International Journal of Computational Geometry & Applications

Estimating Surface Normals in Noisy Point Cloud Data*

NILOY J. MITRA
Stanford Graphics Laboratory, James H. Clark Center, Room S297, 318 Campus Drive, Stanford, CA 94305 USA
niloy@stanford.edu

AN NGUYEN
Stanford Graphics Laboratory, James H. Clark Center, Room S297, 318 Campus Drive, Stanford, CA 94305 USA
anguyen@cs.stanford.edu

LEONIDAS GUIBAS
Stanford Graphics Laboratory, James H. Clark Center, Room S293, 318 Campus Drive, Stanford, CA 94305 USA
guibas@cs.stanford.edu

Received 7 July 2003
Revised 16 November 2003
Communicated by Joe Mitchell

In this paper we describe and analyze a method based on local least square fitting for estimating the normals at all sample points of a point cloud data (PCD) set, in the presence of noise. We study the effects of neighborhood size, curvature, sampling density, and noise on the normal estimation when the PCD is sampled from a smooth curve in $\mathbb{R}^2$ or a smooth surface in $\mathbb{R}^3$, and noise is added. The analysis allows us to find the optimal neighborhood size using other local information from the PCD. Experimental results are also provided.

Keywords: normal estimation; noisy point cloud data; eigen analysis; neighborhood size estimation.

1. Introduction

Modern range sensing technology enables us to make detailed scans of complex objects generating point cloud data (PCD) consisting of millions of points. The data acquired is usually distorted by noise arising out of various physical measurement processes and limitations of the acquisition technologies.

*A preliminary version of this paper appeared in the Proc. of the 19th ACM Symp. on Computational Geometry, 2003. The work was supported by NSF CARGO grant 0138456 and a Stanford Graduate Fellowship.

The traditional way to use PCD is to reconstruct the underlying surface model represented by the PCD, for example as a triangle mesh, and then apply well known methods on that underlying manifold model. However, when the size of the PCD is large, such methods may be expensive. To do surface reconstruction on a PCD, one would first need to filter out the noise from the PCD, usually by some smoothing filter [15]. Such a process may remove sharp features as well, and this may be undesirable. A reconstruction algorithm, such as those proposed by Amenta *et al.* [2,4], then computes a mesh that approximates the noise free PCD. Both the smoothing and the surface reconstruction processes may be computationally expensive. For certain applications like rendering or visualization, such a computation is often unnecessary and direct rendering of PCD has been investigated by the graphics community [17,18].

Alexa *et al.* [1] and Pauly *et al.* [17] have proposed to use PCD as a new modeling primitive. Algorithms for such a paradigm often require information about the normal at each of the points. For example, normals are used in rendering PCD, making visibility computation, answering inside–outside queries, etc. Also some curve (or surface) reconstruction algorithms [6,7] need to have the normal estimates as a part of the input data.

The normal estimation problem has been studied by various communities such as computer graphics, image processing, and mathematics, but mostly in the case of manifold representations of the surface. We would like to estimate the normal at each point in a PCD, given to us only as an unstructured set of points sampled from a smooth curve in $\mathbb{R}^2$ or a smooth surface in $\mathbb{R}^3$, without any additional manifold structure.

Hoppe *et al.* [13] proposed an algorithm where the normal at each point is estimated as the normal to the fitting plane obtained by applying the total least square method to the *k-nearest neighbors* of the point. This method is robust in the presence of noise due to the inherent low pass filtering. In this algorithm, the value of k is a parameter and is chosen manually based on visual inspection of the computed estimates of the normals, and different trial values of k may be needed before a good selection of k is found. Furthermore, the same value of k is used for the normal estimation at all points in the PCD ignoring the variation in curvature and sampling density along the PCD.

We note that the accuracy of the normal estimation using a total least square method depends on (1) the noise in the PCD, (2) the curvature of the underlying manifold, (3) the density and the distribution of the samples, and (4) the neighborhood size used in the estimation process. In this paper, we make precise such dependencies and study the contribution of each of these factors on the normal estimation process. This analysis allows us to find the optimal neighborhood size to be used in the method. The neighborhood size can be computed adaptively at each point based on local information, given some estimates about the noise, the local sampling density, and bounds on the local curvature. The computational complexity of estimating all normals of a PCD with m points is only $O(m \log m)$.

1.1. *Related Work*

In this section, we summarize some of the previous works that are related to the computation of the normal vectors of a PCD. Many current surface reconstruction algorithms [2,4,10] can either compute the normals as part of the reconstruction, or the normals can be trivially approximated once the surface has been reconstructed. As the algorithms require that the input is noise free, a raw PCD with noise needs to go through a smoothing process before these algorithms can be applied.

Fleishman *et al.* [9] and Jones *et al.* [14] have independently proposed the use of edge-preserving filters for designing fast feature-preserving mesh denoising techniques. These methods can be easily extended to PCDs using neighborhood graphs to approximate connectivity information. However, these denoising algorithms have parameters that have to be manually adjusted for good results.

The work of Hoppe *et al.* [13] for surface reconstruction suggests a method to compute the normals for the PCD. The normal estimate at each point is done by fitting a least square plane to its k-nearest neighbors. The value of k is selected experimentally. The same approach has also been adopted by Zwicker *et al.* [20] for local surface estimation. Higher order surfaces have been used by Welch and Witkin [19] for local parameterization. However, as pointed out by Amenta and Bern [3] such algorithms can fail even in cases with arbitrarily dense set of samples. This problem can be resolved by assuming uniformly distributed samples which prevents errors resulting from biased fits. As noted before, all these algorithms work well even in presence of noise because of the inherent filtering effect. The success of these algorithms depends largely on selecting a suitable value for k, but usually little guidance is provided for the selection of this crucial parameter.

1.2. *Paper Overview*

In section 2, we study the normal estimation for PCD which are samples of curves in $\mathbb{R}^2$, and the effects of different parameters of the normal estimation algorithm on the resulting error. In section 3, we derive similar results for PCD which come from surfaces in $\mathbb{R}^3$. In section 4, we provide simulations to illustrate the results obtained in sections 2 and 3. We also provide an algorithm for using our theoretical results on practical data. We conclude in section 5.

2. Normal Estimation in $\mathbb{R}^2$

In this section, we consider the problem of approximating the normals for a point cloud in $\mathbb{R}^2$. Given a set of points, which are noisy samples of a smooth curve in $\mathbb{R}^2$, we can use the following method to estimate the normal to the curve at each of the sample points. For each point O, we find all the points of the PCD inside a circle of radius r centered at O, and then compute the total least square line fitting those points. The normal to the fitting line gives us an approximation to the undirected normal of the curve at O. Note that the orientation of the normals is a global

property of the PCD and thus cannot be computed locally. Once all the undirected normals are computed, an approximation algorithm like the one suggested by Hoppe *et al.* [13] can be applied to obtain a consistent global orientation for all the normals. For the rest of this paper, we only consider the computation of the undirected normals.

We analyze the error of the approximation when the noise is small and the sampling density is high enough around O. Under these assumptions, which we will make precise later, the computed normal approximates well the true normal. We observe that if r is large, the neighborhood of the point cannot be well approximated by a line in the presence of local curvature in the data and we may incur large error. On the other hand, if r is small, the noise in the data can result in significant estimation error. We aim for the optimal r that strikes a balance between these opposing sources of error.

2.1. *Modeling*

Without loss of generality, we consider O as the origin, and the y-axis to be along the normal to the curve at O. We assume that the points of the PCD around O come from a segment of a smooth curve (a 1-D topological disk) of bounded curvature. More precisely, we assume that the segment of the curve near O is locally a graph of a single valued C^2 continuous function $y = g(x)$ defined over some interval R containing $[-r, r]$, and that $|g''(x)| < \kappa$ for all $x \in R$ where κ is some positive constant.

Let $\mathbf{p}_i = (x_i, y_i)$ for $1 \leq i \leq k$ be the points of the PCD that lie inside a circle of radius r centered at O. We assume the following probabilistic model for the points $\mathbf{p}_i$. Assume that x_i's are instances of a random variable X taking values in $[-r, r]$, and $y_i = g(x_i)+n_i$, where the noise terms n_i are independent instances of a random variable N. X and N are assumed to be independent. We assume that the noise N has zero mean and standard deviation σ_n, and takes values in $[-n, n]$.

Using Taylor series, there are numbers ψ_i, $1 \leq i \leq k$ such that $g(x_i) = g''(\psi_i)x_i^2/2$ with $|\psi_i| \leq |x_i| \leq r$. Let $\gamma_i = g''(\psi_i)$, then the bounded curvature assumption implies that $|\gamma_i| \leq \kappa$.

Note that if κr is large, even when there is no noise in the PCD, the normal to the best fit line may not be a good approximation to the tangent as shown in Figure 1. Similarly, if σ_n/r is large and the noise is biased, this normal may not be a good approximation even if the original curve is a straight line, see Figure 2. In order to keep the normal approximation error low, we assume *a priori* that κr and σ_n/r are sufficiently small.

We assume that the samples are *evenly distributed*; there is a radius $r_0 > 0$ (possibly dependent on O) so that any neighborhood of size r_0 in R contains at least two points of the x_i's but no more than some small constant number of them. We observe that the number of points k inside any disk of radius r is bounded from above by $\Theta(1)r\rho$, and also is bounded from below by another $\Theta(1)r\rho$, where ρ is

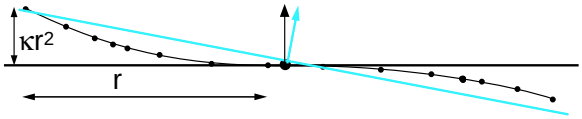


Fig. 1. Curvature causes error in the estimated normal

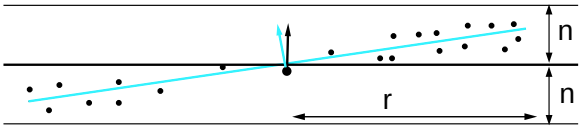


Fig. 2. Noise causes error in the estimated normal

the sampling density of the point cloud. Throughout this paper, we use $\Theta(1)$ to denote some positive constant, and for notational simplicity, different appearances of $\Theta(1)$ may denote different constants. We note that distributions satisfying the (ϵ, δ) sampling condition proposed by Dey *et al.* [8] are evenly distributed in our sense.

Under the above assumptions, we would like to bound the normal estimation error and study the effects of different parameters on the estimation error. The analysis involves probabilistic arguments to account for the random nature of the noise.

2.2. *Total Least Square Method*

The normal to the total least square fitting line (or hyper-plane) of a set of k points $\mathbf{p}_i, 1 \leq i \leq k$ in $\mathbb{R}^d$ for $d \geq 2$ can be obtained by computing the eigenvector corresponding to the smallest eigenvalue of the covariance matrix M defined as $M = \frac{1}{k}\sum_{i=1}^{k}(\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^T$ [13], where $\bar{\mathbf{p}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{p}_i$. We observe that M is always symmetric positive semi-definite, and thus M has non-negative eigenvalues and non-negative diagonal entries.

2.3. *Eigen-analysis of M*

We can write the 2×2 symmetric matrix M defined in the previous section as $\begin{bmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{bmatrix}$. Note that in the absence of noise and curvature, $m_{12} = m_{22} = 0$, and thus zero is the smallest eigenvalue of M with $[0\ 1]^T$ as the corresponding eigenvector. Under our assumption that the noise and the curvature are small, y_i's are small, and thus m_{12} and m_{22} are small. Let $\alpha = (|m_{12}| + m_{22})/m_{11}$. We would like to estimate the smallest eigenvalue of M and its corresponding eigenvector when α is small.

From the *Gershgorin Circle Theorem* [11] it follows that there is an eigenvalue λ_1

such that $|m_{11} - \lambda_1| \leq |m_{12}|$, and an eigenvalue λ_2 such that $|m_{22} - \lambda_2| \leq |m_{12}|$. When $\alpha < 1/2$, we have that $\lambda_1 \geq m_{11} - |m_{12}| > m_{22} + |m_{12}| \geq \lambda_2$. It follows that the two eigenvalues are distinct, and λ_2 is the smallest eigenvalue of M. Let $[v\ 1]^T$ be the eigenvector corresponding to λ_2, then

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{bmatrix} \begin{bmatrix} v \\ 1 \end{bmatrix} = \lambda_2 \begin{bmatrix} v \\ 1 \end{bmatrix},$$
$$\begin{bmatrix} m_{11} - \lambda_2 \\ m_{12} \end{bmatrix} v = - \begin{bmatrix} m_{12} \\ m_{22} - \lambda_2 \end{bmatrix}.$$

Thus,

$$\begin{aligned} v &= -\frac{(m_{11} - \lambda_2)m_{12} + m_{12}(m_{22} - \lambda_2)}{(m_{11} - \lambda_2)^2 + m_{12}^2}, \\ |v| &= \frac{|m_{12}|(m_{11} + m_{22} - 2\lambda_2)}{(m_{11} - \lambda_2)^2 + m_{12}^2} \\ &\leq \frac{\alpha(1+\alpha)}{(1-\alpha)^2}. \end{aligned} \tag{1}$$

Thus, the estimation error, which is the angle between the estimated normal and the true normal (which is $[0\ 1]^T$ in this case), is less than $\tan^{-1}(\alpha(1+\alpha)/(1-\alpha)^2) \approx \alpha$, for small α. Note that we could write the error explicitly in closed form, then bound it. Our approach is more complicated, though as we will show later, it can be extended to obtain the error bound for the 3D case. To bound the estimation error, we need to bound α.

2.4. *Estimating Entries of M*

The assumption that the sample points are evenly distributed in the interval $[-r, r]$ implies that, given any number x in that interval, the number of points $\mathbf{p}_i$'s satisfying $|x_i - x| \geq r/4$ is at least $\Theta(1)k$. It follows easily that $m_{11} = \frac{1}{k}\sum_{i=1}^{k}(x_i - \bar{x})^2 \geq \Theta(1)r^2$. The constant $\Theta(1)$ depends only on the distribution of the random variable X.

For the entries m_{12} and m_{22}, we use $|x_i| \leq r$ and $|y_i| \leq \kappa r^2/2 + n$ to obtain the following trivial bound:

$$\begin{aligned} |m_{12}| &= \left| \frac{1}{k}\sum_{i=1}^{k} x_i y_i - \frac{1}{k^2}\sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i \right| \\ &\leq 2r(\kappa r^2/2 + n), \\ m_{22} &\leq \frac{1}{k}\sum_{i=1}^{k} y_i^2 \\ &\leq 2((\kappa r^2/2)^2 + n^2). \end{aligned}$$

Thus,

$$\begin{aligned}\alpha &\leq \Theta(1)\left(\kappa r + \frac{n}{r} + \kappa^2 r^2 + \frac{n^2}{r^2}\right) \\ &\leq \Theta(1)\left(\kappa r + \frac{n}{r}\right). \end{aligned} \tag{2}$$

This bound illustrates the effects of r, κ and n on the error. For large values of r, the error caused by the curvature κr dominates, while for small neighborhoods the term n/r dictates the error. Nevertheless, the expression depends on the absolute bound n of the noise N. This bound n can be unnecessarily large or unbounded for many distribution models of N. We would like to use our assumption on the distribution of the noise N to further improve our bound on α.

Note that,

$$\begin{aligned}|m_{12}| &= \left|\frac{1}{k}\sum_{i=1}^{k} x_i y_i - \frac{1}{k^2}\sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i\right| \\ &\leq \left|\frac{1}{k}\sum_{i=1}^{k} (\gamma_i x_i^3/2 + x_i n_i)\right| + \left|\frac{1}{k^2}\sum_{i=1}^{k} x_i \sum_{i=1}^{k} (\gamma_i x_i^2/2 + n_i)\right| \\ &\leq \kappa r^3 + \left|\frac{1}{k}\sum_{i=1}^{k} x_i n_i\right| + r\left|\frac{1}{k}\sum_{i=1}^{k} n_i\right|. \end{aligned}$$

Furthermore, under the assumption that X and N are independent, we have $E[x_i n_i] = E[x_i]E[n_i] = 0$ since $E[n_i] = 0$, and $\mathrm{Var}(x_i n_i) = \Theta(1) r^2 \sigma_n^2$ since $\mathrm{Var}(n_i) = \sigma_n^2$. Let ϵ be some small positive number. Using the *Chebyshev Inequality* [16], the following bound on $|m_{12}|$ holds with probability at least $1 - \epsilon$:

$$\begin{aligned}|m_{12}| &\leq \kappa r^3 + \Theta(1)\sqrt{\frac{r^2\sigma_n^2}{\epsilon k}} + \Theta(1) r\sqrt{\frac{\sigma_n^2}{\epsilon k}} \\ &\leq \kappa r^3 + \Theta(1)\sqrt{\frac{r^2\sigma_n^2}{\epsilon r \rho}} + \Theta(1) r\sqrt{\frac{\sigma_n^2}{\epsilon r \rho}} \\ &\leq \kappa r^3 + \Theta(1)\sigma_n\sqrt{\frac{r}{\epsilon\rho}}. \end{aligned} \tag{3}$$

For reasonable noise models, we also have that:

$$\begin{aligned}m_{22} &\leq \frac{1}{k}\sum_{i=1}^{k} 2(\gamma_i^2 x_i^4/4 + n_i^2) \\ &\leq \Theta(1)\kappa^2 r^4 + \Theta(1)\sigma_n^2. \end{aligned}$$

2.5. *Error Bound for the Estimated Normal*

From the previous estimates of the entries of M, we obtain the following bound on α with probability at least $1 - \epsilon$:

$$\alpha \leq c_1 \kappa r + c_2 \frac{\sigma_n}{\sqrt{\epsilon \rho r^3}} + c_3 \frac{\sigma_n^2}{r^2} \tag{4}$$

for some small positive constants c_1, c_2, and c_3 which depend only on the distribution of the points. Note that this bound depends on the standard deviation σ_n of the noise N rather than its magnitude bound n.

For a given set of parameters κ, σ_n, ρ, and ϵ, we can find the optimal r that minimizes the right hand side of the inequality (4). As this optimal value of r is not easily expressed in closed form, let us consider a few extreme cases:

- When there is no curvature ($\kappa = 0$) we can make the bound on α arbitrarily small by increasing r. For sufficiently large r, the bound is linear in σ_n and it decreases as $r^{-3/2}$.
- When there is no noise, we can make the error bound small by choosing r as small as possible, say $r = r_0$.
- When both noise and curvature are present, the error bound cannot be arbitrarily reduced. When the density ρ of the PCD is sufficiently high, $\alpha \leq c_1 \kappa r + (c_2 + c_3)\sigma_n^2 / r^2$. This error bound is minimized when $r = \Theta(1)\sigma_n^{2/3}\kappa^{-1/3}$, in which case $\alpha \leq \Theta(1)\kappa^{2/3}\sigma_n^{2/3}$.
- When there are both noise and curvature, and the density ρ is sufficiently low, $\alpha \leq c_1 \kappa r + (c_2 + c_3)\sigma_n / \sqrt{\epsilon \rho r^3}$. This bound is minimized when $r = \Theta(1)(\sigma_n^2/(\epsilon\rho\kappa^2))^{1/5}$, in which case, $\alpha \leq \Theta(1)(\kappa^3\sigma_n^2/(\epsilon\rho))^{1/5}$.

3. Normal Estimation in $\mathbb{R}^3$

We can extend the results obtained for curves in $\mathbb{R}^2$ to surfaces in $\mathbb{R}^3$. Given a point cloud obtained from a smooth 2-manifold in $\mathbb{R}^3$ and a point O on the surface, we can estimate the normal to the surface at O as follows: find all the points of the PCD inside a sphere of radius r centered at O, then compute the total least square plane fitting those points. The normal vector to the fitting plane is our estimate of the undirected normal at O.

Given a set of k points $\mathbf{p}_i$, $1 \leq i \leq k$, let $M = \frac{1}{k}\sum_{i=1}^{k}(\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^T$ where $\bar{\mathbf{p}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{p}_i$. As pointed out in subsection 2.2, the normal to the total least square plane for this set of k points is the eigenvector corresponding to the minimum eigenvalue of the covariance matrix M. Again, we would like to bound the angle between this eigenvector and the true normal to the surface at O.

3.1. *Modeling*

We model the PCD in a similar fashion as in the $\mathbb{R}^2$ case. We assume that O is the origin, the z-axis is the normal to the surface at O, and that the points of the

PCD in the sphere of radius r around O are samples of a topological disk on the underlying surface that has bounded curvature. We locally represent the surface as the graph of a smooth single valued C^2 continuous function $z = g(\underline{x})$ where $\underline{x} = [x, y]^T$. Using Taylor Theorem, we can write $g(\underline{x}) = \frac{1}{2}\underline{x}^T H \underline{x}$ where H is the Hessian of g at *some* point $\underline{\psi}$ such that $|\underline{\psi}| \leq |\underline{x}|$.

The assumption that the surface has bounded curvature in some neighborhood around O implies that there exists a positive constant κ such that the Hessian H of g satisfies $||H||_2 \leq \kappa$ in that neighborhood.

We write the points $\mathbf{p}_i$ as $\mathbf{p}_i = (x_i, y_i, z_i) = (\underline{x}_i, z_i)$. We assume that $z_i = g(\underline{x}_i) + n_i$, where the n_i's are independent instances of some random variable N with zero mean and standard deviation σ_n. We similarly assume that the points $\underline{x}_i$'s are *evenly distributed* in the xy-plane on a disk D of radius r centered at O, i.e. there is a radius r_0 such that any disk of size r_0 inside D contains at least three points among the $\underline{x}_i$'s but no more than some small constant number of them. We also assume that the noise and the surface curvature are both small.

3.2. *Eigen-analysis in* $\mathbb{R}^3$

We write the covariance matrix M as $\left[\begin{array}{cc|c} m_{11} & m_{12} & m_{13} \\ m_{12} & m_{22} & m_{23} \\ \hline m_{13} & m_{23} & m_{33} \end{array}\right] \triangleq \begin{bmatrix} M_{11} & M_{13} \\ M_{13}^T & m_{33} \end{bmatrix}$. As pointed out in subsection 2.2, M is symmetric and positive semi-definite. Under the assumption that the noise and the curvature are small, and that the points $\underline{x}_i$ are evenly distributed, m_{11} and m_{22} are the two dominant entries in M. We assume, without loss of generality, that $m_{11} \leq m_{22}$. Let $\alpha = (|m_{13}| + |m_{23}| + m_{33})/(m_{11} - |m_{12}|)$. As in the $\mathbb{R}^2$ case, we would like to bound the angle between the computed normal and the true normal to the point cloud in term of α.

Denote by $\lambda_1 \leq \lambda_2$ the eigenvalues of the 2×2 symmetric matrix M_{11}. Using again the Gershgorin Circle Theorem, it is easy to see that $m_{11} - |m_{12}| \leq \lambda_1 \leq \lambda_2 \leq m_{22} + |m_{12}|$.

Let λ be the smallest eigenvalue of M. From the Gershgorin Circle Theorem we have $\lambda \leq |m_{13}| + |m_{23}| + m_{33} = \alpha(m_{11} - |m_{12}|) \leq \alpha\lambda_1$. Let $[\underline{v}^T \ 1]^T$ be the eigenvector of M corresponding to the eigenvalue λ. Then, as with Equation (1), we have that:

$$\begin{aligned} \underline{v} &= -\left((M_{11} - \lambda I)^2 + M_{13}M_{13}^T\right)^{-1}\left((M_{11} - \lambda I)M_{13} + M_{13}(m_{33} - \lambda)\right) \\ &= -(M_{11} - \lambda I)^{-2}\left(I + (M_{11} - \lambda I)^{-2}M_{13}M_{13}^T\right)^{-1} \times \\ &\quad \left((M_{11} - \lambda I)M_{13} + M_{13}(m_{33} - \lambda)\right), \\ ||\underline{v}||_2 &\leq ||(M_{11} - \lambda I)^{-2}||_2 \left|\left|\left(I + (M_{11} - \lambda I)^{-2}M_{13}M_{13}^T\right)^{-1}\right|\right|_2 \times \\ &\quad \left(||(M_{11} - \lambda I)||_2||M_{13}||_2 + ||M_{13}||_2|m_{33} - \lambda|\right). \end{aligned}$$

Note that,

$$\begin{aligned}||(M_{11}-\lambda I)^{-2}M_{13}M_{13}^T||_2 &\leq ||(M_{11}-\lambda I)^{-2}||_2||M_{13}||_2||M_{13}^T||_2\\ &\leq (\lambda_1-\lambda)^{-2}(m_{13}^2+m_{23}^2)\\ &\leq (1-\alpha)^{-2}\alpha^2\,.\end{aligned}$$

Thus,

$$\begin{aligned}\left|\left|\left(I+(M_{11}-\lambda I)^{-2}M_{13}M_{13}^T\right)^{-1}\right|\right|_2 &\leq \frac{1}{1-(1-\alpha)^{-2}\alpha^2}\\ &\leq \frac{(1-\alpha)^2}{1-2\alpha}\,.\end{aligned}$$

It follows that:

$$\begin{aligned}||\underline{v}||_2 &\leq \frac{1}{(1-\alpha)^2\lambda_1^2}\frac{(1-\alpha)^2}{1-2\alpha}\,(\lambda_2\alpha\lambda_1+\alpha\lambda_1\alpha\lambda_1)\\ &\leq \frac{\alpha(1+\alpha)}{1-2\alpha}\,\frac{\lambda_2}{\lambda_1}\,.\end{aligned}$$

When α is small, the right hand side is approximately $(\lambda_2/\lambda_1)\alpha$, and thus the angle between the computed normal and the true normal, $\tan^{-1}||\underline{v}||_2$, is approximately bounded by $(\lambda_2/\lambda_1)\alpha \leq ((m_{22}+|m_{12}|)/(m_{11}-|m_{12}|))\alpha$.

3.3. *Estimation of the entries of M*

As in the $\mathbb{R}^2$ case, from the assumption that the samples are evenly distributed, we can show that $\Theta(1)r^2 \leq m_{11}, m_{22} \leq r^2$. We can also show that $m_{33} \leq \Theta(1)\kappa^2 r^4 + \Theta(1)\sigma_n^2$. Let ρ be the sampling density of the PCD at O, then $k = \Theta(1)\rho r^2$. Again, let ϵ be some small positive number. Using the Chebyshev inequality, we have that $m_{13}, m_{23} \leq \Theta(1)\kappa r^3 + \Theta(1)\sigma_n r/\sqrt{\epsilon k} \leq \Theta(1)\kappa r^3 + \Theta(1)\sigma_n/\sqrt{\epsilon\rho}$ with probability at least $1-\epsilon$. For the term m_{12}, we note that $E[x_i y_i] = 0$ and $Var(x_i y_i) = \Theta(1)r^4$, and so, by the Chebyshev inequality, $m_{12} \leq \Theta(1)r/\sqrt{\epsilon\rho}$ with probability at least $1-\epsilon$.

3.4. *Error Bound for the Estimated Normal*

Let $\beta = m_{12}/m_{11}$. We restrict our analysis to the cases when β is sufficiently less than 1, say $\beta < 1/2$. This restriction simply means that the points $\underline{x_i}$'s are not degenerate, i.e. not all of the points $\underline{x_i}$'s are lying on or near any given line on the xy-plane. With this restriction, it is clear that $(\lambda_2/\lambda_1)\alpha \leq (m_{22}/m_{11})((1+\beta)/(1-\beta))\alpha = \Theta(1)\alpha$.

From the estimations of the entries of M, we obtain the following bound with probability at least $1-\epsilon$:

$$\begin{aligned}\frac{\lambda_2}{\lambda_1}\alpha &\leq \Theta(1)\kappa r + \Theta(1)\frac{\sigma_n}{r^2\sqrt{\epsilon\rho}} + \Theta(1)\kappa^2 r^2 + \Theta(1)\frac{\sigma_n^2}{r^2}\\ &\leq \Theta(1)\kappa r + \Theta(1)\frac{\sigma_n}{r^2\sqrt{\epsilon\rho}} + \Theta(1)\frac{\sigma_n^2}{r^2}.\end{aligned}$$

This is an approximate bound on the angle between the estimated normal and the true normal. To minimize this error bound, it is clear that we should pick

$$r = \left(\frac{1}{\kappa}\left(d_1\frac{\sigma_n}{\sqrt{\epsilon\rho}} + d_2\sigma_n^2\right)\right)^{1/3}, \tag{5}$$

for some constants d_1, d_2. The constants d_1 and d_2 are small and they depend only on the distribution of the PCD.

We notice from the above result, when there is no noise, we should pick the radius r to be as small as possible, say $r = r_0$. When there is no curvature, the radius r should be as large as possible. When the sampling density is high, the optimal value of r that minimizes the error bound is approximately $r = d_2^{1/3}(\sigma_n^2/\kappa)^{1/3}$. This result is similar to that for curves in $\mathbb{R}^2$, and it is not that intuitive.

4. Experiments

In this section, we describe simulations to validate our theoretical results. We then show how to use the theoretical results to obtain good neighborhood sizes for the normal computation using the least square method.

4.1. *Validation*

We consider a family of PCDs whose points are noisy samples of the curves $(x, \kappa x^3/6)$, for $x \in [-1, 1]$ for different choices of κ. We estimate the normals to the curves at the origin by applying the least square method on their corresponding PCD. As the y-axis is known to be the true normal to the curves, the angles between the computed normals and the y-axis give the estimation errors.

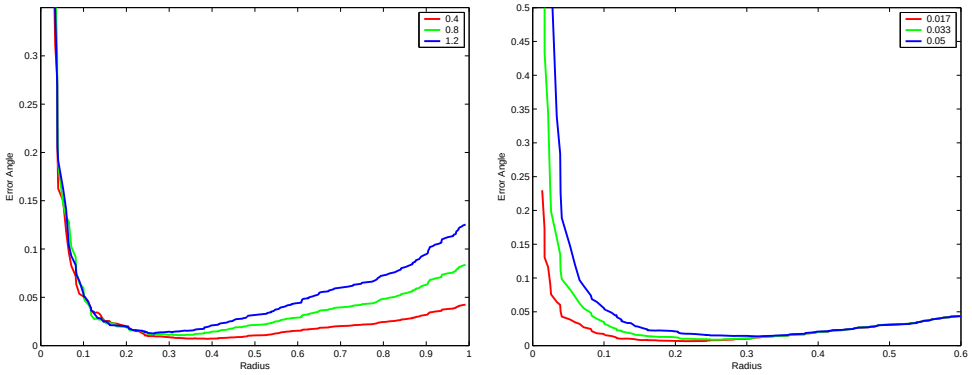
To obtain the PCDs in our experiments, we let the sampling density ρ be 100 points per unit length, and x to be uniformly distributed in the interval $[-1, 1]$. The y-components of the data have been polluted with uniformly random noise in the interval $[-n, n]$, for some value n.

Figure 3(a) shows the error as a function of the neighborhood size r when $n = 0.05$ for three different values of κ, $\kappa = 0.4, 0.8$, and 1.2. As predicted by Equation (4) for large values of r, the error increases as r increases. In the experiments, it can be seen that the error increases as κr for $r > 0.4$.

Figure 3(b) shows the estimation error as a function of the neighborhood size r for small r when $\kappa = 1.2$ for three different values of n, $n = 0.017, 0.033$, and 0.05. We observe that the error tends to decrease as r increases for $r < 0.15$. This is expected as from Equation (4), the bound on the error is a decreasing function of r when r is small. To factor out the random effect of noise, the estimation error curves have been averaged over 50 runs of the experiment.

4.2. *Estimating Neighborhood Size for the Normal Computation*

In this section, we use the results obtained in Section 3 to estimate the normals at all the sample points of a PCD. The data points in the PCD are assumed to be noisy



(a) Error due to curvature dominates for $r > 0.4$. The error–curves behave similarly for different choices of κ.

(b) Error due to noise dominates for $r < 0.15$. The error–curves behave similarly under different amounts of noise.

Fig. 3. Effects of curvature and noise on estimation error for different choices of r.

samples of a smooth surface in $\mathbb{R}^3$. This is the case, for example, for PCD obtained by range scanners. We would like to use Equation (5) to obtain the neighborhood size for the normal computation using the least square method.

We assume that the standard deviation σ_n of the noise has been provided as a part of the input. We estimate the other local parameters in Equation (5), then compute r. Note that this value of r minimizes the bound of the normal computation error, and there is no guarantee that this would minimize the error itself. The constants d_1 and d_2 in the equation depend on the sampling distribution of the PCD. While we can attempt to compute the exact values of d_1 and d_2, we try to estimate the values of d_1 and d_2.

Given a PCD, we estimate the local sampling density ρ as follows. For a given point $\mathbf{p}$ in the PCD, we use the approximate nearest neighbor library ANN [5] to find the distance s from $\mathbf{p}$ to its k_0-th nearest neighbor for some small number k_0. The local sampling density at $\mathbf{p}$ can then approximated as $\rho = k_0/(\pi s^2)$ samples per unit area.

To estimate the maximum local curvature κ, we use the method proposed by Gumhold *et al.* [12]. Let $\mathbf{p}_j$, $1 \leq j \leq k$ be the k-nearest sample points around $\mathbf{p}$, and let μ be the average distance from $\mathbf{p}$ to all the points $\mathbf{p}_j$. We compute the best fit least square plane for those k points, and let d be the distance from $\mathbf{p}$ to that best fit plane. The local curvature at $\mathbf{p}$ can then be approximated as $\kappa = 2d/\mu^2$. This method gives an estimate of the local curvature without any guarantees on the approximation quality.

Once all the parameters are obtained, we compute the neighborhood size r using Equation (5). Note that the estimated value of r can be used to obtain a good value

Algorithm 1 Estimates good normals for all the points of a noisy PCD

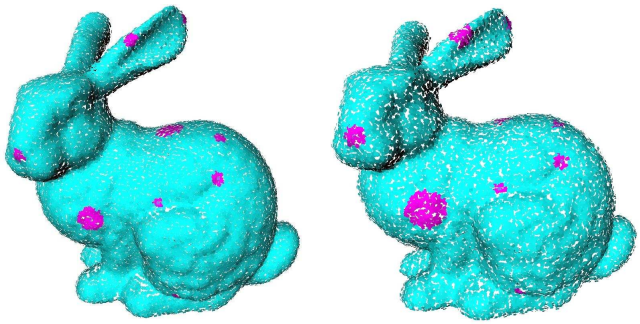
```
estimate d_1, d_2
choose ε
for each point p do
    k ← k_0
    count ← MaxCount
    repeat
        r_old ≈ distance from p to its k-th nearest neighbor
        ρ ← k/πr_old^2
        given k, compute κ locally (Gumhold et al. [12])
        compute r_new using Equation (5)
        k ← ⌈πρr_new^2⌉
        if k > k_threshold then
            break
        end if
        count ← count − 1
    until count ≠ 0
    the normal to the least square fit plane to the k-nearest neighbors of p gives
    a good estimate of the normal at p
end for
```

for k, which, in turn, can be used to re-estimate the local density and the local curvature. This suggests an iterative scheme in which we repeatedly estimate the local density, the local curvature, and the neighborhood size. In our experiments, we found that only a small number of iterations were enough to obtain good values for all the quantities. Algorithm 1 illustrates this iterative scheme. For the following experiments, k_0 was set to 15, and MaxCount was set to 10. The value of ϵ was fixed at 0.1.

We still have the problem of obtaining good estimates for the constants d_1 and d_2. Fortunately, we only have to estimate the constants once for a given PCD, and we can use the same constants for other PCDs with a similar point distribution. We used Figure 6(a) for choosing d_1 and d_2. The PCD was created such that underlying model and hence the exact normals at all points (except those on the edges) are known. Estimation errors can then be computed exactly at almost all the points and this information used to estimate the constants. We found that $d_1 = 1$, $d_2 = 4$ is a good pair of values and the same pair has been be used for the other data sets.

Noisy PCD used in our experiments were obtained by adding noise to the original data. The x, y, and z components of the noise were chosen independently and uniformly random. The magnitude of the added noise was measured in a scale where the average spacing between neighboring points in the mesh representation of the original data was taken as one unit.

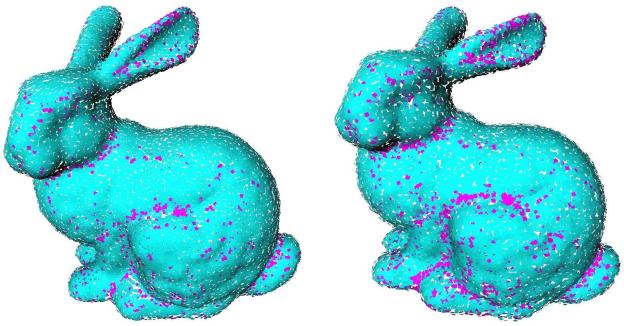
Figure 4 shows the effects of curvature and noise on the choice of neighborhood



(a) 1x noise level (b) 2x noise level

Fig. 4. Effects of curvature and noise on the choice of neighborhood size under different amounts of noise in the input data. The neighbors determined by Algorithm 1 for a few points on the bunny have been highlighted.

size. The neighborhood of a few points are shown in the figures. Figure 4(a) demonstrates that bigger neighborhoods have been selected in flatter regions compared to neighborhoods in regions with more local curvature. Figure 4(b), in comparison to Figure 4(a), shows that a higher noise level results in selection of larger neighborhoods.



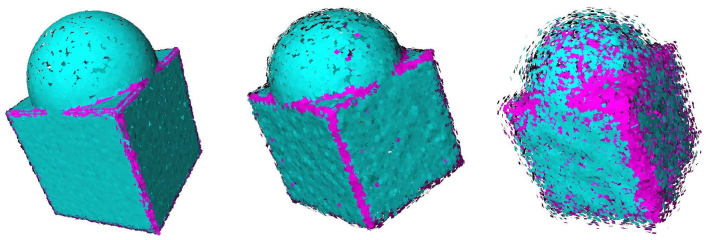
(a) 1x noise level (b) 2x noise level

Fig. 5. Normal estimation errors for the bunny PCD with noise added. Points with more than 5° estimation error have been highlighted.

We compute the normals of the noisy PCD, and use the angles between those normals and the normals of the original PCD as estimates of the normal compu-

tation errors. The normals computed from the mesh representation of the original data set are considered to be the true normals. The mesh representation of the original data sets is not available to Algorithm 1. In Figure 5, we highlight the points with estimation error more than 5° under two different amounts of noise.

Figure 6 shows the performance of the algorithm under different noise conditions. In Figure 6(c), we observe that even in presence of significant noise, the algorithm performs well in flat faces of the object. As noted before, since the underlying surface model is known for this PCD, the true normals used for computing the estimation errors, are specified at almost all the points.



(a) 1x noise (b) 2x noise (c) 4x noise

Fig. 6. Performance of the algorithm under various noisy conditions. Points with more than 5° estimation error have been highlighted.

5. Conclusions

We have analyzed the method of least square fitting to a neighborhood in estimating the normals to a point cloud data derived either from a smooth curve in $\mathbb{R}^2$ or a smooth surface in $\mathbb{R}^3$, with noise added. In both cases, we provided theoretical bounds on the maximum angle between the estimated normal and the true normal of the underlying manifold. This theoretical study allowed us to find an optimal neighborhood size to be used in the least square method. Application of the theoretical study on practical data resulted in satisfactory behavior.

6. Acknowledgments

We are grateful to Tamal K. Dey for his remarks and suggestions. Special thanks to Marc Levoy, Ron Fedkiw for their feedback and comments.

We also want to thank the numerous referees of the previous versions of this paper for their extremely useful suggestions.

References

1. Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T. Silva. Point set surfaces. In *Proceedings of the Conference on Visualization*, pages 21–28. IEEE Computer Society, 2001.
2. N. Amenta, S. Choi, T. K. Dey, and N. Leekha. A simple algorithm for homeomorphic surface reconstruction. In *Proceedings of the Symposium on Computational Geometry*, pages 213–222. ACM Press, 2000.
3. Nina Amenta and Marshall Bern. Surface reconstruction by voronoi filtering. In *Proceedings of the Symposium on Computational Geometry*, pages 39–48, 1998.
4. Nina Amenta, Marshall Bern, and Manolis Kamvysselis. A new Voronoi-based surface reconstruction algorithm. *Computer Graphics*, 32(Annual Conference Series):415–421, 1998.
5. Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, and Angela Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*, 45(6):891–923, 1998.
6. Jean-Daniel Boissonnat and Frederic Cazals. Smooth surface reconstruction via natural neighbour interpolation of distance functions. In *Proceedings of the Symposium on Computational Geometry*, pages 223–232, 2000.
7. J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 67–76. ACM Press, 2001.
8. T. K. Dey, J. Giesen, S. Goswami, and W. Zhao. Shape dimension and approximation from samples. In *Proceedings of the Symposium on Discrete Algorithms*, pages 772–780, 2002.
9. Shachar Fleishman, Iddo Drori, and Daniel Cohen-Or. Bilateral mesh denoising. *ACM Transactions on Graphics (TOG)*, 22(3):950–953, 2003.
10. Stefan Funke and Edgar A. Ramos. Smooth-surface reconstruction in near-linear time. In *Proceedings of the Symposium on Computational Geometry*, pages 781–790. Society for Industrial and Applied Mathematics, 2002.
11. Gene H. Golub and Charles F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, 1996.
12. S. Gumhold, X. Wang, and R. MacLeod. Feature extraction from point clouds. In *10th International Meshing Roundtable, Sandia National Laboratories*, pages 293–305, October 2001.
13. Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Surface reconstruction from unorganized points. *Computer Graphics*, 26(2):71–78, 1992.
14. Thouis R. Jones, Frédo Durand, and Mathieu Desbrun. Non-iterative, feature-preserving mesh smoothing. *ACM Transactions on Graphics (TOG)*, 22(3):943–949, 2003.
15. In-Kwon Lee. Curve reconstruction from unorganized points. *Computer Aided Geometric Design*, 17(2):161–177, 2000.
16. A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addison Wesley, 1994.
17. Mark Pauly, Richard Keiser, Leif P. Kobbelt, and Markus Gross. Shape modeling with point-sampled geometry. *ACM Transaction on Graphics*, 22(3):641–650, 2003.
18. Szymon Rusinkiewicz and Marc Levoy. Qsplat: a multiresolution point rendering system for large meshes. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 343–352. ACM Press/Addison-Wesley

Publishing Co., 2000.
19. William Welch and Andrew Witkin. Free-form shape design using triangulated surfaces. *Computer Graphics*, 28(Annual Conference Series):247–256, 1994.
20. Matthias Zwicker, Mark Pauly, Oliver Knoll, and Markus Gross. Pointshop 3d: an interactive system for point-based surface editing. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 322–329. ACM Press, 2002.