongs to β_2 if the vector ba is parallel to $qa \times q'a$. Thus, a point b in the point cloud is almost on β_2 if the following number is maximal

$$\frac{|\langle ba, qa \times q'a \rangle|}{|ba||qa \times q'a|}, \quad b \in C. \quad (10)$$

Thus, the second principal curvature at a is estimated as:

$$\kappa_2(a) = \min\{r^{-1}(b, a, b') : b \in C \text{ with the property (10)}\}.$$

Therefore, we can propose the following probabilistic procedure for estimation of the principal curvatures and the Gaussian curvature of the surfaces using a point cloud on the surface.

1. Input the point a , small number ϵ , and the expected probability p .
2. Input the point cloud $C = \{x_1, \dots, x_m\}$.
3. Compute the area of the surface s .
4. Compute $\lambda : \lambda = \log \left(\frac{1 - p^{\frac{\pi \epsilon^2}{12s}}}{4} \right) \left(\log \left(1 - \frac{\epsilon^3}{20s} \right) \right)^{-1}$.

In Theorem 2.13, for simplicity take $\theta = \frac{\epsilon}{10}$.

5. If $m > \lambda$, then go to 7.
6. Add more points to C , then go to 3.
7. Find the collection D of all points $\{y\}$ of C such that $d(y, a) < \epsilon$.

For each $y \in D$, put:

$$\begin{aligned} E_y &= \{d(y, a) + d(c, a) - d(y, c) : c \in C\}, \\ E'_y &= \{c \in C : d(y, a) + d(c, a) - d(y, c) = \min(E_y)\}, \\ E''_y &= \{r^{-1}(y, a, c) : c \in E'_y\}, \quad k_y = \min(E''_y). \end{aligned}$$

The first principal curvature is: $\kappa_1(a) = \min\{k_y : y \in D\}$.

8. Select a point $q \in C$ with the property that $r^{-1}(q, a, q') = \kappa_1(a)$. Put

$$\begin{aligned} F &= \left\{ \frac{|\langle ba, qa \times q'a \rangle|}{|ba||qa \times q'a|} : b \in C, b \neq a, qa \neq q'a \neq 0 \right\}, \\ F' &= \left\{ b \in C : \frac{|\langle ba, qa \times q'a \rangle|}{|ba||qa \times q'a|} = \max F \right\}, \end{aligned}$$

$$F'' = \{r^{-1}(b, a, b') : b \in F'\}.$$

The second principal curvature is: $\kappa_2(a) = \min(F'')$.

9. Return Gaussian curvature, and Mean curvature at a respectively, as: $\kappa_1(a)\kappa_2(a)$, $\frac{\kappa_1(a) + \kappa_2(a)}{2}$.

REMARK 3.1. In estimating curvature using a finite set of sample points, in general, it is not possible to establish an upper bound for the estimation error. In fact, for each finite set $B = \{x_1, \dots, x_n\}$ of the points in R^m and each positive number q , there is a submanifold M of R^m , such that $B \subset M$ and the sectional curvature of M at some point exceeds q . This limitation is inherent in all estimation methods based solely on a finite set of sample points from a manifold. Nevertheless, our probabilistic procedures appear to be effective in practical applications.

REMARK 3.2. All necessary code was prepared by the author (using Python). As shown in Tables 1 and 2, the numerical results confirm that the probabilistic procedures provide accurate and reliable estimates across a variety of test cases. Using the code, a finite set of points can be randomly sampled and employed, together with our probabilistic procedures, to estimate curvatures. In our experiments, the parameters ϵ and p were set to 0.1.

Curve	Point	Curvature at point	Estimated curvature at point
$x^2 + y^2 = 25$	(4,3)	0.200000	0.200008
$y = x^3 + 2x$	(1,3)	0.004333	0.004348
$y = 5x^4 + 2x$	(1,7)	0.005617	0.005762
$x^2 + y^2 = 0.0025$	(0,0.05)	20.000000	19.999999
$y = \ln(x^4 + 1) + 5x^2$	(0,5)	10.000000	9.740785

Table 1: Estimation of the curvature of some curves with $\epsilon = 0.1$ and $p = 0.1$

Surface	Point	Curvature at point	Estimated curvature at point
$x^2 + y^2 + z^2 = 25$	(0,0,5)	0.040000	0.039999
$z = x^3 + 2x + y^2 + y$	(1,2,9)	0.004618	0.004204
$z = \sqrt{x^2 + y^2}$	(1, 2, $\sqrt{5}$)	0.027777	0.025705
$z = x^2 + 0.25y^2$	(1,2,2)	0.027777	0.0240704

Table 2: Estimation of the Gaussian curvature of some surfaces with $\epsilon = 0.1$ and $p = 0.1$

REFERENCES

- [1] A. G. Ache, M. W. Warren, *Ricci curvature and the manifold learning problem*, Adv. Math., **21** (2019), 14–66.
- [2] P. J. Besl, R. Jain, *Invariant surface characteristics for 3d object recognition in range images*, Comput. Vis. Graph. Image Process., **33** (1986), 33–80.
- [3] P. Bubenik, M. Hull, D. Patel, B. Whittle, *Persistent homology detects curvature*, Inverse Probl., **36** 025008(2), (2020).
- [4] F. Cazals, M. Pouget, *Estimating differential quantities using polynomial fitting of osculating jets*, Comput. Aided Geom. Des., **22** (2005), 121–146.
- [5] Q. Charles, S. Hao, M. Kaichun, J. G. Leonidas, *Deep learning on point sets for 3d classification and segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017).
- [6] J. Charton et al, *Robust curvature estimation on surface triangulation*, Iadis International Conference Computer Graphics, Visualization, Computer Vision and Image Processing, (2013), 43–45.
- [7] X. Chen, F. Schmitt, *Intrinsic surface properties from surface triangulation*, European Conference on Computer Vision, (1992), 739–743.
- [8] P. Guerrero, Kleiman, M. Ovsjanikov, N. J. Mitra, *Learning local shape properties from raw point clouds*, Comput. Graph. Forum, **37** (2018), 75–85.
- [9] A. Hickok, A. J. Blumberg, *An intrinsic approach to scalar curvature estimation for point clouds*, arXiv:2308.02615v1 [stat.ML] 4 August 2023.
- [10] J. L. Maltret, M. Daniel, *Discrete curvatures and applications: a survey*, Rapport de recherche LSIS.RR.2002.002, Laboratoire des Sciences de l’Information et des Systemes, (2002).
- [11] B. O’Neil, *Elementary differential geometry*, Accademic Press, (1997).
- [12] S. Petitjean, *A survey of methods for recovering quadrics in triangle meshes*, ACM Computing Surveys, **2** (2002), 1–61.
- [13] T. Surazhsky et al, *A comparison of gaussian and mean curvatures estimation methods on triangular meshes*, ICRA, (2003), 1021–1026.
- [14] J. Spang, *Robust estimation of surface curvature information from point cloud data*, arXiv:2306.00299v1 [cs.CG] 1 JUN 2023.

(received 09.04.2025; in revised form 16.02.2026; available online 19.10.2026)

Department of Pure Mathematics, Faculty of Science, Imam Khomeini International University, Qazvin, Iran

E-mail: r.mirzaei@sci.ikiu.ac.ir

ORCID iD: <https://orcid.org/0000-0002-9480-0222>