On the convergence of the estimate of the second fundamental form

(Supplementary material for "Curvature-aware regularization on Riemannian submanifolds")

Kwang In Kim James Tompkin Christian Theobalt

We show the pointwise convergence of the estimate of the second fundamental form (obtained from a point cloud) to the corresponding analytical operator on a manifold as the number of data points grows to infinity. Our proof is based on two assumptions on the regularity of the underlying probability distribution on the manifold M and the boundedness of the corresponding second fundamental form. Please note that we use different notation than in the main paper and so this proof is self-contained.

A toy example. Before we start to discuss the convergence property, we present empirical convergence behavior of our estimate for a toy example. We focus on the estimate of the second fundamental form II at point p where manifold M is given as a hyper-surface in Euclidean space. In this case, locally M can be represented as a graph of a function f, which facilitates the direct comparison between the ground truth II and our estimate $\widehat{II}$. Specifically, we consider a two-dimensional manifold M embedded in $\mathbb{R}^3$, which is given as a graph of f around 0:

$$f(x) = 2x_1^{(3)} - x_2^{(2)} + 0.5x_1x_2, \tag{1}$$

where $x \in \mathbb{R}^2$. The point cloud $\mathcal{X} = \{X_i\}_{i=1}^n \subset \mathbb{R}^3$ is generated by sampling n points from a uniform distribution in an ϵ-neighborhood of 0 in $\mathbb{R}^2$ and evaluating f on them. The error $\mathcal{E}_p$ for an estimate $\widehat{II}_p$ is then calculated by measuring the squared norm of the resulting deviation tensor:

$$\mathcal{E}(\widehat{II}_p) = \|\widehat{II}_p - II_p\|^2_{T_p^*(M)\otimes T_p^*(M)\otimes N_p(M)}, \tag{2}$$

where $T_p^*(M)$ denotes the cotangent space of M at p. We measured the error for $\epsilon = 10, 1, 10^{-1}, 10^{-2}, 10^{-3}$ where n varied accordingly in $10^2, 10^3, 10^4, 10^5, 10^6$. Table 1 summarizes the results of ten different samples of $\mathcal{X}$ for each parameter combination. The error converges toward zero as expected.

	ϵ	10	1	10^{-1}	10^{-2}	10^{-3}
	n	10^2	10^3	10^4	10^5	10^6
Error	Mean	93.03	6.86×10^{-2}	9.22×10^{-5}	1.21×10^{-7}	5.11×10^{-11}
	Std.	56.71	4.21×10^{-2}	7.73×10^{-5}	1.02×10^{-7}	3.91×10^{-11}

Table 1: Estimation error of the second fundamental form for varying ϵ and n (see Eq. 2).

1 Problem statement & proof outline

In Section 3 of the main paper, we adopt an *adapted orthonormal frame* for each p in the ambient manifold $\widetilde{M}$ from which the Riemannian normal coordinates $\{y^i\}$ are constructed. With this, the calculation of the second fundamental form II of M (of dimension d) embedded in $\widetilde{M}$ boils down to the calculation of the Hessians $\{H_{y^i}\}$ of coordinate values $\{y^i\}$ at each p. We assume throughout this document that $\widetilde{M} = \mathbb{R}^{\tilde{d}}$ and so any orthonormal basis in $\mathbb{R}^{\tilde{d}}$ constitutes a normal coordinate system. In particular, our PCA-based coordinate assignments are exact in $\widetilde{M}$. On the other hand, calculating the shape operator explicitly requires the knowledge of the metric g in M (Eq. 9 in the main paper). By introducing the Riemannian normal coordinates x at p in M and accordingly, making

g_p become δ up to the second order, the estimated second fundamental form automatically gives the shape operator given the orthonormal frame in $\mathbb{R}^{\tilde{d}}$ (see "Generalized shape operator" paragraph of the main paper.). However, this introduces an approximation error since our PCA-based coordinate values are, in general, not exact normal coordinates in M.

Suppose that we are given a sample generation process from an underlying probability distribution P on a manifold M such that, at each instance in time, we have a set of data points $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset M \subset \mathbb{R}^{\tilde{d}}$. First, we discuss the convergence of the estimated second fundamental form as $n \to \infty$. Then, the convergence of the estimated shape operator is established by additionally taking into account the approximation error introduced by using PCA coordinates for the normal coordinates x. Since the convergence property is the same for each element of $\{H_{y^i}\}$, we use the symbol f to denote any one element y_i.

At each data point $\mathbf{x}_\alpha \in \mathcal{X}$, the Hessian $H_f|_{\mathbf{x}_\alpha}$ of f is estimated by fitting a quadratic polynomial p_α to $f|_{\mathcal{N}_\epsilon(\mathbf{x}_\alpha)}$, where $\mathcal{N}_\epsilon(\mathbf{x}_\alpha) = \{\mathbf{g}_1, \ldots, \mathbf{g}_l\} := \mathcal{B}(\mathbf{x}_\alpha, \epsilon) \cap \mathcal{X}$, $\mathcal{B}(\mathbf{x}, \epsilon)$ is the ϵ-neighborhood of $\mathbf{x}$,[1] and $h|_\mathcal{S}$ denotes the restriction of a function h on a set $\mathcal{S}$: The Hessian $H_{p_\alpha}|_{\mathbf{x}_\alpha}$ of the polynomial p_α is used as an estimate of $H_f|_{\mathbf{x}_\alpha}$.

The coefficients of polynomial p_α are obtained by solving a weighted least squares problem centered at $\mathbf{x}_\alpha$:

$$\begin{aligned} A_\alpha \approx B_\alpha &= \arg\min_Q \|\mathbf{K}_\alpha(\mathbf{X}_\alpha Q - \mathbf{f})\|^2 \\ &= (\mathbf{X}_\alpha^\top \mathbf{K}_\alpha \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^\top \mathbf{K}_\alpha \mathbf{f}, \end{aligned} \tag{3}$$

where $\mathbf{X}_\alpha$ is the design matrix containing the second-order monomials of the data points in $\mathcal{X}$ centered at $\mathbf{x}_\alpha$ (i.e., each element $\mathbf{x}_i$ of $\mathcal{X}$ is replaced by $\mathbf{x}_i - \mathbf{x}_\alpha$; see Eq. 7):

$$\begin{aligned} A_\alpha &= \frac{1}{2}\left[[H_f|_{\mathbf{x}_\alpha}]_{1,1}, [H_f|_{\mathbf{x}_\alpha}]_{1,2}, \ldots, [H_f|_{\mathbf{x}_\alpha}]_{d,d}\right]^\top, \\ B_\alpha &= \frac{1}{2}\left[[H_p|_{\mathbf{x}_\alpha}]_{1,1}, [H_p|_{\mathbf{x}_\alpha}]_{1,2}, \ldots, [H_p|_{\mathbf{x}_\alpha}]_{d,d}\right]^\top, \\ \mathbf{f} &= [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_l)]^\top, \end{aligned} \tag{4}$$

and $\mathbf{K}_\alpha$ is a diagonal weight matrix with $[\mathbf{K}_\alpha]_{i,i} = K(\mathbf{x}_i - \mathbf{x}_\alpha, \epsilon)$ and the kernel K is defined as:

$$K(\mathbf{x}, h) = \mathbb{1}_{[\|\mathbf{x}\| < h]}. \tag{5}$$

In any coordinate $\{x^i\}$ in M, the zeroth- and the first-order terms of $\{y^i\}$ vanish ($f(\mathbf{x}_\alpha) = 0, \nabla_f|_{\mathbf{x}_\alpha} = 0$) since $\{\frac{\partial}{\partial x^i}\}_{i=1}^d$ spans the tangent space $T_\alpha M$. In this context, the zeroth- and the first-order terms of the fitting polynomial are held fixed at 0.

For notational convenience, henceforth we will assume that the point of evaluation $\mathbf{x}_\alpha$ is 0 unless explicitly stated otherwise and we will omit the index α. All the other locations of interest can be treated in the same way by simply replacing the corresponding locations with the origin.

The point-wise convergence of the second fundamental form is established when $\|A - B\| \to 0$ as $n \to \infty$ and $\epsilon \to 0$. First, we bound it by two multiplicative terms:

$$\|A - B\|^2 \leq \|(\mathbf{X}^\top \mathbf{K}\mathbf{X})^{-1}\|_2 \|\mathbf{K}(\mathbf{X}A - \mathbf{f})\|^2, \tag{6}$$

where the first term depends only on the distribution P on M and it is upper bounded as:

$$\|(\mathbf{X}^\top \mathbf{K}\mathbf{X})^{-1}\|_2 \leq \frac{1}{\|n\epsilon^d \mathcal{E}^{-1}\overline{B}\mathcal{E}^{-1}\|_2} \leq \frac{1}{n\epsilon^{d+4}\lambda_{\overline{B}}}, \tag{7}$$

where:

$$\begin{aligned} \mathcal{E} &= \mathrm{diag}([1/\epsilon^2, \ldots, 1/\epsilon^2]^\top), \\ \overline{B} &= \frac{1}{n\epsilon^d}\sum_{i=1}^n X(\mathbf{x}_i/\epsilon)^\top X(\mathbf{x}_i/\epsilon) K(\mathbf{x}_i, \epsilon), \\ X(\mathbf{x}) &= [\ldots, x^r x^s, \ldots,] \in \mathbb{R}^D (D = \frac{d(d+1)}{2}), \end{aligned}$$

[1]For simplicity, we use the ϵ-neighborhood $\mathcal{B}(\mathbf{x}, \epsilon) := \{\mathbf{y} \in \mathbb{R}^{\tilde{d}} : \|\mathbf{x} - \mathbf{y}\| \leq \epsilon\}$ instead of k nearest neighbors $N_k(\mathbf{x})$. The convergence in the latter case can easily be established by enforcing $N_k(\mathbf{x}) \subset \mathcal{B}(\mathbf{x}, \epsilon)$.

and $\lambda_{\overline{B}}$ is the smallest eigenvalue of $\overline{B}$. In Sec. 3, we quantify $\lambda_{\overline{B}}$ using a certain regularity condition on P: If the distribution P on M satisfies the ***strong density assumption*** [Audibert and Tsybakov, 2007], then with probability larger than $1 - D^2 \exp(-C_2 n\epsilon^d)$ with C_2 being a positive constant, there is a constant $\mu_0 > 0$ such that $\lambda_{\overline{B}} \geq \mu_0$.

Bounding the second term in Eq. 6 requires that the curvature of M is bounded such that the local fitting polynomials constitute good approximations of II in the limit. With this condition, we obtain:

$$\|\mathbf{K}(\mathbf{X}A - \mathbf{f})\|^2 \leq C_1 \gamma^2 l \epsilon^6 \tag{8}$$

with constants $C_1, \gamma > 0$ depending only on the bound on the curvature of M and $l = |\mathcal{N}_\epsilon(0)|$.

Substituting Eq. 7 and Eq. 8 into Eq. 6, we obtain with probability larger than $1 - D^2 \exp(-C_2 n\epsilon^d)$:

$$\|A - B\|^2 \leq \frac{C_1 l \gamma^2}{n\epsilon^{d-2}\mu_0}. \tag{9}$$

This implies that when $\epsilon \to 0$, n grows with a sufficient speed such that $\frac{l}{n\epsilon^d} = \mathcal{O}(1)$ (as shown in Sec. 4), $\|A - B\|^2 \to 0$.

The rest of this document elaborates the construction of the bounds in Eq. 7 and Eq. 8.

2 Bound on $\|\mathbf{K}(\mathbf{X}A - \mathbf{f})\|^2$ (8)

In Section 3 of the main paper, we constructed the coordinates in M using the first d components of the (PCA-based) Riemannian normal coordinates $\{y^1, \ldots, y^{\tilde{d}}\}$ at each point p in $\widetilde{M} = \mathbb{R}^{\tilde{d}}$. Using this coordinate representation and with the *Lipschitz continuity* of the Hessian H_f, we show the pointwise convergence of II. We will use the stronger boundedness condition when we take into account the approximation error caused by using $\{y^1, \ldots, y^d\}$ for the Riemannian normal coordinates in M and show the convergence of the shape operator.

Lemma 1 ([Belward et al., 2008])**.** *Suppose that the Hessian ($H_f(\mathbf{a}) := H_f|_{\mathbf{a}}$) is Lipschitz continuous with the Lipschitz constant γ. Then*

$$\|\mathbf{K}(\mathbf{X}A - \mathbf{f})\|_2^2 = C_1 \gamma^2 l \epsilon^6 \tag{10}$$

with a constant $C_1 > 0$ where l is the size of $\mathcal{N}_\epsilon(0)$.

Proof. Since $\mathcal{N}_\epsilon(0) = \{\mathbf{g}_1, \ldots, \mathbf{g}_l\} = \mathcal{X} \cap \mathcal{B}(0, \epsilon)$, each point $\mathbf{g}_i$ lies in both M and $\mathbb{R}^{\tilde{d}}$. As a point in M, $\mathbf{g}_i$ is assigned with a d-dimensional coordinate values. We represent it with $h_i \mathbf{v}_i$ with $\|\mathbf{v}_i\| = 1$ and $h_i \geq 0$. By construction, $h_i \leq \epsilon$.

Applying the first-order Taylor series remainder formula to f expanded at 0 gives for each point $\mathbf{g}_i$,

$$\begin{aligned} f(h_i \mathbf{v}_i) &= \int_0^1 (1-t) h_i \mathbf{v}_i^\top H_f(h_i \mathbf{v}_i t) h_i \mathbf{v}_i dt, \\ \Leftrightarrow f(h_i \mathbf{v}_i) - \frac{1}{2} h_i \mathbf{v}_i^\top H_f(0) h_i \mathbf{v}_i &= \int_0^1 (1-t) h_i \mathbf{v}_i^\top \left(H_f(h_i \mathbf{v}_i t) - H_f(0)\right) h_i \mathbf{v}_i dt, \end{aligned} \tag{11}$$

where we used the fact that $f(0) = \nabla f|_0 = 0$ when f corresponds to y^i ($i = d+1, \ldots, \tilde{d}$).

Substituting in the definition of A (Equation 4) into (11) gives $[\mathbf{K}(\mathbf{X}A - \mathbf{f})]_i = 0$ when $[\mathbf{K}]_{i,i} = 0$ and

$$\begin{aligned} |[\mathbf{K}(\mathbf{X}A - \mathbf{f})]_i| &= \left| \frac{1}{2} h_i \mathbf{v}_i^\top H_f(0) h_i \mathbf{v}_i - f(h_i \mathbf{v}_i) \right| \\ &\leq \int_0^1 \left| (1-t) h_i \mathbf{v}_i^\top \left(H_f(0) - H_f(h_i \mathbf{v}_i t)\right) h_i \mathbf{v}_i \right| dt \\ &\leq \int_0^1 \left| (1-t) h_i \mathbf{v}_i^\top (\gamma t h_i) h_i \mathbf{v}_i \right| dt \\ &= \frac{1}{6} \gamma h_i^3, \text{ otherewise.} \end{aligned} \tag{12}$$

Then

$$\|\mathbf{K}(\mathbf{X}A - \mathbf{f})\|^2 = \sum_{i=1}^n [\mathbf{K}(\mathbf{X}A - \mathbf{f})]_i^2 \leq \frac{1}{36} l \gamma^2 \epsilon^6, \tag{13}$$

where we used the fact that only l summands are non-zero and $h_i \leq \epsilon$. □

Correction in normal coordinates: convergence of the shape operator estimate (given an orthonormal frame $\{Y_i\}_{i=1}^n$). In general, PCA-based estimates $\{y^1, \ldots, y^d\}$ of the normal coordinate values $\{x^1, \ldots, x^d\}$ at a point in M contain errors of the second order (see the main paper and [Belkin and Niyogi, 2005, Coifman and Lafon, 2006]).[2] Here, we represent the PCA-based estimates and the true Riemannian normal coordinates of $\mathbf{g}_i$ with $\tilde{h}_i\tilde{\mathbf{v}}_i(\|\tilde{\mathbf{v}}_i\| = 1)$ and $h_i\mathbf{v}_i(\|\mathbf{v}_i\| = 1)$, respectively.

Expanding $f(\tilde{h}_i\tilde{\mathbf{v}}_i)$ at 0, we obtain[3]

$$\frac{1}{2}\tilde{h}_i\tilde{\mathbf{v}}_i^\top H_f(0)\tilde{h}_i\tilde{\mathbf{v}}_i - f(\tilde{h}_i\tilde{\mathbf{v}}_i) = \int_0^1 (1-t)\tilde{h}_i\tilde{\mathbf{v}}_i^\top \left(H_f(0) - H_f(\tilde{h}_i\tilde{\mathbf{v}}_i t)\right)\tilde{h}_i\tilde{\mathbf{v}}_i dt,$$

$$\Leftrightarrow \frac{1}{2}\tilde{h}_i\tilde{\mathbf{v}}_i^\top H_f(0)\tilde{h}_i\tilde{\mathbf{v}}_i - f(h_i\mathbf{v}_i) = \int_0^1 (1-t)\tilde{h}_i\tilde{\mathbf{v}}_i^\top \left(H_f(0) - H_f(\tilde{h}_i\tilde{\mathbf{v}}_i t)\right)\tilde{h}_i\tilde{\mathbf{v}}_i dt + \left(f(\tilde{h}_i\tilde{\mathbf{v}}_i) - f(h_i\mathbf{v}_i)\right),$$

$$\Leftrightarrow |[\mathbf{K}(\widetilde{\mathbf{X}}A - \mathbf{f})]_i| \leq \frac{1}{6}\gamma\tilde{h}_i^3 + \left|f(\tilde{h}_i\tilde{\mathbf{v}}_i) - f(h_i\mathbf{v}_i)\right|, \tag{14}$$

where the approximate design matrix $\widetilde{\mathbf{X}}$ is constructed based on the PCA-based estimates of normal coordinate values. Expanding $f(h_i\mathbf{v}_i)$ and $f(\tilde{h}_i\tilde{\mathbf{v}}_i)$ at 0, we obtain

$$f(\tilde{h}_i\tilde{\mathbf{v}}_i) = \int_0^1 (1-t)\tilde{h}_i\tilde{\mathbf{v}}_i^\top \left(H_f(\tilde{h}_i\tilde{\mathbf{v}}_i t) - H_f(0)\right)\tilde{h}_i\tilde{\mathbf{v}}_i dt + \int_0^1 (1-t)\tilde{h}_i\tilde{\mathbf{v}}_i^\top H_f(0)\tilde{h}_i\tilde{\mathbf{v}}_i dt, \tag{15}$$

$$f(h_i\mathbf{v}_i) = \int_0^1 (1-t)h_i\mathbf{v}_i^\top \left(H_f(h_i\mathbf{v}_i t) - H_f(0)\right)h_i\mathbf{v}_i dt + \int_0^1 (1-t)h_i\mathbf{v}_i^\top H_f(0)h_i\mathbf{v}_i dt. \tag{16}$$

With the boundedness of the second fundamental form (in the second summands) and (12), (15) and (16) imply that there is a constant η such that

$$|f(h_i\mathbf{v}_i) - f(\tilde{h}_i\tilde{\mathbf{v}}_i)| \leq \frac{\eta}{2}|h_i^2 - \tilde{h}_i^2| + \frac{\gamma}{6}|h_i^2 - \tilde{h}_i^2|. \tag{17}$$

Substituting this into (14) and noting that $h_i^2 = \tilde{h}_i^2 + \mathcal{O}(h_i^4)$ ([Belkin and Niyogi, 2005, Coifman and Lafon, 2006]), we obtain

$$|[\mathbf{K}(\widetilde{\mathbf{X}}A - \mathbf{f})]_i| = \mathcal{O}(h_i^3). \tag{18}$$

Accordingly, $\|\mathbf{K}(\widetilde{\mathbf{X}}A - \mathbf{f})\|^2$ remains $\mathcal{O}(\epsilon^6)$ since $h_i \leq \epsilon$.

3 Bound on $\lambda_{\overline{B}}$

Here, we adopt the results of [Audibert and Tsybakov, 2007] to construct a lower bound of $\lambda_{\overline{B}}$. Applying this result requires a certain regularity assumption on the underlying probability distribution P on M.

For some constants $c_0, r_0 > 0$, we will say that a Lebesgue measurable set $A \subset \mathbb{R}^d$ is (c_0, r_0)-*regular* if

$$\lambda[A \cap \mathcal{B}(\mathbf{x}, r)] \geq c_0\lambda[\mathcal{B}(\mathbf{x}, r)], \ \forall r \in [0, r_0], \forall \mathbf{x} \in A, \tag{19}$$

where $\lambda[S]$ stands for the Lebesgue measure of $S \subset \mathbb{R}^d$. We fix constants $c_0, r_0 > 0$ and $0 < \mu_{\min} < \mu_{\max} < \infty$ and a compact $\mathcal{C} \subset \mathbb{R}^d$. We say that the *strong density assumption* is satisfied if the distribution P is supported on a compact (c_0, r_0)-regular set $A \subseteq \mathcal{C}$ and has a density μ w.r.t. the Lebesgue measure bounded away from zero and infinity on A (between μ_{min} and μ_{max})

$$\mu_{\min} \leq \mu(\mathbf{x}) \leq \mu_{\max}, \ \forall \mathbf{x} \in A \text{ and } \mu(\mathbf{x}) = 0 \text{ otherwise}. \tag{20}$$

Theorem 1 ([Audibert and Tsybakov, 2007])**.** *Let P satisfies the strong density assumption. Then, there exist constants $C_2, \mu_0 > 0$ such that for any $0 < \epsilon \leq r_0$ and any $n \geq 1$,*

$$P^{\otimes n}(\lambda_{\overline{B}} \leq \mu_0) \leq 2D^2 \exp(-C_2 n\epsilon^d), \tag{21}$$

where $D = \frac{d(d+1)}{2}$ and $P^{\otimes n}$ is the product probability measure according to which the sample is distributed.

[2] The *injectivity radius* $\text{inj}(\mathbf{x}_\alpha)$ of $\mathbf{x}_\alpha \in \mathcal{X}$ is always positive [Klingenberg, 1982]. Here, we assume that $\epsilon(\mathbf{x}_\alpha) \leq \text{inj}(\mathbf{x}_\alpha)$.

[3] It should be noted that although we are constructing the series expansion based on empirical estimates (i.e., with $\tilde{h}_i\tilde{\mathbf{v}}_i$ instead of $h_i\mathbf{v}_i$), the corresponding function is evaluated at $\mathbf{g}_i \in M$ and accordingly, it should be written as $f(h\mathbf{v})$ rather than $f(\tilde{h}\tilde{\mathbf{v}})$.

It should be noted that in the current context, $n = |\mathcal{X}|$ while $l(n, \epsilon) = |\mathcal{N}_\epsilon(0)|$.

Proof. Let A be the support of P, which contains $\mathbf{x}_\alpha$.[4] Consider the matrix $B(\epsilon) := [\widetilde{B}_{(j,k)}]_{D,D} = [\int_{\|\mathbf{u}\|<1} \left[X(\mathbf{u})^\top X(\mathbf{u})\right]_{(j,k)} \mu(\epsilon\mathbf{u})d\mathbf{u}]_{D,D}$.

The smallest eigenvalue $\lambda_{\overline{B}}$ of $\overline{B}$ satisfies

$$\begin{aligned}\lambda_{\overline{B}} &= \min_{\|W\|=1} W^\top \overline{B} W \\ &\geq \min_{\|W\|=1} W^\top B W + \min_{\|W\|=1} W^\top (\overline{B} - B) W \\ &\geq \min_{\|W\|=1} W^\top B W - \sum_{j,k} |\overline{B}_{j,k} - B_{j,k}|. \end{aligned} \tag{22}$$

Let $A_n := \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| \leq 1; \epsilon\mathbf{u} \in A\}$. For any vector W satisfying $\|W\| = 1$, we obtain

$$\begin{aligned} W^\top B W &= \int_{\|\mathbf{u}\|<1} W^\top X(\mathbf{u})^\top X(\mathbf{u}) W \mu(\epsilon\mathbf{u}) d\mathbf{u} \\ &\geq \mu_{\min} \int_{A_n} W^\top X(\mathbf{u})^\top X(\mathbf{u}) W d\mathbf{u}. \end{aligned} \tag{23}$$

By assumption of the theorem, $\epsilon \leq r_0$. Since the support of P is (c_0, r_0)-regular, we get

$$\lambda[A_n] \geq \epsilon^{-d} \lambda[\mathcal{B}(0, \epsilon) \cap A] \geq c_0 \epsilon^{-d} \lambda[\mathcal{B}(0, \epsilon)] = c_0 \upsilon_d,$$

where $\upsilon_d = \lambda[\mathcal{B}(0, 1)]$ is the volume of the unit ball and $c_0 > 0$ is the constant of the (c_0, r_0)-regular set. Let $\mathcal{A}$ denote the class of all compact subsets of $\mathcal{B}(0, 1)$ having Lebesgue measure $c_0 \upsilon_d$. Using the previous displays we obtain

$$\min_{\|W\|=1} W^\top B W \geq \mu_{\min} \min_{\|W\|=1; S \in \mathcal{A}} \int_S W^\top X(\mathbf{u})^\top X(\mathbf{u}) W d\mathbf{u} := 2\mu_0. \tag{24}$$

By the compactness argument, the minimum in (24) exists and is strictly positive.

For $i = 1, \ldots, n$ and any indices (j, k), define

$$\begin{aligned} T_i(j,k) &:= \frac{1}{\epsilon^d} \left[X(\mathbf{x}_i/\epsilon)^\top X(\mathbf{x}_i/\epsilon) K(\mathbf{x}_i, \epsilon)\right]_{(j,k)} - B_{(j,k)} \\ &= \frac{1}{\epsilon^d} \left[X(\mathbf{x}_i/\epsilon)^\top X(\mathbf{x}_i/\epsilon) K(\mathbf{x}_i, \epsilon)\right]_{(j,k)} - \int_{\|\mathbf{u}\|<1} \left(X(\mathbf{u})^\top X(\mathbf{u})\right)_{(j,k)} \mu(\epsilon\mathbf{u}) d\mathbf{u}. \end{aligned} \tag{25}$$

Since for a vector $\|\mathbf{u}\| < 1$ and for any (j, k)

$$\left| \left[X(\mathbf{u})^\top X(\mathbf{u})\right]_{(j,k)} \right| \leq 1, \tag{26}$$

we have expectation $\mathbb{E}[T_i] = 0$, $|T_i| < \frac{2}{\epsilon^d}$, and the following bound for the variance of T_i:[5]

$$\begin{aligned} \mathbb{V}\mathrm{ar}[T_i(j,k)] &\leq \frac{1}{\epsilon^{2d}} \mathbb{E}[\left[X(\mathbf{x}_i/\epsilon)^\top X(\mathbf{x}_i/\epsilon)\right]_{(j,k)}^2 K(\mathbf{x}_i, \epsilon)] \\ &= \frac{1}{\epsilon^d} \int_{\|\mathbf{u}\|<1} \left[X(\mathbf{u})^\top X(\mathbf{u})\right]_{(j,k)}^2 \mu(\epsilon\mathbf{u}) d\mathbf{u} \\ &\leq \frac{\mu_{\max}}{\epsilon^d}. \end{aligned} \tag{27}$$

Using Bernstein's inequality, for any $\rho > 0$, we have

$$P^{\otimes n}(|\overline{B}_{j,k} - B_{j,k}| > \rho) = P^{\otimes n} \left(\left| \frac{1}{n} \sum_{i=1}^n T_i(j,k) \right| > \rho \right) \leq 2 \exp\left(-\frac{n\epsilon^d \rho^2}{2\mu_{\max} + 4\rho/3} \right).$$

This and (22) and (24) imply the claim of the theorem. □

[4]Here, we assume that the point of interest $\mathbf{x}_\alpha$ is contained in A. Otherwise, no sample will be generated at $\mathbf{x}_\alpha$ and accordingly, there's no need to consider this case. Without loss of generality, we will again assume that $\mathbf{x}_\alpha = 0$.

[5]Here, the expectation is taken with respect to the distribution P which is normalized by the measure of $\mathcal{B}(\mathbf{x}, \epsilon)$ (i.e., the distribution is conditioned upon the fact that the events are occurring within $\mathcal{B}(\mathbf{x}, \epsilon)$).

4 Bounding $\frac{l}{n\epsilon^d}$

If we adopt the *strong density assumption* (see Sec. 3), the probability P_ϵ of sampling a data point from the ϵ-neighborhood of $\mathbf{x}_\alpha$ (which is assumed to be zero) is

$$\begin{aligned} P_\epsilon &= \int_A \mu(\mathbf{x}) \mathbb{1}_{[\|\mathbf{x}-\mathbf{x}_\alpha\|<\epsilon]} d\mathbf{x} \\ &\leq \mu_{\max} \int_A \mathbb{1}_{[\|\mathbf{x}-\mathbf{x}_\alpha\|<\epsilon]} d\mathbf{x} \\ &= \mu_{\max} \int_{\mathbb{R}^d} \mathbb{1}_{[\|\mathbf{x}-\mathbf{x}_\alpha\|<\epsilon]} d\mathbf{x} \\ &= \mu_{\max} \upsilon_d \epsilon^d, \end{aligned} \tag{28}$$

where $\upsilon_d = \lambda[\mathcal{B}(0,1)]$ and A is the support of P.

Let's define variables $\{\mathbb{1}_\epsilon(i)\}$

$$\mathbb{1}_\epsilon(i) = \begin{cases} 1 & \text{if } \mathbf{x}_i \in \mathcal{N}_\epsilon(\mathbf{x}_\alpha) \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

Applying Hoeffding's inequality to $\{\mathbb{1}_\epsilon(1), \ldots, \mathbb{1}_\epsilon(n)\}$ yields

$$P\left(\sum_{i=1}^{n} \mathbb{1}_\epsilon(i) - nP_\epsilon \geq t\right) \leq \exp\left(-\frac{2t^2}{n}\right). \tag{30}$$

Substituting (28) into (30) we obtain

$$P\left(l - (\mu_{\max}\upsilon_d) n\epsilon^d \geq t\right) \leq \exp\left(-\frac{2t^2}{n}\right), \tag{31}$$

which states that $\frac{l}{n\epsilon^d} = \mathcal{O}(1)$ and proves that the deviation of empirical Hessian from the true Hessian in (9) converges to zero.

Acknowledgement

The ideas presented in this study have greatly profited from discussions with Matthias Hein.

References

[Audibert and Tsybakov, 2007] Audibert, J.-Y. and Tsybakov, A. B. (2007). Fast learning rates for plug-in classifiers. *The Annals of Statistics*, 35(2):608–633.

[Belkin and Niyogi, 2005] Belkin, M. and Niyogi, P. (2005). Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308.

[Belward et al., 2008] Belward, J. A., Turner, I. W., and Oqielat, M. N. (2008). Numerical investigations of linear least squares methods for derivative estimation. *ANZIAM Journal*, 50:C844–C857.

[Coifman and Lafon, 2006] Coifman, R. R. and Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 20(1):5–30.

[Klingenberg, 1982] Klingenberg, W. (1982). *Riemannian Geometry*. de Gruyter, Berlin.