

**Universität
Stuttgart**

**Fachbereich
Mathematik**

Fully Adaptive Density-Based Clustering

Ingo Steinwart

Preprint 2013/016

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: <http://www.mathematik.uni-stuttgart.de/preprints>

ISSN **1613-8309**

© Alle Rechte vorbehalten. Nachdruck nur mit Genehmigung des Autors.
L^AT_EX-Style: Winfried Geis, Thomas Merkle

Fully Adaptive Density-Based Clustering

Ingo Steinwart
Institute for Stochastics and Applications
Faculty 8: Mathematics and Physics
University of Stuttgart
D-70569 Stuttgart Germany
`ingo.steinwart@mathematik.uni-stuttgart.de`

December 20, 2013

Abstract

Based on the work of Hartigan, the clusters of a distribution are often defined to be the connected components of a density level set. Unfortunately, this definition drastically depends on the user-specified level, and in general finding a reasonable level is a difficult task. In addition, the definition is not rigorous for discontinuous densities, since the topological structure of a density level set may be dramatically changed by modifying the density on a set of measure zero. In this work, we address these issues by first modifying the notion of density level sets in a way that makes the level sets independent of the actual choice of the density. We then propose a simple algorithm for estimating the smallest level at which the modified level sets have more than one connected component. For this algorithm we provide a finite sample analysis, which is then used to show that the algorithm consistently estimates both the smallest level and the corresponding connected components. We further establish rates of convergence for the two estimation problems, and last but not least, we present a simple strategy for determining the width-parameter of the involved density estimator in a data-depending way. The resulting algorithm turns out to be adaptive, that is, it achieves the optimal rates achievable by our analysis without knowing characteristics of the underlying distribution.

1 Introduction

A central and widely studied task in statistical learning theory or machine learning is cluster analysis, where the goal is to find clusters in unlabeled data. Unlike in supervised learning tasks such as classification or regression, a key problem in cluster analysis is already the definition of a learning goal that describes a conceptionally and mathematically convincing definition of clusters. A widely, but by no means generally accepted, definition of clusters has its roots in a paper by Carmichael et al. (1968), who define clusters to be densely populated areas in the input space that are separated by less populated areas. The *non-parametric* mathematical translation of this idea, which goes back to Hartigan (1975), usually assumes that the data $D = (x_1, \dots, x_n) \in X^n$ is generated by some unknown probability measure P on a topological space X that has a density h with respect to some known reference measure μ on X . Given a threshold $\rho \geq 0$, the clusters are then defined to be the connected components of the density level set $\{h \geq \rho\} := \{x \in X : h(x) \geq \rho\}$. Here, one typically considers the case, where $X \subset \mathbb{R}^d$ and μ is the Lebesgue measure on X . In addition, it is often explicitly or implicitly assumed that the density h is continuous, since this avoids various pathologies regarding the topological notion of connectedness that are caused by changes of h on μ -zero sets, see Rigollet (2007) for some illustrations.

Some approaches were made in the past to address these topological pathologies. For Example, Cuevas and Fraiman (1997) introduced a thickness assumption for sets C , that rules out cases, in which neighborhoods of $x \in C$ have not sufficient mass. This thickness assumption excludes some topological pathologies such as topologically connected bridges of zero mass, see e.g. Rigollet (2007), while others such as cuts of measure zero are not addressed. Moreover, Rinaldo and Wasserman

(2010) entirely avoid these issues by considering level sets of convolutions $k * P$ of the underlying distribution P with a continuous kernel k on \mathbb{R}^d having a compact support. Since such convolutions are always continuous, these authors can not only deal with discontinuous densities but also with distributions that do not have a Lebesgue density at all. However, different kernels (or kernel widths) may lead to different clusters, and consequently, their approach introduces another parameter that is in general hard to control by the user.

Defining clusters by the connected components of a level set clearly requires us to estimate the level set in one form or the other. Level set estimation itself is a classical non-parametric problem, which has been considered by various authors such as Devroye and Wise (1980), Hartigan (1987), Müller and Sawitzki (1991), Polonik (1995), Ben-David and Lindenbaum (1997), Tsybakov (1997), Bañilo et al. (2001, 2000), Steinwart et al. (2005), Rigollet and Vert (2009), Singh et al. (2009). In these articles, two different performance measures are considered for assessing the quality of a density level estimate, namely the mass of the symmetric difference between the estimate and the true level set, and the Hausdorff distance between these two sets. Estimators that are consistent with respect to the Hausdorff metric clearly capture all topological structures eventually, so that these estimators form an almost canonical choice for density-based clustering. In contrast, level set estimators that are only consistent with respect to the first performance measure are, in general, not suitable for the cluster problem, since even sets that are equal up to measure zero may have completely different topological properties.

Historically, two distinct questions have been investigated for density-based clustering. The first one is the so-called single level approach, which tries to estimate the connected components of $\{h \geq \rho\}$ for a *single and fixed* level $\rho \geq 0$. The single level approach has been studied by several authors, see, e.g., Hartigan (1975), Cuevas and Fraiman (1997), Rigollet (2007), Maier et al. (2009), Rinaldo and Wasserman (2010) and the references therein. Moreover, we have mentioned above that level set estimators that are consistent with respect to the Hausdorff metric can be easily used for this version of density-based clustering, and thus it seems fair to say that this clustering problem enjoys a good statistical understanding. Unfortunately, however, it suffers from a serious conceptual issue, namely that of determining a good value of ρ . Indeed, it is not hard to see that different values of ρ may lead to different numbers of clusters, see e.g. the illustrations by Chaudhuri and Dasgupta (2010), Rinaldo et al. (2012). On the other hand, it is almost impossible for the user to guess a suitable value for ρ , and using a couple of different candidate values creates the problem of deciding which of the resulting clusterings is best. For this reason, Rinaldo and Wasserman (2010) note that research on data-dependent, automatic methods for choosing ρ (and the width parameter of the involved density estimator) “would be very useful”.

The second approach, which is known as the cluster tree approach, tries to address the issue of finding a good value of ρ by considering all levels simultaneously. In this approach, the focus lies on the identification of the hierarchical structure of the connected components for different levels. To be more precise, let h be a fixed density, which, for the sake of simplicity, is assumed to be continuous, and A be a connected component of $\{h \geq \rho\}$. Then, for every $\rho' \in [0, \rho]$, there exists exactly one connected component B of $\{h \geq \rho'\}$ with $A \subset B$, see e.g. Lemma 2.9. Under some additional assumptions on μ and h , this leads to a *finite* tree, in which each node B is a connected component of some level set $\{h \geq \rho'\}$ and all children of a node B are the connected components of $\{h \geq \rho\}$ for some $\rho > \rho'$ that are contained in B . Results, further definitions, and methods for estimating the structure of this tree can be found in the work by Hartigan (1975), Stuetzle (2003), Chaudhuri and Dasgupta (2010), Stuetzle and Nugent (2010). In particular, Chaudhuri and Dasgupta (2010) show that in a weak sense of Hartigan (1981), a modified single linkage algorithm converges to this tree under some assumptions on the density h . To be more precise, let A and A' be two different connected components of some level set of h , and $D \in X^n$ be a data set from which the tree estimate is constructed. Furthermore, let A_D and A'_D be the smallest clusters in this tree estimate that satisfy $A \cap D \subset A_D$ and $A' \cap D \subset A'_D$, respectively. Then the result by Chaudhuri and Dasgupta (2010) shows that we have $A_D \cap A'_D = \emptyset$ with probability P^n converging to 1 for $n \rightarrow \infty$. Roughly speaking, this means that all parent/child relations of the cluster tree are eventually contained in the tree estimate, and Chaudhuri and Dasgupta (2010), actually show the latter by finite sample guarantees. More recently, Kpotufe and von Luxburg (2011) extended this analysis to a wider range of parameters for the underlying k -NN density estimator. In addition, they

also proposed a simple pruning strategy, that removes connected components that only artificially occur because of finite sample variability. Unfortunately, however, neither of these results tell us *a*) how to find these smallest sets A_D and A'_D in the estimating tree without knowing h , and *b*) how well A_D and A'_D approximate A and A' , respectively. Consequently, it seems fair to say that this notion of consistency reveals more about the cluster structure and less about the actual clusters.

Unlike the papers mentioned above, we neither consider the single level approach nor the cluster tree approach. Instead, we are interested in estimating the infimum ρ^* of all levels at which the density level set consists of more than one connected component. In addition, we wish to estimate the corresponding clusters.

To make these goals mathematically rigorous for discontinuous densities, we first introduce a notion of density level sets that is independent of the actual choice of the density, see (2). Since the above mentioned topological pathologies are all caused by the ambivalence of densities on zero sets, this new notion is immune against such pathologies. In general, our new notion of level sets leads to sets M_ρ that are larger than the classical level sets $\{h \geq \rho\}$, though it will turn out that we always have $M_\rho \Delta \{h \geq \rho\} \subset \partial\{h \geq \rho\}$, see (3) for details. Consequently, both sets are equal up to measure zero, whenever we have $\mu(\partial\{h \geq \rho\}) = 0$ for some density h , and Lemma 2.4 presents another sufficient condition.

With the help of the level sets M_ρ we can then consider the infimum ρ^* over all levels ρ for which M_ρ contains more than connected component. For simplicity, we assume in this paper that there exists some $\rho^{**} > \rho^*$ such that, for all $\rho \in (\rho^*, \rho^{**}]$, the level sets M_ρ contain exactly *two* connected components. Note that the persistence of the cluster structure over a small range of levels $\rho \in (\rho^*, \rho^{**}]$ is assumed either explicitly or implicitly in basically all density-based clustering approaches that deal with several levels ρ , see e.g. Chaudhuri and Dasgupta (2010), Kpotufe and von Luxburg (2011). Intuitively, this restriction is somewhat natural, since without such a persistence it seems impossible to identify the topological structure of a level set with the help of an estimate that is vertically uncertain due to finite sample effects. On the other hand, the restriction to *two* components seems to be quite restrictive at first glance. Surprisingly, however, the opposite is true. To illustrate this, assume for simplicity that $X = [0, 1]$ and $h : X \rightarrow (0, \infty)$ is a continuous density with exactly two distinct strict local minima at say x_1 and x_2 . Now, if, e.g., $h(x_1) < h(x_2)$, then $\rho^* = h(x_1)$ and, for ρ^{**} , we can choose any value with $h(x_1) < \rho^{**} \leq h(x_2)$, since for $\rho \in (\rho^*, h(x_2)]$, the density level set actually contains exactly two connected components. Consequently, our assumption of having two connected components for a small range above ρ^* would only be violated if $h(x_1) = h(x_2)$. Compared to the case $h(x_1) \neq h(x_2)$, the latter seems to be rather singular, in particular, if one considers higher-dimensional analogs.

Besides the assumptions discussed so far, we need to make an additional assumption on the level sets that excludes bridges and cusps that are too thin and long. While this is certainly unpleasant, it seems to be rather necessary, since such an assumption occurs in one form or the other in most articles dealing with density-based clustering and Hausdorff estimation of level sets. Moreover, for Hölder continuous densities it is easy to show that such an assumption *essentially* holds for all levels and this fact is implicitly used, e.g. by Kpotufe and von Luxburg (2011).

In this work we present an algorithm that consistently estimates the level ρ^* and the corresponding clusters with the help of a histogram-based level set estimator. Using finite sample guarantees, we further establish rates of convergence for estimating ρ^* under a rather natural assumption on P that describes how fast the connected components of M_ρ move apart for increasing $\rho \in (\rho^*, \rho^{**}]$. Note that this assumption does not incorporate any sort of continuity, or even smoothness, of h . In particular, it is easy to construct examples of discontinuous densities satisfying this assumption, while we show that it is automatically satisfied for Hölder-continuous densities. We further establish rates of convergence for the problem of estimating the corresponding clusters. Here we additionally need to consider the well-known flatness condition by Polonik (1995) and an assumption that describes the mass of δ -tubes around the boundary of the level sets. Unlike in previous articles however, we do not need to restrict our considerations to (essentially) rectifiable boundaries. Unfortunately, however, all these rates can only be achieved, if the histogram width is chosen in a suitable, distribution dependent way, and therefore we finally propose a simple data-driven parameter selection strategy. For this strategy we show that, in many cases, it achieves the above mentioned rates without knowing characteristics of the underlying distribution.

Since this work strongly builds upon the papers by Steinwart (2011), Sriperumbudur and Steinwart (2012), let us briefly describe our main *new* contributions. Firstly, Steinwart (2011) only establishes consistency of the algorithm considered in this work, that is, no rate of convergence is presented. While Sriperumbudur and Steinwart (2012) do establish such rates, the situation considered by Sriperumbudur and Steinwart (2012) is different. Indeed, Sriperumbudur and Steinwart (2012) considers a different algorithm that uses a Parzen window density estimator to estimate the level sets. However, this algorithm requires the density to be α -Hölder continuous, and in fact, it requires the user to know α . Secondly, neither of the two papers consider a data-dependent way of choosing the width parameter of the involved density estimator. Besides these new contributions, this paper also adds a substantial amount of extra information regarding the imposed assumptions and last but not least polishes many of the results from Steinwart (2011).

The rest of this paper is organized as follows. In Section 2 we introduce our topologically robust notion of density level sets and establish some simple properties of these sets. We further consider maps that relate connected components of different level sets. These maps will be our fundamental tool for comparing the cluster structure of the true density level sets and their empirical estimates. We further make the above notion of clusters rigorous and establish some results about the persistence of the cluster structure under horizontal and vertical uncertainty. Section 3 contains three parts. In the first part we determine the vertical and horizontal uncertainty when estimating density level sets with the help of a standard plug-in histogram approach. We then propose a simple and generic algorithm that receives a family of level set estimates with known uncertainty and that returns both an estimate of the smallest level ρ^* and the resulting clusters. Finally, we present a finite sample analysis for this generic algorithm. In Section 4 we then apply this finite sample analysis to the case in which the algorithm receives the level set estimates of a histogram approach. Here we show the consistency of our algorithm and present learning rates. Section 5 contains the description and the analysis of the data-driven width selection strategy. All proofs as well as many auxiliary results can be found in Section 6. Finally, the appendix contains, as supplemental material, an example of a large class of distributions on \mathbb{R}^2 with continuous densities, that satisfy all the assumptions made in this paper.

2 Preliminaries: Level Sets, Connectivity, and Clusters

In this section we introduce all notions related to the definition and analysis of clusters. We further present various technical result needed throughout the paper.

2.1 Density-Independent Density Level Sets

Unlike to the rest of the paper, where we mostly consider compact metric spaces, we assume throughout this subsection that (X, d) denotes a complete separable metric space. Recall that compact metric spaces are both complete and separable, and hence everything developed in this subsection can actually be used in the remainder of the paper, too. Now, let $\mathcal{B}(X)$ be the Borel σ -algebra on X , μ be a *known* σ -finite measure on $\mathcal{B}(X)$, and P be an *unknown* μ -absolutely continuous probability measure on $\mathcal{B}(X)$. Recall that by Radon-Nykodym's theorem, P has a μ -density $h : X \rightarrow [0, \infty)$, but this density is only μ -almost surely determined and therefore, for $\rho \in [0, \infty)$, the density level set $\{h \geq \rho\}$ is also only μ -almost surely determined. In particular, if we consider a measurable set $A \subset X$ with

$$\mu(A \Delta \{h \geq \rho\}) = 0,$$

then there exists another μ -density $h' : X \rightarrow [0, \infty)$ of P such that $A = \{h' \geq \rho\}$. Now observe that the topological properties such as closedness or connectivity of $\{h' \geq \rho\}$ may be quite different from those of $\{h \geq \rho\}$, since in general these properties may be changed by μ -zero sets. Unfortunately, however, these topological properties play a crucial role in the definition of clusters, and hence we need a notion of “density level sets” that is independent of the particular choice of the density. To achieve this, recall that the support $\text{supp } \nu$ of a measure ν on $(X, \mathcal{B}(X))$ is the complement of the largest open ν -zero set, that is, $\text{supp } \nu$ is the smallest closed subset B of X that satisfies $\nu(X \setminus B) = 0$. Moreover, recall that, for every measure on a complete, separable metric space, the support actually

exists. Now observe that, for every fixed $\rho \in \mathbb{R}$,

$$\mu_\rho(A) := \mu(A \cap \{h \geq \rho\}), \quad A \in \mathcal{B}(X),$$

defines a σ -finite measure μ_ρ on $(X, \mathcal{B}(X))$ that is *independent* of the particular choice of the μ -density h of P . Consequently, the set

$$M_\rho := \text{supp } \mu_\rho,$$

is independent of this choice, too. In the following, we call M_ρ *the density level set to the level ρ* . For any given μ -density h of P , these definitions yield

$$\mu(\{h \geq \rho\} \setminus M_\rho) = \mu(\{h \geq \rho\} \cap (X \setminus M_\rho)) = \mu_\rho(X \setminus M_\rho) = 0, \quad (1)$$

that is, up to μ -zero sets, no density level set $\{h \geq \rho\}$ is larger than M_ρ . Moreover, M_ρ is actually the smallest closed set satisfying this equation. In addition, it is easy to check that we have

$$M_\rho = \{x \in X : \mu_\rho(U) > 0 \text{ for all open neighborhoods } U \text{ of } x\}. \quad (2)$$

Note that if $\text{supp } \mu = X$, we actually have $M_\rho = X$ for all $\rho \leq 0$, but typically we are, of course, interested in the case $\rho > 0$, only. To state our first technical result, which provides a lower and an upper bound for the set M_ρ , we write $\overset{\circ}{A}$ for the interior and \overline{A} for the closure of a set $A \subset X$. Moreover, $\partial A := \overline{A} \setminus \overset{\circ}{A}$ denotes the boundary of a $A \subset X$.

Lemma 2.1. *Let (X, d) be a complete separable metric space, μ be a σ -finite measure on X with $\text{supp } \mu = X$, and P be a μ -absolutely continuous probability measure on X . Then, for all μ -densities h of P and all $\rho \in \mathbb{R}$, we have*

$$\{h \geq \rho\}^\circ \subset M_\rho \subset \overline{\{h \geq \rho\}}.$$

Moreover, if h is continuous, we have $\{h > \rho\} \subset M_\rho \subset \{h \geq \rho\}$ and $\partial M_\rho \subset \{h = \rho\}$.

An immediate consequence of Lemma 2.1 is that the symmetric difference between the sets M_ρ and $\{h \geq \rho\}$ is contained in the boundary of $\{h \geq \rho\}$, that is

$$M_\rho \Delta \{h \geq \rho\} \subset \partial\{h \geq \rho\}. \quad (3)$$

The next lemma shows that the sets M_ρ are ordered the way one would expect density level sets to be ordered.

Lemma 2.2. *Let (X, d) be a complete separable metric space, μ be a σ -finite measure on X , and P be a μ -absolutely continuous probability measure on X . Then, for all $\rho_1 \leq \rho_2$, we have*

$$M_{\rho_2} \subset M_{\rho_1}.$$

It turns out that we will not only need the equality $\mu(\{h \geq \rho\} \setminus M_\rho) = 0$ established in (1), but also the ‘‘converse’’ equality $\mu(M_\rho \setminus \{h \geq \rho\}) = 0$. This is ensured by the following definition.

Definition 2.3. *Let (X, d) be a complete separable metric space, μ be a σ -finite measure on X , and P be a μ -absolutely continuous probability measure on X . For $\rho \in \mathbb{R}$, we say that P is*

i) *upper normal at level $\rho \in \mathbb{R}$, if, for some μ -density (and thus all μ -densities) h of P , we have*

$$\mu(M_\rho \setminus \{h \geq \rho\}) = 0.$$

ii) *lower normal at level $\rho \in \mathbb{R}$, if, for some μ -density (and thus all μ -densities) h of P , we have*

$$\mu(\{h > \rho\} \setminus \overset{\circ}{M}_\rho) = 0.$$

Moreover, we say that P is normal at level ρ if it is both upper and lower normal at level ρ . Finally, P is normal, if it is normal at every level.

The following lemma provides some simple sufficient conditions for normality.

Lemma 2.4. *Let (X, d) be a complete separable metric space, μ be a σ -finite measure on X with $\text{supp } \mu = X$, and P be a μ -absolutely continuous probability measure on X . Then the following statements hold:*

- i) If P has a upper semi-continuous μ -density, then it is upper normal at every level.*
- ii) If P has a lower semi-continuous μ -density, then it is lower normal at every level.*
- iii) If P has a μ -density h such that $\mu(\partial\{h \geq \rho\}) = 0$ for some $\rho \geq 0$, then P is normal at level ρ .*

Let us now assume that P is upper normal at some level ρ . By (1) we then immediately see that

$$\mu(M_\rho \Delta \{h \geq \rho\}) = 0 \quad (4)$$

for all μ -densities h of P . In other words, up to μ -zero measures, M_ρ equals the ρ -level set of all μ -densities h of P . Moreover, if for some $\rho^* > 0$ and $\rho^{**} > \rho^*$, the distribution P is upper normal at every level $\rho \in (\rho^*, \rho^{**}]$, then using the monotonicity of the sets M_ρ and $\{h \geq \rho\}$ in ρ as well as $(\cup_{i \in I} A_i) \Delta (\cup_{i \in I} B_i) \subset \cup_{i \in I} (A_i \Delta B_i)$, we find

$$\mu\left(\{h > \rho^*\} \Delta \bigcup_{\rho > \rho^*} M_\rho\right) \leq \mu\left(\bigcup_{n \in \mathbb{N}} (\{h \geq \rho^* + 1/n\} \Delta M_{\rho^* + 1/n})\right) = 0 \quad (5)$$

for all μ -densities h of P , and if P has a continuous density h , we even have $\bigcup_{\rho > \rho^*} M_\rho = \{h > \rho^*\}$ by an easy consequence of Lemma 2.1. Similarly, if P is lower normal at every level $\rho \in (\rho^*, \rho^{**}]$, we find

$$\mu\left(\{h > \rho^*\} \setminus \bigcup_{\rho > \rho^*} \overset{\circ}{M}_\rho\right) \leq \mu\left(\bigcup_{n \in \mathbb{N}} (\{h > \rho^* + 1/n\} \setminus \overset{\circ}{M}_{\rho^* + 1/n})\right) = 0, \quad (6)$$

and if in addition, (5) holds, we obtain $\mu(\bigcup_{\rho > \rho^*} M_\rho \Delta \bigcup_{\rho > \rho^*} \overset{\circ}{M}_\rho) = 0$.

2.2 Some Notions of Connectivity

We have already mentioned in the introduction that we will follow the idea of defining clusters by connected components. In this subsection, we introduce the necessary topological tools for this approach. Furthermore, we consider another, more quantitative notion of connectivity that is used in our algorithm.

Since in general we cannot expect to estimate the levels set exactly, we need a tool to compare the clusters, i.e. the connected components, of our estimate to the true clusters. Now, connected components form a partition of the (estimated) level set, and thus it seems natural to compare these partitions. It turns out that this idea is so fruitful that it will also help us in other situations. We thus begin by introducing a general approach for comparing partitions.

Definition 2.5. *Let $A \subset B$ be two arbitrary non-empty sets and $\mathcal{P}(A)$ and $\mathcal{P}(B)$ be two partitions of A and B , respectively. We say that $\mathcal{P}(A)$ is comparable to $\mathcal{P}(B)$, if for all $A' \in \mathcal{P}(A)$ there exists a $B' \in \mathcal{P}(B)$ such that $A' \subset B'$.*

Informally speaking, $\mathcal{P}(A)$ is comparable to $\mathcal{P}(B)$ if no cell $A' \in \mathcal{P}(A)$ is broken into pieces in $\mathcal{P}(B)$. Moreover, since $\mathcal{P}(B)$ is a partition, it is clear that we cannot have two distinct cells $B', B'' \in \mathcal{P}(B)$ such that $A' \subset B'$ and $A' \subset B''$. This simple observation immediately leads to the following crucial lemma.

Lemma 2.6. *Let $A \subset B$ be two non-empty sets with partitions $\mathcal{P}(A)$ and $\mathcal{P}(B)$, respectively. Then the following statements are equivalent:*

- i) $\mathcal{P}(A)$ is comparable to $\mathcal{P}(B)$.*
- ii) There exists a map $\zeta : \mathcal{P}(A) \rightarrow \mathcal{P}(B)$ such that for all $A' \in \mathcal{P}(A)$ we have*

$$A' \subset \zeta(A'). \quad (7)$$

Moreover, if one the statements above are true, ζ is uniquely determined by (7). In the following we call ζ the cell relating map (CRM) between A and B , and when we want to emphasize the involved pair (A, B) and the partitions are known from the context we write $\zeta_{A,B} := \zeta$.

Note that a CRM $\zeta : \mathcal{P}(A) \rightarrow \mathcal{P}(B)$ is injective if and only if no two distinct cells of $\mathcal{P}(A)$ are contained in the same cell of $\mathcal{P}(B)$. Conversely, ζ is surjective, if and only if every cell in $\mathcal{P}(B)$ contains a cell of $\mathcal{P}(A)$. As a consequence, ζ is bijective, if and only if there is a one-to-one relation between the cells of the two partitions. The latter situation is usually the one we seek, which justifies the following definition.

Definition 2.7. *Let $A \subset B$ be two non-empty sets with partitions $\mathcal{P}(A)$ and $\mathcal{P}(B)$, respectively. Then we say that $\mathcal{P}(A)$ is persistent in $\mathcal{P}(B)$, if $\mathcal{P}(A)$ is comparable to $\mathcal{P}(B)$ and the corresponding CRM $\zeta : \mathcal{P}(A) \rightarrow \mathcal{P}(B)$ is bijective.*

As we will see later, the bijectivity of a CRM is often proved with the help of intermediate CRMs. The key to this approach is provided by the following lemma that establishes transitivity for comparable partitions and a composition formula for the involved CRMs.

Lemma 2.8. *Let $A \subset B \subset C$ be three non-empty sets with partitions $\mathcal{P}(A)$, $\mathcal{P}(B)$, and $\mathcal{P}(C)$. Assume that $\mathcal{P}(A)$ is comparable to $\mathcal{P}(B)$ and that $\mathcal{P}(B)$ is comparable to $\mathcal{P}(C)$. Then $\mathcal{P}(A)$ is comparable to $\mathcal{P}(C)$ and the corresponding CRMs satisfy*

$$\zeta_{A,C} = \zeta_{B,C} \circ \zeta_{A,B}.$$

The lemma above shows, for example, that if $\mathcal{P}(A)$ is persistent in $\mathcal{P}(B)$ and $\mathcal{P}(B)$ is persistent in $\mathcal{P}(C)$, then $\mathcal{P}(A)$ is also persistent in $\mathcal{P}(C)$. Conversely, if $\mathcal{P}(A)$ is persistent in $\mathcal{P}(C)$, then $\zeta_{A,B}$ must be injective and $\zeta_{B,C}$ must be surjective. Such arguments will be frequently used in our proofs.

Let us now relate the rather abstract notion of CRMs to the topological notion of connectivity that will be used in the definition of clusters. To this end, we fix a metric space (X, d) . Now recall from topology that an $A \subset X$ is connected, if, for every pair $A', A'' \subset A$ of relatively closed disjoint subsets of A with $A' \cup A'' = A$, we have $A' = \emptyset$ or $A'' = \emptyset$. Moreover, the maximal connected subsets of A are called the connected components of the space, see e.g. (Kelley, 1955, p. 54f). It is well-known that these components form a partition of A and that every component is relatively closed in A . In particular, if A is closed, all connected components of A are closed. In the following, we denote the set of (topologically) connected components of A by $\mathcal{C}(A)$. Clearly, $\mathcal{C}(A)$ is a partition of A and the next lemma shows that for different sets these partitions are comparable.

Lemma 2.9. *Let (X, d) be a metric space and $A \subset B$ be two closed non-empty subsets of X with $|\mathcal{C}(B)| < \infty$. Then $\mathcal{C}(A)$ is comparable to $\mathcal{C}(B)$.*

For path-connected components it is straightforward to see that a statement analogous to Lemma 2.9 holds. In the following we will consider a discrete version of path-connectivity introduced in the following definition.

Definition 2.10. *Let (X, d) be a metric space, $A \subset X$ be a non-empty subset, and $\tau > 0$. We say that $x, x' \in A$ are τ -connected in A , if there exist $x_1, \dots, x_n \in A$ such that $x_1 = x$, $x_n = x'$ and $d(x_i, x_{i+1}) < \tau$ for all $i = 1, \dots, n-1$. Moreover, we say that A is τ -connected, if all $x, x' \in A$ are τ -connected in A .*

It is easy to check that the property of being τ -connected in A gives an equivalence relation for elements in A . We call the resulting equivalence classes the τ -connected components of A and denote the set of all τ -connected components of A by $\mathcal{C}_\tau(A)$. Obviously, the τ -connected components of $A \subset X$ are τ -connected. To formulate some more properties of τ -connected components, we write

$$d(x, A) := \inf_{x' \in A} d(x, x')$$

for the distance between some $x \in X$ and $A \subset X$, and $d(A, B) := \inf\{d(x, y) : x \in A, y \in B\}$ for the distance between A and another set $B \subset X$.

With these preparations we can now collect some useful facts about $\mathcal{C}_\tau(A)$ in the following lemma.

Lemma 2.11. *Let (X, d) be a metric space, $A \subset X$ be a non-empty subset and $\tau > 0$. Then we have $d(A', A'') \geq \tau$ for all $A', A'' \in \mathcal{C}_\tau(A)$ with $A' \neq A''$. Moreover, if A is closed, all $A' \in \mathcal{C}_\tau(A)$ are closed, and if X is compact we have $|\mathcal{C}_\tau(A)| < \infty$.*

The next lemma shows that a statement analogous to Lemma 2.9 also holds for τ -connectivity.

Lemma 2.12. *Let (X, d) be a metric space, $A \subset B$ be two non-empty subsets of X and $\tau > 0$. Then $\mathcal{C}_\tau(A)$ is comparable to $\mathcal{C}_\tau(B)$.*

It can be easily shown, see Lemma 6.4, that, for compact metric spaces (X, d) , a closed $A \subset X$ is topologically connected, if and only if it is τ -connected for all $\tau > 0$. The following lemma investigates the relation between $\mathcal{C}_\tau(A)$ and $\mathcal{C}(A)$ in more detail.

Lemma 2.13. *Let (X, d) be a compact metric space and $A \subset X$ be a non-empty closed subset. Then the following statements hold:*

- i) *For all $\tau > 0$, $\mathcal{C}(A)$ is comparable to $\mathcal{C}_\tau(A)$ and the CRM $\zeta : \mathcal{C}(A) \rightarrow \mathcal{C}_\tau(A)$ is surjective.*
- ii) *If $|\mathcal{C}(A)| < \infty$, we have*

$$\tau_A^* := \min\{d(A', A'') : A', A'' \in \mathcal{C}(A) \text{ with } A' \neq A''\} > 0, \quad (8)$$

where $\min \emptyset := \infty$. Moreover, for all $\tau \in (0, \tau_A^] \cap (0, \infty)$, we have $\mathcal{C}(A) = \mathcal{C}_\tau(A)$ and, for such τ , the CRM $\zeta : \mathcal{C}(A) \rightarrow \mathcal{C}_\tau(A)$ is bijective. Finally, if $\tau_A^* < \infty$, that is, $|\mathcal{C}(A)| > 1$, we have*

$$\tau_A^* = \max\{\tau > 0 : \mathcal{C}(A) = \mathcal{C}_\tau(A)\}.$$

Note that, in general, a closed subset of A may have infinitely many topologically connected components as, e.g., the Cantor set shows. In this case, the second assertion of the lemma above is, in general, no longer true.

Given two closed subsets $A \subset B$ and a $\tau > 0$, we can consider both the CRM $\zeta : \mathcal{C}(A) \rightarrow \mathcal{C}(B)$ and the CRM $\zeta_\tau : \mathcal{C}_\tau(A) \rightarrow \mathcal{C}_\tau(B)$. To distinguish between them, we sometimes call ζ the top-CRM and the ζ_τ the τ -CRM. The following lemma, which is a direct consequence of Lemma 2.13, and whose proof is therefore omitted, shows that for sufficiently small τ both coincide.

Lemma 2.14. *Let (X, d) be a compact metric space, $A \subset B$ be two non-empty closed subsets of X with $|\mathcal{C}(A)| < \infty$ and $|\mathcal{C}(B)| < \infty$. Let $\zeta : \mathcal{C}(A) \rightarrow \mathcal{C}(B)$ be the top-CRM. Then, for $\tau^* := \min\{\tau_A^*, \tau_B^*\}$ and all $\tau \in (0, \tau^*]$, we have $\zeta = \zeta_\tau$, where $\zeta_\tau : \mathcal{C}_\tau(A) \rightarrow \mathcal{C}_\tau(B)$ is the τ -CRM.*

We will see later that clusters will be defined with the help of connected components, while our clustering algorithm needs to work with τ -connected components to recognize small bridges. The lemma above shows that for sufficiently small τ , both concepts coincide. This already suggests that the quantity τ_A^* will be key to our analysis.

The following lemma establishes monotonicity of τ_A^* in A under a natural regularity assumption.

Lemma 2.15. *Let (X, d) be a compact metric space and $A \subset B$ be two non-empty closed subsets of X with $|\mathcal{C}(A)| < \infty$ and $|\mathcal{C}(B)| < \infty$. If the top-CRM $\zeta : \mathcal{C}(A) \rightarrow \mathcal{C}(B)$ is injective, then we have $\tau_A^* \geq \tau_B^*$.*

2.3 Clusters

Using the concepts developed in the previous subsections we can now introduce our notion of clusters in this subsection.

We begin with the following definition that describes distributions that have clusters.

Definition 2.16. *Let (X, d) be a compact metric space, μ be a finite Borel measure on X , and P be a μ -absolutely continuous and normal Borel probability measure on X . We say that P can be topologically clustered between the critical levels $\rho^* \geq 0$ and $\rho^{**} > \rho^*$, if, for all $\rho \in [0, \rho^{**}]$, the following conditions are satisfied:*

i) We have either $|\mathcal{C}(M_\rho)| = 1$ or $|\mathcal{C}(M_\rho)| = 2$.

ii) If we have $|\mathcal{C}(M_\rho)| = 1$, then $\rho \leq \rho^*$.

iii) If we have $|\mathcal{C}(M_\rho)| = 2$, then $\rho \geq \rho^*$ and the top-CRM $\zeta : \mathcal{C}(M_{\rho^{**}}) \rightarrow \mathcal{C}(M_\rho)$ is bijective.

Definition 2.16 ensures that up to the level ρ^* we only have one connected component, and thus $\mathcal{C}(M'_\rho)$ is persistent in $\mathcal{C}(M_\rho)$ for all $0 \leq \rho \leq \rho' < \rho^*$. Similarly, Condition iii) guarantees that for some, possibly rather small, vertical range $(\rho^*, \rho^{**}]$, $\mathcal{C}(M'_\rho)$ is persistent in $\mathcal{C}(M_\rho)$. We will see later that our algorithm, probably like any other algorithm, needs to deal with some vertical uncertainty caused by finite sample effects. The persistence described above will be crucial for our algorithm to work well under the presence of this uncertainty.

Note that the definition above does not exclude the case $|\mathcal{C}(M_{\rho^*})| = 1$, and hence the elements of $\mathcal{C}(M_{\rho^*})$ cannot be used to define the clusters of P . On the other hand, for $\rho > \rho^*$, each $A \in \mathcal{C}(M_\rho)$ should be a subset of a cluster of P . This idea is used in the following definition, which defines the clusters of P by a limit for $\rho \searrow \rho^*$.

Definition 2.17. Let (X, d) be a compact metric space, μ be a finite Borel measure on X , and P be a μ -absolutely continuous Borel probability measure on X that can be topologically clustered between the critical levels ρ^* and ρ^{**} . For $\rho \in (\rho^*, \rho^{**}]$, we write $\zeta_\rho : \mathcal{C}(M_{\rho^{**}}) \rightarrow \mathcal{C}(M_\rho)$ for the top-CRM. Moreover, let A_1 and A_2 be the topologically connected components of $M_{\rho^{**}}$. Then the sets

$$A_i^* := \bigcup_{\rho \in (\rho^*, \rho^{**}]} \zeta_\rho(A_i), \quad i \in \{1, 2\},$$

are called the topological clusters of P .

By the bijectivity of the maps ζ_ρ , it is straightforward to show that $A_1^* \cap A_2^* = \emptyset$. In general, however, the clusters may touch each other, that is, we may have $d(A_1^*, A_2^*) = 0$. For example, if P is a mixture of two Gaussians with different centers but same variance, then it is easy to check that the two clusters are only separated by a hyperplane, and therefore they do touch each other.

2.4 Cluster Persistence under Horizontal Uncertainty

Using finitely many samples, we can only expect estimates of the level sets M_ρ that are both vertically and horizontally uncertain. While to some extent the vertical uncertainty has already been addressed by the persistence assumed in our cluster definition, the horizontal uncertainty has not been addressed, so far. Therefore, the goal of this subsection is to investigate under which conditions a controlled horizontal uncertainty does not affect the persistence.

To begin with, let us recall some notions that will help us to describe what we mean by horizontal uncertainty. To this end, let us fix a metric space (X, d) and some $A \subset X$. For $\delta > 0$, the δ -tube around A is then defined by

$$A^{+\delta} := \{x \in X : d(x, A) \leq \delta\}.$$

Conceptually similar is the operation of cutting off a δ -tube from A , namely

$$A^{-\delta} := X \setminus (X \setminus A)^{+\delta}.$$

Clearly, we have $A^{-\delta} \subset A \subset A^{+\delta}$, and both $(X \setminus A)^{+\delta} = X \setminus A^{-\delta}$ and $(X \setminus A)^{-\delta} = X \setminus A^{+\delta}$. Despite the notation, however, both operations are anything than inverse to each other in the exponent. Namely, in general we have $(A^{+\delta})^{-\delta} \neq A$ and $(A^{-\delta})^{+\delta} \neq A$. Nonetheless, for $\epsilon > 0$, we have at least $A \subset (A^{+\delta+\epsilon})^{-\delta}$ and $(A^{-\delta-\epsilon})^{+\delta} \subset A$, see Lemma 6.5, which also collects some other useful inclusions related to these operations.

Remark 2.18. In the literature there is another, closely related concept for adding and cutting off δ -tubes, which based on the Minkowski addition and difference of sets. Namely, in generic metric spaces (X, d) , we can define

$$\begin{aligned} A^{\oplus\delta} &:= \{x \in X : \exists y \in A \text{ with } d(x, y) \leq \delta\} \\ A^{\ominus\delta} &:= \{x \in X : B(x, \delta) \subset A\} \end{aligned}$$

for $A \subset X$ and $\delta > 0$, where $B(x, \delta) := \{y \in X : d(x, y) \leq \delta\}$ denotes the closed ball with radius δ and center x . Some simple considerations then show $A^{\ominus(\delta+\epsilon)} \subset A^{-\delta} \subset A^{\ominus\delta}$ and $A^{\oplus\delta} \subset A^{+\delta} \subset A^{\oplus(\delta+\epsilon)}$ for all $\epsilon, \delta > 0$, that is, the operations of both concepts almost coincide. In addition, it is straightforward to check that $A^{\ominus\delta} = X \setminus (X \setminus A)^{\oplus\delta}$.

Usually, the operations $\oplus\delta$ and $\ominus\delta$ are considered for the special case $X := \mathbb{R}^d$ equipped with the Euclidean norm. In this case, we immediately obtain the more common expressions

$$\begin{aligned} A^{\oplus\delta} &= \{x + y : x \in A \text{ and } y \in \delta B_{\ell_2^d}\} \\ A^{\ominus\delta} &= \{x \in \mathbb{R}^d : x + \delta B_{\ell_2^d} \subset A\}, \end{aligned}$$

where $B_{\ell_2^d}$ denotes the closed unit Euclidean ball at the origin. Note that the latter formulas remain true for sufficiently small $\delta > 0$, if we consider the ‘‘relative case’’ $X \subset \mathbb{R}^d$ and subsets $A \subset X$ satisfying $d(A, \mathbb{R}^d \setminus X) \in (0, \infty)$.

In general, it is quite cumbersome to determine the exact forms of $A^{+\delta}$ and $A^{-\delta}$, respectively $A^{\oplus\delta}$ and $A^{\ominus\delta}$ for a given A . For a particular class of sets $A \subset \mathbb{R}^2$, Example 7.1 illustrates this by providing both $A^{\oplus\delta}$ and $A^{\ominus\delta}$.

Let us now assume that we have an algorithm that can only estimate the true level sets M_ρ up to some δ -tube, that is

$$M_\rho^{-\delta} \subset \hat{M}_\rho \subset M_\rho^{+\delta},$$

where \hat{M}_ρ is the estimate generated by the algorithm. To use \hat{M}_ρ for identifying the connected components of M_ρ , it then seems natural to relate the connected components of \hat{M}_ρ to those of $M_\rho^{-\delta}$ and $M_\rho^{+\delta}$. Clearly, this approach can only be successful, if the connected components of $M_\rho^{-\delta}$ and $M_\rho^{+\delta}$ can in turn be related to those of M_ρ . While for sufficiently small δ it can be shown similarly to Lemma 2.13 that $\mathcal{C}(M_\rho^{\pm\delta})$ is persistent in $\mathcal{C}(M_\rho)$, this is no longer true for $M_\rho^{-\delta}$. Indeed, in the presence of thin bridges the cutting operation may split a connected component into two. Consequently, we need a method to carefully glue connected components together. As we will see later, considering τ -connected components instead of connected components, is such a method provided that τ and δ satisfy certain constraints. Keeping this motivation in mind, our goal of this section is thus to investigate under which conditions $\mathcal{C}_\tau(M_\rho^{-\delta})$ and $\mathcal{C}_\tau(M_\rho^{+\delta})$ are persistent to $\mathcal{C}(M_\rho)$.

Let us begin with the following lemma that establishes properties for the τ -CRM between a set A and $A^{+\delta}$. Roughly speaking, it states, that $\mathcal{C}_\tau(A)$ is persistent in $\mathcal{C}_\tau(A^{+\delta})$, if $\tau > 0$ and $\delta > 0$ are sufficiently small.

Lemma 2.19. *Let (X, d) be a compact metric space, and $A \subset X$ be a non-empty subset of X . Then, for all $\delta > 0$ and $\tau > \delta$, the following statements hold:*

- i) *The set $(A')^{+\delta}$ is τ -connected for all $A' \in \mathcal{C}_\tau(A)$.*
- ii) *The τ -CRM $\zeta : \mathcal{C}_\tau(A) \rightarrow \mathcal{C}_\tau(A^{+\delta})$ is surjective.*
- iii) *If A is closed, $|\mathcal{C}(A)| < \infty$, and $\tau \leq \tau_A^*/3$, then the τ -CRM $\zeta : \mathcal{C}_\tau(A) \rightarrow \mathcal{C}_\tau(A^{+\delta})$ is bijective and satisfies*

$$\zeta(A') = (A')^{+\delta}, \quad A' \in \mathcal{C}_\tau(A). \quad (9)$$

With the help of the preceding lemma we can now establish our first persistence result, which compares the τ -connected components of M_ρ with those of $M_\rho^{+\delta}$.

Theorem 2.20. *Let (X, d) be a compact metric space, μ be a finite Borel measure on X and P be a μ -absolutely continuous Borel probability measure on X that can be topologically clustered between the critical levels ρ^* and ρ^{**} . We define the function $\tau^* : (0, \rho^{**} - \rho^*] \rightarrow (0, \infty)$ by*

$$\tau^*(\varepsilon) := \frac{1}{3} \tau_{M_{\rho^* + \varepsilon}}^*. \quad (10)$$

Then τ^ is monotonically increasing. Moreover, for all $\varepsilon^* \in (0, \rho^{**} - \rho^*]$, $\delta > 0$, $\tau \in (\delta, \tau^*(\varepsilon^*])$, and all $\rho \in [0, \rho^{**}]$, the following statements hold:*

i) We have $1 \leq |\mathcal{C}_\tau(M_\rho^{+\delta})| \leq 2$.

ii) If $\rho \geq \rho^* + \varepsilon^*$, then $|\mathcal{C}_\tau(M_\rho^{+\delta})| = 2$ and the CRM $\zeta : \mathcal{C}(M_\rho) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$ is bijective.

iii) If $|\mathcal{C}_\tau(M_\rho^{+\delta})| = 2$, then $\rho \geq \rho^*$ and the τ -CRM $\zeta : \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta}) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$ is bijective.

iv) If the τ -CRM $\zeta^{**} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$ is bijective and $|\mathcal{C}_\tau(M_\rho^{-\delta})| = 1$, then $\rho < \rho^* + \varepsilon^*$.

The first three statements of Theorem 2.20 basically show that for sufficiently small δ and τ the τ -connected component structure of M_ρ is not changed when δ -tubes are added. Not surprisingly, however, the meaning of “sufficiently small”, which is expressed by the function τ^* , changes when we approach the critical level ρ^* from above. Moreover, even for $0 < \delta < \tau \leq \tau^*(\varepsilon^*)$, Theorem 2.20 does not specify the component structure of $\mathcal{C}_\tau(M_\rho^{+\delta})$ for levels close to ρ^* , that is $\rho \in [\rho^*, \rho^* + \varepsilon^*]$. We will see later, that these two facts will complicate our analysis significantly.

The assumed bijectivity of $\zeta^{**} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$ in *iv*) means that the τ -connected component structure of $M_{\rho^{**}}$ is not changed by cutting off δ -tubes, and the corresponding conclusion essentially states that this is actually true for all levels $\rho \in [\rho^* + \varepsilon^*, \rho^{**}]$. Our next goal is to further investigate persistence under the cutting operation. We begin with the following lemma, which investigates situations in which $\mathcal{C}_\tau(A^{-\delta})$ is persistent in $\mathcal{C}(A)$.

Lemma 2.21. *Let (X, d) be a compact metric space, and $A \subset X$ be a non-empty closed subset of X with $|\mathcal{C}(A)| < \infty$. We define the function $\psi_A^* : (0, \infty) \rightarrow [0, \infty]$ by*

$$\psi_A^*(\delta) := \sup_{x \in A} d(x, A^{-\delta}), \quad \delta > 0.$$

Then, for all $\delta > 0$ and all $\tau > 2\psi_A^*(\delta)$, the following statements hold:

i) For all $B' \in \mathcal{C}(A)$, there exists at most one $A' \in \mathcal{C}_\tau(A^{-\delta})$ such that $A' \cap B' \neq \emptyset$.

ii) We have $|\mathcal{C}_\tau(A^{-\delta})| \leq |\mathcal{C}(A)|$.

iii) If $|\mathcal{C}_\tau(A^{-\delta})| = |\mathcal{C}(A)|$, then $\mathcal{C}_\tau(A^{-\delta})$ is persistent in $\mathcal{C}(A)$. Moreover, for all $B', B'' \in \mathcal{C}(A)$ with $B' \neq B''$ we have

$$d(B', B'') \geq \tau - 2\psi_A^*(\delta). \quad (11)$$

Part *iii*) of Lemma 2.21 states that if τ is sufficiently large compared to δ and $|\mathcal{C}_\tau(A^{-\delta})| = |\mathcal{C}(A)|$, then we obtain persistence. Informally speaking this means that gluing δ -cuts by τ -connectivity may preserve the component structure in some situations.

The lemma above suggests that like τ_A^* , the function ψ_A^* will play a central role in our analysis. Let us therefore consider ψ_A^* in some more detail. To this end, we first note that the function ψ_A^* can actually be defined for arbitrary non-empty sets A in arbitrary metric spaces (X, d) . In the following discussion we always consider this general case.

Our first observation is that the definition of ψ_A^* immediately yields $A \subset (A^{-\delta})^{+\psi_A^*(\delta)}$ for all $\delta > 0$ with $\psi_A^*(\delta) < \infty$, and it is also straightforward to see that $\psi_A^*(\delta)$ is the smallest $\psi > 0$, for which this inclusion holds, that is

$$\psi_A^*(\delta) = \min\{\psi \geq 0 : A \subset (A^{-\delta})^{+\psi}\}$$

for all $\delta > 0$. In other words, $\psi_A^*(\delta)$ gives the size of the smallest tube needed to recover a superset of A from $A^{-\delta}$. In particular, if δ is too large, that is $A^{-\delta} = \emptyset$, we obviously have $\psi_A^*(\delta) = \infty$ and no recovery is possible.

Intuitively it is not surprising that ψ_A^* grows at least linearly, that is

$$\psi_A^*(\delta) \geq \delta \quad (12)$$

for all $\delta > 0$ provided that $d(A, X \setminus A) = 0$. Indeed, $\psi_A^*(\delta) < \delta$ for some $\delta > 0$ gives us an $\varepsilon > 0$ such that $d(x, A^{-\delta}) < \delta - \varepsilon$ for all $x \in A$. Since $d(A, X \setminus A) = 0$ there then exists an $x \in A$ with

$d(x, X \setminus A) < \epsilon$, and for this x there exists an $x' \in A^{-\delta}$ with $d(x, x') < \delta - \epsilon$. Now the definition of $A^{-\delta}$ gives $d(x', X \setminus A) > \delta$, and hence we find a contradiction by

$$\delta < d(x', X \setminus A) \leq d(x', x) + d(x, X \setminus A) < \delta.$$

For generic sets A , the function ψ_A^* is usually hard to bound, but for some classes of sets, ψ_A^* can actually be computed precisely. For example, if $I \subset \mathbb{R}$ is a bounded and closed interval, say $I = [a, b]$, then $\psi_I^*(\delta) = \delta$ for all $\delta \in (0, (b-a)/2]$, and $\psi_I^*(\delta) = \infty$, otherwise. Clearly, this example can be extended to finite unions of such intervals and for intervals that are not closed, the only difference occurs at $\delta = (b-a)/2$. In higher dimensions, an interesting class of sets A with linear behavior of ψ_A^* is described by Serra's model, see (Serra, 1982, p. 144), that consist of all compact sets $A \subset \mathbb{R}^d$ for which there is a $\delta_0 > 0$ with

$$A = (A^{\ominus \delta_0})^{\oplus \delta_0} = (A^{\oplus \delta_0})^{\ominus \delta_0}.$$

If, in addition, A is path-connected, then (Walther, 1999, Thm. 1) shows that this relation also holds for all $\delta \in (0, \delta_0]$. In this case, we then obtain by Remark 2.18

$$A = (A^{\ominus(\delta+\epsilon)})^{\oplus(\delta+\epsilon)} \subset (A^{\ominus(\delta+\epsilon)})^{+\delta+\epsilon} \subset (A^{-\delta})^{+\delta+\epsilon}$$

for all $\delta \in (0, \delta_0)$ and $0 < \epsilon \leq \delta_0 - \delta$. In other words, we have $\psi_A^*(\delta) \leq \delta + \epsilon$, and letting $\epsilon \rightarrow 0$, we thus conclude $\psi_A^*(\delta) = \delta$ for all $\delta \in (0, \delta_0)$. In addition, with the help of Lemma 6.5, it is not hard to see that this result generalizes to finite unions of compact, path-connected sets, which has already been observed in Walther (1999). Finally, note that (Walther, 1999, Thm. 1) also provides some useful characterizations of (path-connected) compact sets belonging to Serra's model. In a nutshell, these are the sets whose boundary is a $(d-1)$ -dimensional sub-manifold of \mathbb{R}^d with outward pointing unit normal vectors satisfying a Lipschitz condition.

Our analysis does not require the exact form of ψ_A^* , but only its asymptotic behavior for $\delta \rightarrow 0$. Therefore it is interesting to note that ψ_A^* is also asymptotically invariant against bi-Lipschitz transformations. To be more precise, let (X, d) and (Y, e) be two metric spaces and $I : X \rightarrow Y$ be a bijective map for which there exists a constant $C > 0$ such that

$$C^{-1}e(I(x), I(x')) \leq d(x, x') \leq Ce(I(x), I(x'))$$

for all $x, x' \in X$. For $A \subset X$ and $\delta > 0$, we then we have $I(A^{+\delta/C}) \subset (I(A))^{+\delta} \subset I(A^{+C\delta})$, which in turn implies

$$C^{-1}\psi_A^*(\delta/C) \leq \psi_{I(A)}^*(\delta) \leq C\psi_A^*(C\delta)$$

for all $\delta > 0$. In particular, we have $\psi_A^*(\delta) \preceq \delta^\gamma$ for some $\gamma \in (0, 1]$ if and only if $\psi_{I(A)}^*(\delta) \preceq \delta^\gamma$.

Last but not least we like to mention that based on the sets $A \subset \mathbb{R}^2$ considered in Example 7.1, Example 7.2 estimates ψ_A^* . In particular, this example provides various sets A with $\psi_A^*(\delta) \sim \delta$ that do not belong to Serra's model, and this class of sets can be further expanded by using bi-Lipschitz transformations as discussed above.

Let us now return to the persistence of level sets under the cutting operation. To this end, we need the following definition.

Definition 2.22. *Let (X, d) be a compact metric space, μ be a σ -finite Borel measure on X and P be a μ -absolutely continuous Borel probability measure on X . Then we say that, up to the level $\rho^{**} > 0$, the distribution P has thick levels of order $\gamma \in (0, 1]$, if there exist $c_{\text{thick}} \geq 1$ and $\delta_{\text{thick}} \in (0, 1]$ such that, for all $\delta \in (0, \delta_{\text{thick}}]$, $\rho \in [0, \rho^{**}]$, we have*

$$\psi_{M_\rho}^*(\delta) \leq c_{\text{thick}} \delta^\gamma. \tag{13}$$

In this case, we call $\psi : (0, \infty) \rightarrow (0, \infty)$, defined by $\psi(\delta) := 3c_{\text{thick}}\delta^\gamma$, the thickness function.

With the help of the discussion following Lemma 2.21 it is easy to see that we have $M_\rho \subset (M^{-\delta})^{+\psi(\delta)/2}$ for all $\delta \in (0, \delta_{\text{thick}}]$ and all $\rho \in (0, \rho^{**}]$. In addition, it becomes clear that exponents $\gamma > 1$ are impossible as soon as $d(M_\rho, X \setminus M_\rho) = 0$ for some $\rho \in (0, \rho^{**}]$.

Intuitively, Definition 2.22 excludes thin cusps and bridges, where the thinness and length of both is controlled by γ . Such assumptions have been widely used in the literature on level set estimation and density-based clustering. For example, a basically identical assumption has been made in Singh et al. (2009) for the exponent $\gamma = 1$, which can be taken, if, e.g., the level sets belong to Serra's model. Moreover, level sets belonging to Serra's model have been investigated in Walther (1997). In particular, (Walther, 1997, Thm. 2) shows that most level sets of a C^1 -density with Lipschitz continuous gradient belong to Serra's model. Unfortunately, however, levels at which the density has a saddle point are excluded in this theorem, and some other elementary sets such as cubes in \mathbb{R}^d do not belong to Serra's model, either. For this reason, we allow constants $c_{\text{thick}} > 1$ and exponents $\gamma < 1$. Finally, a less geometric assumption excluding thin features, known as standardness, has been used by various authors, see e.g. Cuevas and Fraiman (1997), Cuevas et al. (2000), Rigollet (2007) and the references therein, and an overview of these and similar assumptions can be found in Cuevas (2009).

Understanding (13) in the one-dimensional case is very simple. Indeed, if $X \subset \mathbb{R}$ is an interval and P can be topologically clustered between the critical levels ρ^* and ρ^{**} , then every level set M_ρ consists of either one or two closed intervals, since intervals are the only topologically connected sets in \mathbb{R} . Using this, the discussion following Lemma 2.21 shows that P actually has thick levels of order $\gamma = 1$ up to the level ρ^{**} . Moreover, a possible thickness function is $\psi(\delta) = 3\delta$ for all $\delta \in (0, \delta_{\text{thick}}]$, where δ_{thick} equals the smaller radius of the two intervals at level ρ^{**} .

Using the discussion following Lemma 2.21 it is not hard to construct distributions with discontinuous densities that have thick levels of order, e.g. $\gamma = 1$. For continuous densities, however, this task is significantly harder due to saddle point effects at the critical level ρ^* . Nonetheless, Example 7.4 provides a large class of such densities in the case $X \subset \mathbb{R}^2$.

Let us now summarize the assumptions that will be used in the following.

Assumption C. *We have a compact metric space (X, d) , a finite Borel measure μ on X with $\text{supp } \mu = X$, and a μ -absolutely continuous probability measure P that can be topologically clustered between the critical levels ρ^* and ρ^{**} . In addition, we assume that up to the level $\rho^{**} > 0$, the distribution P has thick levels of order $\gamma \in (0, 1]$. We denote the corresponding thickness function by ψ and write τ^* for the function defined in (10).*

Our last result of this subsection, which is the counterpart of Theorem 2.20, shows that, for thick clusters, the connected component structure of M_ρ is not changed when cutting off δ -tubes.

Theorem 2.23. *Let Assumption C be satisfied. Then, for all $\varepsilon^* \in (0, \rho^{**} - \rho^*]$, $\delta \in (0, \delta_{\text{thick}}]$, $\tau \in (\psi(\delta), \tau^*(\varepsilon^*)]$, and all $\rho \in [0, \rho^{**}]$, the following statements hold:*

- i) *We have $1 \leq |\mathcal{C}_\tau(M_\rho^{-\delta})| \leq 2$.*
- ii) *The τ -CRM $\zeta^{**} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})$ is bijective.*
- iii) *If $|\mathcal{C}_\tau(M_\rho^{-\delta})| = 2$, then $\rho \geq \rho^*$ and the τ -CRM $\zeta : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_\rho^{-\delta})$ is bijective.*

Intuitively, considering $\mathcal{C}_\tau(M_\rho^{-\delta})$ rather than $\mathcal{C}(M_\rho^{-\delta})$ means that we add a τ -tube around $M_\rho^{-\delta}$. By Theorem 2.23, the thickness of the level sets then ensure that $\mathcal{C}_\tau(M_\rho^{-\delta})$ and M_ρ have the same component structure, or to say it in simple words, considering τ -connected components glues together what has been accidentally cut by removing δ -tubes.

3 The Algorithm and its Finite Sample Behaviour

In this section, we introduce our clustering algorithm and present first results on its clustering ability. In a nutshell, this algorithm first estimates density level sets with the help of a standard histogram-based density estimator. It then identifies τ -connected components of the estimating level sets and prunes certain components. The smallest level, at which we have more than one remaining τ -connected component is then our estimate of the critical level ρ^* .

While the algorithm is conceptionally simple, its analysis turns out to be laborious, mostly because we have to ensure that, at least for most of the considered levels, the estimated τ -connected

component structure is closely related to the connected component structure of the true level set. A key step towards this relation will be presented in Lemma 3.4, which specifies the vertical and horizontal uncertainty of the estimating level set. Theorem 3.5 then gives a simple criterion for the pruning operation, and Theorems 3.6 and 3.8 provide a detailed finite-sample analysis of the clustering algorithm.

Let us begin by recalling that histograms are based on partitions of the input space X . In the following, we need to ensure that the partitions we use are geometrically well-behaved. To this end, we need the diameter of a subset $A \subset X$, that is,

$$\text{diam } A := \sup_{x, x' \in A} d(x, x').$$

Now, the following definition describes partitions that are controlled in size, measure, and shape.

Definition 3.1. *Let (X, d) be a compact metric space and μ be a finite Borel measure on X with $\text{supp } \mu = X$. Moreover, for each $\delta \in (0, 1]$, let $\mathcal{A}_\delta = (A_1, \dots, A_{m_\delta})$ be a finite partition of X . Then $(\mathcal{A}_\delta)_{\delta \in (0, 1]}$ is called a uniform family of d -dimensional partitions of X , where $d > 0$, if there exists a constant $c_{\text{part}} \geq 1$ such that, for all $\delta \in (0, 1]$ and all $i = 1, \dots, m_\delta$, we have*

$$\begin{aligned} \text{diam } A_i &\leq \delta, \\ m_\delta &\leq c_{\text{part}} \delta^{-d}, \\ \mu(A_i) &\geq c_{\text{part}}^{-1} \delta^d. \end{aligned}$$

The easiest yet most important examples of uniform families of partitions are hypercube partitions in combination with the Lebesgue measure. The following example presents the details.

Example 3.2. *Let $X := [0, 1]^d$ be equipped with the metric defined by the supremum norm $\|\cdot\|_{\ell_\infty^d}$, and λ^d be the d -dimensional Lebesgue measure. For $\delta \in (0, 1]$, there then exists a unique $\ell \in \mathbb{N}$ with $\frac{1}{\ell+1} < \delta \leq \frac{1}{\ell}$. We define $h := \frac{1}{1+\ell}$ and write \mathcal{A}_δ for the usual partition of $[0, 1]^d$ into hypercubes of side-length h . Then, for each $A_i \in \mathcal{A}_\delta$, we have $\text{diam } A_i = h \leq \delta$ and $\lambda^d(A_i) = h^d \geq 2^{-d}\delta^d$. Moreover, we obviously have $|\mathcal{A}_\delta| = h^{-d} \leq 2^d \delta^{-d}$. Consequently, $(\mathcal{A}_\delta)_{\delta \in (0, 1]}$ is a uniform family of d -dimensional partitions of X with $c_{\text{part}} := 2^d$.*

In the following, we will mainly deal with situations in which Assumption A is satisfied and we have a uniform family of partitions. For convenience, the following assumption summarizes this.

Assumption A. *Assumption C is satisfied and we have a uniform family $(\mathcal{A}_\delta)_{\delta \in (0, 1]}$ of d -dimensional partitions of X .*

Let us now assume that (X, d) is a compact metric space and μ is a finite Borel measure with $\text{supp } \mu = X$, such that we have a uniform family of d -dimensional partitions $(\mathcal{A}_\delta)_{\delta \in (0, 1]}$ of X . For a fixed $\delta > 0$, we write $\mathcal{A}_\delta = (A_1, \dots, A_m)$. Given a probability measure P on X , we then define the corresponding histogram by

$$h_{P, \delta}(x) := \sum_{j=1}^m \frac{P(A_j)}{\mu(A_j)} \cdot \mathbf{1}_{A_j}(x), \quad x \in X,$$

where $\mathbf{1}_A$ denotes the indicator function of a set A . Let us now assume that we have a data set $D = (x_1, \dots, x_n) \in X^n$. In a slight abuse of notation, we denote the corresponding empirical measure by D , that is $D := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$, where δ_x is the Dirac measure at the point x . For $A \subset X$ this gives

$$D(A) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_A(x_i),$$

and the corresponding (empirical) histogram is

$$h_{D, \delta}(x) = \sum_{j=1}^m \frac{D(A_j)}{\mu(A_j)} \cdot \mathbf{1}_{A_j}(x), \quad x \in X. \quad (14)$$

Our first result in this section shows, that, for i.i.d. observations D , the empirical histogram $h_{D, \mathcal{A}_\delta}$ uniformly approximates the infinite-sample histogram $h_{P, \delta}$ with high probability.

Theorem 3.3. *Let (X, d) be a compact metric space, μ be a finite Borel measure on X with $\text{supp } \mu = X$, and $(\mathcal{A}_\delta)_{\delta \in (0,1]}$ be a uniform family of d -dimensional partitions of X . Moreover, let P be a probability measure on X . Then, for all $n \geq 1$, $\varepsilon > 0$, and $\delta > 0$, we have*

$$P^n \left(\left\{ D \in X^n : \|h_{D,\delta} - h_{P,\delta}\|_\infty < \varepsilon \right\} \right) \geq 1 - 2c_{\text{part}} \exp \left(-d \ln \delta - \frac{2n\varepsilon^2 \delta^{2d}}{c_{\text{part}}^2} \right).$$

In addition, if P is μ -absolutely continuous and there exists a bounded μ -density h of P , then we have

$$P^n \left(\left\{ D \in X^n : \|h_{D,\delta} - h_{P,\delta}\|_\infty < \varepsilon \right\} \right) \geq 1 - 2c_{\text{part}} \exp \left(-d \ln \delta - \frac{3n\varepsilon^2 \delta^d}{c_{\text{part}}(6\|h\|_\infty + 2\varepsilon)} \right).$$

As already indicated above, our clustering algorithm will use the set $\{h_{D,\delta} \geq \rho\}$ to estimate the true level set M_ρ . The following key lemma specifies the accuracy of this approach.

Lemma 3.4. *Let (X, d) be a compact metric space, μ be a finite Borel measure on X with $\text{supp } \mu = X$, and $(\mathcal{A}_\delta)_{\delta \in (0,1]}$ be a uniform family of d -dimensional partitions of X . Moreover, let P be a μ -absolutely continuous probability measure on X and $\hat{h} : X \rightarrow \mathbb{R}$ be a function with $\|\hat{h} - h_{P,\delta}\|_\infty \leq \varepsilon$ for some $\varepsilon \geq 0$. Then, for all $\rho \geq 0$, the following statements hold:*

- i) If P is upper normal at the level $\rho + \varepsilon$, then we have $M_{\rho+\varepsilon}^{-\delta} \subset \{\hat{h} \geq \rho\}$.*
- ii) If P is upper normal at the level $\rho - \varepsilon$, then we have $\{\hat{h} \geq \rho\} \subset M_{\rho-\varepsilon}^{+\delta}$.*

Note that, for $\varepsilon = 0$, the lemma above shows $M_\rho^{-\delta} \subset \{h_{P,\delta} \geq \rho\} \subset M_\rho^{+\delta}$, that is, the horizontal inaccuracy of approximating M_ρ by $\{h_{P,\delta} \geq \rho\}$ can be described by adding and cutting-off δ -tubes to and from M_ρ . The additional error of using an ε -approximate \hat{h} of $h_{P,\delta}$, for example $h_{D,\delta}$, directly translates into the levels $\rho + \varepsilon$ and $\rho - \varepsilon$ of the enclosing sets $M_{\rho+\varepsilon}^{-\delta}$ and $M_{\rho-\varepsilon}^{+\delta}$. In other words, our statistical uncertainty causes vertical uncertainty, while the geometric inaccuracy of our partition is responsible for the horizontal uncertainty.

Motivated by Lemma 3.4, our next goal is to relate the τ -connected components of our estimate $\{\hat{h} \geq \rho\}$ to the τ -connected components of $M_{\rho+\varepsilon}^{-\delta}$. To make the corresponding results reusable for possible future clustering algorithms, we formulate the following theorems for generic estimators. These generic results are then applied to our histogram-based algorithm in Theorem 3.8.

Theorem 3.5. *Let Assumption C be satisfied. Furthermore, let $\varepsilon^* \in (0, \rho^{**} - \rho^*]$, $\delta \in (0, \delta_{\text{thick}}]$, $\tau \in (\psi(\delta), \tau^*(\varepsilon^*)]$, and $\varepsilon \in (0, \varepsilon^*]$. In addition, let $(L_\rho)_{\rho \geq 0}$ be a decreasing family of sets $L_\rho \subset X$ such that*

$$M_{\rho+\varepsilon}^{-\delta} \subset L_\rho \subset M_{\rho-\varepsilon}^{+\delta} \tag{15}$$

*for all $\rho \geq 0$. Then for all $\rho \in [0, \rho^{**} - 3\varepsilon]$, the following disjoint union holds*

$$\mathcal{C}_\tau(L_\rho) = \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})) \cup \{B' \in \mathcal{C}_\tau(L_\rho) : B' \cap L_{\rho+2\varepsilon} = \emptyset\}, \tag{16}$$

where $\zeta : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(L_\rho)$ is the τ -CRM, whose existence is guaranteed by (15) and Lemma 2.12.

Theorem 3.5 shows that, for suitably chosen δ , ε , and τ , all τ -connected components B' of our estimate L_ρ of M_ρ are either contained in $\zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$ or satisfy $B' \cap L_{\rho+2\varepsilon} = \emptyset$. Now, if the latter components are easy to find and remove, we have a device that allows us to identify exactly the τ -connected components B' that are contained in $\zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$. This suggests that, starting with $\rho = 0$, we only need to scan through the values of ρ . Algorithm 3.1 formalizes this idea.

Obviously, the described heuristic of Algorithm 3.1 only makes sense if we can relate the τ -connected component structure of $M_{\rho+\varepsilon}^{-\delta}$ to the topologically connected component structure of M_ρ for “most” ρ . The following result shows that this is indeed the case. It further shows that the level ρ_D^* returned by Algorithm 3.1 estimates ρ^* .

Algorithm 3.1 Clustering with the help of a generic level set estimator

Require: Some $\tau > 0$ and $\varepsilon > 0$.

An algorithm that produces, for all data sets $D \in X^n$, a decreasing family $(L_{D,\rho})_{\rho \geq 0}$ of subsets of X .

Ensure: An estimate of ρ^* and the topological clusters A_1^* and A_2^* .

1: $\rho \leftarrow 0$

2: **repeat**

3: Identify the τ -connected components B'_1, \dots, B'_M of $L_{D,\rho}$ satisfying

$$B'_i \cap L_{D,\rho+2\varepsilon} \neq \emptyset.$$

4: $\rho \leftarrow \rho + \varepsilon$

5: **until** $M \neq 1$

6: $\rho \leftarrow \rho + 2\varepsilon$

7: Identify the τ -connected components B'_1, \dots, B'_M of $L_{D,\rho}$ satisfying

$$B'_i \cap L_{D,\rho+2\varepsilon} \neq \emptyset.$$

8: **return** $\rho_D^* := \rho$ and the sets $B_i(D) := B'_i$ for $i = 1, \dots, M$.

Theorem 3.6. *Let Assumption C be satisfied. Furthermore, let $\varepsilon^* \leq (\rho^{**} - \rho^*)/9$, $\delta \in (0, \delta_{\text{thick}}]$, $\tau \in (\psi(\delta), \tau^*(\varepsilon^*)]$, and $\varepsilon \in (0, \varepsilon^*]$. In addition, let D be a data sets and $(L_{D,\rho})_{\rho \geq 0}$ be a decreasing family satisfying*

$$M_{\rho+\varepsilon}^{-\delta} \subset L_{D,\rho} \subset M_{\rho-\varepsilon}^{+\delta}$$

for all $\rho \geq 0$. Furthermore, assume that Algorithm 3.1 receives the parameters τ , ε , and $(L_{D,\rho})_{\rho \geq 0}$. Then, the following statements are true:

i) The returned level ρ_D^* satisfies $\rho_D^* \in [\rho^* + 2\varepsilon, \rho^* + \varepsilon^* + 5\varepsilon]$.

ii) We have $|\mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta})| = 2$ and the τ -CRM $\zeta : \mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(L_{D,\rho_D^*})$ is injective.

iii) Algorithm 3.1 returns the two τ -connected components of $\zeta(\mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta}))$.

iv) There exist CRMs $\zeta_{\rho^{**}} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}(M_{\rho^{**}})$ and $\zeta_{\rho_D^*+\varepsilon} : \mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}(M_{\rho_D^*+\varepsilon})$ such that the following diagram

$$\begin{array}{ccc} \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) & \xrightarrow{\zeta_{\rho^{**}}} & \mathcal{C}(M_{\rho^{**}}) \\ \zeta_{\rho^{**}, \rho_D^*+\varepsilon} \downarrow & & \downarrow \tilde{\zeta} \\ \mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta}) & \xrightarrow{\zeta_{\rho_D^*+\varepsilon}} & \mathcal{C}(M_{\rho_D^*+\varepsilon}) \end{array}$$

commutes, where $\zeta_{\rho^{**}, \rho_D^*+\varepsilon}$ is the τ -CRM and $\tilde{\zeta}$ is the top-CRM. Moreover, every map in the diagram is bijective.

v) The returned level ρ_D^* satisfies $\tau - \psi(\delta) < 3\tau^*(\rho_D^* - \rho^* + \varepsilon)$.

To illustrate the meaning of part iv) of Theorem 3.6 let us assume that we are in the situation of this theorem. Moreover, let A_1 and A_2 be the two topologically connected components of $M_{\rho^{**}}$, and

$$V_{\rho_D^*+\varepsilon}^i := \zeta_{\rho^{**}, \rho_D^*+\varepsilon}(\zeta_{\rho^{**}}^{-1}(A_i)), \quad i = 1, 2, \quad (17)$$

be the two τ -connected components of $M_{\rho_D^*+\varepsilon}^{-\delta}$. Note part iv) ensures that we can actually make this definition, and, in addition, it shows $V_{\rho_D^*+\varepsilon}^1 \neq V_{\rho_D^*+\varepsilon}^2$. Moreover, by part ii) and iii) we may assume that the sets returned by Algorithm 3.1 are ordered in the sense of $B_i(D) = \zeta(V_{\rho_D^*+\varepsilon}^i)$, that is

$$B_i(D) = \zeta \circ \zeta_{\rho^{**}, \rho_D^*+\varepsilon}(\zeta_{\rho^{**}}^{-1}(A_i)), \quad i = 1, 2. \quad (18)$$

Now, this definition ensures $V_{\rho_D^* + \varepsilon}^i \subset B_i(D)$, while the diagram gives $V_{\rho_D^* + \varepsilon}^i \subset \tilde{\zeta}(A_i) \subset A_i^*$. For $i = 1, 2$, we consequently have

$$V_{\rho_D^* + \varepsilon}^i \subset B_i(D) \cap A_i^*,$$

i.e. each returned component $B_i(D)$ contains a chunk of the desired cluster A_i^* . The following result refines this analysis.

Theorem 3.7. *Suppose that the assumptions of Theorem 3.6 are satisfied, and that the two sets returned by Algorithm 3.1 are ordered in the sense of (18). For $i = 1, 2$, we write $A_{\rho_D^* + \varepsilon}^i := \tilde{\zeta}(A_i)$. Then we have*

$$\begin{aligned} \mu(B_1(D) \Delta A_1^*) + \mu(B_2(D) \Delta A_2^*) &\leq 2\mu(A_1^* \setminus (A_{\rho_D^* + \varepsilon}^1)^{-\delta}) + 2\mu(A_2^* \setminus (A_{\rho_D^* + \varepsilon}^2)^{-\delta}) \\ &\quad + \mu(M_{\rho_D^* - \varepsilon}^{+\delta} \setminus \{h > \rho^*\}). \end{aligned} \quad (19)$$

Let us assume for a moment that we wish to estimate how well the clusters $B_i(D)$ returned by Algorithm 3.1 estimate the true clusters. By part *i*) of Theorem 3.6 we then know $\rho_D^* \in [\rho^* + 2\varepsilon, \rho^* + \varepsilon^* + 5\varepsilon]$, and hence the estimate of Theorem 3.7 yields

$$\begin{aligned} \mu(B_1(D) \Delta A_1^*) + \mu(B_2(D) \Delta A_2^*) &\leq 2\mu(A_1^* \setminus (A_{\rho^* + \varepsilon^* + 6\varepsilon}^1)^{-\delta}) + 2\mu(A_2^* \setminus (A_{\rho^* + \varepsilon^* + 6\varepsilon}^2)^{-\delta}) \\ &\quad + \mu(M_{\rho^* + \varepsilon}^{+\delta} \setminus \{h > \rho^*\}), \end{aligned}$$

where $A_{\rho^* + \varepsilon^* + 6\varepsilon}^i$ are the connected components at level $\rho^* + \varepsilon^* + 6\varepsilon$. Consequently, it suffices to understand how the mass of the three sets on the left-hand side of this estimate behave for $\delta \rightarrow 0$ and $\varepsilon^* \rightarrow 0$.

Having understood the generic clustering Algorithm 3.1 in detail, let us now return to the concrete example of clustering with the help of histogram-based density level set estimators. The following theorem provides a finite sample bound for this approach.

Theorem 3.8. *Let Assumption A be satisfied. For some fixed $\delta \in (0, \delta_{\text{thick}}]$, $\varsigma \geq 1$, $n \geq 1$, and $\tau > \psi(\delta)$, we fix an $\varepsilon > 0$ satisfying the bound*

$$\varepsilon \geq c_{\text{part}} \sqrt{\frac{\varsigma + \ln(2c_{\text{part}}) - d \ln \delta}{2\delta^{2d}n}}, \quad (20)$$

or, if P has a bounded μ -density h , the bound

$$\varepsilon \geq \sqrt{\frac{2c_{\text{part}}(1 + \|h\|_{\infty})(\varsigma + \ln(2c_{\text{part}}) - d \ln \delta)}{\delta^{d}n}} + \frac{2c_{\text{part}}(\varsigma + \ln(2c_{\text{part}}) - d \ln \delta)}{3\delta^{d}n}. \quad (21)$$

We further pick an $\varepsilon^* > 0$ satisfying

$$\varepsilon^* \geq \varepsilon + \inf\{\varepsilon' \in (0, \rho^{**} - \rho^*] : \tau^*(\varepsilon') \geq \tau\}. \quad (22)$$

For each data sets $D \in X^n$, we then feed Algorithm 3.1 with the parameters τ and ε , and with the family $(L_{D,\rho})_{\rho \geq 0}$ given by

$$L_{D,\rho} := \{h_{D,\delta} \geq \rho\}, \quad \rho \geq 0.$$

If $\varepsilon^* \leq (\rho^{**} - \rho^*)/9$, then the probability P^n of having a data set $D \in X^n$ satisfying both the assertions *i*) - *v*) of Theorem 3.6 and (19) is not less than $1 - e^{-\varsigma}$.

At this point we like to emphasize that a finite sample bound of the form of Theorem 3.8 can be derived from our analysis whenever Algorithm 3.1 uses density level set estimator guaranteeing the inclusions $M_{\rho+\varepsilon}^{-\delta} \subset L_{D,\rho} \subset M_{\rho-\varepsilon}^{+\delta}$. A possible example of such an alternative level set estimator seems to be a plug-in approach based on a moving window density estimator, since for the latter it is possible to establish a uniform convergence result similar to Theorem 3.3, see e.g. Sriperumbudur and Steinwart (2012), Giné and Guillou (2002). Unfortunately, however, the resulting level sets become computationally infeasible, and hence we have not included this approach, here. In general it remains an open question, whether sets $L_{D,\rho}$ that are constructed differently from the moving window estimator can address this issue. So far, the only known result in this direction is by Sriperumbudur

and Steinwart (2012) who have constructed such sets for α -Hölder-continuous densities with known α .

If the assumption of Theorem 3.8 are satisfied, then Algorithm 3.1 clearly terminates after a finite number of iterations. To show that it actually always terminates after finitely many iterations, we first observe that

$$\|h_{D,\delta}\|_\infty \leq c_{\text{part}}\delta^{-d} \sum_{i=1}^m D(A_i) = c_{\text{part}}\delta^{-d}$$

for all $\delta \in (0, 1]$. After at most $\lceil c_{\text{part}}\delta^{-d}\varepsilon^{-1} \rceil + 1$ iterations we thus have $L_{D,\rho} = \{h_{D,\delta} \geq \rho\} = \emptyset$ in the loop of Algorithm 3.1, so that the loop is terminated at this stage with $M = 0$, if it has not been terminated earlier.

4 Consistency and Rates

At the end of the previous section we established a finite sample bound for Algorithm 3.1 when it receives a histogram-based plug-in level set estimator. The first goal of this section is to use this finite sample bound to show that this clustering algorithm estimates both ρ^* and the clusters A_i^* in a consistent manner. In a second step we then introduce a couple of additional assumptions on P that make it possible to establish convergence rates for both estimation problems.

Let us begin with the following result that establishes consistency.

Theorem 4.1. *Let Assumption A be satisfied, and let (ε_n) , (δ_n) , and (τ_n) be strictly positive sequences converging to zero such that $\psi(\delta_n) < \tau_n$ for all sufficiently large n and*

$$\frac{\ln \delta_n^{-1}}{n\delta_n^{2d}\varepsilon_n^2} \rightarrow 0. \quad (23)$$

For $n \geq 1$, consider Algorithm 3.1 with the input parameters ε_n , τ_n , and the family $(L_{D,\rho})_{\rho \geq 0}$ given by $L_{D,\rho} := \{h_{D,\delta_n} \geq \rho\}$. Then, for all $\epsilon > 0$, we have

$$\lim_{n \rightarrow \infty} P^n \left(\{D \in X^n : 0 < \rho_D^* - \rho^* \leq \epsilon\} \right) = 1$$

and, if $\mu(\overline{A_i^* \cup A_2^*} \setminus (A_1^* \cup A_2^*)) = 0$, we also have

$$\lim_{n \rightarrow \infty} P^n \left(\{D \in X^n : \mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) \leq \epsilon\} \right) = 1,$$

where, for $B_1(D)$ and $B_2(D)$, we use the numbering described in Theorem 3.7.

Before we discuss the consequences of Theorem 4.1, let us briefly illustrate the additional assumption $\mu(\overline{A_i^* \cup A_2^*} \setminus (A_1^* \cup A_2^*)) = 0$. To this end, we fix a μ -density h of P . Then Lemma 2.1 tells us that

$$A_i^* \cup A_2^* = \bigcup_{\rho > \rho^*} M_\rho \subset \bigcup_{\rho > \rho^*} \overline{\{h \geq \rho\}} \subset \overline{\bigcup_{\rho > \rho^*} \{h \geq \rho\}} = \overline{\{h > \rho^*\}}.$$

Using (5), which is ensured by the assumed normality in Assumption A, we then obtain

$$\mu(\overline{A_i^* \cup A_2^*} \setminus (A_1^* \cup A_2^*)) \leq \mu(\overline{\{h > \rho^*\}} \setminus \{h > \rho^*\}) \leq \mu(\partial\{h > \rho^*\}) = \mu(\partial\{h \leq \rho^*\}).$$

Consequently, the additional assumption is satisfied, if there exists a μ -density h of P such that $\mu(\partial\{h \leq \rho^*\}) = 0$. In this respect recall, that Lemma 2.4 showed that P is normal, if, for all $\rho \in \mathbb{R}$, we have a μ -density h of P with $\mu(\partial\{h \geq \rho\}) = 0$. In other words, the additional assumption is somewhat similar to the already assumed normality.

Theorem 4.1 shows that for suitably chosen parameters Algorithm 3.1 asymptotically recovers both ρ^* and the clusters A_1^* and A_2^* , whenever the distribution P has levels that are thicker than a pre-described order γ . In other words, as soon as we assume a minimal thickness, we are able to recover the clusters. To be more precise, suppose that we choose $\delta_n \sim n^{-\alpha}$ and $\varepsilon_n \sim n^{-\beta}$ for some $\alpha, \beta > 0$. Then it is easy to check that (23) is satisfied if and only if $2(\alpha d + \beta) < 1$. For

$\tau_n \sim n^{-\alpha\gamma} \ln n$, we then have $\psi(\delta_n) < \tau_n$ for all sufficiently large n , and therefore, Algorithm 3.1 recovers the clusters for all distributions P that have thick levels of order γ . Similarly, the choice $\tau_n \sim (\ln n)^{-1}$ leads to consistency for all distributions P that have thick levels of some order $\gamma > 0$. Finally note that (23) can be replaced by

$$\frac{\ln \delta_n^{-1}}{n \delta_n^d \varepsilon_n^2} \rightarrow 0,$$

if we restrict our consideration to distributions with bounded μ -densities. The proof of this is a straight forward modification of the proof of Theorem 4.1, and hence omitted.

Let us now give two simple though concrete examples. For the first one, recall that for the one-dimensional case in which $X \subset \mathbb{R}$ is a compact interval, we automatically have thickness of order $\gamma = 1$ with $\psi(\delta) = 3\delta$ for all $\delta \in (0, \delta_{\text{thick}}]$. Consequently, Algorithm 3.1 asymptotically recovers the clusters for all distributions P on X that can be topologically clustered, if, for example, we choose $\delta_n \sim n^{-\alpha}$, $\varepsilon_n \sim n^{-\beta}$, and $\tau_n = 4\delta_n$ for $\alpha, \beta > 0$ satisfying $2(\alpha + \beta) < 1$. Note that it is easy to construct distributions in this class that either do or do not have a continuous density. In the two dimensional case, as well as in any higher dimension, it is also straight forward to construct distributions with thick levels of order $\gamma = 1$ and *discontinuous* densities. The construction of a corresponding rich family of *continuous* densities is, however, significantly more complicated due to possible saddle points at the level ρ^* . For $d = 2$, we have thus included the construction of such a family in the appendix. Note that in this case, sequences $\delta_n \sim n^{-\alpha}$, $\varepsilon_n \sim n^{-\beta}$, and $\tau_n \sim n^{-\alpha} \ln n$ with $\alpha, \beta > 0$ and $4\alpha + 2\beta < 1$ guarantee consistency, and for bounded densities the latter can be weakened to $2\alpha + 2\beta < 1$.

Our next goal is to establish rates of convergence for both $\rho_D^* \rightarrow \rho^*$ and $\mu(B_i(D) \Delta A_i^*) \rightarrow 0$. As usual in nonparametric statistics, such rates require some assumptions on P . Let us begin by introducing an assumption that leads to rates for the estimation of ρ^* .

Definition 4.2. *Let (X, d) be a compact metric space, μ be a finite Borel measure on X , and P be a μ -absolutely continuous probability measure on X that can be clustered between the critical levels ρ^* and ρ^{**} . We say that the clusters of P have separation exponent $\kappa \in (0, \infty]$, if there exists a constant $\underline{c}_{\text{sep}} > 0$ such that, for all $\varepsilon \in (0, \rho^{**} - \rho^*]$, we have*

$$\tau^*(\varepsilon) \geq \underline{c}_{\text{sep}} \varepsilon^{1/\kappa}.$$

*Moreover, we say that the separation exponent κ is exact, if there exists another constant $\bar{c}_{\text{sep}} > 0$ such that, for all $\varepsilon \in (0, \rho^{**} - \rho^*]$, we have*

$$\tau^*(\varepsilon) \leq \bar{c}_{\text{sep}} \varepsilon^{1/\kappa}.$$

Roughly speaking, the separation exponent describes how fast the connected components of the level sets M_ρ move apart for increasing $\rho > \rho^*$. It is easy to check that the separation exponent is monotone in the sense that a distribution having separation exponent κ also has separation exponent κ' for all $\kappa' < \kappa$. In particular, the “best” separation exponent is $\kappa = \infty$ and this exponent describes distributions, for which we have $d(A_1^*, A_2^*) \geq \underline{c}_{\text{sep}}$, i.e. the clusters A_1^* and A_2^* do not touch each other.

Because of our horizontal uncertainty described by $M_{\rho+\varepsilon}^- \subset L_{D,\rho} \subset M_{\rho-\varepsilon}^+$ and our gluing with the help of τ -connected components, it seems quite natural that the separation exponent has an influence on how well ρ^* can be estimated by our algorithm. Before we present a finite-sample result in this direction, let us first provide a concrete class of distributions having a separation exponent.

Lemma 4.3. *Let $X \subset \mathbb{R}^d$ be a compact and convex subset, $\|\cdot\|$ some norm on \mathbb{R}^d , and P be a Lebesgue absolutely continuous distribution on X that can be clustered between the critical levels ρ^* and ρ^{**} . Assume that P has a continuous density h and that there exists constants $c > 0$ and $\theta \in (0, \infty)$ such that*

$$|h(x) - h(x')| \leq c \|x - x'\|^\theta \tag{24}$$

for all $x \in \{h \leq \rho^\}$, $\rho \in (\rho^*, \rho^{**}]$, and $x' \in \partial_X M_\rho$, where $\partial_X M_\rho$ denotes the boundary of M_ρ in X . Then the clusters of P have at least the separation exponent θ .*

Note that (24) holds, if the density h in Lemma 4.3 is actually θ -Hölder-continuous, and it is easy to see that the converse is, in general, not true. Moreover, using the inclusion $\partial_X M_\rho \subset \{h = \rho\}$ established in Lemma 2.1, it is easy to check that (24) is *equivalent* to

$$|h(x) - \rho| \leq c d(x, \partial_X M_\rho)^\theta \quad (25)$$

for all $x \in \{h \leq \rho^*\}$ and $\rho \in (\rho^*, \rho^{**}]$. A localized but two-sided version of this condition has been used by Singh et al. (2009) for a level set estimator that is adaptive with respect to the Hausdorff metric.

The exponent given in Lemma 4.3 may or may not be exact. To illustrate this, consider $X := -[3, 3]$ and, for $\theta, \beta \in (0, \infty]$ and $\rho^* \geq 0$, the symmetric density

$$h_{\theta, \beta}(x) := c_{\theta, \beta}(\rho^* + \mathbf{1}_{[0, 1]}(|x|)|x|^\theta + \mathbf{1}_{[1, 2]}(|x|) + \mathbf{1}_{[2, 3]}(|x|)(3 - |x|)^\beta), \quad x \in X \quad (26)$$

where $c_{\theta, \beta}$ is a constant ensuring that $h_{\theta, \beta}$ is a probability density with respect to the Lebesgue measure on X . Obviously, ρ^* is the first critical level and $h_{\theta, \beta}$ is $\min\{1, \theta, \beta\}$ -Hölder-continuous if $\theta < \infty$ and $\beta < \infty$. In addition, it is not hard to check, that $h_{\theta, \beta}$ always has exact separation exponent θ . The reason for this mismatch is that the separation exponent only describes the steepness of h in the “valley” between the “peaks” of the clusters, while the Hölder-continuity, or (24), describes the steepness globally.

To some extent, the distributions of the form (26) are archetypal for smooth densities on \mathbb{R} since they provide simple examples of arbitrary polynomial behaviour in the upper vicinity of the critical level ρ^* . In particular, for C^2 -densities h whose first derivative h' has exactly one zero x_0 in the set $\{h = \rho^*\}$ and whose second derivative satisfies $h''(x_0) > 0$, one can easily show with the help of Taylor’s theorem that their behaviour in the upper vicinity of the critical level ρ^* is, in terms of exponents, identical to that of (26) for $\theta = 2$ and $\beta = 1$. Moreover, larger values for θ can be achieved by assuming that higher derivatives of h vanish at x_0 . Last but not least note that the class of continuous densities on \mathbb{R}^2 that are considered in the appendix have separation exponent $\kappa = 2$, see Example 7.4 for details, and similarly to the 1-dimensional case, the construction can be modified to achieve other exponents, too.

Let us now return to our investigations on the speed of $\rho_D^* \rightarrow \rho^*$. The following theorem, which forms the base of the rates we will establish, provides finite-sample guarantees that bound the error $\rho_D^* - \rho^*$ from above and below.

Theorem 4.4. *Let Assumption A be satisfied and assume additionally that P has a bounded μ -density h and its clusters have separation exponent $\kappa \in (0, \infty]$. For some fixed $\delta \in (0, \delta_{\text{thick}}]$, $\varsigma \geq 1$, $n \geq 1$, and $\tau \geq 2\psi(\delta)$, we pick an $\varepsilon > 0$ satisfying (21), that is*

$$\varepsilon \geq \sqrt{\frac{2c_{\text{part}}(1 + \|h\|_\infty)(\varsigma + \ln(2c_{\text{part}}) - d \ln \delta)}{\delta^d n}} + \frac{2c_{\text{part}}(\varsigma + \ln(2c_{\text{part}}) - d \ln \delta)}{3\delta^d n}.$$

Let us assume that $\varepsilon^* := \varepsilon + (\tau/\underline{c}_{\text{sep}})^\kappa$ satisfies $\varepsilon^* \leq (\rho^{**} - \rho^*)/9$. Then, if Algorithm 3.1 receives the input parameters ε , τ , and the family $(L_{D, \rho})_{\rho \geq 0}$ given by $L_{D, \rho} := \{h_{D, \delta} \geq \rho\}$, the probability P^n of having a data set $D \in X^n$ satisfying

$$\varepsilon < \rho_D^* - \rho^* \quad (27)$$

$$\rho_D^* - \rho^* \leq (\tau/\underline{c}_{\text{sep}})^\kappa + 6\varepsilon, \quad (28)$$

is not less than $1 - e^{-\varsigma}$. Moreover, if the separation exponent κ is exact and $\kappa < \infty$, then we can replace (27) by

$$\frac{1}{4} \left(\frac{\tau}{6\bar{c}_{\text{sep}}} \right)^\kappa + \varepsilon < \rho_D^* - \rho^*. \quad (29)$$

The finite sample guarantees of Theorem 4.4 can be easily used to derive (exact) rates for $\rho_D^* \rightarrow \rho^*$. The following corollary illustrates this.

Corollary 4.5. *Let Assumption A be satisfied and assume that P has bounded μ -density and its clusters have separation exponent $\kappa \in (0, \infty)$. Furthermore, let (ε_n) , (δ_n) , and (τ_n) be sequences with*

$$\varepsilon_n \sim \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{\gamma \kappa}{2\gamma \kappa + d}}, \quad \delta_n \sim \left(\frac{\ln n}{n} \right)^{\frac{1}{2\gamma \kappa + d}}, \quad \text{and} \quad \tau_n \sim \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{\gamma}{2\gamma \kappa + d}},$$

and assume that, for $n \geq 1$, Algorithm 3.1 receives the input parameters ε_n , τ_n , and the family $(L_{D,\rho})_{\rho \geq 0}$ given by $L_{D,\rho} := \{h_{D,\delta_n} \geq \rho\}$. Then there exists a constant $\bar{K} \geq 1$ such that for all sufficiently large n we have

$$P^n \left(\left\{ D \in X^n : \rho_D^* - \rho^* \leq \bar{K} \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{\gamma \kappa}{2\gamma \kappa + d}} \right\} \right) \geq 1 - \frac{1}{n}. \quad (30)$$

Moreover, if the separation exponent κ is exact, there exists another constant $\underline{K} \geq 1$ such that for all sufficiently large n we have

$$P^n \left(\left\{ D \in X^n : \underline{K} \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{\gamma \kappa}{2\gamma \kappa + d}} \leq \rho_D^* - \rho^* \leq \bar{K} \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{\gamma \kappa}{2\gamma \kappa + d}} \right\} \right) \geq 1 - \frac{1}{n}. \quad (31)$$

Finally, if $\kappa = \infty$, then, for

$$\varepsilon_n \sim \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{1}{2}}, \quad \delta_n \sim (\ln \ln n)^{-\frac{1}{2d}}, \quad \text{and} \quad \tau_n \sim (\ln \ln n)^{-\frac{\gamma}{3d}}$$

the estimates (30) and (31) hold for all $n \geq 3$, respectively all sufficiently large n .

Note that Theorem 4.4 also makes it possible to consider alternative choices of (ε_n) , (δ_n) , and (τ_n) , which lead to other rates of convergence for $\rho_D^* \rightarrow \rho^*$. The choice in Corollary 4.5 ensures that modulo the (double) logarithmic factors the rates are the best ones we can derive from Theorem 4.4. So far, we don't know whether these rates are (essentially) minmax optimal, or at least (essentially) optimal for our algorithm, but we conjecture that they are, since both the statistical analysis from Theorem 3.3 as well as the subsequent analysis of Algorithm 3.1 do not seem to allow much improvement. However, a detailed analysis of this question is clearly out of the scope of this work.

To illustrate the rates above, let us recall that for the one-dimensional distributions of the form (26) we can set $\gamma = 1$ and $\kappa = \theta$, so that the exponent in the rates becomes $\frac{\theta}{2\theta+1}$. In particular, for the C^2 -case discussed there, we had $\theta = 2$ and thus we get a rate with exponent $2/5$, while for $\theta \rightarrow \infty$ the exponent converges to $1/2$. Similarly, for the two-dimensional distributions considered in the appendix, we have $\gamma = 1$, $\kappa = 2$, and $d = 2$, and hence the exponent in the rate becomes $1/3$.

Our next goal is to establish rates for $\mu(B_i(D) \triangle A_i^*) \rightarrow 0$. Since this is a modified level set estimation problem, let us recall some assumptions on P , which have been used in the context of level set estimation. The first assumption in this direction is the following flatness-assumption, which in its two-sided variant goes back to Polonik (1995).

Definition 4.6. Let μ be a finite measure on X and P be a distribution on X that has a μ -density h . For a given level $\rho \geq 0$, we say that P has flatness exponent $\vartheta \in (0, \infty]$, if there exists a constant $c_{\text{flat}} > 0$ such that

$$\mu(\{0 < h - \rho < s\}) \leq (c_{\text{flat}} s)^{\vartheta}, \quad s > 0. \quad (32)$$

Obviously, (32) is independent of the actual choice of h . Moreover, note the larger ϑ is, the less h is concentrated in the upper vicinity of the level ρ , and hence the steeper h must behave above the level ρ . In particular, for $\vartheta = \infty$, the density h is allowed to take the value ρ but is otherwise bounded away from ρ .

To provide an illustrative example on how the steepness of h influences the flatness exponent, consider the density in (26). Some simple calculations then show that the corresponding distribution has flatness exponent $\vartheta = \min\{1/\theta, 1/\beta\}$ if $\theta < \infty$ and $\beta < \infty$ and flatness exponent $\vartheta = \infty$ if $\theta = \beta = \infty$. Finally, for the two-dimensional examples considered in the appendix, the flatness exponent is not fully determined because of the variability in this class of distributions, but some simple calculations show that we always have $\vartheta \in [0, 1]$.

Our second assumption describes in some sense the roughness of the boundary of the clusters.

Definition 4.7. Let Assumption C be satisfied. Given some $\alpha \in (0, 1]$, we say that the clusters have an α -smooth boundary, if there exists a constant $c_{\text{bound}} > 0$ such that, for all $\rho \in (\rho^*, \rho^{**}]$, $\delta \in (0, \delta_{\text{thick}}]$, and $i = 1, 2$, we have

$$\mu((A_\rho^i)^{+\delta} \setminus (A_\rho^i)^{-\delta}) \leq c_{\text{bound}} \delta^\alpha, \quad (33)$$

where A_ρ^1 and A_ρ^2 denote the two connected components of the level set M_ρ .

In the Euclidean case, the following lemma shows that assuming $\alpha > 1$ does not make sense. In addition it shows that, for each level set with rectifiable boundaries in the sense of (Federer, 1969, 3.2.14), the bound (33) holds with $\alpha = 1$. Therefore, the α -smoothness of the boundary enforces a uniform version of this.

Lemma 4.8. *Let λ^d be the d -dimensional Lebesgue measure, \mathcal{H}^{d-1} the $(d-1)$ -dimensional Hausdorff measure on \mathbb{R}^d , and σ_d be the volume of the d -dimensional unit Euclidean ball in \mathbb{R}^d . Then, for every non-empty, bounded, and measurable subset $A \subset \mathbb{R}^d$ the following statements hold:*

i) *There exists a $\delta_A > 0$, such that for $\underline{c}_A := d\sigma_d^{1/d}\lambda^d(\overline{A})^{1-1/d}/2$ and all $\delta \in (0, \delta_A]$, we have*

$$\lambda^d(A^{+\delta} \setminus A^{-\delta}) \geq \underline{c}_A \cdot \delta.$$

ii) *If ∂A is $(d-1)$ -rectifiable and $\mathcal{H}^{d-1}(\partial A) > 0$, then there exists a $\delta_A > 0$, such that, for all $\delta \in (0, \delta_A]$, we have*

$$\lambda^d(A^{+\delta} \setminus A^{-\delta}) \leq 4\mathcal{H}^{d-1}(\partial A) \cdot \delta.$$

Before we proceed, let us finally mention that the clusters of the distributions in Example (26) have an α -smooth boundary with $\alpha = 1$ and $c_{\text{bound}} = 4$. Furthermore, the two-dimensional distributions discussed in the appendix also have smoothness boundary $\alpha = 1$, see Example 7.5 for details.

The following simple lemma shows that a bound of the form (33) together with a regular behavior of h around the level of interest ensures a non-trivial flatness exponent.

Lemma 4.9. *Let (X, d) be a complete, separable metric space, μ be a finite Borel measure on X with $\text{supp } \mu = X$, and P be a μ -absolutely continuous distribution on X . Furthermore, let $\rho \geq 0$ be a level and h be a μ -density of P for which exist constants $c > 0$, $\alpha \in (0, 1]$, $\delta_0 > 0$, and $\theta \in (0, \infty)$ such that*

$$\mu(M_\rho^{+\delta} \setminus M_\rho^{-\delta}) \leq c\delta^\alpha \quad (34)$$

for all $\delta \in (0, \delta_0]$ and

$$d(x, \partial M_\rho)^\theta \leq c|h(x) - \rho| \quad (35)$$

for all $x \in \{h > \rho\}$. Then P has flatness exponent α/θ at level ρ .

Recall that a two-sided version of condition (35) has been used in the above mentioned paper by Singh et al. (2009). We will come back to this after we have established our rates.

The following assumption collects all conditions we need to impose on P to get rates for estimating the clusters.

Assumption R. *Assumption A is satisfied and P has a bounded μ -density h . Moreover, P has flatness exponent $\vartheta \in (0, \infty]$ at level ρ^* , its clusters have an α -smooth boundary for some $\alpha \in (0, 1]$, and its clusters have separation exponent $\kappa \in (0, \infty]$.*

With these preparations we can now investigate how well our algorithm estimates the clusters A_1^* and A_2^* . As usual, we begin with a finite-sample estimate, which will then lead to rates of convergence.

Theorem 4.10. *Let Assumption R be satisfied and assume that $\delta, \varepsilon, \tau, \varepsilon^*, \varsigma, n$, and $(L_{D, \rho})_{\rho \geq 0}$ are as in Theorem 4.4. Then the probability P^n of having a data set $D \in X^n$ satisfying (27), (28), and*

$$\mu(B_1(D) \Delta A_1^*) + \mu(B_2(D) \Delta A_2^*) \leq 6c_{\text{bound}}\delta^\alpha + (c_{\text{flat}}(\tau/\underline{c}_{\text{sep}})^\kappa + 7c_{\text{flat}}\varepsilon)^\vartheta \quad (36)$$

is not less than $1 - e^{-\varsigma}$, where the sets $B_1(D)$ and $B_2(D)$ are ordered according to (18). Moreover, if the separation exponent κ is exact and satisfies $\kappa < \infty$, then (29) also holds for these data sets D .

Note that, for finite values of ϑ and κ , the right-hand side of (36) behaves like $\delta^\alpha + \tau^{\vartheta\kappa} + \varepsilon^\vartheta$, and in this case it is thus easy to derive the best convergence rates our analysis yields. The following corollary presents corresponding results and also provides rates for the cases $\vartheta = \infty$ or $\kappa = \infty$.

Corollary 4.11. *Assume that Assumption R be is satisfied and write $\varrho := \min\{\alpha, \vartheta\gamma\kappa\}$. Furthermore, let (ε_n) , (δ_n) , and (τ_n) be sequences with*

$$\varepsilon_n \sim \left(\frac{\ln n}{n}\right)^{\frac{\varrho}{2e^{+\vartheta d}}} (\ln \ln n)^{-\frac{\vartheta d}{8e^{+4\vartheta d}}}, \quad \delta_n \sim \left(\frac{\ln n \cdot \ln \ln n}{n}\right)^{\frac{\varrho}{2e^{+\vartheta d}}}, \quad \text{and} \quad \tau_n \sim \left(\frac{\ln n \cdot (\ln \ln n)^2}{n}\right)^{\frac{\vartheta\gamma}{2e^{+\vartheta d}}}.$$

Assume that, for $n \geq 1$, Algorithm 3.1 receives the input parameters ε_n , τ_n , and the family $(L_{D,\rho})_{\rho \geq 0}$ given by $L_{D,\rho} := \{h_{D,\delta_n} \geq \rho\}$. Then there exists a constant $\bar{K} \geq 1$ such that for all $n \geq 1$ we have

$$P^n \left(\left\{ D \in X^n : \mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) \leq K \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{\vartheta\varrho}{2e^{+\vartheta d}}} \right\} \right) \geq 1 - \frac{1}{n},$$

where the sets $B_1(D)$ and $B_2(D)$ are ordered according to (18).

Let us now compare the established rates for estimating the level ρ^* and the clusters. To this end, we restrict ourselves to the most important case $\alpha = 1$. Then, if $\vartheta\gamma\kappa \leq 1$, we obtain $\varrho = \vartheta\gamma\kappa$ in Corollary 4.11, and the exponent in the asymptotic behavior of the optimal (δ_n) becomes $\frac{1}{2\gamma\kappa+d}$. Since this equals the exponent in Corollary 4.5, and, modulo the extra $\ln \ln n$ terms, we also have the same behavior for (ε_n) and (τ_n) in both corollaries, we conclude that we obtain the rates in Corollaries 4.5 and 4.11 with (essentially) the *same* controlling sequences (ε_n) , (δ_n) , and (τ_n) of Algorithm 3.1. In the case $\vartheta\gamma\kappa \leq 1$ we can thus achieve the best rates for estimating ρ^* and the clusters *simultaneously*. Unfortunately, this changes if $\vartheta\gamma\kappa > 1$. Indeed, while the exponent for (δ_n) in Corollary 4.5 remains the same, it changes from $\frac{1}{2\gamma\kappa+d}$ to $\frac{\vartheta}{2+\vartheta d}$ in Corollary 4.11, and a similar effect takes place for the sequences (ε_n) and (τ_n) . Roughly speaking, the reason for this difference is that in the case $\vartheta\gamma\kappa > 1$ the estimation of ρ^* is easier than the estimation of the level set M_{ρ^*} , and since for estimating the clusters we need to do both, the level set estimation rate determines the rate for estimating the clusters.

To illustrate the difference between the estimation of ρ^* and the clusters in more detail, let us consider the case $\theta = \beta = \infty$ in the toy model (26), that is $\kappa = \infty$. Then the clusters are stumps and the corresponding levels sets M_ρ do not change between the levels ρ^* and ρ^{**} . Intuitively, the best choice for estimating ρ^* are then sufficiently small but fixed values for δ_n and τ_n , so that ε_n converges to 0 as fast as possible. In Corollary 4.5 this is mimicked by choosing very slowly decaying sequences (δ_n) and (τ_n) . Let us now consider the estimation of the clusters. Since the connected components of M_ρ do not change for $\rho \in (\rho^*, \rho^{**}]$, it clearly suffices to find an arbitrary $\rho \in (\rho^*, \rho^{**}]$ and to estimate the connected components for this ρ . The best way to achieve this is to use a sufficiently small but fixed value for ε_n and sequences (δ_n) and (τ_n) that converge to zero as fast as possible. In Corollary 4.11 this is mimicked by choosing a very slowly decaying sequence (ε_n) and fast decaying sequences (δ_n) and (τ_n) .

Interestingly, the case $\vartheta\gamma\kappa > 1$ seems to be rather rare for one-dimensional data. To illustrate this, let us have another look at our toy model (26). We have already seen that $\gamma = 1$, $\kappa = \theta$ and $\vartheta = \min\{1/\theta, 1/\beta\}$. Consequently, if $\theta < \beta \leq \infty$, we find $\vartheta\gamma\kappa = \theta/\beta < 1$, while $\beta \leq \theta < \infty$ implies $\vartheta\gamma\kappa = \theta/\theta = 1$. In other words, the only case, in which we have $\vartheta\gamma\kappa > 1$, is the one with $\theta = \infty$, that is, in the case in which the two clusters do not touch each other. Finally note that in higher dimensions, the situation becomes more interesting. Indeed, for the two-dimensional distributions considered in the appendix, both cases $\vartheta\gamma\kappa > \alpha$ and $\vartheta\gamma\kappa \leq \alpha$ are possible depending on the value of the flatness exponent ϑ , and it is straightforward to modify these distributions so that other values for γ and κ occur, too. On the other hand, the toy model (26) can be easily made two-dimensional by letting the density to be constant in the second dimension. It is easy to check, that in this case the phenomena of the one-dimensional model, which we observed above, are preserved.

As for estimating the critical level ρ^* , we do not know so far, whether our rates for estimating the clusters are minmax optimal. Again, our conjecture is that they are optimal modulo the logarithmic terms. To motivate our conjecture, let us consider the case $\alpha = \gamma = 1$. Moreover, assume that two-sided versions of (25) and (35) hold for all $\rho \in (\rho^*, \rho^{**}]$, respectively $\rho = \rho^*$. Then we have $\kappa = \theta$ and $\vartheta = 1/\theta$ by Lemmas 4.3 and 4.9, and thus we find $\varrho = 1$. Consequently, the rates in Corollary 4.11 have the exponent $\frac{1}{2\theta+d}$. This is exactly the same exponent as the one obtained by Singh et al. (2009) for minmax optimal and adaptive Hausdorff estimation of a fixed level set. In addition, it seems that their lower bound, which is based on Tsybakov (1997), is, modulo logarithmic factors, the

same for assessing the estimator in the way we have done it in Corollary 4.11. While this coincidence indicates that our rates may be (essentially) optimal, it is, of course, not a rigorous argument. A detailed analysis is, however, out of the scope of this paper. Another interesting question, which is also out of the scope, is, whether the estimates $B_i(D)$ approximate the true clusters A_i^* in the Hausdorff metric, too, and if so, whether we can achieve the rates reported by Singh et al. (2009).

5 Data-Dependent Parameter Selection

In the last section we derived rates of convergence for both the estimation of ρ^* and the clusters. In both cases, the best rates we could achieve required parameters ε_n , δ_n , and τ_n that do depend on some properties of the unknown distribution P , namely κ , ϑ , and strictly speaking, also α . Of course, these parameters are not available to us in practice, and therefore the obtained rates are of little practical value. The goal of this final section is to address this issue by proposing a simple data-dependent parameter selection strategy, that is able to recover the rates of Corollary 4.5 without the above mentioned knowledge about P . We further show that this parameter selection strategy also recovers the rates of Corollary 4.11 in the case of $\vartheta\gamma\kappa \leq \alpha$, i.e. in the case in which the estimation of the level is harder than the estimation of the level set.

Let us begin by presenting the parameter selection strategy. To this end, let $\Delta \subset (0, 1]$ be a finite set and $n \geq 1$, $\varsigma \geq 1$. For $\delta \in \Delta$, we fix some $\tau_{\delta,n} > 0$ and define

$$\varepsilon_{\delta,n} := C \sqrt{\frac{c_{\text{part}}(\varsigma + \ln(2c_{\text{part}}|\Delta|) - d \ln \delta) \ln \ln n}{\delta^{d_n}}} + \frac{2c_{\text{part}}(\varsigma + \ln(2c_{\text{part}}|\Delta|) - d \ln \delta)}{3\delta^{d_n}}, \quad (37)$$

where $C \geq 1$ is some user-specified constant. Now assume that, for each $\delta \in \Delta$, we run Algorithm 3.1 with the parameters $\varepsilon_{\delta,n}$ and $\tau_{\delta,n}$, and the family $(L_{D,\rho})_{\rho \geq 0}$ given by $L_{D,\rho} := \{h_{D,\delta} \geq \rho\}$.

We write $\rho_{D,\delta}^*$ for the corresponding level returned by Algorithm 3.1. Let us consider a width $\delta_{D,\Delta}^* \in \Delta$ that achieves the smallest returned level, that is

$$\delta_{D,\Delta}^* \in \arg \min_{\delta \in \Delta} \rho_{D,\delta}^*. \quad (38)$$

Note that in general, this width may not be uniquely determined, so that in the following we need to additionally assume that we have a well-defined choice, e.g. the smallest $\delta \in \Delta$ satisfying (38), whenever there is some ambiguity. Moreover, we write

$$\rho_{D,\Delta}^* := \rho_{D,\delta_{D,\Delta}^*}^* = \min_{\delta \in \Delta} \rho_{D,\delta}^* \quad (39)$$

for the smallest returned level. Note that unlike $\delta_{D,\Delta}^*$, the level $\rho_{D,\Delta}^*$ is always unique. Finally, we define $\varepsilon_{D,\Delta} := \varepsilon_{\delta_{D,\Delta}^*,n}$ and $\tau_{D,\Delta} := \tau_{\delta_{D,\Delta}^*,n}$.

Our first goal is to show that $\rho_{D,\Delta}^*$ achieves the rates of Corollary 4.5 provided that Δ and $\tau_{\delta,n}$ are suitably chosen. We begin with a finite sample guarantee.

Theorem 5.1. *Assume that Assumption A is satisfied, that P has a bounded μ -density h , and that the two clusters of P have separation exponent $\kappa \in (0, \infty]$. For a fixed finite subset $\Delta \subset (0, \delta_{\text{thick}}]$, and $n \geq 1$, $\varsigma \geq 1$, and $C \geq 1$, we define $\varepsilon_{\delta,n}$ by (37) and assume that we have chosen $\tau_{\delta,n}$ such that $\tau_{\delta,n} \geq 2\psi(\delta)$ for all $\delta \in \Delta$. Furthermore, assume that $C^2 \ln \ln n \geq 2(1 + \|h\|_\infty)$ and $\varepsilon_\delta^* := \varepsilon_{\delta,n} + (\tau_{\delta,n}/\underline{c}_{\text{sep}})^\kappa \leq (\rho^{**} - \rho^*)/9$ for all $\delta \in \Delta$. Then we have*

$$P^n \left(\left\{ D \in X^n : \varepsilon_{D,\Delta} < \rho_{D,\Delta}^* - \rho^* \leq \min_{\delta \in \Delta} \left((\tau_{\delta,n}/\underline{c}_{\text{sep}})^\kappa + 6\varepsilon_{\delta,n} \right) \right\} \right) \geq 1 - e^{-\varsigma}.$$

Moreover, if the separation exponent κ is exact and $\kappa < \infty$, then the assumptions above actually guarantee

$$P^n \left(\left\{ D \in X^n : \min_{\delta \in \Delta} (c_1 \tau_{\delta,n}^\kappa + \varepsilon_{\delta,n}) < \rho_{D,\Delta}^* - \rho^* \leq \min_{\delta \in \Delta} (c_2 \tau_{\delta,n}^\kappa + 6\varepsilon_{\delta,n}) \right\} \right) \geq 1 - e^{-\varsigma},$$

where $c_1 := \frac{1}{4}(6\bar{c}_{\text{sep}})^{-\kappa}$ and $c_2 := \bar{c}_{\text{sep}}^{-\kappa}$, and similarly

$$P^n \left(\left\{ D \in X^n : c_1 \tau_{D,\Delta}^\kappa + \varepsilon_{D,\Delta} < \rho_{D,\Delta}^* - \rho^* \leq c_2 \tau_{D,\Delta}^\kappa + 6\varepsilon_{D,\Delta} \right\} \right) \geq 1 - e^{-\varsigma}.$$

Roughly speaking, Theorem 5.1 establishes the same finite sample guarantees for the estimator $\rho_{D,\Delta}^*$ as Theorem 4.4 did for the simpler estimator ρ_D^* . Therefore, it is not surprising that for suitable choices of Δ , the rates of Corollary 4.5 can be recovered, too. The following corollary shows that this can actually be achieved for candidate sets Δ that are completely independent of P .

Corollary 5.2. *Assume that Assumption A is satisfied, that P has a bounded μ -density h , and that the two clusters of P have separation exponent $\kappa \in (0, \infty]$. For $n \geq 16$, we consider the interval*

$$I_n := \left[\left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{1}{d}}, \left(\frac{1}{\ln \ln n} \right)^{\frac{1}{d}} \right]$$

and fix some $n^{-1/d}$ -net $\Delta_n \subset I_n$ of I_n with $|\Delta_n| \leq n$. Furthermore, for some fixed $C \geq 1$, we define $\tau_{\delta,n} := \delta^\gamma \ln \ln \ln n$ and

$$\varepsilon_{\delta,n} := C \sqrt{\frac{c_{\text{part}} (\ln(2c_{\text{part}}|\Delta_n|n) - d \ln \delta) \ln \ln n}{\delta^{d_n}}} + \frac{2c_{\text{part}} (\ln(2c_{\text{part}}|\Delta_n|n) - d \ln \delta)}{3\delta^{d_n}},$$

for all $\delta \in \Delta_n$ and all $n \geq 16$. Then there exists a constant \bar{K} such that, for all sufficiently large n , we have

$$P^n \left(\left\{ D \in X^n : \varepsilon_{D,\Delta_n} < \rho_{D,\Delta_n}^* - \rho^* \leq \bar{K} \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{\gamma\kappa}{2\gamma\kappa+d}} \right\} \right) \geq 1 - \frac{1}{n}. \quad (40)$$

Moreover, if the separation exponent κ is exact and $\kappa < \infty$, then the assumptions above actually guarantee the existence of another constant \underline{K} such that for all sufficiently large n we have

$$P^n \left(\left\{ D \in X^n : \underline{K} \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{\gamma\kappa}{2\gamma\kappa+d}} \leq \rho_{D,\Delta_n}^* - \rho^* \leq \bar{K} \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{\gamma\kappa}{2\gamma\kappa+d}} \right\} \right) \geq 1 - \frac{1}{n}. \quad (41)$$

Finally, we show that the parameter selection strategy (38) in the set-up of Corollary 5.2 also partially recover the rates for estimating the clusters A_i^* obtained in Corollary 4.11.

Corollary 5.3. *Assume that Assumption R be is satisfied with $\alpha \geq \vartheta\gamma\kappa$ and that separation exponent κ is exact. Then, for the procedure considered in Lemma 5.2, there exists a constant $K \geq 1$ such that, for all sufficiently large n , we have*

$$P^n \left(\left\{ D \in X^n : \mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) \leq K \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{\vartheta\gamma\kappa}{2\gamma\kappa+\vartheta d}} \right\} \right) \geq 1 - \frac{1}{n},$$

where the sets $B_1(D)$ and $B_2(D)$ are ordered according to (18).

Unfortunately, the simple parameter selection strategy (38) is not adaptive in the case $\alpha < \vartheta\gamma\kappa$, i.e. in the case in which the estimation of ρ^* is easier than the estimation of the corresponding clusters. It is unclear to us, whether in this case a two-stage procedure that first estimates ρ^* by ρ_{D,Δ_n}^* as above, and then uses a different strategy to estimate the connected components at the level ρ_{D,Δ_n}^* can be made adaptive.

6 Proofs

6.1 Proofs Related to the Definition of Level Sets

Proof of Lemma 2.1: By definition, M_ρ is the smallest closed set A satisfying $\mu(\{h \geq \rho\} \setminus A) = 0$. Moreover, we obviously have

$$\mu(\{h \geq \rho\} \setminus \overline{\{h \geq \rho\}}) = 0,$$

and hence we obtain $M_\rho \subset \overline{\{h \geq \rho\}}$. To show the other inclusion, we fix an $x \in \overset{\circ}{\{h \geq \rho\}}$ and an open set $U \subset X$ with $x \in U$. Then $\{h \geq \rho\} \cap U$ is open and non-empty, and hence $\text{supp } \mu = X$ yields

$$\mu_\rho(U) = \mu(\{h \geq \rho\} \cap U) \geq \mu(\overset{\circ}{\{h \geq \rho\}} \cap U) > 0.$$

By (2) we conclude that $x \in M_\rho$, that is, we have shown $\{h \overset{\circ}{\geq} \rho\} \subset M_\rho$.

Let us now assume that h is continuous. Clearly, we have $\{h > \rho\} \subset \{h \geq \rho\}$ and since $\{h > \rho\}$ is open, we conclude that $\{h > \rho\} \subset \overline{\{h \geq \rho\}} \subset M_\rho$ by the previously shown inclusion. Moreover, since $\{h \geq \rho\}$ is closed, we find $M_\rho \subset \overline{\{h \geq \rho\}} = \{h \geq \rho\}$. Recalling that M_ρ is closed by definition, we further find $\partial M_\rho \subset M_\rho \subset \{h \geq \rho\}$, and thus it remains to show $\partial M_\rho \subset \{h \leq \rho\}$. Let us assume the converse, that is, that there exists an $x \in \partial M_\rho$ such that $h(x) > \rho$. By the continuity we then find an open neighbourhood U of x such that $U \subset \{h > \rho\}$. Since $x \in \partial M_\rho$, we further find an $y \in U \setminus M_\rho$, while our construction together with the previously shown $\{h > \rho\} \subset M_\rho$ yields the contradicting statement $U \setminus M_\rho \subset \{h > \rho\} \setminus M_\rho = \emptyset$. \square

Proof of Lemma 2.2: We fix an $x \in M_{\rho_2}$ and an open set $U \subset X$ with $x \in U$. Moreover, we fix a μ -density h of P . Then we obtain

$$\mu_{\rho_1}(U) = \mu(\{h \geq \rho_1\} \cap U) \geq \mu(\{h \geq \rho_2\} \cap U) = \mu_{\rho_2}(U) > 0,$$

and hence we obtain $x \in M_{\rho_1}$ by (2). \square

Proof of Lemma 2.4: *i).* Let h be an upper semi-continuous μ -density of P . Then $\{h \geq \rho\}$ is closed, and hence Lemma 2.1 shows $M_\rho \subset \overline{\{h \geq \rho\}} = \{h \geq \rho\}$. Thus, P is upper normal at level ρ .

ii). Let h be a lower semi-continuous μ -density of P . By Lemma 2.1 we then know $\{h > \rho\} = \{h \overset{\circ}{>} \rho\} \subset \{h \overset{\circ}{\geq} \rho\} \subset \overset{\circ}{M}_\rho$. This yields the assertion.

iii). The upper normality follows from (3). To see that P is lower normal, we use the inclusion $\{h > \rho\} \setminus \overset{\circ}{M}_\rho \subset \overline{\{h \geq \rho\}} \setminus \{h \overset{\circ}{\geq} \rho\} = \partial\{h \geq \rho\}$ which follows from Lemma 2.1. \square

6.2 Proofs Related to Basic Properties of Connected Components

Proof of Lemma 2.6: *ii) \Rightarrow i).* Trivial.

i) \Rightarrow ii). For $A' \in \mathcal{P}(A)$ we find a $B' \in \mathcal{P}(B)$ such that $A' \subset B'$. Defining $\zeta(A') := B'$ then gives the desired Property (7).

Finally, assume that *ii)* is true. To show that ζ is unique, assume the converse. Then there exist $A' \in \mathcal{P}(A)$ and $B', B'' \in \mathcal{P}(B)$ with $B' \neq B''$ and both $A' \subset B'$ and $A' \subset B''$. Since $A' \neq \emptyset$, this yields $B' \cap B'' \neq \emptyset$, which in turn implies $B' = B''$ as $\mathcal{P}(B)$ is a partition, i.e. we have found a contradiction. \square

Proof of Lemma 2.8: Clearly, $\zeta := \zeta_{B,C} \circ \zeta_{A,B}$ maps from $\mathcal{P}(A)$ to $\mathcal{P}(C)$. Moreover, for $A' \in \mathcal{P}(A)$ we have $A' \subset \zeta_{A,B}(A')$ and for $B' := \zeta_{A,B}(A') \in \mathcal{P}(B)$ we have $B' \subset \zeta_{B,C}(B')$. Combining these inclusions we find

$$A' \subset \zeta_{A,B}(A') \subset \zeta_{B,C}(\zeta_{A,B}(A')) = \zeta_{B,C} \circ \zeta_{A,B}(A') = \zeta(A')$$

for all $A' \in \mathcal{P}(A)$. Consequently, $\mathcal{P}(A)$ is comparable to $\mathcal{P}(C)$ and by Lemma 2.6 we see that ζ is the CRM $\zeta_{A,C}$, that is $\zeta_{A,C} = \zeta = \zeta_{B,C} \circ \zeta_{A,B}$. \square

Proof of Lemma 2.9: Let us fix an $A' \in \mathcal{C}(A)$. Since $A \subset B$ and $|\mathcal{C}(B)| < \infty$ there then exist an $m \geq 1$ and mutually distinct $B_1, \dots, B_m \in \mathcal{C}(B)$ with $A' \subset B_1 \cup \dots \cup B_m$ and $A' \cap B_i \neq \emptyset$ for all $i = 1, \dots, m$. Since A and B are closed, A' and the sets $A' \cap B_i$ are also closed. Consequently, the sets $A' \cap B_i$ are also closed in A' with respect to the relative topology of A' . Let us now assume that $m > 1$. Then $A' \cap B_1$ and $(A' \cap B_2) \cup \dots \cup (A' \cap B_m)$ are two disjoint relatively closed non-empty subsets of A' whose union equals A' . Consequently A' is not connected, which contradicts $A' \in \mathcal{C}(A)$. In other words, we have $m = 1$, that is, $\mathcal{C}(A)$ is comparable to $\mathcal{C}(B)$. \square

Proof of Lemma 2.11: Let $A' \neq A''$ be two τ -connected components of A . Then we have $d(x', x'') \geq \tau$ for all $x' \in A'$ and $x'' \in A''$, since otherwise x' and x'' would be τ -connected in A . Consequently, we have $d(A', A'') \geq \tau$, and from the latter and the compactness of X , it is straightforward to conclude that $|\mathcal{C}_\tau(A)| < \infty$. Finally, let $(x_i) \subset A'$ be a sequence in some component $A' \in \mathcal{C}_\tau(A)$ such that $x_i \rightarrow x$ for some $x \in X$. Since A is closed, we have $x \in A$, and hence $x \in A''$ for some $A'' \in \mathcal{C}_\tau(A)$. By construction we find $d(A', A'') = 0$, and hence we obtain $A' = A''$ by the assertion that has been shown first. \square

Lemma 6.1. *Let (X, d) be a metric space, $A \subset X$ be a non-empty subset and $\tau > 0$. Then the following statements are equivalent:*

- i) A is τ -connected.
- ii) For all non-empty subsets A^+ and A^- of A with $A^+ \cup A^- = A$ and $A^+ \cap A^- = \emptyset$ we have $d(A^+, A^-) < \tau$.

Proof of Lemma 6.1: *i) \Rightarrow ii).* Let us fix two non-empty subsets A^+ and A^- of A with $A^+ \cup A^- = A$ and $A^+ \cap A^- = \emptyset$. Let us further fix two points $x^+ \in A^+$ and $x^- \in A^-$. Since A is τ -connected, there then exist $x_1, \dots, x_n \in A$ such that $x_1 = x^-$, $x_n = x^+$ and $d(x_i, x_{i+1}) < \tau$ for all $i = 1, \dots, n-1$. Then, $x^+ \in A^+$ and $x^- \in A^-$ imply the existence of an $i \in \{1, \dots, n-1\}$ with $x_i \in A^-$ and $x_{i+1} \in A^+$. This yields $d(A^+, A^-) \leq d(x_i, x_{i+1}) < \tau$.

ii) \Rightarrow i). Assume that A is not τ -connected, that is $|\mathcal{C}_\tau(A)| > 1$. We pick an $A^+ \in \mathcal{C}_\tau(A)$ and write $A^- := A \setminus A^+$. Since $|\mathcal{C}_\tau(A)| > 1$, both sets are non-empty, and our construction ensures that they are also disjoint and satisfy $A^+ \cup A^- = A$. Moreover, for every $A' \in \mathcal{C}_\tau(A)$ with $A' \neq A^+$ we know $d(A^+, A') \geq \tau$ by Lemma 2.11 and since A^- is the union of such A' , we conclude $d(A^+, A^-) \geq \tau$. \square

Corollary 6.2. *Let (X, d) be a metric space, $A \subset B \subset X$ be non-empty subsets and $\tau > 0$. If A is τ -connected, then there exists exactly one τ -connected component B' of B with $A \cap B' \neq \emptyset$. Moreover, B' is the only τ -connected component B'' of B that satisfies $A \subset B''$.*

Proof of Corollary 6.2: The second assertion is a direct consequence of the first, and hence it suffice to show the first assertion. Let us assume the first is not true. Since $A \subset B$ there then exist $B', B'' \in \mathcal{C}_\tau(B)$ with $B' \neq B''$, $A \cap B' \neq \emptyset$, and $A \cap B'' \neq \emptyset$. We write $A^- := A \cap B'$ and $A^+ := A \cap (B \setminus B')$. Since $B'' \subset B \setminus B'$, we obtain $A^+ \neq \emptyset$, and therefore, Lemma 6.1 shows $d(A^-, A^+) < \tau$. Consequently, there exist $x^- \in A^-$ and $x^+ \in A^+$ with $d(x^+, x^-) < \tau$. Now we obviously have $x^- \in B'$, and by construction, we also find a $B''' \in \mathcal{C}_\tau(B)$ with $x^+ \in B'''$. Our previous inequality then yields $d(B', B''') < \tau$, while Lemma 2.11 shows $d(B', B''') \geq \tau$, that is, we have found a contradiction. \square

Proof of Lemma 2.12: For $A' \in \mathcal{C}_\tau(A)$, Corollary 6.2 shows that there exists exactly $B' \in \mathcal{C}_\tau(B)$ with $A' \subset B'$. Consequently, $\mathcal{C}_\tau(A)$ is comparable to $\mathcal{C}_\tau(B)$. \square

Lemma 6.3. *Let (X, d) be a metric space, $A \subset X$ be a non-empty subset and $\tau > 0$. Then, for a partition A_1, \dots, A_m of A , the following statements are equivalent:*

- i) $\mathcal{C}_\tau(A) = \{A_1, \dots, A_m\}$.
- ii) For all $i = 1, \dots, m$, the set A_i is τ -connected and $d(A_i, A_j) \geq \tau$ for all $i \neq j$.

Proof of Lemma 6.3: *i) \Rightarrow ii).* Follows from Lemma 2.11.

ii) \Rightarrow i). Let us fix an $A' \in \mathcal{C}_\tau(A)$ and an A_i with $A_i \cap A' \neq \emptyset$. Since A_i is τ -connected and $A' \in \mathcal{C}_\tau(A)$, Corollary 6.2 applied to the sets $A_i \subset A \subset X$ yields $A_i \subset A'$. Moreover, A_1, \dots, A_m is a partition of A , and thus we conclude that

$$A' = \bigcup_{i \in I} A_i,$$

where $I := \{i : A_i \cap A' \neq \emptyset\}$. Now let us assume that $|I| \geq 2$. We fix an $i_0 \in I$ and write $A^+ := A_{i_0}$ and $A^- := \bigcup_{i \in I \setminus \{i_0\}} A_i$. Since $|I| \geq 2$, we obtain $A^- \neq \emptyset$, and hence Lemma 6.1 shows $d(A^+, A^-) < \tau$. On the other hand, our assumption ensures $d(A^+, A^-) \geq \tau$, and hence $|I| \geq 2$ cannot be true. Consequently, there exists a unique index i with $A' = A_i$, that is, we have shown the assertion. \square

Lemma 6.4. *Let (X, d) be a compact metric space and $A \subset X$ be a non-empty closed subset. Then the following statements are equivalent:*

- i) A is connected.

ii) A is τ -connected for all $\tau > 0$.

Proof of Lemma 6.4: $i) \Rightarrow ii)$. Let us assume that A is not τ -connected for some $\tau > 0$. Then, by Lemma 2.11, there are finitely many τ -connected components A_1, \dots, A_m of A with $m > 1$. We write $A' := A_1$ and $A'' := A_2 \cup \dots \cup A_m$. Then A' and A'' are non-empty, disjoint and $A' \cup A'' = A$ by construction. Moreover, Lemma 2.11 shows that A' and A'' are closed since A is closed, and hence A cannot be connected.

$ii) \Rightarrow i)$. Let us assume that A is not connected. Then there exist two non-empty closed disjoint subsets of A with $A' \cup A'' = A$. Since X is compact, A' and A'' are also compact, and hence $A' \cap A'' = \emptyset$ implies $\tau := d(A', A'') > 0$. Lemma 6.1 then shows that A is not τ -connected. \square

Proof of Lemma 2.13: $i)$. Let $A' \in \mathcal{C}(A)$ and $\tau > 0$. Since A is closed, so is A' , and hence A' is τ -connected by Lemma 6.4. Consequently, Corollary 6.2 shows that there exists an $A'' \in \mathcal{C}_\tau(A)$ with $A' \subset A''$, i.e. $\mathcal{C}(A)$ is comparable to $\mathcal{C}_\tau(A)$. Now we fix an $A'' \in \mathcal{C}_\tau(A)$. Then there exists an $x \in A''$, and to this x , there exists an $A' \in \mathcal{C}(A)$ with $x \in A'$. This yields $A' \cap A'' \neq \emptyset$, and since A' is τ -connected by Lemma 6.4, Corollary 6.2 shows $A' \subset A''$, i.e. we obtain $\zeta(A') = A''$. In other words, ζ is surjective.

$ii)$. Let A_1, \dots, A_m be the topologically connected components of A . Then the components are closed, and since A is a closed and thus compact subset of X , the components are compact, too. This shows $d(A_i, A_j) > 0$ for all $i \neq j$, and consequently we obtain $\tau_A^* > 0$. Let us fix a $\tau \in (0, \tau_A^*] \cap (0, \infty)$. Then, Lemma 6.4 shows that each A_i is τ -connected, and therefore Lemma 6.3 together with $d(A_i, A_j) \geq \tau_A^* \geq \tau$ for all $i \neq j$ yields $\mathcal{C}_\tau(A) = \{A_1, \dots, A_m\}$. Consequently, we have proved $\mathcal{C}(A) = \mathcal{C}_\tau(A)$. The bijectivity of ζ now follows from its surjectivity. For the proof of the last equation, we define $\tau^* := \sup\{\tau > 0 : \mathcal{C}(A) = \mathcal{C}_\tau(A)\}$. Then we have already seen that $\tau_A^* \leq \tau^*$. Now suppose that $\tau_A^* < \tau^*$. Then there exists a $\tau \in (\tau_A^*, \tau^*)$ with $\mathcal{C}(A) = \mathcal{C}_\tau(A)$. On the one hand, we then find $d(A_i, A_j) \geq \tau$ for all $i \neq j$ by Lemma 2.11, while on the other hand $\tau > \tau_A^*$ shows that there exist $i_0 \neq j_0$ with $d(A_{i_0}, A_{j_0}) < \tau$. In other words, the assumption $\tau_A^* < \tau^*$ leads to a contradiction, and hence we have $\tau_A^* = \tau^*$. \square

Proof of Lemma 2.15: Let us fix some $A', A'' \in \mathcal{C}(A)$ with $A' \neq A''$. Since ζ is injective, we then obtain $\zeta(A') \neq \zeta(A'')$. Combining this with $A' \subset \zeta(A')$ and $A'' \subset \zeta(A'')$, we find

$$d(A', A'') \geq d(\zeta(A'), \zeta(A'')) \geq \tau_B^*,$$

where the last inequality follows from the definition of τ_B^* in Lemma 2.13. Taking the infimum over all A' and A'' with $A' \neq A''$ yields the assertion. \square

6.3 Proofs Related to Cluster Persistence

Lemma 6.5. *Let (X, d) be a metric space and $A, B \subset X$ be two subsets. Then the following statements hold:*

- $i)$ If A is compact, then $A^{+\delta} = A^{\oplus\delta}$.
- $ii)$ We have $d(A, B) \leq d(A^{+\delta}, B^{+\delta}) + 2\delta$.
- $iii)$ We have

$$\bigcap_{\delta > 0} A^{+\delta} = \overline{A}. \quad (42)$$

- $iv)$ We have $(A \cup B)^{+\delta} = A^{+\delta} \cup B^{+\delta}$ and $(A \cap B)^{+\delta} \subset A^{+\delta} \cap B^{+\delta}$.
- $v)$ We have $A^{-\delta} \cup B^{-\delta} \subset (A \cup B)^{-\delta}$ and, if $d(A, B) > \delta$, we actually have $A^{-\delta} \cup B^{-\delta} = (A \cup B)^{-\delta}$.
- $vi)$ For $A_1, A_2 \subset X$ with $A_1 \cap A_2 = \emptyset$ and $B_i \subset A_i$ with $d(B_1, B_2) > \delta$, we have

$$(A_1^{-\delta} \setminus B_1^{-\delta}) \cup (A_2^{-\delta} \setminus B_2^{-\delta}) \subset (A_1 \cup A_2)^{-\delta} \setminus (B_1 \cup B_2)^{-\delta},$$

and equality holds, if $d(A_1, A_2) > \delta$.

vii) For all $\delta > 0$ and $\epsilon > 0$, we have $A \subset (A^{+\delta+\epsilon})^{-\delta}$ and $(A^{-\delta-\epsilon})^{+\delta} \subset A$.

viii) For all $\delta > 0$ and $\epsilon > 0$, we have $(\partial A)^{+\delta} \subset A^{+\delta+\epsilon} \setminus A^{-\delta-\epsilon}$.

Proof of Lemma 6.5: *i).* Clearly, it suffices to prove $A^{+\delta} \subset A^{\oplus\delta}$. To prove this inclusion, we fix an $x \in A^{+\delta}$. Then there exists a sequence $(x_n) \subset A$ with $d(x, x_n) \leq \delta + 1/n$ for all $n \geq 1$. Since A is compact, we may assume without loss of generality that (x_n) converges to some $x' \in A$. Now we easily obtain the assertion from $d(x, x') \leq d(x, x_n) + d(x_n, x')$.

ii). Let us fix an $x \in A^{+\delta}$ and an $y \in B^{+\delta}$. Then there exist two sequences $(x_n) \subset A$ and $(y_n) \subset B$ such that $d(x, x_n) \leq \delta + 1/n$ and $d(y, y_n) \leq \delta + 1/n$ for all $n \geq 1$. Now this construction yields

$$d(A, B) \leq d(x_n, y_n) \leq d(x_n, x) + d(x, y) + d(y, y_n) \leq d(x, y) + 2\delta + 2/n \quad n \geq 1,$$

and by first letting $n \rightarrow \infty$ and then taking the infimum over all $x \in A^{+\delta}$ and $y \in B^{+\delta}$, we obtain the assertion.

iii). To show the inclusion \supset , we fix an $x \in \bar{A}$. Then there exists a sequence $(x_n) \subset A$ with $x_n \rightarrow x$ for $n \rightarrow \infty$. For $\delta > 0$ there then exists an n_δ such that $d(x, x_n) \leq \delta$ for all $n \geq n_\delta$. This shows $d(x, A) \leq \delta$, i.e. $x \in A^{+\delta}$. To show the converse inclusion \subset , we fix an $x \in X$ that satisfies $x \in A^{+1/n}$ for all $n \geq 1$. Then there exists a sequence $(x_n) \subset A$ with $d(x, x_n) \leq 1/n$, and hence we find $x_n \rightarrow x$ for $n \rightarrow \infty$. This shows $x \in \bar{A}$.

iv). If $x \in (A \cup B)^{+\delta}$, there exists a sequence $(x_n) \subset A \cup B$ with $d(x, x_n) \leq \delta + 1/n$. Without loss of generality we may assume that $(x_n) \subset A$, which immediately yields $x \in A^{+\delta}$. The converse inclusion $A^{+\delta} \cup B^{+\delta} \subset (A \cup B)^{+\delta}$ and the inclusion $(A \cap B)^{+\delta} \subset A^{+\delta} \cap B^{+\delta}$ are trivial.

v). The first inclusion follows from part *iv)* and simple set algebra, namely

$$\begin{aligned} A^{-\delta} \cup B^{-\delta} &= X \setminus ((X \setminus A)^{+\delta} \cap (X \setminus B)^{+\delta}) \subset X \setminus ((X \setminus A) \cap (X \setminus B))^{+\delta} \\ &= X \setminus (X \setminus (A \cup B))^{+\delta} \\ &= (A \cup B)^{-\delta}. \end{aligned}$$

To show the converse inclusion, we fix an $x \in (A \cup B)^{-\delta}$. Since $(A \cup B)^{-\delta} \subset A \cup B$, we may assume without loss of generality that $x \in A$. It then remains to show that $x \in A^{-\delta}$, that is $d(x, X \setminus A) > \delta$. Obviously, $A \cap B = \emptyset$, which follows from $d(A, B) > \delta$, implies

$$X \setminus A = ((X \setminus A) \cap (X \setminus B)) \cup ((X \setminus A) \cap B) = (X \setminus (A \cup B)) \cup B,$$

and hence we obtain

$$d(x, X \setminus A) = d(x, X \setminus (A \cup B)) \wedge d(x, B) > \delta \wedge \delta = \delta,$$

where we used both $x \in (A \cup B)^{-\delta}$ and $d(A, B) > \delta$.

vi). Using the formula $(A_1 \cup A_2) \setminus (B_1 \cup B_2) = (A_1 \setminus B_1) \cup (A_2 \setminus B_2)$, which easily follows from $A_i \setminus B_j = A_i$ for $i \neq j$, we obtain

$$(A_1^{-\delta} \setminus B_1^{-\delta}) \cup (A_2^{-\delta} \setminus B_2^{-\delta}) = (A_1^{-\delta} \cup A_2^{-\delta}) \setminus (B_1^{-\delta} \cup B_2^{-\delta}) \subset (A_1 \cup A_2)^{-\delta} \setminus (B_1 \cup B_2)^{-\delta},$$

where in the last step we used *v)*. The second assertion also follows from *v)*.

vii). Obviously, $A \subset (A^{+\delta+\epsilon})^{-\delta}$ is equivalent to $(X \setminus A^{+\delta+\epsilon})^{+\delta} \subset X \setminus A$. To prove the latter, we fix an $x \in (X \setminus A^{+\delta+\epsilon})^{+\delta}$. Then there exists a sequence $(x_n) \subset X \setminus A^{+\delta+\epsilon}$ with $d(x, x_n) \leq \delta + 1/n$ for all $n \geq 1$. Moreover, $(x_n) \subset X \setminus A^{+\delta+\epsilon}$ implies $d(x_n, x') > \delta + \epsilon$ for all $n \geq 1$ and $x' \in A$. Now assume that we had $x \in A$. For an index n with $1/n \leq \epsilon$, we would then obtain $\delta + \epsilon < d(x_n, x) \leq \delta + \epsilon$, and hence $x \in A$ cannot be true.

To show the second inclusion we fix an $x \in (A^{-\delta-\epsilon})^{+\delta}$. Then there exists a sequence $(x_n) \subset A^{-\delta-\epsilon}$ such that $d(x, x_n) \leq \delta + 1/n$ for all $n \geq 1$. This time, $x_n \in A^{-\delta-\epsilon}$ implies $x_n \notin (X \setminus A)^{+\delta+\epsilon}$, that is $d(x_n, x') > \delta + \epsilon$ for all $n \geq 1$ and $x' \in X \setminus A$. Again, choosing an n with $1/n \leq \epsilon$, we then find $x \in A$.

viii). Let us fix an $x \in (\partial A)^{\oplus\delta}$. By definition, there then exists an $x' \in \partial A$ with $d(x, x') \leq \delta$. Moreover, by the definition of the boundary, there exists an $x'' \in A$ with $d(x', x'') \leq \epsilon$, and hence

we conclude that $d(x, x'') \leq \delta + \epsilon$, that is $x \in A^{+\delta+\epsilon}$. Since $\partial A = \partial(X \setminus A)$, the same argument yields $x \in (X \setminus A)^{+\delta+\epsilon}$, that is $x \notin A^{-\delta-\epsilon}$. Consequently, we have shown $(\partial A)^{\oplus\delta} \subset A^{+\delta+\epsilon} \setminus A^{-\delta-\epsilon}$. Using $(\partial A)^{+\delta} \subset (\partial A)^{\oplus(\delta+\epsilon)}$ and a simple change of variables then yields the assertion. \square

Proof of Lemma 2.19: *i*). Since $\tau > \delta$, there exist an $\varepsilon > 0$ with $\delta + \varepsilon < \tau$. For $x \in (A')^{+\delta}$, there thus exists an $x' \in A'$ with $d(x, x') \leq \delta + \varepsilon < \tau$, i.e. x and x' are τ -connected. Since A' is τ -connected, it is then easy to show that every pair $x, x'' \in (A')^{+\delta}$ is τ -connected.

ii). Let us fix an $A' \in \mathcal{C}_\tau(A^{+\delta})$ and an $x \in A'$. For $n \geq 1$ there then exists an $x_n \in A$ with $d(x, x_n) \leq \delta + 1/n$ and since by Lemma 2.11 there only exist finitely many τ -connected components of A , we may assume without loss of generality that there exists an $A'' \in \mathcal{C}_\tau(A)$ with $x_n \in A''$ for all $n \geq 1$. This yields $d(x, A'') \leq \delta + 1/n$ for all $n \geq 1$, and hence $d(x, A'') \leq \delta$. Consequently, we obtain $x \in (A'')^{+\delta}$, i.e. we have $(A'')^{+\delta} \cap A' \neq \emptyset$. Since $(A'')^{+\delta} \subset A^{+\delta}$, we then conclude that $(A'')^{+\delta} \subset A'$ by Corollary 6.2 and part *i*). Furthermore, we clearly have $A'' \subset (A'')^{+\delta}$, and hence $\zeta(A'') = A'$.

iii). Let us first consider the case $|\mathcal{C}(A)| = 1$. In this case, part *i*) of Lemma 2.13 shows $|\mathcal{C}_\tau(A)| = 1$, and thus $|\mathcal{C}_\tau(A^{+\delta})| = 1$ by the already established part *ii*). This makes the assertion obvious.

In the case $|\mathcal{C}(A)| > 1$ we write A_1, \dots, A_m for the τ -connected components of A . By part *iv*) of Lemma 6.5 we then obtain

$$A^{+\delta} = \bigcup_{i=1}^m A_i^{+\delta}. \quad (43)$$

Since $|\mathcal{C}(A)| > 1$, we further have $\tau_A^* < \infty$, and hence part *ii*) of Lemma 2.13 yields $\mathcal{C}(A) = \mathcal{C}_\tau(A)$. The definition of τ_A^* thus gives $d(A_i, A_j) \geq \tau_A^* \geq 3\tau$ for all $i \neq j$. Our first goal is to show that

$$d(A_i^{+\delta}, A_j^{+\delta}) \geq \tau, \quad i \neq j. \quad (44)$$

To this end, we fix $i \neq j$ and both an $x_i \in A_i^{+\delta}$ and an $x_j \in A_j^{+\delta}$. Now, the compactness of X yields the compactness of A_i and A_j by Lemma 2.11, and hence part *i*) of Lemma 6.5 shows that there exist $x'_i \in A_i$ and $x'_j \in A_j$ with $d(x_i, x'_i) \leq \delta$ and $d(x_j, x'_j) \leq \delta$. This yields

$$3\tau \leq d(x'_i, x'_j) \leq d(x'_i, x_i) + d(x_i, x_j) + d(x_j, x'_j) \leq 2\delta + d(x_i, x_j),$$

and the latter together with $\delta < \tau$ implies (44).

Now part *i*) showed that each $A_i^{+\delta}$, $i = 1, \dots, m$, is τ -connected. Combining this with (43), (44), and Lemma 6.3, we thus see that $A_1^{+\delta}, \dots, A_m^{+\delta}$ are the τ -connected components of $A^{+\delta}$. The bijectivity of ζ then follows from the surjectivity and a simple cardinality argument, and (9) becomes obvious. \square

Proof of Theorem 2.20: Let us first show the assertions related to the function τ^* . To this end, we first observe that for $\varepsilon \in (0, \rho^{**} - \rho^*]$ we have $|\mathcal{C}(M_{\rho^*+\varepsilon})| = |\mathcal{C}(M_{\rho^{**}})| = 2$ by Definition 2.16. This shows $\tau^*(\varepsilon) < \infty$.

Let us now fix $\varepsilon_1, \varepsilon_2 \in (0, \rho^{**} - \rho^*]$ with $\varepsilon_1 \leq \varepsilon_2$. Then Definition 2.16 guarantees that both $M_{\rho^*+\varepsilon_1}$ and $M_{\rho^*+\varepsilon_2}$ have two topologically connected components and that the top-CRM $\zeta : \mathcal{C}(M_{\rho^*+\varepsilon_2}) \rightarrow \mathcal{C}(M_{\rho^*+\varepsilon_1})$ is bijective. From Lemma 2.15 we thus obtain

$$\tau^*(\varepsilon_2) = \frac{1}{3}\tau_{M_{\rho^*+\varepsilon_2}}^* \geq \frac{1}{3}\tau_{M_{\rho^*+\varepsilon_1}}^* = \tau^*(\varepsilon_1).$$

i). Since $\emptyset \neq M_\rho \subset M_\rho^{+\delta}$, we find $|\mathcal{C}_\tau(M_\rho^{+\delta})| \geq 1$. On the other hand, since $\tau > \delta$, part *ii*) of Lemma 2.19 and part *i*) of Lemma 2.13 yield

$$|\mathcal{C}_\tau(M_\rho^{+\delta})| \leq |\mathcal{C}_\tau(M_\rho)| \leq |\mathcal{C}(M_\rho)| \leq 2. \quad (45)$$

ii). Let us fix a $\rho \in [\rho^* + \varepsilon^*, \rho^{**}]$. For $\varepsilon := \rho - \rho^*$, the monotonicity of τ^* then gives $\tau^*(\varepsilon^*) \leq \tau^*(\varepsilon)$, and hence we obtain

$$\tau \leq \frac{1}{3}\tau_{M_{\rho^*+\varepsilon^*}}^* \leq \frac{1}{3}\tau_{M_\rho}^* < \infty.$$

Part *ii*) of Lemma 2.13 thus shows that the CRM $\zeta_\rho : \mathcal{C}(M_\rho) \rightarrow \mathcal{C}_\tau(M_\rho)$ is bijective. Furthermore, $\delta < \tau \leq \tau_{M_\rho}^*/3$ together with part *iii*) of Lemma 2.19 shows that the τ -CRM $\zeta_\delta : \mathcal{C}_\tau(M_\rho) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$

is bijective. Consequently, the CRM $\zeta = \zeta_\delta \circ \zeta_\rho : \mathcal{C}(M_\rho) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$ is bijective, and from the latter we conclude that $|\mathcal{C}_\tau(M_\rho^{+\delta})| = |\mathcal{C}(M_\rho)| = 2$.

iii). Since $|\mathcal{C}_\tau(M_\rho^{+\delta})| = 2$, the already established (45) implies $|\mathcal{C}(M_\rho)| = 2$, and hence Definition 2.16 yields both $\rho \geq \rho^*$ and the bijectivity of the top-CRM $\zeta^{**} : \mathcal{C}(M_{\rho^{**}}) \rightarrow \mathcal{C}(M_\rho)$. Moreover, for ρ^{**} , the already established part *ii*) shows that the τ -CRM $\zeta_M : \mathcal{C}_\tau(M_{\rho^{**}}) \rightarrow \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})$ is bijective, and the proof of *ii*) further showed $\mathcal{C}(M_{\rho^{**}}) = \mathcal{C}_\tau(M_{\rho^{**}})$. Consequently, ζ_M equals the CRM $\mathcal{C}(M_{\rho^{**}}) \rightarrow \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})$. In addition, $\delta < \tau$ together with part *ii*) of Lemma 2.19 and part *i*) of Lemma 2.13 shows that the CRM $\zeta_\rho : \mathcal{C}(M_\rho) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$ is surjective. Now, by Lemma 2.8 these maps commute in the sense of the following diagram

$$\begin{array}{ccc} \mathcal{C}(M_{\rho^{**}}) & \xrightarrow{\zeta^{**}} & \mathcal{C}(M_\rho) \\ \zeta_M \downarrow & & \downarrow \zeta_\rho \\ \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta}) & \xrightarrow{\zeta} & \mathcal{C}_\tau(M_\rho^{+\delta}) \end{array}$$

and consequently, ζ is surjective. Since $|\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| = |\mathcal{C}(M_{\rho^{**}})| = 2$ and $|\mathcal{C}_\tau(M_\rho^{+\delta})| = 2$, we then conclude that ζ is bijective.

iv). We proceed by contraposition. To this end, we fix an $\rho \in [\rho^* + \varepsilon^*, \rho^{**}]$. By the already established part *ii*) we then find $|\mathcal{C}_\tau(M_\rho^{+\delta})| = 2$, and part *iii*) thus shows that the τ -CRM $\zeta_M : \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta}) \rightarrow \mathcal{C}_\tau(M_\rho^{+\delta})$ is bijective. Moreover, Lemma 2.8 yields the following diagram

$$\begin{array}{ccc} \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) & \xrightarrow{\zeta^{**}} & \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta}) \\ \zeta_V \downarrow & & \downarrow \zeta_M \\ \mathcal{C}_\tau(M_\rho^{-\delta}) & \xrightarrow{\zeta_{V,M}} & \mathcal{C}_\tau(M_\rho^{+\delta}) \end{array}$$

where ζ_V and $\zeta_{V,M}$ are the corresponding τ -CRMs. Now our assumption guarantees that ζ^{**} is bijective, and hence the diagram shows that $\zeta_{V,M} \circ \zeta_V$ is bijective. Consequently, ζ_V is injective, and from the latter we obtain $2 = |\mathcal{C}_\tau(M_\rho^{+\delta})| = |\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| \leq |\mathcal{C}_\tau(M_\rho^{-\delta})|$. \square

Proof of Lemma 2.21: *i*). Let us fix a $\psi > 2\psi_A^*(\delta)$ with $\psi < \tau$ and a $\tau' \in (0, \tau_A^*)$ such that $\psi + \tau' < \tau$, where τ_A^* is the constant defined in Lemma 2.13. Moreover, we fix a $B' \in \mathcal{C}(A)$. By Lemma 2.13 we then see that $\mathcal{C}(A) = \mathcal{C}_{\tau'}(A)$, and hence B' is τ' -connected. Now let A_1, \dots, A_m be the τ -connected components of $A^{-\delta}$. Clearly, Lemma 2.11 yields $d(A_i, A_j) \geq \tau$ for all $i \neq j$. Assume that *i*) is not true, that is, there exist indices i_0, j_0 with $i_0 \neq j_0$ such that $A_{i_0} \cap B' \neq \emptyset$ and $A_{j_0} \cap B' \neq \emptyset$. Consequently, there exist $x' \in A_{i_0} \cap B'$ and $x'' \in A_{j_0} \cap B'$, and since B' is τ' -connected, there further exist $x_0, \dots, x_{n+1} \in B' \subset A$ with $x_0 = x', x_{n+1} = x''$ and $d(x_i, x_{i+1}) < \tau'$ for all $i = 0, \dots, n$. Moreover, our assumptions guarantee $d(x_i, A^{-\delta}) < \psi/2$ for all $i = 0, \dots, n+1$. For all $i = 0, \dots, n+1$, there thus exists an index ℓ_i such that

$$d(x_i, A_{\ell_i}) < \psi/2.$$

In addition, we have $x_0 \in A_{i_0}$ and $x_{n+1} \in A_{j_0}$ by construction, and hence we may actually choose $\ell_0 = i_0$ and $\ell_{n+1} = j_0$. Since we assumed $\ell_0 \neq \ell_{n+1}$, there then exists an $i \in \{0, \dots, n\}$ with $\ell_i \neq \ell_{i+1}$. For this index, our construction now yields

$$d(A_{\ell_i}, A_{\ell_{i+1}}) \leq d(x_i, A_{\ell_i}) + d(x_i, x_{i+1}) + d(x_{i+1}, A_{\ell_{i+1}}) < \psi + \tau' < \tau,$$

which contradicts the earlier established $d(A_{\ell_i}, A_{\ell_{i+1}}) \geq \tau$.

ii). Since $A^{-\delta} \subset A$, there exists, for every $A' \in \mathcal{C}_\tau(A^{-\delta})$, a $B' \in \mathcal{C}(A)$ with $A' \cap B' \neq \emptyset$. We pick one such B' and define $\zeta(A') := B'$. Now part *i*) shows that $\zeta : \mathcal{C}_\tau(A^{-\delta}) \rightarrow \mathcal{C}(A)$ is injective, and hence we conclude $|\mathcal{C}_\tau(A^{-\delta})| \leq |\mathcal{C}(A)|$.

iii). As mentioned in part *ii*), we have an injective map $\zeta : \mathcal{C}_\tau(A^{-\delta}) \rightarrow \mathcal{C}(A)$ that satisfies

$$A' \cap \zeta(A') \neq \emptyset, \quad A' \in \mathcal{C}_\tau(A^{-\delta}). \quad (46)$$

Now, $|\mathcal{C}_\tau(A^{-\delta})| = |\mathcal{C}(A)|$ together with the assumed $|\mathcal{C}(A)| < \infty$ implies that ζ is actually bijective. Let us first show that ζ is the only map that satisfies (46). To this end, assume the converse, that is, for some $A' \in \mathcal{C}_\tau(A^{-\delta})$, there exists an $B' \in \mathcal{C}(A)$ with $B' \neq \zeta(A')$ and $A' \cap B' \neq \emptyset$. Since ζ is bijective, there then exists an $A'' \in \mathcal{C}_\tau(A^{-\delta})$ with $\zeta(A'') = B'$, and hence we have $A' \cap B' \neq \emptyset$ by (46). By part *i*), we conclude that $A' = A''$, which in turn yields $\zeta(A') = \zeta(A'') = B'$. In other words, we have found a contradiction, and hence ζ is indeed the only map that satisfies (46).

Let us now show that $\mathcal{C}_\tau(A^{-\delta})$ is persistent in $\mathcal{C}(A)$. Since we assumed $|\mathcal{C}_\tau(A^{-\delta})| = |\mathcal{C}(A)|$, it suffices to prove that the injective map $\zeta : \mathcal{C}_\tau(A^{-\delta}) \rightarrow \mathcal{C}(A)$ defined by (46) is a CRM, i.e. it satisfies

$$A' \subset \zeta(A'), \quad A' \in \mathcal{C}_\tau(A^{-\delta}). \quad (47)$$

To show (47), we pick an $A' \in \mathcal{C}_\tau(A^{-\delta})$ and write B_1, \dots, B_m for the topologically connected components of A . Since $A^{-\delta} \subset A$, we then have $A' \subset B_1 \cup \dots \cup B_m$, where the latter union is disjoint. Now, we have just seen that $\zeta(A') \in \{B_1, \dots, B_m\}$ is the only component satisfying $A' \cap \zeta(A') \neq \emptyset$, and therefore we can conclude $A' \subset \zeta(A')$.

Finally, let us show (11). To this end, we first prove that, for all $A' \in \mathcal{C}_\tau(A^{-\delta})$ and $x \in \zeta(A')$ we have

$$d(x, A') \leq \psi_A^*(\delta), \quad (48)$$

where $\zeta : \mathcal{C}_\tau(A^{-\delta}) \rightarrow \mathcal{C}(A)$ is the bijective CRM considered above. Let us assume that (48) is not true, that is, there exist an $A' \in \mathcal{C}_\tau(A^{-\delta})$ and an $x \in \zeta(A')$ such that $d(x, A') > \psi_A^*(\delta)$. Since $d(x, A^{-\delta}) \leq \psi_A^*(\delta)$, there further exists an $A'' \in \mathcal{C}_\tau(A^{-\delta})$ with $d(x, A'') \leq \psi_A^*(\delta)$. Obviously, this yields $A' \neq A''$. Let us fix a $\tau' \in (0, \tau_A^*)$ such that $2\psi_A^*(\delta) + \tau' < \tau$, and an $x' \in A'$. For $B' := \zeta(A')$, we then have $x' \in B'$ by (47), and our construction guarantees $x \in B'$. Now, the rest of the proof is similar to that of *i*). Namely, since B' is τ' -connected, there exist $x_0, \dots, x_{n+1} \in B'$ with $x_0 = x$, $x_{n+1} = x'$ and $d(x_i, x_{i+1}) < \tau'$ for all $i = 0, \dots, n$. Now let A_1, \dots, A_m be the τ -connected components of $A^{-\delta}$. Then, for all $i = 0, \dots, n+1$, there further exists an index ℓ_i such that

$$d(x_i, A_{\ell_i}) \leq \psi_A^*(\delta),$$

where again we may choose $A_{\ell_0} = A''$ and $A_{\ell_{n+1}} = A'$. Since $\ell_0 \neq \ell_{n+1}$, there then exists an $i \in \{0, \dots, n\}$ with $\ell_i \neq \ell_{i+1}$. For this index, our construction now yields

$$\tau \leq d(A_{\ell_i}, A_{\ell_{i+1}}) \leq d(x_i, A_{\ell_i}) + d(x_i, x_{i+1}) + d(x_{i+1}, A_{\ell_{i+1}}) < 2\psi_A^*(\delta) + \tau' < \tau,$$

and hence we have found a contradiction.

To prove (11), we again assume the converse, that is, that there exist $B', B'' \in \mathcal{C}(A)$ with $B' \neq B''$ and $d(B', B'') < \tau - 2\psi_A^*(\delta)$. Then there exist $x' \in B'$ and $x'' \in B''$ such that $d(x', x'') < \tau - 2\psi_A^*(\delta)$. Now, since ζ is bijective, there exists $A', A'' \in \mathcal{C}_\tau(A^{-\delta})$ with $A' \neq A''$, $B' = \zeta(A')$, and $B'' = \zeta(A'')$. Using (48), we then obtain

$$\tau \leq d(A', A'') \leq d(x', A') + d(x', x'') + d(x'', A'') < 2\psi_A^*(\delta) + \tau - 2\psi_A^*(\delta) = \tau,$$

i.e. we again have found a contradiction. \square

Proof of Theorem 2.23: *i*). To show that $|\mathcal{C}_\tau(M_\rho^{-\delta})| \geq 1$, we first observe that $\delta \leq \delta_{\text{thick}}$ implies

$$\sup_{x \in M_\rho} d(x, M_\rho^{-\delta}) = \psi_{M_\rho}^*(\delta) \leq c_{\text{thick}} \delta^\gamma < \infty,$$

and thus $M_\rho^{-\delta} \neq \emptyset$. This yields $|\mathcal{C}_\tau(M_\rho^{-\delta})| \geq 1$. Conversely, we have $|\mathcal{C}_\tau(M_\rho^{-\delta})| \leq |\mathcal{C}(M_\rho)| \leq 2$, where the first inequality was established in part *ii*) of Lemma 2.21 and the second is ensured by Definition 2.16.

ii). The monotonicity of τ^* established in Theorem 2.20 yields $\delta < \psi(\delta) < \tau \leq \tau^*(\varepsilon^*) \leq \tau_{M_{\rho^{**}}}^*/3$. By part *iii*) of Lemma 2.19 we then conclude that the τ -CRM $\mathcal{C}_\tau(M_{\rho^{**}}) \rightarrow \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})$ is bijective, and part *ii*) of Theorem 2.20 shows $|\mathcal{C}_\tau(M_{\rho^{**}})| = |\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| = 2$. By Lemma 2.8 it thus suffices to show that the τ -CRM $\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho^{**}})$ is bijective. Furthermore, if $|\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| = 1$, this map is automatically injective, and if $|\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| = 2$, the injectivity follows from the surjectivity and the

above proven $|\mathcal{C}_\tau(M_{\rho^{**}})| = 2$. Consequently, it actually suffices to show that ζ is surjective. To this end, we fix a $B' \in \mathcal{C}_\tau(M_{\rho^{**}})$ and an $x \in B'$. Then our assumption ensures $d(x, M_{\rho^{**}}^{-\delta}) < \psi(\delta)$, and hence there exists an $A' \in \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta})$ with $d(x, A') < \psi(\delta)$. Therefore, $\psi(\delta) < \tau$ implies that x and A' are τ -connected, which yields $x \in A'$. In other words, we have shown $A' \cap B' \neq \emptyset$. By Lemma 6.2 and the definition of ζ , we conclude that $\zeta(A') = B'$.

iii). We have $2 = |\mathcal{C}_\tau(M_{\rho}^{-\delta})| \leq |\mathcal{C}(M_\rho)| \leq 2$, where the first inequality was shown in part *ii*) of Lemma 2.21 and the second is guaranteed by Definition 2.16. We conclude that $|\mathcal{C}(M_\rho)| = 2$, and hence the definition of topological clustering ensures both $\rho \geq \rho^*$ and the bijectivity of the top-CRM $\zeta_{\text{top}} : \mathcal{C}(M_{\rho^{**}}) \rightarrow \mathcal{C}(M_\rho)$. Furthermore, $|\mathcal{C}_\tau(M_{\rho}^{-\delta})| = |\mathcal{C}(M_\rho)|$, which has been shown above, together with part *iii*) of Lemma 2.21 yields a bijective CRM $\zeta_\rho : \mathcal{C}_\tau(M_{\rho}^{-\delta}) \rightarrow \mathcal{C}(M_\rho)$. Moreover, part *ii*) of Theorem 2.20 shows $|\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| = 2$, and hence the already established bijectivity of $\zeta^{**} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})$ gives $|\mathcal{C}_\tau(M_{\rho^{**}}^{-\delta})| = |\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| = 2 = |\mathcal{C}(M_{\rho^{**}})|$. Consequently, part *iii*) of Lemma 2.21 yields a bijective CRM $\zeta_{\rho^{**}} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}(M_{\rho^{**}})$. Then the τ -CRM $\zeta : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho}^{-\delta})$ enjoys the following diagram

$$\begin{array}{ccc} \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) & \xrightarrow{\zeta_{\rho^{**}}} & \mathcal{C}(M_{\rho^{**}}) \\ \downarrow \zeta & & \downarrow \zeta_{\text{top}} \\ \mathcal{C}_\tau(M_{\rho}^{-\delta}) & \xrightarrow{\zeta_\rho} & \mathcal{C}(M_\rho) \end{array}$$

whose commutativity follows from Lemma 2.8. Then the bijectivity of $\zeta_{\rho^{**}}$, ζ_{top} , and ζ_ρ yields the bijectivity of ζ , which completes the proof. \square

6.4 Proofs Related to Basic Properties of Histograms

Proof of Theorem 3.3: We fix an $A \in \mathcal{A}_\delta$ and write $f := \mu(A)^{-1}\mathbf{1}_A$. Then f is non-negative, bounded, and our assumptions ensure $\|f\|_\infty \leq c_{\text{part}}\delta^{-d}$. Consequently, Hoeffding's inequality, see e.g. (Devroye et al., 1996, Theorem 8.1), yields

$$P^n \left(\left\{ D \in X^n : \left| \frac{1}{n} \sum_{i=1}^n f(x_i) - \mathbb{E}_P f \right| < \varepsilon \right\} \right) \geq 1 - 2 \exp \left(-\frac{2n\varepsilon^2\delta^{2d}}{c_{\text{part}}^2} \right)$$

for all $n \geq 1$ and $\varepsilon > 0$, where we assumed $D = (x_1, \dots, x_n)$. Furthermore, we have $\frac{1}{n} \sum_{i=1}^n f(x_i) = \mu(A)^{-1}D(A)$ and $\mathbb{E}_P f = \mu(A)^{-1}P(A)$. By a union bound argument and $|\mathcal{A}_\delta| \leq c_{\text{part}}\delta^{-d}$, we thus obtain

$$P^n \left(\left\{ D \in X^n : \sup_{A \in \mathcal{A}_\delta} \left| \frac{D(A)}{\mu(A)} - \frac{P(A)}{\mu(A)} \right| < \varepsilon \right\} \right) \geq 1 - 2c_{\text{part}}\delta^{-d} \exp \left(-\frac{2n\varepsilon^2\delta^{2d}}{c_{\text{part}}^2} \right).$$

Since, for $x \in X$ and $A \in \mathcal{A}_\delta$ with $x \in A$, we have $h_{D,\delta}(x) = \mu(A)^{-1}D(A)$ and $h_{P,\delta}(x) = \mu(A)^{-1}P(A)$, we then find the first assertion.

To show the second inequality, we fix an $A \in \mathcal{A}_\delta$ and write $f := \mu(A)^{-1}(\mathbf{1}_A - P(A))$. This yields $\mathbb{E}_P f = 0$, $\|f\|_\infty \leq c_{\text{part}}\delta^{-d}$, and

$$\mathbb{E}_P f^2 \leq \mu(A)^{-2}P(A) \leq \mu(A)^{-1}\|h\|_\infty \leq c_{\text{part}}\delta^{-d}\|h\|_\infty.$$

Consequently, Bernstein's inequality, see e.g. (Devroye et al., 1996, Theorem 8.2), yields

$$P^n \left(\left\{ D \in X^n : \left| \frac{1}{n} \sum_{i=1}^n f(x_i) \right| < \varepsilon \right\} \right) \geq 1 - 2 \exp \left(-\frac{3n\varepsilon^2\delta^d}{c_{\text{part}}(6\|h\|_\infty + 2\varepsilon)} \right).$$

Using $\frac{1}{n} \sum_{i=1}^n f(x_i) = (D(A) - P(A))\mu(A)^{-1}$, the rest of the proof follows the lines of the proof of the first inequality. \square

Proof of Lemma 3.4: *i).* We will show the equivalent inclusion $\{\hat{h} < \rho\} \subset (X \setminus M_{\rho+\varepsilon})^{+\delta}$. To this end, we fix an $x \in X$ with $\hat{h}(x) < \rho$. If $x \in X \setminus M_{\rho+\varepsilon}$, we immediately obtain $x \in (X \setminus M_{\rho+\varepsilon})^{+\delta}$, and hence we may restrict our considerations to the case $x \in M_{\rho+\varepsilon}$. Then, $\hat{h}(x) < \rho$ together with $\|\hat{h} - h_{P,\delta}\|_\infty \leq \varepsilon$ implies $h_{P,\delta}(x) \leq \hat{h}(x) + \varepsilon < \rho + \varepsilon$. Now let A be the unique cell of the partition \mathcal{A}_δ satisfying $x \in A$. The definition of $h_{P,\delta}$ together with the assumed $0 < \mu(A) < \infty$ then yields

$$\int_A h d\mu = P(A) = h_{P,\delta}(x)\mu(A) < (\rho + \varepsilon)\mu(A), \quad (49)$$

where $h : X \rightarrow [0, \infty)$ is an arbitrary μ -density of P . Our next goal is to show that there exists an $x' \in (X \setminus M_{\rho+\varepsilon}) \cap A$. Suppose the converse, that is $A \subset M_{\rho+\varepsilon}$. Then the upper normality of P at the level $\rho + \varepsilon$ yields $\mu(A \setminus \{h \geq \rho + \varepsilon\}) \leq \mu(M_{\rho+\varepsilon} \setminus \{h \geq \rho + \varepsilon\}) = 0$, and hence we conclude that $\mu(A \cap \{h \geq \rho + \varepsilon\}) = \mu(A)$. This leads to

$$\int_A h d\mu = \int_{A \cap \{h \geq \rho + \varepsilon\}} h d\mu + \int_{A \setminus \{h \geq \rho + \varepsilon\}} h d\mu = \int_{A \cap \{h \geq \rho + \varepsilon\}} h d\mu \geq (\rho + \varepsilon)\mu(A).$$

However, this inequality contradicts (49), and hence there does exist an $x' \in (X \setminus M_{\rho+\varepsilon}) \cap A$. This implies

$$d(x, X \setminus M_{\rho+\varepsilon}) \leq d(x, x') \leq \text{diam } A \leq \delta,$$

i.e. we have shown $x \in (X \setminus M_{\rho+\varepsilon})^{+\delta}$.

ii). Let us fix an $x \in X$ with $\hat{h}(x) \geq \rho$. If $x \in M_{\rho-\varepsilon}$, we immediately obtain $x \in M_{\rho-\varepsilon}^{+\delta}$, and hence it remains to consider the case $x \in X \setminus M_{\rho-\varepsilon}$. Clearly, if $\rho - \varepsilon \leq 0$, this case is impossible, and hence we may additionally assume $\rho - \varepsilon > 0$. Then, $\hat{h}(x) \geq \rho$ together with $\|\hat{h} - h_{P,\delta}\|_\infty \leq \varepsilon$ yields $h_{P,\delta}(x) \geq \hat{h}(x) - \varepsilon \geq \rho - \varepsilon$. Now let A be the unique cell of the partition \mathcal{A}_δ satisfying $x \in A$. By the definition of $h_{P,\delta}$ and $\mu(A) > 0$ we then obtain

$$\int_A h d\mu = P(A) = h_{P,\delta}(x)\mu(A) \geq (\rho - \varepsilon)\mu(A), \quad (50)$$

where, again, $h : X \rightarrow [0, \infty)$ is an arbitrary μ -density of P . Our next goal is to show that there exists an $x' \in M_{\rho-\varepsilon} \cap A$. Suppose the converse holds, that is $A \subset X \setminus M_{\rho-\varepsilon}$. Then the assumed upper normality of P at the level $\rho - \varepsilon$ yields

$$\mu(M_{\rho-\varepsilon} \Delta \{h \geq \rho - \varepsilon\}) = 0,$$

and using $A \Delta B = (X \setminus A) \Delta (X \setminus B)$, we thus find $\mu((X \setminus M_{\rho-\varepsilon}) \Delta \{h < \rho - \varepsilon\}) = 0$. Combining the latter with the assumed $A \subset X \setminus M_{\rho-\varepsilon}$ we obtain

$$\mu(A \setminus \{h < \rho - \varepsilon\}) \leq \mu((X \setminus M_{\rho-\varepsilon}) \setminus \{h < \rho - \varepsilon\}) = 0,$$

and this implies

$$\int_A h d\mu = \int_{A \cap \{h < \rho - \varepsilon\}} h d\mu + \int_{A \setminus \{h < \rho - \varepsilon\}} h d\mu = \int_{A \cap \{h < \rho - \varepsilon\}} h d\mu < (\rho - \varepsilon)\mu(A).$$

However, this inequality contradicts (50), and hence there does exist an $x' \in M_{\rho-\varepsilon} \cap A$. This yields

$$d(x, M_{\rho-\varepsilon}) \leq d(x, x') \leq \text{diam } A \leq \delta,$$

i.e. we have shown $x \in M_{\rho-\varepsilon}^{+\delta}$. □

Lemma 6.6. *Let (X, d) be a compact metric space and μ be a finite Borel measure on X with $\text{supp } \mu = X$. Moreover, let P be a μ -absolutely continuous probability measure on X , and $(L_\rho)_{\rho \geq 0}$ be a decreasing family of sets $L_\rho \subset X$ such that*

$$M_{\rho+\varepsilon}^{-\delta} \subset L_\rho \subset M_{\rho-\varepsilon}^{+\delta}$$

for some fixed $\delta > 0$, $\varepsilon \geq 0$, and all $\rho \geq 0$. For some fixed $\rho \geq 0$ and $\tau > 0$, let $\zeta : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(L_\rho)$ be the τ -CRM. Then the following statements hold:

i) For all $A' \in \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})$ with $A' \cap M_{\rho+3\varepsilon}^{-\delta} \neq \emptyset$ we have $\zeta(A') \cap L_{\rho+2\varepsilon} \neq \emptyset$.

ii) For all $B' \in \mathcal{C}_\tau(L_\rho)$ with $B' \notin \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$, we have

$$B' \subset (X \setminus M_{\rho+\varepsilon})^{+\delta} \cap M_{\rho-\varepsilon}^{+\delta} \quad (51)$$

$$B' \cap L_{\rho+2\varepsilon} \subset (X \setminus M_{\rho+\varepsilon})^{+\delta} \cap M_{\rho+\varepsilon}^{+\delta}. \quad (52)$$

Proof of Lemma 6.6: *i)* This assertion follows

$$\emptyset \neq A' \cap M_{\rho+3\varepsilon}^{-\delta} \subset \xi(A') \cap L_{\rho+2\varepsilon},$$

where we used the τ -CRM property $A' \subset \zeta(A')$ and the inclusion $M_{\rho+3\varepsilon}^{-\delta} \subset L_{\rho+2\varepsilon}$.

ii) Let fix a $B' \in \mathcal{C}_\tau(L_\rho)$ with $B' \notin \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$. For $x \in B'$ we then have

$$x \notin \bigcup_{A' \in \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})} \zeta(A'),$$

and hence the τ -CRM property yields

$$x \notin \bigcup_{A' \in \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})} A' = M_{\rho+\varepsilon}^{-\delta}.$$

This shows $x \in (X \setminus M_{\rho+\varepsilon})^{+\delta}$, i.e. we have proved $B' \subset (X \setminus M_{\rho+\varepsilon})^{+\delta}$. Now, (51) follows from $B' \subset L_\rho \subset M_{\rho-\varepsilon}^{+\delta}$, and (52) follows from $B' \cap L_{\rho+2\varepsilon} \subset L_{\rho+2\varepsilon} \subset M_{\rho+\varepsilon}^{+\delta}$. \square

Proof of Theorem 3.5: Our first goal is to establish the following *disjoint union*:

$$\begin{aligned} \mathcal{C}_\tau(L_\rho) &= \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})) \\ &\cup \{B' \in \mathcal{C}_\tau(L_\rho) \setminus \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})) : B' \cap L_{\rho+2\varepsilon} \neq \emptyset\} \\ &\cup \{B' \in \mathcal{C}_\tau(L_\rho) : B' \cap L_{\rho+2\varepsilon} = \emptyset\}. \end{aligned} \quad (53)$$

We begin by showing the auxiliary result

$$A' \cap M_{\rho+3\varepsilon}^{-\delta} \neq \emptyset, \quad A' \in \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}). \quad (54)$$

To this end, we observe that parts *i)* and *ii)* of Theorem 2.20 yield $|\mathcal{C}_\tau(M_{\rho^{**}}^{+\delta})| = 2$, and hence part *ii)* of Theorem 2.23 implies $|\mathcal{C}_\tau(M_{\rho^{**}}^{-\delta})| = 2$. Let W' and W'' be the two τ -connected components of $M_{\rho^{**}}^{-\delta}$. We first assume that $M_{\rho+\varepsilon}^{-\delta}$ has exactly one τ -connected component A' , i.e. $A' = M_{\rho+\varepsilon}^{-\delta}$. Then $\rho + 3\varepsilon \leq \rho^{**}$ and $\rho + \varepsilon \leq \rho + 3\varepsilon$ imply

$$\emptyset \neq M_{\rho^{**}}^{-\delta} \subset M_{\rho+3\varepsilon}^{-\delta} = M_{\rho+\varepsilon}^{-\delta} \cap M_{\rho+3\varepsilon}^{-\delta} = A' \cap M_{\rho+3\varepsilon}^{-\delta},$$

i.e. we have shown (54). Let us now assume that $M_{\rho+\varepsilon}^{-\delta}$ has more than one τ -component. Then it has exactly two such components A' and A'' by $\rho + \varepsilon < \rho^{**}$ and part *i)* of Theorem 2.23. By part *iii)* of Theorem 2.23 we may then assume without loss of generality that we have $W' \subset A'$ and $W'' \subset A''$. Since $\rho + 3\varepsilon \leq \rho^{**}$ implies $M_{\rho^{**}}^{-\delta} \subset M_{\rho+3\varepsilon}^{-\delta}$, these inclusions yield $\emptyset \neq W' = W' \cap M_{\rho^{**}}^{-\delta} \subset A' \cap M_{\rho+3\varepsilon}^{-\delta}$ and $\emptyset \neq W'' = W'' \cap M_{\rho^{**}}^{-\delta} \subset A'' \cap M_{\rho+3\varepsilon}^{-\delta}$. Consequently, we have proved (54) in this case, too.

Now, from (54) we conclude by part *i)* of Lemma 6.6 that $B' \cap L_{\rho+2\varepsilon} \neq \emptyset$ for all $B' \in \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$. This yields

$$\{B' \in \mathcal{C}_\tau(L_\rho) \setminus \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})) : B' \cap L_{\rho+2\varepsilon} = \emptyset\} = \{B' \in \mathcal{C}_\tau(L_\rho) : B' \cap L_{\rho+2\varepsilon} = \emptyset\},$$

which in turn implies (53).

Let us now show (16). Using (53) we first observe that it remains to show

$$B' \cap L_{\rho+2\varepsilon} = \emptyset,$$

for all $B' \in \mathcal{C}_\tau(L_\rho) \setminus \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$. Let us assume the converse, that is, there exists some $B' \in \mathcal{C}_\tau(L_\rho)$ with $B' \notin \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$ and $B' \cap L_{\rho+2\varepsilon} \neq \emptyset$. Since $L_{\rho+2\varepsilon} \subset M_{\rho+\varepsilon}^{+\delta}$, there then exists an $x \in B' \cap M_{\rho+\varepsilon}^{+\delta}$. By part *i*) of Lemma 6.5 the latter yields an $x' \in M_{\rho+\varepsilon}$ with $d(x, x') \leq \delta$, and since P has thick clusters we obtain

$$d(x', M_{\rho+\varepsilon}^{-\delta}) \leq \psi_{M_{\rho+\varepsilon}}^*(\delta) \leq c_{\text{thick}}\delta^\gamma < 2c_{\text{thick}}\delta^\gamma.$$

From this inequality we conclude that there exists an $x'' \in M_{\rho+\varepsilon}^{-\delta}$ satisfying $d(x', x'') < 2c_{\text{thick}}\delta^\gamma$. Let $A'' \in \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})$ be the unique τ -connected component satisfying $x'' \in A''$. The τ -CRM property then yields $x'' \in A'' \subset \zeta(A'') =: B''$, and hence, using $c \geq 1$, we find

$$d(B', B'') \leq d(x, x'') \leq d(x, x') + d(x', x'') < \delta + 2c_{\text{thick}}\delta^\gamma \leq 3c_{\text{thick}}\delta^\gamma < \tau.$$

However, since $B' \notin \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$ and $B'' \in \zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$ we obtain $B' \neq B''$, and therefore, Lemma 2.11 yields $d(B', B'') \geq \tau$, i.e. we have found a contradiction. \square

Proof of Theorem 3.6: We begin with some general observations. To this end, let $\rho \in [0, \rho^{**} - 4\varepsilon]$ be the level that is currently considered in Line 3 of Algorithm 3.1. Then, Theorem 3.5 shows that Algorithm 3.1 identifies exactly the τ -connected components of $L_{D,\rho}$ that belong to the set $\zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$, where $\zeta : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(L_{D,\rho})$ is the τ -CRM. In the following, we thus consider the set $\zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$. Moreover, we note that the returned level ρ_D^* always satisfies $\rho_D^* \geq \rho + 3\varepsilon$ by Line 4 and Line 6, and equality holds if and only if $|\zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))| \neq 1$.

i). Let us first consider the case $\rho \in [0, \rho^* - \varepsilon]$. Then $\rho + \varepsilon < \rho^*$ together with part *i*) and *iii*) of Theorem 2.23 shows $|\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})| = 1$, and hence $|\zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))| = 1$. Our initial consideration then show, that Algorithm 3.1 does not leave its loop, and thus $\rho_D^* \geq \rho^* + 2\varepsilon$.

Let us now consider the case $\rho \in [\rho^* + \varepsilon^* + \varepsilon, \rho^* + \varepsilon^* + 2\varepsilon]$. Then we first note that Algorithm 3.1 actually inspects such an ρ , since it iteratively inspects all $\rho = i\varepsilon$, $i = 0, 1, \dots$, and the width of the interval above is ε . Moreover, our assumptions on ε^* and ε guarantee $\rho^* + \varepsilon^* + 2\varepsilon \leq \rho^{**} - 4\varepsilon$ and hence we have $\rho \in [\rho^* + \varepsilon^* + \varepsilon, \rho^{**} - 4\varepsilon]$, that is, we are in the situation described at the beginning of the proof. Let us write $\zeta_V : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})$, $\zeta_M : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho-\varepsilon}^{+\delta})$, and $\zeta_{V,M} : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho-\varepsilon}^{+\delta})$ for the τ -CRMs between the involved sets. Using Lemma 2.8 twice, we then obtain the following diagram:

$$\begin{array}{ccc} \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) & \xrightarrow{\zeta_{V,M}} & \mathcal{C}_\tau(M_{\rho-\varepsilon}^{+\delta}) \\ \zeta_V \uparrow & & \uparrow \zeta_M \\ \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) & \xrightarrow{\zeta^{**}} & \mathcal{C}_\tau(M_{\rho^{**}}^{+\delta}) \end{array}$$

where the τ -CRM ζ^{**} is bijective by part *ii*) of Theorem 2.23. Moreover, $\rho - \varepsilon \geq \rho^* + \varepsilon^*$ together with part *ii*) of Theorem 2.20 shows $|\mathcal{C}_\tau(M_{\rho-\varepsilon}^{+\delta})| = 2$, and by *iii*) of Theorem 2.20 we conclude that ζ_M is bijective. Similarly, $\rho + \varepsilon \geq \rho^* + \varepsilon^*$ and the bijectivity of ζ^{**} show by *iv*) of Theorem 2.20 that $|\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})| = 2$, and thus ζ_V is bijective by part *iii*) of Theorem 2.23. Consequently, $\zeta_{V,M}$ is bijective. Let us further consider the τ -CRM $\zeta : \mathcal{C}_\tau(L_{D,\rho}) \rightarrow \mathcal{C}_\tau(M_{\rho-\varepsilon}^{+\delta})$. Then Lemma 2.8 yields another diagram:

$$\begin{array}{ccc} \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) & \xrightarrow{\zeta_{V,M}} & \mathcal{C}_\tau(M_{\rho-\varepsilon}^{+\delta}) \\ & \searrow \zeta & \nearrow \zeta' \\ & \mathcal{C}_\tau(L_{D,\rho}) & \end{array}$$

Since $\zeta_{V,M}$ is bijective, we then find that ζ is injective, and since we have already seen that $M_{\rho+\varepsilon}^{-\delta}$ has two τ -connected components, we conclude that $\zeta(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$ contains two elements. Consequently, the stopping criterion of Algorithm 3.1 is satisfied, that is, $\rho_D^* = \rho + 3\varepsilon \leq \rho^* + \varepsilon^* + 5\varepsilon$.

ii). Theorem 3.5 shows that in its last run through the loop Algorithm 3.1 identifies exactly the τ -connected components of $L_{D,\rho}$ that belong to the set $\zeta_{-3\varepsilon}(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))$, where $\rho := \rho_D^* - 3\varepsilon$ and $\zeta_{-3\varepsilon} : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}_\tau(L_{D,\rho})$ is the τ -CRM. Moreover, since Algorithm 3.1 stops at $\rho_D^* - 3\varepsilon$, we have $|\zeta_{-3\varepsilon}(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))| \neq 1$ by our remarks at the beginning of the proof, and thus $|\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})| \neq 1$. From the already proven part *i*) we further know that $\rho + \varepsilon = \rho_D^* - 2\varepsilon \leq \rho^* + \varepsilon^* + 3\varepsilon \leq \rho^* + 4\varepsilon^* \leq \rho^{**}$, and part *i*) of Theorem 2.23 hence gives $|\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})| = 2$. For later purposes, note that the latter together with $|\zeta_{-3\varepsilon}(\mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}))| \neq 1$ implies the injectivity of $\zeta_{-3\varepsilon}$. Now, part *iii*) of Theorem 2.23 shows that the τ -CRM $\zeta_{\rho^{**},\rho+\varepsilon} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta})$ is bijective. Let us consider the following commutative diagram:

$$\begin{array}{ccc} \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) & \xrightarrow{\zeta_{\rho^{**},\rho+\varepsilon}} & \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \\ & \searrow \zeta_{\rho^{**},\rho_D^*+\varepsilon} & \nearrow \tilde{\zeta} \\ & & \mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta}) \end{array}$$

where the remaining two maps are the corresponding τ -CRMs, whose existence is guaranteed by $\rho_D^* + \varepsilon \leq \rho_D^* + 7\varepsilon^* \leq \rho^{**}$ and $\rho + \varepsilon \leq \rho_D^* + \varepsilon$, respectively. Now the bijectivity of $\zeta_{\rho^{**},\rho+\varepsilon}$ shows that $\zeta_{\rho^{**},\rho_D^*+\varepsilon}$ is injective. Moreover, $\rho_D^* + \varepsilon \leq \rho^{**}$ implies $|\mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta})| \leq 2$ by part *i*) of Theorem 2.23, while $\rho^{**} \geq \rho^* + \varepsilon^*$ implies $|\mathcal{C}_\tau(M_{\rho^{**}}^{-\delta})| = 2$ by part *iv*) of Theorem 2.20 and part *ii*) of Theorem 2.23. Therefore, $\zeta_{\rho^{**},\rho_D^*+\varepsilon}$ is actually bijective. This yields both $|\mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta})| = 2$, which is the first assertion, and the bijectivity of $\tilde{\zeta}$. Let us consider yet another commutative diagram

$$\begin{array}{ccc} \mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta}) & \xrightarrow{\tilde{\zeta}} & \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \\ \zeta \downarrow & & \downarrow \zeta_{-3\varepsilon} \\ \mathcal{C}_\tau((L_{D,\rho_D^*})) & \xrightarrow{\zeta'} & \mathcal{C}_\tau(L_{D,\rho}) \end{array}$$

where again, all occurring maps are the τ -CRMs between the respective sets. Now we have already shown that $\zeta_{-3\varepsilon}$ is injective and that $\tilde{\zeta}$ is bijective. Consequently, ζ is injective.

iii). This assertions follows from Theorem 3.5 and the inequality $\rho_D^* \leq \rho^{**} - 3\varepsilon$, which follows from part *i*).

iv). We have already seen in the proof of part *ii*) that $|\mathcal{C}_\tau(M_{\rho^{**}}^{-\delta})| = 2$, and consequently part *iii*) of Lemma 2.21 shows that there exists a bijective CRM $\zeta_{\rho^{**}} : \mathcal{C}_\tau(M_{\rho^{**}}^{-\delta}) \rightarrow \mathcal{C}(M_{\rho^{**}})$. Moreover, part *ii*) shows $|\mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta})| = 2$, thus part *iii*) of Lemma 2.21 yields another bijective CRM $\zeta_{\rho_D^*+\varepsilon} : \mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}(M_{\rho_D^*+\varepsilon})$. Furthermore, in the proof of part *ii*) we have already seen that τ -CRM $\zeta_{\rho^{**},\rho_D^*+\varepsilon}$ is bijective. This gives the diagram.

v). By parts *iii*) and *iv*) we know $2 = |\mathcal{C}_\tau(M_{\rho_D^*+\varepsilon}^{-\delta})| = |\mathcal{C}(M_{\rho_D^*+\varepsilon})|$. Let us thus write B_1 and B_2 for the two topologically connected components of $M_{\rho_D^*+\varepsilon}$. Furthermore, part *i*) together with $\varepsilon^* \leq (\rho^{**} - \rho^*)/9$ ensures $\rho_D^* + \varepsilon \leq \rho^{**}$. Therefore, part *iii*) of Lemma 2.21, namely (11) shows

$$d(B_1, B_2) \geq \tau - 2\psi_{M_{\rho_D^*+\varepsilon}^*}^*(\delta) \geq \tau - 2c_{\text{thick}}\delta^\gamma > \tau - \psi(\delta).$$

On the other hand, the definition of $\tau_{M_{\rho_D^*+\varepsilon}^*}^*$ in Lemma 2.13 together with the definition of τ^* in (10) gives

$$3\tau^*(\rho_D^* - \rho^* + \varepsilon) = \tau_{M_{\rho_D^*+\varepsilon}^*}^* = d(B_1, B_2).$$

Combining both we find the assertion. \square

Proof of Theorem 3.7: To simplify notation in the following calculations, we write $B_i := B_i(D)$ for $i \in \{1, 2\}$ and $\rho := \rho_D^*$. Consequently, $A_{\rho+\varepsilon}^1$ and $A_{\rho+\varepsilon}^2$ are the two connected components of $M_{\rho+\varepsilon} = M_{\rho_D^*+\varepsilon}$. Moreover, we define $V_{\rho+\varepsilon}^1$ and $V_{\rho+\varepsilon}^2$ by (17), so that $V_{\rho+\varepsilon}^1$ and $V_{\rho+\varepsilon}^2$ become the two

τ -connected components of $M_{\rho+\varepsilon}^{-\delta} = M_{\rho_D^*+\varepsilon}^{-\delta}$. As discussed in front of Theorem 3.7, we then have $A_{\rho+\varepsilon}^i \subset A_i^*$, and the assumed (18) ensures $V_{\rho+\varepsilon}^i \subset B_i$ for $i = 1, 2$. For $i \in \{1, 2\}$, we further write $W_{\rho+\varepsilon}^i := (A_{\rho+\varepsilon}^i)^{-\delta}$. Our first goal is to show that

$$W_{\rho+\varepsilon}^i \subset V_{\rho+\varepsilon}^i, \quad i \in \{1, 2\}. \quad (55)$$

To this end, we fix an $x \in W_{\rho+\varepsilon}^1$. Since $W_{\rho+\varepsilon}^1 \subset A_{\rho+\varepsilon}^1$ and $W_{\rho+\varepsilon}^1 \subset M_{\rho+\varepsilon}^{-\delta}$, where the latter follows from $A_{\rho+\varepsilon}^1 \subset M_{\rho+\varepsilon}$, we then have $x \in A_{\rho+\varepsilon}^1$ and $x \in V_{\rho+\varepsilon}^1 \cup V_{\rho+\varepsilon}^2$. Let us assume that $x \in V_{\rho+\varepsilon}^2$. Then we have $V_{\rho+\varepsilon}^2 \cap A_{\rho+\varepsilon}^1 \neq \emptyset$. Now, the diagram of Theorem 3.6 shows that $\zeta_{\rho+\varepsilon} : \mathcal{C}_\tau(M_{\rho+\varepsilon}^{-\delta}) \rightarrow \mathcal{C}(M_{\rho+\varepsilon})$ satisfies $\zeta_{\rho+\varepsilon}(V_{\rho+\varepsilon}^2) = A_{\rho+\varepsilon}^2$, and hence we have $V_{\rho+\varepsilon}^2 \subset A_{\rho+\varepsilon}^2$. Consequently, $V_{\rho+\varepsilon}^2 \cap A_{\rho+\varepsilon}^1 \neq \emptyset$ implies $A_{\rho+\varepsilon}^1 \cap A_{\rho+\varepsilon}^2 \neq \emptyset$, which is a contradiction. Therefore, we have $x \in V_{\rho+\varepsilon}^1$, that is, we have shown (55) for $i = 1$. The case $i = 2$ can be shown analogously.

With the help of (55) we now find $W_{\rho+\varepsilon}^i \subset V_{\rho+\varepsilon}^i \subset B_i$, and thus $\mu(A_i^* \setminus B_i) \leq \mu(A_i^* \setminus W_{\rho+\varepsilon}^i)$ for $i = 1, 2$. Conversely, using the equation $\mu(B \setminus A) = \mu(B) - \mu(A \cap B)$ twice, we obtain

$$\begin{aligned} \mu(B_1 \setminus (A_1^* \cup A_2^*)) &= \mu(B_1) - \mu(B_1 \cap (A_1^* \cup A_2^*)) \\ &\geq \mu(B_1) - \mu(B_1 \cap A_1^*) - \mu(B_1 \cap A_2^*) \\ &= \mu(B_1 \setminus A_1^*) - \mu(B_1 \cap A_2^*). \end{aligned}$$

Since $B_1 \cap B_2 = \emptyset$ implies $B_1 \cap A_2^* \subset A_2^* \setminus B_2$, we thus find

$$\begin{aligned} \mu(B_1 \Delta A_1^*) &= \mu(B_1 \setminus A_1^*) + \mu(A_1^* \setminus B_1) \\ &\leq \mu(B_1 \setminus (A_1^* \cup A_2^*)) + \mu(A_2^* \setminus B_2) + \mu(A_1^* \setminus B_1) \\ &\leq \mu(B_1 \setminus \{h > \rho^*\}) + \mu(A_1^* \setminus W_{\rho+\varepsilon}^1) + \mu(A_2^* \setminus W_{\rho+\varepsilon}^2), \end{aligned}$$

where in the last estimate we also used (5). Repeating this estimate for $\mu(B_2 \Delta A_2^*)$ and using both $B_1 \cup B_2 \subset L_{D,\rho} \subset M_{\rho-\varepsilon}^{+\delta}$ then yields the assertion. \square

Proof of Theorem 3.8: Let $D \in X^n$ be a dataset that satisfies $\|h_{D,\delta} - h_{P,\delta}\|_\infty < \varepsilon$. By the first estimate of Theorem 3.3 we easily check that the probability P^n of such a D is not smaller than $1 - e^{-\varsigma}$. In the case of a bounded density and (21) the same holds by the second estimate of Theorem 3.3 and

$$\begin{aligned} &\sqrt{\frac{6c_{\text{part}}\|h\|_\infty\varsigma + \ln(2c_{\text{part}}) - d \ln \delta}{3\delta^{d_n}} + \left(\frac{2c_{\text{part}}\varsigma}{3\delta^{d_n}}\right)^2} + \frac{c_{\text{part}}\varsigma}{3\delta^{d_n}} \\ &\leq \sqrt{\frac{6c_{\text{part}}\|h\|_\infty\varsigma + \ln(2c_{\text{part}}) - d \ln \delta}{3\delta^{d_n}} + \frac{2c_{\text{part}}\varsigma}{3\delta^{d_n}}} \\ &\leq \sqrt{\frac{2c_{\text{part}}(1 + \|h\|_\infty)(\varsigma + \ln(2c_{\text{part}}) - d \ln \delta)}{\delta^{d_n}} + \frac{2c_{\text{part}}(\varsigma + \ln(2c_{\text{part}}) - d \ln \delta)}{3\delta^{d_n}}}, \end{aligned}$$

where in the last step we utilized $\ln(2c_{\text{part}}) \geq d \ln \delta$. Now, Lemma 3.4 shows

$$M_{\rho+\varepsilon}^{-\delta} \subset L_{D,\rho} \subset M_{\rho-\varepsilon}^{+\delta}$$

for all $\rho \geq 0$. Let us check that the remaining assumptions of Theorem 3.6 are also satisfied, if $\varepsilon^* \leq (\rho^{**} - \rho^*)/9$. Clearly, we have $\delta \in (0, \delta_{\text{thick}}]$, $\varepsilon \in (0, \varepsilon^*]$, and $\psi(\delta) < \tau$. To show the remaining $\tau \leq \tau^*(\varepsilon^*)$ we write $E := \{\varepsilon' \in (0, \rho^{**} - \rho^*] : \tau^*(\varepsilon') \geq \tau\}$. Since $\varepsilon^* < \infty$, we first observe that $E \neq \emptyset$ by the definition of ε^* . Consequently, there exists an $\varepsilon' \in E$ with $\varepsilon' \leq \inf E + \varepsilon = \varepsilon^*$. Using the monotonicity of τ^* established in Theorem 2.20 we then conclude that $\tau \leq \tau^*(\varepsilon') \leq \tau^*(\varepsilon^*)$. \square

6.5 Proofs Related to Consistency and Rates

Lemma 6.7. *Let (X, d) be a metric space, μ be a finite Borel measure on X , and $(A_\rho)_{\rho \in \mathbb{R}}$ be a decreasing family of closed subsets of X . For $\rho^* \in \mathbb{R}$, we write*

$$\dot{A}_{\rho^*} := \bigcup_{\rho > \rho^*} \dot{A}_\rho \quad \text{and} \quad \hat{A}_{\rho^*} := \bigcup_{\rho > \rho^*} A_\rho.$$

Then we have

$$\dot{A}_{\rho^*} = \bigcup_{\rho > \rho^*} \bigcup_{\varepsilon > 0} \bigcup_{\delta > 0} A_{\rho+\varepsilon}^{-\delta}.$$

Moreover, the following statements are equivalent:

i) $\mu(\dot{A}_{\rho^*} \setminus \dot{A}_{\rho^*}) = 0.$

ii) For all $\varepsilon > 0$, there exists a $\rho_\varepsilon > \rho^*$ such that, for all $\rho \in (\rho^*, \rho_\varepsilon]$, we have $\mu(A_\rho \setminus \dot{A}_\rho) \leq \varepsilon.$

Proof of Lemma 6.7: To show the first equality, we observe that (42) implies

$$\bigcap_{\rho > \rho^*} \bigcap_{\varepsilon > 0} \bigcap_{\delta > 0} (X \setminus A_{\rho+\varepsilon})^{+\delta} = \bigcap_{\varepsilon > 0} \bigcap_{\rho > \rho^*} \overline{X \setminus A_{\rho+\varepsilon}} = \bigcap_{\rho > \rho^*} \overline{X \setminus A_\rho}.$$

Moreover, every set $A \subset X$ satisfies $\overline{X \setminus A} = X \setminus \dot{A}$, and hence we obtain

$$\bigcap_{\rho > \rho^*} \bigcap_{\varepsilon > 0} \bigcap_{\delta > 0} (X \setminus A_{\rho+\varepsilon})^{+\delta} = \bigcap_{\rho > \rho^*} \overline{X \setminus A_\rho} = \bigcap_{\rho > \rho^*} (X \setminus \dot{A}_\rho) = X \setminus \bigcup_{\rho > \rho^*} \dot{A}_\rho.$$

Therefore, by taking the complement we find

$$\bigcup_{\rho > \rho^*} \dot{A}_\rho = X \setminus \left(\bigcap_{\rho > \rho^*} \bigcap_{\varepsilon > 0} \bigcap_{\delta > 0} (X \setminus A_{\rho+\varepsilon})^{+\delta} \right) = \bigcup_{\rho > \rho^*} \bigcup_{\varepsilon > 0} \bigcup_{\delta > 0} (X \setminus (X \setminus A_{\rho+\varepsilon})^{+\delta}) = \bigcup_{\rho > \rho^*} \bigcup_{\varepsilon > 0} \bigcup_{\delta > 0} A_{\rho+\varepsilon}^{-\delta}.$$

i) \Rightarrow ii). Let us fix an $\varepsilon > 0$. Since $\dot{A}_\rho = \bigcup_{\rho' \geq \rho} \dot{A}_{\rho'} \nearrow \dot{A}_{\rho^*}$ for $\rho \searrow \rho^*$, the σ -continuity of finite measures yields a $\rho_\varepsilon > \rho^*$ such that $\mu(\dot{A}_{\rho^*} \setminus \dot{A}_\rho) \leq \varepsilon$ for all $\rho \in (\rho^*, \rho_\varepsilon]$. Using $A_\rho \subset \dot{A}_{\rho^*}$ for $\rho > \rho^*$, we then obtain the assertion $\mu(A_\rho \setminus \dot{A}_\rho) \leq \mu(\dot{A}_{\rho^*} \setminus \dot{A}_\rho) \leq \varepsilon.$

ii) \Rightarrow i). Let us fix an $\varepsilon > 0$. For $\rho \in (\rho^*, \rho_\varepsilon]$, we then have $\dot{A}_\rho \subset \dot{A}_{\rho^*}$, and hence our assumption yields $\mu(A_\rho \setminus \dot{A}_{\rho^*}) \leq \varepsilon$. In other words, we have $\lim_{\rho \searrow \rho^*} \mu(A_\rho \setminus \dot{A}_{\rho^*}) = 0$. Moreover, we have $A_\rho \nearrow \dot{A}_{\rho^*}$ for $\rho \searrow \rho^*$, and hence the σ -continuity of μ yields $\lim_{\rho \searrow \rho^*} \mu(A_\rho \setminus \dot{A}_{\rho^*}) = \mu(\dot{A}_{\rho^*} \setminus \dot{A}_{\rho^*}) = 0.$ \square

Lemma 6.8. Let $f : (0, 1] \rightarrow (0, \infty)$ be a monotonously increasing function and $g : (0, f(1)] \rightarrow [0, 1]$ be its generalized inverse, that is

$$g(y) := \inf \{x \in (0, 1] : f(x) \geq y\}, \quad y \in (0, 1].$$

Then we have $\lim_{y \rightarrow 0^+} g(y) = 0.$

Proof of Lemma 6.8: Let $(y_n) \subset (0, f(1)]$ be a sequence with $y_n \rightarrow 0$. For $n \geq 1$, we write $E_n := \{x \in (0, 1] : f(x) \geq y_n\}$. Let us fix an $\varepsilon \in (0, 1]$. Since f is strictly positive, we then find $f(\varepsilon) > 0$, and hence there exists an $n_0 \geq 1$ such that $f(\varepsilon) \geq y_n$ for all $n \geq n_0$. Consequently, we have $\varepsilon \in E_n$ for all $n \geq n_0$, and from the latter we conclude that $g(y_n) = \inf E_n \leq \varepsilon$ for such n . \square

Proof of Theorem 4.1: Let us fix an $\varepsilon > 0$. For given $n \geq 1$ and $\tau := \tau_n$, $\varepsilon := \varepsilon_n$ we define ε_n^* by the right hand-side of (22). Then, Lemma 6.8 shows $0 < \varepsilon_n^* \leq \varepsilon \wedge (\rho^{**} - \rho^*)/9$ for all sufficiently large n . In addition, δ_n and ε_n satisfy (20) for sufficiently large n by (23), and we also have $\delta_n \leq \delta_{\text{thick}}$ for sufficiently large n . Consequently, there exists an $n_0 \geq 1$ such that, for all $n \geq n_0$, the values ε_n , δ_n , τ_n and ε_n^* satisfy the assumptions of Theorem 3.8 as well as $\varepsilon_n^* \leq \varepsilon.$

Let us now consider an $n \geq n_0$ and a data set $D \in X^n$ satisfying both the assertions i) - v) of Theorem 3.6 and (19). By Theorem 3.8 and our previous considerations we then know that the probability P^n of D is not less than $1 - e^{-\varepsilon}$. Now, part i) of Theorem 3.6 yields $\rho_D^* - \rho^* \geq 2\varepsilon_n > 0$ and

$$\rho_D^* - \rho^* \leq \varepsilon_n^* + 5\varepsilon_n \leq 6\varepsilon_n^* \leq 6\varepsilon,$$

i.e. we have shown the first convergence.

In order to prove the second convergence, we write A_i , $i = 1, 2$, for the two topologically connected components of $M_{\rho^{**}}$. Moreover, for $\rho \in (\rho^*, \rho^{**})$, we define $A_\rho^i := \zeta_\rho(A_i)$, where $\zeta_\rho : \mathcal{C}(M_{\rho^{**}}) \rightarrow$

$\mathcal{C}(M_\rho)$ is the top-CRM. In addition, we write $A_\rho^i := \emptyset$ for $\rho > \rho^{**}$ and $A_\rho^i := X$ for $\rho \leq \rho^*$. Our first goal is to show that

$$\mu(\hat{A}_{\rho^*}^i \setminus \dot{A}_{\rho^*}^i) = 0 \quad (56)$$

for $i = 1, 2$, where we used the notation of Lemma 6.7. To this end, we fix an $\epsilon > 0$. Since P is lower and upper normal at every level $\rho \in [\rho^*, \rho^{**}]$ we find, for an arbitrary μ -density h of P ,

$$\mu(\hat{M}_{\rho^*} \setminus \dot{M}_{\rho^*}) = \mu(\{h > \rho^*\} \setminus \dot{M}_{\rho^*}) = 0,$$

where we used (5), (6), and the notation of Lemma 6.7. Lemma 6.7 then shows that there exists a $\rho_\epsilon > \rho^*$ such that

$$\mu(M_\rho \setminus \dot{M}_\rho) \leq \epsilon \quad (57)$$

for all $\rho \in (\rho^*, \rho_\epsilon]$, where we may assume without loss of generality that $\rho_\epsilon \leq \rho^{**}$. Let us now fix a $\rho \in (\rho^*, \rho_\epsilon]$. Then we obviously have $\dot{A}_\rho^1 \cup \dot{A}_\rho^2 \subset \dot{M}_\rho$. To prove that the converse inclusion also holds, we pick an $x \in \dot{M}_\rho$. Without loss of generality we may assume that $x \in A_\rho^1$. Since A_ρ^2 is closed and thus compact, we then have $\varepsilon := d(x, A_\rho^2) > 0$. Moreover, since \dot{M}_ρ is open, there exists a $\delta \in (0, \varepsilon)$ such that $B(x, \delta) \subset \dot{M}_\rho$. This yields $B(x, \delta) \subset A_\rho^1 \cup A_\rho^2$, and by $d(x, A_\rho^2) > \delta$, we conclude that $B(x, \delta) \subset A_\rho^1$. This shows $x \in \dot{A}_\rho^1$, and hence we indeed have $\dot{M}_\rho = \dot{A}_\rho^1 \cup \dot{A}_\rho^2$. Now we use this equality to obtain

$$M_\rho \setminus \dot{M}_\rho = (A_\rho^1 \setminus (\dot{A}_\rho^1 \cup \dot{A}_\rho^2)) \cup (A_\rho^2 \setminus (\dot{A}_\rho^1 \cup \dot{A}_\rho^2)) = (A_\rho^1 \setminus \dot{A}_\rho^1) \cup (A_\rho^2 \setminus \dot{A}_\rho^2).$$

Using (57), this implies $\mu(A_\rho^i \setminus \dot{A}_\rho^i) \leq \epsilon$, and therefore Lemma 6.7 shows (56).

Let us now fix an $\epsilon > 0$ and a $\varsigma \geq 1$. By the equality of Lemma 6.7 and the σ -continuity of finite measures there then exist $\delta_\epsilon > 0$, $\varepsilon_\epsilon > 0$, and $\rho_\epsilon \in (\rho^*, \rho^{**}]$ such that, for all $\varepsilon \in (0, \varepsilon_\epsilon]$, $\delta \in (0, \delta_\epsilon]$, $\rho \in (\rho^*, \rho_\epsilon]$, and $i = 1, 2$, we have $\mu(\dot{A}_{\rho^*}^i \setminus (A_{\rho+\varepsilon}^i)^{-\delta}) \leq \epsilon$. Combining this with $A_i^* = \hat{A}_{\rho^*}^i$, which holds by the definition of the clusters A_i^* , and Equation (56) we then obtain

$$\mu(A_i^* \setminus (A_{\rho+\varepsilon}^i)^{-\delta}) = \mu(\hat{A}_{\rho^*}^i \setminus (A_{\rho+\varepsilon}^i)^{-\delta}) = \mu(\dot{A}_{\rho^*}^i \setminus (A_{\rho+\varepsilon}^i)^{-\delta}) \leq \epsilon. \quad (58)$$

Moreover, our assumption $\mu(\overline{A_1^* \cup A_2^*} \setminus (A_1^* \cup A_2^*)) = 0$ means $\mu(\overline{\hat{M}_{\rho^*}} \setminus \hat{M}_{\rho^*}) = 0$, and since by part *iii*) of Lemma 6.5 we know that

$$\bigcap_{\delta > 0} \left(\bigcup_{\rho > \rho^*} M_\rho \right)^{+\delta} = \overline{\bigcup_{\rho > \rho^*} M_\rho} = \overline{\hat{M}_{\rho^*}}$$

we find

$$\mu\left(\left(\bigcup_{\rho > \rho^*} M_\rho\right)^{+\delta} \setminus \hat{M}_{\rho^*}\right) \leq \epsilon$$

for all sufficiently small $\delta > 0$. From this it is easy to conclude that

$$\mu(M_{\rho-\varepsilon}^{+\delta} \setminus \hat{M}_{\rho^*}) \leq \epsilon \quad (59)$$

for all sufficiently small $\varepsilon > 0$, $\delta > 0$ and all $\rho > \rho^* + \varepsilon$. Without loss of generality, we may thus assume that (59) also holds for all $\varepsilon \in (0, \varepsilon_\epsilon]$, $\delta \in (0, \delta_\epsilon]$ and all $\rho > \rho^* + \varepsilon$.

For given $\tau := \tau_n$ and $\varepsilon := \varepsilon_n$ we now define ε_n^* by the right hand-side of (22). Then, Lemma 6.8 shows $\varepsilon_n^* \rightarrow 0$, and hence we obtain $\varepsilon_n^* \leq \min\{\frac{\rho_\epsilon - \rho^*}{9}, \varepsilon, \varepsilon_\epsilon\}$ for all sufficiently large n . In addition, δ_n and ε_n satisfy (20) for sufficiently large n by (23), and we also have $\varepsilon_n \leq \varepsilon \wedge \varepsilon_\epsilon$ and $\delta_n \leq \delta_\epsilon \wedge \delta_{\text{thick}}$ for sufficiently large n . Consequently, there exists an $n_0 \geq 1$ such that, for all $n \geq n_0$, the values ε_n , δ_n , τ_n and ε_n^* satisfy the assumptions of Theorem 3.8 as well as $\varepsilon_n \leq \varepsilon \wedge \varepsilon_\epsilon$ and $\delta_n \leq \delta_\epsilon$.

Let us now consider an $n \geq n_0$ and a data set $D \in X^n$ satisfying both the assertions *i*) - *v*) of Theorem 3.6 and (19). By Theorem 3.8 and our previous considerations we then know that the probability P^n of D is not less than $1 - e^{-\varsigma}$. Now, part *i*) of Theorem 3.6 gives both $\rho_D^* \geq \rho^* + 2\varepsilon_n >$

$\rho^* + \varepsilon_n$ and $\rho_D^* \leq \rho^* + \varepsilon_n^* + 5\varepsilon_n \leq \rho^* + 6\varepsilon_n^* \leq \rho_\varepsilon$, and hence (58) and (59) hold for $\varepsilon := \varepsilon_n$, $\delta := \delta_n$, and $\rho := \rho_D^*$. Consequently, (19) shows

$$\begin{aligned} \mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) &\leq 2\mu(A_1^* \setminus (A_{\rho+\varepsilon}^1)^{-\delta}) + 2\mu(A_2^* \setminus (A_{\rho+\varepsilon}^2)^{-\delta}) \\ &\quad + \mu(M_{\rho-\varepsilon}^{+\delta} \setminus \{h > \rho^*\}). \\ &\leq 4\varepsilon + \mu(M_{\rho-\varepsilon}^{+\delta} \setminus \hat{M}_{\rho^*}) \\ &\leq 5\varepsilon, \end{aligned}$$

where in the second to last step we also used (6). \square

Proof of Lemma 4.3: Let $\varepsilon \in (0, \rho^{**} - \rho^*]$ and A_1 and A_2 be the connected components of $M_{\rho^*+\varepsilon}$. Since A_1 and A_2 are closed, they are compact, and hence there exist $x_1 \in A_1$ and $x_2 \in A_2$ with

$$a := \|x_1 - x_2\| = d(A_1, A_2), \quad (60)$$

where we note that $A_1 \cap A_2 = \emptyset$ implies $a > 0$. For $t \in [0, 1]$, we now consider

$$x(t) := tx_1 + (1-t)x_2.$$

Since X is convex, we note that $x(t) \in X$ for all $t \in [0, 1]$. Our first goal is to show that $x_i \in \partial_X M_{\rho^*+\varepsilon}$ for $i = 1, 2$. To this end, we assume the converse, e.g. $x_2 \in \overset{\circ}{M}_{\rho^*+\varepsilon}$. Then there exists an $\epsilon \in (0, a)$ with $B_X(x_2, \epsilon) \subset \overset{\circ}{A}_2$, where $B_X(x_2, \epsilon) := \{x \in X : \|x - x_2\| \leq \epsilon\}$ denotes the closed ball with center x_2 and radius ϵ in X . Now $\|x(\epsilon/a) - x_2\| = \epsilon$ implies $x(\epsilon/a) \in A_2$, while $\|x(\epsilon/a) - x_1\| = a - \epsilon$ shows $\|x(\epsilon/a) - x_1\| < d(A_1, A_2)$. Together this contradicts (60).

For what follows, let us now observe that $t \mapsto x(t)$ is a continuous map on $[0, 1]$, and since h is continuous, there exists a $t^* \in [0, 1]$ with $h(x(t^*)) = \min_{t \in [0, 1]} h(x(t))$. Our next goal is to show that

$$h(x(t^*)) \leq \rho^*. \quad (61)$$

To this end, we assume the converse, that is $h(x(t^*)) > \rho^*$. Then there exists a $\delta \in (0, \varepsilon]$ such that $h(x(t)) > \rho^* + \delta$ for all $t \in [0, 1]$, and therefore an application of Lemma 2.1 using the continuity of h yields $x(t) \in M_{\rho^*+\delta}$ for all $t \in [0, 1]$. In other words, x_1 and x_2 are path-connected in $M_{\rho^*+\delta}$, and since the connecting path is a straight line, it is easy to see that x_1 and x_2 are τ -connected for all $\tau > 0$. Let us pick a $\tau \leq 3\tau^*(\delta) = \tau_{M_{\rho^*+\delta}}^*$. Since $|\mathcal{C}(M_{\rho^*+\delta})| = 2$, part *ii*) of Lemma 2.13 then shows $\mathcal{C}(M_{\rho^*+\delta}) = \mathcal{C}_\tau(M_{\rho^*+\delta})$. Let \tilde{A}_1 and \tilde{A}_2 be the two topologically connected components of $M_{\rho^*+\delta}$. Our previous considerations then showed that \tilde{A}_1 and \tilde{A}_2 are also the two τ -connected components of $M_{\rho^*+\delta}$. Now, $\delta \leq \varepsilon$ gives a top-CRM $\zeta : \mathcal{C}(M_{\rho^*+\varepsilon}) \rightarrow \mathcal{C}(M_{\rho^*+\delta})$, which is bijective, since P can be clustered between ρ^* and ρ^{**} . Without loss of generality we may thus assume that $\zeta(A_i) = \tilde{A}_i$ for $i = 1, 2$. This yields $x_i \in A_i \subset \tilde{A}_i$, i.e. x_1 and x_2 do not belong to the same τ -connected component of $M_{\rho^*+\delta}$. Clearly, this contradicts our observation that x_1 and x_2 are τ -connected, and hence (61) is proven.

Let us now assume without loss of generality that $t^* \in [1/2, 1)$. Since we have already seen that $x_1 \in \partial_X M_{\rho^*+\varepsilon}$, our assumption (24) together with (61) yields

$$|h(x(t^*)) - h(x_1)| \leq c \|x(t^*) - x_1\|^\theta.$$

In addition, Lemma 2.1 together with the continuity of h shows $x_1 \in M_{\rho^*+\varepsilon} \subset \{h \geq \rho^* + \varepsilon\}$. Combining these estimates with (60) and $d(A_1, A_2) = \tau_{M_{\rho^*+\varepsilon}}^* = 3\tau^*(\varepsilon)$, we find

$$\begin{aligned} \rho^* + \varepsilon &\leq h(x_1) \\ &\leq h(x(t^*)) + c \|x(t^*) - x_1\|^\theta \\ &\leq \rho^* + c \|x(t^*) - x_1\|^\theta \\ &= \rho^* + c(1-t^*)^\theta \|x_1 - x_2\|^\theta \\ &\leq \rho^* + c 2^{-\theta} d^\theta(A_1, A_2) \\ &= \rho^* + c(3/2)^{-\theta} \tau^*(\varepsilon)^\theta, \end{aligned}$$

and from the latter the assertion easily follows. \square

Proof of Theorem 4.4: Let us begin by checking the conditions of Theorem 3.8. Obviously, ε is chosen this way, and the definition of ε^* together with the assumption $\varepsilon^* \leq (\rho^{**} - \rho^*)/9$ yields

$$(\tau/\underline{c}_{\text{sep}})^\kappa \leq \varepsilon^* < \rho^{**} - \rho^*. \quad (62)$$

Since the clusters have separation exponent κ , we find in the case $\kappa < \infty$ that

$$\begin{aligned} \varepsilon + \inf\{\tilde{\varepsilon} \in (0, \rho^{**} - \rho^*) : \tau^*(\tilde{\varepsilon}) \geq \tau\} &\leq \varepsilon + \inf\{\tilde{\varepsilon} \in (0, \rho^{**} - \rho^*) : \underline{c}_{\text{sep}}\tilde{\varepsilon}^{1/\kappa} \geq \tau\} \\ &= \varepsilon + (\tau/\underline{c}_{\text{sep}})^\kappa. \end{aligned}$$

Consequently, (22) holds in the case $\kappa < \infty$. Moreover, in the case $\kappa = \infty$, (62) together with $\rho^{**} < \infty$ implies $\tau \leq \underline{c}_{\text{sep}}$. In addition, the separation exponent $\kappa = \infty$ ensures $\tau^*(\tilde{\varepsilon}) \geq \underline{c}_{\text{sep}}$ for all $\tilde{\varepsilon} > 0$, and hence we obtain

$$\varepsilon + \inf\{\tilde{\varepsilon} \in (0, \rho^{**} - \rho^*) : \tau^*(\tilde{\varepsilon}) \geq \tau\} = \varepsilon \leq \varepsilon^*,$$

that is, (22) is also established in the case $\kappa = \infty$. Now, applying Theorem 3.8, we see that $\rho_D^* \in [\rho^* + 2\varepsilon, \rho^* + \varepsilon^* + 5\varepsilon]$ with probability P^n not less than $1 - e^{-\zeta}$, that is, (27) is proved. In addition, the definition of ε^* yield

$$\rho_D^* - \rho^* \leq \varepsilon^* + 5\varepsilon \leq (\tau/\underline{c}_{\text{sep}})^\kappa + 6\varepsilon,$$

and therefore, we also obtain (28). Let us finally show (29). To this end, we first observe that Theorem 3.8, or more precisely, part *v*) of Theorem 3.6 further ensures

$$\tau/2 \leq \tau - \psi(\delta) < 3\tau^*(\rho_D^* - \rho^* + \varepsilon) \leq 3\bar{c}_{\text{sep}}(\rho_D^* - \rho^* + \varepsilon)^{1/\kappa} < 3\bar{c}_{\text{sep}}2^{1/\kappa}(\rho_D^* - \rho^*)^{1/\kappa},$$

where in the last step we used the already established (27). By some elementary transformations we conclude

$$\frac{1}{2} \left(\frac{\tau}{6\bar{c}_{\text{sep}}} \right)^\kappa < \rho_D^* - \rho^*.$$

and combining this $2\varepsilon \leq \rho_D^* - \rho^*$ we obtain the assertion. \square

Proof of Corollary 4.5: Our first goal is to show (31) for $\kappa < \infty$ and sufficiently large n with the help of Theorem 4.4. To this end, we define $\varepsilon_n^* := \varepsilon_n + (\tau_n/\underline{c}_{\text{sep}})^\kappa$ for $n \geq 1$. Since the sequences (ε_n) , (δ_n) , and (τ_n) converge to 0, we then have $\delta_n \in (0, \delta_{\text{thick}}]$ and $\varepsilon_n^* \leq (\rho^{**} - \rho^*)/9$ for all sufficiently large n . Furthermore, our definitions ensure $\tau_n/\delta_n^\gamma \rightarrow \infty$, and thus we have $\tau_n \geq 6c_{\text{thick}}\delta_n^\gamma = 2\psi(\delta_n)$ for all sufficiently large n , too. Before we can apply Theorem 4.4, it thus remains to show (21) for sufficiently large n . To this end, we observe that, for $\varsigma_n := \ln n$, we have

$$\begin{aligned} \varepsilon_n' &:= \sqrt{\frac{2c_{\text{part}}(1 + \|h\|_\infty)(\varsigma_n + \ln(2c_{\text{part}}) - d \ln \delta_n)}{\delta_n^d n}} + \frac{2c_{\text{part}}(\varsigma_n + \ln(2c_{\text{part}}) - d \ln \delta_n)}{3\delta_n^d n} \\ &\preceq \left(\frac{\ln n}{n} \right)^{\frac{\gamma\kappa}{2\gamma\kappa+d}}. \end{aligned}$$

Using $\varepsilon_n \cdot \left(\frac{\ln n}{n}\right)^{-\frac{\gamma\kappa}{2\gamma\kappa+d}} \rightarrow \infty$, we then see that $\varepsilon_n \geq \varepsilon_n'$ for all sufficiently large n . Now, applying Theorem 4.4, namely (28), we obtain an $n_0 \geq 1$ and a constant \bar{K} such that (30) holds for all $n \geq n_0$. Moreover, if κ is exact, (29) yields a constant \underline{K} such that (31) holds for all $n \geq n_0$.

Let us now consider the case $\kappa = \infty$. In this case, we first observe that $\varepsilon_n^* := \varepsilon_n + (\tau_n/\underline{c}_{\text{sep}})^\kappa$ satisfies $\varepsilon_n^* = \varepsilon_n$ for all n with $\tau_n < \underline{c}_{\text{sep}}$, that is, for all sufficiently large n . Moreover, we have $\tau_n/\delta_n^\gamma \rightarrow \infty$, and, like in the case $\kappa < \infty$, it thus suffices to show (21) for sufficiently large n . To this end, we observe that, for $\varsigma_n := \ln n$ and ε_n' as above, we find that, for all sufficiently large n ,

$$\varepsilon_n' \leq c_2 \left(\frac{\ln n \cdot \sqrt{\ln \ln n}}{n} \right)^{\frac{1}{2}} \leq \varepsilon_n,$$

where c_2 is a suitable constant independent of n . Consequently, (27) and (28) yield (31) for all sufficiently large n . \square

Lemma 6.9. *Let $X \subset \mathbb{R}^d$ be compact and convex and d be a metric on X that is defined by a norm on \mathbb{R}^d . Then, for all $A \subset X$ and $x \in \bar{A}$, we have*

$$d(x, \partial_X A) \leq d(x, X \setminus A),$$

where $\partial_X A$ denotes the boundary of A in X .

Proof of Lemma 6.9: Before we begin with the proof let us recall that the $\overline{B}^X = \overline{B}^{\mathbb{R}^d}$ for all $B \subset X$ by the closedness of X , that is, taking the closure with respect to X or \mathbb{R}^d is the same. Like in the statement of the lemma, we thus omit the superscript in the following. Let us now write $\delta := d(x, X \setminus A)$. Then there exists a sequence $(x_n) \subset X \setminus A$ such that $d(x, x_n) \rightarrow \delta$. Since X is assumed to be compact, so is $\overline{X \setminus A}$, and thus there exists an $x_\infty \in \overline{X \setminus A}$ such that $d(x, x_\infty) \leq \delta$. Obviously, it suffices to show $x_\infty \in \partial_X A$. Let us assume the converse. Since $\partial_X A = \overline{A} \cap \overline{X \setminus A}$, we then have $x_\infty \notin \overline{A}$, that is $x_\infty \in X \setminus \overline{A}$. Now, the latter set is open in X , and hence there exists an $\varepsilon > 0$ such that $B_X(x_\infty, \varepsilon) \subset X \setminus \overline{A}$, where $B_X(x_\infty, \varepsilon)$ denotes the closed ball in X that has center x_∞ and radius ε . This ε must satisfy $\varepsilon < \delta$, since otherwise we would find a contradiction to $x \in \bar{A}$ by $x \in B_X(x_\infty, \delta) \subset B_X(x_\infty, \varepsilon) \subset X \setminus \bar{A}$. For $t := \varepsilon/\delta \in (0, 1)$ we now define $x' := tx + (1-t)x_\infty$. The convexity of X implies $x' \in X$, and since d is defined by a norm, we have $d(x_\infty, x') = td(x, x_\infty) \leq \varepsilon$. Together, this yields $x' \in B_X(x_\infty, \varepsilon) \subset X \setminus \bar{A} \subset X \setminus A$. Consequently, $d(x, x') = (1-t)d(x, x_\infty) \leq (1-t)\delta < \delta$ implies $d(x, X \setminus A) < \delta$, which contradicts the definition of δ . \square

Lemma 6.10. *Let $X \subset \mathbb{R}^d$ be compact and convex and d be a metric on X that is defined by a norm on \mathbb{R}^d . Then, for all $A \subset X$ and $\delta > 0$, we have*

$$A^{+\delta} \setminus A^{-\delta} \subset (\partial_X A)^{+\delta},$$

where the operations $A^{+\delta}$ and $A^{-\delta}$ as well as the boundary $\partial_X A$ are with respect to the metric space (X, d) .

Proof of Lemma 6.10: Let us fix an $x \in A^{+\delta} \setminus A^{-\delta} = A^{+\delta} \cap (X \setminus A)^{+\delta}$. If $x \in \bar{A}$, then Lemma 6.9 immediately yields $d(x, \partial_X A) \leq d(x, X \setminus A) \leq \delta$, that is $x \in (\partial_X A)^{+\delta}$. It thus suffices to consider the case $x \notin \bar{A}$. Then we find $x \in X \setminus \bar{A} \subset X \setminus A \subset \overline{X \setminus A}$, and hence another application of Lemma 6.9 yields $d(x, \partial_X (X \setminus A)) \leq d(x, A) \leq \delta$. Now the assertion easily follows from $\partial_X (X \setminus A) = \overline{X \setminus A} \cap \overline{X \setminus (X \setminus A)} = \overline{X \setminus A} \cap \bar{A} = \partial_X A$. \square

Proof of Lemma 4.8: Before we begin the actual proof let us recall that for an integer $0 \leq m \leq d$ the upper and lower Minkowski content of a set $B \subset \mathbb{R}^d$ is defined by

$$\begin{aligned} \mathcal{M}^{*m}(B) &:= \limsup_{\varepsilon \rightarrow 0^+} \frac{\lambda^d(B^{+\varepsilon})}{\sigma_{d-m} \varepsilon^{d-m}} \\ \mathcal{M}_*^m(B) &:= \liminf_{\varepsilon \rightarrow 0^+} \frac{\lambda^d(B^{+\varepsilon})}{\sigma_{d-m} \varepsilon^{d-m}}, \end{aligned}$$

where σ_{d-m} denotes the λ^{d-m} -volume of the unit Euclidean ball in \mathbb{R}^{d-m} . It is straightforward to check that these definitions coincide with those in (Federer, 1969, 3.2.37).

i). Since in the case $\lambda^d(\bar{A}) = 0$ there is nothing to prove, we restrict our considerations to the case $\lambda^d(\bar{A}) > 0$. Now, A is assumed to be bounded, and hence we have $\lambda^d(\bar{A}) < \infty$. The isoperimetric inequality in the form of (Federer, 1969, 3.2.43) thus yields

$$d\sigma_d^{1/d} \lambda^d(\bar{A})^{1-1/d} \leq \mathcal{M}_*^{d-1}(\partial A),$$

and consequently, there exists a $\delta_A > 0$, such that, for all $\delta \in (0, \delta_A]$, we have

$$\frac{d\sigma_d^{1/d} \lambda^d(\bar{A})^{1-1/d}}{2} \leq \frac{\lambda^d((\partial A)^{+\delta})}{\sigma_1 \delta} \leq \frac{\lambda^d(A^{+2\delta} \setminus A^{-2\delta})}{2\delta},$$

where in the last estimate we used part *viii)* of Lemma 6.5 and $\sigma_1 = 2$.

ii). Since ∂A is closed and $(d-1)$ -rectifiable in the sense of (Federer, 1969, 3.2.14), we find

$$\mathcal{M}^{*(d-1)}(\partial A) = \mathcal{H}^{d-1}(\partial A)$$

by (Federer, 1969, 3.2.39). Moreover, since ∂A is bounded, the boundary is contained in a compact set $X \subset \mathbb{R}^d$ such that the relative boundary $\partial_X A$ of A in X equals ∂A and the sets $A^{+\delta}$ and $A^{-\delta}$ considered in X equal the sets $A^{+\delta}$ and $A^{-\delta}$ when considered in \mathbb{R}^d for all $\delta \in (0, 1]$. By Lemma 6.10 there thus exists a $\delta_A > 0$ such that

$$\frac{\lambda^d(A^{+\delta} \setminus A^{-\delta})}{2\delta} \leq \frac{\lambda^d((\partial A)^{+\delta})}{\sigma_1 \delta} \leq 2\mathcal{H}^{d-1}(\partial A)$$

for all $\delta \in (0, \delta_A]$. □

Lemma 6.11. *Let the assumptions of Theorem 3.7 be satisfied. Then we have*

$$\begin{aligned} \mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) &\leq \mu(M_{\rho_D^* - \varepsilon}^{+\delta} \setminus M_{\rho_D^* - \varepsilon}) + \mu(\{\rho^* < h < \rho_D^* + \varepsilon\}) \\ &\quad + 2\mu(A_{\rho_D^* + \varepsilon}^1 \setminus (A_{\rho_D^* + \varepsilon}^1)^{-\delta}) + 2\mu(A_{\rho_D^* + \varepsilon}^2 \setminus (A_{\rho_D^* + \varepsilon}^2)^{-\delta}). \end{aligned}$$

Proof of Lemma 6.11: We will use Inequality (19) established in Theorem 3.7. To this end, we first observe that (5) implies

$$\mu(M_{\rho - \varepsilon}^{+\delta} \setminus \{h > \rho^*\}) = \mu\left(M_{\rho - \varepsilon}^{+\delta} \setminus \bigcup_{\rho' > \rho^*} M_{\rho'}\right) \leq \mu(M_{\rho - \varepsilon}^{+\delta} \setminus M_{\rho - \varepsilon}).$$

To bound the remaining terms on the right-hand side of (19), we further observe that the disjoint relation $A \cap B^{+\delta} = (A \cap (B^{+\delta} \setminus B)) \cup (A \cap B)$ applied to $B := X \setminus A_{\rho + \varepsilon}^i$ yields

$$\begin{aligned} \mu(A_i^* \setminus (A_{\rho + \varepsilon}^i)^{-\delta}) &= \mu(A_i^* \cap (X \setminus A_{\rho + \varepsilon}^i)^{+\delta}) \\ &= \mu(A_i^* \cap (X \setminus A_{\rho + \varepsilon}^i)^{+\delta} \cap A_{\rho + \varepsilon}^i) + \mu(A_i^* \setminus A_{\rho + \varepsilon}^i) \\ &= \mu(A_{\rho + \varepsilon}^i \cap (X \setminus A_{\rho + \varepsilon}^i)^{+\delta}) + \mu(A_i^* \setminus A_{\rho + \varepsilon}^i) \\ &= \mu(A_{\rho + \varepsilon}^i \setminus (A_{\rho + \varepsilon}^i)^{-\delta}) + \mu(A_i^* \setminus A_{\rho + \varepsilon}^i). \end{aligned}$$

Moreover, $A_{\rho + \varepsilon}^i \subset A_i^*$ and $A_1^* \cap A_2^* = \emptyset$ together with (4) and (5) imply

$$\mu(A_1^* \setminus A_{\rho + \varepsilon}^1) + \mu(A_2^* \setminus A_{\rho + \varepsilon}^2) = \mu((A_1^* \cup A_2^*) \setminus (A_{\rho + \varepsilon}^1 \cup A_{\rho + \varepsilon}^2)) = \mu(\{\rho^* < h < \rho + \varepsilon\}).$$

Combining all estimates with (19), we obtain the assertion. □

Proof of Lemma 4.9: Let us fix an $s > 0$. For $x \in \{0 < h - \rho < s\}$ we then find $d(x, \partial M_\rho)^\theta \leq cs$ by (35), that is $x \in (\partial M_\rho)^{+\delta}$ for $\delta := (cs)^{1/\theta}$. Using part *viii)* of Lemma 6.5, we conclude that $x \in M_\rho^{+2\delta} \setminus M_\rho^{-2\delta}$. In the case $2\delta \leq \delta_0$, we thus obtain

$$\mu(\{0 < h - \rho < s\}) \leq \mu(M_\rho^{+2\delta} \setminus M_\rho^{-2\delta}) \leq 2^\alpha c \delta^\alpha = 2^\alpha c^{1+\alpha/\theta} s^{\alpha/\theta},$$

and since μ is a finite measure, it is then easy to see that we can increase the constant on the right-hand side of this estimate so that it holds for all $s > 0$. □

Proof of Theorem 4.10: Since Assumption R includes the assumptions made in Theorem 4.4, it suffices to prove (36). Furthermore, recall that the proofs of Theorems 4.4 and 3.8 showed that the probability P^n of having a dataset $D \in X^n$ satisfying the assumptions of Theorem 3.6 is not less than $1 - e^{-\varsigma}$. For such D , Lemma 6.11 is applicable, and hence we obtain

$$\begin{aligned} \mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) &\leq \mu(M_{\rho_D^* - \varepsilon}^{+\delta} \setminus M_{\rho_D^* - \varepsilon}) + \mu(\{\rho^* < h < \rho_D^* + \varepsilon\}) \\ &\quad + 2\mu(A_{\rho_D^* + \varepsilon}^1 \setminus (A_{\rho_D^* + \varepsilon}^1)^{-\delta}) + 2\mu(A_{\rho_D^* + \varepsilon}^2 \setminus (A_{\rho_D^* + \varepsilon}^2)^{-\delta}) \\ &\leq \mu(M_{\rho_D^* - \varepsilon}^{+\delta} \setminus M_{\rho_D^* - \varepsilon}) + \mu(\{0 < h - \rho^* < \rho_D^* - \rho^* + \varepsilon\}) \\ &\quad + 4C_{\text{bound}}\delta^\alpha, \end{aligned}$$

where in the second estimate we used that the clusters have an α -smooth boundary by Assumption R. Moreover, the α -smooth boundaries also yield

$$\begin{aligned}\mu(M_{\rho_D^* - \varepsilon}^{+\delta} \setminus M_{\rho_D^* - \varepsilon}) &\leq \mu((A_{\rho_D^* - \varepsilon}^1)^{+\delta} \setminus M_{\rho_D^* - \varepsilon}) + \mu((A_{\rho_D^* - \varepsilon}^2)^{+\delta} \setminus M_{\rho_D^* - \varepsilon}) \\ &\leq \mu((A_{\rho_D^* - \varepsilon}^1)^{+\delta} \setminus A_{\rho_D^* - \varepsilon}^1) + \mu((A_{\rho_D^* - \varepsilon}^2)^{+\delta} \setminus A_{\rho_D^* - \varepsilon}^2) \\ &\leq 2c_{\text{bound}}\delta^\alpha.\end{aligned}$$

Finally, by (28) and the flatness exponent ϑ from Assumption R we find

$$\mu(\{0 < h - \rho^* < \rho_D^* - \rho^* + \varepsilon\}) \leq (c_{\text{flat}}(\rho_D^* - \rho^* + \varepsilon))^\vartheta \leq ((\tau/c_{\text{sep}})^\kappa + 7\varepsilon)^\vartheta.$$

Combining these three estimates, we then obtain (36). \square

Proof of Corollary 4.11: Clearly, our goal is to apply Theorem 4.10 for sufficiently large n . To this end, it suffices to check that the ε_n , δ_n , and τ_n satisfy the assumptions of Theorem 4.4 for $\varsigma_n := \ln n$ and all sufficiently large n . To this end, we observe that, for $\varsigma_n := \ln n$, we have

$$\begin{aligned}\varepsilon_n' &:= \sqrt{\frac{2c_{\text{part}}(1 + \|h\|_\infty)(\varsigma_n + \ln(2c_{\text{part}}) - d \ln \delta_n)}{\delta_n^d n}} + \frac{2c_{\text{part}}(\varsigma_n + \ln(2c_{\text{part}}) - d \ln \delta_n)}{3\delta_n^d n} \\ &\asymp \left(\frac{\ln n}{n}\right)^{\frac{\varrho}{2\varrho + \vartheta d}} (\ln \ln n)^{-\frac{\vartheta d}{4\varrho + 2\vartheta d}}.\end{aligned}$$

Using $\varepsilon_n \cdot \left(\frac{\ln n}{n}\right)^{-\frac{\varrho}{2\varrho + \vartheta d}} (\ln \ln n)^{\frac{\vartheta d}{4\varrho + 2\vartheta d}} \rightarrow \infty$, we then see that $\varepsilon_n \geq \varepsilon_n'$ for all sufficiently large n . Moreover, the remaining conditions on ε_n , δ_n , and τ_n from Theorem 4.4 are clearly satisfied for all sufficiently large n , and hence we can apply Theorem 4.10 for such n . This yields

$$\mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) \leq 6c_{\text{bound}}\delta_n^\alpha + (c_{\text{flat}}(\tau_n/c_{\text{sep}})^\kappa + 7c_{\text{flat}}\varepsilon_n)^\vartheta$$

with probability P^n not smaller than $1 - 1/n$ for all sufficiently large n . Some elementary calculations then show that

$$P^n \left(\left\{ D \in X^n : \mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) \leq K \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{\vartheta \varrho}{2\varrho + \vartheta d}} \right\} \right) \geq 1 - \frac{1}{n}$$

holds for a suitable constant K and all sufficiently large n . Moreover, since we always have

$$\mu(B_1(D) \triangle A_1^*) + \mu(B_2(D) \triangle A_2^*) \leq 2\mu(X) < \infty$$

it is an easy exercise to suitably increase K such that the desired inequality actually holds for all $n \geq 1$. \square

6.6 Proofs Related to Adaptivity

Proof of Theorem 5.1: We first observe that $C^2 \ln(\ln n) \geq 2(1 + \|h\|_\infty)$ guarantees that all $\varepsilon_{\delta,n}$ satisfy (21) for $\varsigma' := \varsigma + \ln |\Delta|$. Consequently, Theorem 4.4, namely (27) and (28), yields

$$P^n \left(\left\{ D \in X^n : \varepsilon_{\delta,n} < \rho_{D,\delta}^* - \rho^* \leq (\tau_{\delta,n}^\gamma/c_{\text{sep}})^\kappa + 6\varepsilon_{\delta,n} \right\} \right) \geq 1 - e^{-\varsigma - \ln |\Delta|}.$$

for all $\delta \in \Delta$. Applying the union bound, we thus find

$$P^n \left(\left\{ D \in X^n : \varepsilon_{\delta,n} < \rho_{D,\delta}^* - \rho^* \leq (\tau_{\delta,n}^\gamma/c_{\text{sep}})^\kappa + 6\varepsilon_{\delta,n} \text{ for all } \delta \in \Delta \right\} \right) \geq 1 - e^{-\varsigma}.$$

Let us now consider a data set $D \in X^n$ such that $\varepsilon_{\delta,n} < \rho_{D,\delta}^* - \rho^* \leq (\tau_{\delta,n}^\gamma/c_{\text{sep}})^\kappa + 6\varepsilon_{\delta,n}$ for all $\delta \in \Delta$. Then, the definitions of $\rho_{D,\Delta}^*$ and $\varepsilon_{D,\Delta}$, see (39), imply

$$\rho_{D,\Delta}^* - \rho^* = \min_{\delta \in \Delta} \rho_{D,\delta}^* - \rho^* \in \left(\min_{\delta \in \Delta} \varepsilon_{\delta,n}, \min_{\delta \in \Delta} ((\tau_{\delta,n}^\gamma/c_{\text{sep}})^\kappa + 6\varepsilon_{\delta,n}) \right]$$

and $\varepsilon_{D,\Delta} = \varepsilon_{\delta_{D,\Delta}^*,n}^* < \rho_{D,\delta_{D,\Delta}^*}^* - \rho^* = \rho_{D,\Delta}^* - \rho^*$, that is, we have shown the first assertion. To show the remaining assertions, we first observe that a literal repetition of the argument above, in which we only replace the use of (27) by that of (29), yields

$$P^n \left(\left\{ D \in X^n : c_1 \tau_{\delta,n}^k + \varepsilon_{\delta,n} < \rho_{D,\delta}^* - \rho^* \leq c_2 \tau_{\delta,n}^k + 6\varepsilon_{\delta,n} \text{ for all } \delta \in \Delta \right\} \right) \geq 1 - e^{-\varsigma}.$$

Using (39) we then immediately obtain the second assertion, while considering $\delta = \delta_{D,\Delta}^*$ gives the third assertion. \square

Proof of Corollary 5.2: Let us fix an $n \geq 16$. For later use we note that this choice implies $I_n \subset (0, 1]$. For $\varsigma := \ln n$, we further see that the definition of $\varepsilon_{\delta,n}$ is consistent with (37). Our first goal is to show that we can apply Theorem 5.1 for sufficiently large n . To this end, we first observe that $\max \Delta_n = (\ln \ln n)^{-d} \rightarrow 0$ for $n \rightarrow \infty$, and hence we obtain $\Delta_n \subset (0, \delta_{\text{thick}}]$ for all sufficiently large n . Analogously, $\max \Delta_n \ln \ln n \rightarrow 0$ implies $\max_{\delta \in \Delta_n} (\tau_{\delta,n}/c_{\text{sep}})^k \leq (\rho^{**} - \rho^*)/18$ for all sufficiently large n , and the definition of $\tau_{\delta,n}$ ensures $\min_{\delta \in \Delta_n} \tau_{\delta,n} \geq 2\psi(\delta)$ for all sufficiently large n . Let us now show that eventually we also have $\max_{\delta \in \Delta_n} \varepsilon_{\delta,n} \leq (\rho^{**} - \rho^*)/18$. To this end, note that the derivative of the function $g_n : (0, \infty) \rightarrow \mathbb{R}$ defined by

$$g_n(\delta) := \frac{\ln(2c_{\text{part}}|\Delta_n|n) - d \ln \delta}{\delta^d n}$$

is given by

$$g_n'(\delta) = -\frac{d(1 + \ln(2c_{\text{part}}|\Delta_n|n) - d \ln \delta)}{\delta^{1+d} n},$$

and using $c_{\text{part}} \geq 1$, we thus find that g_n is monotonically decreasing on $(0, 1]$ for all $n \geq 1$. In addition, using $|\Delta_n| \leq n$ we obtain

$$\begin{aligned} g_n(\min I_n) &= g_n \left(\left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{1}{d}} \right) = \frac{\ln(2c_{\text{part}}|\Delta_n|n) + \ln n - \ln \ln n - 2 \ln \ln \ln n}{\ln n \cdot (\ln \ln n)^2} \\ &\leq \frac{4 \ln n - \ln \ln n - 2 \ln \ln \ln n}{\ln n \cdot (\ln \ln n)^2} \\ &\leq \frac{4}{(\ln \ln n)^2} \end{aligned}$$

for all $n \geq \max\{16, 2c_{\text{part}}\}$, and hence $g_n(\min I_n) \ln \ln n \rightarrow 0$ for $n \rightarrow \infty$. Since the definition of $\varepsilon_{\delta,n}$ gives $\varepsilon_{\delta,n} = C \sqrt{c_{\text{part}} g_n(\delta) \ln \ln n} + \frac{2}{3} c_{\text{part}} g_n(\delta)$, we can thus conclude that

$$\begin{aligned} \max_{\delta \in \Delta_n} \varepsilon_{\delta,n} &\leq \max_{\delta \in \Delta_n} C \sqrt{c_{\text{part}} g_n(\delta) \ln \ln n} + \max_{\delta \in \Delta_n} c_{\text{part}} g_n(\delta) \\ &\leq C \sqrt{c_{\text{part}} g_n(\min I_n) \ln \ln n} + c_{\text{part}} g_n(\min I_n) \\ &\rightarrow 0 \end{aligned}$$

for $n \rightarrow \infty$. This ensures the desired $\max_{\delta \in \Delta_n} \varepsilon_{\delta,n} \leq (\rho^{**} - \rho^*)/18$ for all sufficiently large n . Combining this with our previous estimate, we find

$$\max_{\delta \in \Delta_n} ((\tau_{\delta,n}/c_{\text{sep}})^k + \varepsilon_{\delta,n}) \leq (\rho^{**} - \rho^*)/9$$

for all sufficiently large n , and therefore we can indeed apply Theorem 5.1 for such n .

Before we proceed, let us now fix an $n \geq 16$ and assume that without loss of generality that Δ_n is of the form $\Delta = \{\delta_1, \dots, \delta_m\}$ with $\delta_{i-1} < \delta_i$ for all $i = 2, \dots, m$. We write $\delta_0 := \min I_n$ and $\delta_{m+1} := \max I_n$. Our intermediate goal is to show that

$$\delta_i - \delta_{i-1} \leq 2n^{-1/d}, \quad i = 1, \dots, m+1. \quad (63)$$

To this end, we fix an $i \in \{1, \dots, m\}$ and write $\bar{\delta} := (\delta_i + \delta_{i-1})/2 \in I_n$. Since Δ_n is an $n^{-1/d}$ -net of I_n , we then have $\delta_i - \bar{\delta} \leq n^{-1/d}$ or $\bar{\delta} - \delta_{i-1} \leq n^{-1/d}$, and from both (63) follows. Moreover, to

show (63) in the case $i = m + 1$, we first observe that there exists an $\delta_i \in \Delta_n$ with $\delta_i - \delta_m \leq n^{-1/d}$ since Δ_n is an $n^{-1/d}$ -net of I_n . Using our ordering of Δ_n , we can assume without loss of generality that $i = m$, which immediately implies (63).

Let us now prove the first assertion in the case $\kappa < \infty$. To this end, we write

$$\delta_n^* := \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{1}{2\gamma\kappa+d}},$$

where we note that for sufficiently large n we have $\delta_n^* \in I_n$. In the following we thus restrict our considerations to such n . Then there exists an index $i \in \{1, \dots, m+1\}$ such that $\delta_{i-1} \leq \delta_n^* \leq \delta_i$, and by (63) we conclude that $\delta_n^* \leq \delta_i \leq \delta_n^* + 2n^{-1/d}$. Clearly, this yields

$$\begin{aligned} \min_{\delta \in \Delta_n} (c_2 \tau_{\delta,n}^\kappa + 6\varepsilon_{\delta,n}) &= \min_{\delta \in \Delta_n} (c_2 \delta^{\gamma\kappa} (\ln \ln \ln n)^\kappa + 6\varepsilon_{\delta,n}) \\ &\leq c_2 \delta_i^{\gamma\kappa} (\ln \ln \ln n)^\kappa + 6\varepsilon_{\delta_i,n} \\ &\leq c_2 (\delta_n^* + 2n^{-1/d})^{\gamma\kappa} (\ln \ln \ln n)^\kappa + 6\varepsilon_{\delta_i,n} \\ &\leq 6c_2 \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{1}{2\gamma\kappa+d}} + 6\varepsilon_{\delta_i,n} \end{aligned} \quad (64)$$

for all sufficiently large n , where $c_2 := \underline{c}_{\text{sep}}^{-\kappa}$ is the constant from Theorem 5.1. Moreover, using $|\Delta_n| \leq n$ and the monotonicity of g_n , we further find

$$\begin{aligned} g_n(\delta_i) \leq g_n(\delta_n^*) &= \frac{\ln(2c_{\text{part}}|\Delta_n|n) - d \ln \delta_n^*}{(\delta_n^*)^d n} \leq \frac{\ln(2c_{\text{part}}) + 2 \ln n - d \ln \delta_n^*}{(\delta_n^*)^d n} \\ &\leq \frac{4 \ln n}{(\delta_n^*)^d n} \\ &\leq \frac{4}{(\ln \ln n)^{\frac{d}{2\gamma\kappa+d}}} \cdot \left(\frac{\ln n}{n} \right)^{\frac{2\gamma\kappa}{2\gamma\kappa+d}} \end{aligned} \quad (65)$$

for all sufficiently large n . By the relation between $\varepsilon_{\delta,n}$ and $g_n(\delta)$ we then find

$$\varepsilon_{\delta_i,n} \leq 2C \sqrt{c_{\text{part}}} \left(\frac{\ln n \cdot \ln \ln n}{n} \right)^{\frac{\gamma\kappa}{2\gamma\kappa+d}} + 3c_{\text{part}} \left(\frac{\ln n}{n} \right)^{\frac{2\gamma\kappa}{2\gamma\kappa+d}},$$

and combining this estimate with (64) and Theorem 5.1, we obtain the first assertion in the case $\kappa < \infty$.

Let us now consider the case $\kappa = \infty$ for sufficiently large n . To this end, we fix a sample size n such that

$$\delta_n^* := \left(\frac{1}{\ln \ln n} \right)^{\frac{1}{d}}$$

satisfies $(\delta_n^* + 2n^{-1/d})^\gamma \ln \ln \ln n < \underline{c}_{\text{sep}}$, and thus $((\delta_n^* + 2n^{-1/d})^\gamma \ln \ln \ln n / \underline{c}_{\text{sep}})^\kappa = 0$. Since $\delta_n^* \in I_n$ there also exists an index $i \in \{1, \dots, m+1\}$ such that $\delta_{i-1} \leq \delta_n^* \leq \delta_i$, and by (63) we again conclude $\delta_n^* \leq \delta_i \leq \delta_n^* + 2n^{-1/d}$. Clearly, the latter implies

$$\begin{aligned} \min_{\delta \in \Delta_n} ((\tau_{\delta,n} / \underline{c}_{\text{sep}})^\kappa + 6\varepsilon_{\delta,n}) &\leq (\delta_i^\gamma \ln \ln \ln n / \underline{c}_{\text{sep}})^\kappa + 6\varepsilon_{\delta_i,n} \leq ((\delta_n^* + 2n^{-1/d})^\gamma \ln \ln \ln n / \underline{c}_{\text{sep}})^\kappa + 6\varepsilon_{\delta_i,n} \\ &= 6\varepsilon_{\delta_i,n} \end{aligned}$$

by our assumptions on n . Analogously to (65) we further find, for sufficiently large n , that

$$g_n(\delta_i) \leq g_n(\delta_n^*) \leq \frac{3 \ln n - d \ln \delta_n^*}{(\delta_n^*)^d n} = \frac{3 \ln n + \ln \ln \ln n}{n (\ln \ln n)^{-1}} \leq 4 \frac{\ln n \cdot \ln \ln n}{n},$$

and by the relation between $\varepsilon_{\delta,n}$ and $g(\delta)$ we then find the assertion with the help of Theorem 5.1.

Let us finally prove (41). To this end we first recall that we have already seen that, for sufficiently large n , we can apply Theorem 5.1. Consequently, it suffices to find a lower bound for the right-hand-side of

$$\min_{\delta \in \Delta_n} (c_1 \tau_{\delta,n}^\kappa + \varepsilon_{\delta,n}) \geq \min\{1, c_1\} \cdot \min_{\delta \in \Delta_n} (\tau_{\delta,n}^\kappa + \varepsilon_{\delta,n}), \quad (66)$$

where c_1 is the constant appearing in Theorem 5.1. Now, for $n \geq 16$, we have $I_n \subset (0, 1]$, and thus we find $\delta \in (0, 1]$ for all $\delta \in \Delta_n$. For sufficiently large n this yields

$$\begin{aligned} \min_{\delta \in \Delta_n} (\tau_{\delta,n}^\kappa + \varepsilon_{\delta,n}) &= \min_{\delta \in \Delta_n} \left(\delta^{\gamma\kappa} (\ln \ln n)^\kappa + C \sqrt{c_{\text{part}} g_n(\delta) \ln \ln n} + \frac{2}{3} c_{\text{part}} g_n(\delta) \right) \\ &\geq \min_{\delta \in \Delta_n} \left(\delta^{\gamma\kappa} + C \sqrt{c_{\text{part}} g_n(\delta) \ln \ln n} \right) \\ &\geq \min_{\delta \in \Delta_n} \left(\delta^{\gamma\kappa} + C \sqrt{\frac{c_{\text{part}} \ln n \cdot \ln \ln n}{\delta^{d_n}}} \right) \\ &\geq \min_{\delta \in (0,1]} \left(\delta^{\gamma\kappa} + C \sqrt{\frac{c_{\text{part}} \ln n \cdot \ln \ln n}{\delta^{d_n}}} \right). \end{aligned}$$

An elementary application of calculus then yields the assertion. \square

Proof of Corollary 5.3: We have seen in the proof of Corollary 5.2 that for sufficiently large n Inequality (40) follows from the fact that the procedure satisfies the assumptions of Theorem 5.1 for such n and $\varsigma := \ln n$. Consequently, for sufficiently large n , the probability P^n of having a data set $D \in X^n$ satisfying both (40) and the third inequality of Theorem 5.1 is not less than $1 - 1/n$. Let us fix such a data set D . Then we have

$$c_1 \tau_{D,\Delta}^\kappa + \varepsilon_{D,\Delta} \leq \rho_{D,\Delta_n}^* - \rho^* \leq \bar{K} \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{\gamma\kappa}{2\gamma\kappa+d}}. \quad (67)$$

Moreover, an elementary estimate yields

$$c_1 \tau_{D,\Delta}^\kappa + \varepsilon_{D,\Delta} \geq \min\{1/7, c_1 \underline{c}_{\text{sep}}^\kappa\} \cdot ((\tau_{D,\Delta}/\underline{c}_{\text{sep}})^\kappa + 7\varepsilon_{D,\Delta}),$$

and setting $c := \min\{1/7, c_1 \underline{c}_{\text{sep}}^\kappa\}$, we hence obtain

$$(\tau_{D,\Delta}/\underline{c}_{\text{sep}})^\kappa + 7\varepsilon_{D,\Delta} \leq c^{-1} \bar{K} \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{\gamma\kappa}{2\gamma\kappa+d}}. \quad (68)$$

In addition, for sufficiently large n , Inequality (67) implies

$$\delta_{D,\Delta}^* \leq \tau_{D,\Delta}^{1/\gamma} \leq (4\bar{K})^{\frac{1}{\gamma\kappa}} (6\bar{c}_{\text{sep}})^{\frac{1}{\gamma}} \left(\frac{\ln n \cdot (\ln \ln n)^2}{n} \right)^{\frac{1}{2\gamma\kappa+d}}. \quad (69)$$

Now we have already seen in the proof of Theorem 5.1 and Corollary 5.2 that, for sufficiently large n , the assumptions on δ , $\varepsilon_{\delta,n}$, $\varepsilon_{\delta,n}^*$, τ_n , $\varsigma_n := \ln n$, and n of Theorem 4.4 are satisfied for all $\delta \in \Delta_n$ simultaneously. Consequently, we can combine (68) and (69) with Theorem 4.10 to obtain the assertion. \square

References

- A. Baíllo, A. Cuevas, and A. Justel. Set estimation and nonparametric detection. *Canad. J. Statist.*, 28:765–782, 2000.
- A. Baíllo, J.A. Cuesta-Albertos, and A. Cuevas. Convergence rates in nonparametric estimation of level sets. *Statist. Probab. Lett.*, 53:27–35, 2001.
- S. Ben-David and M. Lindenbaum. Learning distributions by their density levels: a paradigm for learning without a teacher. *J. Comput. System Sci.*, 55:171–182, 1997.
- J.W. Carmichael, G.A. George, and R.S. Julius. Finding natural clusters. *Systematic Zoology*, 17: 144–150, 1968.
- K. Chaudhuri and S. Dasgupta. Rates of convergence for the cluster tree. In J. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 343–351. 2010.

- A. Cuevas. Set estimation: another bridge between statistics and geometry. *Boletín de Estadística e Investigación Operativa*, 25:71–85, 2009.
- A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *Ann. Statist.*, 25:2300–2312, 1997.
- A. Cuevas, M. Febrero, and R. Fraiman. Estimating the number of clusters. *Canad. J. Statist.*, 28:367–382, 2000.
- L. Devroye and G.L. Wise. Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.*, 38:480–488, 1980.
- L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, New York, 1996.
- H. Federer. *Geometric Measure Theory*. Springer, Berlin, 1969.
- E. Giné and A. Guillaou. Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.*, 38:907–921, 2002. En l’honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.
- J. A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, New York, 1975.
- J. A. Hartigan. Estimation of a convex density contour in 2 dimensions. *J. Amer. Statist. Assoc.*, 82:267–270, 1987.
- J.A. Hartigan. Consistency of single linkage for high-density clusters. *J. Amer. Statist. Assoc.*, 76:388–394, 1981.
- J. L. Kelley. *General Topology*. D. Van Nostrand, Toronto, 1955.
- S. Kpotufe and U. von Luxburg. Pruning nearest neighbor cluster trees. In L. Getoor and T. Scheffer, editors, *Proceedings of the 28th International Conference on Machine Learning*, pages 225–232. ACM, New York, NY, USA, 2011.
- M. Maier, M. Hein, and U. von Luxburg. Optimal construction of k-nearest neighbor graphs for identifying noisy clusters. *Theoret. Comput. Sci*, 410:1749–1764, 2009.
- D. W. Müller and G. Sawitzki. Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.*, 86:738–746, 1991.
- W. Polonik. Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.*, 23:855–881, 1995.
- P. Rigollet. Generalization error bounds in semi-supervised classification under the cluster assumption. *J. Mach. Learn. Res.*, 8:1369–1392, 2007.
- P. Rigollet and R. Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15:1154–1178, 2009.
- A. Rinaldo and L. Wasserman. Generalized density clustering. *Ann. Statist.*, 38:2678–2722, 2010.
- A. Rinaldo, A. Singh, R. Nugent, and L. Wasserman. Stability of density-based clustering. *J. Mach. Learn. Res.*, 13:905–948, 2012.
- J. Serra. *Image Analysis and Mathematical Morphology*, volume Vol. 1. Academic Press, San Diego, 1982.
- A. Singh, C. Scott, and R. Nowak. Adaptive Hausdorff estimation of density level sets. *Ann. Statist.*, 37:2760–2782, 2009.

- B.K. Sriperumbudur and I. Steinwart. Consistency and rates for clustering with DBSCAN. In N. Lawrence and M. Girolami, editors, *JMLR Workshop and Conference Proceedings Volume 22: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics 2012*, pages 1090–1098. 2012.
- I. Steinwart. Adaptive density level set clustering. In S. Kakade and U. von Luxburg, editors, *JMLR Workshop and Conference Proceedings Volume 19: Proceedings of the 24th Conference on Learning Theory 2011*, pages 703–738. 2011.
- I. Steinwart, D. Hush, and C. Scovel. A classification framework for anomaly detection. *J. Mach. Learn. Res.*, 6:211–232, 2005.
- W. Stuetzle. Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample. *Journal of Classification*, 20:25–47, 2003.
- W. Stuetzle and R. Nugent. A generalized single linkage method for estimating the cluster tree of a density. *Journal of Computational and Graphical Statistics*, 19:397–418, 2010.
- A. B. Tsybakov. On nonparametric estimation of density level sets. *Ann. Statist.*, 25:948–969, 1997.
- G. Walther. Granulometric smoothing. *Ann. Statist.*, 25:2273–2299, 1997.
- G. Walther. On a generalization of Blaschke’s rolling theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.*, 22:301–316, 1999.

7 Appendix: Continuous Densities in two Dimensions

In this appendix we present a couple of two-dimensional examples that show that the assumptions imposed in this paper are not only met by many discontinuous densities, but also by many continuous densities.

We begin with an example of a set $A \subset \mathbb{R}^2$, for which we can compute $A^{\oplus\delta}$ and $A^{\ominus\delta}$ explicitly. This example will be the base of all further examples.

Example 7.1. Let $X := [-1, 1] \times [-2, 2]$ be equipped with the metric defined by the supremums norm. Furthermore, for $x^\pm \in (-0.6, -0.4)$ and $x_\pm^\pm \in (0.4, 0.6)$ we fix two continuous functions $f^-, f^+ : [-1, 1] \rightarrow [-1, 1]$ such that f^+ is increasing on $[-1, x^\pm] \cup [0, x_\pm^+]$ and decreasing on $[x^\pm, 0] \cup [x_\pm^+, 1]$, while f^- is decreasing on $[-1, x^\pm] \cup [0, x_\pm^-]$ and increasing on $[x^\pm, 0] \cup [x_\pm^-, 1]$. In addition, assume that $\{f^- < 0\} = \{f^+ > 0\}$ and $\{f^- = 0\} = \{f^+ = 0\}$ as well as $f^-(\pm 0.5) < 0$ and $f^+(\pm 0.5) > 0$. Now consider the (non-empty) set A enveloped by f^\pm , that is

$$A := \{(x, y) \in X : x \in [-1, 1] \text{ and } f^-(x) \leq y \leq f^+(x)\}.$$

To describe $A^{\ominus\delta}$ for $\delta \in (0, 0.1]$, we define functions $f_{-\delta}^\pm : [-1, 1] \rightarrow [-1, 1]$ by

$$f_{-\delta}^\pm(x) := \begin{cases} f^\pm(-1) & \text{if } x \in [-1, -1 + \delta] \\ f^\pm(0) & \text{if } x \in [-\delta, +\delta] \\ f^\pm(1) & \text{if } x \in [1 - \delta, 1] \end{cases}$$

and $f_{-\delta}^-(x) := f^-(x - \delta) \vee f^-(x + \delta)$, respectively $f_{-\delta}^+(x) := f^+(x - \delta) \wedge f^+(x + \delta)$ for the remaining $x \in [-1, 1]$. Then we have

$$A^{\ominus\delta} = \{(x, y) \in X : x \in [-1, 1] \text{ and } f_{-\delta}^-(x) + \delta \leq y \leq f_{-\delta}^+(x) - \delta\}.$$

Moreover, to describe $A^{\oplus\delta}$, we define the

$$\begin{aligned} x_{0,-1} &:= \min\{x \in [-1, -0.5] : f^+(x) - f^-(x) \geq 0\} \\ x_{0,-0} &:= \max\{x \in [-0.5, 0] : f^+(x) - f^-(x) \geq 0\} \\ x_{0,+0} &:= \min\{x \in [0, 0.5] : f^+(x) - f^-(x) \geq 0\} \\ x_{0,+1} &:= \max\{x \in [0.5, 1] : f^+(x) - f^-(x) \geq 0\}, \end{aligned}$$

where we note that the minima are actually attained by the continuity of f^\pm and the fact that all sets are non-empty. Furthermore, we define two functions $f_{+\delta}^\pm : [-1, 1] \rightarrow [-1, 1]$ by

$$f_{+\delta}^\pm(x) := \begin{cases} f^\pm(x + \delta) & \text{if } x \in [-1 \vee (x_{0,-1} - \delta), x^\pm - \delta] \\ f^\pm(x_\pm^\pm) & \text{if } x \in [x^\pm - \delta, x_\pm^\pm + \delta] \\ f^\pm(x_\pm^\pm) & \text{if } x \in [x_\pm^\pm - \delta, x_\pm^\pm + \delta] \\ f^\pm(x - \delta) & \text{if } x \in [x_\pm^\pm + \delta, (x_{0,+1} + \delta) \wedge 1] \end{cases}$$

as well as

$$\begin{aligned} f_{+\delta}^-(x) &:= f^-(x - \delta) \wedge f^-(x + \delta) \\ f_{+\delta}^+(x) &:= f^+(x - \delta) \vee f^+(x + \delta) \end{aligned}$$

for $x \in [x^\pm + \delta, x_\pm^\pm - \delta] \setminus (x_{0,-0} + \delta, x_{0,+0} - \delta)$ and $f_{+\delta}^\pm(x) := -2\delta$ for the remaining $x \in [-1, 1]$. Then we have

$$A^{\oplus\delta} = \{(x, y) \in X : x \in [-1, 1] \text{ and } f_{+\delta}^-(x) - \delta \leq y \leq f_{+\delta}^+(x) + \delta\}.$$

Finally, we have $|\mathcal{C}(A)| \leq 2$ with $|\mathcal{C}(A)| = 2$ if and only if $x_{0,-0} < x_{0,+0}$, and in the latter case we further have $\tau_A^* = x_{0,+0} - x_{0,-0}$.

Proof of Example 7.1: Let us fix a $\delta \in (0, 1/10]$. To simplify notations, we further write $g^- := f^-_{-\delta} + \delta$ and $g^+ := f^+_{-\delta} - \delta$.

Proof of “ $A^{\oplus\delta} \subset \dots$ ”. By the relation $A^{\oplus\delta} = X \setminus (X \setminus A)^{\oplus\delta}$ it suffices to show that

$$\{(x, y) \in X : x \in [-1, 1] \text{ and } (y < g^-(x) \text{ or } y > g^+(x))\} \subset (X \setminus A)^{\oplus\delta}.$$

By symmetry, it further suffices to consider the case $x \geq 0$ and $y > g^+(x)$. Moreover, to show the inclusion above, it finally suffices to find $x' \in [-1, 1]$ and $y' \in [-2, 2]$ with $|x - x'| \leq \delta$, $|y - y'| \leq \delta$ and $y' > f^+(x')$. However, this task is straightforward. Indeed, we can always set $y' := (y + \delta) \wedge 2$, and if $x \in [0, \delta]$ then $x' := 0$ works, since $y' = (y + \delta) \wedge 2 > g^+(x) + \delta = f^+(0) = f^+(x')$, while for $x \in [1 - \delta, 1]$, the choice $x' := 1$ does by an analogous argument. Finally, if $x \in (\delta, 1 - \delta)$, we set $x' := x - \delta$ if $g^+(x) = f^+(x - \delta) - \delta$ and $x' := x + \delta$ if $g^+(x) = f^+(x + \delta) - \delta$.

Proof of “ $A^{\oplus\delta} \supset \dots$ ”. Again, it suffices to consider $x \geq 0$ due to symmetry. Let us fix an y with $g^-(x) \leq y \leq g^+(x)$. Then, our goal is to show that $(x, y) \notin (X \setminus A)^{\oplus\delta}$, that is,

$$\|(x, y) - (x', y')\|_{\infty} > \delta \tag{70}$$

for all $(x', y') \in X \setminus A$. In the following, we thus fix a pair $(x', y') \in X \setminus A$ for which (70) is not true and show that this leads to a contradiction. We begin by considering the case $x \in [0, \delta]$. Since (70) is not true, we find $|x - x'| \leq \delta$, and hence $x_{-}^{\pm} \leq x' \leq x_{+}^{\pm}$. Then, if $y' > f^+(x')$, this leads to

$$y \leq g^+(x) = f^+(0) - \delta \leq f^+(x') - \delta < y' - \delta,$$

which contradicts the assumed $|y - y'| \leq \delta$. The case $y' < f^-(x')$ analogously leads to a contradiction. Let us now consider the case $x \in [1 - \delta, 1]$. Then $|x - x'| \leq \delta$ implies $x' \geq x_{+}^{\pm}$. Consequently, $y' > f^+(x')$ leads to another contradiction by

$$y \leq g^+(x) = f^+(1) - \delta \leq f^+(x') - \delta < y' - \delta,$$

and the case $y' < f^-(x')$ can be treated analogously. It thus remains to consider the case $x \in [\delta, 1 - \delta]$. Then $|x - x'| \leq \delta$ implies $x - \delta \leq x' \leq x + \delta$. For $x' \leq x_{+}^{\pm}$ we thus find $f^+(x - \delta) \leq f^+(x')$, while for $x' \geq x_{+}^{\pm}$ we find $f^+(x + \delta) \leq f^+(x')$. Consequently, for $y' > f^+(x')$ we obtain a contradiction by

$$y \leq g^+(x) = (f^+(x - \delta) \wedge f^+(x + \delta)) - \delta \leq f^+(x') - \delta < y' - \delta,$$

and, again, the case $y' < f^-(x')$ can be shown similarly.

Proof of “ $A^{\oplus\delta} \subset \dots$ ”. Let us fix a pair $(x, y) \in A^{\oplus\delta}$. Without loss of generality we restrict our considerations to the case $y \geq 0$ and $x \in [-1, 0]$. To show that $y \leq f^+_{+\delta}(x) + \delta$ we assume the converse, that is $y > f^+_{+\delta}(x) + \delta$. Since $(x, y) \in A^{\oplus\delta}$ we then find $(x', y') \in A$ with $\|(x, y) - (x', y')\|_{\infty} \leq \delta$. From the latter we infer that both $x - \delta \leq x' \leq x + \delta$ and

$$y' \geq y - \delta > f^+_{+\delta}(x). \tag{71}$$

Now suppose that $x \in [-1, -1 \vee (x_{0,-1} - \delta)]$. Then we obtain a contradiction since $(x', y') \in A$ implies $x \geq x' - \delta \geq x_{0,-1} - \delta$. Moreover, for $x \in [-1 \vee (x_{0,-1} - \delta), x_{-}^{\pm} - \delta]$, we obtain

$$f^+_{+\delta}(x) = f^+(x + \delta) \geq f^+(x') \geq y',$$

which contradicts (71). Analogously, in the case $x \in [x_{-}^{\pm} - \delta, x_{-}^{\pm} + \delta]$ we obtain a contradiction from

$$f^+_{+\delta}(x) = f^+(x_{-}^{\pm}) \geq f^+(x') \geq y'.$$

Moreover, for $x \in [x_{-}^{\pm} + \delta, 0 \wedge (x_{0,-0} + \delta)]$ we have

$$f^+_{+\delta}(x) = f^+(x - \delta) \vee f^+(x + \delta) \geq f^+(x - \delta) \geq f^+(x') \geq y'$$

which again contradicts (71). Finally, if $x \in (0 \wedge x_{0,-0} + \delta, 0]$ we obtain a contradiction from $x > x_{0,-0} + \delta \geq x' + \delta$.

Proof of “ $A^{\oplus\delta} \supset \dots$ ”. Let us fix a pair $(x, y) \in X$ with $f_{+\delta}^-(x) - \delta \leq y \leq f_{+\delta}^+(x) + \delta$. Without loss of generality we again consider the case $y \geq 0$ and $x \in [-1, 0]$, only. To show $(x, y) \in A^{\oplus\delta}$ we need to find a pair $(x', y') \in A$ with $\|(x, y) - (x', y')\|_\infty \leq \delta$. Let us assume that we have found an x' with $|x - x'| \leq \delta$ and $f(x') \geq y - \delta$. For y' defined by

$$y' := f(x') \wedge (y + \delta)$$

we then immediately obtain $y' \leq y + \delta$. Moreover, if we actually have $y' = y + \delta$, then we clearly obtain $|y - y'| \leq \delta$, while in the case $y' < y + \delta$ we find $y' = f(x') \geq y - \delta$, that is again $|y - y'| \leq \delta$. Consequently, it suffices to find an x' with the properties above. To this end, we first observe that we can exclude the case $x \in [-1, -1 \vee (x_{0,-1} - \delta))$, since for such x we have $0 \leq y \leq f_{+\delta}^+(x) + \delta = -\delta$. Analogously, we can exclude the case $x \in (0 \wedge (x_{0,-0} + \delta), 0]$. Let us now consider the case $x \in [-1 \vee (x_{0,-1} - \delta), x_-^+ - \delta]$. For $x' := x + \delta$ we then have

$$f(x') = f(x + \delta) = f_{+\delta}^+(x) \geq y - \delta,$$

and hence x' satisfies the desired properties. Moreover, for $x \in [x_-^+ - \delta, x_-^+ + \delta]$ we define $x' := x_-^+$, which clearly gives $|x - x'| \leq \delta$. In addition, we again have

$$f(x') = f(x_-^+) = f_{+\delta}^+(x) \geq y - \delta.$$

Finally, let us consider the case $x \in [x_-^+ + \delta, 0 \wedge (x_{0,-0} + \delta)]$. Let us first assume that $f(x - \delta) \geq f(x + \delta)$. For $x' := x - \delta$ we then obtain

$$f(x') = f(x - \delta) = f_{+\delta}^+(x) \geq y - \delta.$$

Analogously, if $f(x - \delta) \leq f(x + \delta)$, then $x' := x + \delta$ has the desired properties.

Finally, $|\mathcal{C}(A)| \leq 2$ is obvious, and so is the equivalence between $|\mathcal{C}(A)| = 2$ and $x_{0,-0} < x_{0,+0}$. In the latter case, $A_1 := \{(x, y) \in A : x \leq x_{0,-0}\}$ and $A_2 := \{(x, y) \in A : x \geq x_{0,+0}\}$ are the two components of A , and from this it is easy to conclude that $\tau_A^* = x_{0,+0} - x_{0,-0}$. \square

Our next example shows how to estimate the function ψ_A^* for the sets considered in Example 7.1

Example 7.2. *Let us consider the situation of Example 7.1. To simplify the presentation, let us additionally assume that the monotonicity of f^+ and f^- is actually strict and that A has sufficiently thick parts on both sides of the y -axis in the sense of*

$$[-0.8, -0.2] \cup [0.2, 0.8] \subset \{f^- \leq -0.2\} \cap \{f^+ \geq 0.2\}. \quad (72)$$

Note that, for all $\delta \in (0, 0.1]$, this condition in particular ensures that $A^{\ominus\delta}$ contains open neighborhoods around the points $(-0.5, 0)$ and $(0, 0.5)$. Moreover, for $\delta \in [0, 0.1]$ we define

$$\begin{aligned} x_{\delta,-1} &:= \min\{x \in [-1, -0.8] : f^+(x) - f^-(x) \geq 2\delta\} \\ x_{\delta,-0} &:= \max\{x \in [-0.2, 0] : f^+(x) - f^-(x) \geq 2\delta\} \\ x_{\delta,+0} &:= \min\{x \in [0, 0.2] : f^+(x) - f^-(x) \geq 2\delta\} \\ x_{\delta,+1} &:= \max\{x \in [0.8, 1] : f^+(x) - f^-(x) \geq 2\delta\}, \end{aligned}$$

where we note that the minima and maxima are actually attained by (72) and the assumed continuity of f^\pm , and for the same reason we further have $x_{\delta,-1} < -0.8$, $x_{\delta,-0} > -0.2$, $x_{\delta,+0} < 0.2$, and $x_{\delta,+1} > 0.8$. Then, the function $f_{-\delta}^+$ has exactly two local maxima $x_{\delta,-}^+$ and $x_{\delta,+}^+$, satisfying $x_{\delta,-}^+ \in [-1, 0]$ and $x_{\delta,+}^+ \in [0, 1]$, and the function $f_{-\delta}^-$ has exactly two local minima $x_{\delta,-}^-$ and $x_{\delta,+}^-$, satisfying $x_{\delta,-}^- \in [-1, 0]$ and $x_{\delta,+}^- \in [0, 1]$. Moreover, for all $\delta \in (0, 0.1]$ we have

$$\psi_A^*(\delta) \leq \delta + \left(\max\{|x_{\delta,i} - x_{0,i}| : i \in \{-1, -0, +0, +1\}\} \vee \max\{|f^i(x_j^i) - f_{-\delta}^i(x_{\delta,j}^i)| : i, j \in \{-, +\}\} \right).$$

Finally, the right hand-side of this inequality can be further estimated under some regularity assumptions on f^\pm . Indeed, if there exists constants $c > 0$ and $\gamma \in (0, 1]$ such that

$$|f^\pm(x_\pm^\pm) - f^\pm(x)| \leq c|x_\pm^\pm - x|^\gamma, \quad x \in [x_\pm^\pm - 0.1, x_\pm^\pm + 0.1], \quad (73)$$

then, for all $\delta \in (0, 0.1]$, we can bound the second maximum by

$$\max\{|f^i(x_j^i) - f_{-\delta}^i(x_{\delta,j}^i)| : i, j \in \{-, +\}\} \leq c\delta^\gamma.$$

In addition, if, for some $i \in \{-1, -0, +0, +1\}$, we write $2\delta_0 := f^+(x_{0,i}) - f^-(x_{0,i})$, then $|x_{\delta,i} - x_{0,i}| = 0$ for all $\delta \in (0, \delta_0]$, i.e. the corresponding term in the first maximum disappears for these δ . If $\delta_0 < 0.1$, and we additionally assume, for example, that

$$|f^\pm(x)| \geq c^{-1/\gamma} |x_{0,-1} - x|^{1/\gamma} \quad (74)$$

for all $x \in [x_{0,-1}, -0.8]$, then we have $|x_{\delta,-1} - x_{0,-1}| \leq c\delta^\gamma$ for all $\delta \in (\delta_0, 0.1]$. Combining these assumptions we easily obtain a variety of sets A satisfying $\psi_A^*(\delta) \leq (c+1)\delta^\gamma$ for all $\delta \in (0, 0.1]$, and these examples of sets can be even further extended by considering bi-Lipschitz transformations of X .

Before we can prove the assertions made in the example above, we need to establish the following technical lemma.

Lemma 7.3. *Let $x^* \in [2/5, 3/5]$ and $f : [0, 1] \rightarrow \mathbb{R}$ be a continuous function that is strictly increasing on $[0, x^*]$ and strictly decreasing on $[x^*, 1]$. For $\delta \in (0, 1/8]$ we define $f_{-\delta} : [0, 1] \rightarrow \mathbb{R}$ by*

$$f_{-\delta}(x) := \begin{cases} f(0) & \text{if } x \in [0, \delta] \\ f(x - \delta) \wedge f(x + \delta) & \text{if } x \in [\delta, 1 - \delta] \\ f(1) & \text{if } x \in [1 - \delta, 1]. \end{cases}$$

Then there exists exactly one $x_\delta^* \in [0, 1]$ such that $f_{-\delta}(x_\delta^*) \geq f_{-\delta}(x)$ for all $x \in [0, 1]$. Moreover, we have $x_\delta^* \in (x^* - \delta, x^* + \delta)$ and x_δ^* is the only element $x \in [\delta, 1 - \delta]$ that satisfies $f(x - \delta) = f(x + \delta)$. Finally, we have

$$f_{-\delta}(x) = \begin{cases} f(x - \delta) & \text{if } x \in [\delta, x_\delta^*] \\ f(x + \delta) & \text{if } x \in [x_\delta^*, 1 - \delta]. \end{cases}$$

Proof of Lemma 7.3: Let us first show that there exists an $x_0 \in (x^* - \delta, x^* + \delta)$ such that $f(x_0 - \delta) = f(x_0 + \delta)$. To this end, we observe $g : [x^* - \delta, x^* + \delta] \rightarrow \mathbb{R}$ defined by $g := f(\cdot - \delta) - f(\cdot + \delta)$ is continuous, and since $g(x^* - \delta) = f(x^* - 2\delta) - f(x^*) < 0$ and $g(x^* + \delta) = f(x^*) - f(x^* + 2\delta) > 0$, we find an $x_0 \in (x^* - \delta, x^* + \delta)$ such that $g(x_0) = 0$ by the intermediate value theorem.

Let us now show that $f(x - \delta) < f(x + \delta)$ for all $x \in [\delta, x_0]$ and $f(x - \delta) > f(x + \delta)$ for all $x \in [x_0, 1 - \delta]$. Clearly, for $x \in [\delta, x^* - \delta]$, the strict monotonicity of f on $[0, x^*]$ yields $f(x - \delta) < f(x + \delta)$. Moreover, for $x \in (x^* - \delta, x_0)$, we have $f(x - \delta) < f(x_0 - \delta) = f(x_0 + \delta) < f(x + \delta)$ since $f(\cdot - \delta) : [x^* - \delta, x^* + \delta] \rightarrow \mathbb{R}$ is strictly increasing, while $f(\cdot + \delta) : [x^* - \delta, x^* + \delta] \rightarrow \mathbb{R}$ is strictly decreasing. This shows the assertion for $x \in [\delta, x_0]$, and the assertion for $x \in [x_0, 1 - \delta]$ can be shown analogously.

Combining the two results above, we find that there exists exactly one $x_0 \in [\delta, 1 - \delta]$ satisfying $f(x_0 - \delta) = f(x_0 + \delta)$, and for this x_0 we further know $x_0 \in (x^* - \delta, x^* + \delta)$. In addition, these results show

$$f_{-\delta}(x) = \begin{cases} f(x - \delta) & \text{if } x \in [\delta, x_0] \\ f(x + \delta) & \text{if } x \in [x_0, 1 - \delta]. \end{cases}$$

Let us now return to global maximizers of $f_{-\delta}$. To this end, we first observe that the existence of a global maximum of $f_{-\delta}$ follows from the continuity of $f_{-\delta}$ and the compactness of $[0, 1]$. Let us now fix an $x_\delta \in [0, 1]$ at which this global maximum is attained by $f_{-\delta}$. We first observe that $x_\delta \in (\delta, 1 - \delta)$. Indeed, if, e.g., we had $x_\delta \geq 1 - \delta$, we would obtain $f(1) = f_{-\delta}(x_\delta) \geq f_{-\delta}(1 - 2\delta) = f(1 - 3\delta) \wedge f(1 - \delta) = f(1 - \delta) > f(1)$ using $1 - 3\delta > x^*$, and $x_\delta \leq \delta$ would similarly lead to a contradiction. We next show that we actually have $x_\delta \in [x^* - \delta, x^* + \delta]$. To this end, we observe that it suffices to show

$$x_\delta \geq x^* - \delta \quad \iff \quad x_\delta \leq x^* + \delta. \quad (75)$$

To show the latter, assume that $x_\delta \geq x^* - \delta$. Since $f_{-\delta}$ attains its maximum at x_δ , we then obtain

$$f(x_\delta + \delta) \geq f(x_\delta - \delta) \wedge f(x_\delta + \delta) = f_{-\delta}(x_\delta) \geq f_{-\delta}(x^* + \delta) = f(x^*) \wedge f(x^* + 2\delta) = f(x^* + 2\delta).$$

Now $x_\delta + \delta \leq x^* + 2\delta$ follows from the assumed $x_\delta + \delta \geq x^*$ and the strict monotonicity of f on $[x^*, 1]$. Analogously, $x_\delta \leq x^* + \delta \Rightarrow x_\delta \geq x^* - \delta$ can be shown, and hence (75) is indeed true.

Finally, we can prove the remaining assertion. To this end, we pick again an x_δ at which $f_{-\delta}$ attains its maximum. Then we have already seen that $x_\delta \in [x^* - \delta, x^* + \delta]$. Now observe that assuming $x_\delta < x_0$ leads to $f(x_\delta - \delta) < f(x_0 - \delta) = f(x_0 + \delta) < f(x_\delta + \delta)$ using $x_0, x_\delta \in [x^* - \delta, x^* + \delta]$, which in turn yields the contradiction

$$f_{-\delta}(x_\delta) = f(x_\delta - \delta) \wedge f(x_\delta + \delta) = f(x_\delta - \delta) < f(x_0 - \delta) = f(x_0 - \delta) \wedge f(x_0 + \delta) = f_{-\delta}(x_0).$$

Analogously, we find a contradiction assuming $x_\delta > x_0$, and hence we have $x_\delta = x_0$. Consequently, x_δ is unique and solves $f(x - \delta) = f(x + \delta)$. \square

Proof of Example 7.2: Let us first note that the existence and uniqueness of the local extrema is guaranteed by Lemma 7.3. In addition, this lemma actually shows $x_{\delta,-}^+ \in (x_-^+ - \delta, x_-^+ + \delta)$, $x_{\delta,-}^- \in (x_-^- - \delta, x_-^- + \delta)$, $x_{\delta,+}^+ \in (x_+^+ - \delta, x_+^+ + \delta)$, and $x_{\delta,+}^- \in (x_+^- - \delta, x_+^- + \delta)$. Moreover, since $A^{-\delta} \subset A^{\ominus\delta}$, we have

$$\psi_A^*(\delta) = \sup_{z \in A} d(z, A^{-\delta}) \leq \sup_{z \in A} d(z, A^{\ominus\delta}).$$

In the following, we thus estimate $d(z, A^{\ominus\delta})$ for $z := (x, y) \in A$.

Let us begin with the case $x \in [-1, x_{\delta,-1}]$. For later purposes, we note that the definition of A then yields $x \geq x_{0,-1}$. By the monotonicity of f^\pm on $[-1, -0.8 + \delta]$ we further know $f_\delta^\pm(x + \delta) = f^\pm(x)$. We write $x' := x_{\delta,-1} + \delta$ and

$$y' := \begin{cases} f^-(x_{\delta,-1}) + \delta & \text{if } y \leq f^-(x_{\delta,-1}) + \delta \\ y & \text{if } y \in [f^-(x_{\delta,-1}) + \delta, f^+(x_{\delta,-1}) - \delta] \\ f^+(x_{\delta,-1}) - \delta & \text{if } y \geq f^+(x_{\delta,-1}) - \delta. \end{cases}$$

If $y \leq f^-(x_{\delta,-1}) + \delta$, we then obtain $y \leq y'$ and $y' = f^-(x_{\delta,-1}) + \delta \leq f^-(x) + \delta \leq y + \delta$, that is $|y - y'| \leq \delta$, and it is easy to check that the same is true in the two other cases. Consequently, we have $\|(x, y) - (x', y')\|_\infty = x_{\delta,-1} + \delta - x$, and our construction further ensures

$$y' \in [f^-(x_{\delta,-1}) + \delta, f^+(x_{\delta,-1}) - \delta] = [f_{-\delta}^-(x') + \delta, f_{-\delta}^+(x') - \delta].$$

By Example 7.1 we conclude $(x', y') \in A^{\ominus\delta}$, and from this we easily find

$$d(z, A^{\ominus\delta}) \leq \delta + x_{\delta,-1} - x \leq \delta + x_{\delta,-1} - x_{0,-1}. \quad (76)$$

Let us now show that the latter inequality is also true in the case $x \in [x_{\delta,-1}, -0.8 + \delta]$. To this end, we first observe that the monotonicity of f^\pm on $[-1, -0.8 + 2\delta]$ then yields

$$f^+(x) - f^-(x) \geq f^+(x_{\delta,-1}) - f^-(x_{\delta,-1}) \geq 2\delta,$$

and consequently, we can define

$$y' := \begin{cases} f^-(x) + \delta & \text{if } y \leq f^-(x) + \delta \\ y & \text{if } y \in [f^-(x) + \delta, f^+(x) - \delta] \\ f^+(x) - \delta & \text{if } y \geq f^+(x) - \delta. \end{cases}$$

If $y \leq f^-(x) + \delta$ we then obtain $y \leq y'$ and $y' = f^-(x) + \delta \leq y + \delta$, that is $|y - y'| \leq \delta$, and again it is easy to check that the same is true in the two other cases. Writing $x' := x + \delta$, we thus have $\|(x, y) - (x', y')\|_\infty = \delta$. Moreover, the construction together with $f_\delta^\pm(x + \delta) = f^\pm(x)$ ensures

$$y' \in [f^-(x) + \delta, f^+(x) - \delta] = [f_{-\delta}^-(x') + \delta, f_{-\delta}^+(x') - \delta],$$

and hence we find $(x', y') \in A^{\ominus\delta}$ by Example 7.1. Consequently, we have shown

$$d(z, A^{\ominus\delta}) \leq \delta \leq \delta + x_{\delta, -1} - x_{0, -1},$$

that is, (76) is true for all $x \in [-1, -0.8 + \delta]$.

Let us now consider the case $x \in [-0.8 + \delta, -0.2 - \delta]$. Here, we will focus on the sub-case $y \geq 0$, the subcase $y \leq 0$ can be shown analogously. For later purposes, note that we have $f^-(x \pm \delta) \leq -2\delta$. Now suppose that we actually have $x \in [-0.8 + \delta, x_{\delta, -}^+ - \delta]$. Then we set $x' := x + \delta$ and $y' := y \wedge (f^+(x) - \delta)$. This gives $y' \leq y$ and $y - \delta \leq f^+(x) - \delta \leq y'$, and hence we again have $\|(x, y) - (x', y')\|_\infty = \delta$. Moreover, our constructions together with Lemma 7.3 ensures

$$y' \in [-\delta, f^+(x) - \delta] = [-\delta, f_{-\delta}^+(x') - \delta] \subset [f_{-\delta}^-(x') + \delta, f_{-\delta}^+(x') - \delta],$$

that is $(x', y') \in A^{\ominus\delta}$, and hence (76) is true in this case, too. The next case, we consider, is $x \in [x_{\delta, -}^+ - \delta, x_{\delta, -}^+ + \delta]$. In this case we set $x' := x_{\delta, -}^+$ and $y' := y \wedge (f_{-\delta}^+(x_{\delta, -}^+) - \delta)$. This implies

$$y' \in [-\delta, f_{-\delta}^+(x_{\delta, -}^+) - \delta] \subset [f_{-\delta}^-(x') + \delta, f_{-\delta}^+(x') - \delta],$$

and hence $(x', y') \in A^{\ominus\delta}$. Moreover, we clearly have $|x - x'| \leq \delta$ and, if $y \leq f_{-\delta}^+(x_{\delta, -}^+) - \delta$, we also have $|y - y'| = 0$. Conversely, if $y \geq f_{-\delta}^+(x_{\delta, -}^+) - \delta$, we find

$$y \leq f^+(x) \leq f^+(x_{\delta, -}^+) = f^+(x_{\delta, -}^+) - (f_{-\delta}^+(x_{\delta, -}^+) - \delta) + y',$$

that is $|y - y'| \leq \delta + f^+(x_{\delta, -}^+) - f_{-\delta}^+(x_{\delta, -}^+)$. Combining the latter two cases, we therefore obtain $\|(x, y) - (x', y')\|_\infty \leq \delta + f^+(x_{\delta, -}^+) - f_{-\delta}^+(x_{\delta, -}^+)$, that is

$$d(z, A^{\ominus\delta}) \leq \delta + f^+(x_{\delta, -}^+) - f_{-\delta}^+(x_{\delta, -}^+).$$

Since all remaining cases can be treated analogously, the proof of the general estimate of $\psi_A^*(\delta)$ is finished.

Let us now consider the additional assumptions of f^\pm . For example, suppose that we have

$$|f^+(x_{\delta, -}^+) - f^+(x)| \leq c|x_{\delta, -}^+ - x|^\gamma$$

for all $x \in [x_{\delta, -}^+ - 0.1, x_{\delta, -}^+ + 0.1]$. By Lemma 7.3 we know $x_{\delta, -}^+ \in (x_{\delta, -}^+ - \delta, x_{\delta, -}^+ + \delta)$. Without loss of generality, we may assume that $x_{\delta, -}^+ \in [x_{\delta, -}^+, x_{\delta, -}^+ + \delta)$. Using Lemma 7.3 and $x_{\delta, -}^+ - \delta \in [x_{\delta, -}^+ - \delta, x_{\delta, -}^+) \subset [x_{\delta, -}^+ - 0.1, x_{\delta, -}^+ + 0.1]$, we then obtain

$$|f^+(x_{\delta, -}^+) - f_{-\delta}^+(x_{\delta, -}^+)| = |f^+(x_{\delta, -}^+) - f^+(x_{\delta, -}^+ - \delta)| \leq c|x_{\delta, -}^+ - x_{\delta, -}^+ + \delta|^\gamma \leq c\delta^\gamma.$$

Now let us assume that for e.g. $i := -1$ we have $\delta_0 > 0$. For $\delta \in (0, \delta_0]$ we then find $f^+(x_{0, -1}) - f^-(x_{0, -1}) \geq 2\delta$, and thus $x_{0, -1} = x_{\delta, -1} = -1$. Conversely, let $\delta \in (\delta_0, 0.1]$. Then we have $f^+(x_{0, -1}) - f^-(x_{0, -1}) < 2\delta$ and a simple application of the intermediate value theorem thus yields $f^+(x_{\delta, -1}) - f^-(x_{\delta, -1}) = 2\delta$. Using the additional assumption on f^\pm around the point $x_{0, -1}$, we then find

$$2c^{-1/\gamma}|x_{\delta, -1} - x_{0, -1}|^{1/\gamma} \leq |f^-(x_{\delta, -1})| + |f^+(x_{\delta, -1})| = f^+(x_{\delta, -1}) - f^-(x_{\delta, -1}) = 2\delta,$$

that is $|x_{\delta, -1} - x_{0, -1}| \leq c\delta^\gamma$. □

Our next example, which represents the main result of this appendix, shows that many continuous distributions satisfy our thickness assumption.

Example 7.4. Let $X := [-1, 1] \times [-2, 2]$ be equipped with the metric defined by the supremums norm. Moreover, let P be a Lebesgue absolutely continuous probability measure that has a continuous density h . Furthermore, assume that there exists a $\rho^{**} > 0$, such that, for all $\rho \in (0, \rho^{**}]$, the level set M_ρ is of the form considered in Example 7.2. In addition, we assume that there is a constant $K \in (0, 1)$ such that

$$|h(x, y) - \rho^* - x^2 + y^2| \leq K(x^2 + y^2) \tag{77}$$

for some $\rho^* \in [0, \rho^{**})$ and all $(x, y) \in \{h > 0\} \cap ([-0.2, 0.2] \times (-1.1, 1.1))$. Moreover, assume that h is continuously differentiable on the sets

$$\begin{aligned} A_1 &:= \{h > 0\} \cap \left(((-0.7, -0.3) \cup (0.3, 0.7)) \times ((-1.1, -0.2) \cup (0.2, 1.1)) \right) \\ A_2 &:= \{h > 0\} \cap \left(((-1, -0.8) \cup (0.8, 1)) \times ((-1.1, 0) \cup (0.2, 1.1)) \right) \\ A_3 &:= \{h > 0\} \cap \left\{ (x, y) \in X : x \in (-0.2, 0) \cup (0, 0.2) \text{ and } |y| < \sqrt{\frac{1+K}{1-K}}|x| \right\} \end{aligned}$$

with $h_y := \frac{\partial h}{\partial y} \neq 0$ on A_1 and $h_x := \frac{\partial h}{\partial x} \neq 0$ on $A_2 \cup A_3$. Finally, assume that there exists another constant $C > 0$ such that $|h_x| \leq C|h_y|$ on A_1 and $|h_y| \leq C|h_x|$ on $A_2 \cup A_3$. Then the distribution P has thick levels of order $\gamma = 1$ with $\delta_{\text{thick}} = 0.1$ and

$$c_{\text{thick}} = 1 + \max \left\{ C, \sqrt{\frac{1+K}{1-K}} \right\}.$$

Moreover, P can be topologically clustered between the critical levels ρ^* and ρ^{**} and for all $\varepsilon \in (0, \rho^{**} - \rho]$ we have

$$\frac{2}{\sqrt{1-K}}\sqrt{\varepsilon} \leq \tau_{M_{\rho^*+\varepsilon}}^* \leq \frac{2}{\sqrt{1+K}}\sqrt{\varepsilon}. \quad (78)$$

Proof of Example 7.4: Since we consider the Lebesgue measure on X , we have $M_0 = X$. Moreover, we have $X^{-\delta} = X$ since we consider the operation in X , and from this, we immediately see $\psi_X^*(\delta) = 0$ for all $\delta > 0$. Consequently, there is nothing to prove for $\rho = 0$.

Let us now fix some $\rho \in (0, \rho^{**}]$. Moreover, let $f^\pm : [-1, 1] \rightarrow [-1, 1]$ be the two functions satisfying the assumptions of Example 7.2 and

$$M_\rho = \{(x, y) \in X : x \in [-1, 1] \text{ and } f^-(x) \leq y \leq f^+(x)\}.$$

Let us now pick an $(x, y) \in M_\rho$ with $y = f^+(x)$ or $y = f^-(x)$. Then we find $(x, y) \in \partial M_\rho$, and thus we have $h(x, y) = \rho$ by Lemma 2.1, that is $h(x, f^\pm(x)) = \rho$.

Our first goal is to verify (73). To this end, we solely focus without loss of generality to the case x_\pm^+ and f^+ , since the other cases can be treated analogously. Let us fix an $x \in [x_\pm^+ - 0.1, x_\pm^+ + 0.1]$. Then we have $x \in (0.3, 0.7)$ and thus $f^+(x) \in (0.2, 1.1)$ by (72). Consequently, h is continuously differentiable in $(x, f^+(x))$. By the implicit function theorem and the previously shown $h(x', f^+(x')) = \rho$ for all $x' \in (0.3, 0.7)$ we then conclude that f^+ is continuously differentiable at x and

$$(f^+(x))' = - \left(\frac{\partial h}{\partial y}(x, f^+(x)) \right)^{-1} \cdot \frac{\partial h}{\partial x}(x, f^+(x)) = \frac{h_x(x, f^+(x))}{h_y(x, f^+(x))}. \quad (79)$$

Using $|h_x| \leq C|h_y|$ on A_1 , we thus find $|(f^+(x))'| \leq C$, and hence f^+ is Lipschitz continuous on $(0.3, 0.7)$ with Lipschitz constant smaller than or equal to C . This implies (73) with constant C and exponent $\gamma = 1$.

Let us now consider the endpoints $x_{0,\pm 1}$, where again it suffices to consider one case, say $x_{0,-1}$, due to symmetry. Let us write $2\delta_0 := f^+(x_{0,-1}) - f^-(x_{0,-1})$. Then, if $\delta_0 \geq 0.1$, we have $|x_{\delta,-1} - x_{0,-1}| = 0$ for all $\delta \in (0, 0.1]$ by Example 7.2, and hence it suffices to show (74) in the case $\delta_0 < 0.1$. Observing that it actually suffices to show (74) for all $x \in (x_{0,-1}, -0.8)$ by continuity, we begin by fixing such an x . By monotonicity we then have $0 < f^+(x) < f^+(0.8) < 1.1$, and consequently, h is continuously differentiable at $(x, f^+(x))$. The implicit function theorem and the previously shown $h(x', f^+(x')) = \rho$ for all $x' \in (x_{0,-1}, -0.8)$, then shows that f^+ is continuously differentiable at x and (79) holds. Using $|h_y| \leq C|h_x|$ on A_2 , we then find $|(f^+(x))'| \geq 1/C$, and the fundamental theorem of calculus thus yields

$$|f^+(x') - f^+(x)| = \left| \int_x^{x'} (f^+(t))' dt \right| \geq C^{-1}|x' - x|$$

for all $x, x' \in (x_{0,-1}, -0.8)$. Now, letting $x' \rightarrow x_{0,-1}$, we obtain, for all $x \in (x_{0,-1}, -0.8)$,

$$|f^+(x)| \geq f^+(x) - f^+(x_{0,-1}) = |f^+(x) - f^+(x_{0,-1})| \geq C^{-1}|x_{0,-1} - x|,$$

that is (74) holds with constant C and exponent $\gamma = 1$.

Finally, let us consider the points $x_{0,\pm 0}$, where yet another time, we only focus on one case, say $x_{0,+0}$. For $x \in [x_{0,+0}, 0.2]$, we then have

$$\rho = h(x, f^+(x)) \leq \rho^* + (1+K)x^2 + (K-1)(f^+(x))^2, \quad (80)$$

that is $(f^+(x))^2 \leq \frac{\rho^* - \rho}{1-K} + \frac{1+K}{1-K}x^2$. Analogously, we can find a lower bound on $(f^+(x))^2$, so that we end up having

$$(f^+(x))^2 \in \left[\frac{\rho^* - \rho}{1+K} + \frac{1-K}{1+K}x^2, \frac{\rho^* - \rho}{1-K} + \frac{1+K}{1-K}x^2 \right], \quad (81)$$

and an analogue result holds for $(f^-(x))^2$. Again, our goal is to show an analogue of (74). To this end, we first consider the case $\rho \in (0, \rho^*]$. By (77), we then know that $h(0,0) = \rho^* \geq \rho$, and hence $f^+(0) \geq 0$. Analogously, we find $f^-(0) \leq 0$, which together implies $x_{0,+0} = 0$. Furthermore, for $x \in [x_{0,+0}, 0.2]$, (81) gives

$$f^+(x) \geq \sqrt{\frac{\rho^* - \rho}{1+K} + \frac{1-K}{1+K}x^2} \geq \sqrt{\frac{1-K}{1+K}}|x| = \sqrt{\frac{1-K}{1+K}}|x_{0,+0} - x|,$$

that is (74) holds with constant $\sqrt{\frac{1+K}{1-K}}$ and exponent $\gamma = 1$. Let us now consider the case $\rho \in (\rho^*, \rho^{**}]$. For $x \in (x_{0,+0}, 0.2)$, (81) then yields

$$f^+(x) \leq \sqrt{\frac{\rho^* - \rho}{1-K} + \frac{1+K}{1-K}x^2} < \sqrt{\frac{1+K}{1-K}}|x|,$$

and thus we find $(x, f^+(x)) \in A_3$. Consequently, h is continuously differentiable at $(x, f^+(x))$, and (79) holds. As for $x_{0,-1}$, we can then show that (74) holds with constant C and exponent $\gamma = 1$.

In order to show that P can be topologically clustered between the critical levels ρ^* and ρ^{**} , we first note that the assumed continuity of h guarantees that P is normal by Lemma 2.4. Let us now fix a $\rho \in (\rho^*, \rho^{**}]$. Since from (77) we infer that $h(0,0) = \rho^*$, we then obtain $(0,0) \notin M_\rho$. The latter implies $x_{0,-0} < 0 < x_{0,+0}$, where $x_{0,-0}$ and $x_{0,+0}$ are the points defined in Example 7.2 for the set M_ρ . By Example 7.1 we then see that $|\mathcal{C}(M_\rho)| = 2$. Analogously, for $\rho \in [0, \rho^*]$, the equality $h(0,0) = \rho^*$ implies $x_{0,-0} = 0 = x_{0,+0}$, which shows $|\mathcal{C}(M_\rho)| = 1$. Finally, the bijectivity of $\zeta : \mathcal{C}(M_{\rho^{**}}) \rightarrow \mathcal{C}(M_\rho)$ follows from the form of the connected components described in Example 7.1.

Let us finally prove (78). To this end, we fix an $\varepsilon \in (0, \rho^{**} - \rho]$ and define $\rho := \rho^* + \varepsilon$. Then we have already observed that $x_{0,-0} < 0 < x_{0,+0}$, and hence $f^\pm(x_{0,\pm 0}) = 0$. For $x := x_{0,+0}$ we then obtain

$$\rho = h(x, f^+(x)) \leq \rho^* + (1+K)x^2$$

by (80), and applying some simple transformations we thus find $x_{0,+0} = x \geq \sqrt{\frac{\rho - \rho^*}{1+K}} = \sqrt{\frac{\varepsilon}{1+K}}$. For $x := x_{0,+0}$ we further have

$$\rho = h(x, f^+(x)) \geq \rho^* + (1-K)x^2,$$

and hence $x_{0,+0} \leq \sqrt{\frac{\varepsilon}{1-K}}$. Since analogous estimates can be derived for $x_{0,-0}$, the formula $\tau_{M_{\rho^* + \varepsilon}}^* = x_{0,+0} - x_{0,-0}$ found in Example 7.1 then gives (78). \square

The last example of this appendix shows that the distributions from the previous example have a smooth boundary.

Example 7.5. *Let X and P be as in Example 7.4. Then the clusters have an α -smooth boundary for $\alpha = 1$ and*

$$c_{\text{bound}} = 8 \left(10 + C + \sqrt{\frac{1+K}{1-K}} \right).$$

Proof of Example 7.5: Let us first consider the case $0 < \delta \leq 0.1$. To this end, we fix a $\rho \in (\rho^*, \rho^{**}]$. Without loss of generality, we only consider the connected component A with $x < 0$ for all $(x, y) \in A$. From Remark 2.18 we know that $A^{+\delta/2} \setminus A^{-\delta/2} \subset A^{\oplus\delta} \setminus A^{\ominus\delta}$ and the latter two sets have been

calculated in Example 7.1. In the following, we will only estimate $\lambda^2(\{(x, y) : y \geq 0\} \cap A^{\oplus\delta} \setminus A^{\ominus\delta})$, the case $y \leq 0$ can be treated analogously. Our first intermediate result towards the desired estimate is

$$\begin{aligned} \lambda^2([-1 \vee (x_{0,-1} - \delta), x_{\delta,-1}] \times [0, 2] \cap A^{\oplus\delta} \setminus A^{\ominus\delta}) &\leq 2|(x_{0,-1} - \delta) - x_{\delta,-1}| \\ &\leq 2\delta + 2|x_{0,-1} - x_{\delta,-1}| \\ &\leq 2(1 + C)\delta, \end{aligned}$$

where in the last step we used that the proof of Example 7.4 showed (74) for $c = C$ and $\gamma = 1$. Moreover, we have

$$\begin{aligned} \lambda^2([x_{\delta,-1}, x_{-}^+ - \delta] \times [0, 2] \cap A^{\oplus\delta} \setminus A^{\ominus\delta}) &= \int_{x_{\delta,-1}}^{x_{-}^+ - \delta} f^+(x + \delta) - f^+(x - \delta) + 2\delta \, dx \\ &\leq 2\delta + \int_{x_{-}^+ - \delta}^{x_{-}^+ + \delta} f(x) \, dx \\ &\leq 4\delta \end{aligned}$$

and analogously we obtain $\lambda^2([x_{-}^+ + \delta, x_{\delta,-0}] \times [0, 2] \cap A^{\oplus\delta} \setminus A^{\ominus\delta}) \leq 4\delta$. In addition, we easily find $\lambda^2([x_{-}^+ - \delta, x_{-}^+ + \delta] \times [0, 2] \cap A^{\oplus\delta} \setminus A^{\ominus\delta}) \leq 4\delta$ and finally, we have

$$\lambda^2([x_{\delta,-0}, 0 \wedge (x_{0,-0} + \delta)] \times [0, 2] \cap A^{\oplus\delta} \setminus A^{\ominus\delta}) \leq 2|x_{\delta,-0} - x_{0,-0} - \delta| \leq 2\delta + 2\sqrt{\frac{1+K}{1-K}}\delta,$$

where we used that the proof of Example 7.4 showed (74) for $c = \sqrt{\frac{1+K}{1-K}}$ and $\gamma = 1$. Combining all these estimates we obtain

$$\lambda^2([-1, 0] \times [0, 2] \cap A^{\oplus\delta} \setminus A^{\ominus\delta}) \leq 4\left(6 + C + \sqrt{\frac{1+K}{1-K}}\right)\delta$$

for all $\delta \in (0, 0.05]$. Moreover, for $\delta \in [0.05, 1]$ we easily obtain

$$\lambda^2([-1, 0] \times [0, 2] \cap A^{\oplus\delta} \setminus A^{\ominus\delta}) \leq 2 \leq 40\delta$$

Combining both estimates and adding the case $y \leq 0$, we then obtain the assertion. \square

Ingo Steinwart
Universität Stuttgart
Fachbereich Mathematik
Pfaffenwaldring 57
70569 Stuttgart
Germany

E-Mail: Ingo.Steinwart@mathematik.uni-stuttgart.de

WWW: <http://www.isa.uni-stuttgart.de/Steinwart>

Erschienenene Preprints ab Nummer 2007/2007-001

Komplette Liste: <http://www.mathematik.uni-stuttgart.de/preprints>

- 2013-016 *Steinwart, I.:* Fully Adaptive Density-Based Clustering
- 2013-015 *Steinwart, I.:* Some Remarks on the Statistical Analysis of SVMs and Related Methods
- 2013-014 *Rohde, C.; Zeiler, C.:* A Relaxation Riemann Solver for Compressible Two-Phase Flow with Phase Transition and Surface Tension
- 2013-013 *Moroianu, A.; Semmelmann, U.:* Generalized Killing spinors on Einstein manifolds
- 2013-012 *Moroianu, A.; Semmelmann, U.:* Generalized Killing Spinors on Spheres
- 2013-011 *Kohls, K; Rösch, A.; Siebert, K.G.:* Convergence of Adaptive Finite Elements for Control Constrained Optimal Control Problems
- 2013-010 *Corli, A.; Rohde, C.; Schleper, V.:* Parabolic Approximations of Diffusive-Dispersive Equations
- 2013-009 *Nava-Yazdani, E.; Polthier, K.:* De Casteljau's Algorithm on Manifolds
- 2013-008 *Bächle, A.; Margolis, L.:* Rational conjugacy of torsion units in integral group rings of non-solvable groups
- 2013-007 *Knarr, N.; Stroppel, M.J.:* Heisenberg groups over composition algebras
- 2013-006 *Knarr, N.; Stroppel, M.J.:* Heisenberg groups, semifields, and translation planes
- 2013-005 *Eck, C.; Kutter, M.; Sändig, A.-M.; Rohde, C.:* A Two Scale Model for Liquid Phase Epitaxy with Elasticity: An Iterative Procedure
- 2013-004 *Griesemer, M.; Wellig, D.:* The Strong-Coupling Polaron in Electromagnetic Fields
- 2013-003 *Kabil, B.; Rohde, C.:* The Influence of Surface Tension and Configurational Forces on the Stability of Liquid-Vapor Interfaces
- 2013-002 *Devroye, L.; Ferrario, P.G.; Györfi, L.; Walk, H.:* Strong universal consistent estimate of the minimum mean squared error
- 2013-001 *Kohls, K.; Rösch, A.; Siebert, K.G.:* A Posteriori Error Analysis of Optimal Control Problems with Control Constraints
- 2012-018 *Kimmerle, W.; Konovalov, A.:* On the Prime Graph of the Unit Group of Integral Group Rings of Finite Groups II
- 2012-017 *Stroppel, B.; Stroppel, M.:* Desargues, Doily, Dualities, and Exceptional Isomorphisms
- 2012-016 *Moroianu, A.; Pilca, M.; Semmelmann, U.:* Homogeneous almost quaternion-Hermitian manifolds
- 2012-015 *Steinke, G.F.; Stroppel, M.J.:* Simple groups acting two-transitively on the set of generators of a finite elation Laguerre plane
- 2012-014 *Steinke, G.F.; Stroppel, M.J.:* Finite elation Laguerre planes admitting a two-transitive group on their set of generators
- 2012-013 *Diaz Ramos, J.C.; Dominguez Vázquez, M.; Kollross, A.:* Polar actions on complex hyperbolic spaces
- 2012-012 *Moroianu, A.; Semmelmann, U.:* Weakly complex homogeneous spaces
- 2012-011 *Moroianu, A.; Semmelmann, U.:* Invariant four-forms and symmetric pairs
- 2012-010 *Hamilton, M.J.D.:* The closure of the symplectic cone of elliptic surfaces
- 2012-009 *Hamilton, M.J.D.:* Iterated fibre sums of algebraic Lefschetz fibrations
- 2012-008 *Hamilton, M.J.D.:* The minimal genus problem for elliptic surfaces

- 2012-007 *Ferrario, P.*: Partitioning estimation of local variance based on nearest neighbors under censoring
- 2012-006 *Stroppel, M.*: Buttons, Holes and Loops of String: Lacing the Doily
- 2012-005 *Hantsch, F.*: Existence of Minimizers in Restricted Hartree-Fock Theory
- 2012-004 *Grundhöfer, T.; Stroppel, M.; Van Maldeghem, H.*: Unitals admitting all translations
- 2012-003 *Hamilton, M.J.D.*: Representing homology classes by symplectic surfaces
- 2012-002 *Hamilton, M.J.D.*: On certain exotic 4-manifolds of Akhmedov and Park
- 2012-001 *Jentsch, T.*: Parallel submanifolds of the real 2-Grassmannian
- 2011-028 *Spreer, J.*: Combinatorial 3-manifolds with cyclic automorphism group
- 2011-027 *Griesemer, M.; Hantsch, F.; Wellig, D.*: On the Magnetic Pekar Functional and the Existence of Bipolarons
- 2011-026 *Müller, S.*: Bootstrapping for Bandwidth Selection in Functional Data Regression
- 2011-025 *Felber, T.; Jones, D.; Kohler, M.; Walk, H.*: Weakly universally consistent static forecasting of stationary and ergodic time series via local averaging and least squares estimates
- 2011-024 *Jones, D.; Kohler, M.; Walk, H.*: Weakly universally consistent forecasting of stationary and ergodic time series
- 2011-023 *Györfi, L.; Walk, H.*: Strongly consistent nonparametric tests of conditional independence
- 2011-022 *Ferrario, P.G.; Walk, H.*: Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors
- 2011-021 *Eberts, M.; Steinwart, I.*: Optimal regression rates for SVMs using Gaussian kernels
- 2011-020 *Frank, R.L.; Geisinger, L.*: Refined Semiclassical Asymptotics for Fractional Powers of the Laplace Operator
- 2011-019 *Frank, R.L.; Geisinger, L.*: Two-term spectral asymptotics for the Dirichlet Laplacian on a bounded domain
- 2011-018 *Hänel, A.; Schulz, C.; Wirth, J.*: Embedded eigenvalues for the elastic strip with cracks
- 2011-017 *Wirth, J.*: Thermo-elasticity for anisotropic media in higher dimensions
- 2011-016 *Höllig, K.; Hörner, J.*: Programming Multigrid Methods with B-Splines
- 2011-015 *Ferrario, P.*: Nonparametric Local Averaging Estimation of the Local Variance Function
- 2011-014 *Müller, S.; Dippon, J.*: k-NN Kernel Estimate for Nonparametric Functional Regression in Time Series Analysis
- 2011-013 *Knarr, N.; Stroppel, M.*: Unitals over composition algebras
- 2011-012 *Knarr, N.; Stroppel, M.*: Baer involutions and polarities in Moufang planes of characteristic two
- 2011-011 *Knarr, N.; Stroppel, M.*: Polarities and planar collineations of Moufang planes
- 2011-010 *Jentsch, T.; Moroianu, A.; Semmelmann, U.*: Extrinsic hyperspheres in manifolds with special holonomy
- 2011-009 *Wirth, J.*: Asymptotic Behaviour of Solutions to Hyperbolic Partial Differential Equations
- 2011-008 *Stroppel, M.*: Orthogonal polar spaces and unitals
- 2011-007 *Nagl, M.*: Charakterisierung der Symmetrischen Gruppen durch ihre komplexe Gruppenalgebra

- 2011-006 *Solanes, G.; Teufel, E.:* Horo-tightness and total (absolute) curvatures in hyperbolic spaces
- 2011-005 *Ginoux, N.; Semmelmann, U.:* Imaginary Kählerian Killing spinors I
- 2011-004 *Scherer, C.W.; Köse, I.E.:* Control Synthesis using Dynamic D -Scales: Part II — Gain-Scheduled Control
- 2011-003 *Scherer, C.W.; Köse, I.E.:* Control Synthesis using Dynamic D -Scales: Part I — Robust Control
- 2011-002 *Alexandrov, B.; Semmelmann, U.:* Deformations of nearly parallel G_2 -structures
- 2011-001 *Geisinger, L.; Weidl, T.:* Sharp spectral estimates in domains of infinite volume
- 2010-018 *Kimmerle, W.; Konovalov, A.:* On integral-like units of modular group rings
- 2010-017 *Gauduchon, P.; Moroianu, A.; Semmelmann, U.:* Almost complex structures on quaternion-Kähler manifolds and inner symmetric spaces
- 2010-016 *Moroianu, A.; Semmelmann, U.:* Clifford structures on Riemannian manifolds
- 2010-015 *Grafarend, E.W.; Kühnel, W.:* A minimal atlas for the rotation group $SO(3)$
- 2010-014 *Weidl, T.:* Semiclassical Spectral Bounds and Beyond
- 2010-013 *Stroppel, M.:* Early explicit examples of non-desarguesian plane geometries
- 2010-012 *Effenberger, F.:* Stacked polytopes and tight triangulations of manifolds
- 2010-011 *Györfi, L.; Walk, H.:* Empirical portfolio selection strategies with proportional transaction costs
- 2010-010 *Kohler, M.; Krzyżak, A.; Walk, H.:* Estimation of the essential supremum of a regression function
- 2010-009 *Geisinger, L.; Laptev, A.; Weidl, T.:* Geometrical Versions of improved Berezin-Li-Yau Inequalities
- 2010-008 *Poppitz, S.; Stroppel, M.:* Polarities of Schellhammer Planes
- 2010-007 *Grundhöfer, T.; Krinn, B.; Stroppel, M.:* Non-existence of isomorphisms between certain unitals
- 2010-006 *Höllig, K.; Hörner, J.; Hoffacker, A.:* Finite Element Analysis with B-Splines: Weighted and Isogeometric Methods
- 2010-005 *Kaltenbacher, B.; Walk, H.:* On convergence of local averaging regression function estimates for the regularization of inverse problems
- 2010-004 *Kühnel, W.; Solanes, G.:* Tight surfaces with boundary
- 2010-003 *Kohler, M.; Walk, H.:* On optimal exercising of American options in discrete time for stationary and ergodic data
- 2010-002 *Gulde, M.; Stroppel, M.:* Stabilizers of Subspaces under Similitudes of the Klein Quadric, and Automorphisms of Heisenberg Algebras
- 2010-001 *Leitner, F.:* Examples of almost Einstein structures on products and in cohomogeneity one
- 2009-008 *Griesemer, M.; Zenk, H.:* On the atomic photoeffect in non-relativistic QED
- 2009-007 *Griesemer, M.; Moeller, J.S.:* Bounds on the minimal energy of translation invariant n -polaron systems
- 2009-006 *Demirel, S.; Harrell II, E.M.:* On semiclassical and universal inequalities for eigenvalues of quantum graphs
- 2009-005 *Bächle, A.; Kimmerle, W.:* Torsion subgroups in integral group rings of finite groups
- 2009-004 *Geisinger, L.; Weidl, T.:* Universal bounds for traces of the Dirichlet Laplace operator

- 2009-003 *Walk, H.:* Strong laws of large numbers and nonparametric estimation
- 2009-002 *Leitner, F.:* The collapsing sphere product of Poincaré-Einstein spaces
- 2009-001 *Brehm, U.; Kühnel, W.:* Lattice triangulations of E^3 and of the 3-torus
- 2008-006 *Kohler, M.; Krzyżak, A.; Walk, H.:* Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps
- 2008-005 *Kaltenbacher, B.; Schöpfer, F.; Schuster, T.:* Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems
- 2008-004 *Leitner, F.:* Conformally closed Poincaré-Einstein metrics with intersecting scale singularities
- 2008-003 *Effenberger, F.; Kühnel, W.:* Hamiltonian submanifolds of regular polytope
- 2008-002 *Hertweck, M.; Höfert, C.R.; Kimmerle, W.:* Finite groups of units and their composition factors in the integral group rings of the groups $PSL(2, q)$
- 2008-001 *Kovarik, H.; Vugalter, S.; Weidl, T.:* Two dimensional Berezin-Li-Yau inequalities with a correction term
- 2007-006 *Weidl, T.:* Improved Berezin-Li-Yau inequalities with a remainder term
- 2007-005 *Frank, R.L.; Loss, M.; Weidl, T.:* Polya's conjecture in the presence of a constant magnetic field
- 2007-004 *Ekholm, T.; Frank, R.L.; Kovarik, H.:* Eigenvalue estimates for Schrödinger operators on metric trees
- 2007-003 *Lesky, P.H.; Racke, R.:* Elastic and electro-magnetic waves in infinite waveguides
- 2007-002 *Teufel, E.:* Spherical transforms and Radon transforms in Moebius geometry
- 2007-001 *Meister, A.:* Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions