Estimating the boundary of a point cloud with applications to solving PDEs

Jeff Calder

School of Mathematics
University of Minnesota

IPAM HJ2020 Reunion Workshop
Jan 20, 2022

Joint work with:
Sangmin Park (CMU), Dejan Slepčev (CMU)

Research supported by NSF-DMS 1713691, 1944925, the Alfred P. Sloan Foundation, and the McKnight Foundation
Goal: Identify “boundary points” of a point cloud, in a way that allows setting boundary conditions for solving PDEs.
Previous work

To fix some notation, \( \mathcal{X} = \{x^1, \ldots, x^n\} \) is an i.i.d. sample from \( \Omega \subset \mathbb{R}^d \) with density \( \rho \).

1 [Devroye & Wise, 1980] set

\[
\Omega_n = \bigcup_{i=1}^{n} B(x^i, r) \quad \text{and} \quad \hat{\partial} \Omega_n = \partial \Omega_n.
\]
Previous work

To fix some notation, $\mathcal{X} = \{x^1, \ldots, x^n\}$ is an i.i.d. sample from $\Omega \subset \mathbb{R}^d$ with density $\rho$.

1 [Devroye & Wise, 1980] set

$$\Omega_n = \bigcup_{i=1}^{n} B(x^i, r_n) \text{ and } \widehat{\partial\Omega_n} = \partial\Omega_n.$$ 

* [Cuevas and Rodriguez-Casal, 2004] showed that

$$d_H(\partial\Omega_n, \partial\Omega) \sim (n^{-1} \log(n))^{\frac{1}{d}}$$

provided $r_n \sim (n^{-1} \log(n))^{\frac{1}{d}}$.

* Computation of $\Omega_n$ is via alpha-shapes, which are only computationally feasible in $d = 2, 3$ dimensions.

* [Casal 2007] and [Aamari, Aaron, & Levrard, 2021] improve the rate by interpolating the boundary points better.

2 [Cuevas and Fraiman, 1997] use kernel density estimators to detect the boundary as a level set of $\hat{\rho}$.

3 [Lachiéze-Rey & Vega, 2017] Voronoi-cell based boundary estimator (similar to alpha-shapes for complexity).
Previous work

4 [Wu & Wu, 2019] and [Aaron & Cholaquidis, 2020] use the size of the vector

\[ \sum_{j: |x^i - x^j| \leq r} (x^i - x^j). \]

There are many other works that use similar ideas, but without theoretical guarantees

- BORDER [Xia et al., 2006] and BRIM [Qiu et al., 2007].
Posing the problem

There are 2 different ways to pose the problem:

1. Estimate $\partial \Omega$ from the i.i.d. sample $\mathcal{X}$.
   - Computationally very hard in high dimensions.

2. Estimate the points in the sample $\mathcal{X}$ that are close (within $\varepsilon$) of the boundary.
   - As we will show, this is tractable in high dimensions.
   - This is all we need to set boundary conditions for solving PDEs on $\mathcal{X}$.

Distance to the boundary

We first change gears and look at estimating the distance to the boundary

\[ d_\Omega(x) = \text{dist}(x, \partial \Omega). \]

Provided \( B(x, r) \cap \partial \Omega \) is not empty

\[
\begin{align*}
    d_\Omega(x) &= \max_{y \in B(x, r) \cap \Omega} \{d_\Omega(x) - d_\Omega(y)\} \\
    &= \max_{y \in B(x, r) \cap \Omega} \{\nabla d_\Omega(x) \cdot (x - y)\} + O(r^2) \\
    &= \max_{y \in B(x, r) \cap \Omega} \{\nu(x) \cdot (x - y)\} + O(r^2),
\end{align*}
\]

since \( \nabla d_\Omega(x) = \nu(x) \) is the inward normal vector.
Distance to the boundary

To get a second order estimator, we use the higher order Taylor expansion

\[ d_\Omega(x) - d_\Omega(y) = \frac{1}{2} (\nabla d_\Omega(x) + \nabla d_\Omega(y)) \cdot (x - y) + O(r^3). \]

This yields, provided \( B(x, r) \cap \partial \Omega \) is not empty

\[ d_\Omega(x) = \max_{y \in B(x, r) \cap \Omega} \{ d_\Omega(x) - d_\Omega(y) \} \]

\[ = \max_{y \in B(x, r) \cap \Omega} \left\{ \frac{1}{2} (\nabla d_\Omega(x) + \nabla d_\Omega(y)) \cdot (x - y) \right\} + O(r^3) \]

\[ = \max_{y \in B(x, r) \cap \Omega} \left\{ \frac{1}{2} (\nu(x) + \nu(y)) \cdot (x - y) \right\} + O(r^3). \]

**Note:** Estimating \( d_\Omega \) boils down to estimating the inward normal vector \( \nu(x) \).
First order normal vector estimation

Let \( r > 0 \) and \( \mathcal{X} = \{x^1, x^2, \cdots, x^n\} \) be the set of i.i.d. points distributed according to \( \rho \) on \( \Omega \subset \mathbb{R}^d \). For each \( x^0 \in \mathcal{X} \) we define the first order normal vector estimator

\[
\hat{v}_r(x^0) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{B(x^0, r)}(x^i)(x^i - x^0), \quad \hat{\nu}_r(x^0) = \frac{\hat{v}_r(x^0)}{|\hat{v}_r(x^0)|}.
\]

We also define the corresponding population level estimator

\[
\bar{v}_r(x^0) = \int_{\Omega \cap B(x^0, r)} (x - x^0) \rho(x) \, dx, \quad \bar{\nu}_r(x^0) = \frac{\bar{v}_r(x^0)}{|\bar{v}_r(x^0)|}.
\]
Second order normal vector estimation

For each $x^0 \in \mathcal{X}$ we define the second order normal vector estimator

$$\hat{v}^n_r(x^0) = \frac{1}{n} \sum_{i=1}^{n} \frac{1_B(x^0, r)(x^i)}{\hat{\theta}(x^i)} (x^i - x^0), \quad \nu^n_r(x^0) = \frac{\hat{v}^n_r(x^0)}{|\hat{v}^n_r(x^0)|},$$

$$\hat{\theta}(x) = \frac{1}{\omega_d n} \left( \frac{2}{r} \right)^d \sum_{j=1}^{n} 1_B(x, r/2)(x^j).$$

At the population level our estimator takes the form

$$\nu^n_r(x^0) = \int_{B(x^0, r) \cap \Omega} \frac{\rho(x)}{\theta(x)} (x - x^0) dx, \quad \nu^n_r(x^0) = \frac{\nu^n_r(x^0)}{|\nu^n_r(x^0)|},$$

(1)

$$\theta(x) = \frac{2^d}{\omega_d r^d} \int_{B(x, r/2) \cap \Omega} \rho(z) dz.$$
Let $r > 0$ and $\mathcal{X} = \{x_1, x_2, \cdots, x^n\} \subset \Omega$. We define the first order distance function estimator $\hat{d}_{d_1^r} : \mathcal{X} \rightarrow \mathbb{R}$ by

$$
\hat{d}_{d_1^r}(x^0) = \max_{x^i \in B(x^0, r) \cap \mathcal{X}} (x^0 - x^i) \cdot \hat{\nu}_r(x^0).
$$
Second order estimator of $d_\Omega$

Our Taylor expansion would suggest the second order estimator

$$
\max_{x^i \in B(x^0, r) \cap X_n} (x^0 - x^i) \cdot \frac{1}{2} (\hat{\nu}^n_r(x^0) + \hat{\nu}^n_r(x^i))
$$

This test has difficulties with false postives at interior points, where $\hat{\nu}^n_r(x^0)$ and $\hat{\nu}^n_r(x^i)$ are not reliable, and can cancel each other out.

To avoid this problem, we define the second order estimator with cutoff

$$
\hat{d}_r^2(x^0) = \max_{x^i \in B(x^0, r) \cap X} (x^0 - x^i) \cdot \left[ \hat{\nu}^n_r(x^0) + \frac{\hat{\nu}^n_r(x^i) - \hat{\nu}^n_r(x^0)}{2} \mathbb{1}_{\mathbb{R}^+}(\hat{\nu}^n_r(x^i) \cdot \hat{\nu}^n_r(x^0)) \right].
$$
Estimating the boundary

Let us define
\[ \partial_r \Omega = \{ x \in \Omega : d_\Omega(x) \leq r \}. \]

We want an estimator of the boundary points \( \partial \hat{\Omega} \) that satisfies
\[ \partial_\varepsilon \subset \partial \hat{\Omega} \subset \partial_{2\varepsilon} \Omega. \]

This is motivated by setting boundary conditions for solving PDEs, where the boundary should:

- Identify sufficiently many boundary points so that BC are continuously attained as \( n \to \infty \).
- Not identify any interior points as boundary points.

Given an empirical estimator \( \hat{d}_r \) we define the test \( \hat{T}_{\varepsilon,r} : X \to \{0, 1\} \) by

\[
\hat{T}_{\varepsilon,r}(x^0) = \begin{cases} 
1 & \text{if } \hat{d}_r(x^0) < \frac{3\varepsilon}{2} \\
0 & \text{otherwise.}
\end{cases}
\]
Assumptions

The assumptions we make on the geometric parameters are as follows.

Assumption

\[ \frac{\varepsilon}{r} \leq \frac{1}{3\sqrt{d}}. \]

Assumption

\[ r^2 \leq R\varepsilon. \]

\( R \) is the reach of \( \partial \Omega \), which is assumed positive.

We also assume \( \rho \) is Lipschitz continuous and bounded above and below by positive constants

\[ 0 < \rho_{\text{min}} \leq \rho \leq \rho_{\text{max}}. \]
Error estimates for the normal

Theorem (Error estimates for the estimated normal vector)

Let \( x^0 \in \mathcal{X} \) with \( d_\Omega(x^0) \leq 2\varepsilon \), \( \gamma > 2 \) and \( \varepsilon, r > 0 \) satisfy Assumption 2. Let \( r \) and \( n \) satisfy

\[
\left( \frac{3\gamma \rho_{\text{max}} d^2 \omega_d R^2 \log n}{C_x^2 \rho_{\text{min}}^2 n} \right)^{\frac{1}{d+2}} \leq r \leq \frac{RC_y}{2C_x}.
\]

Then

\[
P \left( |\hat{\nu}_r(x^0) - \nu(x^0)| \geq \frac{13C_x}{RC_y} r \right) \leq 2dn^{-\gamma}
\]
Theorem (Error estimates for the distance estimator)

Let $\varepsilon, r > 0$ satisfy Assumptions 1 and 2. Suppose $\gamma > 2$, and $x^0 \in \mathcal{X}$ such that $d_\Omega(x^0) \leq 2\varepsilon$, and $r$ satisfy

\begin{equation}
RC_r \left( \frac{\log n}{n} \right)^{\frac{1}{d+2}} \leq r \leq \frac{RC_y}{2C_x}
\end{equation}

Then

\begin{equation}
d_\Omega(x^0) - \frac{13C_x}{RC_y} r^2 \leq \hat{d}_r(x^0) \leq d_\Omega(x^0) + \left( \frac{13C_x}{RC_y} + \frac{1}{R} \right) r^2
\end{equation}

with probability at least $1 - (2d + 1)n^{-\gamma}$. 
Corollary (Accuracy of the boundary test)

Let $x^0 \in X$, $\gamma > 2$ and $\varepsilon, r > 0$ satisfy Assumptions 1 and 2. If $\varepsilon, n, \gamma$ satisfy

\begin{equation}
RC_{\varepsilon} \left( \frac{\log n}{n} \right)^{\frac{2}{d+2}} \leq \varepsilon
\end{equation}

and $r$ satisfies

\begin{equation}
r \leq \frac{RC_y}{2C_x}
\end{equation}

then

\begin{equation}
\mathbb{P}(\hat{T}_{\varepsilon, r}^1(x^0) = 1 \mid d_\Omega(x^0) \geq 2\varepsilon) + \mathbb{P}(\hat{T}_{\varepsilon, r}^1(x^0) = 0 \mid d_\Omega(x^0) \leq \varepsilon) \leq (2d + 1)n^{-\gamma}.
\end{equation}
Experiments

- Blue points satisfy $d_\Omega \leq \varepsilon$.
- Green points satisfy $\varepsilon < d_\Omega \leq 2\varepsilon$
- Red points are identified by our second order test.
Comparison with other methods

- 2nd
- 1/α=0.05
- 1/α=0.1

- BRIM, 15%
- BRIM, 7%

- WuWu, 15%
- WuWu, 10%

Calder (UMN)
Extension to the manifold setting

We can extend our method to the manifold setting by:
- Estimating the tangent space with PCA.
- Projecting the normal estimation onto the estimated tangent space.

Figure: Boundary points of a point cloud on a manifold identified the second order test, \( n = 3000, r = 0.21, \varepsilon = 0.05 \). The point cloud is represented by blue dots, and the boundary points identified are circled in red.
Solving PDEs on point clouds

Let us recall the notation: For $\varepsilon > 0$ we define

$$\partial_\varepsilon \Omega = \{ x \in \Omega : \text{dist}(x, \partial \Omega) \leq \varepsilon \}$$

and set $\Omega_\varepsilon = \Omega \setminus \partial_\varepsilon \Omega$.

We now turn to solving PDEs on point clouds. We assume that we have computed a boundary set $\partial_\varepsilon \mathcal{X} \subset \mathcal{X}$ that satisfies

$$(10) \quad \mathcal{X}_\varepsilon \subset \Omega_\varepsilon \quad \text{and} \quad \partial_\varepsilon \mathcal{X} \subset \partial_2 \varepsilon \Omega,$$

where $\mathcal{X}_\varepsilon = \mathcal{X} \setminus \partial_\varepsilon \mathcal{X}$.

We will show with a series of examples that (10) is sufficient to ensure that boundary conditions (Dirichlet/Neumann/Robin) are preserved in the limit as $n \to \infty$ and $\varepsilon \to 0$. 
An eikonal equation

First, we consider extending our main results to the estimation of the distance function

(11) \[ d_{\Omega}(x) := \text{dist}(x, \partial \Omega) \]

on the whole point cloud \( \mathcal{X} \). We can do this by solving the graph eikonal equation

(12) \[
\begin{align*}
\min_{y \in B_0(x^i, \varepsilon) \cap \mathcal{X}} \left\{ u_\varepsilon(y) - u_\varepsilon(x^i) + |y - x^i| \right\} &= 0, \quad \text{if } x^i \in \mathcal{X}_\varepsilon \\
u_\varepsilon(x^i) &= 0, \quad \text{if } x^i \in \partial_\varepsilon \mathcal{X},
\end{align*}
\]

where we write \( B_0(x, \varepsilon) := B(x, \varepsilon) \setminus \{x\} \) for the punctured ball.
Error estimates

**Theorem**

Assume $\varepsilon \leq \frac{R}{8}$. Let $u_\varepsilon$ solve (12) and let $0 < t \leq \min\{\frac{1}{d}, \frac{1}{2} - \frac{4\varepsilon}{R}\}$. Then

\begin{equation}
-2\varepsilon \leq u_\varepsilon - d_\Omega \leq 2d_\Omega \left( t + \frac{4\varepsilon}{R} \right) \text{ on } X
\end{equation}

holds with probability at least $1 - 2n \exp\left( -\frac{\omega d^{-1}}{4(d+1)} \rho_{\text{min}} \varepsilon^d (2t)^{-\frac{d+1}{2}} \right)$.

**Remark**

To obtain the conditions under which the error rate is linear in $\varepsilon$ we take $t = \varepsilon$ and obtain

\[-2\varepsilon \leq u_\varepsilon - d_\Omega \leq 2d_\Omega \left( 1 + \frac{4}{R} \right) \varepsilon\]

holds with probability at least $1 - 2n^{-2}$ provided that the length scale $\varepsilon$ satisfies:

\begin{equation}
\varepsilon \geq \left( \frac{6(d + 1) \log(n)}{2^{\frac{d+1}{2}} \omega_{d-1} \rho_{\text{min}} n} \right)^{\frac{2}{3d+1}}.
\end{equation}
Numerical results

Figure: Plots of the solution to the graph eikonal equation (12) for \( n = 10^4 \) for both the box and ball domains, and error plots for varying \( \varepsilon \) averaged over 100 trials. The red points indicate the detected boundary points used in solving (12).
Second order equations

To proceed in generality, we assume there exists $C_{\nu}$ such that

$$|\hat{\nu}_\varepsilon(x^i) - \nu(x^i)| \leq C_{\nu}\varepsilon$$  \hspace{1cm} (15)

for all $x^i \in \mathcal{X} \cap \partial_2\varepsilon\Omega$. The graph PDEs we solve will involve the graph Laplacian

$$\mathcal{L}_\varepsilon u(x^i) = \frac{2}{\sigma_\eta n\varepsilon^{d+2}} \sum_{j=1}^{n} \eta \left( \frac{|x^i - x^j|}{\varepsilon} \right) (u(x^j) - u(x^i)),$$  \hspace{1cm} (16)

where $\sigma_\eta = \int_{\mathbb{R}^d} \eta(|z|)z_1^2 \, dz$, $\eta$ is compactly supported on $[0, 1]$, and $\int_{\mathbb{R}^d} \eta(|z|) \, dz = 1$.

We define the normal derivative $\nabla_\nu u(x) = \nabla u(x) \cdot \nu$ and the approximate normal derivative $\hat{\nabla}_\nu$ by

$$\hat{\nabla}_\nu u(x^i) = \frac{u(p_n(x^i + \varepsilon\hat{\nu}_\varepsilon(x^i))) - u(x^i)}{\varepsilon},$$  \hspace{1cm} (17)

where $p_n : \Omega \to \mathcal{X}$ is the closest point map.
Robin-type boundary conditions

We consider the following graph Poisson equation with Robin-type boundary conditions

\[
\begin{align*}
\mathcal{L}_\varepsilon u(x^i) &= f(x^i), \quad \text{if } x^i \in \mathcal{X}_\varepsilon \\
\gamma u(x^i) - (1 - \gamma) \hat{\nabla}_\nu u(x^i) &= g(x^i), \quad \text{if } x^i \in \partial \varepsilon \mathcal{X}.
\end{align*}
\]

Here, $\gamma \in (0, 1]$ and $f$ and $g$ are given smooth functions.

We will show that the solution of (18) converges as $n \to \infty$ and $\varepsilon \to 0$ to the solution of the Robin problem

\[
\begin{align*}
-\rho^{-1} \text{div}(\rho^2 \nabla u) &= f, \quad \text{in } \Omega \\
\gamma u - (1 - \gamma) \nabla_\nu u &= g, \quad \text{on } \partial \Omega.
\end{align*}
\]
Error estimate

Theorem

Let \( \varepsilon > 0 \) and assume \( C_\nu \varepsilon \leq 1 \). Let \( u \) be the solution of (19) with \( \gamma > 0 \), and let \( u_\varepsilon \) satisfy (18). Then for any \( 0 < \lambda \leq \varepsilon^{-1} \) and \( t > 0 \), the event that

\[
|u(x^i) - u_\varepsilon(x^i)| \leq C \left( \| \gamma u - (1 - \gamma) \nabla_\nu u - g \|_{L^\infty(\partial_{2\varepsilon} \Omega)} + (1 - \gamma)(t + C_\nu \varepsilon + \varepsilon) + \varepsilon^2 + \lambda \right)
\]

holds for all \( x^i \in X \) has probability at least

\[
1 - n \exp \left( -\frac{1}{6} \omega_d \rho_{\text{min}} n \varepsilon^d t^d \right) - 2n \exp \left( -C n \varepsilon^{d+2} \lambda^2 \right).
\]

Remark

The rate is \( O(\varepsilon) \) by choosing \( t = \lambda = \varepsilon \). However, when \( \gamma = 1 \) (the Dirichlet problem), it is possible to obtain an \( O(\varepsilon^2) \) rate provided \( g \) can be extended so that

\[
\| u - g \|_{L^\infty(\partial_{2\varepsilon} \Omega)} \leq C \varepsilon^2.
\]
Numerical results

Figure: On the left, plots of the solution to the Robin problem and principal Dirichlet eigenvector for $n = 10^5$ points on the disk, compared to the exact solutions of each problem. On the right we show an error plot for varying $\varepsilon$ averaged over 100 trials.
Dirichlet eigenfunction

Figure: First 7 Laplacian Dirichlet eigenfunctions on the disk computed via approximation with graph Laplacian eigenvectors with $n = 10^5$ points.
MNIST

(a) Random digits  
(b) Boundary digits

Figure: MNIST experiments.
Figure: MNIST experiments.
FashionMNIST

(a) Random images

(b) Boundary images

**Figure:** FashionMNIST experiments.
FashionMNIST

Figure: FashionMNIST experiments.

- GitHub Code: https://github.com/sangmin-park0/BoundaryTest
- Python Notebook Example: https://colab.research.google.com/drive/1tW0SZ9vZEAZ08T248EAl0CNtpzmpDFDT?usp=sharing