

From Small Scales to Large Scales: Distance-to-Measure Density based Geometric Analysis of Complex Data

Katharina Proksch*

K.PROKSCH@UTWENTE.NL

*Faculty of Electrical Engineering, Mathematics & Computer Science
University of Twente
Hallenweg 19, 7522NH Enschede, Netherlands*

Christoph A. Weitkamp*

CWEITKA@MATHEMATIK.UNI-GOETTINGEN.DE

*Institute for Mathematical Stochastics
University of Göttingen
Goldschmidtstraße 7, 37077 Göttingen, Germany*

Thomas Staudt

THOMAS.STAUDT@UNI-GOETTINGEN.DE

*Institute for Mathematical Stochastics
University of Göttingen
Goldschmidtstraße 7, 37077 Göttingen, Germany*

Benoît Lelandais

BENOIT.LELANDAIS@PASTEUR.FR

*Institut Pasteur
Imaging and Modeling Unit
UMR 3691, CNRS, Paris, France*

Christophe Zimmer

CZIMMER@PASTEUR.FR

*Institut Pasteur
Imaging and Modeling Unit
UMR 3691, CNRS, Paris, France*

Editor: Florence d'Alche-Buc

Abstract

How can we tell complex point clouds with different small scale characteristics apart, while disregarding global features? Can we find a suitable transformation of such data in a way that allows to discriminate between differences in this sense with statistical guarantees?

In this paper, we consider the analysis and classification of complex point clouds as they are obtained, e.g., via single molecule localization microscopy. We focus on the task of identifying differences between noisy point clouds based on small scale characteristics, while disregarding large scale information such as overall size. We propose an approach based on a transformation of the data via the so-called Distance-to-Measure (DTM) function, a transformation which is based on the average of nearest neighbor distances. For each data set, we estimate the probability density of average local distances of all data points and use the estimated densities for classification. While the applicability is immediate and the practical performance of the proposed methodology is very good, the theoretical study of the density estimators is quite challenging, as they are based on *non-i.i.d.* observations that have been obtained via a complicated transformation. In fact, the transformed data are stochastically dependent in a non-local way that is not captured by commonly considered dependence measures. Nonetheless, we show that the asymptotic behaviour of the

*. Katharina Proksch and Christoph Weitkamp contributed equally to this work.

density estimator is driven by a kernel density estimator of certain i.i.d. random variables by using theoretical properties of U -statistics, which allows to handle the dependencies via a Hoeffding decomposition. We show via a numerical study and in an application to simulated single molecule localization microscopy data of chromatin fibers that unsupervised classification tasks based on estimated DTM-densities achieve excellent separation results.

Keywords: geometric data analysis, distance-to-measure signature, kernel density estimators, nearest neighbor distributions

1. Introduction

The analysis and extraction of information from complex point clouds has become a main task in many applications. Prominent examples can be found in geomorphology, where structure in point-clouds obtained from laser scanners is investigated to infer the shape of the Earth (Vosselman et al., 2004; Hayakawa and Oguchi, 2016), or in cosmology, where the Cosmic Web is analysed based on a discrete set of points from N -body simulations or galaxy studies (Libeskind et al., 2017). Related questions also arise in biology, when data from single molecule localization microscopy (SMLM), which is based on the localization of fluorescent molecules that appear at different times, are analyzed (Nicovich et al., 2017; Lelek et al., 2021). Data obtained in SMLM are 2D or 3D point clouds, where the points correspond to particular molecular localization events. In this paper, we consider a specific example which is related to the analysis of super-resolution visualization of human chromosomal regions as it has recently been investigated in Hao et al. (2021). In this application, the goal is to better understand the 3D organization of the chromatin fiber in cell nuclei, which plays a key role in the regulation of gene expression.

In all aforementioned examples, it is important to identify significant differences between noisy point clouds, where a focus is on general structure and small scale information rather than on global features such as the overall shape of a point cloud.

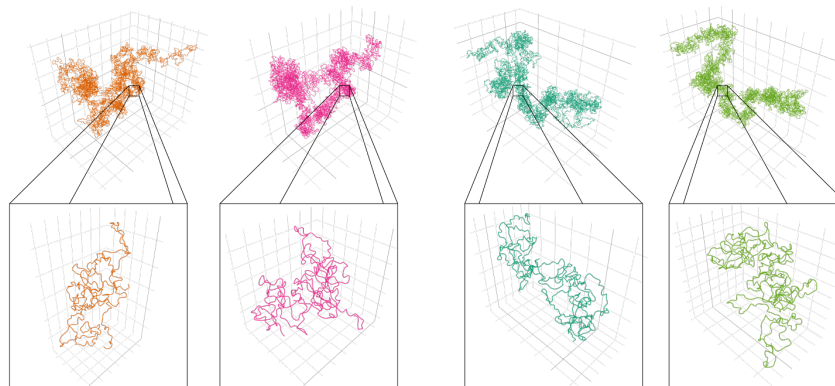


Figure 1: **Example Data:** Four different simulated chromatin fibers in two different conditions: Condition A (orange (far left) and blue-green (middle right)) and Condition B (pink (middle left) and green (far right)) for the purpose of comparison.

For illustration, Figure 1 shows four simulated chromatin fibers in two different conditions. The displayed structures form loops of different sizes and frequencies, based on the condition under which they were simulated. The differences between the conditions are so subtle that they are not clearly distinguishable by eye. In the application considered in this paper, we analyse noisy samples of such simulated structures. The noise accounts for localization errors as they are present in real SMLM data. The loops are of sizes comparable to the resolution of the images (see Section 4 and Hao et al. (2021) for more details), which makes the problem tractable but difficult. The aim is to classify the point clouds based on their loop distribution (i.e., based on their small scale characteristics), while disregarding their total size or other large scale features. It is natural to transform such complicated data prior to the analysis, in particular when one has a clear objective in mind. In the above reference, the statistical analysis of the simulated and real data was based on a transformation of each data cloud onto a set of two parameters, capturing smoothness and local curvature of the point clouds. While this transformation provided a clear discrimination between different groups, the amount of information preserved in a two-dimensional parameter is not sufficient as a basis for point-by-point classification. In this paper, we propose an approach which is similar in spirit, but which provides a transformation into a curve, with different characteristics for the different conditions. In our analysis, the whole curves are then used as features. To this end, we perform the following two steps.

- (i) A transformation of the point cloud based on the *Distance-to-Measure (DTM) signature* (Chazal et al., 2016, 2017) to a one-dimensional data set.
- (ii) The analysis of the distribution of the DTM-transformed data via their estimated probability density.

The DTM signature is closely related to certain nearest neighbor distributions, which makes this approach very intuitive. In particular, this framework allows for a comprehensive exploratory analysis of complex data, for which we might seek a simple graphical representation that captures and summarizes the local structural information well.

1.1 The DTM-Density as a Representation for Local Features

We now introduce the statistical framework of the paper and carefully define the previously mentioned DTM-signature. Throughout the following, we consider random point clouds as samples from a *Euclidean metric measure space* $\mathcal{X} = (\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$, i.e., a triple, where $\mathcal{X} \subset \mathbb{R}^d$ denotes a compact set, $\|\cdot\|$ stands for the Euclidean distance and $\mu_{\mathcal{X}}$ denotes a probability measure that is fully supported on the compact set \mathcal{X} . If, additionally, $\mu_{\mathcal{X}}$ has a Lipschitz continuous density with respect to the d -dimensional Lebesgue measure, then we call \mathcal{X} a *regular Euclidean metric measure space*. For a metric measure space \mathcal{X} , we define the corresponding *Distance-to-Measure (DTM) function* with *mass parameter* $m \in (0, 1]$ for $x \in \mathbb{R}^d$ as

$$d_{\mathcal{X},m}^2(x) = \frac{1}{m} \int_0^m F_x^{-1}(u) du, \quad (1)$$

where $F_x(t) = P(\|X - x\|^2 \leq t)$, $X \sim \mu_{\mathcal{X}}$, and F_x^{-1} denotes the corresponding quantile function. The DTM-function, which is essential for the definition of the DTM-signature, is a population quantity that is generally unknown in practice and thus has to be estimated

from the data. In order to do so, we replace the quantile function in definition (1) by its empirical version as follows. Let $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and denote the corresponding empirical measure by $\hat{\mu}_{\mathcal{X}}$. We define for $t \geq 0$

$$\hat{F}_{x,n}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\|x - X_i\|^2 \leq t\}} \quad (2)$$

and denote by $\hat{F}_{x,n}^{-1}$ the corresponding quantile function, giving rise to a plug-in estimator $\delta_{\mathcal{X},m}^2(x)$ for the Distance-to-Measure function $d_{\mathcal{X},m}^2(x)$:

$$\delta_{\mathcal{X},m}^2(x) = \frac{1}{m} \int_0^m \hat{F}_{x,n}^{-1}(u) du. \quad (3)$$

In the special case that $m = \frac{k}{n}$, it is possible to rewrite (3) as a *nearest neighbor statistic* as follows

$$\delta_{\mathcal{X},m}^2(x) = \frac{1}{k} \sum_{X_i \in N_k(x)} \|X_i - x\|^2, \quad (4)$$

where $N_k(x)$ is the set containing the k nearest neighbors of x among the data points X_1, \dots, X_n .

As discussed previously, we require a good descriptor for the small scale behavior of our data. Hence, in a similar spirit as Br echeteau (2019), we reduce the potentially complex Euclidean metric measure space to a one-dimensional probability distribution by considering the *Distance-to-Measure (DTM) signature* $d_{\mathcal{X},m}^2(X)$, where $X \sim \mu_{\mathcal{X}}$. That is, the deterministic point $x \in \mathcal{X}$ is replaced by the random variable X representing our observations. The distribution of $d_{\mathcal{X},m}^2(X)$ captures the relative frequency of the mean of the distances of a random point in \mathcal{X} to its “ $m \cdot 100\%$ nearest neighbors”. We will empirically illustrate that the distribution of $d_{\mathcal{X},m}^2(X)$ is a good descriptor for the small scale behavior of the considered data for small values of m and verify that it is well-suited for chromatin loop analysis. Furthermore, it is easy to see that for $m = 1$ the random quantity $d_{\mathcal{X},1}^2(X)$ is closely related to the lower bound FLB_p of the Gromov-Wasserstein distance defined in M emoli (2011):

$$\text{FLB}_p = \frac{1}{2} \inf_{\mu \in \mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \left(\int_{\mathcal{X} \times \mathcal{Y}} |s_{\mathcal{X},p}(x) - s_{\mathcal{Y},p}(y)|^p \mu(dx, dy) \right)^{\frac{1}{p}},$$

where $s_{\mathcal{X},p} = \|d_{\mathcal{X}}(x, \cdot)\|_{L^p(\mu_{\mathcal{X}})}$ and $\mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ denotes the set of all couplings of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$. This motivates why $d_{\mathcal{X},1}^2(X)$ is well suited for object discrimination with a focus on large scale characteristics. Although this case is not of interest in our specific data example, we include it in our analysis, since variants of $d_{\mathcal{X},1}^2(X)$ have been proven very useful for pose invariant object discrimination when large scale differences are crucial (Hamza and Krim, 2003; Gellert et al., 2019), see also Section 3.2 for an example.

Since we propose to reduce (possibly complex) multi-dimensional metric measure spaces to a one-dimensional probability distribution, the next step is to visualize and investigate

these distributions. It is well known that probability densities (if they exist) can provide a useful visual insight into the probability distributions considered. In this regard, they are usually better suited than cumulative distribution functions (see, e.g., Chen and Pokojovy (2018)). Therefore, we focus on the estimation of the density of $d_{\mathcal{X},m}^2(X)$ in this paper. A natural estimator for the density of $d_{\mathcal{X},m}^2(X)$, in the following denoted as *DTM-density*, in case of a known DTM-function, is given by

$$\widehat{f}_{d_{\mathcal{X},m}^2}(y) = \frac{1}{nh} \sum_{i=1}^n K \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right). \quad (5)$$

However, since $\mu_{\mathcal{X}}$ is unknown, we cannot compute $d_{\mathcal{X},m}^2$ in practice and consequently it is generally not feasible to estimate $f_{d_{\mathcal{X},m}^2}$ via $\widehat{f}_{d_{\mathcal{X},m}^2}$. Instead, we propose to replace $d_{\mathcal{X},m}^2$ by its empirical version $\delta_{\mathcal{X},m}^2$ and estimate $f_{d_{\mathcal{X},m}^2}$ based on the plug-in estimator

$$\widehat{f}_{\delta_{\mathcal{X},m}^2}(y) = \frac{1}{nh} \sum_{i=1}^n K \left(\frac{\delta_{\mathcal{X},m}^2(X_i) - y}{h} \right). \quad (6)$$

It is important to note that, in contrast to $\widehat{f}_{d_{\mathcal{X},m}^2}$, the plug-in estimator $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ is based on the *non-i.i.d.* observations $\delta_{\mathcal{X},m}^2(X_1), \dots, \delta_{\mathcal{X},m}^2(X_n)$. In fact, for each $i \neq j$, $\delta_{\mathcal{X},m}^2(X_i)$ and $\delta_{\mathcal{X},m}^2(X_j)$ are stochastically dependent. The asymptotic behaviour of kernel density estimators under dependence has been studied extensively in the literature for various mixing and linear processes connected to weakly dependent time series (Castellana and Leadbetter, 1986; Robinson, 1983; Liebscher, 1996; Lu, 2001; Wu and Mielniczuk, 2002). In all these settings, results on asymptotic normality similar to the i.i.d. case can be derived. Related results for spatial processes can be found, e.g., in Hallin et al. (2004). For *long-range dependent* data, the asymptotic behaviour of kernel density estimators changes drastically. Here, the empirical density process (based on kernel estimators of the marginal densities) converges weakly to a tight limit (see Csörgő and Mielniczuk (1995)). For the sequence $\delta_{\mathcal{X},m}^2(X_1), \dots, \delta_{\mathcal{X},m}^2(X_n)$, however, a structure as in the above examples (in space or time) is not given. For each $i \neq j$, $\delta_{\mathcal{X},m}^2(X_i)$ and $\delta_{\mathcal{X},m}^2(X_j)$ are stochastically dependent in a way that is not captured by the dependency models considered in the literature discussed above. Curiously perhaps, $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ allows a decomposition into a non-degenerate U -statistic $U_n(X_1, \dots, X_n)$ of order 2 and negligible remainder terms (Step 2 in the proof of Theorem 7, see Lemma C.2 in Section C.1). U -statistics are averages of symmetric functions *applied to all tuples* of fixed size (here 2) in the sample and are a concept from classical mathematical statistics due to Hoeffding (1948). By proper projection we can find an *i.i.d.* approximation to $U_n(X_1, \dots, X_n)$ which drives the asymptotics and ensures a Gaussian limit (Step 3 in the proof of Theorem 7, see Lemma C.4 in Section C.1, see also Chapter 12.1 in Van der Vaart (2000) for mathematical properties of U -statistics in general).

1.2 Main Results

The main theoretical contribution of the paper is the distributional limit of the kernel density estimator defined in (6). More precisely, we prove (cf. Theorem 7), given certain

regularity conditions on $f_{d_{\mathcal{X},m}^2}$, $d_{\mathcal{X},m}^2(y)$ and \mathcal{X} , (see Condition 2 in Section 2.1) that for $n \rightarrow \infty$, $h = o(n^{-1/5})$ and $nh \rightarrow \infty$

$$\sqrt{nh} \left(\widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - f_{d_{\mathcal{X},m}^2}(y) \right) \Rightarrow N \left(0, f_{d_{\mathcal{X},m}^2}(y) \int K^2(u) du \right). \quad (7)$$

This means that, although the kernel density estimator $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ is based on transformed, dependent random variables, asymptotically, it behaves precisely as the inaccessible kernel density estimator $\widehat{f}_{d_{\mathcal{X},m}^2}$ based on independent random variables. This entails that many methods which are feasible for kernel density estimators based on i.i.d. data, can be applied in this much more complex setting as well, with the same asymptotic justification.

1.3 Application

Chromosomes, which consist of chromatin fibres, are essential parts of cell nuclei in human beings and carry the genetic information important for heredity transmission. It is known by now that there are small scale self-interacting genomic regions, so called topologically associating domains (TADs) which are often associated with loops in the chromatin fibers (Nuebler et al., 2018). As an application, we consider chromatin loop analysis, one aspect of which is to study the presence or absence of loops in the chromatin (see Section 4).

The local loop structure is very well characterized by local nearest neighbor means as illustrated on the right of Figure 2 and hence we propose to use DTM-signatures for tackling this issue. Figure 2 shows the pipeline for the data transformation (left) and the resulting kernel density estimators ($m = 1/250$, biweight kernel, bandwidth selection as in Section 4) for the four data sets shown in Figure 1 (right of Figure 2, same coloring). It shows that the kernel density estimators mainly differ between the different conditions and not between the corresponding chromatin fibers and that the differences between the conditions are clearly pronounced. This demonstrates that the transformation is well suited for a qualitative analysis of the data.

1.4 Related Work

The use of the DTM-signature for the purpose of pose invariant object discrimination was proposed by Bréchet (2019), who in particular established a relation between the DTM-signature and the Gromov-Wasserstein distance (see Mémoli (2011) for a definition). In the aforementioned work, the author considers the asymptotic behavior of the Wasserstein distance between sub-sampled estimates of the DTM-signatures for two different spaces. While the approaches are related, one big advantage of our method of estimating the DTM-densities is that it does not require sub-sampling and all data points can be used for the analysis, guaranteeing a powerful procedure without loss of information.

As illustrated in Section 1.1, the DTM-signature is based on the DTM-function (see Eq. 1). This function has been thoroughly studied and applied in the context of support estimation and topological data analysis (Chazal et al., 2011, 2013; Buchet et al., 2014) and for its sample counterpart (see Eq. 3) many consistency properties have been established in Chazal et al. (2016, 2017).

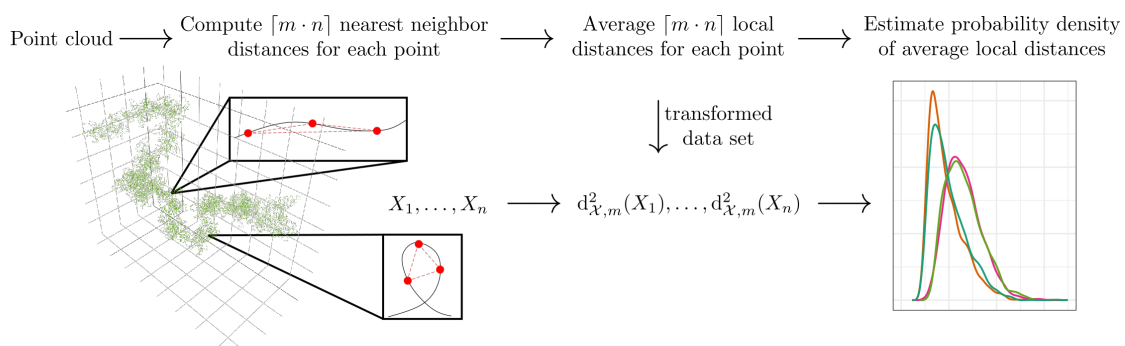


Figure 2: **Data analysis pipeline:** Illustration of the different steps in the proposed data analysis. The red dots in the details of the image represent data points, the red lines show the point-to-point distances, whereas the underlying chromatin structure is depicted by a black line. Right: The resulting DTM-density estimates of the point clouds illustrated in Figure 1 (same coloring).

Distance based signatures for object discrimination have been applied and studied in a variety of settings (Osada et al., 2002; Gelfand et al., 2005; Belongie et al., 2006; Shi et al., 2007; Brinkman and Olver, 2012; Berrendero et al., 2016). Recently, lower bounds of the Gromov-Wasserstein distance (see Mémoli (2011)) have received some attention in applications (Gellert et al., 2019) and in the investigation of their discriminating properties and their statistical behavior (Mémoli and Needham, 2022; Weitkamp et al., 2024).

Furthermore, it is noteworthy that nearest neighbor distributions are of great interest in various fields in biology (Zou and Wu, 1995; Meng et al., 2020) as well as in physics (Torquato et al., 1990; Bhattacharjee, 2003; Hsiao et al., 2020). In these fields it is quite common to consider the (mean of the) distribution of all nearest neighbors for data analysis. This case corresponds to $m = 1/n$ in our method (see Remark 8 for a discussion of different types of sampling mechanisms). We would like to emphasize that taking the mean over a certain percentage of nearest neighbors makes our method a lot more robust against noise, leading to a more reliable performance in the analysis of noisy point clouds.

In the analysis of SMLM images, methods from spatial statistics are often employed. Related to the global distribution of all distances is Ripley’s K, which is used to infer the amount and the degree of clustering in a given data set as compared to a point cloud generated by a homogeneous Poisson point process (see, e.g., Nicovich et al. (2017) for the application of Ripley’s K in this context). Despite the connection via certain distributions of distances, the objectives and underlying models are quite different to the setting of this paper, such that a direct comparison is not sensible.

Kernel density estimation from dependent data is a broad and well investigated topic. In addition to the references provided in Section 1.1, kernel density estimators of symmetric functions of the data and dyadic undirected data have been considered (Frees, 1994; Graham et al., 2022). In these settings, the summands of the corresponding kernel density estimators admit a “ U -statistic like” dependency structure that has to be accounted for.

While this is more closely related to the dependency structure which we are encountering in our analysis, the structure of the statistics that appear in the decomposition of the kernel density estimator (6) is quite different, such that those results cannot directly be transferred to our setting.

1.5 Organization of the Paper

In Section 2 we state the main results and are concerned with the derivation of distributional limit (7) and the assumptions required. Afterwards, in Section 3 we illustrate our findings via simulations. In Section 4, we apply our methodology to the classification within the framework of chromatin loop analysis.

1.6 Notation:

Throughout the following, we denote the d -dimensional Lebesgue measure by λ^d and the $(d-1)$ -dimensional surface measure in \mathbb{R}^d by σ^{d-1} . We write $B(x, r)$ for the open ball in \mathbb{R}^d (equipped with $\|\cdot\|$) with center x and radius r . Given a function f or a measure μ , we write $\text{supp}(f)$ and $\text{supp}(\mu)$ to denote their respective support. Let F be a distribution function with compact support $[a, b]$ and let F^{-1} denote the corresponding quantile function. As frequently done, we set $F^{-1}(0) = a$ and $F^{-1}(1) = b$. Let $U \subseteq \mathbb{R}^{d_1}$ be an open set. We denote by $C^k(U, \mathbb{R}^{d_2})$ the set of all k -times continuously differentiable functions from U to \mathbb{R}^{d_2} . Further, we denote by $C^{k,1}(U, \mathbb{R}^{d_2})$ the set of all k -times continuously differentiable functions from U to \mathbb{R}^{d_2} , whose k th derivative is Lipschitz continuous. For $d_2 = 1$, we abbreviate this to $C^k(U)$ and $C^{k,1}(U)$. If the domain and range of a function g are clear from the context, we will usually write $g \in C^k$ or $g \in C^{k,1}$.

2. Distributional Limits

In this section, we recall the setting, establish the conditions required for our asymptotic theory and ensure that these are met in some simple examples before we state our main theoretical results, upon which our statistical methodology is based. To this end, we show that $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ is a reasonable estimator for the density of the DTM-signature by proving the distributional limit (7).

2.1 Setting and Assumptions

First of all, we summarize the setting introduced in Section 1.1.

Setting 1 *Let $(\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$ denote a regular Euclidean metric measure space. For $x \in \mathcal{X}$ let $d_{\mathcal{X},m}^2(x)$ denote the corresponding Distance-to-Measure function with mass parameter $m \in (0, 1]$. Let $X \sim \mu_{\mathcal{X}}$ and assume that the Distance-to-Measure signature $d_{\mathcal{X},m}^2(X)$ admits a density $f_{d_{\mathcal{X},m}^2}$. Let $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and denote by $\widehat{f}_{d_{\mathcal{X},m}^2}$ and $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ the kernel density estimators defined in (5) and (6), respectively.*

It is noteworthy that the assumption that $d_{\mathcal{X},m}^2(X)$ admits a Lebesgue density is slightly restrictive. The probability measure $\mu_{d_{\mathcal{X},m}^2}$ of the DTM-signature can have a pure point

component $\mu_{d_{\mathcal{X},m}^2, \text{pp}}$ in addition to the continuous component $\mu_{d_{\mathcal{X},m}^2, \text{cont}}$, if the spaces considered have very little local structure (for examples, see Section 2.2). That is,

$$\mu_{d_{\mathcal{X},m}^2} = \mu_{d_{\mathcal{X},m}^2, \text{pp}} + \mu_{d_{\mathcal{X},m}^2, \text{cont}}.$$

If we define $f_{d_{\mathcal{X},m}^2}$ to be the Radon-Nikodym derivative of the absolutely continuous component $\mu_{d_{\mathcal{X},m}^2, \text{cont}}$, i.e., $f_{d_{\mathcal{X},m}^2} d\lambda = d\mu_{d_{\mathcal{X},m}^2, \text{cont}}$, the pointwise asymptotic analysis of $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ performed in Section 2.3 (see Theorem 7) remains valid for all y with $\mu_{d_{\mathcal{X},m}^2}(\{y\}) = 0$ that meet the corresponding assumptions. This guarantees that our analysis remains meaningful even if parts of our space do not provide local structure that is discriminative.

In order to derive statement (7), we require certain regularity assumptions on the density $f_{d_{\mathcal{X},m}^2}$, the DTM-function $d_{\mathcal{X},m}^2$ and the kernel K . For the sake of completeness, we first recall some facts about the relation of the level sets of a given function. Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ and let $y \in \mathbb{R}$ be such that $g^{-1}(\{y\}) \neq \emptyset$. Suppose that the function g is continuously differentiable in an open neighborhood of $g^{-1}(\{y\})$. Assume further that $\nabla g \neq 0$ on the level set $g^{-1}(\{y\})$. Then, it follows by Cauchy-Lipschitz's theory (see e.g., Hirsch and Smale (1974); Amann (2011)) that there exists a constant $h_0 > 0$, an open set $W \supset g^{-1}([y - h_0, y + h_0])$ and a canonical one parameter family of C^1 -diffeomorphisms $\Phi : [-h_0, h_0] \times W \rightarrow \mathbb{R}^d$ with the following property:

$$\Phi(v, g^{-1}(\{y\})) = g^{-1}(\{y + v\})$$

for all $v \in [-h_0, h_0]$ (for the precise construction of Φ see the proof of Lemma D.3 in Section D.1). Throughout the following, the family $\{\Phi(v, \cdot)\}_{v \in [-h_0, h_0]}$ (also abbreviated to Φ) is referred to as *canonical level set flow of $g^{-1}(\{y\})$* .

Condition 2 Let $f_{d_{\mathcal{X},m}^2}$ be supported on $[D_1, D_2]$ and let $y \in [D_1, D_2]$. Assume that there exists $\epsilon > 0$ such that $f_{d_{\mathcal{X},m}^2}$ is twice continuously differentiable on $(y - \epsilon, y + \epsilon)$. Further, suppose that the function $d_{\mathcal{X},m}^2 : \mathbb{R}^d \rightarrow \mathbb{R}$ is $C^{2,1}$ on an open neighborhood of the level set

$$\Gamma_y := d_{\mathcal{X},m}^2{}^{-1}(\{y\}) = \{x \in \mathbb{R}^d : d_{\mathcal{X},m}^2(x) = y\},$$

that $\nabla d_{\mathcal{X},m}^2 \neq 0$ on Γ_y and that there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$

$$\mathcal{I}_{\mathcal{X}}(y; v) := \int_{\Gamma_y} |\mathbf{1}_{\{x \in \mathcal{X}\}} - \mathbf{1}_{\{\Phi(v, x) \in \mathcal{X}\}}| d\sigma^{d-1}(x) \leq C_y |v|, \quad (8)$$

where $\{\Phi(v, \cdot)\}_{v \in [-h_0, h_0]}$ denotes the canonical level set flow of Γ_y and C_y denotes a finite constant that depends on y and $d_{\mathcal{X},m}^2$. Suppose that the kernel $K : \mathbb{R} \rightarrow \mathbb{R}_+$, is an even, twice continuously differentiable function with $\text{supp}(K) = [-1, 1]$. If $m < 1$, we assume additionally that there are constants $\kappa > 0$ and $1 \leq b < 5$ such that for $u \in (0, 1)$ it holds

$$\omega_{\mathcal{X}}(u) := \sup_{x \in \mathcal{X}} \sup_{t, t' \in (0, 1)^2, |t - t'| < u} |F_x^{-1}(t) - F_x^{-1}(t')| \leq \kappa u^{1/b}. \quad (9)$$

The satisfiability of Condition 2 is an important issue that is difficult to address in general. Hence, in Section 2.2 we will verify that the requirements of Condition 2 are met in several

simple examples. Nevertheless, in order to put Condition 2 into a broader perspective, we first gather some known regularity properties of $d_{\mathcal{X},m}^2$ as well as $\{F_x^{-1}\}_{x \in \mathcal{X}}$ and discuss the technical requirement (8) afterwards.

2.1.1 REGULARITY OF $d_{\mathcal{X},m}^2$ AND $\{F_x^{-1}\}_{x \in \mathcal{X}}$

We distinguish between the cases $m < 1$ and $m = 1$ for the presentation of known regularity results. For $m < 1$, the smoothness of $d_{\mathcal{X},m}^2$ has been investigated in Chazal et al. (2011), where the authors derived the following results.

Lemma 3

1. Let $(\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$ denote an Euclidean metric measure space. Then, the function $d_{\mathcal{X},m}^2 : \mathbb{R}^d \rightarrow \mathbb{R}$ is almost everywhere twice differentiable.
2. If $\mathcal{X} = (\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$ denotes a regular Euclidean metric measure space, then the function $d_{\mathcal{X},m}^2 : \mathbb{R}^d \rightarrow \mathbb{R}$ is differentiable with derivative

$$\nabla d_{\mathcal{X},m}^2(x) = \frac{2}{m} \int [x - y] d\bar{\mu}_x(y),$$

where $\bar{\mu}_x = \mu_{\mathcal{X}}|_{B(x, F_x^{-1}(m))}$.

Another important point for the case $m < 1$ is the verification of inequality (9). This corresponds to bounding a uniform modulus of continuity for the family $\{F_x^{-1}\}_{x \in \mathcal{X}}$. An application of Lemma 3 in Chazal et al. (2016) immediately yields the subsequent result.

Lemma 4 Let $(\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$ be a regular Euclidean metric measure space. Suppose that there are constants $a, b > 0$ such that for all $r > 0$ and all $x \in \mathcal{X}$

$$\mu_{\mathcal{X}}(B(x, r)) \geq 1 \wedge ar^b. \tag{10}$$

Then, it holds that

$$\omega_{\mathcal{X}}(u) \leq 2 \left(\frac{h}{a}\right)^{1/b} \text{diam}(\mathcal{X}).$$

Remark 5 Condition (10) is frequently assumed in the context of shape analysis. Measures that fulfill (10) are often called (a,b) -standard (see Cuevas (2009); Fasy et al. (2014); Chazal et al. (2016) for a detailed discussion of (a,b) -standard measures). In particular, we observe that our assumption (9) is met, whenever $b < 5$.

In the case $m = 1$, it is important to observe that the DTM-function admits the following specific form:

$$d_{\mathcal{X},1}^2(x) = \int_0^1 F_x^{-1}(u) du = \mathbb{E} [\|X - x\|^2], \tag{11}$$

where $X \sim \mu_{\mathcal{X}}$. This identity gives rise to the following lemma.

Lemma 6 Let $\mathcal{X} = (\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$ denote a regular Euclidean metric measure space and let $X \sim \mu_{\mathcal{X}}$. Then, it holds that:

1. The function $d_{\mathcal{X},1}^2 : \mathbb{R}^d \rightarrow \mathbb{R}$ is given as

$$x = (x_1, \dots, x_d) \mapsto (x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_d - c_d)^2 + \zeta, \quad (12)$$

where $c = (c_1, \dots, c_d)^T = \mathbb{E}[X]$ and ζ denotes a finite constant that can be made explicit.

2. The function $d_{\mathcal{X},1}^2 : \mathbb{R}^d \rightarrow \mathbb{R}$ is three times continuously differentiable.

3. We have $\nabla d_{\mathcal{X},1}^2(x) = 0$ if and only if $x = \mathbb{E}[X]$.

4. Consider the representation of $d_{\mathcal{X},1}^2$ in (12). Set $\Gamma_y = d_{\mathcal{X},1}^2{}^{-1}(\{y\})$ and suppose that $(y - \mathbb{E}[\|X - \mathbb{E}[X]\|^2]) > 2h_0 > 0$. In this case, the canonical level set flow $\{\Phi(v, \cdot)\}_{v \in [-h_0, h_0]}$ of Γ_y considered as function from $[-h_0, h_0] \times d_{\mathcal{X},1}^2{}^{-1}([y - h_0, y + h_0])$ to \mathbb{R}^d is for $x = (x_1, \dots, x_d)$ given as

$$(v, x) \mapsto \left((x_1 - c_1) \sqrt{1 + \frac{v}{\|x - c\|^2}} + c_1, \dots, (x_d - c_d) \sqrt{1 + \frac{v}{\|x - c\|^2}} + c_d \right). \quad (13)$$

In order to increase the readability of this section, the proof of Lemma 6 is postponed to Section A in the Appendix.

2.1.2 DISCUSSION OF ASSUMPTION (8) IN CONDITION 2

To conclude this section, we consider the technical assumption (8). First of all, it is obvious (if $d_{\mathcal{X},m}^2$ is nowhere constant) that the assumption only comes into play for $d \geq 2$. Furthermore, we observe that it is trivially fulfilled if there exists some $\epsilon > 0$ such that $\Gamma_{y-\epsilon} \subset \mathcal{X}$, $\Gamma_y \subset \mathcal{X}$ and $\Gamma_{y+\epsilon} \subset \mathcal{X}$. Only if this is not the case, there might be points $y \in \mathcal{X}$ for which (8) is not satisfied. However, the assumption will typically be satisfied for all points of regularity of the density $f_{d_{\mathcal{X},m}^2}$. To provide some intuition on this matter, we will consider the following example.

Example 1 (For $([0, 1]^2, \mathcal{U}([0, 1]^2), \|\cdot\|)$ with $m = 1$ and $y = \frac{5}{12}$, (8) does not hold.)
 Let $\mathcal{X} = [0, 1]^2$ and let $\mu_{\mathcal{X}}$ stand for the uniform distribution on \mathcal{X} . In this case, using relation (11), we obtain for $x = (x_1, x_2) \in \mathcal{X}$

$$d_{\mathcal{X},1}^2(x) = \mathbb{E}[\|X - x\|^2] = \left(x_1 - \frac{1}{2}\right)^2 + \left(x_2 - \frac{1}{2}\right)^2 + \frac{1}{6}.$$

The corresponding DTM-density is supported on $[1/6, 2/3]$ and it is smooth everywhere except for $y = 5/12$, where $f_{d_{\mathcal{X},m}^2}$ has a kink (more details are provided in Section B.3 in the Appendix). The level sets Γ_y ($y \geq 1/6$), are concentric circles centered at $(1/2, 1/2)$ with radii $\sqrt{y - 1/6}$. For all $y < 5/12$ the level sets are fully contained in the open cube $(0, 1)^2$. For all $y > 5/12$, we have $\mathbb{R}^2 \setminus [0, 1]^2 \cap \Gamma_y \neq \emptyset$, i.e., the level sets are at least partly outside of the cube $[0, 1]^2$. This means that $y = 5/12$ is, in a sense, a transition point. In order to check (8) for $y = 5/12$, we observe that $x \in \Gamma_{\frac{5}{12}}$ implies $\mathbf{1}_{\{x \in \mathcal{X}\}} = 1$, while for

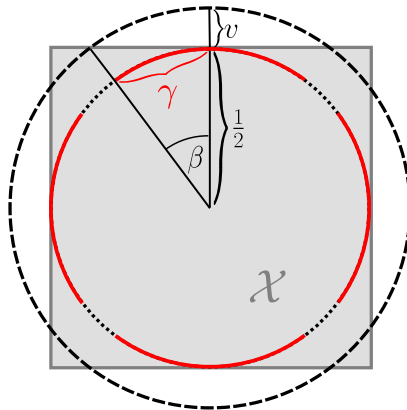


Figure 3: **Tangential Level Set:** Illustration of the behavior of the level sets in a neighborhood of tangential intersection point with the boundary of \mathcal{X} in the setting of Example 1.

$v > 0$ we find $\mathbb{1}_{\{\Phi(v,x) \in \mathcal{X}\}} = 0$ if and only if x is in one of the neighborhoods around the points $(1/2, 1)$, $(1, 1/2)$, $(1/2, 0)$ and $(0, 1/2)$ that correspond to the red solid lines in Figure 3. Therefore, in order to evaluate integral (8) we need to compute the arc lengths of the red curves in Figure 3 (corresponding to eight times the length of the segment γ). For this, we note that we have

$$\gamma = \beta \cdot \pi = \arccos\left(\frac{1/2}{1/2+v}\right) \cdot \pi.$$

By a Taylor expansion, we obtain

$$0 = \arccos(1) \leq \arccos\left(\frac{1/2}{1/2+v}\right) + \left(1 - \frac{1/2}{1/2+v}\right) \cdot \left(-\frac{1}{\sqrt{1 - \frac{1/4}{(1/2+v)^2}}}\right),$$

where we used that the remainder term in the expansion is negative. Therefore, for $v \leq 1$,

$$\arccos\left(\frac{1/2}{1/2+v}\right) \geq \frac{v}{\sqrt{v+v^2}} \geq \frac{v}{\sqrt{2v}}.$$

This yields

$$\mathcal{I}_{\mathcal{X}}(5/12; v) \geq 8 \cdot \pi \cdot \frac{\sqrt{v}}{\sqrt{2}} > \sqrt{v},$$

which proves that for $y = 5/12$ the requirement (8) is not fulfilled.

We conclude this subsection by noting that the dimension of \mathcal{X} heavily influences the regularity of (8). While it seems to be problematic, if Γ_y intersects tangentially with the boundary $\partial\mathcal{X}$ of \mathcal{X} for $d = 2$, this is not necessarily the case for $d \geq 3$. In particular, if we consider $\mathcal{X} = [0, 1]^3$ equipped with the uniform distribution, we find that for $y = 3/4$ the level set Γ_y tangentially touches $\partial\mathcal{X}$ at 6 points. However, here, it does not cause any problems. Following our considerations from Example 1, one can show that condition (8) holds for all points y in the support of $f_{d^2_{\mathcal{X},m}}$ for $d \geq 3$.

2.2 Examples of DTM-Densities

In the following, we will derive $d_{\mathcal{X},m}^2$ as well as $f_{d_{\mathcal{X},m}^2}$ in several simple examples explicitly and verify that in these settings Condition 2 is met almost everywhere. Since calculating $d_{\mathcal{X},m}^2$ and $f_{d_{\mathcal{X},m}^2}$ explicitly is quite cumbersome (especially for $m < 1$), we concentrate on one- or two-dimensional examples. In order to increase the readability of this section, we postpone the explicit, but lengthy representations of the derived DTM-functions and densities (as well as their derivation) to Section B.

We begin our considerations with the simplest case possible, the interval $[0, 1]$ equipped with the uniform distribution.

Example 2 *Let $\mathcal{X} = [0, 1]$ and let $\mu_{\mathcal{X}}$ denote the uniform distribution on \mathcal{X} . In Section B.1, we derive $d_{\mathcal{X},m}^2$ for general m . There we also compute $f_{d_{\mathcal{X},1}^2}$ (see Figure 4 for an illustration). For $m = 1$, the requirement (8) does not come into play as \mathcal{X} is one-dimensional and $d_{\mathcal{X},1}^2$ is nowhere constant. However, we point out that the density $f_{d_{\mathcal{X},1}^2}$ is unbounded (but twice continuously differentiable in the interior of its support). In the case $m < 1$ things are quite different. The function $d_{\mathcal{X},m}^2$ is constant on $[m/2, 1 - m/2]$ as all balls of radius $m/2$ centered at $x \in [m/2, 1 - m/2]$ have the same structure and hence the distribution of the random variable $d_{\mathcal{X},m}^2(X)$ has a pure point component showing that it does not admit a Lebesgue density.*

It is immediately clear that the DTM-signature can only admit a density with respect to the Lebesgue measure, if the DTM-function defined in (1) is almost nowhere constant. In the next example, we equip $\mathcal{X} = [0, 1]$ with another probability distribution, whose density is not constant on \mathcal{X} , thereby adding structure to the metric measure space. In this case, we will find that also for $m < 1$ the corresponding DTM-signature admits a Lebesgue density.

Example 3 *Let $\mathcal{X} = [0, 1]$ and let $\mu_{\mathcal{X}}$ denote the probability distribution on $[0, 1]$ with density $f(x) = 2x$. Let $m = 0.1$. In Section B.2, we derive $d_{\mathcal{X},0.1}^2$ explicitly and demonstrate that the random variable $d_{\mathcal{X},0.1}^2(X)$, $X \sim \mu_{\mathcal{X}}$, admits a Lebesgue density in this setting (see Figure 4 for an illustration). We observe that $d_{\mathcal{X},0.1}^2$ is continuously differentiable everywhere and three times continuously differentiable almost everywhere. Further, the density $f_{d_{\mathcal{X},0.1}^2}$ admits one discontinuity for $y = \frac{-683}{60} + 18\sqrt{\frac{2}{5}}$ and is C^2 almost everywhere.*

We observe that the DTM-densities derived in Example 2 and Example 3 are both unbounded. This has a simple explanation. Let $(\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$ be a regular Euclidean metric measure space and denote the d -dimensional Lebesgue density of $\mu_{\mathcal{X}}$ by $g_{\mu_{\mathcal{X}}}$. Suppose that $f_{d_{\mathcal{X},m}^2}$ exists. Then, one can show (see e.g. Appendix C of Weitkamp et al. (2024)) that

$$f_{d_{\mathcal{X},m}^2}(y) = \int_{\{x \in \mathcal{X} : d_{\mathcal{X},m}^2(x) = y\}} \frac{g_{\mu_{\mathcal{X}}}(u)}{\|\nabla d_{\mathcal{X},m}^2(u)\|} d\sigma^{d-1}(u). \quad (14)$$

Since $d\sigma^0$ corresponds to integration with respect to the counting measure, the DTM-density of a one-dimensional Euclidean metric measure space is unbounded if there are $u \in \mathcal{X}$ with $\|\nabla d_{\mathcal{X},m}^2(u)\| = 0$ (this is the case in Example 2 and Example 3). However,

it is important to note that this behavior mainly occurs for one-dimensional Euclidean metric measure spaces. For higher dimensional spaces, the area (w.r.t. $d\sigma^{d-1}$) of the set $A = \{x \in \mathcal{X} : \|\nabla d_{\mathcal{X},m}^2(u)\| = 0\}$, is usually a null set. Hence, it is possible that the density $f_{d_{\mathcal{X},m}^2}$ defined in (14) remains bounded even if A is non-empty (see Example 4 and Example 5).

To conclude this section and in order to illustrate that the showcased regularity of the DTM-function $d_{\mathcal{X},m}^2$ and the DTM-density $f_{d_{\mathcal{X},m}^2}$ does not only hold for one-dimensional settings, we consider two simple examples in \mathbb{R}^2 next. As the derivation of the family $(F_x^{-1})_{x \in \mathcal{X}}$ is in general very cumbersome, we restrict ourselves to the case $m = 1$.

Example 4 (Continued) *Recall that $\mathcal{X} = [0, 1]^2$, $\mu_{\mathcal{X}}$ stands for the uniform distribution on \mathcal{X} and that $m = 1$. Based on our previous considerations it is possible to derive $f_{d_{\mathcal{X},1}^2}$ explicitly (see Section B.3 for the derivation). As illustrated in Figure 4, the density $f_{d_{\mathcal{X},1}^2}$ is continuous. Moreover, it is twice continuously differentiable inside its support for $y \neq \frac{5}{12}$, which is also the only point where the requirements of (8) are not met, as discussed previously.*

We note that the density $f_{d_{\mathcal{X},1}^2}$ derived in Example 4 is constant on $[1/6, 5/12]$. This kind of behavior is also expressed when considering a disc in \mathbb{R}^2 equipped with the uniform distribution (it is easy to verify that $f_{d_{\mathcal{X},1}^2}$ is a constant function in this case). It is well known that it is difficult for kernel density estimators to approximate constant pieces or a constant function. However, it is not reasonable to assume that the data stems from a uniform distribution over a compact set in many applications (such as chromatin loop analysis). More often, it is possible to assume that the data generating distribution is more concentrated in the center of the considered set. The final example of this section showcases that in such a case the corresponding DTM-signature admits a density without any constant parts even on the disk.

Example 5 *Let \mathcal{X} denote a disk in \mathbb{R}^2 centered at $(0, 0)$ with radius 1 and let $\mu_{\mathcal{X}}$ denote probability measure with density*

$$f(x_1, x_2) = \begin{cases} -\frac{2}{\pi} (x_1^2 + x_2^2 - 1) & x_1^2 + x_2^2 \leq 1, \\ 0 & \text{else.} \end{cases} \quad (15)$$

In this framework, we derive $d_{\mathcal{X},1}^2$ and $f_{d_{\mathcal{X},1}^2}$ in Section B.4. We observe that the level sets Γ_y (of $d_{\mathcal{X},1}^2$) are contained in \mathcal{X} for any $y \in [1/3, 4/3]$, i.e., condition (8) is met for all $y \in (1/3, 4/3)$ in this setting. Further, we realize that $f_{d_{\mathcal{X},1}^2}$ (see Figure 4 for an illustration) is smooth and nowhere constant on the interior of its support.

2.3 Theoretical Results

In this section we study the asymptotic behavior of the kernel estimator of the DTM-density (6). Clearly, standard methodology implies the following pointwise central limit theorem for the kernel estimator $\widehat{f}_{d_{\mathcal{X},m}^2}$ defined in (5).

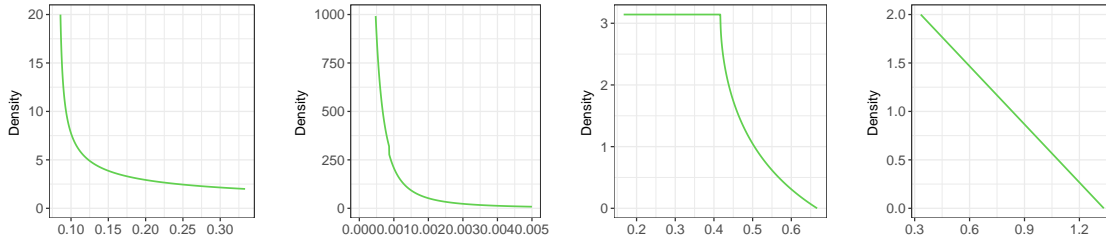


Figure 4: **Distance-to-Measure signature:** Illustration of the densities calculated in Example 2-3, Example 4 and Example 5 (from left to right).

Theorem 6 Assume Setting 1 and suppose that $d_{\mathcal{X},m}^2(X_1)$ admits a density that is twice continuously differentiable in a neighborhood of y . Suppose further that the kernel $K : \mathbb{R} \rightarrow \mathbb{R}_+$, is an even, twice continuously differentiable function with $\text{supp}(K) = [-1, 1]$. Then, it holds for $n \rightarrow \infty$, $nh \rightarrow \infty$ and $h = o(n^{-1/5})$ that

$$\sqrt{nh} \left(\widehat{f}_{d_{\mathcal{X},m}^2}(y) - f_{d_{\mathcal{X},m}^2}(y) \right) \Rightarrow N \left(0, f_{d_{\mathcal{X},m}^2}(y) \int K^2(u) du \right).$$

Surprisingly perhaps, despite the complicated stochastic dependence of the random variables $\delta_{\mathcal{X},m}^2(X_i)$, asymptotically, $\widehat{f}_{d_{\mathcal{X},m}^2}$ and $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ behave equivalently in the following sense.

Theorem 7 Assume Setting 1 and let Condition 2 hold. Then, it holds for $n \rightarrow \infty$, $nh \rightarrow \infty$ and $h = o(n^{-1/5})$ that

$$\sqrt{nh} \left(\widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - f_{d_{\mathcal{X},m}^2}(y) \right) \Rightarrow N \left(0, f_{d_{\mathcal{X},m}^2}(y) \int K^2(u) du \right).$$

As the proof of Theorem 7 is lengthy and quite technical, it has been deferred to Appendix C. There, we will write the density estimator $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ as a U -statistic plus remainder terms as previously discussed at the end of Section 1.1. Then, using a Hoeffding decomposition, the dependencies can be handled. However, showing that the remainder terms vanish is not trivial and requires the application of some tools from geometric measure theory.

3. Simulations

In the following, we investigate the finite sample behavior of $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ in Monte Carlo simulations. To this end, we illustrate the pointwise limit derived in Theorem 7 in the setting of Example 5 and exemplarily highlight the potential of $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ to discriminate between different Euclidean metric measure spaces. All simulations were performed in R (R Core Team (2017)). An R package, that implements the calculation of $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ as well as some basic analytical tools, is available at <https://github.com/cweitkamp3/DTMdemo>.

3.1 Pointwise Limit

We start with the illustration of Theorem 6. To this end, we consider the Euclidean metric measure space $(\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$ from Example 5. Recall that in this setting, \mathcal{X} denotes a disk in \mathbb{R}^2 centered at $(0, 0)$ with radius 1 and that $\mu_{\mathcal{X}}$ denotes the probability measure with density

$$f(x_1, x_2) = \begin{cases} -\frac{2}{\pi} (x_1^2 + x_2^2 - 1) & x_1^2 + x_2^2 \leq 1, \\ 0 & \text{else.} \end{cases}$$

Now, we choose $m = 1$ and consider

$$\widehat{f}_{\delta_{\mathcal{X},1}^2}(y) = \frac{1}{nh_1} \sum_{i=1}^n K_{Bi} \left(\frac{\delta_{\mathcal{X},m}^2(X_i) - y}{h_1} \right),$$

where K_{Bi} denotes the Biweight kernel, i.e.,

$$K_{Bi}(u) = \begin{cases} \frac{15}{16} (1 - u^2)^2 & |u| \leq 1, \\ 0 & \text{else.} \end{cases} \quad (16)$$

Since we have calculated $d_{\mathcal{X},1}^2$ explicitly (see Eq. 19), it is of interest to compare the behavior of $\widehat{f}_{\delta_{\mathcal{X},1}^2}(y)$ to the one of

$$\widehat{f}_{d_{\mathcal{X},1}^2}(y) = \frac{1}{nh_2} \sum_{i=1}^n K_{Bi} \left(\frac{d_{\mathcal{X},1}^2(X_i) - y}{h_2} \right).$$

As discussed previously, $\widehat{f}_{d_{\mathcal{X},1}^2}$ a kernel density estimator based on independent data, whose limit behavior is well understood (see Theorem 6). Nevertheless, for $y \in (1/3, 4/3)$, $\widehat{f}_{d_{\mathcal{X},1}^2}(y)$ and $\widehat{f}_{\delta_{\mathcal{X},1}^2}(y)$ admit the same asymptotic behavior according to Theorem 7, whose requirements can be easily checked in this setting (see Example 5). In order to illustrate this, we generate two independent samples $\{X_i\}_{i=1}^n$ and $\{X'_i\}_{i=1}^n$ of $\mu_{\mathcal{X}}$ and calculate $\Delta_n = \{\delta_{\mathcal{X},1}^2(X_i)\}_{i=1}^n$ as well as $D_n = \{d_{\mathcal{X},1}^2(X_i)\}_{i=1}^n$ for $n = 50, 500, 2500, 5000$. We set

$$h_1 = (1.06 \min\{s(\Delta_n), \text{IQR}(\Delta_n)/1.34\})^{5/4} n^{-1/4}$$

and

$$h_2 = (1.06 \min\{s(D_n), \text{IQR}(D_n)/1.34\})^{5/4} n^{-1/4},$$

where s is the usual sample standard deviation and IQR denotes the inter quartile range. Based on Δ_n and D_n , we choose a central value of y and calculate

$$\sqrt{nh_1}(\widehat{f}_{\delta_{\mathcal{X},1}^2}(y) - f_{d_{\mathcal{X},1}^2}(y)) \text{ and } \sqrt{nh_2}(\widehat{f}_{d_{\mathcal{X},1}^2}(y) - f_{d_{\mathcal{X},1}^2}(y)). \quad (17)$$

For each n , we repeat this process 5,000 times. The finite sample distributions of the quantities defined in (17) are compared to their theoretical normal counter part in Figure 5 (exemplarily for the specific choice of $y = 0.7$). The kernel density estimators displayed highlight that the asymptotic behavior of $\widehat{f}_{\delta_{\mathcal{X},1}^2}(y)$ (red) matches that of $\widehat{f}_{d_{\mathcal{X},1}^2}(y)$ (green). Further, we observe that even for small samples sizes both finite sample distributions strongly resemble their theoretical normal limit distribution (blue).

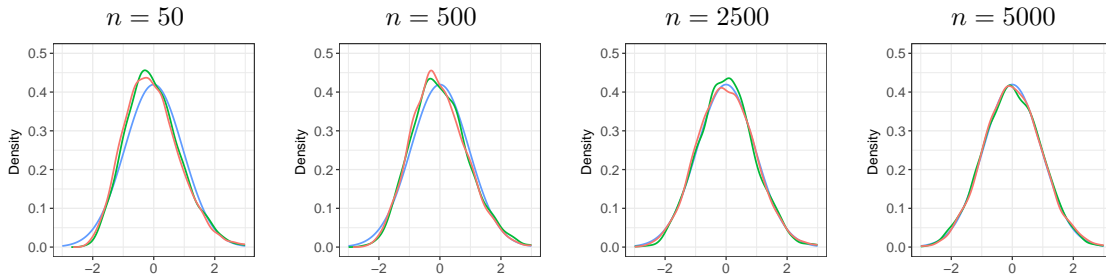


Figure 5: **Pointwise limit distribution:** Kernel density estimators of $\sqrt{nh_1}(\widehat{f}_{d_{\mathcal{X},1}^2}(0.7) - f_{d_{\mathcal{X},1}^2}(0.7))$ (in red) and $\sqrt{nh_2}(\widehat{f}_{d_{\mathcal{X},1}^2}(0.7) - f_{d_{\mathcal{X},1}^2}(0.7))$ (in green) for $n = 50, 500, 2500, 5000$ (from left to right, sample size 5,000) and the normal limiting density (blue).

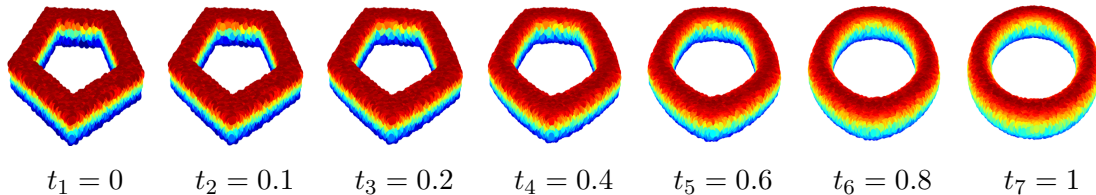


Figure 6: **Metric measure spaces:** Graphical illustration of the metric measure spaces $\{\mathcal{Y}_i\}_{i=1}^7$.

3.2 Discriminating Properties

In the remainder of this section, we will showcase empirically the potential of the DTM-signature for discriminating between different Euclidean metric measure spaces. To this end, let $\mu_{\mathcal{Y}_1}$ stand for the uniform distribution on a 3D-pentagon (inner pentagon side length: 1, Euclidean distance between inner and outer pentagon: 0.4, height: 0.4) and let $\mu_{\mathcal{Y}_7}$ denote the uniform distribution on a torus (center radius: 1.169, tube radius: 0.2) with the same center and orientation (see the plots for $t_1 = 0$ and $t_7 = 1$ in Figure 6). In order to interpolate between these measures, let $\Pi_{\mu_{\mathcal{Y}_1}}^{\mu_{\mathcal{Y}_7}}(t)$, $t \in [0, 1]$, denote the 2-Wasserstein geodesic between $\mu_{\mathcal{Y}_1}$ and $\mu_{\mathcal{Y}_7}$ (see e.g. Santambrogio (2015, Sec. 5.4) for a formal definition). Figure 6 displays the Euclidean metric measure spaces \mathcal{Y}_i , $1 \leq i \leq 7$, corresponding to $\mu_{\mathcal{Y}_i} = \Pi_{\mu_{\mathcal{Y}_1}}^{\mu_{\mathcal{Y}_7}}(t_i)$ for $t_i \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$ (the geodesic has been approximated discretely based on 40,000 points with the WSGeometry-package (Heinemann and Bonneel, 2021)).

In this example, we are not interested in only finding local changes, but we want to distinguish between Euclidean metric measure spaces that differ globally. Hence, $m = 1$ seems to be the most reasonable choice. At the end of this section, we will illustrate the influence of the parameter m in the present setting. We draw independent samples of size n from $\mu_{\mathcal{Y}_i}$, denoted as $\{Y_{j,n,i}\}_{j=1}^n$, and calculate $\Delta_{n,i} = \{\delta_{\mathcal{X},1}^2(Y_{j,n,i})\}_{j=1}^n$ and $\widehat{f}_{\delta_{\mathcal{Y}_i,1}^2}$

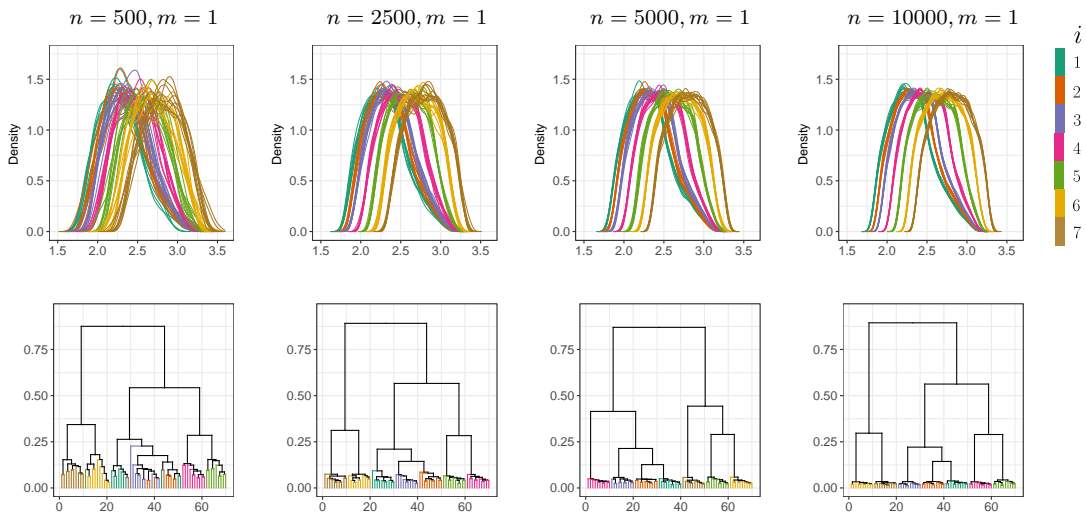


Figure 7: **Discriminating between Euclidean metric measure spaces:** Upper row: Ten realizations of the kernel density estimators $\hat{f}_{\delta_{y_{1,1}}^2}$ (blue-green), $\hat{f}_{\delta_{y_{2,1}}^2}$ (orange), $\hat{f}_{\delta_{y_{3,1}}^2}$ (blue), $\hat{f}_{\delta_{y_{4,1}}^2}$ (pink), $\hat{f}_{\delta_{y_{5,1}}^2}$ (green), $\hat{f}_{\delta_{y_{6,1}}^2}$ (yellow) and $\hat{f}_{\delta_{y_{7,1}}^2}$ (brown) for $n = 500, 2500, 5000, 10000$ (from left to right). Lower row: The results of an average linkage clustering of the considered kernel density estimators based on the L^1 -distance (same coloring).

based on each of these samples for $1 \leq i \leq 7$ and $n = 500, 2500, 5000, 10000$ (Biweight kernel, $h_i = 1.06 \min\{s(\Delta_n), \text{IQR}(\Delta_n)/1.34\}n^{-1/5}$). We repeat this procedure for each i and n 10 times and display the resulting kernel density estimators in the upper row of Figure 7. While it is not possible to reliably distinguish between the realizations of $\hat{f}_{\delta_{y_{1,1}}^2}$ (blue-green), $\hat{f}_{\delta_{y_{2,1}}^2}$ (orange) and $\hat{f}_{\delta_{y_{3,1}}^2}$ (blue) by eye for $n = 500$, this is very simple for $n \geq 2500$. Now that we have estimated the densities, we can choose a suitable notion of distance between densities (e.g. the L^1 -distance) and perform a linkage clustering in order to showcase that the illustrations in the upper row are not deceptive and that it is indeed possible to discriminate between the Euclidean metric measure spaces considered based on the kernel density estimators of the respective DTM-densities. To this end, we calculate the L_1 -distance between the kernel density estimators considered and perform an average linkage clustering on the resulting distance matrix for each n . The results are showcased in the lower row of Figure 7. The average linkage clustering confirms our previous observations.

To conclude this section, we illustrate the influence of the choice of m . For this purpose, we repeat the above procedure with $n = 10,000$ and $m = 0.2, 0.4, 0.6, 0.8$ (this means that we can use the alternative representation of $d_{\mathcal{X},m}^2$ in (4) with $k = 2000, 4000, 6000, 8000$). The resulting kernel density estimators are displayed in the upper row and the corresponding average clustering in the bottom row of Figure 8 (same coloring as previously). As we consider the transformation of μ_{y_1} into μ_{y_7} along a 2-Wasserstein geodesic, it is intuitive that choosing m too small is not informative in this setting (the goal is to distinguish

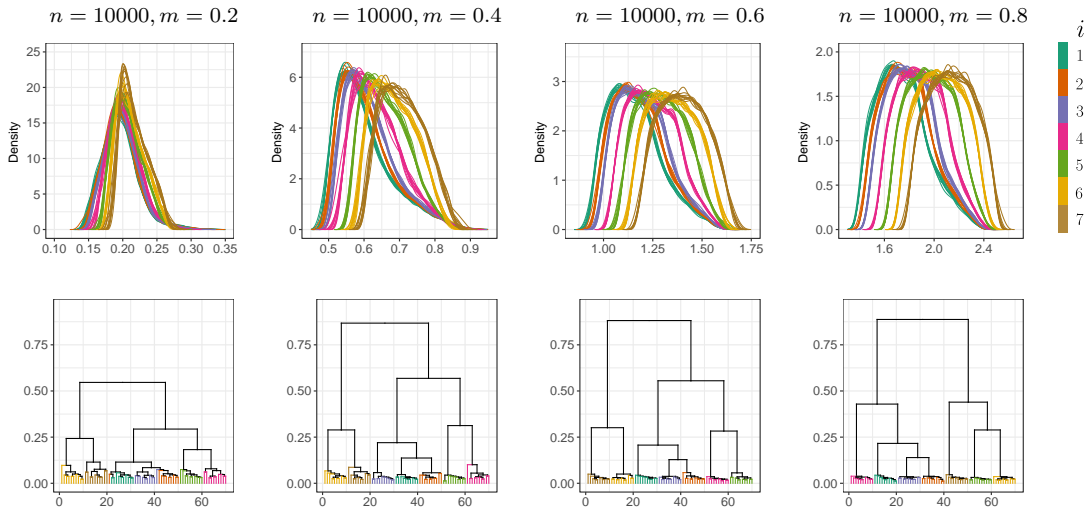


Figure 8: **The influence of m :** Upper row: Ten realizations of the kernel density estimators $\hat{f}_{\delta_{\mathcal{Y}_1,m}^2}$ (blue-green), $\hat{f}_{\delta_{\mathcal{Y}_2,m}^2}$ (orange), $\hat{f}_{\delta_{\mathcal{Y}_3,m}^2}$ (blue), $\hat{f}_{\delta_{\mathcal{Y}_4,m}^2}$ (pink), $\hat{f}_{\delta_{\mathcal{Y}_5,m}^2}$ (green), $\hat{f}_{\delta_{\mathcal{Y}_6,m}^2}$ (yellow) and $\hat{f}_{\delta_{\mathcal{Y}_7,m}^2}$ (brown) for $n = 10000$ and $m = 0.2, 0.4, 0.6, 0.8$ (from left to right). Lower row: The results of an average linkage clustering of the considered kernel density estimators based on the L^1 -distance (same coloring).

between the whole spaces). Indeed, this is exactly, what we observe. For $m = 0.2$ the kernel density estimators strongly resemble each other and in particular the Euclidean metric measure spaces \mathcal{Y}_1 and \mathcal{Y}_2 are hardly distinguishable (see the corresponding dendrogram in the lower row of Figure 8). For $m \geq 0.4$ the kernel density estimators are better separated and the corresponding dendrograms highlight that it is possible to discriminate between the spaces \mathcal{Y}_i based on the kernel density estimators $\hat{f}_{\delta_{\mathcal{Y}_i,m}^2}$, $i = 1, \dots, 7$ and $m = 0.4, 0.6, 0.8$. It is noteworthy that although the form of the kernel density estimators drastically changes between $m = 0.4$ and $m = 1$, the quality of the corresponding clustering only increases slightly with increasing m .

4. Chromatin Loop Analysis

In this section, we will highlight how to use the DTM-density-transformation for chromatin loop analysis. First, we briefly recall some important facts about chromatin fibers, state the goal of this analysis and precisely describe the data used here.

For human beings, chromosomes are essential parts of cell nuclei. They carry the genetic information important for heredity transmission and consist of chromatin fibers. Learning the topological 3D structure of the chromatin fiber in cell nuclei is important for a better understanding of the human genome. As discussed in Section 1.3, TADs are self-interacting genomic regions, which are often associated with loops in the chromatin fibers. These

domains have been estimated to the range of 100-300 nm (Nuebler et al., 2018). Hi-C data (Lieberman-Aiden et al., 2009) allow to construct spatial proximity maps of the human genome and are often used to analyze genome-wide chromatin organization and to identify TADs. However, spatial size and form, and how frequently chromatin loops and domains exist in single cells, cannot directly be answered based on Hi-C data, whereas in 3D visualization of chromosomal regions via SMLM with a sufficiently high resolution, this information might be more easily accessible (Hao et al., 2021). Therefore, in the above reference, such an approach is considered, in which two groups of images of chromatin fibers were produced: Chromatin with supposedly fully intact loop structures and chromatin, which had been treated with auxin prior to imaging. Auxin is known to cause a degrading of the loops. Therefore, in the second set of images, the loops are expected to be mostly dissolved. The obtained resolution in these images was of the order of 150 nm, i.e., below the diffraction limit and comparable to the typical sizes of TADs. This means that the analysis of chromatin loops based on these images is tractable but difficult as we will not see detailed loops when zooming in.

In this paper we analyse simulated SMLM data of chromatin fibers that mimic the chromatin structure with loops as local features and compare them to simulated data that mimic the progressive degradation of loop structures in five steps. The simulated structures mimic the first chromosome (of 23 in total) of the human genome, which is the longest with approximately 249 megabases (Mb, corresponding to 249,000,000 nucleotides). Each step corresponds to a loop density with a different parameter, which we denote by c . The value of c is the number of loops per megabase. A value of $c = 25$ corresponds to a high loop density with 2490 loops in total and corresponds to the setting without the application of auxin. Values of $c = 10, 6, 4, 2$ correspond to decreasing states of resolved loops (1494, 996 and 498 loops) and $c = 0$ encodes the fully resolved state. These simulated images provide a controlled setting in which we can investigate the applicability of our methods and in which we can explore how small a difference in loop density our method can still pick up and when it starts to break down. Here, we only consider classification into the different conditions based on the estimated DTM density. While it is clear from the results described below that information on loop size and frequency is encoded in these densities, a quantification of these parameters requires a deeper study of the proposed methods and is beyond the scope of this paper.

In our study, we consider 102 synthetic, noisy samples of size 49800 of 6 different loop densities each and denote the corresponding samples as $\mathcal{X}_{i,c}$, $c = 25, 10, 6, 4, 2, 0$, $1 \leq i \leq 102$. These samples are created by first discretizing the chromatin structure such that the distance between two points along the chromatin structure corresponds to 45 nm. Then, we add independent, centered Gaussian errors with covariance matrix

$$\Sigma = \begin{pmatrix} 45 & 0 & 0 \\ 0 & 45 & 0 \\ 0 & 0 & 90 \end{pmatrix}$$

to each point (see Figure 9 for an illustration of data obtained in this fashion). This high level of noise is chosen to match the experimental data obtained in Hao et al. (2021). Throughout the following, we consider the data on a scale of 1:45. We stress once again that the goal of our analysis is to distinguish between the respective loop conditions and

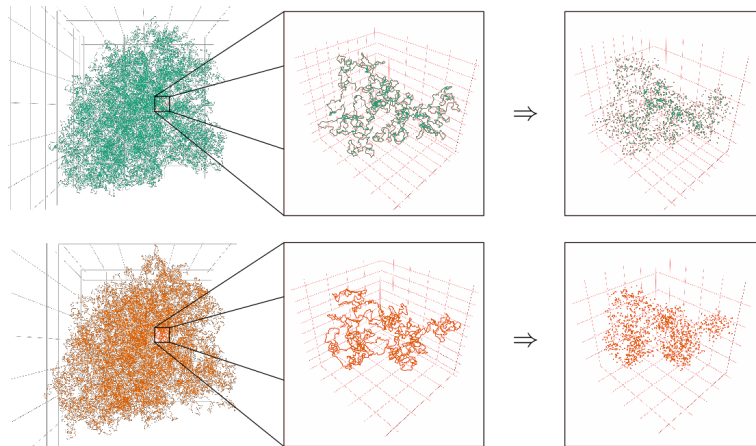


Figure 9: **Chromatin Loops:** Upper row: Illustration of a chromatin structure and the corresponding sample with loop density $c = 25$. Lower row: Illustration of the same chromatin and the corresponding sample with loop density $c = 10$.

not between chromatin fibers from which the points are sampled (the overall form of the chromatin fibers within one condition can be quite different). We demonstrate in the following that the corresponding DTM-signatures, or more precisely the corresponding kernel density estimators $\hat{f}_{\delta_{\mathcal{X}_{i,c},m}^2}$, $1 \leq i \leq 102$, (m chosen suitably small) represent a useful transformation of the data that allows discrimination between the different loop densities, while disregarding the overall shape of the chromatin fiber. To this end, we follow the strategy proposed in Section 1.3 and calculate $\Delta_{i,c} = \{\delta_{\mathcal{X},m}^2(X_{j,i,c}) : X_{j,i,c} \in \mathcal{X}_{i,c}\}$ for $1 \leq i \leq 102$, $c \in \{25, 10, 6, 4, 2, 0\}$ and $m \in \{1/9960, 1/4980, 1/1245\}$. These particular choices of m entail that in order to calculate $\delta_{\mathcal{X},m}^2(X_{j,i,c})$ we need to take the mean over the distances to the $k = 5, 10, 40$ nearest neighbors of $X_{j,i,c} \in \mathcal{X}_{i,c}$ (recall the representation of $\delta_{\mathcal{X},m}^2$ in (4)).

Remark 8 (Choice of the parameters m and k) *It is important to note that the practical choice of the parameter m (resp. the parameter k) strongly depends on the mathematical model and the sampling mechanism used. In many mathematical models, increasing the sample size corresponds to an increase in sampling density, while the object under consideration remains fixed. In this case, DTM-densities with the same mass parameter m , which determines the proportion of points considered, are comparable, although the number of nearest neighbors considered varies with n . However, this is not necessarily the case in many applications. In our biological example, varying n means changing the size of the structure under consideration since the distance between two observable points along the polymer remains fixed (up to random errors). In such a case, the corresponding DTM-densities are only comparable if the parameter k is fixed (resulting in variations of m). Note that our synthetic data sets are all of the same size and therefore both viewpoints are equivalent in this particular case.*

We determine $\hat{f}_{\delta_{\mathcal{X}_{i,c},m}^2}$ (Biweight kernel, $h = 1.06 \min\{s(\Delta_n), \text{IQR}(\Delta_n)/1.34\}n^{-1/5}$) based on each of the samples $\Delta_{i,c}$. The resulting kernel density estimators are displayed in Figure

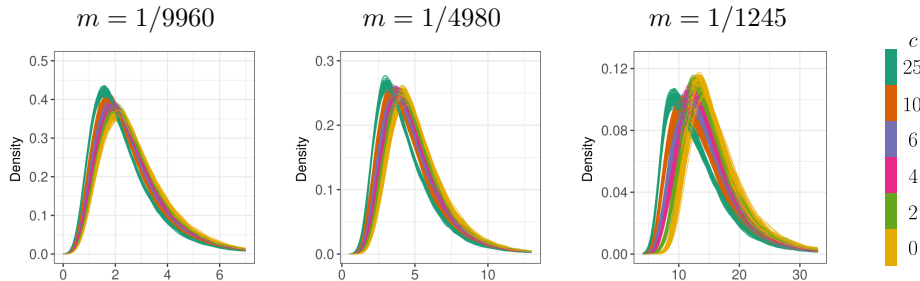


Figure 10: **Chromatin Loop Analysis I:** Illustration of the DTM-density estimators $\hat{f}_{\delta_{\mathcal{X}_{i,c},m}^2}$, $1 \leq i \leq 102$, for $c = 25$ (blue-green), $c = 10$ (orange), $c = 6$ (blue), $c = 4$ (pink), $c = 2$ (green) and $c = 0$ (yellow) and $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right).

10. Generally, the kernel density estimators based on the different samples with the same loop density strongly resemble each other and it is possible to roughly distinguish between the different values of c . For all values of m considered, the DTM-density estimators based on $\mathcal{X}_{i,25}$, $1 \leq i \leq 102$, (here the respective chromatin fibers form many loops) are well separated from the other kernel density estimators and the estimators based on the samples $\Delta_{i,2}$ and $\Delta_{i,0}$ (which correspond to the lowest loop densities considered) are the most similar when comparing the different loop densities. In order to make a more qualitative comparison between the estimators $\hat{f}_{\delta_{\mathcal{X}_{i,c},m}^2}$, we use the strategy developed in Section 3.2 and perform an average linkage clustering based on the L_1 -distance between the estimated densities. For clarity, we restrict ourselves to the comparison of the loop density $c = 25$ against $c = 10$ as well as $c = 2$ against $c = 0$ and point out that the comparison between the setting $c = 2$ against $c = 0$ is very difficult as the loop frequencies are very low. The dendrograms in the upper row of Figure 11 illustrate the comparison of $c = 25$ and $c = 10$. It is remarkable that for each m the correct clusters are obtained. The lower row of Figure 11 showcases the dendrograms for the comparison of the estimators $\hat{f}_{\delta_{\mathcal{X}_{i,2},m}^2}$ and $\hat{f}_{\delta_{\mathcal{X}_{i,0},m}^2}$, $1 \leq i \leq 102$ and $m \in \{1/9960, 1/4980, 1/1245\}$. For $m \in \{1/9960, 1/4980\}$, we obtain (up to one exception) the correct clusters, although they are much closer (w.r.t. the L^1 -distance) than the clusters for the previous comparisons. However, for $m = 1/1245$, it is no longer possible to reliably identify two clusters that correspond to $c = 2$ and $c = 0$. It seems that in this case m is too large to yield a perfect discrimination.

Furthermore, we investigate whether classification based on the DTM-density estimates $\hat{f}_{\delta_{\mathcal{X}_{i,c},m}^2}$ is possible. Here, we restrict ourselves once again to the comparison of $c = 25$ with $c = 10$ as well as of $c = 2$ with $c = 0$. For each comparison, we randomly select 5%/10% (rounded up) of the density estimates for each considered loop density and classify the remaining ones according to the majority of the labels of their $k = 1, 3, 5$ nearest neighbors in the randomly selected sample. We repeat this procedure for both comparisons 10,000 times and report the relative number of misclassifications in Table 1. The upper row of said

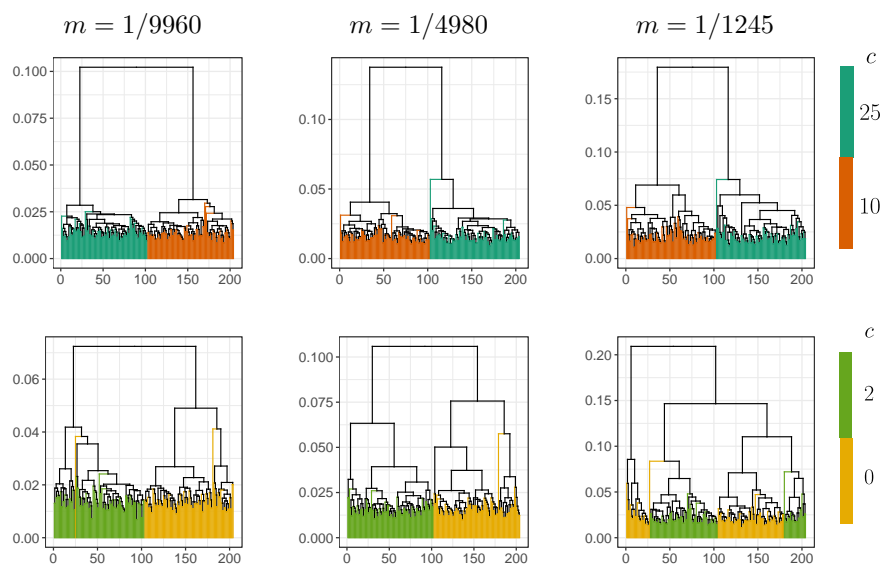


Figure 11: **Chromatin Analysis II:** Upper row: The results of an average linkage clustering of the kernel density estimators $\hat{f}_{\delta_{X_{i,25},m}}^2$ (blue-green) and $\hat{f}_{\delta_{X_{i,10},m}}^2$ (orange), $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right) based on the L^1 -distance. Lower row: The results of an average linkage clustering of the kernel density estimators $\hat{f}_{\delta_{X_{i,2},m}}^2$ (green) and $\hat{f}_{\delta_{X_{i,0},m}}^2$ (yellow), $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right) based on the L^1 -distance.

	$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$
5%	0.000	0.000	0.000	5%	0.000	0.000	0.000	5%	0.000	0.000	0.000
10%	0.000	0.000	0.000	10%	0.000	0.000	0.000	10%	0.000	0.000	0.000
	$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$
5%	0.029	0.049	0.069	5%	0.012	0.019	0.039	5%	0.029	0.064	0.100
10%	0.020	0.033	0.043	10%	0.001	0.007	0.012	10%	0.008	0.021	0.025

Table 1: **Chromatin Analysis III:** Upper row: The relative number of missclassifications of a k -nearest neighbor classification (w.r.t. the L^1 -distance) based on the kernel density estimators $\hat{f}_{\delta_{\mathcal{X}_{i,25},m}^2}$ and $\hat{f}_{\delta_{\mathcal{X}_{i,10},m}^2}$, $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right). Lower row: The relative number of missclassifications of a k -nearest neighbor classification (w.r.t. the L^1 -distance) based on the kernel density estimators $\hat{f}_{\delta_{\mathcal{X}_{i,2},m}^2}$ and $\hat{f}_{\delta_{\mathcal{X}_{i,0},m}^2}$, $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right).

table highlights that in the comparison of $c = 25$ and $c = 10$ the DTM-density estimates are always classified correctly. Things change in the comparison of $c = 2$ with $c = 0$. While for all m at least 90% of the classifications are correct, there is a noticeable difference between the individual values of m . We observe that $m = 1/4980$ yields by far the best performance in this setting. It is clear that the loop distributions of the respective chromatin fibers for these two values of c are extremely similar (the chromatin admits few to no loops). Hence, choosing m too large incorporates too much global (non-loop) structure and makes it difficult to discriminate between these two loop densities. On the other hand, choosing m too small seems to incorporate too little structure.

To conclude this section, we find that it is possible for a suitable choice of m to clearly distinguish between the different loop densities based on the DTM-density estimators $\hat{f}_{\delta_{\mathcal{X}_{i,c},m}^2}$. We have illustrated that these estimators yield a good summary of the data and can be used to approach the (already quite difficult) problem of chromatin loop analysis for noisy synthetic data.

4.1 Sensitivity Regarding the Choice of m

The parameter m determining the percentage of nearest neighbors, which are included in the analysis, is set by the experimenter and can therefore be seen as tuning parameter. Naturally, the question arises in which sense and to which extent the proposed methodology is robust regarding the choice of m . In this paper, we have considered an example in which spaces differ regarding their large scale characteristics (see Section 3.2, spaces $\mathcal{Y}_1 \dots, \mathcal{Y}_7$ in Figure 6) and the example of chromatin loops in Section 4, where structures differ locally and globally, but only local features are of interest. General information of this sort is typically known to the experimenter prior to data analysis.

We have seen in Section 3.2 that large scale analyses require large values of m . There, it is shown that $m = 1$ provides very good classification results already for small sample sizes (Figure 7). However, as sample sizes increase, values of m as small as $m = 0.2$ also permit

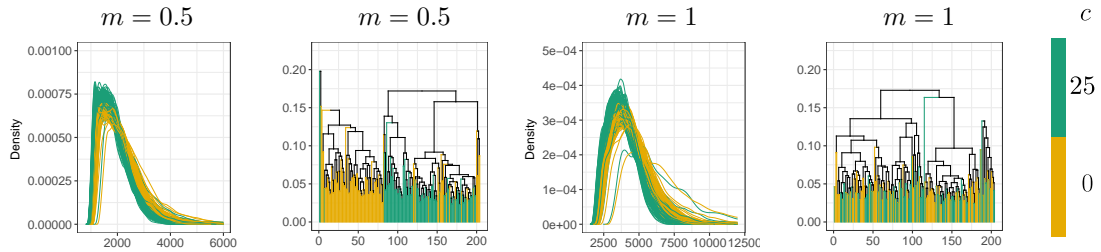


Figure 12: **Chromatin Analysis IV**: The results of an average linkage clustering of the kernel density estimators $\hat{f}_{\delta_{\mathcal{X}_{i,0},m}^2}$ (yellow) and $\hat{f}_{\delta_{\mathcal{X}_{i,25},m}^2}$ (blue green), $1 \leq i \leq 102$, for $m \in \{0.5, 1\}$ (from left to right).

perfect classification between the spaces (Figure 8).

Considering small scale characteristics on the other hand, the parameter m should be chosen small, as indicated by the results presented in Section 4. In the example considered there, robustness with respect to the choice of m is also given, albeit in a smaller range (see Figures 10 and 11). The global behaviour of distances ($m = 1$) is not discriminative in this case at all as can be seen in Figure 12 (right hand side), where the two extreme conditions $c = 0$ and $c = 25$ cannot be separated with $m = 1$. Yet, the two plots on the left of Figure 12 for $m = 0.5$ show that the conditions $c = 0$ and $c = 25$ can still be decently separated, which is quite impressive given the difficulty of the problem.

Overall, we establish that the methodology provided in this paper does depend on the choice of m , but not critically so, as stability over suitable parameter ranges seems to be well maintained in general.

Acknowledgments

We thank the Action Editor and two anonymous referees for their careful reading of our manuscript and their many insightful and constructive comments and suggestions. C. Weitkamp gratefully acknowledges support by the DFG Research Training Group 2088.

Appendix A. Proof of Lemma 6

In this section, we state the full proof of Lemma 6.

Proof [Lemma 6] Let $X = (X_1, \dots, X_d) \sim \mu_{\mathcal{X}}$ and $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

1. We observe that

$$\begin{aligned} d_{\mathcal{X},1}^2(x) &= \mathbb{E} [\|X - x\|^2] = \sum_{i=1}^d (\mathbb{E} [X_i^2] - 2x_i \mathbb{E} [X_i] + x_i^2) \\ &= \sum_{i=1}^d \left((x_i - \mathbb{E} [X_i])^2 + \mathbb{E} [X_i^2] - (\mathbb{E} [X_i])^2 \right). \end{aligned}$$

Setting $c_i = \mathbb{E}[X_i]$ and $\zeta = \sum_{i=1}^d (\mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2)$ yields the claim.

2. This follows directly from the first statement.
3. The first statement implies that $\nabla d_{\mathcal{X},1}^2(x) = 2(x - \mathbb{E}[X])$. Clearly, this is zero if and only if $x = \mathbb{E}[X]$.
4. By the second and third statement $d_{\mathcal{X},1}^2$ is three times continuously differentiable and $\nabla d_{\mathcal{X},1}^2 > 0$ on $d_{\mathcal{X},1}^2{}^{-1}([y - 2h_0, y + 2h_0])$. In consequence, there exists an open set $U \supset d_{\mathcal{X},1}^2{}^{-1}([y - h_0, y + h_0])$ such that the function

$$\varphi : U \subset \mathbb{R}^d \rightarrow \mathbb{R}^d; \quad x \mapsto \frac{\nabla d_{\mathcal{X},1}^2(x)}{\|\nabla d_{\mathcal{X},1}^2(x)\|^2}$$

is $C^2(U, \mathbb{R}^d)$. By Theorem 2 in Chapter 15 of Hirsch and Smale (1974) there is a unique flow $\Phi^* : [-h_0, h_0] \times W \rightarrow \mathbb{R}^d$ with

$$\begin{cases} \frac{\partial}{\partial v} \Phi^*(v, x) = \frac{\nabla d_{\mathcal{X},1}^2(\Phi^*(v, x))}{\|\nabla d_{\mathcal{X},1}^2(\Phi^*(v, x))\|^2} \\ \Phi^*(0, x) = x, \end{cases} \quad (18)$$

where $W \subset \mathbb{R}^d$ is an open set that contains $d_{\mathcal{X},1}^2{}^{-1}([y - h_0, y + h_0])$. Differentiating the function $v \mapsto d_{\mathcal{X},1}^2(\Phi^*(v, x))$ immediately shows that $d_{\mathcal{X},1}^2(\Phi^*(v, x)) = d_{\mathcal{X},1}^2(x) + v$. This implies that $\Phi^*(v, d_{\mathcal{X},1}^2{}^{-1}(\{y\})) = d_{\mathcal{X},1}^2{}^{-1}(\{y + v\})$. In consequence, it only remains to prove that Φ defined in the statement is a solution of the ordinary differential equation (18), which is straightforward. ■

Appendix B. Additional Details on Example 1-5

In this section, we will provide additional details on the examples considered in Section 2.2.

B.1 Example 2

Let $\mathcal{X} = [0, 1]$ and let $\mu_{\mathcal{X}}$ denote the uniform distribution on \mathcal{X} . First, we derive $d_{\mathcal{X},1}^2$ and $f_{d_{\mathcal{X},1}^2}$. To this end, we observe that for $x \in \mathcal{X}$ and $X \sim \mu_{\mathcal{X}}$

$$d_{\mathcal{X},1}^2(x) = \int_0^1 F_x^{-1}(t) dt = \mathbb{E}[(X - x)^2] = \frac{1}{3} - x + x^2.$$

Based on this, we derive that

$$f_{d_{\mathcal{X},1}^2}(t) = \begin{cases} \frac{2\sqrt{3}}{\sqrt{12t-1}} & \frac{1}{12} < t \leq \frac{1}{3}, \\ 0 & \text{else.} \end{cases}$$

Next, we come to general m . In order to calculate $d_{\mathcal{X},m}^2$ (and potentially $f_{d_{\mathcal{X},m}^2}$), it is necessary to derive the family $(F_x^{-1})_{x \in \mathcal{X}}$ explicitly. This is straightforward in this setting. We find that

$$d_{\mathcal{X},m}^2(x) = \frac{1}{m} \int_0^m F_x^{-1}(u) du = \begin{cases} x^2 - mx + \frac{m^2}{3} & \text{for } 0 \leq x < \frac{m}{2}, \\ \frac{m^2}{12} & \text{for } \frac{m}{2} \leq x \leq 1 - \frac{m}{2}, \\ (1-x)^2 - m(1-x) + \frac{m^2}{3} & \text{for } 1 - \frac{m}{2} \leq x \leq 1. \end{cases}$$

Since $d_{\mathcal{X},m}^2$ is constant for $x \in [m/2, 1 - m/2]$ it is immediately clear that the corresponding distribution function $F_{d_{\mathcal{X},0.1}^2}$ is not continuous. Indeed, we find that

$$F_{d_{\mathcal{X},m}^2}(y) = \mathbb{P}(d_{\mathcal{X},m}^2(X) \leq y) = \begin{cases} 0 & \text{for } y < \frac{m^2}{12}, \\ 1 - m + 2\sqrt{y - \frac{m^2}{12}} & \text{for } \frac{m^2}{12} \leq y \leq \frac{m^2}{3}, \\ 1 & \text{for } y > \frac{m^2}{3}. \end{cases}$$

B.2 Example 3

Let $\mathcal{X} = [0, 1]$ and let $\mu_{\mathcal{X}}$ denote the probability distribution on $[0, 1]$ with density $f(x) = 2x$. Let $m = 0.1$. As previously, we have to explicitly calculate the family $(F_x^{-1})_{x \in \mathcal{X}}$. This is again fairly simple. We find that

$$d_{\mathcal{X},0.1}^2(x) = \frac{1}{0.1} \int_0^{0.1} F_x^{-1}(u) du = \begin{cases} x^2 - \frac{2}{3}\sqrt{\frac{2}{5}}x + \frac{1}{20} & 0 \leq x \leq \frac{\sqrt{0.1}}{2}, \\ \frac{1}{4800x^2} & \frac{\sqrt{0.1}}{2} \leq x \leq \frac{1}{2} + \frac{3}{2\sqrt{10}}, \\ x^2 + \left(18\sqrt{\frac{2}{5}} - \frac{40}{3}\right)x + \frac{19}{20} & \frac{1}{2} + \frac{3}{2\sqrt{10}} < x \leq 1. \end{cases}$$

It is obvious that in this case $d_{\mathcal{X},0.1}^2$ is almost nowhere constant. Furthermore, this allows us to derive that

$$f_{d_{\mathcal{X},0.1}^2}(y) = \begin{cases} \frac{80\sqrt{5}-108\sqrt{2}}{\sqrt{13661-4320\sqrt{10}+180y}} & -\frac{13661}{180} + 24\sqrt{10} < y \leq \frac{19-6\sqrt{10}}{120}, \\ 1 + \frac{40\sqrt{5}-54\sqrt{2}}{\sqrt{13661-4320\sqrt{10}+180y}} + \frac{1}{4800y^2} & \frac{19-6\sqrt{10}}{120} < y \leq \frac{-683}{60} + 18\sqrt{\frac{2}{5}}, \\ \frac{1}{4800y^2} & \frac{-683}{60} + 18\sqrt{\frac{2}{5}} < y \leq \frac{1}{120} \\ -1 + \frac{2}{\sqrt{-\frac{1}{2}+90y}} & \frac{1}{120} < y \leq \frac{1}{20} \\ 0 & \text{else.} \end{cases}$$

B.3 Example 4

Let $\mathcal{X} = [0, 1]^2$ and let $\mu_{\mathcal{X}}$ stand for the uniform distribution on \mathcal{X} . Choose $m = 1$ and let $\mathcal{X} \sim \mu_{\mathcal{X}}$. Then, it is possible to derive that for $x = (x_1, x_2) \in \mathcal{X}$

$$d_{\mathcal{X},1}^2(x) = \mathbb{E}[||X - x||^2] = x_1^2 + x_2^2 - x_1 - x_2 + \frac{2}{3}.$$

Based on this, we derive that

$$f_{d_{\mathcal{X},1}^2}(y) = \begin{cases} \pi & \frac{1}{6} \leq y \leq \frac{5}{12}, \\ 2\operatorname{arccot}\left(2\sqrt{y - \frac{5}{12}}\right) - 2\operatorname{arctan}\left(\sqrt{4y - \frac{5}{3}}\right) & \frac{5}{12} < y \leq \frac{2}{3}, \\ 0 & \text{else.} \end{cases}$$

B.4 Example 5

Let \mathcal{X} denote a disk in \mathbb{R}^2 centered at $(0,0)$ with radius 1 and let $\mu_{\mathcal{X}}$ denote probability measure with density f defined in (15). In this case, it is for $m = 1$ straightforward to derive that for $x = (x_1, x_2) \in \mathcal{X}$

$$d_{\mathcal{X},1}^2(x) = x_1^2 + x_2^2 + \frac{1}{3}. \tag{19}$$

The corresponding DTM-density is given as

$$f_{d_{\mathcal{X},1}^2}(y) = \begin{cases} -2y + \frac{8}{3} & \frac{1}{3} < y \leq \frac{4}{3}, \\ 0 & \text{else.} \end{cases} \tag{20}$$

Appendix C. Proof of Theorem 7

In this section, we give the full proof of Theorem 7. The proof is composed of four steps, each of which formulated as an independent lemma (see Section C.1).

Step 1: Replacement of $\delta_{\mathcal{X},m}^2(X_i)$ by $d_{\mathcal{X},m}^2(X_i)$ (Lemma C.1).

We provide a decomposition of $\widehat{f}_{\delta_{\mathcal{X},m}^2}$ in a sum of two leading terms in which $\delta_{\mathcal{X},m}^2(X_i)$ is replaced by $d_{\mathcal{X},m}^2(X_i)$ in the argument of the kernel K and we show that the remainder terms are negligible.

Step 2: Introducing U -statistics (Lemma C.2).

It is shown that the leading terms obtained in Step 1 can be written as a (sum of two) U -statistic(s) asymptotically.

Step 3: Hoeffding decomposition (Lemma C.4).

Applying a Hoeffding decomposition allows to derive a representation of the (sum of two) U -statistic(s) of step 2 as a sum of a deterministic term (expectation), a stochastic leading term consisting of a sum of independent random variables and a remainder term.

Step 4: CLT for the leading term of Step 3 (Lemma C.7).

Since the leading term of Step 3 is a sum of centered independent random variables, we can apply a standard CLT to show its asymptotic normality.

C.1 Auxiliary lemmas representing Step 1 - Step 4

Before we come to the proof of Theorem 7, we will establish several auxiliary results. In order to highlight the overall proof strategy, the corresponding proofs are deferred to Section C.3. We begin this section, by addressing Step 1.

Lemma C.1 (Step 1) *Assume that Setting 1 holds and let Condition 2 be met. Then, it follows that*

$$\begin{aligned} \widehat{f}_{\delta_{\mathcal{X},m}^2}(y) &= \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{h} K \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) + \frac{1}{h^2} K' \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) A_n(X_i) \right] \\ &\quad + \mathcal{O}_P \left(\frac{1}{nh^3} \right) + o_P \left(\frac{\log(n)^{1/(2b)}}{n^{1/2+1/(2b)}h} \right), \end{aligned}$$

where

$$A_n(x) := \frac{1}{m} \int_0^{F_x^{-1}(m)} F_x(t) - \widehat{F}_{x,n}(t) dt.$$

As a direct consequence of Lemma C.1, we find that for $h \in o(1/n^{1/5})$ the statistic

$$\begin{aligned} V_n(y) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) + \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} K' \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) A_n(X_i) \\ &=: V_n^{(1)}(y) + V_n^{(2)}(y) \end{aligned} \quad (21)$$

drives the limit behavior of $\sqrt{nh} \left(\widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - f_{\delta_{\mathcal{X},m}^2}(y) \right)$. Next, we will establish that the statistic $V_n(y)$ can, up to asymptotically negligible terms, be written as a U -statistic (see e.g. Van der Vaart (2000) for more information on U -statistics).

Lemma C.2 (Introduction of U -statistics, Step 2) *Assume Setting 1 and let $V_n^{(1)}$ and $V_n^{(2)}$ be as defined in (21). Then, we have*

$$V_n^{(1)}(y) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,h}^{(1)}(X_i, X_j),$$

where

$$g_{y,h}^{(1)}(x_1, x_2) = \frac{1}{2h} \left(K \left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h} \right) + K \left(\frac{d_{\mathcal{X},m}^2(x_2) - y}{h} \right) \right).$$

Furthermore,

$$V_n^{(2)}(y) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,h}^{(2)}(X_i, X_j) + \mathcal{O}_P \left(\frac{1}{nh^2} \right),$$

where

$$\begin{aligned} g_{y,h}^{(2)}(x_1, x_2) &= \frac{1}{2mh^2} \left[K' \left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h} \right) \int_0^{F_{x_1}^{-1}(m)} F_{x_1}(t) - \mathbf{1}_{\{\|x_1 - x_2\|^2 \leq t\}} dt \right. \\ &\quad \left. + K' \left(\frac{d_{\mathcal{X},m}^2(x_2) - y}{h} \right) \int_0^{F_{x_2}^{-1}(m)} F_{x_2}(t) - \mathbf{1}_{\{\|x_1 - x_2\|^2 \leq t\}} dt \right]. \end{aligned}$$

Remark C.3 *It is important to note that $g_{y,h}^{(1)}$ and $g_{y,h}^{(2)}$ are symmetric by definition, i.e., $V_n^{(1)}(x)$ is a U -statistic and $V_n^{(2)}(x)$ can be decomposed into a U -statistic and an asymptotically negligible remainder term.*

Combining Lemma C.1 and Lemma C.2, we see that

$$\widehat{f}_{\delta_{\mathcal{X},m}^2}(y) = U_n + \mathcal{O}_P\left(\frac{1}{nh^3}\right) + o_P\left(\frac{\log(n)^{1/(2b)}}{n^{1/2+1/(2b)}h}\right), \quad (22)$$

where $U_n = U_n(y)$ denotes the U -statistic with kernel function $g_{y,h}(x_1, x_2) := g_{y,h}^{(1)}(x_1, x_2) + g_{y,h}^{(2)}(x_1, x_2)$. Before we use (22) to finalize the proof of Theorem 7, we establish two further auxiliary results. Next, we rewrite U_n using the Hoeffding decomposition (see Van der Vaart (2000, Sec. 11.4)), which is the key ingredient to handling the stochastic dependencies introduced by the terms $A_n(X_i)$.

Lemma C.4 (Hoeffding decomposition, Step 3) *Assume Setting 1. Let U_n be the U -statistic with kernel function $g_{y,h}(x_1, x_2) = g_{y,h}^{(1)}(x_1, x_2) + g_{y,h}^{(2)}(x_1, x_2)$. Then, it follows that*

$$U_n = \Theta_{y,h} + \frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,h,2}(X_i, X_j).$$

Here, we have that

$$\Theta_{y,h} = \int K(v) f_{\mathcal{X},m}^2(x+vh) dv.$$

Furthermore, let $Z_1, Z_2 \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$. Then, it holds that

$$g_{y,h,1}(x_1) = \frac{1}{2h} K\left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h}\right) - \frac{1}{2} \Theta_{y,h} + \mathbb{E}_{Z_1} \left[\frac{1}{2mh^2} K'\left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h}\right) \Psi(x_1, Z_1) \right],$$

where

$$\Psi(x_1, x_2) := \|x_1 - x_2\|^2 \wedge F_{x_2}^{-1}(m) - \mathbb{E}_{Z_2} [\|Z_2 - x_2\|^2 \wedge F_{x_2}^{-1}(m)], \quad (23)$$

and

$$g_{y,h,2}(x_1, x_2) = g_{y,h}(x_1, x_2) - g_{y,h,1}(x_1) - g_{y,h,1}(x_2) - \Theta_{y,h}. \quad (24)$$

Remark C.5 *It is well known that the mean zero random variables $(g_{y,h,2}(X_i, X_j))_{1 \leq i < j \leq n}$ are uncorrelated (see Van der Vaart (2000, Sec. 11.4)).*

For our later considerations it is important to derive a certain regularity for the function Ψ defined in (23).

Lemma C.6 *Assume Setting 1 and let Ψ be the function defined in (23). Then, the function $z \mapsto \Psi(x_1, z)$ is Lipschitz continuous for all $x_1 \in \mathcal{X}$ and the corresponding Lipschitz constant does not depend on the choice of x_1 , i.e., it holds*

$$|\Psi(x_1, z_1) - \Psi(x_1, z_2)| \leq C \|z_1 - z_2\|,$$

for all $z_1, z_2 \in \mathcal{X}$, where the constant $0 < C < \infty$ does not depend on x_1 .

The next step in the proof of Theorem 7 is to derive for $n \rightarrow \infty$ and $h \rightarrow 0$ the limit distribution of $\sqrt{nh} \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) \right)$.

Lemma C.7 *Assume Setting 1 and let Condition 2 be met. Let $n \rightarrow \infty, h \rightarrow 0$ such that $nh \rightarrow \infty$ and recall that*

$$g_{y,h,1}(x_1) = \frac{1}{2h} K \left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h} \right) - \frac{1}{2} \Theta_{y,h} + \mathbb{E}_{Z_1} \left[\frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \Psi(x_1, Z_1) \right],$$

It holds

$$\frac{2\sqrt{h}}{\sqrt{n}} \sum_{i=1}^n g_{y,h,1}(X_i) \Rightarrow N \left(0, f_{d_{\mathcal{X},m}^2}(y) \int K^2(u) du \right). \quad (25)$$

With all auxiliary results required established, we can finally come to the proof of Theorem 7. The proof strategy is to demonstrate that the limit of $\sqrt{nh} \left(\widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - f_{d_{\mathcal{X},m}^2}(y) \right)$ coincides with the limit of $\frac{2\sqrt{h}}{\sqrt{n}} \sum_{i=1}^n g_{y,h,1}(X_i)$.

C.2 Proof of Theorem 7

The proof of Theorem 7 is now a consequence of the lemmas provided in the previous subsection.

Proof [Theorem 7] We find that

$$\begin{aligned} & \left| \sqrt{nh} \left(\widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - f_{d_{\mathcal{X},m}^2}(y) \right) - \sqrt{nh} \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) \right) \right| \\ & \leq \sqrt{nh} \left| \widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) + \Theta_{y,h} \right) \right| + \sqrt{nh} \left| f_{d_{\mathcal{X},m}^2}(y) - \Theta_{y,h} \right|. \end{aligned} \quad (26)$$

In the following, we consider both summands separately.

First summand: For the first summand, we obtain that

$$\begin{aligned} \mathcal{S}_n(y) &= \sqrt{nh} \left| \widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) + \Theta_{y,h} \right) \right| \\ &\leq \sqrt{nh} \left(\left| \widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - V_n(y) \right| + |V_n(y) - U_n(y)| + \left| U_n(y) - \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) + \Theta_{y,h} \right) \right| \right), \end{aligned}$$

where $V_n(y)$ and $U_n(y)$ are defined in(21) and (22) respectively. By Lemma C.1 and $h = o(n^{-1/5})$ we obtain that

$$\sqrt{nh} \left| \widehat{f}_{\delta_{\mathcal{X},m}^2}(y) - V_n(y) \right| = \mathcal{O}_P \left(\frac{\sqrt{nh}}{nh^3} \right) + o_P \left(\frac{\sqrt{nh} \log(n)^{1/(2b)}}{n^{1/2+1/(2b)}h} \right) = o_P(1).$$

Similarly, we get by Lemma C.2 and $h = o(n^{-1/5})$ that

$$\sqrt{nh} |V_n(y) - U_n(y)| = \mathcal{O}_P \left(\frac{\sqrt{nh}}{nh^2} \right) = o_P(1).$$

Hence, it remains to consider

$$\sqrt{nh} \left| U_n(y) - \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) + \Theta_{y,h} \right) \right| = \sqrt{nh} \left| \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,h,2}(X_i, X_j) \right|,$$

where the last equality follows by Lemma C.4. Considering the definition of $g_{y,h,2}(x_1, x_2)$ in (24), we recognize that $g_{y,h,2}(x_1, x_2) \in \mathcal{O}\left(\frac{1}{h^2}\right)$, as $h \rightarrow 0$. Let now $g_{y,2}^*(x_1, x_2) = h^2 g_{y,h,2}(x_1, x_2)$. Then, $g_{y,h,2}^*(x_1, x_2) = \mathcal{O}(1)$, as $h \rightarrow 0$. Furthermore, we have by Remark C.5 that the random variables $\{g_{y,h,2}(X_i, X_j)\}_{1 \leq i < j \leq n}$ are uncorrelated, whence the same holds for the random variables $\{g_{y,h,2}^*(X_i, X_j)\}_{1 \leq i < j \leq n}$. In consequence, we obtain that

$$\text{Var} \left(\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,2}^*(X_i, X_j) \right) = \mathcal{O}(n^{-2}).$$

This in turn implies by Chebyshev's inequality that

$$\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,2}^*(X_i, X_j) = \mathcal{O}_P(n^{-1}).$$

Therefore, we obtain with $h = o(n^{-1/5})$ that

$$\begin{aligned} \frac{2\sqrt{h}}{\sqrt{n(n-1)}} \sum_{1 \leq i < j \leq n} g_{y,h,2}(X_i, X_j) &= \frac{\sqrt{nh}}{h^2} \left(\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,2}^*(X_i, X_j) \right) \\ &= \mathcal{O}_P \left(\frac{1}{n^{1/2} h^{1.5}} \right) = o_P(1). \end{aligned}$$

Thus, we have shown that $\mathcal{S}(y) = o_P(1)$.

Second summand: Finally, we come to the second summand in (26). First of all, we observe that

$$\Theta_{y,h} = \int K(v) f_{d_{\mathcal{X},m}^2}(x+vh) dv = \mathbb{E} \left[\frac{1}{nh} \sum_{i=1}^n K \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) \right],$$

where $\{d_{\mathcal{X},m}^2(X_i)\}_{i=1}^n$ is a collection of i.i.d. random variables with density $f_{d_{\mathcal{X},m}^2}$. Since $f_{d_{\mathcal{X},m}^2}$ is assumed to be twice differentiable on $(y - \epsilon, y + \epsilon)$ and K is symmetric, i.e., $\int uK(u) du = 0$, it follows by a straightforward adaptation of Proposition 1.2 of Tsybakov (2008) that

$$\left| \Theta_{y,h} - f_{d_{\mathcal{X},m}^2}(y) \right| \leq Ch^2.$$

Here, C denotes a constant independent of n and h . We get that

$$\sqrt{nh} \left| \Theta_{y,h} - f_{d_{\mathcal{X},m}^2}(y) \right| = \mathcal{O}_P(\sqrt{nh^5}) = o_P(1),$$

as $h = o(n^{-1/5})$.

In conclusion, we have shown that

$$\left| \sqrt{nh} \left(\widehat{f}_{\delta_{\mathcal{X},m}^2} (y) - f_{d_{\mathcal{X},m}^2} (y) \right) - \sqrt{nh} \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) \right) \right| = o_P(1),$$

which yields that

$$\sqrt{nh} \left(\widehat{f}_{\delta_{\mathcal{X},m}^2} (y) - f_{d_{\mathcal{X},m}^2} (y) \right) \Rightarrow N \left(0, f_{d_{\mathcal{X},m}^2} (y) \int_{\mathbb{R}} K^2(u) du \right)$$

as claimed. ■

C.3 Proofs of the Auxiliary Lemmas from Section C.1

In this section, we gather the full proofs of Lemma C.1-C.7.

C.3.1 PROOF OF LEMMA C.1

In the course of this proof we have to differentiate between the cases $0 < m < 1$ and $m = 1$.

The case $0 < m < 1$: By assumption the kernel K is twice continuously differentiable. Using a Taylor series approximation, we find that

$$\begin{aligned} K \left(\frac{\delta_{\mathcal{X},m}^2(X_i) - y}{h} \right) &= K \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) + \frac{1}{h} K' \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) (\delta_{\mathcal{X},m}^2(X_i) - d_{\mathcal{X},m}^2(X_i)) \\ &\quad + \frac{1}{2h^2} K'' \left(\frac{\zeta_i - y}{h} \right) (\delta_{\mathcal{X},m}^2(X_i) - d_{\mathcal{X},m}^2(X_i))^2, \end{aligned}$$

for some ζ_i between $d_{\mathcal{X},m}^2(X_i)$ and $\delta_{\mathcal{X},m}^2(X_i)$. By Theorem 9 in Chazal et al. (2017) (whose conditions are met by assumption) it holds

$$\sup_{x \in \mathcal{X}} |\delta_{\mathcal{X},m}^2(x) - d_{\mathcal{X},m}^2(x)| = \mathcal{O}_P(1/\sqrt{n}). \quad (27)$$

In consequence, we obtain that

$$\begin{aligned} \widehat{f}_{\delta_{\mathcal{X},m}^2} (y) &= \frac{1}{nh} \sum_{i=1}^n \left[K \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) + \frac{1}{h} K' \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) (\delta_{\mathcal{X},m}^2(X_i) - d_{\mathcal{X},m}^2(X_i)) \right] \\ &\quad + \mathcal{O}_P(1/(nh^3)). \end{aligned}$$

Furthermore, it has been shown (see the proof of Theorem 5 in Chazal et al. (2017)) that for any $x \in \mathcal{X}$

$$\begin{aligned} \delta_{\mathcal{X},m}^2(x) - d_{\mathcal{X},m}^2(x) &= \frac{1}{m} \int_0^{F_x^{-1}(m)} F_x(t) - \widehat{F}_{x,n}(t) dt + \frac{1}{m} \int_{F_x^{-1}(m)}^{\widehat{F}_{x,n}^{-1}(m)} m - \widehat{F}_{x,n}(t) dt \\ &=: A_n(x) + R_n(x). \end{aligned} \quad (28)$$

In consequence, it remains to estimate $\sup_{x \in \mathcal{X}} |R_n(x)|$. Clearly, we have that

$$|R_n(x)| \leq \frac{1}{m} |S_n(x)| |T_n(x)|, \quad (29)$$

where

$$S_n(x) = \left| F_x^{-1}(m) - \widehat{F}_{x,n}^{-1}(m) \right| \text{ and } T_n(x) = \sup_t \left| F_x(t) - \widehat{F}_{x,n}(t) \right|.$$

Claim 1: It holds that

$$\sup_{x \in \mathcal{X}} |S_n(x)| = o_P \left(\left(\frac{\log(n)}{n} \right)^{1/(2b)} \right) \text{ as well as } \sup_{x \in \mathcal{X}} |T_n(x)| = \mathcal{O}_P \left(\sqrt{\frac{d}{n}} \right),$$

where $1 \leq b < 5$.

Combining Claim 1 with (29) yields $\sup_{x \in \mathcal{X}} |R_n(x)| = o_P \left(\frac{\log(n)^{1/(2b)}}{n^{1/2+1/(2b)}} \right)$, which gives the statement for $0 < m < 1$.

Proof of Claim 1: It has already been established in the proof of Theorem 9 in Chazal et al. (2017) that under the assumptions made

$$\sup_{x \in \mathcal{X}} |T_n(x)| = \mathcal{O}_P \left(\sqrt{\frac{d}{n}} \right).$$

Hence, it only remains to demonstrate the first equality. To this end, let $\xi_i \stackrel{i.i.d.}{\sim} \text{Uniform}(0,1)$, $1 \leq i \leq n$, and denote by H_n their empirical distribution function. Define $k = mn$. Then, it holds that $\widehat{F}_{x,n}^{-1}(m) \stackrel{\mathcal{D}}{=} F_x^{-1}(\xi_{(k)}) = F_x^{-1}(H_n^{-1}(m))$. Here, $\xi_{(k)}$ is the k -th order statistic and $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. Hence, we have for any $m > 0$ and $x \in \mathcal{X}$ that

$$\mathbb{P}(|S_n(x)| > \epsilon) = \mathbb{P}(|F_x^{-1}(H_n^{-1}(m)) - F_x^{-1}(m)| > \epsilon) \leq \mathbb{P}(\omega_x(|m - H_n^{-1}(m)|) > \epsilon),$$

where ω_x denote the modulus of continuity for F_x^{-1} . This means that for $u \in (0, 1)$

$$\omega_x(u) := \sup_{t, t' \in (0,1)^2, |t-t'| < u} |F_x^{-1}(t) - F_x^{-1}(t')|.$$

By assumption, there exists a constant $\kappa \in \mathbb{R}$ such that $\omega_{\mathcal{X}}(u) = \sup_{x \in \mathcal{X}} \omega_x(u) \leq \kappa u^{1/b}$ for all $u \in (0, 1)$. Hence, we find that

$$\begin{aligned} \mathbb{P}(\omega_x(|m - H_n^{-1}(m)|) > \epsilon) &\leq \mathbb{P}\left(|m - H_n^{-1}(m)| > \left(\frac{\epsilon}{\kappa}\right)^b\right) \\ &\leq 2 \exp\left(-\frac{n \left(\frac{\epsilon}{\kappa}\right)^{2b}}{m} \frac{1}{1 + \frac{2\left(\frac{\epsilon}{\kappa}\right)^b}{3m}}\right), \end{aligned} \quad (30)$$

where the last line follows from Shorack and Wellner (2009) (Inequality 1 on Page 453 and Proposition 1, page 455). Next, we observe that

$$\mathbb{P}\left(\sup_{x \in \mathcal{X}} |S_n(x)| > \epsilon\right) \leq \mathbb{P}\left(\sup_{x \in \mathcal{X}} \omega_x(|m - H_n^{-1}(m)|) > \epsilon\right) \leq \mathbb{P}\left(\kappa |m - H_n^{-1}(m)|^{1/b} > \epsilon\right).$$

Using (30), we find that

$$\mathbb{P}\left(\sup_{x \in \mathcal{X}} |S_n(x)| > \epsilon\right) \leq 2 \exp\left(-\frac{n\left(\frac{\epsilon}{\kappa}\right)^{2b}}{m} \frac{1}{1 + \frac{2\left(\frac{\epsilon}{\kappa}\right)^b}{3m}}\right) \leq 2 \exp\left(-\frac{3n}{5} \left(\frac{\epsilon}{\kappa}\right)^{2b}\right),$$

where we used that $m \leq 1$ and $\epsilon/\kappa < 1$ for epsilon small enough. Let $\epsilon = \tau \left(\frac{\log(n)}{n}\right)^{1/(2b)}$. It follows that

$$\mathbb{P}\left(\sup_{x \in \mathcal{X}} |S_n(x)| > \epsilon\right) \leq 2 \exp\left(-\frac{3}{5} \left(\frac{\tau}{\kappa}\right)^{2b} \log(n)\right),$$

and thus

$$\sup_{x \in \mathcal{X}} |S_n(x)| = o_P\left(\left(\frac{\log(n)}{n}\right)^{1/(2b)}\right),$$

which yields Claim 1.

The case $m = 1$: Similar as for $0 < m < 1$, we find that

$$\begin{aligned} K\left(\frac{\delta_{\mathcal{X},1}^2(X_i) - y}{h}\right) &= K\left(\frac{d_{\mathcal{X},1}^2(X_i) - y}{h}\right) + \frac{1}{h} K' \left(\frac{d_{\mathcal{X},1}^2(X_i) - y}{h}\right) (\delta_{\mathcal{X},1}^2(X_i) - d_{\mathcal{X},1}^2(X_i)) \\ &\quad + \frac{1}{2h^2} K'' \left(\frac{\zeta_i - y}{h}\right) (\delta_{\mathcal{X},1}^2(X_i) - d_{\mathcal{X},1}^2(X_i))^2, \end{aligned}$$

for some ζ_i between $d_{\mathcal{X},1}^2(X_i)$ and $\delta_{\mathcal{X},1}^2(X_i)$. While Theorem 9 in Chazal et al. (2017) is only stated for $m < 1$, it is straightforward to adapt its proof to the case $m = 1$. Hence, we obtain

$$\sup_{x \in \mathcal{X}} |\delta_{\mathcal{X},1}^2(x) - d_{\mathcal{X},1}^2(x)| = \mathcal{O}_P(1/\sqrt{n}). \quad (31)$$

Furthermore, we note that for $x \in \mathcal{X}$

$$\delta_{\mathcal{X},1}^2(x) - d_{\mathcal{X},1}^2(x) = \int_0^{D_x} F_x(t) - \widehat{F}_{x,n}(t) dt, \quad (32)$$

where $[0, D_x]$ denotes the support of F_x . In combination with our previous considerations, we find that

$$\widehat{f}_{\delta_{\mathcal{X},1}^2}(y) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{d_{\mathcal{X},1}^2(X_i) - y}{h}\right) + \frac{1}{h} K' \left(\frac{d_{\mathcal{X},1}^2(X_i) - y}{h}\right) A_n(x) \right] + \mathcal{O}_P\left(\frac{1}{nh^3}\right),$$

which yields the claim. ■

C.3.2 PROOF OF LEMMA C.2

First, we consider $V_n^{(1)}(x)$. Clearly, we have

$$\begin{aligned} V_n^{(1)}(y) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \frac{1}{2h} \left(K \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) + K \left(\frac{d_{\mathcal{X},m}^2(X_j) - y}{h} \right) \right) \\ &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,h}^{(1)}(X_i, X_j). \end{aligned}$$

Next, we come to $V_n^{(2)}(y)$. We have that

$$\begin{aligned} V_n^{(2)}(y) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} K' \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) \frac{1}{m} \int_0^{F_{X_i}^{-1}(m)} F_{X_i}(t) - \widehat{F}_{X_i,n}(t) dt \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) \int_0^{F_{X_i}^{-1}(m)} F_{X_i}(t) - \mathbb{1}_{\{\|X_i - X_j\|^2 \leq t\}} dt. \end{aligned}$$

Further, we obtain that

$$V_n^{(2)}(y) = \frac{1}{n^2} \sum_{1 \leq i < j \leq n} g_{y,h}^{(2)}(X_i, X_j) + \frac{1}{n^2} \sum_{i=1}^n \frac{1}{mh^2} \left[K' \left(\frac{d_{\mathcal{X},m}^2(X_i) - y}{h} \right) \int_0^{F_{X_i}^{-1}(m)} F_{X_i}(t) - 1 dt \right].$$

We note that K is twice differentiable and \mathcal{X} is compact, i.e.,

$$\left| \int_0^{F_{x_1}^{-1}(m)} F_{x_2}(t) - \mathbb{1}_{\{\|x_1 - x_2\|^2 \leq t\}} dt \right| \leq \text{diam}(\mathcal{X}) < \infty$$

for all $x_1, x_2 \in \mathcal{X}$. This yields that

$$\begin{aligned} V_n^{(2)}(y) &= \frac{2}{n^2} \sum_{1 \leq i < j \leq n} g_{y,h}^{(2)}(X_i, X_j) + \mathcal{O}_P \left(\frac{1}{nh^2} \right) \\ &= \frac{n-1}{n} \left(\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,h}^{(2)}(X_i, X_j) \right) + \mathcal{O}_P \left(\frac{1}{nh^2} \right) \\ &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g_{y,h}^{(2)}(X_i, X_j) + \mathcal{O}_P \left(\frac{1}{nh^2} \right). \end{aligned}$$

■

C.3.3 PROOF OF LEMMA C.4

Since $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$, the claim follows by the Hoeffding decomposition (Van der Vaart, 2000, Lemma 11.11) once we have shown that

1. $\Theta_{y,h} = \mathbb{E}[U_n]$

2. $g_{y,h,1}(x_1) = \mathbb{E} [g_{y,h}(x_1, Z_1)] - \Theta_{y,h}$, where $Z_1 \sim \mu_{\mathcal{X}}$.

First equality: We start by verifying the first equality. Clearly,

$$\mathbb{E} [U_n] = \mathbb{E} \left[g_{y,h}^{(1)}(X_1, X_2) \right] + \mathbb{E} \left[g_{y,h}^{(2)}(X_1, X_2) \right].$$

Since $X_1, X_2 \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$, we obtain that

$$\begin{aligned} \mathbb{E} \left[g_{y,h}^{(1)}(X_1, X_2) \right] &= \mathbb{E} \left[\frac{1}{2h} \left(K \left(\frac{d_{\mathcal{X},m}^2(X_1) - y}{h} \right) + K \left(\frac{d_{\mathcal{X},m}^2(X_2) - y}{h} \right) \right) \right] \\ &= \int \frac{1}{h} K \left(\frac{u - y}{h} \right) d(d_{\mathcal{X},m}^2 \# \mu_{\mathcal{X}})(u). \end{aligned}$$

Here, the last equality follows by the change-of-variables formula ($d_{\mathcal{X},m}^2 \# \mu_{\mathcal{X}}$ denotes the pushforward measure of $\mu_{\mathcal{X}}$ with respect to $d_{\mathcal{X},m}^2$). By assumption the measure $d_{\mathcal{X},m}^2 \# \mu_{\mathcal{X}}$ possesses a density $f_{d_{\mathcal{X},m}^2}$ with respect to the Lebesgue measure. Hence,

$$\mathbb{E} \left[g_{y,h}^{(1)}(X_1, X_2) \right] = \int \frac{1}{h} K \left(\frac{u - y}{h} \right) f_{d_{\mathcal{X},m}^2}(u) du = \int K(v) f_{d_{\mathcal{X},m}^2}(y + vh) dv.$$

As $X_1, X_2 \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$, we obtain for the second summand that

$$\mathbb{E} \left[g_{y,h}^{(2)}(X_1, X_2) \right] = \mathbb{E} \left[\frac{1}{mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(X_1) - y}{h} \right) \int_0^{F_{X_1}^{-1}(m)} F_{X_1}(t) - \mathbf{1}_{\{\|X_1 - X_2\|^2 \leq t\}} dt \right]$$

Since X_1 and X_2 are independent, the Theorem of Tonelli/Fubini (Billingsley, 2013, Thm. 18) yields that

$$\begin{aligned} \mathbb{E} \left[g_{y,h}^{(2)}(X_1, X_2) \right] &= \mathbb{E}_{X_1} \left[\frac{1}{mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(X_1) - y}{h} \right) \int_0^{F_{X_1}^{-1}(m)} F_{X_1}(t) - \mathbb{E}_{X_2} [\mathbf{1}_{\{\|X_1 - X_2\|^2 \leq t\}}] dt \right] \\ &= 0. \end{aligned}$$

Combining our results, we find that $\mathbb{E} [U_n] = \Theta_{y,h}$.

Second equality: Recall that $Z_1 \sim \mu_{\mathcal{X}}$. We demonstrate that

$$g_{y,h,1}(x_1) = \mathbb{E} [g_{y,h}(x_1, Z_1)] - \Theta_{y,h} = \mathbb{E} \left[g_{y,h}^{(1)}(x_1, Z_1) \right] + \mathbb{E} \left[g_{y,h}^{(2)}(x_1, Z_1) \right] - \Theta_{y,h}.$$

Once again, we consider the two summands separately. We observe that

$$\mathbb{E} \left[g_{y,h}^{(1)}(x_1, Z_1) \right] = \frac{1}{2h} K \left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h} \right) + \frac{1}{2} \Theta_{y,h}.$$

Here, the last equality follows by our previous considerations for $\mathbb{E}[U_n]$. For the second summand, it follows that

$$\begin{aligned} \mathbb{E} \left[g_{y,h}^{(2)}(x_1, Z_1) \right] &= \mathbb{E} \left[\frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h} \right) \int_0^{F_{x_1}^{-1}(m)} F_{x_1}(t) - \mathbf{1}_{\{\|x_1 - Z_1\|^2 \leq t\}} dt \right] \\ &\quad + \mathbb{E} \left[\frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \int_0^{F_{Z_1}^{-1}(m)} F_{Z_1}(t) - \mathbf{1}_{\{\|Z_1 - x_1\|^2 \leq t\}} dt \right] =: T_1 + T_2. \end{aligned}$$

The Theorem of Tonelli/Fubini (Billingsley, 2013, Thm. 18) shows that

$$T_1 = \frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h} \right) \int_0^{F_{x_1}^{-1}(m)} F_{x_1}(t) - \mathbb{E} \left[\mathbf{1}_{\{\|x_1 - Z_1\|^2 \leq t\}} \right] dt = 0.$$

Furthermore, we obtain for $Z_2 \sim \mu_{\mathcal{X}}$ independent of Z_1 that

$$\begin{aligned} T_2 &= \mathbb{E}_{Z_1} \left[\frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \int_0^{F_{Z_1}^{-1}(m)} \mathbb{E}_{Z_2} \left[\mathbf{1}_{\{\|Z_1 - Z_2\|^2 \leq t\}} \right] - \mathbf{1}_{\{\|Z_1 - x_1\|^2 \leq t\}} dt \right] \\ &= \mathbb{E}_{Z_1} \left[\frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \mathbb{E}_{Z_2} \left[\int_0^{F_{Z_1}^{-1}(m)} \mathbf{1}_{\{\|Z_1 - Z_2\|^2 \leq t\}} - \mathbf{1}_{\{\|Z_1 - x_1\|^2 \leq t\}} dt \right] \right], \end{aligned}$$

where the last step follows by the theorem of Tonelli/Fubini (Billingsley, 2013, Thm. 18). Moreover, the determination of the integral in the above expression yields that

$$T_2 = \mathbb{E}_{Z_1} \left[\frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \left(\|x_1 - Z_1\|^2 \wedge F_{Z_1}^{-1}(m) - \mathbb{E}_{Z_2} \left[\|Z_1 - Z_2\|^2 \wedge F_{Z_1}^{-1}(m) \right] \right) \right].$$

Combining all of our results, we finally get that

$$g_{y,h,1}(x_1) = \frac{1}{2h} K \left(\frac{d_{\mathcal{X},m}^2(x_1) - y}{h} \right) - \frac{1}{2} \Theta_{y,h} + \mathbb{E}_{Z_1} \left[\frac{1}{2mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \Psi(x_1, Z_1) \right],$$

as claimed. ■

C.3.4 PROOF OF LEMMA C.6

Let $x_1 \in \mathcal{X}$ be arbitrary. We observe that for any $z_1, z_2 \in \mathcal{X}$

$$\begin{aligned} |\Psi(x_1, z_1) - \Psi(x_1, z_2)| &\leq \left| \|x_1 - z_1\|^2 \wedge F_{z_1}^{-1}(m) - \|x_1 - z_2\|^2 \wedge F_{z_2}^{-1}(m) \right| \\ &\quad + \left| \mathbb{E}_{Z_2} \left[\|Z_2 - z_1\|^2 \wedge F_{z_1}^{-1}(m) \right] - \mathbb{E}_{Z_2} \left[\|Z_2 - z_2\|^2 \wedge F_{z_2}^{-1}(m) \right] \right| \\ &=: \Psi_1(z_1, z_2) - \Psi_2(z_1, z_2). \end{aligned}$$

In the following, we consider Ψ_1 and Ψ_2 separately. We have that for $z_1, z_2 \in \mathcal{X}$

$$\begin{aligned} \Psi_1(z_1, z_2) &\leq \left| \|x_1 - z_1\|^2 - \|x_1 - z_2\|^2 \right| + \left| F_{z_1}^{-1}(m) - F_{z_2}^{-1}(m) \right| \\ &\leq D \left| \|x_1 - z_1\| - \|x_1 - z_2\| \right| + 2\sqrt{D} \|z_1 - z_2\|, \end{aligned}$$

where the last inequality follows with $D = \text{diam}(\mathcal{X}) < \infty$ and Lemma 8 in Chazal et al. (2017). In particular, note that in the current setting we have that

$$\sup_{t \in (0,1)} \sup_{x \in \mathcal{X}} F_x^{-1}(t) \leq D < \infty.$$

In consequence, we obtain that for $z_1, z_2 \in \mathcal{X}$

$$\Psi_1(z_1, z_2) \leq D \|\|z_1 - z_2\|\| + 2\sqrt{D} \|z_1 - z_2\| \leq C \|z_1 - z_2\|,$$

where C denotes a constant that only depends on \mathcal{X} .

Next, we observe

$$\Psi_2(z_1, z_2) \leq \mathbb{E}_{Z_2} \left[\left| \|Z_2 - z_1\|^2 \wedge F_{z_1}^{-1}(m) - \|Z_2 - z_2\|^2 \wedge F_{z_2}^{-1}(m) \right| \right].$$

Considering our previous calculation, we immediately obtain that

$$\Psi_2(z_1, z_2) \leq \mathbb{E}_{Z_2} [C \|z_1 - z_2\|] = C \|z_1 - z_2\|,$$

where C denotes the same constant as previously. Combining our results, we find that

$$|\Psi(x_1, z_1) - \Psi(x_1, z_2)| \leq C \|z_1 - z_2\|,$$

where the constant C only depends on \mathcal{X} and not on x_1 . This yields the claim. ■

C.3.5 PROOF OF LEMMA C.7

Next, we derive (25) using Lyapunov's Central Limit Theorem for triangular arrays (Billingsley, 1979, Sec. 27). To this end, we define $z_{in} := 2g_{y,h,1}(X_i)$, $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_{in}$ and $\sigma_{in}^2 := \text{Var}(z_{in})$. Clearly, for n fixed the z_{in} 's are independent and identically distributed. In order to check the assumptions of Lyapunov's Central Limit Theorem, it remains to find $r > 2$ such that

$$\rho_{1n} := \mathbb{E} [|z_{1n} - \mathbb{E}[z_{1n}]|^r] < \infty \tag{33}$$

and

$$\frac{n\rho_{1n}}{(n\sigma_{1n}^2)^{r/2}} \rightarrow 0, \tag{34}$$

as $n \rightarrow \infty$.

Calculation of σ_{1n}^2 : The next step is to consider σ_{1n}^2 . As $\mathbb{E}[z_{1n}] = 0$ by construction, we find that

$$\sigma_{1n} = \mathbb{E} [|z_{1n}|^2] = \mathbb{E} \left[|2g_{y,h,1}(X_1)|^2 \right] = \mathbb{E} \left[|T_3 + T_4|^2 \right],$$

where

$$T_3 := \frac{1}{h} K \left(\frac{d_{\mathcal{X},m}^2(X_1) - y}{h} \right) - \Theta_{y,h} \tag{35}$$

and

$$T_4 := \mathbb{E}_{Z_1} \left[\frac{1}{mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \Psi(X_1, Z_1) \right]. \quad (36)$$

Here, $\Psi(x_1, x_2)$ is the function defined in (23). Obviously, we obtain that

$$\sigma_{1n}^2 = \mathbb{E} [T_3^2] + \mathbb{E} [T_4^2] + 2\mathbb{E} [T_3 T_4].$$

In the following, we treat each of these summands separately.

First summand: Considering the first term, we see that

$$\mathbb{E} [T_3^2] = \mathbb{E} \left[\left| \frac{1}{h} K \left(\frac{d_{\mathcal{X},m}^2(X_1) - y}{h} \right) - \Theta_{y,h} \right|^2 \right].$$

which is essentially the variance of the kernel density estimator of the real valued random variable $d_{\mathcal{X},m}^2(X_1)$. Hence, one can show using standard arguments (see e.g. Silverman (2018, Sec. 3.3)) that

$$\mathbb{E} [|T_3|^2] = \frac{f_{d_{\mathcal{X},m}^2}(y)}{h} \int |K(u)|^2 du + o\left(\frac{1}{h}\right).$$

as $h \rightarrow 0$.

Second summand: Next, we consider $\mathbb{E} [|T_4|^2]$. We have that

$$\mathbb{E} [|T_4|^2] = \mathbb{E} \left[\left| \mathbb{E}_{Z_1} \left[\frac{1}{mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \Psi(X_1, Z_1) \right] \right|^2 \right].$$

Recall that $Z_1 \sim \mu_{\mathcal{X}}$ and that $\mu_{\mathcal{X}}$ has, by assumption, a Lipschitz continuous Lebesgue density. Denote this density by $g_{\mu_{\mathcal{X}}}$. Then, it follows that

$$\begin{aligned} & \mathbb{E}_{Z_1} \left[\frac{1}{mh^2} K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \Psi(X_1, Z_1) \right] \\ &= \frac{1}{mh^2} \int_{\mathcal{X}} K' \left(\frac{d_{\mathcal{X},m}^2(z) - y}{h} \right) \Psi(X_1, z) g_{\mu_{\mathcal{X}}}(z) d\lambda^d(z) \end{aligned} \quad (37)$$

Next, we realize that

$$\sup_{x_1, x_2} |\Psi(x_1, x_2)| \leq D$$

and hence there is a constant $0 < C < \infty$ such that

$$\max \left\{ \sup_{x \in \mathcal{X}} g_{\mu_{\mathcal{X}}}(x), \sup_{x_1, x_2} |\Psi(x_1, x_2)| \right\} < C. \quad (38)$$

Further, it follows by Lemma C.6 that the function $\Psi_{x_1}^* : \mathcal{X} \rightarrow \mathbb{R}$, $z \mapsto \Psi(x_1, z)$ is Lipschitz continuous for all $x_1 \in \mathcal{X}$ with a Lipschitz constant that does not depend on x_1 . This in combination with the Lipschitz continuity of $g_{\mu_{\mathcal{X}}}$ and (38) implies that the function

$$\psi_{x_1} : \mathcal{X} \rightarrow \mathbb{R}, \quad z \mapsto \Psi(x_1, z)g_{\mu_{\mathcal{X}}}(z)$$

is Lipschitz continuous for all $x_1 \in \mathcal{X}$ with a Lipschitz constant that does not depend on x_1 . We have that the function $x \mapsto d_{\mathcal{X},m}^2(x)$ is coercive, that $d_{\mathcal{X},m}^2$ is $C^{2,1}$ on an open neighborhood of $\Gamma_y = d_{\mathcal{X},m}^{-2}(y)$ and that $\nabla d_{\mathcal{X},m}^2 \neq 0$ on Γ_y by assumption. By Condition 2, there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$

$$\int_{\Gamma_y} \left| \mathbb{1}_{\{\Phi(0,x) \in \mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x) \in \mathcal{X}\}} \right| d\mathcal{H}^{d-1}(x) \leq C_y |v|,$$

where Φ denotes the canonical level set flow of Γ_y and C_y denotes a finite constant that depends on y and $d_{\mathcal{X},m}^2$. Furthermore, the kernel K is twice continuously differentiable and $\text{supp}(K) = [-1, 1]$. Since K is also even, by assumption, it follows that K' is odd, i.e.

$$\int_{-1}^1 K'(z) dz = 0.$$

Thus, we find by Theorem D.1 that there exists some constants $c_y > 0$ and $h_0 > 0$ (depending on $d_{\mathcal{X},m}^2$, y and \mathcal{X}) such that for any $h < h_0$ we obtain that

$$\left| \mathbb{E}_{Z_1} \left[K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \Psi(X_1, Z_1) \right] \right| \leq c_y h^2.$$

In consequence, we find that for h small enough

$$\mathbb{E} [|T_4|^2] \leq \mathbb{E} \left[\frac{1}{m^2 h^4} \left| \mathbb{E}_{Z_1} \left[K' \left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h} \right) \Psi(X_1, Z_1) \right] \right|^2 \right] \leq \frac{c_y}{m^2}.$$

This in particular shows that $\mathbb{E} [|T_4|^2] = \mathcal{O}(1)$ as $h \rightarrow 0$.

Third summand: By Hölder's inequality, we obtain that

$$\mathbb{E} [T_3 T_4] \leq \mathbb{E} [|T_3 T_4|] \leq (\mathbb{E} [|T_3|^2])^{1/2} (\mathbb{E} [|T_4|^2])^{1/2}.$$

Plugging in our previous findings, we find that

$$\mathbb{E} [T_3 T_4] = \mathcal{O} \left(\frac{1}{\sqrt{h}} \right) \cdot \mathcal{O}(1) = \mathcal{O} \left(\frac{1}{\sqrt{h}} \right).$$

as $h \rightarrow 0$. In consequence, we find that

$$\sigma_{1n}^2 = \frac{\int d_{\mathcal{X},m}^2(x)}{h} \int |K(u)|^2 du + o \left(\frac{1}{h} \right).$$

This concludes our consideration of σ_{1n}^2 .

Calculation of third moments: We choose $r = 3$ and consider $\rho_{1n} = \mathbb{E}[|z_{1n} - \mathbb{E}[z_{1n}]|^r]$. By construction $\mathbb{E}[z_{1n}] = 0$. Thus, we obtain

$$\rho_{1n} = \mathbb{E}[|z_{1n}|^3] = \mathbb{E}[|g_{y,h,1}(X_1)|^3] = \mathbb{E}[|T_3 + T_4|^3],$$

where T_3 and T_4 denote the terms defined in (35) and (36), respectively. Furthermore, it follows that

$$\rho_{1n} \leq \mathbb{E}[(|T_3| + |T_4|)^3] \leq 8\mathbb{E}[|T_3|^3] + 8\mathbb{E}[|T_4|^3].$$

Considering the first summand, this yields that

$$\mathbb{E}[|T_3|^3] = \mathbb{E}\left[\left|\frac{1}{h}K\left(\frac{d_{\mathcal{X},m}^2(X_1) - y}{h}\right) - \Theta_{y,h}\right|^3\right],$$

which is the third moment of the kernel density estimator of the real valued random variable $d_{\mathcal{X},m}^2(X_1)$. In particular, one can show using standard arguments that

$$\mathbb{E}[|T_3|^3] \leq \frac{8f_{d_{\mathcal{X},m}^2}(x)}{h^2} \int |K(u)|^3 du + o\left(\frac{1}{h^2}\right).$$

It remains to consider $\mathbb{E}[|T_4|^3]$. We have already shown that for $h \rightarrow 0$

$$\left|\frac{1}{2mh^2}\mathbb{E}_{Z_1}\left[K'\left(\frac{d_{\mathcal{X},m}^2(Z_1) - y}{h}\right)\Psi(X_1, Z_1)\right]\right| = \mathcal{O}(1).$$

Consequently, this implies that $\mathbb{E}[|T_4|^3] = \mathcal{O}(1)$. Hence, we obtain that

$$\rho_{1n} \leq \frac{8f_{d_{\mathcal{X},m}^2}(y)}{h^2} \int_{-1}^1 |K(u)|^3 du + o\left(\frac{1}{h^2}\right).$$

Applying Lyapunov's CLT: Now that we have calculated ρ_{1n} and σ_{1n}^2 , we can verify the remaining assumption of Lyapunov's Central Limit Theorem for triangular array's (Billingsley, 1979, Sec. 26). First of all, we observe that $\rho_{1n} < \infty$, since K , K' and Ψ are continuous and compactly supported. Furthermore, we obtain

$$\frac{n\rho_{1n}}{(n\sigma_{1n}^2)^{3/2}} \leq \frac{\frac{8nf_{d_{\mathcal{X},m}^2}(x)}{h^2} \int |K(u)|^3 du + o\left(\frac{n}{h^2}\right)}{\left(\frac{nf_{d_{\mathcal{X},m}^2}(x)}{2h} \int |K(u)|^2 du + o\left(\frac{n}{h}\right)\right)^{3/2}} = \mathcal{O}\left((nh)^{-1/2}\right) \rightarrow 0,$$

if $nh \rightarrow \infty$. In consequence, Lyapunov's Central Limit Theorem for triangular arrays is applicable. It yields that

$$\frac{\bar{z}_n - \mathbb{E}[\bar{z}_n]}{\sqrt{\text{Var}(\bar{z}_n)}} \xrightarrow{D} N(0, 1).$$

This in turn implies that

$$\sqrt{nh} \left(\frac{2}{n} \sum_{i=1}^n g_{y,h,1}(X_i) \right) \Rightarrow N \left(0, f_{d_{\mathcal{X},m}^2}(y) \int_{-1}^1 |K(u)|^2 du \right)$$

which gives the claim. ■

Appendix D. Some Geometric Measure Theory

In the proof of Lemma C.7, we need to bound the term (37):

$$\mathcal{I}(y) := \int_{\mathcal{X}} K' \left(\frac{d_{\mathcal{X},m}^2(z) - y}{h} \right) \Psi(X_1, z) g_{\mu_{\mathcal{X}}}(z) d\lambda^d(z),$$

where $g_{\mu_{\mathcal{X}}}$ denotes the Lebesgue density of $\mu_{\mathcal{X}}$ (which exists by assumption) and $X_1 \sim \mu_{\mathcal{X}}$. Since the kernel K and thus also its derivative K' are supported on $[-1, 1]$, we obtain

$$\mathcal{I}(y) = \int_{A_h(y)} K' \left(\frac{d_{\mathcal{X},m}^2(z) - y}{h} \right) \Psi(X_1, z) g_{\mu_{\mathcal{X}}}(z) d\lambda^d(z),$$

where

$$A_h(y) := \{z \in \mathcal{X} \mid y - h \leq d_{\mathcal{X},m}^2(z) \leq y + h\} = (d_{\mathcal{X},m}^2)^{-1} [y - h, y + h] \cap \mathcal{X}. \quad (39)$$

In the following, we will show how to control such integrals over thickened level sets such as $A_h(y)$ for small h . More precisely, we prove the subsequent theorem that has already been applied to bound the term $\mathcal{I}(y)$ in the proof of Lemma C.7.

Theorem D.1 *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact set. Let $g: \mathcal{X} \rightarrow [-\alpha, \alpha]$ be α -Lipschitz continuous and suppose that $k: \mathbb{R} \rightarrow [-\alpha, \alpha]$ for some $\alpha > 0$. Assume that $\text{supp}(k) = [-1, 1]$ and $\int k(s) ds = 0$. Let $d: \mathbb{R}^d \rightarrow \mathbb{R}$ be a coercive function, i.e., $\lim_{\|x\| \rightarrow \infty} d(x) = \infty$, with level sets $\Gamma_y = d^{-1}\{y\}$ for $y \in \mathbb{R}$. Call $y \in \mathbb{R}$ a $C^{2,1}$ -regular bounded value of d with respect to \mathcal{X} if*

C.1 Γ_y has an open neighborhood on which d is $C^{2,1}$,

C.2 $\nabla d \neq 0$ on Γ_y .

C.3 There exists $h_0^* > 0$ and such that for all $-h_0^* < v < h_0^*$

$$\int_{\Gamma_y} |\mathbb{1}_{\{\Phi(0,x) \in \mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x) \in \mathcal{X}\}}| d\mathcal{H}^{d-1}(x) \leq C_y |v|,$$

where Φ denotes the canonical level set flow of Γ_y and C_y denotes a constant that only depends on the function d , the variable y and the underlying space \mathcal{X} .

If y is a $C^{2,1}$ -regular bounded value of d with respect to \mathcal{X} , then

$$\left| \int_{\mathcal{X}} k \left(\frac{\mathfrak{d}(x) - y}{h} \right) g(x) d\lambda^d(x) \right| \leq c_y h^2 \quad (40)$$

for some $c_y > 0$ and any $0 < h < h_0$, where c_y and $h_0 > 0$ only depend on d , y , α and \mathcal{X} (and not on k and g explicitly).

The proof of Theorem D.1 consists of two steps, each of which is formulated as an independent lemma (see Section D.1).

Step 1: Splitting the integration (Lemma D.2).

We first note that integration over $A_h(y)$ can be split into integrating first over the surface $(\mathfrak{d}_{\mathcal{X},m}^2)^{-1}(v) \cap \mathcal{X}$ with respect to the $(d-1)$ -dimensional Hausdorff measure \mathcal{H}^{d-1} (see Federer (1969); Morgan (2016) for an introduction) and afterwards over $v \in [y-h, y+h]$.

Step 2: Local Lipschitz continuity (Lemma D.3).

We prove that the integral of a bounded, α -Lipschitz function $g : \mathcal{X} \subset \mathbb{R}^d \rightarrow [-\alpha, \alpha]$ over the level set of a $C^{2,1}$ -regular bounded value \mathfrak{d} with respect to \mathcal{X} , denoted as y , is locally Lipschitz continuous in y . More precisely, we prove that there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$ it holds that

$$\left| \int_{\Gamma_{y+v} \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) \right| \leq C_y |v|,$$

where $C_y > 0$ denotes a constant that only depends on \mathfrak{d} , y , α and \mathcal{X} .

D.1 Auxiliary Lemmas Representing Step 1 and Step 2

Lemma D.2 *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a Lipschitz continuous function. Let $h > 0$, $\mathcal{X} \subset \mathbb{R}^d$ a compact space and $g : \mathbb{R}^d \rightarrow \mathbb{R}$ such that the function*

$$x \mapsto \frac{|g(x)|}{\|\nabla f(x)\|} \mathbb{1}_{\{x \in \mathcal{X} : |f(x)| \leq h\}} \quad (41)$$

is integrable with respect to λ^d . Then, it follows that

$$\int_{\{x \in \mathcal{X} : |f(x)| \leq h\}} g(x) d\lambda^d(x) = \int_{-h}^h \int_{f^{-1}(v) \cap \mathcal{X}} \frac{g(x)}{\|\nabla f(x)\|} d\mathcal{H}^{d-1}(x) dv,$$

where \mathcal{H}^{d-1} denotes the $(d-1)$ -dimensional Hausdorff measure.

Proof First of all, we observe that

$$\int_{\{x \in \mathcal{X} : |f(x)| \leq h\}} g(x) d\lambda^d(x) = \int_{\mathbb{R}^d} \frac{g(x)}{\|\nabla f(x)\|} \mathbb{1}_{\{x \in \mathcal{X} : |f(x)| \leq h\}} \|\nabla f(x)\| d\lambda^d(x).$$

Since the function defined in (41) is integrable, it follows by the *Co-Area Formula* (see Federer (1969, Thm. 3.2.12), where the *k-dimensional Jacobian* of f is $\|\nabla f\|$ in this setting) that

$$\begin{aligned} & \int_{\mathbb{R}^d} \frac{g(x)}{\|\nabla f(x)\|} \mathbb{1}_{\{x \in \mathcal{X} : |f(x)| \leq h\}} \|\nabla f(x)\| d\lambda^d(x) \\ &= \int_{-\infty}^{\infty} \int_{f^{-1}(v)} \frac{g(x)}{\|\nabla f(x)\|} \mathbb{1}_{\{x \in \mathbb{R}^d : -h \leq f(x) \leq h\}} \mathbb{1}_{\{x \in \mathcal{X}\}} d\mathcal{H}^{d-1}(x) dv \\ &= \int_{-h}^h \int_{f^{-1}(v) \cap \mathcal{X}} \frac{g(x)}{\|\nabla f(x)\|} d\mathcal{H}^{d-1}(x) dv. \end{aligned}$$

This yields the claim. ■

Lemma D.3 *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact set. Let $g : \mathcal{X} \rightarrow [-\alpha, \alpha]$ be an α -Lipschitz function for some $\alpha > 0$. Let $\mathfrak{d} : \mathbb{R}^d \rightarrow \mathbb{R}$ be a coercive function and $y \in \mathbb{R}$ a $C^{2,1}$ -regular bounded value of \mathfrak{d} with respect to \mathcal{X} . Let $\Gamma_y = \mathfrak{d}^{-1}\{y\}$. Then, it holds that*

$$\left| \int_{\Gamma_{y+v} \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) \right| \leq C_y |v|, \quad (42)$$

for some $C_y > 0$ and any $-h_0 < v < h_0$, where C_y and $h_0 > 0$ only depend on \mathfrak{d} , y , α and \mathcal{X} (and not on g explicitly).

Proof Before we prove (42), we ensure that the statement is well defined and prove that under the assumptions made

$$\int_{\Gamma_y \cap \mathcal{X}} |g(x)| d\mathcal{H}^{d-1}(x) \leq \alpha \int_{\Gamma_y \cap \mathcal{X}} d\mathcal{H}^{d-1}(x) < \infty.$$

To this end, we observe that $\mathcal{H}^{d-1}(\mathfrak{d}^{-1}(\{y\}) \cap \mathcal{X}) \leq \mathcal{H}^{d-1}(\mathfrak{d}^{-1}(\{y\}))$. As \mathfrak{d} is coercive it follows that the set $\mathfrak{d}^{-1}([0, y])$ is bounded. Hence, the same holds true for $\mathfrak{d}^{-1}(\{y\})$. Furthermore, as \mathfrak{d} is $C^{2,1}$ in an open neighborhood of the level set Γ_y and $\nabla \mathfrak{d} \neq 0$ on Γ_y , it follows that Γ_y is a compact C^1 -manifold of dimension $d - 1$ (Villanacci et al., 2002, Thm. 9), which obviously has finite volume (and hence finite $(d - 1)$ -dimensional Hausdorff measure (Federer, 1969; Morgan, 2016)).

Now, we focus on proving the statement (42). By assumption, \mathfrak{d} is $C^{2,1}$ on an open neighborhood U of Γ_y with $\|\nabla \mathfrak{d}\| > 0$ on Γ_y . In consequence, there exists $h'_0 > 0$ such that $\mathfrak{d}^{-1}([y - h'_0, y + h'_0]) \subset U$ and $\|\nabla \mathfrak{d}\| > 0$ on $\mathfrak{d}^{-1}([y - h'_0, y + h'_0])$. This means that the function

$$\varphi(u) : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad u \mapsto \frac{\nabla \mathfrak{d}(u)}{\|\nabla \mathfrak{d}(u)\|^2}$$

is $C^{1,1}(\mathfrak{d}^{-1}([y - h'_0, y + h'_0]), \mathbb{R}^d)$. By Theorem A.6 in Eldering (2013) (or more generally by Cauchy-Lipschitz's theory (Hirsch and Smale, 1974; Amann, 2011)) there exists $0 < h_0 \leq h'_0$

such that one can construct a flow $\Phi : [-h_0, h_0] \times W \rightarrow \mathbb{R}^d$ with

$$\begin{cases} \frac{\partial}{\partial t} \Phi(t, x) = \frac{\nabla \mathfrak{d}(\Phi(t, x))}{\|\nabla \mathfrak{d}(\Phi(t, x))\|^2} \\ \Phi(0, x) = x, \end{cases}$$

where $W \subset \mathbb{R}^d$ is an open set that contains $\mathfrak{d}^{-1}([y - h_0, y + h_0])$. Differentiating the function $t \mapsto \mathfrak{d}(\Phi(t, x))$ immediately shows that $\mathfrak{d}(\Phi(t, x)) = \mathfrak{d}(x) + t$. This implies that $\Phi(t, \mathfrak{d}^{-1}(\{y\})) = \mathfrak{d}^{-1}(\{y + t\})$. In particular, Eldering (2013, Thm. A.6) yields that Φ is in $C^{1,1}$. Consequently, we find that

$$\begin{aligned} & \left| \int_{\Gamma_{y+v} \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) \right| \\ &= \left| \int_{\mathfrak{d}^{-1}(\{y\})} g(x) \mathbb{1}_{\{x \in \mathcal{X}\}} d\mathcal{H}^{d-1}(x) - \int_{\Phi(v, \mathfrak{d}^{-1}(\{y\}))} g(x) \mathbb{1}_{\{x \in \mathcal{X}\}} d\mathcal{H}^{d-1}(x) \right| \\ &\leq \int_{\mathfrak{d}^{-1}(\{y\})} \left| g(\Phi(0, x)) J_{\Phi(0, \cdot)}(x) \mathbb{1}_{\{\Phi(0, x) \in \mathcal{X}\}} - g(\Phi(v, x)) J_{\Phi(v, \cdot)}(x) \mathbb{1}_{\{\Phi(v, x) \in \mathcal{X}\}} \right| d\mathcal{H}^{d-1}(x), \end{aligned}$$

where $J_{\Phi(v, \cdot)}$ denotes the *Jacobian determinant* of $\Phi(v, \cdot)$. The last line follows by a change of variables (see e.g. Merigot and Thibert (2021, Thm. 56)) and the fact that $\Phi(0, \cdot)$ is the identity. By Kirszbraun's Theorem (Federer, 1969, Thm. 2.10.43) we can extend $g : \mathcal{X} \rightarrow [-\alpha, \alpha]$ to a Lipschitz continuous function $\tilde{g} : \mathbb{R}^d \rightarrow \mathbb{R}$, that has the same Lipschitz constant α . Obviously, it holds that

$$\begin{aligned} & \int_{\mathfrak{d}^{-1}(\{y\})} \left| g(\Phi(0, x)) J_{\Phi(0, \cdot)}(x) \mathbb{1}_{\{\Phi(0, x) \in \mathcal{X}\}} - g(\Phi(v, x)) J_{\Phi(v, \cdot)}(x) \mathbb{1}_{\{\Phi(v, x) \in \mathcal{X}\}} \right| d\mathcal{H}^{d-1}(x) \\ &= \int_{\mathfrak{d}^{-1}(\{y\})} \left| \tilde{g}(\Phi(0, x)) J_{\Phi(0, \cdot)}(x) \mathbb{1}_{\{\Phi(0, x) \in \mathcal{X}\}} - \tilde{g}(\Phi(v, x)) J_{\Phi(v, \cdot)}(x) \mathbb{1}_{\{\Phi(v, x) \in \mathcal{X}\}} \right| d\mathcal{H}^{d-1}(x). \end{aligned}$$

Therefore, we find that

$$\begin{aligned} & \left| \int_{\Gamma_{y+v} \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) \right| \\ &\leq \int_{\mathfrak{d}^{-1}(\{y\})} \left| \tilde{g}(\Phi(0, x)) J_{\Phi(0, \cdot)}(x) - \tilde{g}(\Phi(v, x)) J_{\Phi(v, \cdot)}(x) \right| \mathbb{1}_{\{\Phi(0, x) \in \mathcal{X}\}} d\mathcal{H}^{d-1}(x) \\ &\quad + \int_{\mathfrak{d}^{-1}(\{y\})} \left| \mathbb{1}_{\{\Phi(0, x) \in \mathcal{X}\}} - \mathbb{1}_{\{\Phi(v, x) \in \mathcal{X}\}} \right| \left| \tilde{g}(\Phi(v, x)) J_{\Phi(v, \cdot)}(x) \right| d\mathcal{H}^{d-1}(x). \end{aligned}$$

Since Φ is in $C^{1,1}([-h_0, h_0] \times W)$, it follows that $(v, x) \mapsto \tilde{g}(\Phi(v, x))$ and $(v, x) \mapsto J_{\Phi(v, \cdot)}(x)$ are Lipschitz continuous functions. We observe that for $(v, x) \in [-h_0, h_0] \times \mathcal{X}$

$$\begin{aligned} |\tilde{g}(\Phi(v, x))| &\leq |\tilde{g}(\Phi(0, x))| + |\tilde{g}(\Phi(v, x)) - \tilde{g}(\Phi(0, x))| \leq \alpha + \alpha \|\Phi(v, x) - \Phi(0, x)\| \\ &\leq \alpha + \alpha L_\Phi h_0, \end{aligned}$$

where L_Φ denotes the Lipschitz constant of Φ . This implies immediately that the function $(v, x) \mapsto \tilde{g}(\Phi(v, x)) J_{\Phi(v, \cdot)}(x)$ is Lipschitz continuous on $[-h_0, h_0] \times \mathcal{X}$ with a Lipschitz constant that only depends on \mathfrak{d}, y, α and \mathcal{X} . Further, we realize that

$$\left| \mathbb{1}_{\{\Phi(0, x) \in \mathcal{X}\}} - \mathbb{1}_{\{\Phi(v, x) \in \mathcal{X}\}} \right| > 0 \tag{43}$$

implies that either $x \in \mathcal{X}$ or $\Phi(v, x) \in \mathcal{X}$ (but not both). Given (43), our previous calculations show that

$$|\tilde{g}(\Phi(v, x))J_{\Phi(v, \cdot)}(x)| \leq C_y,$$

where C_y denotes a finite constant that depends only on \mathfrak{d}, y, α as well as \mathcal{X} . For the remainder of this proof, this constant may vary from line to line. We obtain that

$$\begin{aligned} & \left| \int_{\Gamma_{y+v} \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) \right| \\ & \leq \int_{\mathfrak{d}^{-1}(\{y\})} C_y |v| \mathbb{1}_{\{x \in \mathcal{X}\}} d\mathcal{H}^{d-1}(x) + C_y \int_{\mathfrak{d}^{-1}(\{y\})} |\mathbb{1}_{\{\Phi(0, x) \in \mathcal{X}\}} - \mathbb{1}_{\{\Phi(v, x) \in \mathcal{X}\}}| d\mathcal{H}^{d-1}(x). \end{aligned}$$

Since y is a $C^{2,1}$ -regular value of \mathfrak{d} with respect to \mathcal{X} , we find that (by potentially adjusting h_0) there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$

$$\left| \int_{\Gamma_{y+v} \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} g(x) d\mathcal{H}^{d-1}(x) \right| \leq C_y |v| \mathcal{H}^{d-1}(\Gamma_y) + C_y |v| \leq C_y |v|.$$

This gives the claim. ■

D.2 Proof of Theorem D.1

By assumption, we have that $\nabla \mathfrak{d} \neq 0$ on the level set Γ_y . Furthermore, we have assumed that the function \mathfrak{d} is $C^{2,1}$ on an open neighborhood of Γ_y . Thus, there exists $h_0 > 0$ such that $\|\nabla \mathfrak{d}\| > 0$ on

$$\mathfrak{d}^{-1}[y - h_0, y + h_0] = \{x \in \mathbb{R}^d : y - h_0 \leq \mathfrak{d}(x) \leq y + h_0\}. \quad (44)$$

Throughout the following let $0 < h < h_0$. As $\text{supp}(k) = [-1, 1]$, we get that

$$\left| \int_{\mathcal{X}} k \left(\frac{\mathfrak{d}(x) - y}{h} \right) g(x) d\lambda^d(x) \right| = \left| \int_{\{x \in \mathcal{X} : |\mathfrak{d}(x) - y| \leq h\}} k \left(\frac{\mathfrak{d}(x) - y}{h} \right) g(x) d\lambda^d(x) \right|.$$

Since $\|\nabla \mathfrak{d}(x)\| > 0$ for $x \in \mathfrak{d}^{-1}[y - h_0, y + h_0]$ and $|g(x)| \leq \alpha$ for all x , we obtain that

$$\sup_{x \in \mathbb{R}^d} \left| \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} \mathbb{1}_{\{x \in \mathcal{X} : |\mathfrak{d}(x) - y| \leq h\}} \right| < C_y, \quad (45)$$

where C_y denotes a constant that only depends on \mathfrak{d}, y, α and \mathcal{X} (in particular it can be chosen independently from h). In the following, C_y may vary from line to line. Clearly, (45) implies that the function

$$x \mapsto \frac{|g(x)|}{\|\nabla \mathfrak{d}(x)\|} \mathbb{1}_{\{x \in \mathcal{X} : |\mathfrak{d}(x) - y| \leq h\}}$$

is λ^d -integrable for any $0 \leq h \leq h_0$. Therefore, it follows by Lemma D.2 in combination with (45) that

$$\begin{aligned} \left| \int_{\mathcal{X}} k \left(\frac{\mathfrak{d}(x) - y}{h} \right) g(x) d\lambda^d(x) \right| &= \left| \int_{-h}^h \int_{\{x \in \mathcal{X} : \mathfrak{d}(x) - y = v\}} k \left(\frac{\mathfrak{d}(x) - y}{h} \right) \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) dv \right| \\ &= \left| \int_{-h}^h k \left(\frac{v}{h} \right) \int_{\{x \in \mathcal{X} : \mathfrak{d}(x) - y = v\}} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) dv \right|. \end{aligned}$$

We note that

$$\{x \in \mathcal{X} : \mathfrak{d}(x) - y = v\} = \mathfrak{d}^{-1}(y + v) \cap \mathcal{X} = \Gamma_{y+v} \cap \mathcal{X}.$$

This yields that

$$\begin{aligned} &\left| \int_{\mathcal{X}} k \left(\frac{\mathfrak{d}(x) - y}{h} \right) g(x) d\lambda^d(x) \right| \\ &\leq \left| \int_{-h}^h k \left(\frac{v}{h} \right) \int_{\Gamma_y \cap \mathcal{X}} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) dv \right| \\ &+ \left| \int_{-h}^h k \left(\frac{v}{h} \right) \left(\int_{\Gamma_{y+v} \cap \mathcal{X}} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) \right) dv \right| =: T_5 + T_6. \end{aligned}$$

Next, we consider both summands separately. First of all, we observe that the integral

$$\int_{\Gamma_y \cap \mathcal{X}} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x)$$

does not depend on v . Consequently, we obtain that

$$T_5 \stackrel{(i)}{\leq} C_y \left| \int_{-h}^h k \left(\frac{v}{h} \right) dv \right| \left| \int_{\Gamma_y \cap \mathcal{X}} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) \right| \stackrel{(ii)}{\leq} C_y \left| \int_{-h}^h k \left(\frac{v}{h} \right) dv \right|.$$

Here, (i) follows by (45) and (ii) follows since $\mathcal{H}^{d-1}(\Gamma_y \cap \mathcal{X}) \leq C < \infty$ for some constant C , as already argued in the proof of Lemma D.3. Setting $u = v/h$ and using $\int k(u) du = 0$ gives that

$$T_5 \leq C_y h \left| \int_{-1}^1 k(u) du \right| = 0.$$

Hence, it only remains to consider the second summand T_6 . Let $\mathcal{X}^* = \mathcal{X} \cap \mathfrak{d}^{-1}([y - h_0, y + h_0])$. Since $h \leq h_0$, we obtain that

$$T_6 \leq \int_{-h}^h \left| k \left(\frac{v}{h} \right) \right| \left| \int_{\Gamma_{y+v} \cap \mathcal{X}^*} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}^*} \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|} d\mathcal{H}^{d-1}(x) \right| dv$$

We realize that the function

$$g^* : \mathcal{X}^* \rightarrow \mathbb{R}, \quad x \mapsto \frac{g(x)}{\|\nabla \mathfrak{d}(x)\|}$$

is Lipschitz continuous, as $\|\nabla d(x)\| > 0$ for $x \in d^{-1}([y - h_0, y + h_0])$, the function $\|\nabla d(x)\|$ is in $C^{1,1}(d^{-1}(y - h_0, y + h_0))$ and g is Lipschitz continuous and bounded by assumption. As y is a $C^{2,1}$ -regular bounded value of d with respect to \mathcal{X} , it is straightforward to verify that it is also one with respect to \mathcal{X}^* . Thus, the requirements of Lemma D.3 are met. By potentially decreasing h_0 , we find for all h small enough that

$$T_6 \leq C_y \int_{-h}^h \left| k\left(\frac{v}{h}\right) \right| v dv.$$

Setting $u = v/h$ gives that

$$T_6 \leq C_y h^2 \int_{-1}^1 |k(u)| u du \leq c_y h^2,$$

where $c_y > 0$ depends only on d, y, α and \mathcal{X} .

All in all, this gives

$$\left| \int_{\mathcal{X}} k\left(\frac{d(x) - y}{h}\right) g(z) d\lambda^d(z) \right| \leq T_5 + T_6 \leq c_y h^2,$$

which yields the claim. ■

References

- Herbert Amann. *Ordinary Differential Equations: An Introduction to Nonlinear Analysis*, volume 13. Walter de Gruyter, 2011.
- Serge Belongie, Greg Mori, and Jitendra Malik. Matching with shape contexts. In *Statistics and Analysis of Shapes*, pages 81–105. Springer, 2006.
- José R. Berrendero, Antonio Cuevas, and Beatriz Pateiro-López. Shape classification based on interpoint distance distributions. *Journal of Multivariate Analysis*, 146:237–247, 2016.
- Biplab Bhattacharjee. nth-nearest-neighbor distribution functions of an interacting fluid from the pair correlation function: A hierarchical approach. *Physical Review E*, 67:041208, 2003.
- Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 1979.
- Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.
- Claire Bréchet. A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature. *Electronic Journal of Statistics*, 13(1):795–849, 2019.
- Daniel Brinkman and Peter J. Olver. Invariant histograms. *The American Mathematical Monthly*, 119(1):4–24, 2012.
- Mickaël Buchet, Frédéric Chazal, Tamal K. Dey, Fengtao Fan, Steve Y. Oudot, and Yusu Wang. Topological analysis of scalar fields with outliers, 2014. <https://arxiv.org/abs/1412.1680>.

- J. V. Castellana and M. R. Leadbetter. On smoothed probability density estimation for stationary processes. *Stochastic Processes and their Applications*, 21(2):179–193, 1986.
- Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011.
- Frédéric Chazal, Leonidas J Guibas, Steve Y. Oudot, and Primoz Skraba. Persistence-based clustering in Riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):1–38, 2013.
- Frédéric Chazal, Pascal Massart, Bertrand Michel, et al. Rates of convergence for robust geometric inference. *Electronic journal of statistics*, 10(2):2243–2286, 2016.
- Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance to a measure and kernel distance. *The Journal of Machine Learning Research*, 18(1):5845–5884, 2017.
- Su Chen and Michael Pokojovy. Modern and classical k-sample omnibus tests. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(1):e1418, 2018.
- Sándor Csörgő and Jan Mielniczuk. Density estimation under long-range dependence. *The Annals of Statistics*, 23(3):990–999, 1995.
- Antonio Cuevas. Set estimation: Another bridge between statistics and geometry. *Boletín de Estadística e Investigación Operativa*, 25(2):71–85, 2009.
- Jaap Eldering. *Normally hyperbolic invariant manifolds: The noncompact case*, volume 2. Springer, 2013.
- Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6):2301–2339, 2014.
- Herbert Federer. *Geometric Measure Theory*. Springer, 1969.
- Edward W. Frees. Estimating densities of functions of observations. *Journal of the American Statistical Association*, 89(426):517–525, 1994.
- Natasha Gelfand, Niloy J. Mitra, Leonidas J. Guibas, and Helmut Pottmann. Robust global registration. In *Symposium on geometry processing*, volume 2, page 5. Vienna, Austria, 2005.
- Manuela Gellert, Md Faruq Hossain, Felix Jacob Ferdinand Berens, Lukas Willy Bruhn, Claudia Urbainsky, Volkmar Liebscher, and Christopher Horst Lillig. Substrate specificity of thioredoxins and glutaredoxins—towards a functional classification. *Heliyon*, 5(12):e02943, 2019.
- Bryan S Graham, Fengshi Niu, and James L Powell. Kernel density estimation for undirected dyadic data. *Journal of Econometrics*, page 105336, 2022.

- Marc Hallin, Zudi Lu, and Lanh T. Tran. Kernel density estimation for spatial processes: The ll theory. *Journal of Multivariate Analysis*, 88(1):61–75, 2004.
- A. Ben Hamza and Hamid Krim. Geodesic object representation and recognition. In *International conference on discrete geometry for computer imagery*, pages 378–387. Springer, 2003.
- Xian Hao, Jyotsana J. Parmar, Benoît Lelandais, Andrey Aristov, Wei Ouyang, Christian Weber, and Christophe Zimmer. Super-resolution visualization and modeling of human chromosomal regions reveals cohesin-dependent loop structures. *Genome Biology*, 22, 2021.
- Yuichi S. Hayakawa and Takashi Oguchi. Applications of terrestrial laser scanning in geomorphology. *Journal of Geography (Chigaku Zasshi)*, 125(3):299–324, 2016.
- Florian Heinemann and Nicholas Bonneel. *WSGeometry: Compute Wasserstein barycenters, geodesics, PCA and distances*, 2021. URL <https://CRAN.R-project.org/package=WSGeometry>. R package version 1.0.
- Morris Hirsch and Stephen Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Springer, 1974.
- Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948. ISSN 0003-4851.
- Shih-Si Hsiao, Ko-Tang Chen, and A Yu Ite. Mean field theory of weakly-interacting rydberg polaritons in the EIT system based on the nearest-neighbor distribution. *Optics Express*, 28(19):28414–28429, 2020.
- M. Lelek, M.T. Gyparaki, and G. Beliu et al. Single-molecule localization microscopy. *Nature Reviews Methods Primers*, 1, 2021.
- Noam I. Libeskind et al. Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473(1):1195–1217, 08 2017.
- E. Lieberman-Aiden et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.
- Eckhard Liebscher. Strong convergence of sums of α -mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications*, 65(1):69–80, 1996.
- Zudi Lu. Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics*, 53(3):447–468, 2001.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>.
- Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.

- Facundo Mémoli and Tom Needham. Distance distributions and inverse problems for metric measure spaces. *Studies in Applied Mathematics*, 149(4):943–1001, 2022.
- Hui Meng, Yuan Gao, Xin Yang, Kun Wang, and Jie Tian. K-nearest neighbor based locally connected network for fast morphological reconstruction in fluorescence molecular tomography. *IEEE transactions on medical imaging*, 39(10):3019–3028, 2020.
- Quentin Merigot and Boris Thibert. Optimal transport: Discretization and algorithms. In *Handbook of Numerical Analysis*, volume 22, pages 133–212. Elsevier, 2021.
- Frank Morgan. *Geometric Measure Theory: A Beginner’s Guide*. Academic press, 2016.
- P. Nicovich, D. Owen, and K. Gaus. Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nature Protocols*, 12:453–460, 2017.
- Johannes Nuebler, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A. Mirny. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, 115(29):E6697–E6706, 2018.
- Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4):807–832, 2002.
- Peter M. Robinson. Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3):185–207, 1983.
- Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 55. Birkäuser, NY, 2015.
- Yonggang Shi, Paul M Thompson, Greig I. de Zubicaray, Stephen E. Rose, Zhuowen Tu, Ivo Dinov, and Arthur W. Toga. Direct mapping of hippocampal surfaces with intrinsic shape context. *NeuroImage*, 37(3):792–807, 2007.
- Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics*. SIAM, 2009.
- Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. Routledge, 2018.
- S. Torquato, B. Lu, and J. Rubinstein. Nearest-neighbor distribution functions in many-body systems. *Physical Review A*, 41(4):2059, 1990.
- Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.
- Aad W. Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge university press, 2000.
- Antonio Villanacci, Laura Carosi, Pierluigi Benevieri, and Andrea Battinelli. *Differential Topology and General Equilibrium with Complete and Incomplete Markets*. Springer Science & Business Media, 2002.

George Vosselman, Ben Gorte, George Sithole, and Tahir Rabbani. Recognising structure in laser scanner point clouds. *International archives of photogrammetry, remote sensing and spatial information sciences*, 46(8):33–38, 2004.

Christoph Alexander Weitkamp, Katharina Proksch, Carla Taveling, and Axel Munk. Distribution of distances based object matching: Asymptotic inference. *Journal of the American Statistical Association*, 119(545):538–551, 2024.

Wei Biao Wu and Jan Mielniczuk. Kernel density estimation for linear processes. *The Annals of Statistics*, 30(5):1441 – 1459, 2002.

Guangzhou Zou and Hsin-I Wu. Nearest-neighbor distribution of interacting biological entities. *Journal of Theoretical Biology*, 172(4):347–353, 1995.