ERROR ESTIMATION OF THE WEIGHTED NONLOCAL LAPLACIAN ON RANDOM POINT CLOUD

ZUOQIANG SHI*, BAO WANG †, AND STANLEY OSHER ‡

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

Keywords: 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 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 formulated as: Let $P = \{\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n\}$ and $S = \{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_m\}$ 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 S, denoted as $b(\boldsymbol{s})$ for any $\boldsymbol{s} \in S$. The interpolation methods are used to compute u over the whole point cloud $\bar{P}$ from the given values over S.

In nonlocal Laplacian, which is widely used in nonlocal methods for image processing [1, 2, 6, 7], the interpolating function is obtained by minimizing the following energy functional

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

with the constraint

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

Here $w(\boldsymbol{x}, \boldsymbol{y})$ is a given weight function, typically chosen to be Gaussian, i.e., $w(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{\sigma^2})$ with $\sigma > 0$ being a parameter, and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$. In graph theory and machine learning literature, nonlocal Laplacian is also called graph Laplacian [3, 18].

Graph Laplacian works very well with 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 [12, 11]. WNLL is proposed to fix this issue. In WNLL, energy functional in (1.1) is modified by adding a weight, $\frac{|\bar{P}|}{|S|}$, to balance the labeled and unlabeled terms, which

*Department of Mathematical Sciences, Tsinghua University, Beijing, China, 100084. *Email: zqshi@tsinghua.edu.cn.*

†Department of Mathematics, University of California, Los Angeles, CA 90095, USA, *Email: wangbaonj@gmail.com*

‡Department of Mathematics, University of California, Los Angeles, CA 90095, USA, *Email: sjo@math.ucla.edu*

resulting in

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

with the constraint

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

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

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

This linear system can be solved efficiently by conjugate gradient iteration. The advantages of the WNLL compared to the graph Laplacian has been shown evidently in image inpainting [12, 11], scientific data interpolation [17], and more recently deep learning [14, 13, 15].

1.1. Main Result. We consider the error of the WNLL in a model problem. 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 subset of $\mathcal{M}$ which has been labeled, and S is a uniform sample of $\mathcal{D}$. In S, we have $u(\boldsymbol{x})=b(\boldsymbol{x})$. An illustration of the computational domain and the point cloud is shown in Fig. 1. In WNLL, we solve the following linear system, (1.4), to extend the label function u to the entire domain P.

$$\begin{aligned}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})\left(u_\delta(\boldsymbol{x})-u_\delta(\boldsymbol{y})\right)+\mu\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})(u_\delta(\boldsymbol{x})-b(\boldsymbol{y}))=0,\quad \boldsymbol{x}\in P,\\ u_\delta(\boldsymbol{x})=b(\boldsymbol{x}),\quad \boldsymbol{x}\in S.\end{aligned} \tag{1.4}$$

where $R_\delta(\boldsymbol{x},\boldsymbol{y})$, $K_\delta(\boldsymbol{x},\boldsymbol{y})$ are kernel functions given as

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

where $C_\delta=\frac{1}{(4\pi\delta^2)^{k/2}}$ is the normalization factor. $R, K\in C^2(\mathbb{R}^+)$ are two kernel functions satisfying the conditions listed in Assumption 1.1.

In this paper, we consider the convergence of WNLL as δ goes to 0 and $n=|P|$ goes to infinity. In the analysis, we assume that δ is small enough and n is large enough. More specifically, we require that

$$\delta\le T_0\quad\text{and}\quad \frac{1}{\delta^{k+3}\sqrt{n}}\left(\ln n-2\ln\delta+1\right)^{1/2}\le C_0, \tag{1.6}$$

where $T_0 > 0$ is a constant only depends on $\mathcal{M}$ and $C_0 > 0$ is another constant depends on $\mathcal{M}, \mathcal{D}, R, K$.

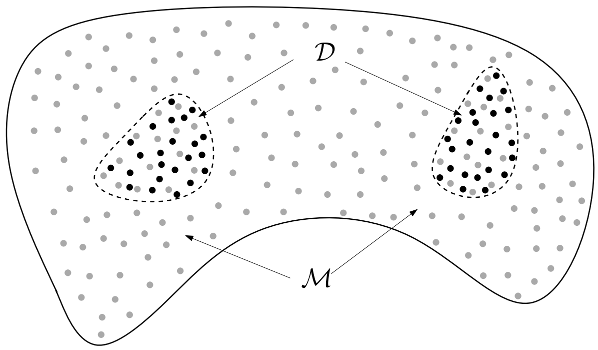

FIG. 1. *Illustration of the computational domain. Gray points: sample of $\mathcal{M}$; Black points: sample of $\mathcal{D} \subset \mathcal{M}$.*

As the continuous counterpart, we consider the Laplace-Beltrami equation on a closed smooth manifold $\mathcal{M}$

$$
\left\{ \begin{array}{rcll} \Delta_{\mathcal{M}} u(\boldsymbol{x}) & = & 0, & \boldsymbol{x} \in \mathcal{M} \backslash \mathcal{D}, \\ u(\boldsymbol{x}) & = & b(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{D}, \end{array} \right. \tag{1.7}
$$

where $\Delta_{\mathcal{M}} = \text{div}(\nabla)$ is the Laplace-Beltrami operator on $\mathcal{M}$. Let $\Phi : \Omega \subset \mathbb{R}^k \to \mathcal{M} \subset \mathbb{R}^d$ be a local parametrization of $\mathcal{M}$ and $\theta \in \Omega$. For any differentiable function $f : \mathcal{M} \to \mathbb{R}$, we define the gradient on the manifold

$$
\nabla f(\Phi(\theta)) = \sum_{i,j=1}^{m} g^{ij}(\theta) \frac{\partial \Phi}{\partial \theta_i}(\theta) \frac{\partial f(\Phi(\theta))}{\partial \theta_j}(\theta). \tag{1.8}
$$

And for the vector field $F : \mathcal{M} \to T_{\boldsymbol{x}}\mathcal{M}$ on $\mathcal{M}$, where $T_{\boldsymbol{x}}\mathcal{M}$ is the tangent space of $\mathcal{M}$ at $\boldsymbol{x} \in \mathcal{M}$, the divergence is defined as

$$
\text{div}(F) = \frac{1}{\sqrt{\det G}} \sum_{k=1}^{d} \sum_{i,j=1}^{m} \frac{\partial}{\partial \theta_i} \left(\sqrt{\det G}\, g^{ij} F^k(\Phi(\theta)) \frac{\partial \Phi^k}{\partial \theta_j} \right) \tag{1.9}
$$

where $(g^{ij})_{i,j=1,\cdots,k} = G^{-1}$, $\det G$ is the determinant of matrix G and $G(\theta) = (g_{ij})_{i,j=1,\cdots,k}$ is the first fundamental form with

$$
g_{ij}(\theta) = \sum_{k=1}^{d} \frac{\partial \Phi_k}{\partial \theta_i}(\theta) \frac{\partial \Phi_k}{\partial \theta_j}(\theta), \quad i, j = 1, \cdots, m. \tag{1.10}
$$

and $(F^1(\boldsymbol{x}), \cdots, F^d(\boldsymbol{x}))^T$ is the representation of F in the embedding coordinates.

To prove the convergence, we need the following assumptions.

Assumption 1.1.

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

- Assumptions on the kernel functions:
  (a) Smoothness: $K(r), R(r) \in C^2(\mathbb{R}^+)$;
  (b) Nonnegativity: $R(r), K(r) \geq 0$ for any $r \geq 0$.
  (c) Compact support: $R(r) = 0$ for $\forall r > 1$; $K(r) = 0$ for $\forall r > r_0 \geq 2$.
  (d) Nondegeneracy: $\exists \delta_0 > 0$ such that $R(r) \geq \delta_0$ for $0 \leq r \leq 1/2$ and $K(r) \geq \delta_0$ for $0 \leq r \leq 2$.
- Assumptions on the point cloud: P and S are uniformly distributed on $\mathcal{M}$ and $\mathcal{D}$, respectively.

In this paper, we use the notation C to denote any constant which may be different in different places. The main contribution of this paper is to analyze relation between the solutions of the Laplace-Beltrami equation (1.7) and the WNLL (1.4). More precisely, we prove the following theorem:

Theorem 1.1. *Let u_δ solves* (1.4) *and u solves* (1.7)*. Under* (1.6) *and Assumption 1.1, with probability at least $1 - 1/(2n)$, where $n = |P|$, we have*

$$|u_\delta - u| \leq C\delta,$$

as long as

$$\mu \sum_{\boldsymbol{y} \in S} K_\delta(\boldsymbol{x}, \boldsymbol{y}) \geq C \sum_{\boldsymbol{y} \in P} R_\delta(\boldsymbol{x}, \boldsymbol{y}), \quad \boldsymbol{x} \in P \cap \mathcal{D}_\delta. \tag{1.11}$$

$\mathcal{D}_\delta = \{\boldsymbol{x} \in \mathcal{M} : dist(\boldsymbol{x}, \mathcal{D}) \leq 2\delta\}$, $C = C(\mathcal{M}, \mathcal{D}, R, K) > 0$ is a constant independent of δ, P and S.

In above theorem, (1.11) actually gives a condition for weight μ. Notice that

$$\frac{1}{n} \sum_{\boldsymbol{y} \in P} R_\delta(\boldsymbol{x}, \boldsymbol{y}) \approx \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} R_\delta(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y} = O(1), \quad \boldsymbol{x} \in P \cap \mathcal{D}_\delta.$$

Here $O(1)$ denotes a number between $1/C$ and C with constant $C > 0$ independents on δ, P and S.

Moerover, if S is dense enough, we have that

$$\frac{1}{|S|} \sum_{\boldsymbol{y} \in S} K_\delta(\boldsymbol{x}, \boldsymbol{y}) \approx \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} K_\delta(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y}, \quad \boldsymbol{x} \in P \cap \mathcal{D}_\delta.$$

Here, we need the assumption on K such that $K(r) \geq \delta_0 > 0, \forall 0 \leq r \leq 2$. This implies that

$$\int_{\mathcal{D}} K_\delta(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y} = O(1), \quad \boldsymbol{x} \in P \cap \mathcal{D}_\delta.$$

Hence, from (1.11), we have

$$\mu \sim \frac{|P|}{|S|}.$$

This explains the scaling of μ in WNLL.

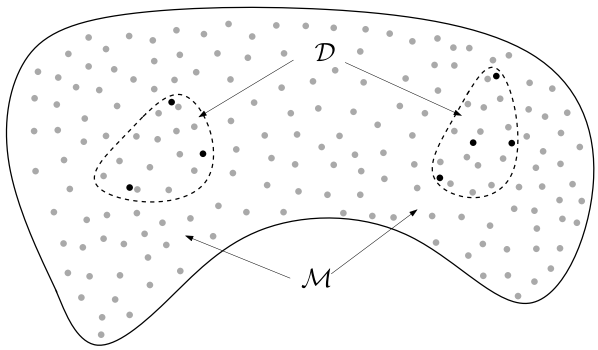

FIG. 2. *Illustration of the computational domain with extremely low labeling rate.*

On the other hand, if sample S is extremely sparse such that $\bigcup_{\boldsymbol{x}\in S} B(\boldsymbol{x};4\delta)$ does not cover $\mathcal{D}_\delta$ as shown in Fig. 2, $\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})$ may be zero for some $\boldsymbol{x}\in P\cap\mathcal{D}_\delta$. In this case, condition (1.11) does not hold. Then we can not guarantee the convergence even in the WNLL. With extremely low labeling rate, actually, the whole framework of harmonic extension fails [10, 16]. We should use other approach to get a smooth interpolation.

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.*

$$|L_{\delta,n}u(\boldsymbol{x})|\le L_{\delta,n}v(\boldsymbol{x}) \quad\rightarrow\quad |u|\le v,$$

where

$$L_{\delta,n}u(\boldsymbol{x})=\sum_{\boldsymbol{y}\in P}R_\delta(\boldsymbol{x},\boldsymbol{y})\left(u(\boldsymbol{x})-u(\boldsymbol{y})\right)+\mu\sum_{\boldsymbol{y}\in S}K_\delta(\boldsymbol{x},\boldsymbol{y})u(\boldsymbol{x}),\quad \boldsymbol{x}\in P. \tag{1.12}$$

THEOREM 1.3. *Let* u_δ *and* u *solve* (1.4) *and* (1.7) *respectively.* v *is the solution of* (1.13),

$$\left\{\begin{array}{rcll} -\Delta_{\mathcal{M}}v(\boldsymbol{x}) &=& 1, & \boldsymbol{x}\in\mathcal{M}\backslash\mathcal{D},\\ v(\boldsymbol{x}) &=& 1, & \boldsymbol{x}\in\mathcal{D}.\end{array}\right. \tag{1.13}$$

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

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

as long as

$$\mu\sum_{\boldsymbol{y}\in S}K_\delta(\boldsymbol{x},\boldsymbol{y})\ge C\sum_{\boldsymbol{y}\in P}R_\delta(\boldsymbol{x},\boldsymbol{y}),\quad \boldsymbol{x}\in P\cap\mathcal{D}_\delta.$$

$\mathcal{D}_\delta=\{\boldsymbol{x}\in\mathcal{M}: dist(\boldsymbol{x},\mathcal{D})\le 2\delta\}$, $C=C(\mathcal{M},\mathcal{D},R,K)>0$ *is a constant independent on* δ, P *and* S.

The above two theorems will be proved in Section 2 and Section 3, respectively. In Section 4, we prove a technical theorem used in the analysis. Some discussions are made in Section 5.

2. Maximum Principle (Theorem 1.2). First, we introduce some notations. For any two points $\boldsymbol{x}, \boldsymbol{y} \in P$, we say that they are neighbors if and only if $R_\delta(\boldsymbol{x}, \boldsymbol{y}) > 0$, denoted as $\boldsymbol{x} \sim \boldsymbol{y}$. For $\boldsymbol{x} \in S, \boldsymbol{y} \in P \cup S$, they are neighbors if and only if $K_\delta(\boldsymbol{x}, \boldsymbol{y}) > 0$, denoted also by $\boldsymbol{x} \sim \boldsymbol{y}$ or $\boldsymbol{y} \sim \boldsymbol{x}$. $\boldsymbol{x}$ and $\boldsymbol{y}$ are connected if there exist $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m \in P \cup S$ such that

$$\boldsymbol{x} \sim \boldsymbol{z}_1 \sim \cdots \sim \boldsymbol{z}_m \sim \boldsymbol{y}.$$

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

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

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

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

and consequently

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

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

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

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

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

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

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

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

In the rest of this section, we will prove that with high probability, P is S-connected. To prove this, we need a theorem from the empirical process theory [9].

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

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

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

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

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

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

This theorem will be proved in Section 4.

Suppose P is not S-connected. Let

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

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

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

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

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

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

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

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

It follows that

$$I_n(R_{\delta/4}(\boldsymbol{x}_0, \cdot)) = 0.$$

On the other hand, $I(R_{\delta/4}(\boldsymbol{x}_0, \cdot)) = O(1)$. Using Theorem 2.1 and assumption (1.6), we know that the probability is less than $1/(2n)$, which proves that P is S-connected with probability at least $1 - 1/(2n)$. So far, we have proved Theorem 1.2.

3. Error Estimate (Theorem 1.3). Let $e_\delta(\boldsymbol{x}) = u_\delta(\boldsymbol{x}) - u(\boldsymbol{x})$. u_δ and u solve (1.4) and (1.7) respectively.

Direct calculation shows that

(3.1)
$$L_{\delta,n} e_\delta(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x}) - u(\boldsymbol{y})) + \mu \sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x}) - b(\boldsymbol{y}))\mathrm{d}\boldsymbol{y}, \quad \boldsymbol{x} \in P,$$

(3.2)
$$e_\delta(\boldsymbol{x}) = 0, \quad \boldsymbol{x} \in S.$$

Next, we will find an upper bound of the right hand side of (3.1).

An upper bound of the second term of (3.1) is relatively easy to find by using the smoothness of u and b:

$$\left|\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x}) - b(\boldsymbol{y}))\right| \leq C\delta \sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y}) \tag{3.3}$$

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

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

$$I_{bd} = \sum_{j=1}^{d} \int_{\partial\mathcal{M}} n^j(\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y}) \cdot \nabla(\nabla^j u(\boldsymbol{y}))\bar{R}_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\tau_{\boldsymbol{y}}, \tag{3.4}$$

and

$$I_{in} = \frac{1}{\delta^2}\int_{\mathcal{M}} R_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x})-u(\boldsymbol{y}))\mathrm{d}\boldsymbol{y} + \int_{\mathcal{M}} \bar{R}_\delta(\boldsymbol{x},\boldsymbol{y})\Delta_{\mathcal{M}} u(\boldsymbol{y})\mathrm{d}\boldsymbol{y} - \int_{\partial\mathcal{M}} \bar{R}_\delta(\boldsymbol{x},\boldsymbol{y})\frac{\partial u}{\partial \mathbf{n}}(\boldsymbol{y})\mathrm{d}\tau_{\boldsymbol{y}} - I_{bd}.$$

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

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

$$|I_{in}| \le C\delta \|u\|_{C^3(\mathcal{M})}, \tag{3.5}$$

as long as $\delta \le T_0$.

According to above theorem, for u solves (1.7), we have

$$\frac{1}{\delta^2}\left|\int_{\mathcal{M}} R_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x})-u(\boldsymbol{y}))\mathrm{d}\boldsymbol{y}\right| \le C\delta, \quad \boldsymbol{x} \in \mathcal{M}\backslash\mathcal{D}_\delta, \tag{3.6}$$

where $\mathcal{D}_\delta = \{\boldsymbol{x} \in \mathcal{M} : \mathrm{dist}(\boldsymbol{x},\mathcal{D}) \le 2\delta\}$. Notice that for $\boldsymbol{x} \in \mathcal{M}\backslash\mathcal{D}$, $R_\delta(\boldsymbol{x},\boldsymbol{y})$ has no intersection with $\mathcal{D}$, so all boundary terms vanish.

To get an upper bound of the first term in (3.1), we need to estimate the difference between $\int_{\mathcal{M}} R_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x})-u(\boldsymbol{y}))\mathrm{d}\boldsymbol{y}$ and $\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x})-u(\boldsymbol{y}))$. This is given by the following theorem.

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

$$\sup_{f\in\bar{\mathcal{R}}_\delta} |I(f) - I_n(f)| \le \frac{C}{\delta^k\sqrt{n}}\left(\ln n - 2\ln\delta + 1\right)^{1/2}, \tag{3.7}$$

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

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

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

$\bar{\mathcal{R}}_\delta = \{R_\delta(\boldsymbol{x},\cdot), R_\delta(\boldsymbol{x},\cdot)u(\cdot), R_\delta(\boldsymbol{x},\cdot)v(\cdot) : \boldsymbol{x}\in\mathcal{M},$ *u and v solves (1.7) and (1.13) respectively.*$\}$

This theorem will be proved in Section 4 using the empirical process theory [9].

Using Theorem 3.2, we have

$$\left|\frac{1}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x})-u(\boldsymbol{y}))\right| \le \frac{C}{\delta^k\sqrt{n}}(\ln n - 2\ln\delta + 1)^{1/2} + C\delta^3, \quad \boldsymbol{x} \in P\cap(\mathcal{M}\backslash\mathcal{D}_\delta). \tag{3.8}$$

For $\boldsymbol{x} \in P \cap \mathcal{D}_\delta$, the bound is straightforward, just using the smoothness of u,

$$\left|\frac{1}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})(u(\boldsymbol{x})-u(\boldsymbol{y}))\right| \le \frac{C\delta}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y}), \quad \boldsymbol{x} \in P\cap\mathcal{D}_\delta. \tag{3.9}$$

Substituting (3.3), (3.8) and (3.9) in (3.1), we have

(3.10)
$$|L_{\delta.n}e_\delta(\boldsymbol{x})| \le C\delta \left(\frac{1}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})\right) + C\delta\left(\frac{\mu}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})\right), \quad \boldsymbol{x}\in P\cap\mathcal{D}_\delta.$$

and

$$|L_{\delta.n}e_\delta(\boldsymbol{x})| \le \frac{C}{\delta^k\sqrt{n}}(\ln n - 2\ln\delta + 1)^{1/2} + C\delta^3 + C\delta\left(\frac{\mu}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})\right), \quad \boldsymbol{x}\in P\cap(\mathcal{M}\backslash\mathcal{D}_\delta).$$

Furthermore, using (1.6), we have

$$|L_{\delta.n}e_\delta(\boldsymbol{x})| \le C\delta^3 + C\delta\left(\frac{\mu}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})\right), \quad \boldsymbol{x}\in P\cap(\mathcal{M}\backslash\mathcal{D}_\delta). \tag{3.11}$$

(3.10) and (3.11) give an upper bound for $|L_{\delta.n}e_\delta(\boldsymbol{x})|$.

Next, we want to get a lower bound of $L_{\delta.n}v(\boldsymbol{x})$ with v given in (1.13).

By Theorem 3.1 and (1.13), we have

$$\frac{1}{\delta^2}\int_{\mathcal{M}} R_\delta(\boldsymbol{x},\boldsymbol{y})(v(\boldsymbol{x})-v(\boldsymbol{y}))\mathrm{d}\boldsymbol{y} \ge \int_{\mathcal{M}} \bar{R}_\delta(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y} - C\delta, \quad \boldsymbol{x}\in\mathcal{M}\backslash\mathcal{D}_\delta. \tag{3.12}$$

Also using Theorem 3.2

$$\begin{aligned}&\frac{1}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})(v(\boldsymbol{x})-v(\boldsymbol{y}))\\ \ge\ &\bar{w}_\delta\delta^2 - C\delta^3 - \frac{C}{\delta^k\sqrt{n}}(\ln n - 2\ln\delta+1)^{1/2} \ge \bar{w}_\delta\delta^2/2, \quad \boldsymbol{x}\in P\cap(\mathcal{M}\backslash\mathcal{D}_\delta).\end{aligned} \tag{3.13}$$

with $\bar{w}_\delta = \min\limits_{\boldsymbol{x}\in\mathcal{M}\backslash\mathcal{D}_\delta}\frac{1}{|\mathcal{M}|}\int_{\mathcal{M}}\bar{R}_\delta(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y}$. Here, we also use the assumption that n is large enough, (1.6).

In $P\cap\mathcal{D}_\delta$, we have

$$\frac{1}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y})(v(\boldsymbol{x})-v(\boldsymbol{y})) \ge -\frac{C\delta}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y}), \quad \boldsymbol{x}\in P\cap\mathcal{D}_\delta, \tag{3.14}$$

this is due to the smoothness of v.

Also notice that

$$\frac{1}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y})v(\boldsymbol{x}) = \frac{v(\boldsymbol{x})}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y}) \ge \frac{1}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y}). \tag{3.15}$$

Combining (3.13), (3.14) and (3.15), we obtain

$$L_{\delta.n}v(\boldsymbol{x}) \ge \frac{\mu}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y}) - \frac{C\delta}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y}), \quad \boldsymbol{x}\in P\cap\mathcal{D}_\delta. \tag{3.16}$$

and

$$L_{\delta,n}v(\boldsymbol{x}) \geq \frac{\bar{w}_\delta}{2}\delta^2 + \frac{\mu}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y}), \quad \boldsymbol{x} \in P \cap (\mathcal{M}\backslash\mathcal{D}_\delta). \tag{3.17}$$

Comparing (3.11) and (3.17), we have

$$|L_{\delta,n}e_\delta(\boldsymbol{x})| \leq C\delta L_{\delta.n}v(\boldsymbol{x}), \quad \boldsymbol{x} \in P \cap (\mathcal{M}\backslash\mathcal{D}_\delta). \tag{3.18}$$

Meanwhile, (3.10) and (3.16) show that

$$|L_{\delta,n}e_\delta(\boldsymbol{x})| \leq C\delta L_{\delta.n}v(\boldsymbol{x}), \quad \boldsymbol{x} \in P \cap \mathcal{D}_\delta, \tag{3.19}$$

as long as

$$\frac{\mu}{n}\sum_{\boldsymbol{y}\in S} K_\delta(\boldsymbol{x},\boldsymbol{y}) \geq \frac{C}{n}\sum_{\boldsymbol{y}\in P} R_\delta(\boldsymbol{x},\boldsymbol{y}), \quad \boldsymbol{x} \in P \cap \mathcal{D}_\delta. \tag{3.20}$$

The proof of Theorem 1.3 is completed.

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 $\{y_1, \cdots, y_m\} \subset Y$ with the property that every $f \in F$ has some y_i such that $d(f, y_i) < \epsilon$. The set $\{y_1, \cdots, y_m\}$ is called an ϵ-cover of F . The logarithm of the covering numbers is called the entropy of the set. For every sample $\{x_1, \cdots, x_n\}$, let μ_n be the empirical measure supported on that sample. For $1 \leq p < \infty$ and a function f , put $\|f\|_{L_p(\mu_n)} = \left(\frac{1}{n}\sum_{i=1}^n |f(x_i)|^p\right)^{1/p}$ 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 ϵ 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 μ to be a probability measure on $\mathcal{M}$. Let $(\boldsymbol{x}_i)_{i=1}^\infty$ be independent random variables distributed according to μ. For any $\epsilon > 0$ and every $n \geq 8/\epsilon^2$,*

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

Note that

$$L_1(\mu_n) \leq L_\infty(\mu_n) \leq L_\infty$$

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

Corollary 4.2. *Let F be a class of functions from $\mathcal{M}$ to $[-1,1]$ and set μ to be a probability measure on $\mathcal{M}$. Let $(\boldsymbol{x}_i)_{i=1}^\infty$ be independent random variables distributed according to μ. For any $\epsilon > 0$ and every $n \geq 8/\epsilon^2$,*

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

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

COROLLARY 4.3. *Let F be a class of functions from $\mathcal{M}$ to $[-1,1]$. Let $(\boldsymbol{x}_i)_{i=1}^\infty$ be independent random variables distributed according to p, where p is the probability distribution. Then with probability at least $1-\delta$, we have*

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

where

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

Proof. Using Corollary 4.2, with probability at least $1-\delta$,

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

where ϵ_δ is determined by

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

Obviously,

$$\epsilon_\delta \ge \sqrt{\frac{128}{n}} = 8\sqrt{\frac{2}{n}}$$

which gives that

$$N(\epsilon_\delta/8, F, L_\infty) \le N(\sqrt{\frac{2}{n}}, F, L_\infty)$$

Then, we have

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

which proves the corollary. $\Box$

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 $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{M}$

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

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

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

Using the Corollary 4.3, 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 (1.7)) and v (solution of (1.13)) 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 weight is critical in the convergence 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 the WNLL.

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_∞ norm of the gradient instead of the L_2 norm, i.e., to solve the following optimization problem

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

with the constraint

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

This approach is closely related to the infinity Laplacian [5, 4]. 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 in our future work.

REFERENCES

[1] A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.*, 4:490–530, 2005.

[2] A. Buades, B. Coll, and J.-M. Morel. Neighborhood filters and pde's. *Numer. Math.*, 105:1–34, 2006.

[3] F. R. K. Chung. *Spectral Graph Theory.* American Mathematical Society, 1997.

[4] L. Z. L. O. Elmoataz Abderrahim, Desquesnes Xavier. Nonlocal infinity laplacian equation on graphs with applications in image processing and machine learning. *Mathematics and Computers in Simulation*, 102:153–163, 2014.

[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.*, 6:595–630, 2007.

[7] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Model. Simul.*, 7:1005–1028, 2008.

[8] Z. Li and Z. Shi. A convergent point integral method for isotropic elliptic equations on point cloud. *SIAM: Multiscale Modeling Simulation*, 14:874–905, 2016.

[9] S. Mendelson. A few notes on statistical learning theory. In *Lecture Notes in Computer Science*, volume 2600, pages 1–40, 2003.

[10] B. Nadler, N. Srebro, and X. Zhou. Semi-supervised learning with the graph laplacian: The limit of infinite unlabelled data. *NIPS*, 2009.

[11] Z. Shi, S. Osher, and W. Zhu. Weighted nonlocal laplacian on interpolation from sparse data. *Journal of Scientific Computing*, 73(2):1164–1177, Dec 2017.

[12] Z. Shi, J. Sun, and M. Tian. Harmonic extension on point cloud. *SIAM Multiscale Modeling & Simulation*, 16:215–247, 2018.
[13] B. Wang, A. T. Lin, Z. Shi, W. Zhu, P. Yin, A. L. Bertozzi, and S. J. Osher. Adversarial Defense via Data Dependent Activation Function and Total Variation Minimization. *arXiv preprint arXiv:1809.08516*, 2018.
[14] B. Wang, X. Luo, Z. Li, W. Zhu, Z. Shi, and S. J. Osher. Deep neural nets with interpolating function as output activation. In *The Thirty-second Annual Conference on Neural Information Processing Systems (NIPS)*, 2018.
[15] B. Wang and S. J. Osher. Graph Interpolating Activation Improves Both Natural and Robust Accuracies in Data-Efficient Deep Learning. *arXiv preprint arXiv:1907.06800*, 2019.
[16] X. Zhou and M. Belkin. Semi-supervised learning by higher order regularization. *NIPS*, 2011.
[17] W. Zhu, B. Wang, R. Barnard, C. D. Hauck, F. Jenko, and S. Osher. Scientific data interpolation with low dimensional manifold model. *Journal of Computational Physics*, 352(1):213–245, 2018.
[18] X. 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*, pages 912–919, 2003.