Bernoulli **27**(1), 2021, 155–191
https://doi.org/10.3150/20-BEJ1232

Nonparametric estimation of surface integrals on level sets

WANLI QIAO

Department of Statistics, George Mason University, 4400 University Drive, MS 4A7, Fairfax, VA 22030, USA. E-mail: wqiao@gmu.edu

Surface integrals on density level sets often appear in asymptotic results of nonparametric level set estimation, such as for confidence regions and bandwidth selection. Also surface integrals can be used to describe the shape of level sets. Assuming the integrands are known, we consider three estimators of the surface integrals on density level sets, one as a direct plug-in estimator, and the other two based on different neighborhoods of level sets. For all the three estimators, we derive the rates of convergence and asymptotic distributions.

Keywords: curvatures; kernel density estimation; level sets; positive reach; surface integrals

1. Introduction

The c-level set of a density function f on $\mathbb{R}^d$ is defined as

$$\mathcal{M}_c = f^{-1}(c) = \{x \in \mathbb{R}^d : f(x) = c\}.$$

The set $\mathcal{L}_c = f^{-1}[c, \infty) = \{x \in \mathbb{R}^d : f(x) \geq c\}$ is called the super level set. Level set estimation finds its application in many areas such as clustering [10], classification [38], and anomaly detection [58]. It has received extensive study in the literature. See, for example, [9,17,31,43,48,49,55,57,60,61].

We consider $d \geq 2$ in this paper. For simplicity of notation, the subscript c is often omitted for $\mathcal{M}_c$ and $\mathcal{L}_c$. When f has no flat part at the level c, $\mathcal{M}$ is a $(d-1)$-dimensional submanifold of $\mathbb{R}^d$. Given an i.i.d. sample $X_1, \ldots, X_n$ from the density f, we investigate the estimation of the surface integral

$$\lambda_c(f, g) = \lambda(f, g) = \int_{\mathcal{M}} g(x)\, d\mathscr{H}(x),$$

for some known integrable function g (for example, when $g \equiv 1$), where $\mathscr{H}$ is the $(d-1)$-dimensional normalized Hausdorff measure. Our estimators are based on the kernel density estimator of f given by

$$\widehat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where $h > 0$ is the bandwidth and K is a d-dimensional kernel function. The plug-in estimators of $\mathcal{M}$ and $\mathcal{L}$ are given by $\widehat{\mathcal{M}} = \{x \in \mathbb{R}^d : \widehat{f}(x) = c\}$ and $\widehat{\mathcal{L}} = \{x \in \mathbb{R}^d : \widehat{f}(x) \geq c\}$, respectively. We consider the following three estimators for $\lambda(f, g)$:

$$\lambda(\widehat{f}, g) = \int_{\widehat{f}^{-1}(c)} g(x)\, d\mathscr{H}(x), \tag{1.2}$$

$$\lambda_{\varepsilon_n}^{(1)}(\widehat{f}, g) := \frac{1}{2\varepsilon_n} \int_{\widehat{f}^{-1}[c-\varepsilon_n, c+\varepsilon_n]} g(x) \|\nabla \widehat{f}(x)\|\, dx, \quad \text{and} \tag{1.3}$$

1350-7265

$$\lambda^{(2)}_{\varepsilon_n}(\widehat{f},g) := \frac{1}{2\varepsilon_n}\int_{\widehat{f}^{-1}(c)\oplus\varepsilon_n} g(x)\,dx, \tag{1.4}$$

for some $\varepsilon_n > 0$, where $\nabla\widehat{f}$ is the gradient of $\widehat{f}$, and $\widehat{f}^{-1}(c)\oplus\varepsilon_n$ is the union of all balls with radius ε_n and centers at $\widehat{f}^{-1}(c)$. Here $\lambda(\widehat{f},g)$ is a direct plug-in estimator, and $\lambda^{(j)}_{\varepsilon_n}(\widehat{f},g)$, $j=1,2$, can be viewed as local averages of surface integrals close to $\lambda(\widehat{f},g)$ when ε_n is small. The integration in $\lambda^{(2)}_{\varepsilon_n}(\widehat{f},g)$ is over a band or a tube of constant width around $\widehat{f}^{-1}(c)$, while the domain of integration in $\lambda^{(1)}_{\varepsilon_n}(\widehat{f},g)$ has varying width. See Section 2.2 for more comparison among the three estimators. The main results of this paper include the rates of convergence and asymptotic normality of these estimators.

The surface integral $\lambda(f,g)$ is an important quantity that is involved in asymptotic theory for level set estimation. For example, it appears in [9] as the convergence limit of the set-theoretic measure of $\mathcal{L}\Delta\widehat{\mathcal{L}} := (\mathcal{L}\setminus\widehat{\mathcal{L}})\cup(\widehat{\mathcal{L}}\setminus\mathcal{L})$, which is the symmetric difference between $\mathcal{L}$ and $\widehat{\mathcal{L}}$. Using a similar measure as a risk criterion, [45] shows that the optimal bandwidth for nonparametric level set estimation is determined by a ratio of two surface integrals in the form of $\lambda(f,g)$. [47] develops large sample confidence regions for $\mathcal{M}$ and $\mathcal{L}$, for which the surface area of $\mathcal{M}$ (a special form of $\lambda(f,g)$ when $g\equiv 1$) is the only unknown quantity that needs to be estimated. The quantity $\lambda(f,g)$ is also a key component in the concept of vertical density representation [59]. Surface integrals on (regression) level sets appear in optimal tuning parameter selection for nearest neighbour classifiers [11,29,51].

Some important concepts in differential geometry are in the form of surface integrals. For example, the Willmore energy of a $(d-1)$-dimensional smooth submanifold $\mathcal{S}$ embedded in $\mathbb{R}^d$ is defined as $W(\mathcal{S}) = \int_{\mathcal{S}} |H(x)|^2\,d\mathscr{H}(x)$, where $H(x)$ is the mean curvature of $\mathcal{S}$ at x [63]. $W(\mathcal{S})$ measures the total elastic bending of $\mathcal{S}$ from a sphere if $d=3$. The Willmore energy is widely used in studying the shape of biological cell membranes [53]. When $\mathcal{S}$ is a level set, the Willmore energy finds its applications in image inpainting [13] and segmentation of spinal vertebrae [36]. Other examples of surface integrals on manifolds include Minkowski functionals, where the integrands are some functions of the principal curvatures. When the manifolds are level sets, Minkowski functionals are widely used as morphological descriptors (shape statistics) in studying cosmic microwave background. See Chapter 10.3 of [39], and [33,44,52] and the references therein. Minkowski functionals (and their ratios) can be used to characterize different shapes, for example, planarity, filamentarity, clusters etc.

Surface area has been used in the definition of “contour index”, which is the ratio between perimeter and the square root of area, and has appeared in medical imaging and remote sensing [12,50]. The estimation of surface area also has extensive applications in stereology [5,6,26]. Other examples of $\lambda(f,g)$ arise where g is observable temperature, humidity, or the density of some non-homogeneous material, and one is interested in the surface integrals of these quantities on an unknown manifold [32].

In the literature [18] obtains a consistency result for the estimation of the surface area of $\mathcal{M}$, that is, when $g\equiv 1$. There exists some recent work on the estimation of the surface area of the boundary of an unknown body $S\subset G$ where G is a bounded set given a sample on G. The work is relevant to the study in this paper but in a setting different from what we consider here. There the surface area is defined as the Minkowski content, which coincides with the normalized Hausdorff measure in regular cases [1]. [4,24] obtain asymptotic normality results for the surface area (or perimeter) of ∂S in a framework that assumes a uniform distribution on G and the binary labels for S and $G\setminus S$ are observed with the sample. The sampling scheme is called the “inside-outside” model in [21], and the binary labels for S and $G\setminus S$ contain important information of the location of ∂S. In contrast, we study the estimation of surface integral $\lambda(f,g)$, which is more general than surface area. More importantly, the location of $f^{-1}(c)$ is completely unknown and needs to be estimated. As a result even for $g\equiv 1$ the approach we take is very different from the above work. In the setting of the “inside-outside” model, the rates of convergence of the estimators for the surface area of ∂S are derived under different shape assumptions (see [19,41,42]). The consistency of the surface integral estimation on ∂S has been considered by [32], where the integrand is assumed to be observable on the sample points. Assuming the i.i.d. sampling scheme only on S, [3] estimates the perimeter of ∂S using the alpha-shape for $d = 2$ and derive the rate of convergence. As shown in Section 2.1, $\lambda(\widehat{f}, g) - \lambda(f, g)$ and $d_g(f, \widehat{f}) := \int_{\mathcal{L}\Delta\widehat{\mathcal{L}}} g(x)\,dx$ are related, the latter having been studied in, see, for example, [7–9,20,40].

We organize the paper as follows. In the rest of this section we introduce the notation, geometric concepts and assumptions used for our results. In Section 2, the rates of convergence and asymptotic distributions of the three estimators are given in Theorems 2.1 and 2.2. Sections 2.1 and 2.2 are dedicated to the estimation of the surface integral $\lambda(f, g)$ using the direct plug-in estimator $\lambda(\widehat{f}, g)$, and neighborhood-based estimators $\lambda_{\varepsilon_n}^{(j)}(\widehat{f}, g)$, $j = 1, 2$, respectively. The proofs of all the results are left to Section 4.

1.1. Notation and geometric concepts

For any $x \in \mathbb{R}^d$ and $A \subset \mathbb{R}^d$, let $d(x, A) = \inf_{y\in A} \|x - y\|$. The Hausdorff distance between any two sets $A, B \subset \mathbb{R}^d$ is

$$d_H(A, B) = \max\Big\{\sup_{x\in B} d(x, A), \sup_{x\in A} d(x, B)\Big\}.$$

The ball with center x and radius ε is denoted by $\mathcal{B}(x, \varepsilon) = \{y \in \mathbb{R}^d : \|x - y\| \le \varepsilon\}$. For any set $A \subset \mathbb{R}^d$ and $\varepsilon > 0$, we denote $A \oplus \varepsilon = \bigcup_{x\in A} \mathcal{B}(x, \varepsilon)$. Let the normal projection of x onto A be $\pi_A(x) = \{y \in A : \|x - y\| = d(x, A)\}$, which may not be a single point.

We will also use the concept of *reach* of a manifold. For a set $\mathcal{S} \subset \mathbb{R}^d$, let $\mathrm{Up}(\mathcal{S})$ be the set of points $x \in \mathbb{R}^d$ such that $\pi_{\mathcal{S}}(x)$ is a single point. The reach of $\mathcal{S}$ is defined as

$$\rho(\mathcal{S}) = \sup\big\{\delta > 0 : \mathcal{S} \oplus \delta \subset \mathrm{Up}(\mathcal{S})\big\}.$$

See [23]. A positive reach is related to the concepts of "r-convexity" and "rolling condition" (see [18]). These are common regularity conditions for the estimation of surface area. For two manifolds A and B, if the normal projections $\pi_A : B \to A$ and $\pi_B : A \to B$ are homeomorphisms, then A and B are called *normal compatible*. See [16] and Figure 1. For a matrix M and compatible vectors u and v, denote $\langle u, v\rangle_M = u^T M v$ and $\|u\|_M^2 = \langle u, u\rangle_M$. Let $\|M\|_F$ be the Frobenius norm of a matrix M. For a function $g : \mathbb{R}^d \to \mathbb{R}$, let $\mathrm{supp}(g)$ be the support of g. If g is twice differentiable, let ∇g and $\nabla^2 g$ be the gradient and Hessian matrix of g, respectively. Denote $\|g\|_\infty = \sup_{x\in\mathbb{R}^d} |g(x)|$ and $\|g\|_p = [\int_{\mathbb{R}^d} |g(x)|^p\,dx]^{1/p}$ for $p \ge 1$. For a vector or matrix M and a positive integer r, let $M^{\otimes r}$ be the rth Kronecker power of M.

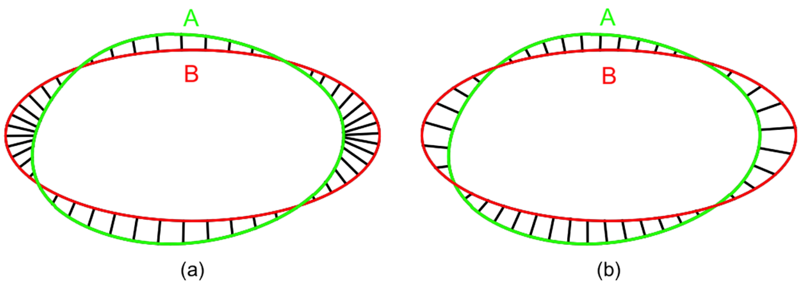

Figure 1. This figure shows the normal compatibility between two curves A (green) and B (red). (a) shows the normal projection from A to B. (b) shows the normal projection from B to A.

The vector of the rth derivatives of g (if they exist) is defined as $\nabla^{\otimes r} g(x) = \frac{\partial^r g}{(\partial x)^{\otimes r}}(x) \in \mathbb{R}^{d^r}$, where we apply the rth Kronecker power to the operator ∇. For a bandwidth $h > 0$, let $\gamma_{n,h}^{(k)} = \sqrt{\frac{\log n}{nh^{d+2k}}}$, which, under standard assumptions, is the uniform rate of convergence of the kernel estimator of the kth density derivatives after being centered at their expectation. We denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$ for $a, b \in \mathbb{R}$.

1.2. Assumptions and their discussion

A function K is called a νth ($\nu \geq 2$) order kernel if $\int_{\mathbb{R}^d} u^{\otimes \nu} K(u)\, du \neq 0$ and is finite while

$$\int_{\mathbb{R}^d} u^{\otimes l} K(u)\, du = \begin{cases} 1, & \text{if } l = 0, \\ 0, & \text{if } l = 1, \ldots, \nu - 1, \end{cases} \tag{1.5}$$

where we use the convention $u^{\otimes 0} = 1$. For the Kronecker power used in matrices and multivariable Taylor expansion as well as high-order kernels, we refer the reader to [14].

We introduce some assumptions that will be used in this paper. For $\delta > 0$, denote $\mathcal{I}(\delta) = \mathcal{I}_f(\delta) = f^{-1}([c - \delta, c + \delta]) = \{x \in \mathbb{R}^d : f(x) \in [c - \delta, c + \delta]\}$, which is a neighborhood of $\mathcal{M}$.

Assumptions

- (F) The density function f is continuous on $\mathbb{R}^d$ and has continuous partial derivatives up to order ν on $\mathcal{I}(2\delta_0)$ for $\nu \geq 2$ and some $\delta_0 > 0$. There exists $\varepsilon_0 > 0$ such that $\|\nabla f(x)\| \geq \varepsilon_0$ for all $x \in \mathcal{I}(2\delta_0)$.
- (K) K is a twice continuously differentiable kernel of νth order, with $\text{supp}(K) \subset \mathcal{B}(0, 1)$.
- (H) The bandwidth h depends on n such that $\gamma_{n,h}^{(2)} \to 0$ and $h \to 0$ as $n \to \infty$.

Discussion of the assumptions

1. The assumption $\|\nabla f(x)\| \geq \varepsilon_0$ for $x \in \mathcal{I}(2\delta_0)$ in (F) implies that the Lebesgue measure of $\mathcal{M}$ on $\mathbb{R}^d$ is zero. The level c is called a regular value of f and it is implied that $\inf_{x \in \mathbb{R}} f(x) < c < \sup_{x \in \mathbb{R}} f(x)$. This is a typical assumption in the literature of level set estimation (see, e.g., [9,20,37,40]), which guarantees that $\mathcal{M}$ has no flat parts and is a compact $(d-1)$-dimensional manifold (see Theorem 2 in [61]). In particular, under assumption (F) the following well-known margin condition (first introduced in [43]) is satisfied: $P(|f(X) - c| < \varepsilon) \leq C\varepsilon$ for some positive constant C and small ε (see Lemma 4 in [49]).
2. The geometric construction in the derivation of our theory requires $\mathcal{M}$ to have positive reach. In general, for functions with Liptschitz continuous gradient, their level sets at regular values have positive reach (See Lemma 4.11 and Theorem 4.12 in [23]). So assumption (F) can guarantee that $\mathcal{M}$ has positive reach. The requirement for the νth partial derivatives of f is mainly used to deal with the bias in the kernel estimation.
3. The estimator $\widehat{f}$ using a high-order kernel ($\nu > 2$) can take negative values, and therefore loses its interpretability for practitioners (see, e.g., page 69 of [54], and [30]). However, this is not a problem for our estimator, since we are only interested in the level set of f at a positive level, and our estimators $\widehat{\mathcal{M}}$ and $\widehat{\mathcal{L}}$ do not directly use negative values of $\widehat{f}$.
4. Assumption (H) guarantees the strong uniform consistency of the second partial derivatives of $\widehat{f}$. This is used to establish the normal compatibility between $\mathcal{M}$ and $\widehat{\mathcal{M}}$, so that we can explicitly define a homeomorphism between them. A similar assumption has been used in [17] and [47] for the construction of confidence regions of level sets.

2. Main results

For $\lambda(f, g)$ where g is a known function, the three estimators $\lambda(\widehat{f}, g)$, and $\lambda_{\varepsilon_n}^{(j)}(\widehat{f}, g)$, $j = 1, 2$, defined in (1.2)–(1.4) are investigated. In general, we use $\lambda_{\varepsilon_n}^{*}(\widehat{f}, g)$ to denote any one of them. In particular, when $\lambda_{\varepsilon_n}^{*}$ is used to denote λ, we take $\varepsilon_n \equiv 0$.

Theorem 2.1. *Let $g : \mathcal{I}(2\delta_0) \to \mathbb{R}$ be a function with bounded continuous second partial derivatives. Let $\lambda_{\varepsilon_n}^{*}$ be either λ, $\lambda_{\varepsilon_n}^{(1)}$ or $\lambda_{\varepsilon_n}^{(2)}$. Under assumptions* (F) *and* (K), *there exist constants $A > 0$, $C > 0$, $\delta_1 > 0$ and $\delta_2 > 0$ such that for all $n \geq 1$, $h > 0$, $\varepsilon_n \geq 0$ satisfying $h^{\nu} + \varepsilon_n^2 + \frac{|\log h|}{nh^{d+4}} \leq \delta_1$ and for all $0 \leq \varepsilon \leq \delta_2$,*

$$\mathbb{P}\big(\big|\lambda_{\varepsilon_n}^{*}(\widehat{f}, g) - \lambda(f, g)\big| \leq \varepsilon \vee C\big(\varepsilon_n^2 + h^{\nu} + \big(\gamma_{n,h}^{(1)}\big)^2\big)\big)$$
$$\geq 1 - A\exp\big(-nh\varepsilon^2/A\big) - A\exp\big(-nh^{d+2}\varepsilon/A\big) - A\exp\big(-nh^{d+4}/A\big). \quad (2.1)$$

Remark 2.1. Under the additional assumption (H), by applying the Borel–Cantelli lemma with $\varepsilon = \sqrt{\frac{2A\log n}{nh}} \vee \frac{2A\log n}{nh^{d+2}}$, we get $|\lambda_{\varepsilon_n}^{*}(\widehat{f}, g) - \lambda(f, g)| = O(\sqrt{\frac{\log n}{nh}} + \varepsilon_n^2 + h^{\nu} + (\gamma_{n,h}^{(1)})^2)$ almost surely.

We give an asymptotic normality result for $\lambda_{\varepsilon_n}^{*}(\widehat{f}, g) - \lambda(f, g)$ in the following theorem. To formulate the result, we need to introduce more notation and geometric concepts, especially the curvatures on level sets. We denote the tangent space of $\mathcal{M}$ at $x \in \mathcal{M}$ by $T_x(\mathcal{M})$. Let $\boldsymbol{N}(x) = \frac{\nabla f(x)}{\|\nabla f(x)\|}$ be the normalized gradient. For any $t \in \mathbb{R}$, let $T_{x,t}(\mathcal{M}) = \{y + t\boldsymbol{N}(x) : y \in T_x(\mathcal{M})\}$, which is an affine tangent plane. If $-\boldsymbol{N}$ is chosen as the surface normal of $\mathcal{M}$, then $\nabla\boldsymbol{N}(x)$ is called the shape operator (or Weingarten map). Let $G(x) = \boldsymbol{I}_d - \boldsymbol{N}(x)\boldsymbol{N}(x)^T$. Direct calculation shows that (see [34])

$$\nabla\boldsymbol{N}(x) = \big\|\nabla f(x)\big\|^{-1} G(x)\nabla^2 f(x). \quad (2.2)$$

The following matrix is equivalent to $\nabla\boldsymbol{N}(x)$ as a linear map on $T_x(\mathcal{M})$ (see (4.30)) and hence is also called the shape operator of $\mathcal{M}$ at x:

$$S(x) = \nabla\boldsymbol{N}(x)G(x) = \big\|\nabla f(x)\big\|^{-1} G(x)\nabla^2 f(x)G(x). \quad (2.3)$$

The principal curvatures of $\mathcal{M}$ at x, denoted by $\kappa_1(x) \geq \cdots \geq \kappa_{d-1}(x)$, are $d-1$ eigenvalues of the shape operator $S(x)$ associated with eigenvectors orthogonal to $\boldsymbol{N}(x)$. The mean curvature, denoted by $H(x)$ is the sum of the $(d-1)$ principal curvatures, i.e., $H(x) = \mathbf{tr}[S(x)]$. Note that both the sum and the average of the principal curvatures are often called the mean curvature in the literature, and we use the former for convenience. Consider the plug-in estimators of the above quantities. Let $\widehat{\boldsymbol{N}}(x) = \frac{\nabla\widehat{f}(x)}{\|\nabla\widehat{f}(x)\|}$, $\widehat{G}(x) = \boldsymbol{I}_d - \widehat{\boldsymbol{N}}(x)\widehat{\boldsymbol{N}}(x)^T$, $\widehat{S}(x) = \nabla\widehat{\boldsymbol{N}}(x)\widehat{G}(x)$, and $\widehat{H}(x) = \mathbf{tr}[\widehat{S}(x)]$. For a kernel function K and $x \in \mathcal{M}$, let $R_K(x) = \int_{\mathbb{R}}(\int_{T_{x,t}(\mathcal{M})} K(u)\,d\mathscr{H}(u))^2\,dt$. If K is spherically symmetric, then $R_K(x)$ is a constant for all $x \in \mathcal{M}$ and we write $R(K) = R_K(x)$. For two sequences a_n and b_n, we denote $a_n \gg b_n$ if $a_n/b_n \to \infty$ as $n \to \infty$. We use $\xrightarrow{P}$ and $\xrightarrow{D}$ to denote convergences in probability and in distribution.

Theorem 2.2. *Let $g : \mathcal{I}(2\delta_0) \to \mathbb{R}$ be a function with bounded continuous second partial derivatives. Let $\lambda_{\varepsilon_n}^{*}$ be either λ, $\lambda_{\varepsilon_n}^{(1)}$ or $\lambda_{\varepsilon_n}^{(2)}$ with $\varepsilon_n \to 0$. Under assumptions* (K), (F) *and* (H), *with μ_1 and μ_2 being constants given in* (4.87) *and* (4.72), *and $\mu^{(j)}$ given in* (4.80) *for $j = 1, 2$, we have the following results depending the relationship among the rates $\frac{1}{nh^{d+2}}$, h^{ν}, $\frac{1}{\sqrt{nh}}$, and ε_n^2.*

(i) *If* $\frac{1}{nh^{d+2}} \gg \max(h^\nu, \frac{1}{\sqrt{nh}}, \varepsilon_n^2)$, *then* $nh^{d+2}[\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f, g)] \xrightarrow{P} \mu_1$.

(ii) *If* $h^\nu \gg \max(\frac{1}{nh^{d+1}}, \frac{1}{\sqrt{nh}}, \varepsilon_n^2)$, *then* $h^{-\nu}[\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f, g)] \xrightarrow{P} \mu_2$.

(iii) *If* $\varepsilon_n^2 \gg \max(\frac{1}{nh^{d+1}}, h^\nu, \frac{1}{\sqrt{nh}})$, *then* $\varepsilon_n^{-2}[\lambda^{(j)}_{\varepsilon_n}(\widehat{f}, g) - \lambda(f, g)] \xrightarrow{P} \mu^{(j)}$ *for* $j = 1, 2$.

(iv) *If* $\frac{1}{\sqrt{nh}} \gg \max(\frac{1}{nh^{d+1}}, h^\nu, \varepsilon_n^2)$, *then*

$$\sqrt{nh}\big[\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f, g)\big] \xrightarrow{D} \mathcal{N}\big(0, \sigma^2\big), \tag{2.4}$$

where $\sigma^2 = c\lambda(f, w_g^2 R_K)$ *with* $w_g(x) = \|\nabla f(x)\|^{-1}[\boldsymbol{N}(x)^T \nabla g(x) + H(x)g(x)]$.

(v) *Suppose we choose to use a spherically symmetric kernel function* K, *and then* $\sigma^2 = cR(K)\lambda(f, w_g^2)$ *in* (*iv*). *For a nonnegative sequence* $\tau_n = o(1)$, *let* $\widehat{\sigma}^2_{\tau_n} = cR(K)\lambda^*_{\tau_n}(\widehat{f}, \widehat{w}_g^2)$, *where* $\widehat{w}_g(x) = \|\nabla \widehat{f}(x)\|^{-1}[\widehat{\boldsymbol{N}}(x)^T \nabla g(x) + \widehat{H}(x)g(x)]$. *Suppose* $\sigma^2 \neq 0$. *Under the assumption in* (*iv*), *we have*

$$\sqrt{nh}\widehat{\sigma}^{-1}_{\tau_n}\big[\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f, g)\big] \xrightarrow{D} \mathcal{N}(0, 1). \tag{2.5}$$

Remark 2.2.

(a) We can write $\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f, g) = \mathrm{I}_n + \mathrm{II}_n$, where $\mathrm{I}_n = \lambda^*_{\varepsilon_n}(\widehat{f}, g) - \mathbb{E}[\lambda^*_{\varepsilon_n}(\widehat{f}, g)]$ is a centered stochastic term, and $\mathrm{II}_n = \mathbb{E}[\lambda^*_{\varepsilon_n}(\widehat{f}, g)] - \lambda(f, g)$ is the bias. The results in (i)–(iii) are formulated in a way that shows the asymptotic expressions of the bias II_n in different scenarios. In fact, $(nh^{d+1})^{-1}\mu_1 + h^\nu \mu_2 + \varepsilon_n^2 \mu^{(j)}$ is the leading term in the bias II_n. The result in (iv) occurs when the bias is dominated by I_n. See (4.77) in the proof. We point out that the signs of the components μ_1, μ_2 and $\mu^{(j)}$ in the leading term of the bias may not be the same, depending on the unknown f. Therefore in some cases one may choose h and ε_n to cancel out the leading terms in the bias, which is unlike the typical variance-bias tradeoff strategy used for bandwidth selection for kernel density estimators.

(b) The result in (v) can be used to construct a confidence interval for $\lambda(f, g)$. For $0 < \alpha < 1$, let $z_{\alpha/2}$ be the $(1 - \alpha/2)$ quantile of $\mathcal{N}(0, 1)$. With rates of h, τ_n and ε_n chosen complying with the conditions in (v), a $(1 - \alpha)$ asymptotic confidence interval for $\lambda(f, g)$ is given by

$$\left[\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \frac{1}{\sqrt{nh}}\widehat{\sigma}_{\tau_n} z_{\alpha/2}, \lambda^*_{\varepsilon_n}(\widehat{f}, g) + \frac{1}{\sqrt{nh}}\widehat{\sigma}_{\tau_n} z_{\alpha/2}\right].$$

Note that the choice of λ^* used in $\widehat{\sigma}_{\tau_n}$ does not have to be the same one as in $\lambda^*_{\varepsilon_n}(\widehat{f}, g)$.

(c) The theorem implies that $\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f, g) = O_p(\alpha_{n,h})$, where $\alpha_{n,h} = \frac{1}{nh^{d+2}} + h^\nu + \frac{1}{\sqrt{nh}} + \varepsilon_n^2$. Below we give an interpretation of the rates in $\alpha_{n,h}$. Here $O_p(\varepsilon_n^2)$ accounts for the difference between $\lambda^*_{\varepsilon_n}(\widehat{f}, g)$ and $\lambda(\widehat{f}, g)$, that is, the error in approximating the surface integral by an integral over a neighborhood of width ε_n around the level set. We then focus on $\lambda(\widehat{f}, g) - \lambda(f, g)$. It turns out that we have the approximation $\lambda(\widehat{f}, g) - \lambda(f, g) \approx T_{n,1} + T_{n,2}$, where $T_{n,1} = \lambda(f, w_g \times (f - \widehat{f}))$ and $T_{n,2} = \lambda(f, p(\nabla \widehat{f} - \nabla f))$ with $p(\nabla \widehat{f} - \nabla f)$ a quadratic form of $\nabla \widehat{f} - \nabla f$. It is known that $\widehat{f}(x) - \mathbb{E}\widehat{f}(x)$ has a standard rate of $1/\sqrt{nh^d}$. Due to the integrals on a $(d-1)$-dimensional manifold $\mathcal{M}$, we gain $(d-1)$ powers of h, which results in the rate $1/\sqrt{nh}$ for the stochastic part of $T_{n,1}$. The fact that $(nh^{d+2})^{-1/2}$ is the rate of convergence of $\nabla \widehat{f}(x) - \mathbb{E}\nabla \widehat{f}(x)$ explains $(nh^{d+2})^{-1}$ as a rate for the bias part of $T_{n,2}$. The rate h^ν accounts for the remaining bias in the overall estimation. To have the asymptotic normality in (2.4), we make the bias asymptotically negligible using the condition $\frac{1}{\sqrt{nh}} \gg \max(\frac{1}{nh^{d+1}}, h^\nu, \varepsilon_n^2)$, which under assumption (H) requires $\nu > d+1$.

(d) We discuss some special cases of g. If $g \equiv 1$, then $w_g(x) = \|\nabla f(x)\|^{-1} H(x)$ and the surface integral $\lambda(f, w_g^2)$ in σ^2 is a weighted Willmore energy (see the Introduction section). The asymptotic normal distribution in part (iv) can be degenerate in some extreme cases for such g that $w_g \equiv 0$, which are excluded from consideration in part (v).

In order to prove the rates of convergence and asymptotic distributions in Theorem 2.2, we derive a few results that are also interesting in their own right. We consider the direct plug-in estimator $\lambda(\widehat{f}, g)$ in Section 2.1, and then $\lambda_{\varepsilon_n}^{(j)}(\widehat{f}, g)$, $j = 1, 2$, in Section 2.2.

2.1. Direct plug-in estimation

The proof of Theorem 2.2 when $\lambda_{\varepsilon_n}^* = \lambda$ is built upon the results below in this section. One of the main challenges is that the domains of integrals in $\lambda(f, g)$ and $\lambda(\widehat{f}, g)$ are not the same, which makes the comparison difficult. Briefly speaking, our strategy is to establish a diffeomorphism between the two domains and utilize the area formula in differential geometry to convert $\lambda(\widehat{f}, g)$ into a surface integral with the same integral domain as $\lambda(f, g)$. We provide some heuristic for this procedure, which also explains the role of the curvature of level sets in σ^2.

In Figure 2, we consider a circle with radius r as our level set $\mathcal{M}$. An ideal estimator $\widehat{\mathcal{M}}$ is a circle with the same center and radius $r + \Delta r$. Here we allow Δr to be either positive or negative, and the two dotted circles on the graph are two possible versions of $\widehat{\mathcal{M}}$. In this heuristic we only consider the perimeter of the circles, corresponding to $g \equiv 1$ in the surface integral on level sets. From elementary geometry it is known that the difference between the perimeters of $\widehat{\mathcal{M}}$ and $\mathcal{M}$ is $2\pi\Delta r$. We focus on the local geometry to better understand the behavior of this difference. We consider short arcs $\widehat{BC}$ on $\mathcal{M}$ and $\widehat{B^*C^*}$ on $\widehat{\mathcal{M}}$, where both BB^* and CC^* can be extended to go through the center of the circles. The difference between the lengths of the two arcs is $|\widehat{B^*C^*}| - |\widehat{BC}| = \Delta r r^{-1} |\widehat{BC}|$, where r^{-1} can be understood as the curvature of $\mathcal{M}$, and we denote it by H (the same notation for mean curvature). Now imagine that both $\mathcal{M}$ and $\widehat{\mathcal{M}}$ are slightly deformed from circles, and we allow both Δr and H to depend on location x, then the difference of arc lengths can be approximated by a surface integral $\int_{\widehat{BC}} \Delta r(x) H(x)\, d\mathscr{H}(x)$. It is straightforward to extend this approximation to the entire surface, i.e., $\mathscr{H}(\widehat{\mathcal{M}}) - \mathscr{H}(\mathcal{M}) \approx \int_{\mathcal{M}} \Delta r(x) H(x)\, d\mathscr{H}(x)$. Another key aspect is $\Delta r(x) \approx \frac{f(x) - \widehat{f}(x)}{\|\nabla f(x)\|}$, where $\|\nabla f(x)\|$ reflects a ratio between vertical and horizontal variations in level set estimation. See

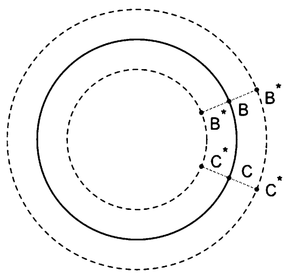

Figure 2. This figure provides some heuristic to understand the connection and difference between the estimation of Hausdorff measure of $\mathcal{M}$ and the Lebesgue measure of $\widehat{\mathcal{L}} \Delta \mathcal{L}$.

Lemma 2.2 below and its remark for details. Note that this approximation has included the sign of Δr into consideration. So overall

$$\mathscr{H}(\widehat{\mathcal{M}}) - \mathscr{H}(\mathcal{M}) \approx \int_{\mathcal{M}} \frac{f(x) - \widehat{f}(x)}{\|\nabla f(x)\|} H(x)\, d\mathscr{H}(x). \tag{2.6}$$

The above approximation explains the role of curvature in the estimation of surface integrals on level sets. In Remark 2.2 c) we have given the approximation $\lambda(\widehat{f}, g) - \lambda(f, g) \approx T_{n,1} + T_{n,2}$, which coincides with (2.6) when $g \equiv 1$ if $T_{n,2}$ is ignored.

Using Figure 2 again, we heuristically explain the difference and connection between $\mathscr{H}(\widehat{\mathcal{M}}) - \mathscr{H}(\mathcal{M})$ and the volume of the symmetric difference $\widehat{\mathcal{L}}\Delta\mathcal{L}$, which corresponds to the band between $\mathcal{M}$ and $\widehat{\mathcal{M}}$. From elementary geometry, it is known that the volume of this band is approximately $\pi r|\Delta r|$, if Δr is small. Using the arc notation, we have that Volume$(\widehat{BCC^*B^*}) \approx |\Delta r||\widehat{BC}|$. So overall

$$\text{Volume}(\widehat{\mathcal{L}}\Delta\mathcal{L}) \approx \int_{\mathcal{M}} \frac{|f(x) - \widehat{f}(x)|}{\|\nabla f(x)\|}\, d\mathscr{H}(x), \tag{2.7}$$

which is an L_1-type integral (see [45]). The approximations in (2.6) and (2.7) provide some insight into the connection and difference between estimating the surface integral on $\mathcal{M}$ and the set-theoretic measure of $\widehat{\mathcal{L}}\Delta\mathcal{L}$: both can be approximated by surface integrals on $\mathcal{M}$ and the integrands are related to $f - \widehat{f}$, but the latter integrates an absolute value and only the former is (asymptotically) impacted by the curvatures of the level sets.

Below we formulate the results that have been outlined in the heuristic above. Essentially we need to establish a homeomorphism between $\mathcal{M}$ and $\widehat{\mathcal{M}}$, which is given in Lemma 2.1 below. We show the approximation $\Delta r(x) \approx \frac{f(x)-\widehat{f}(x)}{\|\nabla f(x)\|}$ in Lemma 2.2. Theorem 2.3 gives an exact expression of $\lambda(\widehat{f}, g)$ as a surface integral on $\mathcal{M}$. The asymptotic normality of $\lambda(f, w_g \times (\widehat{f} - \mathbb{E}\widehat{f}))$ is given in Theorem 2.4.

The following lemma is regarding the reach of the true and estimated density level sets as well as their normal compatibility (see Section 1.1 for these geometric concepts). It is similar to Lemma 1 in [17], but our result is under slightly different assumptions and holds uniformly for a collection of density level sets. Also see Theorems 1 and 2 in [61] for relevant results. With a twice differentiable kernel K, let

$$\beta_n^{(0)} = \sup_{x\in\mathbb{R}^d} \left|\widehat{f}(x) - f(x)\right|,$$

$$\beta_n^{(1)} = \sup_{x\in\mathcal{I}(2\delta_0)} \left\|\nabla\widehat{f}(x) - \nabla f(x)\right\|,$$

$$\beta_n^{(2)} = \sup_{x\in\mathcal{I}(2\delta_0)} \left\|\nabla^2\widehat{f}(x) - \nabla^2 f(x)\right\|_F.$$

It can be shown that $\beta_n^{(0)} = O_{\text{a.s.}}(\gamma_{n,h}^{(0)}) + O(h^\nu)$, $\beta_n^{(1)} = O_{\text{a.s.}}(\gamma_{n,h}^{(1)}) + o(h^{\nu-1})$, and $\beta_n^{(2)} = O_{\text{a.s.}}(\gamma_{n,h}^{(2)}) + o(h^{\nu-2})$ under the assumptions (F), (K) and (H). See Lemmas 2 and 3 in [2] and Lemma 4.1 in Section 4.

Lemma 2.1. *Under the assumption* (F), *we have the following results.*

(i) *There exists a constant* $t_0 > 0$ *such that* $(\mathcal{I}(\delta_0) \oplus t_0) \subset \mathcal{I}(2\delta_0)$.
(ii) *There exists a constant* $r_0 > 0$, *such that the reach* $\rho(\mathcal{M}_\tau) > r_0$ *for all* $\tau \in [c - \delta_0, c + \delta_0]$.

(iii) *There exist constants* $c_0 > 0$ *and* $\eta_0 > 0$ *such that when* $\beta_n^{(0)} \leq c_0$ *we have*

$$\sup_{\tau\in[c-\delta_0,c+\delta_0]} d_H(\widehat{\mathcal{M}}_\tau, \mathcal{M}_\tau) \leq \eta_0\beta_n^{(0)}.$$

(iv) *There exists a constant* $C_0 > 0$ *such that when* $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \leq C_0$, *the following* (iv-1) *and* (iv-2) *hold.*
  (iv-1) *There exists a constant* $r_1 > 0$ *such that* $\rho(\widehat{\mathcal{M}}_\tau) > r_1$ *for all* $\tau \in [c - \delta_0, c + \delta_0]$.
  (iv-2) $\widehat{\mathcal{M}}_\tau$ *and* $\mathcal{M}_\tau$ *are normal compatible for all* $\tau \in [c - \delta_0, c + \delta_0]$.

With this result, we can explicitly define homeomorphisms between level sets and their plug-in estimators. For $x \in \mathcal{M}_\tau$ with $|\tau - c| \leq \delta_0$, and $t \in \mathbb{R}$, let

$$\zeta_x(t) = x + t\boldsymbol{N}(x). \tag{2.8}$$

Furthermore, let $t_n(x) = \operatorname{argmin}_t\{|t| : \zeta_x(t) \in \widehat{\mathcal{M}}_\tau\}$, which is the first time point when $\zeta_x(t)$ hits $\widehat{\mathcal{M}}_\tau$ starting from a point x on $\mathcal{M}_\tau$, and

$$P_n(x) = \zeta_x\big(t_n(x)\big). \tag{2.9}$$

Under the assumptions in Lemma 2.1, $\widehat{\mathcal{M}}_\tau$ and $\mathcal{M}_\tau$ are normal compatible. Hence, $t_n(x)$ is uniquely defined, and P_n is a homeomorphism between $\mathcal{M}_\tau$ and $\widehat{\mathcal{M}}_\tau$. The following lemma gives an approximation for $t_n(x)$.

Lemma 2.2. *Suppose that the assumption* (F) *holds. When* $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \leq C_0$, *where* C_0 *is given in Lemma* 2.1, *for any point* $x \in \mathcal{I}(\delta_0)$, *we have*

$$t_n(x) = \big\|\nabla f(x)\big\|^{-1}\big[f(x) - \widehat{f}(x)\big] + \delta_n(x), \tag{2.10}$$

where for some constant $C_1 > 0$

$$\sup_{x\in\mathcal{I}(\delta_0)} \big|\delta_n(x)\big| \leq C_1\big(\beta_n^{(0)}\beta_n^{(1)} + \big(\beta_n^{(0)}\big)^2\big). \tag{2.11}$$

Remark 2.3. Note that $|t_n(x)| = \|P_n(x) - x\|$. Then the above result links the local horizontal variation $t_n(x)$ with the local vertical variation $f(x) - \widehat{f}(x)$. Here $\|\nabla f(x)\|$ on the right-hand side of (2.10) can be understood as a directional derivative of f in the gradient direction, that is, $\|\nabla f(x)\| = \langle\nabla f(x), \boldsymbol{N}(x)\rangle$, and it reflects the asymptotic rate of change between the local vertical and horizontal variations, as $P_n(x) - x$ is parallel to the direction of $\boldsymbol{N}(x)$.

In the next theorem, we find a function $s_{n,g}$, which depends on g, P_n, and the principal curvatures κ_i on $\mathcal{M}$ such that $\lambda(\widehat{f}, g) = \lambda(f, s_{n,g})$.

Theorem 2.3. *Suppose that the assumption* (F) *holds. Assume that* g *is integrable on* $\mathcal{I}(2\delta_0)$. *When* $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \leq C_0$, *where* C_0 *is given in Lemma* 2.1, *we have*

$$\lambda(\widehat{f}, g) = \lambda\bigg(f, \frac{(g\circ P_n) \times \prod_{i=1}^{d-1}[1 + t_n\kappa_i]}{\langle\widehat{\boldsymbol{N}}\circ P_n, \boldsymbol{N}\rangle}\bigg). \tag{2.12}$$

As indicated in Remark 2.2(c), one of the leading terms in the approximation of $\lambda(\widehat{f}, g) - \lambda(f, g)$ is $\lambda(f, w_g \times (f - \widehat{f}))$. The asymptotic form of the bias in the estimation of $\lambda(f, g)$ is implicitly given in Theorem 2.2 and discussed in Remark 2.2(a). Next, we focus on the asymptotic normality of the centered stochastic variation in the estimation of $\lambda(f, g)$, essentially $\lambda(f, w_g \times (\mathbb{E}\widehat{f} - \widehat{f}))$, which assists assertion (iv) in Theorem 2.2.

Theorem 2.4. *Suppose that the assumptions* (F) *and* (K) *hold. Let* $w : \mathcal{M} \to \mathbb{R}$ *be a bounded continuous function. There exist constants* $C_2 > 0$, $h_0 > 0$ *and* $t_0 > 0$ *such that for all* $n \geq 1$, $h \in (0, h_0]$ *and* $\varepsilon \in [0, t_0]$,

$$\mathbb{P}\big(\big|\lambda\big(f, w \times (\widehat{f} - \mathbb{E}\widehat{f})\big)\big| \geq \varepsilon\big) \leq 2\exp\big(-C_2 nh\varepsilon^2\big). \tag{2.13}$$

If additionally $nh \to \infty$ *and* $h \to 0$, *then with* $\sigma^2 = c\lambda(f, w^2 R_K)$ *we have*

$$\sqrt{nh}\lambda\big(f, w \times (\widehat{f} - \mathbb{E}\widehat{f})\big) \xrightarrow{D} \mathscr{N}\big(0, \sigma^2\big). \tag{2.14}$$

2.2. Estimation using integrals over neighborhoods of level sets

We consider $\lambda_{\varepsilon_n}^{(1)}(\widehat{f}, g)$ and $\lambda_{\varepsilon_n}^{(2)}(\widehat{f}, g)$ as two alternative estimators of $\lambda(f, g)$, where $\lambda_{\varepsilon_n}^{(1)}$ and $\lambda_{\varepsilon_n}^{(2)}$ are given in (1.3) and (1.4). It can be seen that (by Proposition A.1 of [9] and Section 3.4 of [27])

$$\lambda_{\varepsilon_n}^{(1)}(\widehat{f}, g) = \frac{1}{2\varepsilon_n}\int_{-\varepsilon_n}^{\varepsilon_n} \lambda_{c+\varepsilon}(\widehat{f}, g)\, d\varepsilon, \tag{2.15}$$

$$\lambda_{\varepsilon_n}^{(2)}(\widehat{f}, g) = \frac{1}{2\varepsilon_n}\int_{0}^{\varepsilon_n} \int_{\partial(\widehat{f}^{-1}(c)\oplus\varepsilon)} g(x)\, d\mathscr{H}(x)\, d\varepsilon, \tag{2.16}$$

where $\partial(\widehat{f}^{-1}(c) \oplus \varepsilon) = \{x : d(x, \widehat{f}^{-1}(c)) = \varepsilon\}$. So $\lambda_{\varepsilon_n}^{(1)}(\widehat{f}, g)$ and $\lambda_{\varepsilon_n}^{(2)}(\widehat{f}, g)$ can be viewed as two local averages of surface integrals in a neighborhood of $\widehat{f}^{-1}(c)$. They may have some computational advantage over $\lambda(\widehat{f}, g)$, because numerical approximation for integrals over subsets of $\mathbb{R}^d$ in (1.3) and (1.4) is less challenging than numerical approximation to a surface integral in (1.2) for $d \geq 3$.

Even though both of $\lambda_{\varepsilon_n}^{(1)}(\widehat{f}, g)$ and $\lambda_{\varepsilon_n}^{(2)}(\widehat{f}, g)$ are based on integration over some neighborhoods of $\widehat{f}^{-1}(c)$, their integration domains have different shapes, which can potentially impact their performance. The region $\widehat{f}^{-1}[c - \varepsilon_n, c + \varepsilon_n]$ is an implicit tube defined through variation of the vertical levels of $\widehat{f}$, while the tube $\widehat{f}^{-1}(c) \oplus \varepsilon_n$ has a constant radius, grown through horizontal variation from $\widehat{f}^{-1}(c)$. Note that the roles of ε_n in $\lambda_{\varepsilon_n}^{(1)}$ and $\lambda_{\varepsilon_n}^{(2)}$ are the magnitude of the vertical and horizontal variations, respectively, and the relevant geometric interpretation for why $\|\nabla\widehat{f}\|$ appears in the integrand in (1.3) can be found in Remark 2.3. When ε_n is appropriately chosen, the two different types of tubes have been used as asymptotic confidence regions for $\widehat{f}^{-1}(c)$ in [17] and [47]. Finite sample performance of these two types of confidence regions are compared in [47] and it has been observed that the two types confidence regions behave differently, especially when c is close to 0 or or a critical value of f. We expect these scenarios also impact the integration in $\lambda_{\varepsilon_n}^{(j)}(\widehat{f}, g)$, $j = 1, 2$.

Let $\lambda_{\varepsilon_n}^{*}$ be one of $\lambda_{\varepsilon_n}^{(1)}$ and $\lambda_{\varepsilon_n}^{(2)}$. Then it can be seen that $\lambda_{\varepsilon_n}^{*}(\widehat{f}, g) - \lambda(f, g) = [\lambda_{\varepsilon_n}^{*}(\widehat{f}, g) - \lambda(\widehat{f}, g)] + [\lambda(\widehat{f}, g) - \lambda(f, g)]$. The results developed in Section 2.1 are useful to study $\lambda(\widehat{f}, g) - \lambda(f, g)$. The difference $\lambda_{\varepsilon_n}^{*}(\widehat{f}, g) - \lambda(\widehat{f}, g)$ accounts for the extra error in the estimation caused by replacing a surface integral by an integration over a tubular neighborhood of $\widehat{f}^{-1}(c)$ with (vertical or horizontal) radius ε_n. Similar to (2.8), for $x \in \widehat{\mathcal{M}}_c$, define $\widehat{\zeta}_x(t) = x + t\widehat{N}(x)$, $t \in \mathbb{R}$, and let $\widehat{t}_\varepsilon(x) = \operatorname{argmin}_t\{|t| :$

$\widehat{\zeta}_x(t) \in \widehat{\mathcal{M}}_{c+\varepsilon}\}$. Define $P_\varepsilon(x) = \widehat{\zeta}_x(\widehat{t_\varepsilon}(x)) = x + \widehat{t_\varepsilon}(x)\widehat{\boldsymbol{N}}(x)$. The following lemma is analogous to Lemmas 2.1 and 2.2, which implies that $\widehat{t_\varepsilon}$ is uniquely defined and P_ε is a diffeomorphism between $\widehat{\mathcal{M}}_c$ and $\widehat{\mathcal{M}}_{c+\varepsilon}$, when ε is small enough.

We need to introduce more notation. For $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, let $|\alpha| = \alpha_1 + \cdots + \alpha_d$. For a smooth function $g : \mathbb{R}^d \to \mathbb{R}$, define $\partial^{(\alpha)} g(x) = \frac{\partial^{|\alpha|}}{\partial^{\alpha_1} x_1 \cdots \partial^{\alpha_d} x_d} g(x)$, $x \in \mathbb{R}^d$. Here $\partial^{(\alpha)} g = g$ if $|\alpha| = 0$ by convention. Let k_0 be a nonnegative integer and $\varepsilon \geq 0$. For a k_0 times differentiable function $\hbar$: $[\mathcal{I}(\delta_0) \oplus \varepsilon] \to \mathbb{R}$, define the modulus of continuity

$$\psi_\hbar^{(k)}(\varepsilon) = \max_{\alpha \in \mathbb{N}^d : |\alpha| = k} \sup_{x \in \mathcal{I}(\delta_0)} \sup_{x' \in \mathcal{B}(x,\varepsilon)} \left|\partial^{(\alpha)} \hbar(x) - \partial^{(\alpha)} \hbar(x')\right|, \quad \text{for } k = 0, 1, \dots, k_0. \tag{2.17}$$

Note that assumption (F) guarantees that $\psi_f^{(k)}(\varepsilon) = O(\varepsilon)$ for $k = 0, 1, \dots, \nu - 1$ and $\psi_f^{(\nu)}(\varepsilon) = o(1)$, as $\varepsilon \to 0$.

Lemma 2.3. *Under the assumption* (F), *when* $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \leq C_0$, *where* C_0 *is given in Lemma* 2.1, *there exists a constant* $r_2 > 0$ *such that for all* $\varepsilon \in [-r_2, r_2]$, (*i*) $\widehat{\mathcal{M}}_c$ *and* $\widehat{\mathcal{M}}_{c+\varepsilon}$ *are normal compatible and* $\widehat{\mathcal{M}}_{c+\varepsilon} \subset \mathcal{I}(2\delta_0)$; *and* (ii)

$$\widehat{t_\varepsilon}(x) = \varepsilon \left\|\nabla \widehat{f}(x)\right\|^{-1} - \frac{1}{2}\varepsilon^2 \left\|\nabla \widehat{f}(x)\right\|^{-3} \left\|\widehat{\boldsymbol{N}}(x)\right\|^2_{\nabla^2 \widehat{f}(x)} + \eta_\varepsilon(x), \tag{2.18}$$

where $|\eta_\varepsilon(x)| \leq C_3(|\varepsilon|^3 + \varepsilon^2(\beta_n^{(2)} + \psi_f^{(2)}(\eta_0|\varepsilon|)))$ *for all* $x \in \widehat{\mathcal{M}}_c$, *for some constant* $C_3 > 0$.

The following theorem gives an approximation to the difference between $\lambda_{\varepsilon_n}^{(j)}(\widehat{f}, g)$ and $\lambda(\widehat{f}, g)$. In particular, this difference is shown to be of the order of ε_n^2.

Theorem 2.5. *Let* $g : \mathcal{I}(2\delta_0) \to \mathbb{R}$ *be a function with bounded continuous second derivatives. Let* $\phi_{n,\varepsilon} = [\beta_n^{(1)} + \beta_n^{(2)} + |\varepsilon| + \psi_f^{(2)}(\eta_0|\varepsilon|) + \psi_g^{(2)}((\eta_0 \vee 1)|\varepsilon|)]\varepsilon^2$. *Under the assumption* (F), *when* $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \leq C_0$, *where* C_0 *is given in Lemma* 2.1, *there exists a constant* $r_3 > 0$ *such that for* $\varepsilon_n \in (0, r_3]$, *we have*

$$\lambda_{\varepsilon_n}^{(j)}(\widehat{f}, g) = \lambda(\widehat{f}, g) + \varepsilon_n^2 \mu_n^{(j)} + \rho_n^{(j)}, \quad j = 1, 2, \tag{2.19}$$

with $\mu_n^{(j)} = \frac{1}{3}\lambda(\widehat{f}, \theta_n^{(j)})$, *where* $\theta_n^{(1)}$ *and* $\theta_n^{(2)}$ *are given in* (4.50) *and* (4.51), *and* $|\rho_n^{(j)}| \leq C_4 \phi_{n,\varepsilon_n}$ *for some constant* $C_4 > 0$.

As a slight generalization, the next corollary is a consistence result of surface integrals when the integrand is unknown and can be consistently estimated. The corollary is useful, for example, in deriving the asymptotic normality in (2.5), where one needs to show that $\widehat{\sigma}_{\tau_n}^2$ is a consistent estimator of σ^2 as $\tau_n \to 0$.

Corollary 2.1. *Let* $\lambda_{\varepsilon_n}^*$ *be either* λ, $\lambda_{\varepsilon_n}^{(1)}$ *or* $\lambda_{\varepsilon_n}^{(2)}$. *Let* $p : \mathcal{I}(2\delta_0) \to \mathbb{R}$ *be a bounded continuous function and* p_n *be a sequence of functions on* $\mathcal{I}(2\delta_0)$. *Under the assumption* (F), *when* $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \leq C_0$ *and* $\varepsilon_n \in [0, r_3]$, *where* C_0 *is given in Lemma* 2.1, *and* r_3 *is given in Theorem* 2.5, *we have*

$$\begin{aligned}\left|\lambda_{\varepsilon_n}^*(\widehat{f}, p_n) - \lambda(f, p)\right| \leq{}& C_5\big(\eta_{p,n} + \psi_p^{(0)}\big(\eta_0 \beta_n^{(0)}\big) + \beta_n^{(0)} + \big(\beta_n^{(1)}\big)^2 \\ &+ \varepsilon_n + \psi_p^{(0)}\big((\eta_0 \vee 1)\varepsilon_n\big)\big),\end{aligned} \tag{2.20}$$

for some constant $C_5 > 0$, *where* $\eta_{p,n} = \sup_{x\in\mathcal{I}(2\delta_0)} |p_n(x) - p(x)|$. *Furthermore, if* $\eta_{p,n} = o_p(1)$, *then under assumptions* (*K*) *and* (*H*), *we have*

$$\lambda^*_{\varepsilon_n}(\widehat{f}, p_n) - \lambda(f, p) = o_p(1). \tag{2.21}$$

3. Discussion

In this paper, we consider nonparametric estimation of $\lambda(f, g)$, which is a surface integral on density level set $f^{-1}(c)$, where the integrand g is assumed to be known (except in Corollary 2.1). We study three types of estimators: $\lambda(\widehat{f}, g)$, $\lambda^{(1)}_{\varepsilon_n}(\widehat{f}, g)$ and $\lambda^{(2)}_{\varepsilon_n}(\widehat{f}, g)$, among which $\lambda(\widehat{f}, g)$ is a direct plug-in estimator, and the estimators using $\lambda^{(1)}_{\varepsilon_n}$ and $\lambda^{(2)}_{\varepsilon_n}$ are based on different neighborhoods of $\widehat{f}^{-1}(c)$. Our main results are the rates of convergence and asymptotic normality for these estimators. Apparently our methods can be extended to level set estimation in regression problems and the intersection of multiple level sets (see [46]).

To make the bias asymptotically negligible, the bandwidth used in the estimation of surface integrals on level sets needs to be appropriately selected. This is especially important if one would like to construct confidence intervals for the population surface integrals using the asymptotic normality results (parts (iv) and (v) in Theorem 2.2). An alternative approach is to explicitly correct the bias by subtracting its estimators from $\lambda^*_{\varepsilon_n}(\widehat{f}, g)$. This leads to the problem of estimating surface integrals on level sets with unknown integrands, the examples of which include all the leading terms in the bias (μ_1, μ_2, $\mu^{(1)}$, and $\mu^{(2)}$ in Theorem 2.2). The estimation of surface integrals with unknown integrands has many other applications, such as the estimation of Willmore energy, and Minkowski functionals (see the Introduction section). Also the bandwidth selection for level set estimation is another application [45]. For the estimation of surface integrals with unknown integrands, our Corollary 2.1 gives a consistence result, and we will address the rate of convergence and asymptotic normality elsewhere.

In a different context, a consistency result in the estimation of Minkowski functionals is recently obtained by [21], where they consider the (compact) support S of a density function f. They impose on S an assumption called the polynomial volume property. In particular, if S is a set with positive reach, then the Lebesgue measure of $S \oplus \varepsilon$ can be expanded as a polynomial of ε of degree d with coefficients being the Minkowski functionals of S (up to some constants) when $0 < \varepsilon < \rho(S)$ (Theorem 5.6, [23]). Given an i.i.d. sample $\mathcal{X}_n$ of f, [21] estimates the Minkowski functionals of S by fitting the polynomial with the Lebesgue measure of $\mathcal{X}_n \oplus \varepsilon$ for a range of ε. Their method can be adapted to the context of this paper to estimate the Minkowski functionals of $\mathcal{L}$, where one may use the Lebesgue measure of $\widehat{\mathcal{L}} \oplus \varepsilon$ for a range of ε to fit a polynomial of degree d, assuming that $\mathcal{L}$ has positive reach. We note the it is still an open problem to study the rate of convergence and asymptotic distributions for the estimators given in [21].

Another open problem is the minimax rates of estimating the surface integrals on level sets. We leave the study of this interesting question to future work.

4. Proofs

We denote $\partial^{(\alpha)} K_h(\cdot) = \partial^{(\alpha)} K(\cdot/h)$ for $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ in all the proofs. We also denote $Q_0 = \sup_{x\in\mathcal{I}(2\delta_0)} \|\nabla^2 f(x)\|_F$, $J_0 = \sup_{x\in\mathcal{I}(2\delta_0)} |g(x)|$, $J_1 = \sup_{x\in\mathcal{I}(2\delta_0)} \|\nabla g(x)\|$, and $J_2 = \sup_{x\in\mathcal{I}(2\delta_0)} \|\nabla^2 g(x)\|_F$.

Lemma 4.1. *Fix* $\alpha \in \mathbb{N}^d$ *with* $|\alpha| \in \{0, 1, 2\}$. *We denote* $\mathcal{S} = \mathbb{R}^d$ *when* $|\alpha| = 0$, *and* $\mathcal{S} = \mathcal{I}(\delta_0)$ *when* $|\alpha| = 1, 2$. *Suppose that assumptions* (*K*) *and* (*F*) *hold. There exists a constant* $C > 0$ *such that for all*

$n \geq 1$, $h \in (0,1)$, and $\tau > 0$ satisfying $nh^{d+2|\alpha|} \geq \tau$ and $nh^{d+2|\alpha|} \geq |\log h|$, we have

$$\mathbb{P}\left(\sup_{x\in\mathcal{S}}\left|\partial^{(\alpha)}\widehat{f}(x) - \mathbb{E}\partial^{(\alpha)}\widehat{f}(x)\right| < C\sqrt{\frac{\tau \vee |\log h|}{nh^{d+2|\alpha|}}}\right) \geq 1 - e^{-\tau}. \tag{4.1}$$

With $t_0 > 0$ given in Lemma 2.1, there exists a constant $C' > 0$ such that for all $h \in (0, t_0]$,

$$\sup_{x\in\mathcal{S}}\left|\mathbb{E}\partial^{(\alpha)}\widehat{f}(x) - \partial^{(\alpha)}f(x)\right| \leq C'h^{\nu-|\alpha|}\phi(h), \tag{4.2}$$

where $\phi(h) = 1$ when $|\alpha| = 0$ and $\phi(h) = \psi_f^{(\nu)}(h)$ when $|\alpha| = 1, 2$.

Proof. The proof of (4.1) follows similar arguments given in the proof of Proposition A.5 in [56]. In particular, notice that under the assumption (K), for $\alpha \in \mathbb{N}^d$ and $|\alpha| \leq 2$, $\partial^{(\alpha)}K$ is of bounded variation and the classes

$$\mathcal{G}_\alpha = \left\{x \mapsto \partial^{(\alpha)}K_h(y - x) : y \in \mathcal{S}\right\}, \quad |\alpha| = 0, 1, 2,$$

are of VC-type (see [25]) and the functions in these classes are uniformly bounded. Using Theorems A.1 and A.2 in [56], we can show that there exists a constant $C'' > 0$ such that for all $n \geq 1$, $h > 0$, and $\tau > 0$, with $\zeta_{n,h} = \frac{C''}{nh^{d+2|\alpha|}}\log\frac{C''}{h}$,

$$\mathbb{P}\left(\sup_{x\in\mathcal{S}}\left|\partial^{(\alpha)}\widehat{f}(x) - \mathbb{E}\partial^{(\alpha)}\widehat{f}(x)\right| < \zeta_{n,h} + \sqrt{\zeta_{n,h}} + \frac{\tau C''}{nh^{d+2|\alpha|}} + \frac{C''\sqrt{\tau}}{\sqrt{nh^{d+2|\alpha|}}}\right) \geq 1 - e^{-\tau}.$$

Then (4.1) is a consequence under the given conditions for n, h and τ in this lemma.

To see (4.2), notice that for $\alpha \in \mathbb{N}^d$ and $|\alpha| \leq 2$, using change of variable $z = (x - y)/h$ and integration by part we get for all $x \in \mathcal{S}$

$$\begin{aligned}\mathbb{E}\partial^{(\alpha)}\widehat{f}(x) - \partial^{(\alpha)}f(x) &= \frac{1}{h^{d+|\alpha|}}\int_{\mathbb{R}^d}\partial^{(\alpha)}K_h(x - y)f(y)\,dy - \partial^{(\alpha)}f(x)\\ &= \int_{\mathbb{R}^d}K(z)\left[\partial^{(\alpha)}f(x - hz) - \partial^{(\alpha)}f(x)\right]dz. \end{aligned} \tag{4.3}$$

With the smoothness condition of f in assumption (F), we have the Taylor expansion

$$\begin{aligned}&\partial^{(\alpha)}f(x - hz) - \partial^{(\alpha)}f(x)\\ &\quad= \sum_{j=1}^{\nu-|\alpha|-1}\frac{1}{j!}(-h)^j\left(z^{\otimes j}\right)^T\nabla^{\otimes j}\partial^{(\alpha)}f(x)\\ &\qquad+ \frac{1}{(\nu-|\alpha|)!}(-h)^{\nu-|\alpha|}\left(z^{\otimes(\nu-|\alpha|)}\right)^T\int_0^1\nabla^{\otimes(\nu-|\alpha|)}\partial^{(\alpha)}f(x - ahz)\,da.\end{aligned}$$

Plugging this into (4.3) and using the definition of νth order kernel functions, we have

$$\begin{aligned}&\mathbb{E}\partial^{(\alpha)}\widehat{f}(x) - \partial^{(\alpha)}f(x)\\ &= \frac{1}{(\nu-|\alpha|)!}(-h)^{\nu-|\alpha|}\int_{\mathbb{R}^d}K(z)\left(z^{\otimes(\nu-|\alpha|)}\right)^T\int_0^1\nabla^{\otimes(\nu-|\alpha|)}\partial^{(\alpha)}f(x - ahz)\,da\,dz.\end{aligned}$$

When $|\alpha| = 0$, noticing that ν is an even number, we have

$$\mathbb{E}\widehat{f}(x) - f(x) = \frac{1}{\nu!} h^{\nu} \int_{\mathbb{R}^d} K(z)\big(z^{\otimes\nu}\big)^T dz \nabla^{\otimes\nu} f(x) + r_h(x), \tag{4.4}$$

where $r_h(x) = \frac{1}{\nu!} h^{\nu} \int_{\mathbb{R}^d} K(z)(z^{\otimes\nu})^T \int_0^1 [\nabla^{\otimes\nu} f(x - ahz) - \nabla^{\otimes\nu} f(x)]\, da\, dz$. It follows that

$$\big|r_h(x)\big| \le \frac{2^{\nu}}{\nu!} \|K\|_1 h^{\nu} \psi_f^{(\nu)}(h), \tag{4.5}$$

using the assumption that $\mathrm{supp}(K) \subset \mathcal{B}(0, 1)$ and f has continuous νth partial derivatives on $\mathcal{I}(2\delta_0)$. This leads to (4.2) when $|\alpha| = 0$. When $|\alpha| = 1, 2$,

$$\begin{aligned}
&\mathbb{E}\partial^{(\alpha)}\widehat{f}(x) - \partial^{(\alpha)} f(x) \\
&= \frac{1}{(\nu - |\alpha|)!} (-h)^{\nu - |\alpha|} \\
&\quad \times \int_{\mathbb{R}^d} \int_0^1 \big[\nabla^{\otimes(\nu-|\alpha|)} \partial^{(\alpha)} f(x - ahz) - \nabla^{\otimes(\nu-|\alpha|)} \partial^{(\alpha)} f(x)\big]^T da \big(z^{\otimes(\nu-|\alpha|)}\big) \\
&\quad \times K(z)\, dz.
\end{aligned} \tag{4.6}$$

Since $\mathrm{supp}(K) \subset \mathcal{B}(0, 1)$, from (4.6) we have that with $c_K := \int_{\mathbb{R}^d} \|z^{\otimes(\nu-|\alpha|)}\| |K(z)|\, dz$,

$$\begin{aligned}
&\big|\mathbb{E}\partial^{(\alpha)}\widehat{f}(x) - \partial^{(\alpha)} f(x)\big| \\
&\le \frac{c_K}{(\nu - |\alpha|)!} h^{\nu-|\alpha|} \sup_{z \in \mathcal{B}(0,1)} \left\| \int_0^1 \nabla^{\otimes(\nu-|\alpha|)} \partial^{(\alpha)} f(x - ahz)\, da - \nabla^{\otimes(\nu-|\alpha|)} \partial^{(\alpha)} f(x) \right\| \\
&\le \frac{c_K 2^{\nu-|\alpha|}}{(\nu - |\alpha|)!} h^{\nu-|\alpha|} \psi_f^{(\nu)}(h).
\end{aligned}$$

Then we get (4.2) when $|\alpha| = 1, 2$. □

Proof of Lemma 2.1. Define functions $u(x, t) = |f(\zeta_x(t)) - f(x)|$ and $q(t) = \sup_{x \in \mathcal{I}(\delta_0)} u(x, t)$. Since u is a continuous function on $\mathcal{I}(2\delta_0) \times \mathbb{R}$, and $\mathcal{I}(\delta_0)$ is a compact set, q is a continuous function on $\mathbb{R}$. Let $t_0 = \sup\{s > 0 : \sup_{t \in [-s,s]} q(s) \le \delta_0\}$. Since $q(0) = 0$, we have $t_0 > 0$ and $(\mathcal{I}(\delta_0) \oplus t_0) \subset \mathcal{I}(2\delta_0)$. This is assertion (i).

To show assertion (ii), let us first recall a useful result in [23]. For any twice differentiable function η on $\mathbb{R}^d$, let $A_\eta = \{x : \eta(x) = 0\}$. For $\varepsilon > 0$, define

$$\iota_\eta(\varepsilon) = \frac{\varepsilon}{2} \wedge \frac{\inf_{x \in A_\eta \oplus \varepsilon} \|\nabla \eta(x)\|}{\sup_{A_\eta \oplus (2\varepsilon)} \|\nabla^2 \eta(x)\|_F}. \tag{4.7}$$

The proof of Lemma 4.11 in [23] shows that $\rho(A_\eta) \ge \iota_\eta(\varepsilon)$ for all $\varepsilon > 0$ such that $\iota_\eta(\varepsilon)$ is well defined and positive. Hence under assumption (F), we have

$$\inf_{\tau \in [c-\delta_0, c+\delta_0]} \rho(\mathcal{M}_\tau) \ge \frac{t_0}{4} \wedge \frac{\varepsilon_0}{Q_0} =: r_0 > 0. \tag{4.8}$$

Next, we prove assertion (iii). For any $x \in \mathcal{I}(\delta_0)$ and $t \in (0, r_0]$, we have the Taylor expansion

$$f\big(\zeta_x(t)\big) = f(x) + \int_0^t \big[\nabla f\big(\zeta_x(s)\big)\big]^T N(x)\,ds = f(x) + \int_0^t \big[\|\nabla f(x)\| + v(x,s)\big]\,ds,$$

where $v(x,s) = \int_0^s [N(x)]^T \nabla^2 f(\zeta_x(v)) N(x)\,dv$. Note that $|v(x,s)| \le s Q_0$ for all $s \in [0, r_0]$ and $x \in \mathcal{I}(\delta_0)$. Therefore for all $x \in \mathcal{I}(\delta_0)$ and $t \in (0, r_0]$,

$$\big|f\big(\zeta_x(t)\big) - f(x)\big| \ge \varepsilon_0 t - \frac{1}{2} Q_0 t^2 \ge \frac{1}{2}\varepsilon_0 t. \tag{4.9}$$

Let $\delta_1 = \frac{1}{2}\varepsilon_0 r_0$. For any $\tau_1 \in [c - \delta_0, c + \delta_0]$ and $|\tau_2 - \tau_1| \le \delta_1$, it follows from (4.9) that $\sup_{x \in \mathcal{M}_{\tau_2}} d(x, \mathcal{M}_{\tau_1}) \le \frac{2}{\varepsilon_0}|\tau_2 - \tau_1|$. Similarly we can also show that $\sup_{x \in \mathcal{M}_{\tau_1}} d(x, \mathcal{M}_{\tau_2}) \le \frac{2}{\varepsilon_0}|\tau_2 - \tau_1|$. Hence,

$$d_H(\mathcal{M}_{\tau_1}, \mathcal{M}_{\tau_2}) \le \frac{2}{\varepsilon_0}|\tau_2 - \tau_1|. \tag{4.10}$$

Then applying the argument in the proof of Theorem 2 in [20], we get that with $\eta_0 = 12/\varepsilon_0$,

$$\sup_{\tau \in [c-\delta_0, c+\delta_0]} d_H(\widehat{\mathcal{M}}_\tau, \mathcal{M}_\tau) \le \eta_0 \beta_n^{(0)}, \tag{4.11}$$

when $\beta_n^{(0)} \le \delta_1$. Therefore when $\beta_n^{(0)} \le \delta_1 \wedge \frac{t_0}{2\eta_0}$, we have

$$\widehat{\mathcal{I}}(\delta_0) \oplus \frac{t_0}{2} \subset \mathcal{I}(2\delta_0), \tag{4.12}$$

where $\widehat{\mathcal{I}}(\delta_0) = \widehat{f}^{-1}([c - \delta_0, c + \delta_0])$. Furthermore, using (4.7) again, when $\beta_n^{(1)} \le \frac{1}{2}\varepsilon_0$ and $\beta_n^{(2)} \le \frac{1}{2} Q_0$, we have

$$\inf_{\tau \in [c-\delta_0, c+\delta_0]} \rho(\widehat{\mathcal{M}}_\tau) \ge \frac{t_0}{8} \wedge \frac{\inf_{x \in \mathcal{I}(2\delta_0)} \|\nabla \widehat{f}(x)\|}{\sup_{x \in \mathcal{I}(2\delta_0)} \|\nabla^2 \widehat{f}(x)\|_F} \ge \frac{t_0}{8} \wedge \frac{\varepsilon_0}{3Q_0} =: r_1 > 0. \tag{4.13}$$

Therefore using (4.8), (4.11) and (4.13), when $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \le C_0 := \min(\delta_1 \wedge \frac{2-\sqrt{2}}{\eta_0} r_1, \frac{1}{2}\varepsilon_0, \frac{1}{2} Q_0)$, we have

$$\sup_{\tau \in [c-\delta_0, c+\delta_0]} d_H(\widehat{\mathcal{M}}_\tau, \mathcal{M}_\tau) \le (2 - \sqrt{2}) \inf_{\tau \in [c-\delta_0, c+\delta_0]} \min\big(\rho(\widehat{\mathcal{M}}_\tau), \rho(\mathcal{M}_\tau)\big),$$

which by Theorem 1 in [16] implies that $\widehat{\mathcal{M}}_\tau$ and $\mathcal{M}_\tau$ are normal compatible for all $\tau \in [c - \delta_0, c + \delta_0]$. The proof is now completed. □

Proof of Lemma 2.2. The fact that $|t_n(x)| = \|P_n(x) - x\|$ and (4.11) imply

$$\sup_{x \in \mathcal{I}(\delta_0)} \big|t_n(x)\big| \le \sup_{\tau \in [c-\delta_0, c+\delta_0]} d_H(\widehat{\mathcal{M}}_\tau, \mathcal{M}_\tau) \le \eta_0 \beta_n^{(0)}. \tag{4.14}$$

Since $\widehat{f}(P_n(x)) - f(x) = 0$, using Taylor expansion for $\widehat{f}(P_n(x))$ we obtain

$$0 = \widehat{f}\big(x + t_n(x) \boldsymbol{N}(x)\big) - f(x) = \widehat{f}(x) - f(x) + t_n(x) \boldsymbol{N}(x)^T \nabla \widehat{f}(x) + e_n(x), \tag{4.15}$$

where $e_n(x) = \frac{1}{2} t_n(x)^2 \boldsymbol{N}(x)^T \nabla^2 \widehat{f}(x + s_1 t_n(x)\boldsymbol{N}(x))\boldsymbol{N}(x)$, for some $0 < s_1 < 1$. Plugging $\nabla \widehat{f}(x) = \nabla f(x) - [\nabla f(x) - \nabla \widehat{f}(x)]$ into (4.15), we have

$$t_n(x) = \left\|\nabla f(x)\right\|^{-1}\left[f(x) - \widehat{f}(x)\right] + \delta_n(x), \tag{4.16}$$

where

$$\delta_n(x) = \left\|\nabla f(x)\right\|^{-1} \boldsymbol{N}(x)^T \left[\nabla f(x) - \nabla \widehat{f}(x)\right] t_n(x) - \left\|\nabla f(x)\right\|^{-1} e_n(x).$$

Under assumption (F), using (4.14) and $\beta_n^{(2)} \le \frac{1}{2} Q_0$, we then have

$$\sup_{x\in\mathcal{I}(\delta_0)} \left|\delta_n(x)\right| \le \frac{\eta_0}{\varepsilon_0}\beta_n^{(0)}\beta_n^{(1)} + \frac{\eta_0^2}{2\varepsilon_0}\left(\beta_n^{(0)}\right)^2\left(Q_0 + \beta_n^{(2)}\right), \tag{4.17}$$

which gives (2.11). □

Proof of Theorem 2.4. Note that with $\xi_i = \frac{1}{h^d}\int_{\mathcal{M}} w(x)K_h(x - X_i)\, d\mathscr{H}(x)$, we can write $\lambda(f, w \times (\widehat{f} - \mathbb{E}\widehat{f})) = \frac{1}{n}\sum_{i=1}^n [\xi_i - \mathbb{E}\xi_i]$. Suppose that $h \le \frac{3}{16}\rho(\mathcal{M}) =: h_0$. Since $\text{supp}(K) \subset \mathcal{B}(0, 1)$, for any $1 \le i \le n$, if $d(X_i, \mathcal{M}) > h$, then $\xi_i = 0$. Otherwise when $d(X_i, \mathcal{M}) \le h$, by the definition of reach, there exists a unique point $x_i \in \mathcal{M}$ such that $\|x_i - X_i\| = d(X_i, \mathcal{M})$. Using elementary geometry, $\|x - X_i\| \ge h$ for all $x \in \mathcal{M} \setminus \mathcal{B}(x_i, 2h)$. Hence,

$$|\xi_i| \le \frac{1}{h^d}\int_{\mathcal{M}\cap\mathcal{B}(x_i,2h)} \left|K_h(x - X_i)\right| d\mathscr{H}(x) \le \frac{1}{h^d}\|K\|_\infty \mathscr{H}\left(\mathcal{M}\cap\mathcal{B}(x_i, 2h)\right).$$

By Corollary 1.3 in [15], there exists a constant $A > 0$ that only depends on $(d - 1)$ and $\rho(\mathcal{M})$ such that $\mathscr{H}(\mathcal{M}\cap\mathcal{B}(x_i, 2h)) \le A2^{d-1}h^{d-1}$. Hence, $|\xi_i| \le \frac{1}{h}A2^{d-1}\|K\|_\infty$.

Next, we will calculate the variance of ξ_i. We have

$$\begin{aligned}
\mathbb{E}\left(\xi_i^2\right) &= \frac{1}{h^{2d}}\int_{\mathbb{R}^d}\int_{\mathcal{M}} w(x)K\left(\frac{x-z}{h}\right) d\mathscr{H}(x)\int_{\mathcal{M}} w(y)K\left(\frac{y-z}{h}\right) d\mathscr{H}(y) f(z)\, dz \\
&= \frac{1}{h^d}\int_{\mathcal{M}}\int_{\mathcal{M}}\int_{\mathbb{R}^d} w(x)K(u)w(y)K\left(\frac{y-x}{h} + u\right) f(x - hu)\, du\, d\mathscr{H}(y)\, d\mathscr{H}(x),
\end{aligned}$$

where we have used the variable transformation $u = h^{-1}(x - z)$. Then we can write $\mathbb{E}(\xi_i^2) = \frac{1}{h^d}\int_{\mathcal{M}} U_G(x)\mathscr{H}(x)$, where

$$U_G(x) = \int_{\mathcal{M}\cap\mathcal{B}(x,2h)}\int_{\mathbb{R}^d} w(x)K(u)w(y)K\left(\frac{y-x}{h} + u\right) f(x - hu)\, du\, d\mathscr{H}(y), \quad x \in \mathcal{M}.$$

We have $|U_G(x)| \le A2^{2d-1}h^{d-1}\|w\|_\infty^2\|K\|_\infty^2\|f\|_\infty$ for all $x \in \mathcal{M}$ and hence $\text{Var}(\xi_i) \le \mathbb{E}(\xi_i^2) \le \frac{1}{h}A2^{2d-1}\|w\|_\infty^2\|K\|_\infty^2\|f\|_\infty\mathscr{H}(\mathcal{M})$. By Bernstein's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\lambda\left(f, w \times (\widehat{f} - \mathbb{E}\widehat{f})\right)\right| \ge \varepsilon\right) \le 2\exp\left\{-\frac{nh\varepsilon^2}{A2^{2d}(\|w\|_\infty^2\|K\|_\infty^2\|f\|_\infty\mathscr{H}(\mathcal{M}) + \|K\|_\infty\varepsilon/3)}\right\}.$$

Then (2.13) immediately follows.

Next, we assume $nh \to \infty$ and $h \to 0$ and will show (2.14). Let $\mathcal{M} \ominus x = \{y - x : y \in \mathcal{M}\}$, that is, the manifold obtained by shifting $\mathcal{M}$ so that x becomes the origin. Then another variable transformation leads to

$$U_G(x) = \int_{(\mathcal{M}\ominus x)\cap\mathcal{B}(0,2h)} \int_{\mathbb{R}^d} w(x)K(u)w(x+y)K\left(\frac{y}{h}+u\right) f(x-hu)\,du\,d\mathscr{H}(y).$$

Without loss of generality, we may assume that $\mathbf{N}(x) = (0, \ldots, 0, 1)'$ and $T_x(\mathcal{M}) = \mathbb{R}^{d-1} \times \{0\}$. When h is small enough, $(\mathcal{M} \ominus x) \cap \mathcal{B}(0, 2h)$ coincides with the graph $T_x(\mathcal{M}) \ni y = (y_1, \ldots, y_{d-1}, 0) \mapsto \phi(y) := (y_1, \ldots, y_{d-1}, p(y)) \in \mathbb{R}^d$ with a smooth function $\phi(y)$, where for $y \in \mathcal{T}_{x,h} := [T_x(\mathcal{M}) \cap \mathcal{B}(0, 2h)]$, p has a quadratic approximation (see page 141, [35])

$$p(y) = \frac{1}{2}\sum_{i=1}^{d-1} \kappa_i(x)\langle y, p_i(x)\rangle^2 + O\left(h^3\right), \tag{4.18}$$

where $\kappa_i(x)$, $i = 1, \ldots, d-1$ are the principal curvatures of $\mathcal{M}$ at x, and $p_i(x)$, $i = 1, \ldots, d-1$ are the corresponding principal directions. Both $\kappa_i(x)$ and $p_i(x)$ can be obtained as the eigenvalues and eigenvectors of the shape operator $S(x)$ in (2.3). Note that the rate in the $O(h^3)$-term in (4.18) is uniform over $\mathcal{M}$. From (4.18) we have

$$\sup_{y\in\mathcal{T}_{x,h}} \left\|\phi(y) - y\right\| = O\left(h^2\right). \tag{4.19}$$

Then ϕ defines a diffeomorphism between $\mathcal{T}_{x,h}$ and its image under ϕ. The Jacobian determinant of ϕ is

$$J_\phi(y) = \sqrt{1 + \left(\frac{\partial}{\partial y_1}p(y)\right)^2 + \cdots + \left(\frac{\partial}{\partial y_{d-1}}p(y)\right)^2} = 1 + O(h), \tag{4.20}$$

uniformly in $\phi(\mathcal{T}_{x,h})$, where we have used (4.18). Notice that $(\mathcal{M} \ominus x) \cap \mathcal{B}(0, 2h) \subseteq \phi(\mathcal{T}_{x,h})$, and the $U_G(x)$ remains unchanged if its domain of integration is changed from $(\mathcal{M}\ominus x)\cap\mathcal{B}(0, 2h)$ to $\phi(\mathcal{T}_{x,h})$, because $\mathrm{supp}(K) \subset \mathcal{B}(0, 1)$. In view of this and by using (4.19) and (4.20), we get

$$\begin{aligned} U_G(x) &= \int_{\mathcal{T}_{x,h}} \int_{\mathbb{R}^d} w(x)K(u)w\left(x+\phi(y)\right)K\left(\frac{\phi(y)}{h}+u\right) f(x-hu)\,du\,J_\phi(y)\,d\mathscr{H}(y) \\ &= \int_{\mathcal{T}_{x,h}} \int_{\mathbb{R}^d} w(x)K(u)w(x+y)K\left(\frac{y}{h}+u\right) f(x-hu)\,du\,d\mathscr{H}(y)\left(1+o(1)\right), \end{aligned}$$

where the little o is uniform in $x \in \mathcal{M}$. Hence,

$$\begin{aligned} \mathbb{E}\left(\xi_i^2\right) &= \frac{1}{h^d}\int_{\mathcal{M}}\int_{\mathcal{T}_{x,h}}\int_{\mathbb{R}^d} w(x)K(u)w(x+y)K\left(\frac{y}{h}+u\right) \\ &\quad\times f(x-hu)\,du\,d\mathscr{H}(y)\,d\mathscr{H}(x)\left(1+o(1)\right) \\ &= \frac{1}{h}\int_{\mathcal{M}}\int_{T_x(\mathcal{M})}\int_{\mathbb{R}^d} w(x)K(u)w(x+hv)K(v+u) \\ &\quad\times f(x)\,du\,d\mathscr{H}(v)\,d\mathscr{H}(x)\left(1+o(1)\right) \\ &= \frac{1}{h}c\rho(w, K, f)\left(1+o(1)\right), \end{aligned} \tag{4.21}$$

where $\rho(w,K,f)=\int_{\mathcal{M}} w(x)^2\int_{T_x(\mathcal{M})}\int_{\mathbb{R}^d} K(u)K(v+u)\,du\,d\mathscr{H}(v)\mathscr{H}(x)$. Notice that we have $\mathbb{R}^d=\bigcup_{t\in\mathbb{R}} T_{x,t}(\mathcal{M})$ for any $x\in\mathcal{M}$. So we can further write

$$\begin{aligned}\rho(w,K,f)&=\int_{\mathcal{M}} w(x)^2\int_{T_x(\mathcal{M})}\int_{\mathbb{R}}\int_{T_{x,t}(\mathcal{M})} K(u)K(v+u)\,d\mathscr{H}(u)dtd\mathscr{H}(v)\mathscr{H}(x)\\&=\int_{\mathcal{M}} w(x)^2\int_{\mathbb{R}}\int_{T_{x,t}(\mathcal{M})} K(u)\int_{T_x(\mathcal{M})} K(v+u)\,d\mathscr{H}(v)\,d\mathscr{H}(u)\,dt\mathscr{H}(x)\\&=\int_{\mathcal{M}} w(x)^2\int_{\mathbb{R}}\Big(\int_{T_{x,t}(\mathcal{M})} K(u)\,d\mathscr{H}(u)\Big)^2 dt\mathscr{H}(x),\end{aligned}\tag{4.22}$$

where the last step holds because we use change of variables $z=u+v$ and $z\in T_{x,t}(\mathcal{M})$ when $u\in T_{x,t}(\mathcal{M})$ and $v\in T_x(\mathcal{M})$. Similarly, using change of variables and Taylor expansion we get

$$\begin{aligned}\mathbb{E}(\xi_i)&=\frac{1}{h^d}\int_{\mathcal{M}}\int_{\mathbb{R}^d} w(x)K\Big(\frac{x-z}{h}\Big)f(z)\,dz\,d\mathscr{H}(x)\\&=\int_{\mathcal{M}}\int_{\mathbb{R}^d} w(x)K(u)f(x-hu)\,du\mathscr{H}(x)\\&=\int_{\mathcal{M}}\int_{\mathbb{R}^d} w(x)K(u)f(x)\,du\,d\mathscr{H}(x)\big(1+o(1)\big)\\&=c\int_{\mathcal{M}} w(x)\mathscr{H}(x)\big(1+o(1)\big).\end{aligned}\tag{4.23}$$

From (4.21) and (4.23) we have $\operatorname{Var}(\xi_i)=h^{-1}c\rho(w,K,f)\{1+o(1)\}$ and therefore

$$\operatorname{Var}\big(\lambda\big(f,w\times(\widehat{f}-\mathbb{E}\widehat{f})\big)\big)=\frac{\operatorname{Var}(\xi_i)}{n}=\frac{1}{nh}c\rho(w,K,f)\big(1+o(1)\big).$$

It only remains to show the asymptotic normality of $\frac{1}{n}\sum_{i=1}^n(\xi_i-\mathbb{E}\xi_i)$. Note that following a similar procedure of obtaining (4.21), we can show that $\mathbb{E}(|\xi_i|^3)=O(\frac{1}{h^{3d}}h^{3d-2})=O(\frac{1}{h^2})$. Then for the third absolute central moment of ξ_i, we have $\mathbb{E}[|\xi_i-\mathbb{E}(\xi_i)|^3]\le 8\mathbb{E}(|\xi_i|^3)=O(\frac{1}{h^2})$. Therefore,

$$\frac{\{\sum_{i=1}^n\mathbb{E}[|\xi_i-\mathbb{E}(\xi_i)|^3]\}^{1/3}}{[\sum_{i=1}^n\operatorname{Var}(\xi_i)]^{1/2}}=O\Big(\frac{(n/h^2)^{1/3}}{(n/h)^{1/2}}\Big)=O\Big(\frac{1}{n^{1/6}h^{1/6}}\Big)=o(1).$$

With the Liapunov condition satisfied, the asymptotic normality in (2.14) is verified. □

Proof of Theorem 2.3. It is known that $\mathcal{M}$ is a compact $(d-1)$-dimensional submanifold embedded in $\mathbb{R}^d$ with assumption (F) (see Theorem 2 in [61]). It admits an atlas $\{(U_\alpha,\psi_\alpha):\alpha\in\mathscr{A}\}$ indexed by a finite set $\mathscr{A}$, where $\{U_\alpha:\alpha\in\mathscr{A}\}$ is an open cover of $\mathcal{M}$, and for an open set $\Omega_\alpha\subset\mathbb{R}^{d-1}$, $\psi_\alpha:\Omega_\alpha\to U_\alpha$ is a diffeomorphism. Let B_α be the Jacobian matrix of ψ_α. By Lemma 2.1, $P_n(x)=x+t_n(x)\boldsymbol{N}(x)$ in (2.9) is a diffeomorphism between $\mathcal{M}_\tau$ and $\widehat{\mathcal{M}}_\tau$ for $|\tau-c|\le\delta_0$, when $\max(\beta_n^{(0)},\beta_n^{(1)},\beta_n^{(2)})\le C_0$. Let $\widehat{U}_\alpha=\{P_n(x):x\in U_\alpha\}$. Then $\{(\widehat{U}_\alpha,P_n\circ\psi_\alpha):\alpha\in\mathscr{A}\}$ is a finite atlas for $\widehat{\mathcal{M}}$. Let A_n be the Jacobian matrix of P_n, that is, $A_n(x)=\nabla P_n(x)$. By the chain rule, the Jabobian matrix of $P_n\circ\psi_\alpha:\Omega_\alpha\to\widehat{U}$ is given by $(A_n\circ\psi_\alpha)B_\alpha$. The area formula on manifolds (cf. page 117, [22]) leads to

$$\int_{\widehat{U}_\alpha} g(x)\,d\mathscr{H}(x)=\int_{\Omega_\alpha} g\big(P_n\big(\psi_\alpha(\theta)\big)\big)J_{n,\alpha}(\theta)\,d\theta,\tag{4.24}$$

where $J_{n,\alpha}(\theta) = \{\det[B_\alpha(\theta)^T A_n(\psi_\alpha(\theta))^T A_n(\psi_\alpha(\theta))B_\alpha(\theta)]\}^{1/2}$. There exits a partition of unity (see page 52, [28]) of $\widehat{\mathcal{M}}$ subordinate to $\{\widehat{U}_\alpha\}$ denoted by $\{\rho_\alpha : \alpha \in \mathscr{A}\}$, that is, $\mathrm{supp}(\rho_\alpha) \subset \widehat{U}_\alpha$, $\sum_{\alpha\in\mathscr{A}} \rho_\alpha(x) = 1$ for $x \in \widehat{\mathcal{M}}$ and there exist only finitely many α such that $\rho_\alpha(x) \neq 0$ for $x \in \widehat{\mathcal{M}}$. Using (4.24), we have

$$\lambda(\widehat{f}, g) = \sum_{\alpha\in\mathscr{A}} \int_{\widehat{U}_\alpha} \rho_\alpha(x)g(x)\, d\mathscr{H}(x) = \sum_{\alpha\in\mathscr{A}} \int_{\Omega_\alpha} (\rho_\alpha g)\big(P_n\big(\psi_\alpha(\theta)\big)\big) J_{n,\alpha}(\theta)d\theta. \tag{4.25}$$

Recall that $\kappa_i(x)$, $i = 1, \dots, d-1$ are the principal curvatures of $\mathcal{M}$ at x, and $p_i(x)$, $i = 1, \dots, d-1$ are the corresponding principal directions, which form an orthonormal basis of $T_x(\mathcal{M})$. Let $\widetilde{P}(x) = [p_1(x), \dots, p_{d-1}(x)]$. Notice that $B_\alpha(\theta) \in T_x(\mathcal{M})$, for $\theta \in \Omega_\alpha$, because $\boldsymbol{N}(\psi_\alpha(\theta))^T B_\alpha(\theta) = 0$, which can be seen by taking derivatives of the equation $f(\psi_\alpha(\theta)) = c$. For $B_\alpha(\theta)$, there exists a $(d-1) \times (d-1)$ matrix $L_\alpha(\theta)$ such that $B_\alpha(\theta) = \widetilde{P}(\psi_\alpha(\theta))L_\alpha(\theta)$. Notice that $[B_\alpha(\theta)]^T B_\alpha(\theta) = [L_\alpha(\theta)]^T L_\alpha(\theta)$. Therefore

$$\begin{aligned}
J_{n,\alpha}(\theta) &= \Big\{\det\Big[L_\alpha(\theta)^T \widetilde{P}\big(\psi_\alpha(\theta)\big)^T A_n\big(\psi_\alpha(\theta)\big)^T A_n\big(\psi_\alpha(\theta)\big)\widetilde{P}\big(\psi_\alpha(\theta)\big)L_\alpha(\theta)\Big]\Big\}^{1/2} \\
&= \Big\{\det\Big[L_\alpha(\theta)^T L_\alpha(\theta)\Big]\Big\}^{1/2}\Big\{\det\Big[\widetilde{P}\big(\psi_\alpha(\theta)\big)^T A_n\big(\psi_\alpha(\theta)\big)^T A_n\big(\psi_\alpha(\theta)\big)\widetilde{P}\big(\psi_\alpha(\theta)\big)\Big]\Big\}^{1/2} \\
&= \Big\{\det\Big[B_\alpha(\theta)^T B_\alpha(\theta)\Big]\Big\}^{1/2}\Big\{\det\Big[\widetilde{P}\big(\psi_\alpha(\theta)\big)^T A_n\big(\psi_\alpha(\theta)\big)^T A_n\big(\psi_\alpha(\theta)\big)\widetilde{P}\big(\psi_\alpha(\theta)\big)\Big]\Big\}^{1/2}.
\end{aligned}$$

Plugging this into (4.25) and using the area formula again, we get

$$\begin{aligned}
\lambda(\widehat{f}, g) &= \sum_{\alpha\in\mathscr{A}} \int_{U_\alpha} (\rho_\alpha g)\big(P_n(x)\big)\Big\{\det\Big[\widetilde{P}(x)^T A_n(x)^T A_n(x)\widetilde{P}(x)\Big]\Big\}^{1/2} d\mathscr{H}(x) \\
&= \int_{\mathcal{M}} g\big(P_n(x)\big)\Big\{\det\Big[\widetilde{P}(x)^T A_n(x)^T A_n(x)\widetilde{P}(x)\Big]\Big\}^{1/2} d\mathscr{H}(x).
\end{aligned} \tag{4.26}$$

For the rest of the proof, we will evaluate the right-hand side of (4.26). We can write $A_n(x) = \boldsymbol{I}_d + R_n(x)$, $x \in \mathcal{I}(\delta_0)$, where $R_n(x) = \nabla\boldsymbol{N}(x)t_n(x) + \boldsymbol{N}(x)[\nabla t_n(x)]^T$. Recall that $t_n(x) = \langle P_n(x) - x, \boldsymbol{N}(x)\rangle$ and hence

$$\begin{aligned}
\big[\nabla t_n(x)\big]^T &= \big[P_n(x) - x\big]^T \nabla\boldsymbol{N}(x) + \boldsymbol{N}(x)^T\big[\nabla P_n(x) - \boldsymbol{I}_d\big] \\
&= t_n(x)\boldsymbol{N}(x)^T\nabla\boldsymbol{N}(x) + \boldsymbol{N}(x)^T\big[A_n(x) - \boldsymbol{I}_d\big] \\
&= \boldsymbol{N}(x)^T\big[A_n(x) - \boldsymbol{I}_d\big],
\end{aligned}$$

where we have used the fact that $\boldsymbol{N}(x)^T\nabla\boldsymbol{N}(x) = 0$ by differentiating both sides of the equation $\langle\boldsymbol{N}(x), \boldsymbol{N}(x)\rangle = 1$. Overall we write

$$A_n(x) = \boldsymbol{I}_d + \nabla\boldsymbol{N}(x)t_n(x) + \boldsymbol{N}(x)\boldsymbol{N}(x)^T\big[A_n(x) - \boldsymbol{I}_d\big]. \tag{4.27}$$

Furthermore, by taking gradient on both sides of the equation $\widehat{f}(P_n(x)) = f(x)$, $x \in \mathcal{I}(\delta_0)$, we get $[\nabla\widehat{f}(P_n(x))]^T A_n(x) = [\nabla f(x)]^T$. Denote $\widetilde{\boldsymbol{N}}(x) = \|\nabla\widehat{f}(P_n(x))\|^{-1}\nabla f(x)$. Then we can write

$$\big[\widehat{\boldsymbol{N}}\big(P_n(x)\big)\big]^T A_n(x) = \big[\widetilde{\boldsymbol{N}}(x)\big]^T. \tag{4.28}$$

Since $\{p_1(x), \dots, p_{d-1}(x), \boldsymbol{N}(x)\}$ is an orthonormal basis of $\mathbb{R}^d$, we have the following representation of $\widehat{\boldsymbol{N}}(P_n(x))$:

$$\widehat{\boldsymbol{N}}\big(P_n(x)\big) = \sum_{i=1}^{d-1} \big\langle \widehat{\boldsymbol{N}}\big(P_n(x)\big), p_i(x) \big\rangle p_i(x) + \big\langle \widehat{\boldsymbol{N}}\big(P_n(x)\big), \boldsymbol{N}(x) \big\rangle \boldsymbol{N}(x). \tag{4.29}$$

Recall $G(x) = \boldsymbol{I}_d - \boldsymbol{N}(x)\boldsymbol{N}(x)^T$ and (2.3). Note that for any $u, v \in T_x(\mathcal{M})$, we have that

$$u^T S(x) v = u^T \nabla \boldsymbol{N}(x) v = u^T \nabla \boldsymbol{N}(x)^T v. \tag{4.30}$$

The above equality can be derived by observing that the matrix $G(x)$ is symmetric and has the property that $G(x)u = u$ and $u^T G(x) = u^T$, for any $u \in T_x(\mathcal{M})$. Hence, $[p_i(x)]^T \nabla \boldsymbol{N}(x) = [p_i(x)]^T S(x) = \kappa_i(x)[p_i(x)]^T$. Using (4.27) we obtain

$$\big[p_i(x)\big]^T A_n(x) = \big[p_i(x)\big]^T + \big[p_i(x)\big]^T \nabla \boldsymbol{N}(x) t_n(x) = \big[1 + t_n(x)\kappa_i(x)\big]\big[p_i(x)\big]^T. \tag{4.31}$$

Plugging (4.29) and (4.31) into (4.28), we have

$$\begin{aligned}
&\big[\widetilde{\boldsymbol{N}}(x)\big]^T \\
&= \sum_{i=1}^{d-1} \big\langle \widehat{\boldsymbol{N}}\big(P_n(x)\big), p_i(x) \big\rangle \big[p_i(x)\big]^T A_n(x) + \big\langle \widehat{\boldsymbol{N}}\big(P_n(x)\big), \boldsymbol{N}(x) \big\rangle \big[\boldsymbol{N}(x)\big]^T A_n(x) \\
&= \sum_{i=1}^{d-1} \big\langle \widehat{\boldsymbol{N}}\big(P_n(x)\big), p_i(x) \big\rangle \big[1 + t_n(x)\kappa_i(x)\big] \big[p_i(x)\big]^T \\
&\quad + \big\langle \widehat{\boldsymbol{N}}\big(P_n(x)\big), \boldsymbol{N}(x) \big\rangle \big[\boldsymbol{N}(x)\big]^T A_n(x).
\end{aligned} \tag{4.32}$$

Using Lemmas 4.1 and 2.2, we know with probability one $\langle \widehat{\boldsymbol{N}}(P_n(x)), \boldsymbol{N}(x) \rangle \neq 0$ for all $x \in \mathcal{I}(\delta_0)$ for large n and so

$$\big[\boldsymbol{N}(x)\big]^T A_n(x) = \frac{[\widetilde{\boldsymbol{N}}(x)]^T - \sum_{i=1}^{d-1} \langle \widehat{\boldsymbol{N}}(P_n(x)), p_i(x) \rangle [1 + t_n(x)\kappa_i(x)][p_i(x)]^T}{\langle \widehat{\boldsymbol{N}}(P_n(x)), \boldsymbol{N}(x) \rangle}. \tag{4.33}$$

Plugging this into (4.27), we then have

$$\begin{aligned}
&A_n(x) \\
&= \boldsymbol{I}_d + \nabla \boldsymbol{N}(x) t_n(x) - \boldsymbol{N}(x)\big[\boldsymbol{N}(x)\big]^T \\
&\quad + \frac{\boldsymbol{N}(x)[\widetilde{\boldsymbol{N}}(x)]^T - \sum_{i=1}^{d-1} \langle \widehat{\boldsymbol{N}}(P_n(x)), p_i(x) \rangle [1 + t_n(x)\kappa_i(x)] \boldsymbol{N}(x)[p_i(x)]^T}{\langle \widehat{\boldsymbol{N}}(P_n(x)), \boldsymbol{N}(x) \rangle}.
\end{aligned} \tag{4.34}$$

Again recall that for $i = 1, \dots, d-1$ we have $\nabla \boldsymbol{N}(x) p_i(x) = \nabla \boldsymbol{N}(x) G(x) p_i(x) = S(x) p_i(x) = \kappa_i(x) p_i(x)$, $[\boldsymbol{N}(x)]^T p_i(x) = 0$ and $[\widetilde{\boldsymbol{N}}(x)]^T p_i(x) = 0$. Also $[p_i(x)]^T p_j(x) = 0$ if $i \neq j$. With the new expression of $A_n(x)$ given in (4.34), we have for $j = 1, \dots, d-1$,

$$\begin{aligned}
A_n(x) p_j(x) &= \big[1 + \kappa_j(x) t_n(x)\big] p_j(x) \\
&\quad - \frac{\langle \widehat{\boldsymbol{N}}(P_n(x)), p_j(x) \rangle [1 + t_n(x)\kappa_j(x)]}{\langle \widehat{\boldsymbol{N}}(P_n(x)), \boldsymbol{N}(x) \rangle} \boldsymbol{N}(x).
\end{aligned} \tag{4.35}$$

Using (4.35), we have

$$\big[\widehat{\boldsymbol N}\big(P_n(x)\big)\big]^T A_n(x)p_j(x)$$
$$= \big[1+\kappa_j(x)t_n(x)\big]\big\langle \widehat{\boldsymbol N}\big(P_n(x)\big), p_j(x)\big\rangle - \big\langle \widehat{\boldsymbol N}\big(P_n(x)\big), p_j(x)\big\rangle\big[1+t_n(x)\kappa_j(x)\big] = 0,$$

which simply means that $A_n(x)$ as a Jacobian matrix converts any vector in the tangent space $T_x(\mathcal{M})$ to a vector in the tangent space $T_{P_n(x)}(\widehat{\mathcal{M}})$ (recall that $\{p_1(x),\dots,p_{d-1}\}$ is an orthonormal basis that spans $T_x(\mathcal{M})$ and $\widehat{\boldsymbol N}(P_n(x))$ is normal to $T_{P_n(x)}(\widehat{\mathcal{M}})$).

Now using (4.35), for any $1\le i, j\le d-1$, we have

$$\begin{aligned}&\big\langle A_n(x)p_i(x), A_n(x)p_j(x)\big\rangle\\ &\quad= \big[1+\kappa_i(x)t_n(x)\big]\big[1+\kappa_j(x)t_n(x)\big]\\ &\qquad\times\bigg[\big\langle p_i(x), p_j(x)\big\rangle + \frac{\langle \widehat{\boldsymbol N}(P_n(x)), p_i(x)\rangle\langle \widehat{\boldsymbol N}(P_n(x)), p_j(x)\rangle}{\langle \widehat{\boldsymbol N}(P_n(x)), \boldsymbol N(x)\rangle^2}\bigg].\end{aligned} \tag{4.36}$$

Let $\widetilde K(x)=\operatorname{diag}(\kappa_1(x),\dots,\kappa_{d-1}(x))$. Notice that $\langle p_i(x), p_j(x)\rangle=\delta_{ij}$, which is the Kronecker delta. Then the matrix form of (4.36) is given by

$$\begin{aligned}&\big[A_n(x)\widetilde P(x)\big]^T\big[A_n(x)\widetilde P(x)\big]\\ &\quad=\big[\boldsymbol I_{d-1}+t_n(x)\widetilde K(x)\big]\bigg\{\boldsymbol I_{d-1}+\frac{[\widehat{\boldsymbol N}(P_n(x))\widetilde P(x)]^T[\widehat{\boldsymbol N}(P_n(x))\widetilde P(x)]}{\langle \widehat{\boldsymbol N}(P_n(x)), \boldsymbol N(x)\rangle^2}\bigg\}\\ &\qquad\times\big[\boldsymbol I_{d-1}+t_n(x)\widetilde K(x)\big].\end{aligned} \tag{4.37}$$

Hence,

$$\begin{aligned}&\det\big\{\big[A_n(x)\widetilde P(x)\big]^T\big[A_n(x)\widetilde P(x)\big]\big\}\\ &\quad=\prod_{i=1}^{d-1}\big[1+t_n(x)\kappa_i(x)\big]^2\det\bigg\{\boldsymbol I_{d-1}+\frac{[\widehat{\boldsymbol N}(P_n(x))\widetilde P(x)]^T[\widehat{\boldsymbol N}(P_n(x))\widetilde P(x)]}{\langle \widehat{\boldsymbol N}(P_n(x)), \boldsymbol N(x)\rangle^2}\bigg\}\\ &\quad=\prod_{i=1}^{d-1}\big[1+t_n(x)\kappa_i(x)\big]^2\bigg\{1+\frac{\|\widehat{\boldsymbol N}(P_n(x))\widetilde P(x)\|^2}{\langle \widehat{\boldsymbol N}(P_n(x)), \boldsymbol N(x)\rangle^2}\bigg\}\\ &\quad=\frac{1}{\langle \widehat{\boldsymbol N}(P_n(x)), \boldsymbol N(x)\rangle^2}\prod_{i=1}^{d-1}\big[1+t_n(x)\kappa_i(x)\big]^2,\end{aligned} \tag{4.38}$$

where we have used Sylvester's determinant identity and the fact that

$$\sum_{i=1}^{d-1}\big\langle \widehat{\boldsymbol N}\big(P_n(x)\big), p_i(x)\big\rangle^2 + \big\langle \widehat{\boldsymbol N}\big(P_n(x)\big), \boldsymbol N(x)\big\rangle^2 = 1.$$

Plugging (4.38) into (4.26), we then get (2.12). □

Proof of Lemma 2.3. Due to (4.12), $\widehat{\mathcal{M}}_{c+\varepsilon}\subset\mathcal{I}(2\delta_0)$ for $|\varepsilon|\le\delta_0$ when $\beta_n^{(0)}\le\delta_1\wedge\frac{t_0}{2\eta_0}$. Recall $r_1>0$ given in (4.13). Similar to (4.9), when further $\beta_n^{(1)}\le\frac{1}{2}\varepsilon_0$ and $\beta_n^{(2)}\le\frac{1}{2}Q_0$, we have for $x\in\widehat{\mathcal{M}}$ and

$t \in [0, r_1]$,

$$\left|\widehat{f}\big(\widehat{\zeta}_x(t)\big) - \widehat{f}(x)\right| \geq \frac{1}{2}\varepsilon_0 t - \frac{3}{4}Q_0 t^2 \geq \frac{1}{4}\varepsilon_0 t. \tag{4.39}$$

Then similar to (4.10), for $|\varepsilon| \leq \delta_0 \wedge \frac{1}{4}\varepsilon_0 r_1$,

$$d_H(\widehat{\mathcal{M}}_c, \widehat{\mathcal{M}}_{c+\varepsilon}) \leq \frac{4}{\varepsilon_0}|\varepsilon| \leq \eta_0|\varepsilon|. \tag{4.40}$$

Hence with $r_2 = \delta_0 \wedge ((2-\sqrt{2})r_1/\eta_0)$, we have

$$\sup_{\varepsilon \in [-r_2, r_2]} d_H(\widehat{\mathcal{M}}_c, \widehat{\mathcal{M}}_{c+\varepsilon}) \leq (2-\sqrt{2})r_1,$$

which by Theorem 1 in [16] implies that $\widehat{\mathcal{M}}_c$ and $\widehat{\mathcal{M}}_{c+\varepsilon}$ are normal compatible for all $|\varepsilon| \leq r_2$. Therefore $|\widehat{t}_\varepsilon(x)| \leq d_H(\widehat{\mathcal{M}}_c, \widehat{\mathcal{M}}_{c+\varepsilon}) \leq \eta_0|\varepsilon|$ for all $x \in \widehat{f}^{-1}(c)$ and $|\varepsilon| \leq r_2$. Using Taylor expansion, we have

$$\begin{aligned} c + \varepsilon = \widehat{f}\big(P_\varepsilon(x)\big) &= c + \widehat{t}_\varepsilon(x)\big\langle \widehat{\boldsymbol{N}}(x), \nabla \widehat{f}(x)\big\rangle + \delta_\varepsilon(x) \\ &= c + \widehat{t}_\varepsilon(x)\big\|\nabla \widehat{f}(x)\big\| + \delta_\varepsilon(x), \end{aligned} \tag{4.41}$$

where $\delta_\varepsilon(x) = \frac{1}{2}\widehat{t}_\varepsilon(x)^2[\widehat{\boldsymbol{N}}(x)]^T \nabla^2 \widehat{f}(\widehat{\zeta}_x(s_1))\widehat{\boldsymbol{N}}(x)$ for some s_1 between 0 and $\widehat{t}_\varepsilon(x)$. Note that when $\beta_n^{(2)} \leq \frac{1}{2}Q_0$, we have $|\delta_\varepsilon(x)| \leq Q_0\eta_0^2\varepsilon^2$ for all $x \in \widehat{f}^{-1}(c)$. Also we can write $\widehat{t}_\varepsilon(x) = \|\nabla \widehat{f}(x)\|^{-1}(\varepsilon - \delta_\varepsilon(x))$ by (4.41). Plugging this into the expression of $\delta_\varepsilon(x)$, when $\beta_n^{(1)} \leq \frac{1}{2}\varepsilon_0$ and $\beta_n^{(2)} \leq \frac{1}{2}Q_0$ we obtain

$$\big\|\nabla \widehat{f}(x)\big\|^{-1}\delta_\varepsilon(x) = \frac{1}{2}\varepsilon^2\big\|\nabla \widehat{f}(x)\big\|^{-3}\big\|\widehat{\boldsymbol{N}}(x)\big\|^2_{\nabla^2 \widehat{f}(x)} + \eta_\varepsilon(x),$$

where

$$\begin{aligned} \big|\eta_\varepsilon(x)\big| &\leq \frac{8Q_0^2\eta_0^2}{\varepsilon_0^3}|\varepsilon|^3 + \frac{4Q_0^3\eta_0^4}{\varepsilon_0^3}\varepsilon^4 + \frac{4\varepsilon^2}{\varepsilon_0^3}\big(2\beta_n^{(2)} + \big\|\nabla^2 f\big(\widehat{\zeta}_x(s_1)\big) - \nabla^2 f(x)\big\|_F\big) \\ &\leq \frac{(8Q_0^2\eta_0^2 + 4Q_0^3\eta_0^4 r_2) \vee (4d)}{\varepsilon_0^3}\big(|\varepsilon|^3 + \varepsilon^2\beta_n^{(2)} + \varepsilon^2\psi_f^{(2)}\big(\eta_0|\varepsilon|\big)\big). \end{aligned}$$

Then immediately we get (2.18). □

To prove Theorem 2.5, we need the following proposition, where we denote $P_\varepsilon^*(x) = x + \varepsilon\widehat{\boldsymbol{N}}(x)$.

Proposition 4.1. *Let $g : \mathcal{I}(2\delta_0) \to \mathbb{R}$ be an integrable function. Under the assumption* (F), *when $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \leq C_0$ and $\varepsilon_n \in (0, r_1 \wedge r_2]$, where C_0 and r_1 are given in Lemma 2.1, and r_2 is given in Lemma 2.3, we have*

$$\lambda_{\varepsilon_n}^{(1)}(\widehat{f}, g) = \frac{1}{2\varepsilon_n}\int_{-\varepsilon_n}^{\varepsilon_n} \lambda\bigg(\widehat{f}, \frac{(g \circ P_\varepsilon) \times \prod_{i=1}^{d-1}[1 + \widehat{t}_\varepsilon\widehat{\kappa}_i]}{\langle \widehat{\boldsymbol{N}} \circ P_\varepsilon, \widehat{\boldsymbol{N}}\rangle}\bigg)\, d\varepsilon, \tag{4.42}$$

$$\lambda_{\varepsilon_n}^{(2)}(\widehat{f}, g) = \frac{1}{2\varepsilon_n}\int_{-\varepsilon_n}^{\varepsilon_n} \lambda\big(\widehat{f}, \big(g \circ P_\varepsilon^*\big) \times \det(\boldsymbol{I}_d + \varepsilon\widehat{S})\big)\, d\varepsilon. \tag{4.43}$$

Proof of Proposition 4.1. By Lemma 2.3, the map P_ε is a diffeomorphism between $\widehat{\mathcal{M}}_c$ and $\widehat{\mathcal{M}}_{c+\varepsilon}$ for $\varepsilon \in [-r_2, r_2]$. Then following a similar proof of Theorem 2.3, we have for $\varepsilon \in [-r_2, r_2]$,

$$\lambda_{c+\varepsilon}(\widehat{f}, g) = \int_{\widehat{\mathcal{M}}_c} \frac{g(P_\varepsilon(x)) \prod_{i=1}^{d-1} [1 + \widehat{t}_\varepsilon(x)\widehat{\kappa}_i(x)]}{\langle \widehat{\boldsymbol{N}}(P_\varepsilon(x)), \widehat{\boldsymbol{N}}(x)\rangle} \, d\mathscr{H}(x), \tag{4.44}$$

and therefore (4.42) follows by using (2.15).

The set $\widehat{\mathcal{M}}_c \oplus \varepsilon_n$ is tube of radius ε_n around the submanifold $\widehat{\mathcal{M}}_c$. Then $\widehat{\zeta}(x, \varepsilon) := \widehat{\zeta}_x(\varepsilon)$ defines a diffeomorphism between $\widehat{\mathcal{M}}_c \times [-\varepsilon_n, \varepsilon_n]$ and $\widehat{\mathcal{M}}_c \oplus \varepsilon_n$ for $\varepsilon_n \in (0, r_1]$. The Weyl's volume element in the tube given in [62] (also see [27]) is $\det(\boldsymbol{I}_d + \varepsilon \widehat{S}(x))\, d\varepsilon\, d\mathscr{H}(x)$, where $\widehat{S}(x)$ is the shape operator on $\widehat{\mathcal{M}}$ as defined in (2.3). Therefore,

$$\lambda_{\varepsilon_n}^{(2)}(\widehat{f}, g) = \frac{1}{2\varepsilon_n} \int_{\widehat{\mathcal{M}}_c} \int_{-\varepsilon_n}^{\varepsilon_n} g\big(x + \varepsilon \widehat{\boldsymbol{N}}(x)\big) \det\big(\boldsymbol{I}_d + \varepsilon \widehat{S}(x)\big)\, d\varepsilon\, d\mathscr{H}(x). \tag{4.45}$$

This is (4.43). □

Proof of Theorem 2.5. Below we discuss the cases of $j = 1$ and $j = 2$ for $\lambda_{\varepsilon_n}^{(j)}$, respectively.

Case of $\lambda_{\varepsilon_n}^{(1)}$. We continue to use the notation given in the proof of Lemma 2.1. First we will use the following elementary results under the assumption (F). When $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) \le C_0$, for all $x \in \widehat{\mathcal{M}}_c$ and $|\varepsilon| \le r_2$, using (2.2) and Taylor expansion we have

$$\begin{aligned}
\|G(x)\|_F &= \sqrt{d-1},\\
\|\nabla \boldsymbol{N}(x)\|_F &\le \frac{\sqrt{d}\,Q_0}{\varepsilon_0},\\
\|\boldsymbol{N}(x) - \boldsymbol{N}\big(P_\varepsilon(x)\big)\| &\le \frac{\sqrt{d}\,Q_0\eta_0|\varepsilon|}{\varepsilon_0},\\
\big|\|\nabla f(x)\|^{-1} - \|\nabla f\big(P_\varepsilon(x)\big)\|^{-1}\big| &\le \frac{Q_0\eta_0|\varepsilon|}{\varepsilon_0^2},\\
\|G(x) - G\big(P_\varepsilon(x)\big)\|_F &\le \frac{2\sqrt{d}\,Q_0\eta_0|\varepsilon|}{\varepsilon_0},\\
\|\nabla \boldsymbol{N}(x) - \nabla \boldsymbol{N}\big(P_\varepsilon(x)\big)\|_F &\le \frac{3\sqrt{d}\,Q_0^2\eta_0}{\varepsilon_0^2}|\varepsilon| + \frac{d\sqrt{d}}{\varepsilon_0}\psi_f^{(2)}\big(\eta_0|\varepsilon|\big),\\
\big|\|\nabla \widehat{f}(x)\|^{-1} - \|\nabla f(x)\|^{-1}\big| &\le \frac{2\beta_n^{(1)}}{\varepsilon_0^2},\\
\|\widehat{\boldsymbol{N}}(x) - \boldsymbol{N}(x)\| &\le \frac{4\beta_n^{(1)}}{\varepsilon_0},\\
\|\widehat{G}(x) - G(x)\|_F &\le \frac{16\beta_n^{(1)}}{\varepsilon_0} \quad \text{and}\\
\|\nabla \widehat{\boldsymbol{N}}(x) - \nabla \boldsymbol{N}(x)\|_F &\le \frac{(48 + 3\sqrt{d})Q_0}{\varepsilon_0^2}\beta_n^{(1)} + \frac{\sqrt{d}}{\varepsilon_0}\beta_n^{(2)}.
\end{aligned} \tag{4.46}$$

Note that using Taylor expansion and Lemma 2.3, for $x \in \widehat{\mathcal{M}}$ and $|\varepsilon| \le r_2$ we have that

$$\begin{aligned}
&1 - \big\langle \widehat{\boldsymbol{N}}\big(P_\varepsilon(x)\big), \widehat{\boldsymbol{N}}(x)\big\rangle \\
&= \frac{1}{2}\big\|\widehat{\boldsymbol{N}}\big(P_\varepsilon(x)\big) - \widehat{\boldsymbol{N}}(x)\big\|^2 \\
&= \frac{1}{2}\bigg\|\int_0^1 \nabla\widehat{\boldsymbol{N}}\big((1-s)x + sP_\varepsilon(x)\big)\,ds\big[P_\varepsilon(x) - x\big]\bigg\|^2 \\
&= \frac{1}{2}\bigg\|\int_0^1 \big[\nabla\widehat{\boldsymbol{N}}\big((1-s)x + sP_\varepsilon(x)\big) - \nabla\widehat{\boldsymbol{N}}(x)\big]\,ds\big[P_\varepsilon(x) - x\big]\bigg\|^2 \\
&\le \frac{1}{2}\bigg\|\int_0^1 \big[\nabla\widehat{\boldsymbol{N}}\big((1-s)x + sP_\varepsilon(x)\big) - \nabla\widehat{\boldsymbol{N}}(x)\big]\,ds\bigg\|_F^2 \big\|P_\varepsilon(x) - x\big\|^2 \\
&\le \frac{1}{2}\bigg[2\frac{(48+3\sqrt{d})Q_0}{\varepsilon_0^2}\beta_n^{(1)} + \frac{2\sqrt{d}}{\varepsilon_0}\beta_n^{(2)} + \frac{3\sqrt{d}Q_0^2\eta_0}{\varepsilon_0^2}\varepsilon + \frac{d\sqrt{d}}{\varepsilon_0}\psi_f^{(2)}\big(\eta_0|\varepsilon|\big)\bigg]^2\eta_0^2\varepsilon^2 =: \alpha_\varepsilon.
\end{aligned}$$

Here we further restrict $|\varepsilon| \le \widetilde{r}_2$ for some $\widetilde{r}_2 > 0$ such that $\alpha_{\varepsilon_n} < 1$, which then implies

$$1 < \big\langle \widehat{\boldsymbol{N}}\big(P_\varepsilon(x)\big), \widehat{\boldsymbol{N}}(x)\big\rangle^{-1} \le 1 + \alpha_\varepsilon. \tag{4.47}$$

Using (2.18) and Taylor expansion,

$$\begin{aligned}
&g\big(P_\varepsilon(x)\big) \\
&\quad= g(x) + \widehat{t}_\varepsilon(x)\big\langle\nabla g(x), \widehat{\boldsymbol{N}}(x)\big\rangle \\
&\qquad+ \frac{1}{2}\widehat{t}_\varepsilon(x)^2\int_0^1 \big[\widehat{\boldsymbol{N}}(x)\big]^T\nabla^2 g\big((1-s)x + sP_\varepsilon(x)\big)\widehat{\boldsymbol{N}}(x)\,ds \\
&\quad= g(x) + \varepsilon\big\|\nabla\widehat{f}(x)\big\|^{-1}\big\langle\nabla g(x), \widehat{\boldsymbol{N}}(x)\big\rangle + \frac{1}{2}\varepsilon^2\big\|\nabla\widehat{f}(x)\big\|^{-2}\big\|\widehat{\boldsymbol{N}}(x)\big\|_{\nabla^2 g(x)}^2 \\
&\qquad- \frac{1}{2}\varepsilon^2\big\|\nabla\widehat{f}(x)\big\|^{-3}\big\|\widehat{\boldsymbol{N}}(x)\big\|_{\nabla^2\widehat{f}(x)}^2\big\langle\nabla g(x), \widehat{\boldsymbol{N}}(x)\big\rangle + \delta_{\varepsilon,1}(x),
\end{aligned} \tag{4.48}$$

where for all $x \in \widehat{\mathcal{M}}_c$ and $|\varepsilon| \le r_2$, $|\delta_{\varepsilon,1}(x)|$ is bounded from above by

$$J_1C_3\big(|\varepsilon|^3 + \varepsilon^2\beta_n^{(2)} + \varepsilon^2\psi_f^{(2)}\big(\eta_0|\varepsilon|\big)\big) + \bigg(\frac{4Q_0\eta_0^2|\varepsilon|^3}{\varepsilon_0^3} + \frac{2Q_0^2\eta_0^4\varepsilon^4}{\varepsilon_0^4}\bigg)J_2 + \frac{d}{2}\eta_0^2\varepsilon^2\psi_g^{(2)}\big(\eta_0|\varepsilon|\big),$$

with $C_3 > 0$ given in Lemma 2.3. Recall that $\widehat{H}(x) = \sum_{i=1}^{d-1}\widehat{\kappa}_i(x)$. Denote $\widehat{Q}(x) = \frac{1}{2}\sum\sum_{i\neq j}\widehat{\kappa}_i(x) \times \widehat{\kappa}_j(x)$ when $d \ge 3$, and set $\widehat{Q}(x) \equiv 0$ when $d = 2$. Notice that

$$\max_{i=1,\ldots,d-1}\big|\widehat{\kappa}_i(x)\big| \le \big\|\nabla\widehat{\boldsymbol{N}}(x)\big\|_F\big\|\widehat{G}(x)\big\|_F \le \frac{3(8+\sqrt{d})^2Q_0}{\varepsilon_0} =: A_0.$$

Hence $|\widehat{H}(x)| = |\mathbf{tr}[\nabla\widehat{\boldsymbol{N}}(x)\widehat{G}(x)]| \leq \sqrt{d}\|\nabla\widehat{\boldsymbol{N}}(x)\|_F\|\widehat{G}(x)\|_F \leq \sqrt{d}A_0$ and $|\widehat{Q}(x)| \leq \frac{1}{2}|\widehat{H}(x)|^2 \leq \frac{1}{2}d^2A_0^2$. Using (2.18) again, we have

$$\prod_{i=1}^{d-1}\big[1+\widehat{t}_\varepsilon(x)\widehat{\kappa}_i(x)\big] = 1+\widehat{t}_\varepsilon(x)\widehat{H}(x)+\big[\widehat{t}_\varepsilon(x)\big]^2\widehat{Q}(x)+\delta_{\varepsilon,2}(x),$$

where $|\delta_{\varepsilon,2}(x)| \leq A_0^3\eta_0^3|\varepsilon|^3d^2(1+A_0\eta_0|\varepsilon|)^d =: u(\varepsilon)$ for all $\widehat{\mathcal{M}}_c$ and $|\varepsilon| \leq r_2$. Hence,

$$\begin{aligned}\prod_{i=1}^{d-1}\big[1+\widehat{t}_\varepsilon(x)\widehat{\kappa}_i(x)\big] &= 1+\varepsilon\big\|\nabla\widehat{f}(x)\big\|^{-1}\widehat{H}(x)+\varepsilon^2\big\|\nabla\widehat{f}(x)\big\|^{-2}\widehat{Q}(x)\\ &\quad -\frac{1}{2}\varepsilon^2\big\|\nabla\widehat{f}(x)\big\|^{-3}\widehat{H}(x)\big\|\widehat{\boldsymbol{N}}(x)\big\|^2_{\nabla^2\widehat{f}(x)}+\delta_{\varepsilon,3}(x),\end{aligned} \tag{4.49}$$

where

$$\big|\delta_{\varepsilon,3}(x)\big| \leq \sqrt{d}A_0C_3\big(|\varepsilon|^3+\varepsilon^2\beta_n^{(2)}+\varepsilon^2\psi_f^{(2)}(\eta_0|\varepsilon|)\big)+\frac{1}{2}d^2A_0^2\bigg(\frac{Q_0\eta_0^2|\varepsilon|^3}{2\varepsilon_0^2}+\frac{Q_0^2\eta_0^4\varepsilon^4}{4\varepsilon_0^2}\bigg)+u(\varepsilon).$$

Combining (4.47), (4.48), and (4.49), we get

$$\frac{g(P_\varepsilon(x))\prod_{i=1}^{d-1}[1+\widehat{t}_\varepsilon(x)\widehat{\kappa}_i(x)]}{\langle\widehat{\boldsymbol{N}}(P_\varepsilon(x)),\widehat{\boldsymbol{N}}(x)\rangle} = g(x)+\varepsilon\eta_n^{(1)}(x)+\varepsilon^2\theta_n^{(1)}(x)+\delta_{\varepsilon,4}(x),$$

where

$$\begin{aligned}\eta_n^{(1)}(x) &= \big\|\nabla\widehat{f}(x)\big\|^{-1}\big\langle\nabla g(x),\widehat{\boldsymbol{N}}(x)\big\rangle+g(x)\big\|\nabla\widehat{f}(x)\big\|^{-1}\widehat{H}(x),\\ \theta_n^{(1)}(x) &= \big\|\nabla\widehat{f}(x)\big\|^{-2}\widehat{H}(x)\big\langle\nabla g(x),\widehat{\boldsymbol{N}}(x)\big\rangle\\ &\quad +\frac{1}{2}\Big[\big\|\nabla\widehat{f}(x)\big\|^{-2}\big\|\widehat{\boldsymbol{N}}(x)\big\|^2_{\nabla^2 g(x)}\\ &\quad -\big\|\nabla\widehat{f}(x)\big\|^{-3}\big\|\widehat{\boldsymbol{N}}(x)\big\|^2_{\nabla^2\widehat{f}(x)}\big\langle\nabla g(x),\widehat{\boldsymbol{N}}(x)\big\rangle\Big]\\ &\quad +g(x)\Big[\big\|\nabla\widehat{f}(x)\big\|^{-2}\widehat{Q}(x)-\frac{1}{2}\big\|\nabla\widehat{f}(x)\big\|^{-3}\widehat{H}(x)\big\|\widehat{\boldsymbol{N}}(x)\big\|^2_{\nabla^2\widehat{f}(x)}\Big],\end{aligned} \tag{4.50}$$

and $|\delta_{\varepsilon,4}(x)| \leq A_1\phi_{n,\varepsilon}$ for some constant $A_1 > 0$ for all $x \in \widehat{\mathcal{M}}_c$ and $|\varepsilon| \leq \min(r_1, r_2, \widetilde{r}_2) =: r_3$. Therefore from (4.44) we have that when $\varepsilon_n \leq r_3$,

$$\begin{aligned}\lambda^{(1)}_{\varepsilon_n}(\widehat{f}, g) &= \frac{1}{2\varepsilon_n}\int_{-\varepsilon_n}^{\varepsilon_n}\int_{\widehat{\mathcal{M}}_c}\big[g(x)+\varepsilon\eta_n^{(1)}(x)+\varepsilon^2\theta_n^{(1)}(x)+\delta_{\varepsilon,4}(x)\big]\,d\mathscr{H}(x)\,d\varepsilon\\ &= \lambda(\widehat{f}, g)+\varepsilon_n^2\mu_n^{(1)}+\rho_n^{(1)},\end{aligned}$$

where $\mu_n^{(1)} = \frac{1}{3}\lambda(\widehat{f}, \theta_n^{(1)})$ and $|\rho_n^{(1)}| \leq A_1\phi_{n,\varepsilon_n}$, which is (2.19) for $j = 1$.

Case of $\lambda^{(2)}_{\varepsilon_n}$. Next, we derive an approximation to $\lambda^{(2)}_{\varepsilon_n}(\widehat{f}, g)$. By Taylor expansion,

$$g\big(x+\varepsilon\widehat{\boldsymbol N}(x)\big) = g(x) + \varepsilon\big\langle\widehat{\boldsymbol N}(x), \nabla g(x)\big\rangle + \frac{1}{2}\varepsilon^2\big[\widehat{\boldsymbol N}(x)\big]^T\int_0^1 \nabla^2 g\big(x+s\varepsilon\widehat{\boldsymbol N}(x)\big)\,ds\widehat{\boldsymbol N}(x)$$
$$= g(x) + \varepsilon\big\langle\widehat{\boldsymbol N}(x), \nabla g(x)\big\rangle + \frac{1}{2}\varepsilon^2\big\|\widehat{\boldsymbol N}(x)\big\|^2_{\nabla^2 g(x)} + \delta_{\varepsilon,5}(x),$$

where $|\delta_{\varepsilon,5}(x)| \le \frac{d}{2}\psi^{(2)}_g(|\varepsilon|)\varepsilon^2$ for $x\in\widehat{\mathcal M}_c$ and $|\varepsilon|\le r_3$. Also $\det(\boldsymbol I_d+\varepsilon\widehat{S}(x)) = 1+\varepsilon\widehat{H}(x) + \varepsilon^2\widehat{Q}(x)+\delta_{\varepsilon,6}(x)$, where $|\delta_{\varepsilon,6}(x)|\le|\varepsilon|^3 d^2(1+A_0|\varepsilon|)^d$. Hence, $g(x+\varepsilon\widehat{\boldsymbol N}(x))\det(\boldsymbol I_d+\varepsilon\widehat{S}(x)) = g(x)+\varepsilon\eta^{(2)}_n(x)+\varepsilon^2\theta^{(2)}_n(x)+\delta_{\varepsilon,7}(x)$, where $|\delta_{\varepsilon,7}(x)|\le A_2\phi_{n,\varepsilon}$ for some constant $A_2>0$ for all $x\in\widehat{\mathcal M}_c$ and $|\varepsilon|\le r_3$ and

$$\eta^{(2)}_n(x) = \big\langle\widehat{\boldsymbol N}(x), \nabla g(x)\big\rangle + g(x)\widehat{H}(x),$$
$$\theta^{(2)}_n(x) = \widehat{H}(x)\big\langle\widehat{\boldsymbol N}(x), \nabla g(x)\big\rangle + \frac{1}{2}\big\|\widehat{\boldsymbol N}(x)\big\|^2_{\nabla^2 g(x)} + g(x)\widehat{Q}(x). \tag{4.51}$$

From (4.45), we have that when $\varepsilon_n\le r_3$, $\lambda^{(2)}_{\varepsilon_n}(\widehat{f}, g) = \frac{1}{2\varepsilon_n}\int_{-\varepsilon_n}^{\varepsilon_n}\lambda(\widehat{f}, g+\varepsilon\eta^{(2)}_n+\varepsilon^2\theta^{(2)}_n+\delta_{\varepsilon,7})\,d\varepsilon$, which yields (2.19) for $j=2$. □

Proof of Corollary 2.1. We can write $\lambda^*_{\varepsilon_n}(\widehat{f}, p_n) - \lambda(f,p) = D_{n,1}+D_{n,2}+D_{n,3}$, where $D_{n,1} = \lambda^*_{\varepsilon_n}(\widehat{f}, p_n-p)$, $D_{n,2} = \lambda^*_{\varepsilon_n}(\widehat{f}, p) - \lambda(\widehat{f}, p)$, and $D_{n,3} = \lambda(\widehat{f}, p)-\lambda(f,p)$. It follows from the calculation in the proof of Theorems 2.5 and 2.1 that under the assumption of this corollary we can find a constant $C_5>0$ such that $|D_{n,1}|\le C_5\eta_{p,n}$, $|D_{n,2}|\le C_5(\psi^{(0)}_p((\eta_0\vee 1)\varepsilon_n)+\varepsilon_n)$, and $|D_{n,3}|\le C_5(\psi^{(0)}_p(\eta_0\beta^{(0)}_n)+\beta^{(0)}_n+(\beta^{(1)}_n)^2)$. Then we get (2.20), and (2.21) is a consequence of Lemma 4.1. □

Proof of Theorem 2.1. We assume that $\max(\beta^{(0)}_n, \beta^{(1)}_n, \beta^{(2)}_n) < C_0$ and $0<\varepsilon_n\le r_3$, where C_0 is given in Lemma 2.1 and r_3 is given in Theorem 2.5. First, we consider the case $\lambda^*_{\varepsilon_n}=\lambda$. Using Theorem 2.3, we can write $\lambda(\widehat{f}, g)-\lambda(f,g) = \mathrm{I}_n+\mathrm{II}_n+\mathrm{III}_n$, where

$$\mathrm{I}_n = \int_{\mathcal M}\big[g\big(P_n(x)\big)-g(x)\big]\,d\mathscr H(x), \tag{4.52}$$

$$\mathrm{II}_n = \int_{\mathcal M} g(x)\left\{\frac{\prod_{i=1}^{d-1}[1+t_n(x)\kappa_i(x)]}{\langle\widehat{\boldsymbol N}(P_n(x)), \boldsymbol N(x)\rangle}-1\right\}d\mathscr H(x), \tag{4.53}$$

$$\mathrm{III}_n = \int_{\mathcal M}\big[g\big(P_n(x)\big)-g(x)\big]\left\{\frac{\prod_{i=1}^{d-1}[1+t_n(x)\kappa_i(x)]}{\langle\widehat{\boldsymbol N}(P_n(x)), \boldsymbol N(x)\rangle}-1\right\}d\mathscr H(x). \tag{4.54}$$

We first study I_n. Since by definition $P_n(x)-x = t_n(x)\boldsymbol N(x)$, using Taylor expansion and Lemma 2.2 we have that

$$\mathrm{I}_n = \int_{\mathcal M}\big\{\big\|\nabla f(x)\big\|^{-1}\big[f(x)-\widehat{f}(x)\big]\boldsymbol N(x)^T\nabla g(x)\big\}\,d\mathscr H(x) + L_n, \tag{4.55}$$

where for some $0<s<1$ with δ_n given in (2.10),

$$L_n = \int_{\mathcal M}\Big[\delta_n(x)\boldsymbol N(x)^T\nabla g(x) + \frac{1}{2}t_n^2(x)\boldsymbol N(x)^T\nabla^2 g\big(x+s\boldsymbol N(x)t_n(x)\big)\boldsymbol N(x)\Big]\,d\mathscr H(x).$$

We apply Lemmas 2.1 and 2.2 to the integrand of L_n and then obtain

$$|L_n| \le C_1 \lambda(f,1) J_1 \big(\beta_n^{(0)} \beta_n^{(1)} + \big(\beta_n^{(0)}\big)^2\big) + \frac{1}{2} \lambda(f,1) J_2 \eta_0^2 \big(\beta_n^{(0)}\big)^2, \tag{4.56}$$

where C_1 is given in Lemma 2.2.

Next we focus on II_n. We will keep using the elementary results given in (4.46) at the beginning of the proof of Theorem 2.5. Notice that for $x \in \mathcal{M}$ both $\boldsymbol{N}(x)$ and $\langle \widehat{\boldsymbol{N}}(P_n(x))$ are vectors with unit norm and hence $\langle \widehat{\boldsymbol{N}}(P_n(x)), \boldsymbol{N}(x)\rangle = 1 - \frac{1}{2}\|\widehat{\boldsymbol{N}}(P_n(x)) - \boldsymbol{N}(x)\|^2$. Observe that

$$\widehat{\boldsymbol{N}}\big(P_n(x)\big) - \boldsymbol{N}(x) = \big[\widehat{\boldsymbol{N}}(x) - \boldsymbol{N}(x)\big] + \delta_{n,1}(x), \tag{4.57}$$

where $\delta_{n,1}(x) = \{[\widehat{\boldsymbol{N}}(P_n(x)) - \boldsymbol{N}(P_n(x))] - [\widehat{\boldsymbol{N}}(x) - \boldsymbol{N}(x)]\} + [\boldsymbol{N}(P_n(x)) - \boldsymbol{N}(x)]$. Denote $U_n(x) = \widehat{\boldsymbol{N}}(x) - \boldsymbol{N}(x)$. Then using Taylor expansion, we get

$$\delta_{n,1}(x) = \int_0^1 \big[\nabla U_n\big(aP_n(x) + (1-a)x\big) + \nabla \boldsymbol{N}\big(aP_n(x) + (1-a)x\big)\big]\, da \big[P_n(x) - x\big],$$

and therefore by using Lemma 2.1 we get for $x \in \mathcal{M}$

$$\big\|\delta_{n,1}(x)\big\| \le \frac{(24 + 3\sqrt{d}) Q_0 \eta_0}{\varepsilon_0} \beta_n^{(0)}. \tag{4.58}$$

We denote $V_n(x) = \nabla \widehat{f}(x) - \nabla f(x)$ and $V_n^{(j)}(x) = \|\nabla \widehat{f}(x)\|^j - \|\nabla f(x)\|^j$ for $j \in \mathbb{Z}$. Notice that $V_n^{(2)}(x) = 2\langle \nabla f(x), V_n(x)\rangle + \|V_n(x)\|^2$. Furthermore, using Taylor expansion,

$$\begin{aligned} V_n^{(-1)}(x) &= \big\|\nabla f(x)\big\|^{-1} \big\{\big(1 + \big\|\nabla f(x)\big\|^{-2} V_n^{(2)}(x)\big)^{-1/2} - 1\big\} \\ &= -\frac{1}{2}\big\|\nabla f(x)\big\|^{-3} V_n^{(2)}(x) + \delta_{n,2}(x), \end{aligned}$$

where $\delta_{n,2}(x) = \frac{3}{8}(1 + s(x))^{-\frac{5}{2}} \|\nabla f(x)\|^{-5} [V_n^{(2)}(x)]^2$, for some $|s(x)| \le \|\nabla f(x)\|^{-2} |V_n^{(2)}(x)|$. Note that $|s(x)| \le \frac{5\beta_n^{(1)}}{2\varepsilon_0} \le \frac{5}{8}$ when $\beta_n^{(1)} \le \frac{1}{4}\varepsilon_0$, which is also what we assume for the rest of the proof. Therefore, we can write

$$V_n^{(-1)}(x) = -\big\|\nabla f(x)\big\|^{-3} \big\langle \nabla f(x), V_n(x)\big\rangle + \delta_{n,3}(x), \tag{4.59}$$

where $\delta_{n,3}(x) = -\frac{1}{2}\|\nabla f(x)\|^{-3}\|V_n(x)\|^2 + \delta_{n,2}(x)$. We have $|\delta_{n,3}(x)| \le \frac{74}{\varepsilon_0^3}(\beta_n^{(1)})^2$. Recall that $G(x) = \boldsymbol{I}_d - \boldsymbol{N}(x)\boldsymbol{N}(x)^T$. Using (4.59), we have

$$\begin{aligned} \widehat{\boldsymbol{N}}(x) - \boldsymbol{N}(x) &= \big\|\nabla f(x)\big\|^{-1} V_n(x) + V_n^{(-1)}(x) \nabla f(x) + V_n^{(-1)}(x) V_n(x) \\ &= \big\|\nabla f(x)\big\|^{-1} V_n(x) - \big\|\nabla f(x)\big\|^{-3} \big\langle \nabla f(x), V_n(x)\big\rangle \nabla f(x) + \delta_{n,4}(x) \\ &= G(x)\big\|\nabla f(x)\big\|^{-1} V_n(x) + \delta_{n,4}(x), \end{aligned} \tag{4.60}$$

where $\delta_{n,4}(x) = \delta_{n,3}(x)\nabla f(x) + V_n^{(-1)}(x)V_n(x)$. We have $\|\delta_{n,4}(x)\| \le \frac{76}{\varepsilon_0^2}(\beta_n^{(1)})^2$ due to (4.59). Using (4.57), (4.58), and (4.60), we have

$$\begin{aligned}
&\langle \widehat{\boldsymbol{N}}(P_n(x)), \boldsymbol{N}(x)\rangle \\
&= 1 - \frac{1}{2}\big\|\big[\widehat{\boldsymbol{N}}(x) - \boldsymbol{N}(x)\big] + \delta_{n,1}(x)\big\|^2 \\
&= 1 - \frac{1}{2}\big\|G(x)\|\nabla f(x)\|^{-1}V_n(x) + \delta_{n,4}(x) + \delta_{n,1}(x)\big\|^2 \\
&= 1 - \frac{1}{2}\big\|\nabla f(x)\big\|^{-2}\big[V_n(x)\big]^T G(x)\big[V_n(x)\big] + \delta_{n,5}(x),
\end{aligned} \tag{4.61}$$

where

$$\begin{aligned}
|\delta_{n,5}(x)| \le{}& \frac{1}{2}\big[(24+3\sqrt{d})Q_0\eta_0\big]\varepsilon_0^{-2}\big(\beta_n^{(0)}\big)^2 + \frac{1}{2}(24+3\sqrt{d})(19+\sqrt{d})Q_0\eta_0\varepsilon_0^{-2}\beta_n^{(0)}\beta_n^{(1)} \\
&+ 76(3+2\sqrt{d})\varepsilon_0^{-3}\big(\beta_n^{(1)}\big)^3 =: \alpha_{n,1}.
\end{aligned}$$

Also note that $\frac{1}{2}\|\nabla f(x)\|^{-2}|[V_n(x)]^T G(x)[V_n(x)]| \le \frac{\sqrt{d}}{2\varepsilon_0^2}(\beta_n^{(1)})^2 =: \alpha_{n,2}$. By further requiring $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) < \widetilde{C}_0$ for some $0 < \widetilde{C}_0 \le \frac{\varepsilon_0}{4}$ such that $\alpha_{n,1} + \alpha_{n,2} < \frac{1}{2}$, and using Taylor expansion and Lemma 4.1 we have

$$\langle \widehat{\boldsymbol{N}}(P_n(x)), \boldsymbol{N}(x)\rangle^{-1} = 1 + \frac{1}{2}\big\|\nabla f(x)\big\|^{-2}\big[V_n(x)\big]^T G(x)\big[V_n(x)\big] + \delta_{n,6}(x), \tag{4.62}$$

where $|\delta_{n,6}(x)| \le 2(\alpha_{n,1} + \alpha_{n,1}\alpha_{n,2} + (\alpha_{n,2})^2) =: \alpha_{n,3}$. Also note that by Lemma 2.2,

$$\prod_{i=1}^{d-1}\big[1 + t_n(x)\kappa_i(x)\big] = 1 + \big\|\nabla f(x)\big\|^{-1}H(x)\big[f(x) - \widehat{f}(x)\big] + \delta_{n,7}(x), \tag{4.63}$$

where $H(x)$ is the mean curvature of $\mathcal{M}$ at x, and

$$|\delta_{n,7}(x)| \le \frac{d\sqrt{d}Q_0}{\varepsilon_0}C_1\big(\beta_n^{(0)}\beta_n^{(1)} + \big(\beta_n^{(0)}\big)^2\big) + \frac{d^4Q_0^2\eta_0^2(\beta_n^{(0)})^2}{\varepsilon_0^2}\Big(1 + \frac{dQ_0}{\varepsilon_0}\eta_0\beta_n^{(0)}\Big)^d =: \alpha_{n,4}.$$

Denote $M(x) = \frac{1}{2}g(x)\|\nabla f(x)\|^{-2}G(x)$ and

$$\Gamma_n = \int_{\mathcal{M}}\big[\nabla\widehat{f}(x) - \nabla f(x)\big]^T M(x)\big[\nabla\widehat{f}(x) - \nabla f(x)\big]\,d\mathscr{H}(x). \tag{4.64}$$

Then using (4.62) and (4.63), we get

$$\mathrm{II}_n = \int_{\mathcal{M}} g(x)\big\|\nabla f(x)\big\|^{-1}H(x)\big[f(x) - \widehat{f}(x)\big]\,d\mathscr{H}(x) + \Gamma_n + \delta_{n,8}(x), \tag{4.65}$$

where $|\delta_{n,8}(x)| \le A_0[(\beta_n^{(0)})^2 + \beta_n^{(0)}\beta_n^{(1)} + (\beta_n^{(1)})^3]$ for some constant $A_0 > 0$.

Next, we study III_n. Lemma 2.1 leads to

$$\sup_{x\in\mathcal{M}}\big|g\big(P_n(x)\big) - g(x)\big| \le \sup_{x\in\mathcal{I}(2\delta_0)}\big\|\nabla g(x)\big\|\sup_{x\in\mathcal{M}}\big|t_n(x)\big| \le \eta_0 J_1\beta_n^{(0)}. \tag{4.66}$$

It follows from (4.61), (4.63), Lemmas 2.2 and 4.1 that

$$\sup_{x\in\mathcal{M}}\left|\frac{\prod_{i=1}^{d-1}[1+t_n(x)\kappa_i(x)]}{\langle\widehat{\boldsymbol{N}}(P_n(x)),\boldsymbol{N}(x)\rangle}-1\right|\le\alpha_{n,5}+\alpha_{n,6}+\alpha_{n,5}\alpha_{n,6}, \tag{4.67}$$

where $\alpha_{n,5}=\frac{\sqrt{d}}{2\varepsilon_0^2}(\beta_n^{(1)})^2+\alpha_{n,3}$ and $\alpha_{n,6}=\frac{d\sqrt{d}Q_0}{\varepsilon_0^2}\beta_n^{(0)}+\alpha_{n,4}$. By (4.66) and (4.67) we have

$$\begin{aligned}|\mathrm{III}_n|&\le\lambda(f,1)\sup_{x\in\mathcal{M}}\left|g\big(P_n(x)\big)-g(x)\right|\sup_{x\in\mathcal{M}}\left|\frac{\prod_{i=1}^{d-1}[1+t_n(x)\kappa_i(x)]}{\langle\widehat{\boldsymbol{N}}(P_n(x)),\boldsymbol{N}(x)\rangle}-1\right|\\&\le A_1\Big(\big(\beta_n^{(0)}\big)^2+\beta_n^{(0)}\big(\beta_n^{(1)}\big)^2\Big),\end{aligned} \tag{4.68}$$

for some constant $A_1>0$. Now collecting the results for I_n in (4.55), II_n in (4.65) and III_n in (4.68) we then get

$$\lambda(\widehat{f},g)-\lambda(f,g)=\lambda\big(f,w_g\times(f-\widehat{f})\big)+\Gamma_n+\eta_{n,1}, \tag{4.69}$$

where $|\eta_{n,1}|\le A_2(\beta_n^{(0)}\beta_n^{(1)}+(\beta_n^{(0)})^2+(\beta_n^{(1)})^3)$ for some constant $A_2>0$. Clearly,

$$|\Gamma_n|\le\lambda(f,1)\sup_{x\in\mathcal{M}}\left\|M(x)\right\|_F\big(\beta_n^{(1)}\big)^2\le\frac{\sqrt{d}J_0\lambda(f,1)}{2\varepsilon_0^2}\big(\beta_n^{(1)}\big)^2.$$

Also from Theorem 2.5, we have $|\lambda_{\varepsilon_n}^{(j)}(\widehat{f},g)-\lambda(\widehat{f},g)|\le A_3\varepsilon_n^2$ for $j=1,2$, for some constant $A_3>0$. Therefore, we can write

$$\lambda_{\varepsilon_n}^*(\widehat{f},g)-\lambda(f,g)=\lambda\big(f,w_g\times(f-\widehat{f})\big)+\eta_{n,2}, \tag{4.70}$$

for $\lambda_{\varepsilon_n}^*=\lambda$, $\lambda_{\varepsilon_n}^{(1)}$ or $\lambda_{\varepsilon_n}^{(2)}$, where $|\eta_{n,2}|\le A_4((\beta_n^{(0)}+\beta_n^{(1)})^2+\varepsilon_n^2)$ for some constant $A_4>0$. Using (4.4) we have

$$\lambda\big(f,w_g\times(f-\mathbb{E}\widehat{f})\big)=h^\nu\mu_2-\lambda(f,w_gr_h), \tag{4.71}$$

where

$$\mu_2=-\frac{1}{\nu!}\int_{\mathcal{M}}\left[\int_{\mathbb{R}^d}y^{\otimes\nu}K(y)\,dy\right]^T\nabla^{\otimes\nu}f(x)w_g(x)\,d\mathscr{H}(x). \tag{4.72}$$

Here $|\mu_2|<\infty$ and by using (4.5),

$$\left|\lambda(f,w_gr_h)\right|\le\frac{2^\nu}{\nu!}\lambda(f,1)\sup_{x\in\mathcal{M}}\left|w_g(x)\right|\|K\|_1h^\nu\psi_f^{(\nu)}(h).$$

We denote $\pi_n=\lambda(f,w_g\times(\mathbb{E}\widehat{f}-\widehat{f}))$. Since $\lambda(f,w_g\times(f-\widehat{f}))=\lambda(f,w_g\times(f-\mathbb{E}\widehat{f}))+\pi_n$, from (4.70) we can write

$$\lambda_{\varepsilon_n}^*(\widehat{f},g)-\lambda(f,g)=\pi_n+\eta_{n,3}, \tag{4.73}$$

where $|\eta_{n,3}|\le A_5((\beta_n^{(0)}+\beta_n^{(1)})^2+\varepsilon_n^2+h^\nu)$ for some constant $A_5>0$. Let $\widetilde{\beta}_n^{(0)}=\sup_{x\in\mathbb{R}^d}|\widehat{f}(x)-\mathbb{E}\widehat{f}(x)|$ and $\widetilde{\beta}_n^{(1)}=\sup_{x\in\mathcal{I}(2\delta_0)}\|\nabla\widehat{f}(x)-\mathbb{E}\nabla\widehat{f}(x)\|$. Then using Lemma 4.1, $|\eta_{n,3}|\le A_6((\widetilde{\beta}_n^{(0)}+$

$\widetilde{\beta}_n^{(1)})^2 + \varepsilon_n^2 + h^\nu)$ for some constant $A_6 > 0$. Recall that so far we require $\max(\beta_n^{(0)}, \beta_n^{(1)}, \beta_n^{(2)}) < C_0 \wedge \widetilde{C}_0$. Denote this event by $\mathscr{E}$. Note that for any constant $C > 2A_6$ and any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\{|\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f,g)| \le \varepsilon \vee C(\varepsilon_n^2 + h^\nu + (\gamma_{n,h}^{(1)})^2)\} \\
&\quad \ge \mathbb{P}\{[|\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f,g)| \le \varepsilon \vee C(\varepsilon_n^2 + h^\nu + (\gamma_{n,h}^{(1)})^2)] \cap \mathscr{E}\} \\
&\quad \ge \mathbb{P}\left\{\left[|\pi_n| \le \frac{1}{2}\varepsilon\right] \cap \mathscr{E}\right\} + \mathbb{P}\left\{\left[|\eta_{n,3}| \le \frac{1}{2}\varepsilon \vee \frac{1}{2}C((\varepsilon_n^2 + h^\nu + \gamma_{n,h}^{(1)})^2)\right] \cap \mathscr{E}\right\} - \mathbb{P}(\mathscr{E}) \\
&\quad \ge \mathbb{P}\left\{|\pi_n| \le \frac{1}{2}\varepsilon\right\} + \mathbb{P}\left\{(\widetilde{\beta}_n^{(0)} + \widetilde{\beta}_n^{(1)})^2 \le \left(\frac{1}{2A_6} - \frac{1}{C}\right)\varepsilon \vee \frac{C}{2A_6}(\gamma_{n,h}^{(1)})^2\right\} \\
&\qquad + \mathbb{P}(\mathscr{E}) - 2.
\end{aligned} \tag{4.74}$$

Using Theorem 2.4 we get for all $n \ge 1$, $h \in (0, h_0]$ and $\varepsilon \in [0, t_0]$,

$$\mathbb{P}\left(|\pi_n| \le \frac{1}{2}\varepsilon\right) \ge 1 - 2\exp\left(-\frac{C_2}{4}nh\varepsilon^2\right). \tag{4.75}$$

By Lemma 4.1, we have $\mathbb{P}(\mathscr{E}) \ge 1 - A_7 \exp(-nh^{d+4}/A_7)$ and

$$\mathbb{P}\{(\widetilde{\beta}_n^{(0)} + \widetilde{\beta}_n^{(1)})^2 < A_8(t \vee (\gamma_{n,h}^{(1)})^2)\} \ge 1 - A_9 \exp(-nh^{d+2}t/A_9) \tag{4.76}$$

for some constants $A_7, A_8, A_9 > 0$ when $t \in [0, h_1]$, $\frac{|\log h|}{nh^{d+4}} \in (0, h_2]$ and $h \in (0, h_3]$ for some constants $h_1, h_2, h_3 > 0$. Using $t = \frac{1}{A_8}(\frac{1}{2A_6} - \frac{1}{C})\varepsilon$ with $C > (2A_6)(A_8 \vee 1)$ in (4.76), from (4.74) we get the desired result in this theorem. □

Proof of Theorem 2.2. With μ_2 given in (4.72) and μ_1 given in (4.87) below, we will derive the following result. Under the assumption (H) and when $\varepsilon_n \to 0$,

$$\begin{aligned}
\lambda^*_{\varepsilon_n}(\widehat{f}, g) - \lambda(f,g) &= \lambda(f, w_g \times (\mathbb{E}\widehat{f} - \widehat{f})) + \frac{1}{nh^{d+2}}\mu_1 + h^\nu \mu_2 + \varepsilon_n^2 \mu^{(j)} \\
&\quad + o_p\left(h^\nu + \frac{1}{nh^{d+2}} + \frac{1}{\sqrt{nh}} + \varepsilon_n^2\right).
\end{aligned} \tag{4.77}$$

By Theorem 2.4, we have $\sqrt{nh}\lambda(f, w_g \times (\mathbb{E}\widehat{f} - \widehat{f})) \xrightarrow{D} \mathscr{N}(0, \sigma^2)$, where $\sigma^2 = c\lambda(f, w_g^2 R_K)$. Then the assertions in parts (i)–(iv) immediately follow from (4.77). Notice that

$$\sup_{x \in \mathcal{I}(2\delta_0)} |\widehat{w}_g(x) - w_g(x)| = O_p(\gamma_{n,h}^{(2)}) + o_p(h^{\nu-2}) = o_p(1)$$

by Lemma 4.1. Using Corollary 2.1, we get $|\widehat{\sigma}^2_{\tau_n} - \sigma^2| = o_p(1)$ and hence the assertion in part (v) follows. In order to show (4.77), noticing $\lambda^{(j)}_{\varepsilon_n}(\widehat{f}, g) - \lambda(f,g) = [\lambda^{(j)}_{\varepsilon_n}(\widehat{f}, g) - \lambda(\widehat{f}, g)] + [\lambda(\widehat{f}, g) - \lambda(f,g)]$, and by using (4.69), (4.71) and Lemma 4.1, it suffices to show that

$$\lambda^{(j)}_{\varepsilon_n}(\widehat{f}, g) - \lambda(\widehat{f}, g) = \varepsilon_n^2(\mu^{(j)} + o_p(1)), \quad j = 1, 2, \tag{4.78}$$

$$\Gamma_n = \frac{1}{nh^{d+2}}(\mu_1 + o_p(1)) + o_p\left(\frac{1}{\sqrt{nh}}\right) + o(h^{2\nu-2}). \tag{4.79}$$

for Γ_n given in (4.64). First we show (4.78). Recall $\theta_n^{(j)}$ and $\mu_n^{(j)}$ for $j=1,2$, given in Theorem 2.5. Let

$$\mu^{(j)} = \frac{1}{3}\lambda\big(f, \theta^{(j)}\big), \quad j=1,2, \tag{4.80}$$

where with $Q(x) = \frac{1}{2}\sum\sum_{i\neq j}\kappa_i(x)\kappa_j(x)$ when $d\geq 3$ and $Q(x)\equiv 0$ when $d=2$,

$$\begin{aligned}\theta^{(1)}(x) &= \big\|\nabla f(x)\big\|^{-2} H(x)\big\langle \nabla g(x), \boldsymbol{N}(x)\big\rangle \\ &\quad + \frac{1}{2}\big[\|\nabla f(x)\|^{-2}\|\boldsymbol{N}(x)\|^2_{\nabla^2 g(x)} - \|\nabla f(x)\|^{-3}\|\boldsymbol{N}(x)\|^2_{\nabla^2 f(x)}\langle \nabla g(x), \boldsymbol{N}(x)\rangle\big] \\ &\quad + g(x)\Big[\big\|\nabla f(x)\big\|^{-2} Q(x) - \frac{1}{2}\big\|\nabla f(x)\big\|^{-3} H(x)\big\|\boldsymbol{N}(x)\big\|^2_{\nabla^2 f(x)}\Big], \\ \theta^{(2)}(x) &= H(x)\big\langle \boldsymbol{N}(x), \nabla g(x)\big\rangle + \frac{1}{2}\big\|\boldsymbol{N}(x)\big\|^2_{\nabla^2 g(x)} + g(x)Q(x).\end{aligned}$$

It is known by Weyl's inequality that eigenvalues are Lipschitz continuous as functions of symmetric matrices. Hence we get (4.78) by using Corollary 2.1 and Lemma 4.1.

For the rest of the proof, we will show (4.79). We can write $\Gamma_n = \Gamma_n^{(1)} + \Gamma_n^{(2)} + \Gamma_n^{(3)}$, where

$$\Gamma_n^{(1)} = \int_{\mathcal{M}} \big\|\nabla \widehat{f}(x) - \mathbb{E}\nabla \widehat{f}(x)\big\|^2_{M(x)}\, d\mathscr{H}(x), \tag{4.81}$$

$$\Gamma_n^{(2)} = 2\int_{\mathcal{M}} \big\langle \nabla \widehat{f}(x) - \mathbb{E}\nabla \widehat{f}(x), \mathbb{E}\nabla \widehat{f}(x) - \nabla f(x)\big\rangle_{M(x)}\, d\mathscr{H}(x), \tag{4.82}$$

$$\Gamma_n^{(3)} = \int_{\mathcal{M}} \big\|\mathbb{E}\nabla \widehat{f}(x) - \nabla f(x)\big\|^2_{M(x)}\, d\mathscr{H}(x). \tag{4.83}$$

Next, we focus on $\Gamma_n^{(1)}$. Denote $\nabla K_h(\cdot) = \nabla K(\cdot/h)$ and

$$L_h^{i,j}(x) = \big[\nabla K_h(x-X_i) - \mathbb{E}\nabla K_h(x-X_1)\big]^T M(x)\big[\nabla K_h(x-X_j) - \mathbb{E}\nabla K_h(x-X_1)\big].$$

Notice that $L_h^{i,j}(x) = L_h^{j,i}(x)$. Also if $i\neq j$, then $\mathbb{E}[L_h^{i,j}(x)] = 0$. Then we can write $\Gamma_n^{(1)} = \frac{1}{n^2h^{2d+2}}\sum_{i=1}^n\sum_{j=1}^n \lambda(f, L_h^{i,j})$. We first study $\mathbb{E}\Gamma_n^{(1)}$ and can write

$$\mathbb{E}\Gamma_n^{(1)} = \frac{1}{n^2h^{2d+2}}\mathbb{E}\sum_{i=1}^n \lambda\big(f, L_h^{i,i}\big) = \frac{1}{nh^{2d+2}}\mathbb{E}\lambda\big(f, L_h^{1,1}\big) = \frac{1}{nh^{2d+2}}\big[Q_h^{(1)} - Q_h^{(2)}\big], \tag{4.84}$$

where

$$Q_h^{(1)} = \mathbb{E}\int_{\mathcal{M}} \big\|\nabla K_h(x-X)\big\|^2_{M(x)}\, d\mathscr{H}(x), \tag{4.85}$$

$$Q_h^{(2)} = \int_{\mathcal{M}} \big\|\mathbb{E}\nabla K_h(x-X)\big\|^2_{M(x)}\, d\mathscr{H}(x). \tag{4.86}$$

Then using change of variable $z=(x-y)/h$ and Taylor expansion for $f(x-hz)$ we have

$$\begin{aligned}
Q_h^{(1)} &= \int_{\mathcal{M}}\int_{\mathbb{R}^d}\|\nabla K_h(x-y)\|_{M(x)}^2 f(y)dyd\mathscr{H}(x)\\
&= h^d\int_{\mathcal{M}}\int_{\mathbb{R}^d}\|\nabla K(z)\|_{M(x)}^2 f(x-hz)\,dz\,d\mathscr{H}(x) = h^d\mu_1 + o\big(h^d\big),
\end{aligned}$$

where

$$\mu_1 = c\int_{\mathcal{M}}\int_{\mathbb{R}^d}\|\nabla K(z)\|_{M(x)}^2\,dz\,d\mathscr{H}(x). \tag{4.87}$$

Similarly, using change of variable and Taylor expansion for $\nabla f(x-hz)$ we have

$$\begin{aligned}
Q_h^{(2)} &= \int_{\mathcal{M}}\left\|\int_{\mathbb{R}^d}\nabla K_h(x-y)f(y)\,dy\right\|_{M(x)}^2 d\mathscr{H}(x)\\
&= h^{2d}\int_{\mathcal{M}}\left\|\int_{\mathbb{R}^d}\nabla K(z)f(x-hz)\,dz\right\|_{M(x)}^2 d\mathscr{H}(x)\\
&= h^{2d+2}\int_{\mathcal{M}}\left\|\int_{\mathbb{R}^d}K(z)\nabla f(x-hz)\,dz\right\|_{M(x)}^2 d\mathscr{H}(x)\\
&= h^{2d+2}\int_{\mathcal{M}}\|\nabla f(x)\|_{M(x)}^2\,d\mathscr{H}(x) + o\big(h^{2d+2}\big).
\end{aligned}$$

Therefore from (4.84), we get

$$\mathbb{E}\Gamma_n^{(1)} = \frac{1}{nh^{d+2}}\big(\mu_1 + o(1)\big). \tag{4.88}$$

Next, we compute the variance of $\Gamma_n^{(1)}$. We have

$$\mathrm{Var}\big(\Gamma_n^{(1)}\big) = \mathbb{E}\big[\big(\Gamma_n^{(1)}\big)^2\big] - \big[\mathbb{E}\Gamma_n^{(1)}\big]^2 = \Lambda_n^{(1)} + \Lambda_n^{(2)} + \Lambda_n^{(3)}, \tag{4.89}$$

where

$$\Lambda_n^{(1)} = \frac{2}{n^4h^{4d+4}}\mathbb{E}\sum\sum_{i\neq j}\big[\lambda\big(f, L_h^{i,j}\big)\big]^2 = \frac{2(n-1)}{n^3h^{4d+4}}\mathbb{E}\big[\lambda\big(f, L_h^{1,2}\big)\big]^2, \tag{4.90}$$

$$\Lambda_n^{(2)} = \frac{1}{n^4h^{4d+4}}\mathbb{E}\sum_{i=1}^n\big[\lambda\big(f, L_h^{i,i}\big)\big]^2 = \frac{1}{n^3h^{4d+4}}\mathbb{E}\big[\lambda\big(f, L_h^{1,1}\big)\big]^2, \quad \text{and} \tag{4.91}$$

$$\begin{aligned}
\Lambda_n^{(3)} &= \frac{1}{n^4h^{4d+4}}\mathbb{E}\sum\sum_{i\neq j}\big[\lambda\big(f, L_h^{i,i}\big)\lambda\big(f, L_h^{j,j}\big)\big] - \big[\mathbb{E}\Gamma_n^{(1)}\big]^2\\
&= \left[\frac{n(n-1)}{n^2} - 1\right]\big[\mathbb{E}\Gamma_n^{(1)}\big]^2 = -\frac{1}{n}\big[\mathbb{E}\Gamma_n^{(1)}\big]^2.
\end{aligned} \tag{4.92}$$

By (4.88), obviously $\Lambda_n^{(3)} = O(\frac{1}{nh^{2d+4}})$. To find the rates of $\Lambda_n^{(1)}$ and $\Lambda_n^{(2)}$, notice that

$$\mathbb{E}\big[\lambda\big(f, L_h^{1,2}\big)\big]^2 = S_h^{(2)} - 2S_h^{(3)} + \big(Q_h^{(2)}\big)^2, \tag{4.93}$$

$$\mathbb{E}\big[\lambda\big(f, L_h^{1,1}\big)\big]^2 = S_h^{(1)} - 4S_h^{(4)} - 3\big(Q_h^{(2)}\big)^2 + 2Q_h^{(1)} Q_h^{(2)} + 4S_h^{(3)}, \tag{4.94}$$

where $Q_h^{(1)}$ and $Q_h^{(2)}$ are given in (4.85) and (4.86), and

$$\begin{aligned}
S_h^{(1)} &= \mathbb{E}\Big[\int_{\mathcal{M}} \|\nabla K_h(x-X)\|_{M(x)}^2 \, d\mathscr{H}(x)\Big]^2, \\
S_h^{(2)} &= \mathbb{E}\Big[\int_{\mathcal{M}} \big\langle \nabla K_h(x-X_1), \nabla K_h(x-X_2)\big\rangle_{M(x)} \, d\mathscr{H}(x)\Big]^2, \\
S_h^{(3)} &= \mathbb{E}\Big[\int_{\mathcal{M}} \big\langle \nabla K_h(x-X), \mathbb{E}\nabla K_h(x-X)\big\rangle_{M(x)} \, d\mathscr{H}(x)\Big]^2, \\
S_h^{(4)} &= \mathbb{E}\Big\{\Big[\int_{\mathcal{M}} \|\nabla K_h(x-X)\|_{M(x)}^2 \, d\mathscr{H}(x)\Big] \\
&\qquad \times \Big[\int_{\mathcal{M}} \big\langle \nabla K_h(x-X), \mathbb{E}\nabla K_h(x-X)\big\rangle_{M(x)} \, d\mathscr{H}(x)\Big]\Big\}.
\end{aligned}$$

Next we calculate the rates of the above quantities. For $S_h^{(1)}$, using change of variable $u = (x-z)/h$ and following a similar argument for (4.21), we have

$$\begin{aligned}
S_h^{(1)} &= \int_{\mathbb{R}^d} \Big[\int_{\mathcal{M}} \|\nabla K_h(x-z)\|_{M(x)}^2 \, d\mathscr{H}(x)\Big] \Big[\int_{\mathcal{M}} \|\nabla K_h(y-z)\|_{M(y)}^2 \, d\mathscr{H}(y)\Big] f(z)\, dz \\
&= h^d \int_{\mathcal{M}} \int_{\mathcal{M}} \int_{\mathbb{R}^d} \|\nabla K(u)\|_{M(x)}^2 \|\nabla K_h(y-x+hu)\|_{M(y)}^2 f(x-hu)\, du\, d\mathscr{H}(y)\, d\mathscr{H}(x) \\
&= ch^{2d-1} \int_{\mathcal{M}} \int_{T_x(\mathcal{M})} \int_{\mathbb{R}^d} \|\nabla K(u)\|_{M(x)}^2 \|\nabla K(v+u)\|_{M(x)}^2 \, du\, d\mathscr{H}(v)\, d\mathscr{H}(x) \\
&\quad + o\big(h^{2d-1}\big).
\end{aligned}$$

Similarly, using change of variable $u = (x-z)/h$ and $v = (x-w)/h$ we get

$$\begin{aligned}
S_h^{(2)} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Big[\int_{\mathcal{M}} \big\langle \nabla K_h(x-z), \nabla K_h(x-w)\big\rangle_{M(x)} \, d\mathscr{H}(x)\Big] \\
&\qquad \times \Big[\int_{\mathcal{M}} \big\langle \nabla K_h(y-z), \nabla K_h(y-w)\big\rangle_{M(y)} \, d\mathscr{H}(y)\Big] f(z) f(w) dz dw \\
&= h^{2d} \int_{\mathcal{M}} \int_{\mathcal{M}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \big\langle \nabla K_h(y-x+hu), \nabla K_h(y-x+hv)\big\rangle_{M(y)} \\
&\qquad \times \big\langle \nabla K(u), \nabla K(v)\big\rangle_{M(x)} f(x-hu) f(y-hv)\, du\, dv\, d\mathscr{H}(y)\, d\mathscr{H}(x) \\
&= c^2 h^{3d-1} \int_{\mathcal{M}} \int_{T_x(\mathcal{M})} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \big\langle \nabla K(u), \nabla K(v)\big\rangle_{M(x)} \\
&\qquad \times \big\langle \nabla K(w+u), \nabla K(w+v)\big\rangle_{M(x)} \, du\, dv\, d\mathscr{H}(w)\, d\mathscr{H}(x) + o\big(h^{3d-1}\big).
\end{aligned}$$

Using change of variable $z = (x - v)/h$, we get

$$
\begin{aligned}
S_h^{(3)} &= \int_{\mathbb{R}^d} \left[\int_{\mathcal{M}} \int_{\mathbb{R}^d} \langle \nabla K_h(x-u), \nabla K_h(x-v) \rangle_{M(x)} f(v)\, dv\, d\mathscr{H}(x) \right]^2 f(u)\, du \\
&= h^d \int_{\mathbb{R}^d} \left[\int_{\mathcal{M}} \int_{\mathbb{R}^d} \langle \nabla K_h(x-u), \nabla K(z) \rangle_{M(x)} f(x-hz)\, dz\, d\mathscr{H}(x) \right]^2 f(u)\, du \\
&= h^{d+1} \int_{\mathbb{R}^d} \left[\int_{\mathcal{M}} \int_{\mathbb{R}^d} \langle \nabla K_h(x-u), \nabla f(x-hz) \rangle_{M(x)} K(z)\, dz\, d\mathscr{H}(x) \right]^2 f(u)\, du \\
&= h^{2d+2} \int_{\mathbb{R}^d} \left[\int_{\mathcal{M}} \langle \nabla K_h(x-u), \nabla f(x) + o(1) \rangle_{M(x)} d\mathscr{H}(x) \right]^2 f(u)\, du \\
&= ch^{4d+1} \int_{\mathcal{M}} \int_{T_x(\mathcal{M})} \int_{\mathbb{R}^d} \langle \nabla K(u), \nabla f(x) \rangle_{M(x)} \langle \nabla K(u+v), \nabla f(x) \rangle_{M(x)} \\
&\quad du\, d\mathscr{H}(v)\, d\mathscr{H}(x) + o\big(h^{4d+1}\big).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
S_h^{(4)} = ch^{3d} &\int_{\mathcal{M}} \int_{T_x(\mathcal{M})} \int_{\mathbb{R}^d} \|\nabla K(u)\|_{M(x)}^2 \langle \nabla K(u+v), \nabla f(x) \rangle_{M(x)} du\, d\mathscr{H}(v)\, d\mathscr{H}(x) \\
&+ o\big(h^{3d}\big).
\end{aligned}
$$

Plugging the rates of $S_h^{(1)}$, $S_h^{(2)}$, $S_h^{(3)}$, and $S_h^{(4)}$ into (4.93) and (4.94), and noticing (4.90) and (4.91), we have

$$
\Lambda_n^{(1)} = O\left(\frac{1}{n^2 h^{d+5}}\right), \quad \text{and} \quad \Lambda_n^{(2)} = O\left(\frac{1}{n^3 h^{2d+5}}\right),
$$

and hence it follows from (4.88) and (4.89) that with μ_1 given in (4.87),

$$
\Gamma_n^{(1)} = \frac{1}{nh^{d+2}} \big(\mu_1 + o_p(1)\big). \tag{4.95}
$$

Next, we study $\Gamma_n^{(2)}$. Notice that we can write $\Gamma_n^{(2)} = \frac{2}{nh^{d+1}} \sum_{i=1}^n (Y_i - \mathbb{E}Y_i)$, where $Y_i = \int_{\mathcal{M}} \langle \nabla K_h(x - X_i), \mathbb{E}\nabla \widehat{f}(x) - \nabla f(x) \rangle_{M(x)}\, d\mathscr{H}(x)$. Obviously, $\mathbb{E}(\Gamma_n^{(2)}) = 0$. For its variance,

$$
\begin{aligned}
\operatorname{Var}\big(\Gamma_n^{(2)}\big) &\le \frac{4}{nh^{2d+2}} \mathbb{E}\big(Y_1^2\big) \\
&= \frac{4}{nh^{2d+2}} \int_{\mathbb{R}^d} \left[\int_{\mathcal{M}} \langle \nabla K_h(x-z), \mathbb{E}\nabla \widehat{f}(x) - \nabla f(x) \rangle_{M(x)} d\mathscr{H}(x) \right] \\
&\quad \times \left[\int_{\mathcal{M}} \langle \nabla K_h(y-z), \mathbb{E}\nabla \widehat{f}(y) - \nabla f(y) \rangle_{M(y)} d\mathscr{H}(y) \right] f(z)\, dz \\
&= \frac{4}{nh^{d+2}} \int_{\mathcal{M}} \int_{\mathcal{M}} \int_{\mathbb{R}^d} \langle \nabla K(u), \mathbb{E}\nabla \widehat{f}(x) - \nabla f(x) \rangle_{M(x)} \\
&\quad \times \langle \nabla K_h(y - x + hu), \mathbb{E}\nabla \widehat{f}(y) - \nabla f(y) \rangle_{M(y)} f(x-hu)\, du\, d\mathscr{H}(y)\, d\mathscr{H}(x)
\end{aligned}
$$

$$= \frac{4c}{nh^3} \int_{\mathcal{M}} \int_{T_x(\mathcal{M})} \int_{\mathbb{R}^d} \langle \nabla K(u), \mathbb{E}\nabla \widehat{f}(x) - \nabla f(x) \rangle_{M(x)}$$

$$\times \langle \nabla K(v+u), \mathbb{E}\nabla \widehat{f}(x+hv) - \nabla f(x+hv) \rangle_{M(x)}\, du\, d\mathscr{H}(v)\, d\mathscr{H}(x)\big(1+o(1)\big)$$

$$= o\left(\frac{1}{nh^{5-2\nu}}\right),$$

where we have used (4.2) and a similar argument for (4.21). This then implies that $\Gamma_n^{(2)} = o_p(\frac{1}{\sqrt{nh^{5-2\nu}}}) = o_p(\frac{1}{\sqrt{nh}})$. For $\Gamma_n^{(3)}$, we use (4.2) again and have $\Gamma_n^{(3)} = o(h^{2\nu-2})$. Combining the rates for $\Gamma_n^{(2)}$ and $\Gamma_n^{(3)}$ with (4.95), we get (4.79) and hence (4.77). The proof is completed. □

Acknowledgements

The author would like to thank the Associate Editor and the referees for careful reading of the paper and for insightful comments that lead to significant improvements. This work is partially supported by NSF grant DMS 1821154, NSF grant FET 1900061, and the Thomas F. and Kate Miller Jeffress Memorial Trust Award.

References

[1] Ambrosio, L., Colesanti, A. and Villa, E. (2008). Outer Minkowski content for some classes of closed sets. *Math. Ann.* **342** 727–748. MR2443761 https://doi.org/10.1007/s00208-008-0254-z

[2] Arias-Castro, E., Mason, D. and Pelletier, B. (2016). On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. *J. Mach. Learn. Res.* **17** Paper No. 43, 28. MR3491137

[3] Arias-Castro, E. and Rodríguez-Casal, A. (2017). On estimating the perimeter using the alpha-shape. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1051–1068. MR3689960 https://doi.org/10.1214/16-AIHP747

[4] Armendáriz, I., Cuevas, A. and Fraiman, R. (2009). Nonparametric estimation of boundary measures and related functionals: Asymptotic results. *Adv. in Appl. Probab.* **41** 311–322. MR2541178 https://doi.org/10.1239/aap/1246886612

[5] Baddeley, A. and Jensen, E.B.V. (2005). *Stereology for Statisticians. Monographs on Statistics and Applied Probability* **103**. Boca Raton, FL: CRC Press/CRC. MR2107000

[6] Baddeley, A.J., Gundersen, H.J.G. and Cruz-Orive, L.M. (1986). Estimation of surface area from vertical sections. *J. Microsc.* **142** 259–276.

[7] Baíllo, A. (2003). Total error in a plug-in estimator of level sets. *Statist. Probab. Lett.* **65** 411–417. MR2039885 https://doi.org/10.1016/j.spl.2003.08.007

[8] Baíllo, A., Cuevas, A. and Justel, A. (2000). Set estimation and nonparametric detection. *Canad. J. Statist.* **28** 765–782. MR1821433 https://doi.org/10.2307/3315915

[9] Cadre, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.* **97** 999–1023. MR2256570 https://doi.org/10.1016/j.jmva.2005.05.004

[10] Cadre, B., Pelletier, B. and Pudlo, P. (2009). Clustering by estimation of density level sets at a fixed probability. Available at http://hal.archives-ouvertes.fr/docs/00/39/74/37/PDF/tlevel.pdf.

[11] Cannings, T.I., Berrett, T.B. and Samworth, R.J. (2017). Local nearest neighbour classification with applications to semi-supervised learning. Preprint. Available at arXiv:1704.00642.

[12] Canzonieri, V. and Carbone, A. (1998). Clinical and biological applications of image analysis in non-Hodgkin's lymphomas. *Hematol. Oncol.* **16** 15–28. https://doi.org/10.1002/(sici)1099-1069(1998030)16:1<15::aid-hon618>3.0.co;2-j

[13] Caselles, V., Haro, G., Sapiro, G. and Verdera, J. (2008). On geometric variational models for inpainting surface holes. *Comput. Vis. Image Underst.* **111** 351–373.

[14] Chacón, J.E. and Duong, T. (2018). *Multivariate Kernel Smoothing and Its Applications*. *Monographs on Statistics and Applied Probability* **160**. Boca Raton, FL: CRC Press. MR3822372 https://doi.org/10.1201/9780429485572

[15] Chazal, F. (2013). An upper bound for the volume of geodesic balls in submanifolds of Euclidean spaces. Technical report, INRIA.

[16] Chazal, F., Lieutier, A. and Rossignac, J. (2007). Normal-map between normal-compatible manifolds. *Internat. J. Comput. Geom. Appl.* **17** 403–421. MR2362411 https://doi.org/10.1142/S0218195907002422

[17] Chen, Y.-C., Genovese, C.R. and Wasserman, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.* **112** 1684–1696. MR3750891 https://doi.org/10.1080/01621459.2016.1228536

[18] Cuevas, A., Fraiman, R. and Pateiro-López, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.* **44** 311–329. MR2977397 https://doi.org/10.1239/aap/1339878713

[19] Cuevas, A., Fraiman, R. and Rodríguez-Casal, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.* **35** 1031–1051. MR2341697 https://doi.org/10.1214/009053606000001532

[20] Cuevas, A., González-Manteiga, W. and Rodríguez-Casal, A. (2006). Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.* **48** 7–19. MR2234775 https://doi.org/10.1111/j.1467-842X.2006.00421.x

[21] Cuevas, A. and Pateiro-López, B. (2018). Polynomial volume estimation and its applications. *J. Statist. Plann. Inference* **196** 174–184. MR3772767 https://doi.org/10.1016/j.jspi.2017.11.005

[22] Evans, L.C. and Gariepy, R.F. (1992). *Measure Theory and Fine Properties of Functions*. *Studies in Advanced Mathematics*. Boca Raton, FL: CRC Press. MR1158660

[23] Federer, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. MR0110078 https://doi.org/10.2307/1993504

[24] García Trillos, N., Slepčev, D. and von Brecht, J. (2017). Estimating perimeter using graph cuts. *Adv. in Appl. Probab.* **49** 1067–1090. MR3732187 https://doi.org/10.1017/apr.2017.34

[25] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. MR1955344 https://doi.org/10.1016/S0246-0203(02)01128-7

[26] Gokhale, A.M. (1990). Unbiased estimation of curve length in 3-D using vertical slices. *J. Microsc.* **159** 133–141.

[27] Gray, A. (2004). *Tubes*, 2nd ed. *Progress in Mathematics* **221**. Basel: Birkhäuser. With a preface by Vicente Miquel. MR2024928 https://doi.org/10.1007/978-3-0348-7966-8

[28] Guillemin, V. and Pollack, A. (2010). *Differential Topology*. Providence, RI: AMS Chelsea Publishing. MR2680546 https://doi.org/10.1090/chel/370

[29] Hall, P. and Kang, K.-H. (2005). Bandwidth choice for nonparametric classification. *Ann. Statist.* **33** 284–306. MR2157804 https://doi.org/10.1214/009053604000000959

[30] Hall, P. and Murison, R.D. (1993). Correcting the negativity of high-order kernel density estimators. *J. Multivariate Anal.* **47** 103–122. MR1239108 https://doi.org/10.1006/jmva.1993.1073

[31] Hartigan, J.A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267–270. MR0883354

[32] Jiménez, R. and Yukich, J.E. (2011). Nonparametric estimation of surface integrals. *Ann. Statist.* **39** 232–260. MR2797845 https://doi.org/10.1214/10-AOS837

[33] Kerscher, M. (2000). Statistical analysis of large-scale structure in the Universe. In *Statistical Physics and Spatial Statistics*: *The Art of Analyzing and Modeling Spatial Structures and Pattern Formation*. *Lecture Notes in Physics* **554** 36–71. Berlin: Springer. MR1870944 https://doi.org/10.1007/3-540-45043-2_3

[34] Kindlmann, G., Whitaker, R., Tasdizen, T. and Möller, T. (2003). Curvature-based transfer functions for direct volume rendering: Methods and applications. In *Proc. IEEE Visualization* 2003 513–520.

[35] Lee, J.M. (1997). *Riemannian Manifolds*: *An Introduction to Curvature*. *Graduate Texts in Mathematics* **176**. New York: Springer. MR1468735 https://doi.org/10.1007/b98852

[36] Lim, P.H., Bagci, U. and Bai, L. (2013). Introducing Willmore flow into level set segmentation of spinal vertebrae. *IEEE Trans. Biomed. Eng.* **60** 115–122.

[37] Mammen, E. and Polonik, W. (2013). Confidence regions for level sets. *J. Multivariate Anal.* **122** 202–214. MR3189318 https://doi.org/10.1016/j.jmva.2013.07.017

[38] Mammen, E. and Tsybakov, A.B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618 https://doi.org/10.1214/aos/1017939240
[39] Martínez, V.J. and Saar, E. (2001). *Statistics of the Galaxy Distribution*. Roca Raton, FL: CRC Press/CRC.
[40] Mason, D.M. and Polonik, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.* **19** 1108–1142. MR2537201 https://doi.org/10.1214/08-AAP569
[41] Pateiro-López, B. and Rodríguez-Casal, A. (2008). Length and surface area estimation under smoothness restrictions. *Adv. in Appl. Probab.* **40** 348–358. MR2431300
[42] Pateiro-López, B. and Rodríguez-Casal, A. (2009). Surface area estimation under convexity type assumptions. *J. Nonparametr. Stat.* **21** 729–741. MR2549435 https://doi.org/10.1080/10485250902971732
[43] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters—An excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204 https://doi.org/10.1214/aos/1176324626
[44] Pranav, P., Edelsbrunner, H., van de Weygaert, R., Vegter, G., Kerber, M., Jones, B.J.T. and Wintraecken, M. (2017). The topology of the cosmic web in terms of persistent Betti numbers. *Mon. Not. R. Astron. Soc.* **465** 4281–4310.
[45] Qiao, W. (2020). Asymptotics and optimal bandwidth for nonparametric estimation of density level sets. *Electron. J. Stat.* **14** 302–344. MR4048601 https://doi.org/10.1214/19-EJS1668
[46] Qiao, W. (2020). Asymptotic confidence regions for density ridges. Preprint. Available at arXiv:2004.11354.
[47] Qiao, W. and Polonik, W. (2019). Nonparametric confidence regions for level sets: Statistical properties and geometry. *Electron. J. Stat.* **13** 985–1030. MR3934621 https://doi.org/10.1214/19-EJS1543
[48] Rigollet, P. and Vert, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** 1154–1178. MR2597587 https://doi.org/10.3150/09-BEJ184
[49] Rinaldo, A. and Wasserman, L. (2010). Generalized density clustering. *Ann. Statist.* **38** 2678–2722. MR2722453 https://doi.org/10.1214/10-AOS797
[50] Salas, W.A., Boles, S.H., Frolking, S., Xiao, X. and Li, C. (2003). The perimeter/area ratio as an index of misregistration bias in land cover change estimates. *Int. J. Remote Sens.* **24** 1165–1170.
[51] Samworth, R.J. (2012). Optimal weighted nearest neighbour classifiers. *Ann. Statist.* **40** 2733–2763. MR3097618 https://doi.org/10.1214/12-AOS1049
[52] Schmalzing, J. and Górski, K.M. (1998). Minkowski functionals used in the morphological analysis of cosmic microwave background anisotropy maps. *Mon. Not. R. Astron. Soc.* **297** 355–365.
[53] Seifert, U. (1997). Configurations of fluid membranes and vesicles. *Adv. Phys.* **46** 13–137.
[54] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability*. London: CRC Press. MR0848134 https://doi.org/10.1007/978-1-4899-3324-9
[55] Singh, A., Scott, C. and Nowak, R. (2009). Adaptive Hausdorff estimation of density level sets. *Ann. Statist.* **37** 2760–2782. MR2541446 https://doi.org/10.1214/08-AOS661
[56] Sriperumbudur, B. and Steinwart, I. (2012). Consistency and rates for clustering with DBSCAN. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. *PMLR* **22** 1090–1098.
[57] Steinwart, I. (2015). Fully adaptive density-based clustering. *Ann. Statist.* **43** 2132–2167. MR3396981 https://doi.org/10.1214/15-AOS1331
[58] Steinwart, I., Hush, D. and Scovel, C. (2005). A classification framework for anomaly detection. *J. Mach. Learn. Res.* **6** 211–232. MR2249820
[59] Troutt, M.D., Pang, W.K. and Hou, S.H. (2004). *Vertical Density Representation and Its Applications*. River Edge, NJ: World Scientific Co., Inc. MR2051604 https://doi.org/10.1142/9789812562616
[60] Tsybakov, A.B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. MR1447735 https://doi.org/10.1214/aos/1069362732
[61] Walther, G. (1997). Granulometric smoothing. *Ann. Statist.* **25** 2273–2299. MR1604445 https://doi.org/10.1214/aos/1030741072
[62] Weyl, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–472. MR1507388 https://doi.org/10.2307/2371513
[63] Willmore, T.J. (1965). Note on embedded surfaces. *An. Şti. Univ. "Al. I. Cuza" Iaşi Secţ. I a Mat.* (*N.S.*) **11B** 493–496. MR0202066

Received November 2019 and revised April 2020