

## CONVERGENCE OF THE WEIGHTED NONLOCAL LAPLACIAN ON RANDOM POINT CLOUD \*

Zuoqiang Shi<sup>1)</sup>

*Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China*

*Email: zqshi@tsinghua.edu.cn*

Bao Wang

*Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA*

*Email: wangbaonj@gmail.com*

### Abstract

We analyze the convergence of the weighted nonlocal Laplacian (WNLL) on the high dimensional randomly distributed point cloud. Our analysis reveals the importance of the scaling weight,  $\mu \sim |P|/|S|$  with  $|P|$  and  $|S|$  being the number of entire and labeled data, respectively, in WNLL. The established result gives a theoretical foundation of the WNLL for high dimensional data interpolation.

*Mathematics subject classification:* 65D05, 65D25, 41A05.

*Key words:* Weighted nonlocal Laplacian, Laplace-Beltrami operator, Point cloud; High-dimensional interpolation.

### 1. Introduction

In this paper, we consider the convergence of the weighted nonlocal Laplacian (WNLL) on high dimensional randomly distributed data. WNLL is proposed in [11] for high dimensional point cloud interpolation, which successfully resolves the curse of dimensionality issue in the classical basis function-based approaches. High dimensional point cloud interpolation is a fundamental problem in machine learning, which can be mathematically formulated as follows: Let  $P = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$  and  $S = \{\mathbf{s}_1, \dots, \mathbf{s}_m\}$  be two sets of points in  $\mathbb{R}^d$ . Suppose  $u$  is a function defined on the point cloud  $\bar{P} = P \cup S$ , which is known only over the set  $S$ , and we denote the function  $u$  as  $b(\mathbf{s})$  for any  $\mathbf{s} \in S$ . We use interpolation methods, e.g. WNLL, to compute  $u$  over the whole point cloud  $\bar{P}$  leveraging the given values over  $S$ .

Nonlocal Laplacian is widely used in nonlocal methods for image processing [2,3,6,7], and in nonlocal Laplacian, the interpolating function is obtained by minimizing the following energy functional

$$\mathcal{J}(u) = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{y} \in \bar{P}} w(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y}))^2, \quad (1.1)$$

with the constraint

$$u(\mathbf{x}) = b(\mathbf{x}), \quad \mathbf{x} \in S. \quad (1.2)$$

Here,  $w(\mathbf{x}, \mathbf{y})$  is a given weight function, typically chosen to be Gaussian, i.e.  $w(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/\sigma^2)$  with  $\sigma > 0$  being a hyperparameter, and  $\|\cdot\|$  is the Euclidean norm in

---

\* Received December 13, 2020 / Revised version received April 17, 2021 / Accepted April 29, 2021 /

Published online September 13, 2021 /

<sup>1)</sup> Correspondence author

$\mathbb{R}^d$ . In graph theory and machine learning literature, nonlocal Laplacian is also called graph Laplacian [4, 16].

Graph Laplacian works very well with a high labeling rate, i.e., there is a large portion of labeled data. However, when the labeling rate is low, i.e.  $|S|/|\bar{P}| \ll 1$ , the solution of the graph Laplacian is found to be discontinuous at the labeled points [11, 12]. WNLL is devised to fix the issues related to the low-labeling rate, and in WNLL, the energy functional in (1.1) is modified by adding the weight,  $\frac{|\bar{P}|}{|S|}$ , to balance the labeled and unlabeled terms, which resulting in

$$\min_u \sum_{\mathbf{x} \in P} \left( \sum_{\mathbf{y} \in \bar{P}} w(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y}))^2 \right) + \frac{|\bar{P}|}{|S|} \sum_{\mathbf{x} \in S} \left( \sum_{\mathbf{y} \in \bar{P}} w(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y}))^2 \right), \quad (1.3)$$

with the constraint

$$u(\mathbf{x}) = b(\mathbf{x}), \quad \mathbf{x} \in S.$$

When the labeling rate is high, WNLL is close to graph Laplacian. However, when the labeling rate is low, the specially designed weight forces the solution to be close to the given values near the labeled points, such that the discontinuities are removed. Furthermore, The optimization problem (1.3) is easy to solve numerically. With a symmetric weight function, i.e.  $w(\mathbf{x}, \mathbf{y}) = w(\mathbf{y}, \mathbf{x})$ , the corresponding Euler-Lagrange equation of (1.3) is a simple linear system

$$\begin{aligned} 2 \sum_{\mathbf{y} \in P} w(\mathbf{x}, \mathbf{y}) (u(\mathbf{x}) - u(\mathbf{y})) + \left( \frac{|P|}{|S|} + 2 \right) \sum_{\mathbf{y} \in S} w(\mathbf{y}, \mathbf{x})(u(\mathbf{x}) - b(\mathbf{y})) &= 0, \quad \mathbf{x} \in P, \\ u(\mathbf{x}) &= b(\mathbf{x}), \quad \mathbf{x} \in S. \end{aligned}$$

This linear system can be solved efficiently by the conjugate gradient iteration. The advantages of the WNLL over the graph Laplacian have been shown evidently in image inpainting [11, 12], scientific data interpolation [15], and more recently deep learning [13].

### 1.1. Main Result

We consider the error of the WNLL in a model problem, where the whole computational domain is set to be a  $k$ -dimensional closed manifold  $\mathcal{M}$  embedded in  $\mathbb{R}^d$ . The point cloud  $P$ , uniformly distributed on  $\mathcal{M}$ , gives a discrete representation of  $\mathcal{M}$ . Let  $\mathcal{D} \subset \mathcal{M}$  be a labeled subset of  $\mathcal{M}$ , and  $S$  is a uniform sample of  $\mathcal{D}$ . In  $S$ , we have  $u(\mathbf{x}) = b(\mathbf{x})$ . An illustration of the computational domain and the point cloud is shown in Fig. 1.1.

In WNLL, we solve the following linear system, (1.4), to extend the label function  $u$  to the entire domain  $P$ .

$$\sum_{\mathbf{y} \in P} R_\delta(\mathbf{x}, \mathbf{y}) (u_\delta(\mathbf{x}) - u_\delta(\mathbf{y})) + \mu \sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y})(u_\delta(\mathbf{x}) - b(\mathbf{y})) = 0, \quad \mathbf{x} \in P, \quad (1.4a)$$

$$u_\delta(\mathbf{x}) = b(\mathbf{x}), \quad \mathbf{x} \in S, \quad (1.4b)$$

where  $R_\delta(\mathbf{x}, \mathbf{y})$  is kernel function given as

$$R_\delta(\mathbf{x}, \mathbf{y}) = C_\delta R \left( \frac{\|\mathbf{x} - \mathbf{y}\|^2}{4\delta^2} \right), \quad (1.5)$$

where  $C_\delta = \frac{1}{\omega_k \delta^k}$  with  $\omega_k$  is the volume of the unit ball in  $\mathbb{R}^k$ .  $R : [0, +\infty) \rightarrow \mathbb{R}$  is a kernel functions satisfying the conditions listed in Assumption 1.1.

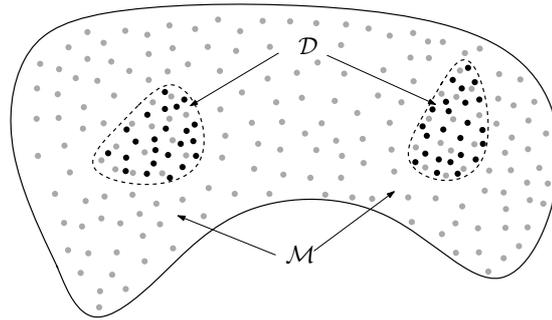


Fig. 1.1. Illustration of the computational domain. Gray points: sample of  $\mathcal{M}$ ; Black points: sample of  $\mathcal{D} \subset \mathcal{M}$ .

In this paper, we will prove that when  $\delta$  goes to 0 and  $n = |P|, m = |S|$  goes to infinity, solution of WNLL converges to the solution of following Laplace-Beltrami equation:

$$\begin{cases} \Delta_{\mathcal{M}}u(\mathbf{x}) = 0, & \mathbf{x} \in \mathcal{M} \setminus \mathcal{D}, \\ u(\mathbf{x}) = b(\mathbf{x}), & \mathbf{x} \in \mathcal{D}, \end{cases} \tag{1.6}$$

where  $\Delta_{\mathcal{M}}$  is the Laplace-Beltrami operator on  $\mathcal{M}$ .

In the analysis, we need to impose following assumptions:

**Assumption 1.1.** • Assumptions on the manifold:  $\mathcal{M}$  be a  $k$ -dimensional closed  $C^\infty$  manifold isometrically embedded in a Euclidean space  $\mathbb{R}^d$ .  $\mathcal{D}$  and  $\partial\mathcal{D}$  are smooth submanifolds of  $\mathbb{R}^d$ . Moreover,  $b(\mathbf{x}) \in C^1(\mathcal{D})$ .

- Assumptions on the kernel functions:
  - (a) Compact support:  $R(r) = 0$  for  $r > 1$  and  $R$  is continuous in  $[0, 1)$ ;
  - (b) Nonnegativity:  $0 \leq R(r) \leq 1$  for any  $r \geq 0$ .
  - (c) Nondegeneracy:  $\exists \delta_0 > 0$  such that  $R(r) \geq \delta_0$  for  $0 \leq r \leq 1/2$ .
  - (d)  $\exists \eta_0 > 0$  such that  $\bar{R}(r) \leq \eta_0 R(r)$  for  $r \geq 0$  with  $\bar{R}(r) = \int_x^1 R(s) ds$
- Assumptions on the point cloud:  $P$  and  $S$  are uniformly distributed on  $\mathcal{M}$  and  $\mathcal{D}$ , respectively.

**Remark 1.1.** The assumption on the kernel function  $R$  is very mild actually. Most often used kernel functions all satisfy these assumptions, for instance,

$$R(r) = \begin{cases} 1, & 0 \leq r < 1, \\ 0, & r \geq 1. \end{cases}$$

When  $\delta$  is small enough, locally the manifold  $\mathcal{M}$ ,  $\mathcal{D}$  and the boundary  $\partial\mathcal{D}$  can be well approximated by Euclidean space. Then it is easy to check that

In the analysis, we consider the limit as  $\delta \rightarrow 0$  and  $n, m \rightarrow \infty$ . So, we assume that  $\delta$  is small enough, i.e.

$$\delta \leq T_0, \tag{1.7}$$

where  $T_0 > 0$  is a constant only depend on  $\mathcal{M}, \mathcal{D}$ . It is easy to check that  $\bar{\omega}_0 > 0$ .

**Proposition 1.1.** *Let*

$$\begin{aligned} \bar{\omega}_0 &= \inf_{0 < \delta \leq T_0} \inf_{\mathbf{x} \in \mathcal{M} \setminus \mathcal{D}} \left( \int_{\mathcal{M} \setminus \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right), \\ \omega_{\mathcal{D}^c} &= \inf_{0 < \delta \leq T_0} \inf_{\mathbf{x} \in \mathcal{D}_\delta \setminus \mathcal{D}} \left( \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\tau_{\mathbf{y}} \right), \\ \omega_0 &= \inf_{0 < \delta \leq T_0} \inf_{\mathbf{x} \in \mathcal{M}} \int_{\mathcal{M}} (R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y}), \\ \omega_1 &= \sup_{0 < \delta \leq T_0} \sup_{\mathbf{x} \in \mathcal{M}} \int_{\mathcal{M}} (R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y}), \\ \bar{\omega}_{\partial \mathcal{D}} &= \sup_{0 < \delta \leq T_0} \sup_{\mathbf{x} \in \mathcal{M}} \left( \delta \int_{\partial \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\tau_{\mathbf{y}} \right), \end{aligned}$$

where  $\mathcal{D}_\delta = \{\mathbf{x} \in \mathcal{M}, \text{dist}(\mathbf{x}, \mathcal{D}) \leq \delta\}$  and

$$\bar{R}_\delta(\mathbf{x}, \mathbf{y}) = C_\delta \bar{R} \left( \frac{\|\mathbf{x} - \mathbf{y}\|^2}{4\delta^2} \right). \tag{1.8}$$

Then we have

$$0 < \bar{\omega}_0, \omega_{\mathcal{D}^c}, \omega_0, \omega_1, \bar{\omega}_{\partial \mathcal{D}} < +\infty.$$

The proof is straightforward and is deferred to appendix.

For samples,  $P$  and  $S$ , we assume they are large enough. More specifically, we assume

$$\frac{1}{\delta^{k+3}\sqrt{n}} (\ln n - 2 \ln \delta + 1)^{1/2} \leq C_0, \quad \frac{1}{\delta^{k+2}\sqrt{m}} (\ln m - 2 \ln \delta + 1)^{1/2} \leq C_0, \tag{1.9}$$

where  $C_0 > 0$  is a constant independent on  $\delta, P$  and  $S$ .

In this paper, we abuse the notation  $C$  to denote any constant independent on  $\delta, P$  and  $S$ . The constant may be different in different places.

The main contribution of this paper is to analyze the relation between the solutions of the Laplace-Beltrami equation (1.6) and the WNLL (1.4). More precisely, we prove the following theorem:

**Theorem 1.1.** *Let  $u_\delta$  solves (1.4) with  $\mu = |P|/|S|$  and  $u$  solves (1.6). Under (1.7) and Assumption 1.1, with probability at least  $1 - 1/(2 \min\{m, n\})$ , where  $n = |P|, m = |S|$ , we have*

$$|u_\delta(\mathbf{x}) - u(\mathbf{x})| \leq C\delta, \quad \forall \mathbf{x} \in P.$$

**Remark 1.2.** In above theorem, the error is bounded in terms of  $\delta$ . We can also bound the error in terms of  $n$  and  $m$  based on assumption (1.9),

$$\frac{1}{\delta^{k+3}\sqrt{n}} (\ln n - 2 \ln \delta + 1)^{1/2} \leq C_0, \quad \frac{1}{\delta^{k+2}\sqrt{m}} (\ln m - 2 \ln \delta + 1)^{1/2} \leq C_0.$$

Theorem 1.1 is a direct consequence of the maximum principle (Theorem 1.2) and the error estimation (Theorem 1.3).

**Theorem 1.2.** *Under the assumptions in Assumption 1.1, with probability at least  $1 - 1/(2n)$ ,  $n = |P|$ ,  $L_{\delta,n}$  has the comparison principle, i.e.*

$$|L_{\delta,n}u(\mathbf{x})| \leq L_{\delta,n}v(\mathbf{x}) \quad \rightarrow \quad |u(\mathbf{x})| \leq v(\mathbf{x}), \quad \forall \mathbf{x} \in P$$

where

$$L_{\delta,n}u(\mathbf{x}) = \sum_{\mathbf{y} \in P} R_{\delta}(\mathbf{x}, \mathbf{y}) (u(\mathbf{x}) - u(\mathbf{y})) + \mu \sum_{\mathbf{y} \in S} R_{\delta}(\mathbf{x}, \mathbf{y}) u(\mathbf{x}), \quad \mathbf{x} \in P. \tag{1.10}$$

Based on the maximum principle, Theorem 1.1 can be proved if we can find an auxiliary function  $v$  such that  $|L_{\delta,n}(u_{\delta} - u)| \leq C\delta L_{\delta,n}v$ .

**Theorem 1.3.** *Let  $u_{\delta}$  and  $u$  solve (1.4) and (1.6) respectively.  $v$  is the solution of (1.11),*

$$\begin{cases} -\Delta_{\mathcal{M}}v(\mathbf{x}) = 1, & \mathbf{x} \in \mathcal{M} \setminus \mathcal{D}, \\ v(\mathbf{x}) = 1, & \mathbf{x} \in \mathcal{D}. \end{cases} \tag{1.11}$$

Under the assumptions in Assumption 1.1, with probability at least  $1 - 1/(2 \min\{m, n\})$ ,  $n = |P|$ ,

$$|L_{\delta,n}(u_{\delta} - u)| \leq C\delta L_{\delta,n}v,$$

$C > 0$  is a constant independent on  $\delta$ ,  $P$  and  $S$ .

Theorems 1.2 and 1.3 will be proved in Sections 2 and 3 respectively. In Section 4, a technical theorem is proved. Some discussions are made in Section 5.

## 2. Maximum Principle (Theorem 1.2)

First, we introduce some notations. For any two points  $\mathbf{x}, \mathbf{y} \in P \cup S$ , we say that they are neighbors if and only if  $R_{\delta}(\mathbf{x}, \mathbf{y}) > 0$ , denoted as  $\mathbf{x} \sim \mathbf{y}$  or  $\mathbf{y} \sim \mathbf{x}$ .  $\mathbf{x}$  and  $\mathbf{y}$  are connected if there exist  $\mathbf{z}_1, \dots, \mathbf{z}_m \in P \cup S$  such that

$$\mathbf{x} \sim \mathbf{z}_1 \sim \dots \sim \mathbf{z}_m \sim \mathbf{y}.$$

We say point cloud  $P$  is  $S$ -connected if for any point  $\mathbf{x} \in P$ , there exists  $\mathbf{y} \in S$ , such that  $\mathbf{x}$  and  $\mathbf{y}$  are connected.

If  $P$  is  $S$ -connected, it is easy to check that  $L_{\delta,n}$  has maximum principle, i.e.

$$L_{\delta,n}u(\mathbf{x}) \geq 0, \quad \mathbf{x} \in P \quad \rightarrow \quad u(\mathbf{x}) \geq 0, \quad \mathbf{x} \in P, \tag{2.1}$$

$$L_{\delta,n}u(\mathbf{x}) \leq 0, \quad \mathbf{x} \in P \quad \rightarrow \quad u(\mathbf{x}) \leq 0, \quad \mathbf{x} \in P, \tag{2.2}$$

and consequently

$$|L_{\delta,n}u(\mathbf{x})| \leq L_{\delta,n}v(\mathbf{x}) \quad \rightarrow \quad |u(\mathbf{x})| \leq v(\mathbf{x}), \quad \forall \mathbf{x} \in P. \tag{2.3}$$

The maximum principle can be proved by contradiction. Suppose that  $L_{\delta,n}u(\mathbf{x}) \geq 0$ ,  $\mathbf{x} \in P$ , but  $u(\mathbf{x}_0) = \min_{\mathbf{x} \in P} u(\mathbf{x}) < 0$ . Since  $P$  is  $S$ -connected, there exist  $\mathbf{z} \in S$  and  $\mathbf{z}_1, \dots, \mathbf{z}_m \in P$ , such that  $\mathbf{x}_0 \sim \mathbf{z}_1 \sim \dots \sim \mathbf{z}_m \sim \mathbf{z}$ .

$L_{\delta,n}u(\mathbf{x}_0) \geq 0$  and  $u(\mathbf{x}_0) = \min_{\mathbf{x} \in P} u(\mathbf{x}) < 0$  imply that  $u(\mathbf{x}) = u(\mathbf{x}_0)$  as long as  $\mathbf{x} \sim \mathbf{x}_0$ . More specifically,  $u(\mathbf{z}_1) = u(\mathbf{x}_0)$ . Then, we move to  $\mathbf{z}_1$  to get that  $u(\mathbf{z}_2) = u(\mathbf{z}_1)$ . Repeating this process, we can show that  $u(\mathbf{z}_m) = \dots = u(\mathbf{z}_1) = u(\mathbf{x}_0) = \min_{\mathbf{x} \in P} u(\mathbf{x})$ .

Now, we compute  $L_{\delta,n}u(\mathbf{z}_m)$ . First,  $u(\mathbf{z}_m) = \min_{\mathbf{x} \in P} u(\mathbf{x})$ , so

$$\sum_{\mathbf{y} \in P} R_{\delta}(\mathbf{x}, \mathbf{y}) (u(\mathbf{z}_m) - u(\mathbf{y})) \leq 0.$$

Moreover,  $u(\mathbf{z}_m) < 0$  and  $\mathbf{z}_m \sim \mathbf{z}$  give that

$$\sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y})u(\mathbf{z}_m) \leq R_\delta(\mathbf{z}_m, \mathbf{z})u(\mathbf{z}_m) < 0.$$

Then we have  $L_{\delta,n}u(\mathbf{z}_m) < 0$  which contradicts  $L_{\delta,n}u(\mathbf{z}_m) \geq 0$ .

In the rest of this section, we will prove that with high probability,  $P$  is  $S$ -connected. To prove this, we need a theorem which will be proved in Section 4.

**Theorem 2.1.** *With probability at least  $1 - 1/(2n)$ ,  $n = |P|$ ,*

$$\sup_{f \in \mathcal{R}_\delta} |I(f) - I_n(f)| \leq \frac{C}{\delta^k \sqrt{n}} (\ln n - 2 \ln \delta + 1)^{1/2}, \tag{2.4}$$

where  $k$  is the dimension of  $\mathcal{M}$ ,

$$I(f) = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} f(\mathbf{x})d\mathbf{x}, \quad I_n(f) = \frac{1}{n} \sum_{\mathbf{x} \in P} f(\mathbf{x}),$$

$|\mathcal{M}|$  is the volume of  $\mathcal{M}$  and  $\mathcal{R}_\delta$  is a function class defined as

$$\mathcal{R}_\delta = \{R_\delta(\mathbf{x}, \cdot) : \mathbf{x} \in \mathcal{M}\}.$$

Suppose  $P$  is not  $S$ -connected. Let

$$\bar{S} = \{\mathbf{x} \in P \cup S : \mathbf{x} \text{ is connected to } S\}, \quad \bar{S}^c = (P \cup S) \setminus \bar{S}.$$

Then  $\bar{S}^c \neq \emptyset$ . Denote

$$\bar{S}_\delta = \left( \bigcup_{\mathbf{x} \in \bar{S}} B(\mathbf{x}; \delta/2) \right) \cap \mathcal{M}, \quad \bar{S}_\delta^c = \left( \bigcup_{\mathbf{x} \in \bar{S}^c} B(\mathbf{x}; \delta/2) \right) \cap \mathcal{M}$$

where  $B(\mathbf{x}; \delta) = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{x} - \mathbf{y}\| \leq \delta\}$ .

Using the definition of  $\bar{S}$  and  $\bar{S}^c$ , we know that  $\bar{S}_\delta \cap \bar{S}_\delta^c = \emptyset$ , hence

$$\partial \bar{S}_\delta \cap \bar{S}_\delta^c = \emptyset,$$

where  $\partial \bar{S}_\delta$  is the boundary of  $\bar{S}_\delta$  in  $\mathbb{R}^d$ . Furthermore, since  $\mathcal{M}$  is connected, we have

$$\partial \bar{S}_\delta \cap \mathcal{M} \neq \emptyset.$$

Choose any  $\mathbf{x}_0 \in \partial \bar{S}_\delta \cap \mathcal{M}$ , we also have that  $\mathbf{x}_0 \notin \bar{S}_\delta^c$ , which implies that

$$R_{\delta/4}(\mathbf{x}_0, \mathbf{y}) = 0, \quad \forall \mathbf{y} \in P.$$

It follows that

$$I_n(R_{\delta/4}(\mathbf{x}_0, \cdot)) = \frac{1}{n} \sum_{\mathbf{x} \in P} R_{\delta/4}(\mathbf{x}_0, \mathbf{x}) = 0.$$

On the other hand, using Theorem 2.1 and Proposition 1.1,

$$I_n(R_{\delta/4}(\mathbf{x}_0, \cdot)) \geq \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} R_{\delta/4}(\mathbf{x}_0, \mathbf{x})d\mathbf{x} - \frac{C}{\delta^k \sqrt{n}} (\ln n - 2 \ln \delta + 1)^{1/2} \geq \frac{1}{2} \omega_0 > 0.$$

Here we also use the assumption that  $n$  is large enough (1.9). Then it has been proved  $P$  is  $S$ -connected with probability at least  $1 - 1/(2n)$  by contradiction.

### 3. Error Estimate (Theorem 1.3)

Let  $u$  and  $v$  solve (1.6) and (1.11) respectively. First, we list several facts and notations which will be used in the subsequent analysis.

- $u$  and  $v$  are both Lipschitz continuous in  $\mathcal{M}$ . The Lipschitz constants are denoted as  $\xi_u$  and  $\xi_v$  respectively.
- $u, v$  are both smooth functions in  $\mathcal{M} \setminus \mathcal{D}$  and  $\mathcal{D}$ . We need that  $u, v \in C^3$  in  $\mathcal{M} \setminus \mathcal{D}$  and  $u, v \in C^1$  in  $\mathcal{D}$ .

Let  $e_\delta(\mathbf{x}) = u_\delta(\mathbf{x}) - u(\mathbf{x})$ .  $u_\delta$  and  $u$  solve (1.4) and (1.6) respectively. Direct calculation shows that for  $\mathbf{x} \in P$

$$L_{\delta,n}e_\delta(\mathbf{x}) = \sum_{\mathbf{y} \in P} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) + \mu \sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - b(\mathbf{y})). \tag{3.1}$$

We need to find an upper bound of the right hand side of (3.1). First, we have

$$\begin{aligned} & \frac{1}{n} |L_{\delta,n}e_\delta(\mathbf{x})| \\ & \leq \left| \frac{1}{n} \sum_{\mathbf{y} \in P} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) \right| + \xi_u \mu \left| \frac{1}{n} \sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y}) \|\mathbf{x} - \mathbf{y}\| \right| \\ & \leq \left| \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} \right| + \xi_u \delta \left| \frac{1}{m} \sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y}) \right| + \Xi_n \\ & \leq \left| \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} \right| + \xi_u \delta \left| \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right| + \Xi_n + \Xi_m, \end{aligned} \tag{3.2}$$

where

$$\Xi_P^u = \sup_{\mathbf{x} \in \mathcal{M}} \left| \frac{1}{n} \sum_{\mathbf{y} \in P} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) - \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} \right|, \tag{3.3}$$

$$\Xi_S = \sup_{\mathbf{x} \in \mathcal{M}} \left| \frac{1}{m} \sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y}) - \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right|. \tag{3.4}$$

To find an upper bound of the first term of (3.2), we need the following theorem which can be found in [8].

**Theorem 3.1.** *Let  $u(\mathbf{x}) \in C^3(\mathcal{M})$  and*

$$\begin{aligned} I_{bd} &= \sum_{j=1}^d \int_{\partial \mathcal{M}} n^j(\mathbf{y})(\mathbf{x} - \mathbf{y}) \cdot \nabla(\nabla^j u(\mathbf{y})) \bar{R}_t(\mathbf{x}, \mathbf{y}) d\tau_{\mathbf{y}}, \\ I_{in} &= \frac{1}{\delta^2} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} + \int_{\mathcal{M}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) \Delta_{\mathcal{M}} u(\mathbf{y}) d\mathbf{y} \\ & \quad - \int_{\partial \mathcal{M}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) \frac{\partial u}{\partial \mathbf{n}}(\mathbf{y}) d\tau_{\mathbf{y}} - I_{bd}, \end{aligned} \tag{3.5}$$

where  $\mathbf{n}(\mathbf{y}) = (n^1(\mathbf{y}), \dots, n^d(\mathbf{y}))$  is the out normal vector of  $\partial\mathcal{M}$  at  $\mathbf{y}$ ,  $\nabla^j$  is the  $j$ th component of gradient  $\nabla$ ,  $\bar{R}_\delta(\mathbf{x}, \mathbf{y}) = C_\delta \bar{R}\left(\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4\delta^2}\right)$  and  $\bar{R}(r) = \int_r^\infty R(s)ds$ .

Then there exist constants  $C, T_0$  depending only on  $\mathcal{M}$ , so that,

$$|I_{in}| \leq C\delta\|u\|_{C^3(\mathcal{M})}, \tag{3.6}$$

as long as  $\delta \leq T_0$ .

Since  $u$  is piecewise smooth in  $\mathcal{M}\setminus\mathcal{D}$  and  $\mathcal{D}$ , we estimate the error in  $\mathcal{M}\setminus\mathcal{D}$  and  $\mathcal{D}$  separately. For  $\mathbf{x} \in \mathcal{M}\setminus\mathcal{D}$ , using above theorem, for  $u$  solves (3.6), we have

$$\begin{aligned} & \left| \int_{\mathcal{M}\setminus\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y}))d\mathbf{y} \right| \\ & \leq C\delta^3\|u\|_{C^3(\mathcal{M}\setminus\mathcal{D})} + \delta^2 \left| \int_{\partial\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) \frac{\partial u}{\partial \mathbf{n}}(\mathbf{y})dS_{\mathbf{y}} \right| + \delta^2 |I_{bd}(\mathcal{D})| \\ & \leq C\delta^3\|u\|_{C^3(\mathcal{M}\setminus\mathcal{D})} + C(\delta^2\|u\|_{C^1(\mathcal{M}\setminus\mathcal{D})} + \delta^3\|u\|_{C^2(\mathcal{M}\setminus\mathcal{D})}) \int_{\partial\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}}, \end{aligned} \tag{3.7}$$

where

$$I_{bd}(\mathcal{D}) = \sum_{j=1}^d \int_{\partial\mathcal{D}} n^j(\mathbf{y})(\mathbf{x} - \mathbf{y}) \cdot \nabla(\nabla^j u(\mathbf{y}))\bar{R}_t(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}}. \tag{3.8}$$

The integral over  $\mathcal{D}$  can be bounded based on the Lipschitz continuity of  $u$ ,

$$\left| \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y}))d\mathbf{y} \right| \leq 2\xi_u\delta \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y})d\mathbf{y}. \tag{3.9}$$

Then, we get for  $\mathbf{x} \in P \cap (\mathcal{M}\setminus\mathcal{D})$ ,

$$\begin{aligned} & \left| \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y}))d\mathbf{y} \right| \\ & \leq C\delta^3\|u\|_{C^3(\mathcal{M}\setminus\mathcal{D})} + C(\delta^2\|u\|_{C^1(\mathcal{M}\setminus\mathcal{D})} + \delta^3\|u\|_{C^2(\mathcal{M}\setminus\mathcal{D})}) \int_{\partial\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}} \\ & \quad + 2\xi_u\delta \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y})d\mathbf{y}. \end{aligned} \tag{3.10}$$

Next, we try to bound  $\int_{\partial\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}}$ . For  $\mathbf{x} \in P \cap (\mathcal{M}\setminus\mathcal{D})$  with  $\text{dist}(x, \partial\mathcal{D}) \leq \delta$ , using proposition 1.1,

$$\int_{\partial\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}} \leq \frac{\bar{\omega}_{\partial\mathcal{D}}}{\delta} \leq \frac{\bar{\omega}_{\partial\mathcal{D}}}{\delta\omega_{\mathcal{D}^c}} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y})d\mathbf{y}. \tag{3.11}$$

For  $\mathbf{x} \in P \cap (\mathcal{M}\setminus\mathcal{D})$  with  $\text{dist}(x, \partial\mathcal{D}) > \delta$

$$\begin{aligned} & \int_{\partial\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}} \leq \int_{\partial\mathcal{D}} \frac{(\mathbf{x} - \mathbf{y}) \cdot \mathbf{n}(\mathbf{y})}{\delta} \bar{R}_\delta(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}} \\ & = \frac{1}{\delta} \int_{\mathcal{D}} \text{div}((\mathbf{x} - \mathbf{y})\bar{R}_\delta(\mathbf{x}, \mathbf{y}))d\mathbf{y} \\ & = \frac{k}{\delta} \int_{\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y})d\mathbf{y} + \frac{1}{\delta} \int_{\mathcal{D}} \frac{\|\mathbf{x} - \mathbf{y}\|^2}{4\delta^2} R_\delta(\mathbf{x}, \mathbf{y})d\mathbf{y} \\ & \leq \frac{C}{\delta} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y})d\mathbf{y}. \end{aligned} \tag{3.12}$$

In the last inequality, we use the assumption that  $\bar{R}(r) \leq \eta_0 R(r)$ . Combining (3.11) and (3.12), we have for  $\mathbf{x} \in P \cap (\mathcal{M} \setminus \mathcal{D})$

$$\int_{\partial \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) dS_{\mathbf{y}} \leq \frac{C}{\delta} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y}. \tag{3.13}$$

Substituting (3.13) in (3.10), we get for  $\mathbf{x} \in P \cap (\mathcal{M} \setminus \mathcal{D})$ ,

$$\begin{aligned} & \left| \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} \right| \\ & \leq C\delta^3 \|u\|_{C^3(\mathcal{M} \setminus \mathcal{D})} + C\delta (\|u\|_{C^1(\mathcal{M} \setminus \mathcal{D})} + \delta \|u\|_{C^2(\mathcal{M} \setminus \mathcal{D})} + 2\xi_u) \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y}. \end{aligned} \tag{3.14}$$

For  $\mathbf{x} \in P \cap \mathcal{D}$ , the bound is straightforward, just using the Lipschitz continuity of  $u$ ,

$$\left| \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y})) d\mathbf{y} \right| \leq \xi_u \delta \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \leq \omega_1 \xi_u \delta, \quad \mathbf{x} \in P \cap \mathcal{D}. \tag{3.15}$$

Substituting (3.14) and (3.15) in (3.2), we have

$$\frac{1}{n} |L_{\delta,n} e_\delta(\mathbf{x})| \leq \frac{\omega_1 \xi_u \delta}{|\mathcal{M}|} + \frac{\xi_u \delta}{|\mathcal{D}|} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} + \Xi_P + \Xi_S, \quad \mathbf{x} \in P \cap \mathcal{D}, \tag{3.16}$$

$$\frac{1}{n} |L_{\delta,n} e_\delta(\mathbf{x})| \leq \frac{C\delta^3}{|\mathcal{M}|} \|u\|_{C^3(\mathcal{M} \setminus \mathcal{D})} + \frac{C\delta}{|\mathcal{D}|} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} + \Xi_P + \Xi_S, \quad \mathbf{x} \in P \cap (\mathcal{M} \setminus \mathcal{D}). \tag{3.17}$$

Next, we want to get a lower bound of  $L_{\delta,n} v(\mathbf{x})$  with  $v$  given in (1.11).

$$\begin{aligned} \frac{1}{n} L_{\delta,n} v(\mathbf{x}) &= \frac{1}{n} \sum_{\mathbf{y} \in P} R_\delta(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y})) + v(\mathbf{x}) \frac{1}{m} \sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y}) \\ &\geq \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y})) d\mathbf{y} + \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} - \Xi_P^v - \Xi_S, \end{aligned} \tag{3.18}$$

where

$$\begin{aligned} \Xi_P^v &= \sup_{\mathbf{x} \in \mathcal{M}} \left| \frac{1}{n} \sum_{\mathbf{y} \in P} R_\delta(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y})) - \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y})) d\mathbf{y} \right|, \\ \Xi_S &= \sup_{\mathbf{x} \in \mathcal{M}} \left| \frac{1}{m} \sum_{\mathbf{y} \in S} R_\delta(\mathbf{x}, \mathbf{y}) - \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right|. \end{aligned} \tag{3.19}$$

Using Theorem 3.1 again, we have for  $\mathbf{x} \in \mathcal{M} \setminus \mathcal{D}$

$$\begin{aligned} & \int_{\mathcal{M} \setminus \mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y})) d\mathbf{y} \\ & \geq \delta^2 \int_{\mathcal{M} \setminus \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} - 2\delta^2 \left| \int_{\partial \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) \frac{\partial v}{\partial \mathbf{n}}(\mathbf{y}) dS_{\mathbf{y}} \right| - C\delta^3 \|v\|_{C^3(\mathcal{M} \setminus \mathcal{D})} \\ & \geq \delta^2 \int_{\mathcal{M} \setminus \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} - C\delta^2 \|v\|_{C^1(\mathcal{M} \setminus \mathcal{D})} \left| \int_{\partial \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) dS_{\mathbf{y}} \right| - C\delta^3 \|v\|_{C^3(\mathcal{M} \setminus \mathcal{D})} \\ & \geq \bar{\omega}_0 \delta^2 - C\delta \|v\|_{C^1(\mathcal{M} \setminus \mathcal{D})} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} - C\delta^3 \|v\|_{C^3(\mathcal{M} \setminus \mathcal{D})}, \end{aligned}$$

and

$$\int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y}))d\mathbf{y} = \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - 1)d\mathbf{y} \geq 0, \quad \mathbf{x} \in \mathcal{M} \setminus \mathcal{D}.$$

This is due to the fact that  $v(\mathbf{x}) \geq 1, \quad \mathbf{x} \in \mathcal{M}$ . Then we have for  $\mathbf{x} \in \mathcal{M} \setminus \mathcal{D}$

$$\begin{aligned} & \int_{\mathcal{M}} R_{\delta}(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y}))d\mathbf{y} \\ & \geq \bar{\omega}_0 \delta^2 - C\delta \|v\|_{C^1(\mathcal{M} \setminus \mathcal{D})} \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})d\mathbf{y} - C\delta^3 \|v\|_{C^3(\mathcal{M} \setminus \mathcal{D})}. \end{aligned} \tag{3.20}$$

In  $P \cap \mathcal{D}$ , using the Lipschitz continuity of  $v$ ,

$$\int_{\mathcal{M}} R_{\delta}(\mathbf{x}, \mathbf{y})(v(\mathbf{x}) - v(\mathbf{y}))d\mathbf{y} \geq -\xi_v \delta \int_{\mathcal{M}} R_{\delta}(\mathbf{x}, \mathbf{y}), \quad \mathbf{x} \in P \cap \mathcal{D}. \tag{3.21}$$

Combining (3.18), (3.20) and (3.21), we obtain

$$\frac{1}{n} L_{\delta,n} v(\mathbf{x}) \geq \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})d\mathbf{y} - \frac{\xi_v \omega_1 \delta}{|\mathcal{M}|} - \Xi_P^v - \Xi_S, \quad \mathbf{x} \in P \cap \mathcal{D}, \tag{3.22}$$

$$\frac{1}{n} L_{\delta,n} v(\mathbf{x}) \geq \frac{1}{2|\mathcal{D}|} \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})d\mathbf{y} + \frac{\bar{\omega}_0 \delta^2}{2|\mathcal{M}|} - \Xi_P^v - \Xi_S, \quad \mathbf{x} \in P \cap (\mathcal{M} \setminus \mathcal{D}). \tag{3.23}$$

Now we need to deal with  $\Xi_P^u, \Xi_P^v$  and  $\Xi_S$  which are given in (3.3), (3.19) and (3.4).

**Theorem 3.2.** *With probability at least  $1 - 1/(2n)$ ,  $n = |P|$ ,*

$$\sup_{f \in \mathcal{R}_{\delta}} |I(f) - I_n(f)| \leq \frac{C}{\delta^k \sqrt{n}} (\ln n - 2 \ln \delta + 1)^{1/2}, \tag{3.24}$$

where  $k$  is the dimension of  $\mathcal{M}$ ,

$$I(f) = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} f(\mathbf{x})d\mathbf{x}, \quad I_n(f) = \frac{1}{n} \sum_{\mathbf{x} \in P} f(\mathbf{x}),$$

$|\mathcal{M}|$  is the volume of  $\mathcal{M}$  and  $\mathcal{R}_{\delta}$  is a function class defined as

$$\bar{\mathcal{R}}_{\delta} = \{R_{\delta}(\mathbf{x}, \cdot), R_{\delta}(\mathbf{x}, \cdot)u(\cdot), R_{\delta}(\mathbf{x}, \cdot)v(\cdot) : \mathbf{x} \in \mathcal{M}, u \text{ and } v \text{ solves (1.6) and (1.11) respectively.}\}$$

This theorem will be proved in Section 4 using the empirical process theory.

$$\frac{1}{\delta^{k+3} \sqrt{n}} (\ln n - 2 \ln \delta + 1)^{1/2} \leq C_0, \quad \frac{1}{\delta^{k+2} \sqrt{m}} (\ln m - 2 \ln \delta + 1)^{1/2} \leq C_0. \tag{3.25}$$

Using Theorem 3.2 and the assumption (1.7) we know that with high probability,

$$\Xi_P^u \leq C_0 \delta^3, \quad \Xi_P^v \leq C_0 \delta^3, \quad \Xi_S \leq C_0 \delta^2.$$

Under the assumption that  $C_0$  is sufficiently small, using (3.22) and (3.23), we have

$$\frac{1}{n} L_{\delta,n} v(\mathbf{x}) \geq \begin{cases} \frac{1}{2|\mathcal{D}|} \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})d\mathbf{y}, & \mathbf{x} \in P \cap \mathcal{D}, \\ \frac{1}{4|\mathcal{M}|} \delta^2 \bar{\omega}_0 + \frac{1}{2|\mathcal{D}|} \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})d\mathbf{y}, & \mathbf{x} \in P \cap (\mathcal{M} \setminus \mathcal{D}). \end{cases}$$

It follows from (3.17) and (3.16) that

$$\frac{1}{n}|L_{\delta,n}e_{\delta}(\mathbf{x})| \leq \begin{cases} C\delta + \frac{C\delta}{|\mathcal{D}|} \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})d\mathbf{y}, & \mathbf{x} \in P \cap \mathcal{D}, \\ C\delta^3 + \frac{C\delta}{|\mathcal{D}|} \int_{\mathcal{D}} R_{\delta}(\mathbf{x}, \mathbf{y})d\mathbf{y}, & \mathbf{x} \in P \cap (\mathcal{M} \setminus \mathcal{D}). \end{cases}$$

Finally we have

$$|L_{\delta,n}e_{\delta}(\mathbf{x})| \leq C\delta L_{\delta,n}v(\mathbf{x}), \quad \mathbf{x} \in P. \tag{3.26}$$

The proof of Theorem 1.3 is completed.  $\square$

### 4. Entropy Bound

In this section, we will prove Theorems 2.1 and 3.2. The method we use is to estimate the covering number of the function classes. First, we introduce the definition of the covering number.

Let  $(Y, d)$  be a metric space and set  $F \subset Y$ . For every  $\epsilon > 0$ , denote by  $N(\epsilon, F, d)$  the minimal number of open balls (with respect to the metric  $d$ ) that are needed to cover  $F$ . That is, the minimal cardinality of the set  $\{y_1, \dots, y_m\} \subset Y$  with the property that every  $f \in F$  has some  $y_i$  such that  $d(f, y_i) < \epsilon$ . The set  $\{y_1, \dots, y_m\}$  is called an  $\epsilon$ -cover of  $F$ . The logarithm of the covering numbers is called the entropy of the set. For every sample  $\{x_1, \dots, x_n\}$ , let  $\mu_n$  be the empirical measure supported on that sample. For  $1 \leq p < \infty$  and a function  $f$ , put  $\|f\|_{L_p(\mu_n)} = (\frac{1}{n} \sum_{i=1}^n |f(x_i)|^p)^{1/p}$  and set  $\|f\|_{\infty} = \max_{1 \leq i \leq n} |f(x_i)|$ . Let  $N(\epsilon, F, L_p(\mu_n))$  be the covering numbers of  $F$  at scale  $\epsilon$  with respect to the  $L_p(\mu_n)$  norm.

We will use following theorem which is well known in empirical process theory.

**Theorem 4.1 (Theorem 2.3 in [9]).** *Let  $F$  be a class of functions from  $\mathcal{M}$  to  $[-1, 1]$  and set  $\mu$  to be a probability measure on  $\mathcal{M}$ . Let  $(\mathbf{x}_i)_{i=1}^{\infty}$  be independent random variables distributed according to  $\mu$ . For any  $\epsilon > 0$  and every  $n \geq 8/\epsilon^2$ ,*

$$\mathbb{P} \left( \sup_{f \in F} \left| \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i) - \int_{\mathcal{M}} f(\mathbf{x})\mu(\mathbf{x})d\mathbf{x} \right| > \epsilon \right) \leq 8\mathbb{E}_{\mu}[N(\epsilon/8, F, L_1(\mu_n))] \exp(-n\epsilon^2/128).$$

Note that

$$L_1(\mu_n) \leq L_{\infty}(\mu_n) \leq L_{\infty},$$

where  $\|f\|_{L_{\infty}} = \max_{\mathbf{x} \in \mathcal{M}} |f(\mathbf{x})|$ . Then we get

$$N(\epsilon, F, L_1(\mu_n)) \leq N(\epsilon, F, L_{\infty}),$$

which implies following corollary.

**Corollary 4.1.** *Let  $F$  be a class of functions from  $\mathcal{M}$  to  $[-1, 1]$  and set  $\mu$  to be a probability measure on  $\mathcal{M}$ . Let  $(\mathbf{x}_i)_{i=1}^{\infty}$  be independent random variables distributed according to  $\mu$ . For any  $\epsilon > 0$  and every  $n \geq 8/\epsilon^2$ ,*

$$\mathbb{P} \left( \sup_{f \in F} \left| \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i) - \int_{\mathcal{M}} f(\mathbf{x})\mu(\mathbf{x})d\mathbf{x} \right| > \epsilon \right) \leq 8N(\epsilon/8, F, L_{\infty}) \exp(-n\epsilon^2/128),$$

where  $N(\epsilon, F, L_{\infty})$  is the covering numbers of  $F$  at scale  $\epsilon$  with respect to the  $L_{\infty}$  norm

To estimate the error, we need to rewrite Corollary 4.1 in another form:

**Corollary 4.2.** *Let  $F$  be a class of functions from  $\mathcal{M}$  to  $[-1, 1]$ . Let  $(\mathbf{x}_i)_{i=1}^\infty$  be independent random variables distributed according to  $p$ , where  $p$  is the probability distribution. Then with probability at least  $1 - \delta$ , we have*

$$\sup_{f \in F} |p(f) - p_n(f)| \leq \sqrt{\frac{128}{n} \left( \ln N(\sqrt{2/n}, F, L_\infty) + \ln \frac{8}{\delta} \right)},$$

where

$$p(f) = \int_{\mathcal{M}} f(\mathbf{x})p(\mathbf{x})d\mathbf{x}, \quad p_n(f) = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i).$$

*Proof.* Using Corollary 4.1, with probability at least  $1 - \delta$ ,

$$\sup_{f \in F} |p(f) - p_n(f)| \leq \epsilon_\delta,$$

where  $\epsilon_\delta$  is determined by

$$\epsilon_\delta = \sqrt{\frac{128}{n} \left( \ln N(\epsilon_\delta/8, F, L_\infty) + \ln \frac{8}{\delta} \right)}.$$

Notice that  $N(\epsilon_\delta/8, F, L_\infty) \geq 1$  and  $0 < \delta < 1$ , we have

$$\epsilon_\delta \geq \sqrt{\frac{128 \ln 8}{n}} > \sqrt{\frac{128}{n}} = 8\sqrt{\frac{2}{n}},$$

which gives that

$$N(\epsilon_\delta/8, F, L_\infty) \leq N\left(\sqrt{\frac{2}{n}}, F, L_\infty\right).$$

Then, we have

$$\epsilon_\delta \leq \sqrt{\frac{128}{n} \left( \ln N\left(\sqrt{\frac{2}{n}}, F, L_\infty\right) + \ln \frac{8}{\delta} \right)},$$

which proves the corollary. □

The above corollaries provide a tool to estimate the integral error on random samples. To apply the above corollaries in our problem, the key point is to obtain the estimates of the covering number of function class  $\mathcal{R}_\delta$ .

Since the kernel  $R \in C^1(\mathcal{M})$  and  $\mathcal{M} \in C^\infty$ , we have for any  $\mathbf{x}, \mathbf{y} \in \mathcal{M}$

$$\left| R\left(\frac{\|\mathbf{x} - \mathbf{y}\|^2}{4\delta^2}\right) - R\left(\frac{\|\mathbf{z} - \mathbf{y}\|^2}{4\delta^2}\right) \right| \leq \frac{C}{\delta} \|\mathbf{x} - \mathbf{z}\|.$$

This gives an easy bound of  $N(\epsilon, \mathcal{R}_\delta, L_\infty)$ ,

$$N(\epsilon, \mathcal{R}_\delta, L_\infty) \leq \left(\frac{C}{\epsilon\delta}\right)^k. \tag{4.1}$$

Using the Corollary 4.2, with probability at least  $1 - 1/(2n)$ ,

$$\sup_{f \in \mathcal{R}_\delta} |p(f) - p_n(f)| \leq \frac{C}{\delta^k \sqrt{n}} (\ln n - 2 \ln \delta + 1)^{1/2}. \tag{4.2}$$

This proves Theorem 2.1. Theorem 3.2 can be proved similarly using the fact that  $u$  (solution of (1.6)) and  $v$  (solution of (1.11)) are both smooth.

### 5. Discussion and Future Works

In this paper, we analyzed the convergence of the weighted nonlocal Laplacian (WNLL) on the random point cloud. The analysis reveals that the scaling weight in WNLL is critical in the convergence guarantee and it should have the same order as  $|P|/|S|$ , i.e.  $\mu \sim |P|/|S|$ . The result in this paper provides a theoretical foundation for WNLL.

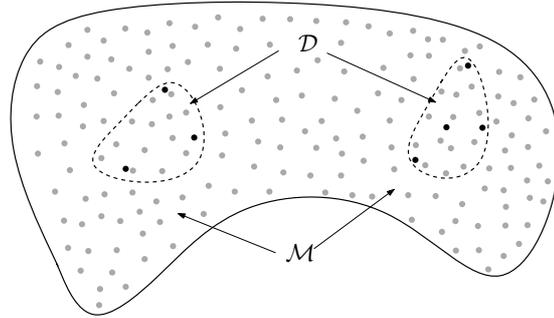


Fig. 5.1. Illustration of the computational domain with extremely low labeling rate.

As illustrated in Fig. 5.1, if sample  $S$  is very sparse such that assumption (1.7) does not hold. Then we cannot guarantee the convergence even in the WNLL. With very low labeling rate, actually, the whole framework of harmonic extension fails [10,14]. We should use other approach to get a smooth interpolation. Furthermore, our analysis also shows that the convergence may fail with extremely low labeling rate. In this case, we should consider other approaches. One interesting option is to minimize  $L_\infty$  norm of the gradient instead of the  $L_2$  norm, i.e. to solve the following optimization problem

$$\min_u \left( \max_{\mathbf{x} \in P \cup S} \left( \sum_{\mathbf{y} \in P \cup S} w(\mathbf{x}, \mathbf{y})(u(\mathbf{x}) - u(\mathbf{y}))^2 \right)^{1/2} \right),$$

with the constraint

$$u(\mathbf{x}) = b(\mathbf{x}), \quad \mathbf{x} \in S.$$

This approach is closely related to the infinity Laplacian [1,5]. The above optimization problem can be solved by the split Bregman iteration. An interesting observation is that the WNLL can accelerate the convergence of the split Bregman iteration and improve its efficiency. We will further explore these aspects in our future work.

#### A. Proof of Proposition 1.1

Since  $\mathcal{M}, \mathcal{D}, \partial\mathcal{D}$  are smooth enough, when  $\delta$  is small, in the support of the kernel functions, they can be well approximated by the tangent space. To prove Proposition 1.1, we only need to compute the integrals with  $\mathcal{M}, \mathcal{D}, \partial\mathcal{D}$  are approximated by tangent space.

With the tangent space approximation, it is easy to check that for  $\mathbf{x} \in \mathcal{M} \setminus \mathcal{D}$

$$\begin{aligned} \int_{\mathcal{M} \setminus \mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} &= \int_{(\mathcal{M} \setminus \mathcal{D}) \cap \text{supp}(\bar{R}_\delta(\mathbf{x}, \cdot))} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &= \int_{(\mathcal{M} \setminus \mathcal{D}) \cap B(\mathbf{x}, 2\delta)} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &\geq \frac{1}{2} s_k \left( \int_0^1 s \bar{R}(s) ds \right) - O(\delta). \end{aligned}$$

For  $\mathbf{x} \in \mathcal{M}$ ,

$$\begin{aligned} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} &= \int_{\mathcal{M} \cap \text{supp}(R_\delta(\mathbf{x}, \cdot))} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &= \int_{\mathcal{M} \cap B(\mathbf{x}, 2\delta)} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &\geq s_k \left( \int_0^1 s R(s) ds \right) - O(\delta), \end{aligned}$$

where  $s_k$  is the area of unit sphere in  $\mathbb{R}^k$ .

Using tangent space approximation, upper bound is also easy to get.

$$\begin{aligned} \int_{\mathcal{M}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} &= \int_{\mathcal{M} \cap \text{supp}(R_\delta(\mathbf{x}, \cdot))} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &= \int_{\mathcal{M} \cap B(\mathbf{x}, 2\delta)} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &\leq s_k \left( \int_0^1 s R(s) ds \right) + O(\delta). \end{aligned}$$

Notice that  $\partial\mathcal{D}$  is  $k - 1$  dimensional manifold, using tangent space of  $\partial\mathcal{D}$  at  $\mathbf{x} \in \partial\mathcal{D}$  to approximate  $\partial\mathcal{D} \cap B(\mathbf{x}, 2\delta)$ , we obtain

$$\begin{aligned} \delta \int_{\partial\mathcal{D}} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\tau_{\mathbf{y}} &= \delta \int_{\partial\mathcal{D} \cap \text{supp}(\bar{R}_\delta(\mathbf{x}, \cdot))} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\tau_{\mathbf{y}} \\ &\leq \delta \int_{\partial\mathcal{D} \cap B(\mathbf{x}, 2\delta)} \bar{R}_\delta(\mathbf{x}, \mathbf{y}) d\tau_{\mathbf{y}} \\ &\leq s_{k-1} \left( \int_0^1 s \bar{R}(s) ds \right) + O(\delta) \end{aligned}$$

where  $s_{k-1}$  is the area of unit sphere in  $\mathbb{R}^{k-1}$ .

For  $\text{dist}(\mathbf{x}, \mathcal{D}) < \delta$ ,

$$\begin{aligned} \int_{\mathcal{D}} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} &= \int_{\mathcal{D} \cap \text{supp}(R_\delta(\mathbf{x}, \cdot))} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &\geq \int_{\mathcal{D} \cap B(\mathbf{x}, \sqrt{2}\delta)} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \\ &\geq \int_{\mathcal{D} \cap B(\bar{\mathbf{x}}, (\sqrt{2}-1)\delta)} R_\delta(\mathbf{x}, \mathbf{y}) d\mathbf{y} \geq \frac{1}{2} (\sqrt{2} - 1)^k \delta_0 - O(\delta) \end{aligned}$$

where  $\bar{\mathbf{x}} = \arg \min_{\mathbf{x} \in \partial\mathcal{D}} |x - \bar{\mathbf{x}}|$ .

**Acknowledgments.** Research supported by NSFC Grant 12071244 and NSF DMS-1924935.

## References

- [1] E. Abderrahim, D. Xavier, L. Zakaria and L. Olivier, Nonlocal infinity laplacian equation on graphs with applications in image processing and machine learning, *Mathematics and Computers in Simulation*, **102** (2014), 153–163.
- [2] A. Buades, B. Coll and J.M. Morel, A review of image denoising algorithms, with a new one, *Multiscale Model. Simul.*, **4** (2005), 490–530.
- [3] A. Buades, B. Coll and J.M. Morel, Neighborhood filters and pde's, *Numer. Math.*, **105** (2006), 1–34.
- [4] F.R.K. Chung, Spectral Graph Theory, American Mathematical Society, 1997.
- [5] M. Ghoniem, A. Elmoataz and O. Lezoray, Discrete infinity harmonic functions: Towards a unified interpolation framework on graphs, in IEEE International Conference on Image Processing, 2011.
- [6] G. Gilboa and S. Osher, Nonlocal linear image regularization and supervised segmentation, *Multiscale Model. Simul.*, **6** (2007), 595–630.
- [7] G. Gilboa and S. Osher, Nonlocal operators with applications to image processing, *Multiscale Model. Simul.*, **7** (2008), 1005–1028.
- [8] Z. Li and Z.Q Shi, A convergent point integral method for isotropic elliptic equations on point cloud, *SIAM: Multiscale Modeling Simulation*, **14** (2016), 874–905.
- [9] Shahar Mendelson, A few notes on statistical learning theory, in *Lecture Notes in Computer Science*, **2600** 2003, 1–40.
- [10] B. Nadler, N. Srebro and X.Y. Zhou, Semi-supervised learning with the graph laplacian: The limit of infinite unlabelled data, NIPS, (2009).
- [11] Z.Q. Shi, S. Osher and W. Zhu, Weighted nonlocal laplacian on interpolation from sparse data, *Journal of Scientific Computing*, **73** (2017), 1164–1177.
- [12] Z.Q. Shi, J. Sun and M.H. Tian, Harmonic extension on point cloud, *SIAM Multiscale Modeling & Simulation*, **16** (2018), 215–247.
- [13] B. Wang, X. Luo, Z. Li, W. Zhu, Z. Shi and S. Osher, Deep neural networks with interpolating function as output activation, The Thirty-second Annual Conference on Neural Information Processing Systems (NeurIPS), (2018).
- [14] X.Y. Zhou and M. Belkin, Semi-supervised learning by higher order regularization, NIPS, (2011).
- [15] W. Zhu, B. Wang, R. Barnard, C. Hauck, F. Jenko and S. Osher, Scientific data interpolation with low dimensional manifold model, *J. Comput. Phys*, **352** (2018), 213–245.
- [16] X.J. Zhu, Z. Ghahramani and J.D. Lafferty, Semi-supervised learning using gaussian fields and harmonic functions, in Machine Learning, Proceedings of the Twentieth International Conference ICML 2003), August 21–24, 2003, Washington, DC, USA, 2003, 912–919.