JMLR: Workshop and Conference Proceedings 19 (2011) 703–737 24th Annual Conference on Learning Theory

Adaptive Density Level Set Clustering

Ingo Steinwart INGO.STEINWART@MATHEMATIK.UNI-STUTTGART.DE
Institute for Stochastics and Applications
Faculty 8: Mathematics and Physics
University of Stuttgart
D-70569 Stuttgart Germany

Editor: Sham Kakade, Ulrike von Luxburg

Abstract

Clusters are often defined to be the connected components of a density level set. Unfortunately, this definition depends on a level that needs to be user specified by some means. In this paper we present a simple algorithm that is able to asymptotically determine the optimal level, that is, the level at which there is the first split in the cluster tree of the data generating distribution. We further show that this algorithm asymptotically recovers the corresponding connected components. Unlike previous work, our analysis does not require strong assumptions on the density such as continuity or even smoothness.

Keywords: Clustering, Density Level Sets

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 $D = (x_1, \ldots, x_n) \in X^n$ 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 assumed that the density h is continuous, since this removes or hides various pathologies regarding the topological notion of connectedness that are caused by changes of h on μ-zero sets.

Historically, two distinct questions have been investigated for this cluster definition. 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, and hence it seems fair to say that it already enjoys a reasonably good statistical understanding. Unfortunately, however, it suffers from a conceptional problem, namely that of determining a good value of ρ, and recently Rinaldo and Wasserman (2010) actually remark that research in this direction “would be very useful”.

The second approach tries to address this issue by considering the hierarchical structure of the connected components for different levels. To be more precise, if h is a fixed density, which, for the sake of simplicity, is assumed to have *closed* density level sets, and A is 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 $A \subset B$. Under some additional assumptions on μ and the density h, this then 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. We refer to Hartigan (1975); Stuetzle (2003); Chaudhuri and Dasgupta (2010); Stuetzle and Nugent (2010) for definitions and methods for estimating the structure of this tree. 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 $n \to \infty$. 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 results. Unfortunately, however, neither of these results tell us *a)* how to find A_D and A'_D 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 approach reveals more about the cluster structure and less about the actual clusters.

In contrast to the these papers on the cluster tree approach this work focusses more on estimating the actual, maximal clusters[1]. Namely, we present a simple algorithm that automatically approximates the smallest possible value of ρ for which the level set contains more than one component. In addition, the algorithm approximates the resulting components arbitrarily well for $n \to \infty$ under minimal and somewhat natural conditions, which include discontinuous densities.

Unlike basically all other papers on density based clustering, with the exception of Rinaldo and Wasserman (2010), we do not assume that the density h is continuous, or even Hölder continuous. While this approach enlarges the class of distributions significantly, it also produces some serious technical difficulties as we no longer have a canonical representative for the only μ-almost surely defined density h. To be more precise, in general the topological properties of the level set $\{h \geq \rho\}$ do dramatically depend on the chosen representative for the density, and hence the entire density based clustering approach becomes ill-defined. To address this problem, we first provide a definition for density level sets that make them actually *independent* of the chosen representative. As a consequence, it becomes mathematically rigorous to consider the infimum ρ^* over all levels ρ for which the

1. This destinction is, however, to some extend artificial, since recursively applying methods that estimate the maximal clusters well, automatically yields a consistent estimate of the cluster tree

corresponding density level sets contain more than connected component. For simplicity, we then assume that there exists some $\rho^{**} > \rho^*$ such that the level sets for all $\rho \in (\rho^*, \rho^{**}]$ 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. 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 $h : X \to (0, \infty)$ 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 ρ^{**} can be any value with $h(x_1) < \rho^{**} < h(x_2)$. Moreover, for $\rho \in (\rho^*, \rho^{**}]$, the density level actually contains exactly two connected components, while for $\rho > \rho^{**}$ the level sets may or may not contain three connected components. In other words, our assumption of 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 unlikely. Moreover, it is needless to say that in higher dimensions similar arguments can be made. Finally, it seems fair to say at this point that one more significant assumption on the level sets need to be made, namely one that excludes bridges and cusps that are too thin and long. However, 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. With the assumptions described so far, our main result, Theorem 26, then shows that a simple histogram based algorithm both approximates ρ^* and the resulting clusters arbitrarily well for sample sizes $n \to \infty$.

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 investigating the cluster structure of the true density sets and their empirical estimates. We further make the above notion of clusters rigorous and establish some results about the stability of the cluster structure under simple operations related to the blurriness of empirical estimates. In Section 3 we then present our algorithm and our main result, Theorem 26, followed by a discussion in Section 4. Finally, all proofs, together with some auxiliary results and some background material can be found in the appendix.

2. Preliminaries: Density 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 and their regularity

Unlike to the rest of the paper, where we focus on 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 $h : X \to [0, \infty)$, 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

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

then there exists another μ-density $h' : X \to [0, \infty)$ 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, observe that, for every $\rho \in \mathbb{R}$,

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

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

$$\begin{aligned} M_\rho &:= \operatorname{supp} \mu_\rho\,, \\ V_\rho &:= X \setminus M_\rho\,, \end{aligned}$$

where $\operatorname{supp} \mu_\rho$ denotes the support of μ_ρ, are independent of this choice, too. In the following, we call M_ρ the density level set to the level ρ. To justify this notation, recall that by definition, the support of a measure is the complement of the largest open zero set, and hence $\operatorname{supp} \mu_\rho$ is the smallest closed subset B of X that satisfies $\mu_\rho(X \setminus B) = 0$. Moreover, recall that, for every measure on a complete, separable metric space, the support actually exists. Consequently, for any given μ-density $h : X \to [0, \infty)$ of P, we have

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

that is, up to μ-zero sets no density level set $\{h \geq \rho\}$ is larger than M_ρ. Moreover, M_ρ is the smallest closed set satisfying this equation, and hence we further obtain

$$M_\rho \subset \overline{\{h \geq \rho\}}\,, \tag{2}$$

where $\overline{A}$ denotes the closure of an $A \subset X$. In addition, it is easy to check that we have

$$M_\rho = \big\{x \in X : \mu_\rho(U) > 0 \text{ for all open neighborhoods } U \text{ of } x\big\}\,. \tag{3}$$

Note that if $\operatorname{supp} \mu = X$, we actually have $V_\rho = \emptyset$ and $M_\rho = X$ for all $\rho \leq 0$, but typically we are, of course, interested in the case $\rho > 0$, only. To state our first result, which provides a lower bound for the set M_ρ, we need to recall that the interior $\mathring{A}$ of a set $A \subset X$ is the largest open subset of A.

Lemma 1 *Let (X, d) be a complete separable metric space, μ be a σ-finite measure on X with* $\operatorname{supp} \mu = X$*, 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*

$$\overline{\{h \mathring{\geq} \rho\}} \subset M_\rho \subset \overline{\{h \geq \rho\}}\,.$$

Lemma 1 in particular shows that $\{h \overset{\circ}{\geq} \rho\} \subset M_\rho \subset \overline{\{h \geq \rho\}}$. Therefore, the difference between the sets M_ρ and $\{h \geq \rho\}$ is contained in the boundary of $\{h \geq \rho\}$, that is

$$M_\rho \vartriangle \{h \geq \rho\} \subset \partial\{h \geq \rho\}\,, \tag{4}$$

where $\partial A := \overline{A} \setminus \mathring{A}$ denotes the boundary of a set A. Moreover, if P has a continuous density, we obtain the following corollary.

Corollary 2 *Let (X, d) be a complete separable metric space, μ be a σ-finite measure on X with $\operatorname{supp}\mu = X$, and P be a μ-absolutely continuous probability measure on X that has a continuous μ-density $h : X \to [0, \infty)$. Then, for all $\rho \in \mathbb{R}$, we have*

$$\overline{\{h > \rho\}} \subset M_\rho \subset \{h \geq \rho\}\,,$$
$$\{h > \rho\} \subset \mathring{M}_\rho \subset \{h \overset{\circ}{\geq} \rho\}\,.$$

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

Lemma 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. Then, for all $\rho_1 \leq \rho_2$, we have*

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

In turns out that we will not only need the equality $\mu(\{h \geq \rho\} \setminus M_\rho) = 0$ but also the “converse” equality $\mu(M_\rho \setminus \{h \geq \rho\}) = 0$. This is ensured by the following definition.

Definition 4 *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. We say that P is regular at level $\rho \in \mathbb{R}$, if*

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

for one μ-density (and thus all μ-densities) $h : X \to [0, \infty)$ of P.

Note that by (4) a probability measure P is regular at level ρ, if the boundary $\partial\{h \geq \rho\}$ for *one* μ-density h is a μ-zero set. Let us now assume that P is regular at some level ρ. By (1) we then immediately see that

$$\mu(M_\rho \vartriangle \{h \geq \rho\}) = 0 \tag{5}$$

for *all* μ-densities h of P. Furthermore, since $V_\rho = X \setminus M_\rho$ and $\{h < \rho\} = X \setminus \{h \geq \rho\}$, the equation $A \vartriangle B = (X \setminus A) \vartriangle (X \setminus B)$, which holds for all $A, B \subset X$, shows

$$\mu(V_\rho \vartriangle \{h < \rho\}) = 0\,. \tag{6}$$

In other words, up to μ-zero measures, M_ρ and V_ρ are the ρ-level sets of *all* μ-densities h of P. The following lemma shows that there even exists a μ-density h of P such that $M_\rho = \{h \geq \rho\}$.

Lemma 5 *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 $\rho \in \mathbb{R}$, the following statements are equivalent:*

i) P is regular at level ρ.

ii) There exists a μ-density $h : X \to [0, \infty)$ of P such that $\{h \geq \rho\}$ is closed.

iii) There exists a μ-density $h : X \to [0, \infty)$ of P such that $M_\rho = \{h \geq \rho\}$.

In particular, if P has an upper semi-continuous μ-density, then P is regular at every level.

The previous results may suggest that for continuous densities h we actually have $M_\rho = \{h \geq \rho\}$. In general, however, this is *not* the case. To see this, consider, e.g. a Lebesgue density that has a *strict* local maximum at $x^* \in X$ and the corresponding level set $\{h \geq \rho\}$ for $\rho := h(x^*)$. Moreover note, that not every probability measure P is regular. Indeed, it is possible to construct a Lebesgue-absolutely continuous probability measure P on $[0, 1]$ that is not regular for a continuous range of levels ρ.

Besides the regularity, we need another notion ensuring that certain topological operations on the level sets do not change their mass.

Definition 6 *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. We say that P is normal at level $\rho^* \geq 0$, if*

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

where $\bar{M}_{\rho^*} := \bigcup_{\rho > \rho^*} M_\rho$ and $\dot{M}_{\rho^*} := \bigcup_{\rho > \rho^*} \mathring{M}_\rho$.

The following two lemmata provide sufficient conditions for normality. We begin with continuous densities.

Lemma 7 *Let (X, d) be a complete separable metric space, μ be a σ-finite measure on X with $\operatorname{supp} \mu = X$, and P be a μ-absolutely continuous probability measure on X that has a continuous μ-density $h : X \to [0, \infty)$. Then, for all $\rho^* \geq 0$, we have*

$$\dot{M}_{\rho^*} = \bar{M}_{\rho^*}\,,$$

and hence P is normal at every level.

The have already mentioned that regularity is ensured if there exists a μ-density h of P with $\mu(\partial\{h \geq \rho\}) = 0$. The next lemma shows that this is also a sufficient for normality.

Lemma 8 *Let (X, d) be a complete separable metric space, μ be a σ-finite measure on X with $\operatorname{supp} \mu = X$, and P be a μ-absolutely continuous probability measure on X that has a μ-density $h : X \to [0, \infty)$ such that there exists a $\rho^* \geq 0$ with*

$$\mu(\partial\{h \geq \rho\}) = 0$$

for all $\rho > \rho^*$. Then P is regular at every level $\rho > \rho^*$ and normal at every level $\rho \geq \rho^*$.

2.2. 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.

Let us begin by introducing some notations. To this end, let (X, d) be a compact metric space. 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 $B \subset X$. Furthermore, for $\delta > 0$, we define the δ-tube around A by

$$T_\delta(A) := \{x \in X : d(x, A) \le \delta\}\,.$$

Lemma 27 in Subsection B collects some simple but useful facts about the δ-tube around A. Let us further recall the definition of τ-connected sets.

Definition 9 *Let (X, d) be a compact 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 $\mathcal{C}_\tau(A)$. In addition, we define $\mathcal{C}_\tau(\emptyset) := \emptyset$.

Not surprisingly, the τ-connected components of $A \subset X$ are τ-connected, see Lemma 28, and we always have $|\mathcal{C}_\tau(A)| < \infty$ and $d(A', A'') \ge \tau$ for all $A', A'' \in \mathcal{C}_\tau(A)$, see Lemma 29. Finally, if A is closed, so are the τ-connected components of A.

In the following, we often have to compare the τ-connected components of subsets $A \subset B$. The next lemma presents the fundamental tool for this task.

Lemma 10 *Let (X, d) be a compact metric space, $A \subset B$ be two non-empty subsets of X and $\tau > 0$. Then there exists exactly one map $\zeta : \mathcal{C}_\tau(A) \to \mathcal{C}_\tau(B)$ such that*

$$A' \subset \zeta(A')\,, \qquad\qquad A' \in \mathcal{C}_\tau(A)\,.$$

We call ζ the τ-connected components relating map (τ-CCRM) between A and B. Moreover, we sometimes write $\zeta_{A,B} := \zeta$ when we have to emphasize the involved pair (A, B).

Note that in general, the map ζ_τ is neither injective or surjective. Informally speaking, ζ_τ is injective, if and only if no τ-connected component of B merges two τ-connected components of A, while ζ_τ is surjective, if and only if B does not possess new τ-connected components, i.e. there is no τ-connected component B' of B with $B' \subset B \setminus A$. The next lemma introduces a very useful arithmetic property of τ-CCRMs.

Lemma 11 *Let (X, d) be a compact metric space, $A \subset B \subset C$ be three non-empty subsets of X and $\tau > 0$. Then the τ-CCRMs of these sets satisfy*

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

Let us now turn to the topological notion of connectivity that will be used in the definition of clusters. To this end, recall from topology that an $A \subset X$ is called connected, if, for every pair $A', A'' \subset A$ of 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. It is well-known that these components form a partition of A and that every component is closed if A itself is closed. In the following, we denote the set of topologically connected components of A by $\mathcal{C}(A)$. Furthermore, to clearer distinct connected sets and components from τ-connected sets and components, we often call the former *topologically* connected.

It can be easily shown, see Lemma 33, that, for compact metric spaces (X, d), an $A \subset X$ 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 12 *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 $\tau > 0$, there exists exactly one map $\zeta : \mathcal{C}(A) \to \mathcal{C}_\tau(A)$ with

$$A' \subset \zeta(A'), \qquad\qquad A' \in \mathcal{C}(A).$$

Moreover, ζ, which we call the connected components relating map (CCRM) of A, is surjective.

ii) If $|\mathcal{C}(A)| < \infty$, we have

$$\tau_A^* := \min\{d(A', A'') : A', A'' \in \mathcal{C}(A) \text{ with } A' \neq A''\} > 0, \tag{7}$$

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 CCRM $\zeta : \mathcal{C}(A) \to \mathcal{C}_\tau(A)$ 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.

Our next goal is to find a connected components relating map for topologically connected components. This is done in the following lemma, which is a direct consequence of Lemma 12, and whose proof is therefore omitted.

Lemma 13 *Let (X, d) be a compact metric space, $A \subset B$ be two non-empty closed subsets of X that both have finitely many topologically connected components. Then there exists exactly one map $\zeta : \mathcal{C}(A) \to \mathcal{C}(B)$ such that*

$$A' \subset \zeta(A'), \qquad\qquad A' \in \mathcal{C}(A).$$

In the following, we call ζ the topologically connected components relating map (top-CCRM) between A and B. For $\tau^* := \min\{\tau_A^*, \tau_B^*\}$ and all $\tau \in (0, \tau^*]$, we have $\zeta = \zeta_\tau$, where ζ_τ is the τ-CCRM between A and B.

Since top-CCRMs are τ-CCRMs for all sufficiently small $\tau > 0$, the composition formula presented in Lemma 11 also holds for top-CCRMs. Moreover, by a straightforward modification of the proof of Lemma 11, we see that it also holds, if some (or all) top-CCRMs or τ-CCRMs are replaced by the CCRMs found in part *i)* of Lemma 12.

The quantity τ_A^* defined in (7) will play a crucial role in our analysis. However, we need to consider this quantity for more than one set, the next lemma establishes a relation between τ_A^* and τ_B^* if $A \subset B$.

Lemma 14 *Let (X, d) be a compact metric space, $A \subset B$ be two non-empty closed subsets of X that both have finitely many topologically connected components. If the top-CCRM $\zeta : \mathcal{C}(A) \to \mathcal{C}(B)$ is injective, we have $\tau_A^* \geq \tau_B^*$.*

The following lemma establishes properties for the τ-CCRM between a set A and $T_\delta(A)$. Roughly speaking, it states, that the τ-connected component structure of $T_\delta(A)$ is identical to that of A, if $\tau > 0$ and $\delta > 0$ are sufficiently small.

Lemma 15 *Let (X, d) be a compact metric space, $A \subset X$ be a non-empty subset of X. Then, for all $\delta > 0$ and $\tau > \delta$, the following statements hold:*

i) *The set $T_\delta(A')$ is τ-connected for all $A' \in \mathcal{C}_\tau(A)$.*

ii) *The τ-CCRM $\zeta : \mathcal{C}_\tau(A) \to \mathcal{C}_\tau(T_\delta(A))$ is surjective.*

iii) *If A is closed and $|\mathcal{C}(A)| < \infty$, there exist $\tau^* > 0$ and $\delta^* > 0$ such that, for all $\tau \in (0, \tau^*]$ and $\delta \in (0, \delta^*]$ with $\tau > \delta$, the τ-CCRM $\zeta : \mathcal{C}_\tau(A) \to \mathcal{C}_\tau(T_\delta(A))$ is actually bijective. Moreover, we can choose τ^* and δ^* by the equation*

$$3\tau^* = 3\delta^* = \tau_A^* . \tag{8}$$

2.3. Clusters

In this subsection, we introduce our notion of clusters. We further present some results describing how robust the clusters are against horizontal blurriness. Finally, we introduce a notion that excludes thin bridges and cusps.

Let us begin with the following definition that describes distributions that have clusters.

Definition 16 *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. Then we say that P can be topologically clustered between the critical levels $\rho^* \geq 0$ and $\rho^{**} > \rho^*$, if P is normal at level ρ^* and, for all $\rho \in [0, \rho^{**}]$, the following conditions hold:*

i) *The set M_ρ has either one or two topologically connected components.*

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

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

iv) *P is regular at level ρ.*

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 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 $\zeta_\rho : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_\rho)$ for the top-CCRM. Moreover, let A_1 and A_2 be the topologically connected components of $M_{\rho^{**}}$. Then the sets*

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

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.

Using finitely many samples, we can only expect estimates of the level sets M_ρ that are both vertically and *horizontally* blurry. To address the latter issue, we define, for $\delta > 0$ and $\rho \geq 0$, the sets

$$\begin{aligned} M_{\rho,\delta} &:= T_\delta(M_\rho) \\ V_{\rho,\delta} &:= X \setminus T_\delta(X \setminus M_\rho) = X \setminus T_\delta(V_\rho)\,. \end{aligned}$$

Note that $M_{\rho,\delta}$ is obtained from M_ρ by adding a δ-tube, while $V_{\rho,\delta}$ is obtained from M_ρ by removing a δ-tube. Our next goal is to investigate, how the component structure of M_ρ is preserved under these operations. The first result in this direction establishes some permanence properties that hold without further assumptions. To appreciate its rather theoretically appearing statements recall from the previous subsection that CCRMs between two sets A and B are bijective, if *a)* the components of A are not glued together in B and *b)* every component in B already appears in A.

Theorem 18 *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 all $\varepsilon^* > 0$ with $\rho^* + \varepsilon^* \leq \rho^{**}$, we define $\delta_{\varepsilon^*} > 0$ and $\tau_{\varepsilon^*} > 0$ by*

$$3\delta_{\varepsilon^*} = 3\tau_{\varepsilon^*} = \tau^*_{M_{\rho^*+\varepsilon^*}}\,, \tag{9}$$

where $\tau^*_{M_{\rho^*+\varepsilon^*}} > 0$ is the quantity considered in Lemma 12. Then, for all $\delta \in (0, \delta_{\varepsilon^*}]$, $\tau \in (0, \tau_{\varepsilon^*}]$ with $\delta < \tau$, and all $\rho \in [0, \rho^{**}]$, the following statements hold:

i) The set $M_{\rho,\delta}$ has either one or two τ-connected components.

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

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

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

The first three statements of Theorem 18 basically show that the connected component structure of M_ρ is not changed when δ-tubes are added. Moreover, the assumed bijectivity of $\zeta^{**} : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}_\tau(M_{\rho^{**},\delta})$ in *iv)* means that the τ-connected component structure of $M_{\rho^{**}}$ is not changed by removing δ-tubes, and the corresponding conclusion essentially states that this is actually true for all levels $\rho \in [\rho^* + \varepsilon^*, \rho^{**}]$.

To ensure the same stability for removing δ-tubes, we need the following additional assumption.

Definition 19 *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 *that can be topologically clustered between the critical levels* ρ^* *and* ρ^{**}*. Then we say that* P *has two thick clusters of order* $\gamma \in (0, 1]$*, if there exist* $c \geq 1$ *and* $\tilde{\delta}_0 \in (0, 1]$ *such that, for all* $\delta \in (0, \tilde{\delta}_0]$*,* $\rho \in [0, \rho^{**}]$*, we have*

$$d(x, V_{\rho,\delta}) < c\,\delta^\gamma\,, \qquad\qquad x \in M_\rho\,.$$

In this case, we call $\psi : (0,\infty) \to (0,\infty)$*, defined by* $\psi(\delta) := 2c\delta^\gamma$*, the corresponding thickness function.*

Roughly speaking, the definition above ensures that every point of M_ρ is close to the set $V_{\rho,\delta}$ that results from removing a δ-tube from M_ρ. Intuitively, this excludes very thin cusps and bridges, where the thinness and length of both is controlled by γ. Note that an assumption of similar spirit is often made in density-based cluster analysis, we refer, e.g., to Cuevas et al. (2000); Rigollet (2007) for some examples in this direction.

Moreover note that, 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, it is then straightforward to show that P actually has two thick clusters of order $\gamma = 1$, and that we can further use every constant $c > 1$. Consequently, we can, e.g., consider the thickness function $\psi(\delta) := 3\delta$, $\delta > 0$.

The next result, which is the counterpart of Theorem 18, shows that, for thick clusters, the connected component structure of M_ρ is not changed when removing δ-tubes.

Theorem 20 *Let* (X, d) *be a compact metric space,* μ *be a finite measure on* X*, and* P *be a* μ*-absolutely continuous probability measure on* X *that has two thick clusters of order* $\gamma \in (0, 1]$ *between the critical levels* ρ^* *and* ρ^{**}*. Let* ψ *be the corresponding thickness function. Moreover, for some fixed* $\varepsilon^* > 0$ *with* $\varepsilon^* \leq \rho^{**} - \rho^*$ *we define* $\delta_{\varepsilon^*} > 0$ *and* $\tau_{\varepsilon^*} > 0$ *by* (9)*. Then, for all* $\delta \in (0, \delta_{\varepsilon^*}]$*,* $\tau \in (0, \tau_{\varepsilon^*}]$ *with* $\psi(\delta) < \tau$ *and* $\delta \leq \tilde{\delta}_0$*, and all* $\rho \in [0, \rho^{**}]$*, the following statements hold:*

i) *The set* $V_{\rho,\delta}$ *has either one or two* τ*-connected components.*

ii) *The* τ*-CCRM* $\zeta^{**} : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}_\tau(M_{\rho^{**},\delta})$ *is bijective.*

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

Intuitively, considering $\mathcal{C}_\tau(V_{\rho,\delta})$ rather than $\mathcal{C}(V_{\rho,\delta})$ means that we add a τ-tube around $V_{\rho,\delta}$. By Theorem 20, the thickness of the level sets then ensure that $\mathcal{C}_\tau(V_{\rho,\delta})$ 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 consistency

In this section, we introduce our clustering algorithm and present our main results on its clustering ability.

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 regularly behaved. To this end, we need the diameter of a subset $A \subset X$, that is,

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

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

Definition 21 *Let* (X, d) *be a compact metric space and* μ *be a finite measure on* X *with* $\operatorname{supp} \mu = X$*. If there exist constants* $d_X > 0$ *and* $\kappa_X > 0$ *such that, for all* $\delta \in (0, 1]$*, there exists a finite partition* $\mathcal{A}_\delta = (A_1, \dots, A_m)$ *of* X *such that*

$$\begin{aligned} \operatorname{diam} A_i &\leq \delta, \\ m &\leq \kappa_X \delta^{-d_X}, \\ \mu(A_i) &\geq \kappa_X^{-1} \delta^{d_X} \end{aligned}$$

for all $i = 1, \dots, m$*, then we say that the triple* (X, d, μ) *admits uniform* d_X*-dimensional partitions. Moreover, we call each such* $\mathcal{A}_\delta$ *a* δ*-uniform partition of* X*.*

The easiest yet most important examples of uniform partitions are hypercube partitions. To be more precise, let $X := [0, 1]^d$, the d-dimensional cube equipped with the metric defined by the supremum norm $\|\cdot\|_{\ell_\infty^d}$. 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 $\mathcal{A}_\delta$ for the usual partition of $[0, 1]^d$ into hypercubes of length h. Then, for each $A_i \in \mathcal{A}_\delta$, we clearly have $\operatorname{diam} A_i = h \leq \delta$ and $\lambda^d(A_i) = h^d \geq 2^{-d}\delta^d$, where λ^d denotes the d-dimensional Lebesgue measure. Moreover, we obviously have $|\mathcal{A}_\delta| = h^{-d} \leq 2^d\delta^{-d}$, where $|\mathcal{A}_\delta|$ denotes the cardinality of $\mathcal{A}$. Consequently, $\mathcal{A}_\delta$ is an δ-uniform partition of X with $d_X := d$ and $\kappa_X := 2^d$.

Let us now assume that (X, d, μ) admits uniform d_X-dimensional partitions. Moreover, for fixed $\delta > 0$, let $\mathcal{A}_\delta = (A_1, \dots, A_m)$ be a δ-uniform partition of X. Given a probability measure P on X, we then define the corresponding histogram by

$$\bar{h}_{P,\mathcal{A}_\delta}(x) := \sum_{j=1}^m \frac{P(A_j)}{\mu(A_j)} \cdot \mathbf{1}_{A_j}(x), \qquad x \in X,$$

where $\mathbf{1}_A$ denotes the indicator function of a set A. If the partition is known from the context, we further write $\bar{h}_{P,\delta} := \bar{h}_{P,\mathcal{A}_\delta}$ to simplify notation. Let us now assume that we have a data set $D = (x_1, \dots, x_n) \in X^n$. In a slight abose of notation, we then denote the corresponding empirical measure by D, that is $D := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, where δ_x denotes 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

$$\bar{h}_{D,\mathcal{A}_\delta}(x) = \sum_{j=1}^m \frac{D(A_j)}{\mu(A_j)} \cdot \mathbf{1}_{A_j}(x)\,, \qquad\qquad x \in X. \tag{10}$$

Our first result in this section shows, that, for i.i.d. observations D, the histogram $\bar{h}_{D,\mathcal{A}_\delta}$ approximates $\bar{h}_{P,\mathcal{A}_\delta}$ *uniformly.*

Theorem 22 *Let (X, d) be a compact metric space and μ be a finite measure on X such that (X, d, μ) admits uniform d_X-dimensional partitions. Moreover, let P be a probability measure on X, $\delta > 0$, and $\mathcal{A}_\delta$ be a δ-uniform partition of X. Then, for all $n \geq 1$ and all $\varepsilon > 0$, we have*

$$P^n\Big(\big\{D \in X^n : \|\bar{h}_{D,\mathcal{A}_\delta} - \bar{h}_{P,\mathcal{A}_\delta}\|_\infty < \varepsilon\big\}\Big) \geq 1 - 2\kappa_X \exp\Big(-d_X \ln\delta - \frac{2\delta^{2d_X}\varepsilon^2 n}{\kappa_X^2}\Big)\,.$$

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

$$P^n\Big(\big\{D \in X^n : \|\bar{h}_{D,\mathcal{A}_\delta} - \bar{h}_{P,\mathcal{A}_\delta}\|_\infty < \varepsilon\big\}\Big) \geq 1 - 2\kappa_X \exp\Big(-d_X \ln\delta - \frac{3\varepsilon^2\delta^{d_X} n}{\kappa_X(6\|h\|_\infty + 2\varepsilon)}\Big)\,.$$

Our clustering algorithm will rely on an empirical histogram. To be more precise, let us assume that, for some fixed $\varepsilon > 0$, we have a function $\hat{h} : X \to \mathbb{R}$ that is a uniform ε-approximate of $\bar{h}_{P,\mathcal{A}_\delta}$, i.e.,

$$\|\hat{h} - \bar{h}_{P,\mathcal{A}_\delta}\|_\infty \leq \varepsilon\,.$$

Note that, by Theorem 22, empirical histograms are such ε-approximates with high probability. We write

$$\hat{f}_\rho := \operatorname{sign}(\hat{h} - \rho)\,, \qquad\qquad \rho \geq 0, \tag{11}$$

where $\operatorname{sign}(\cdot)$ denotes the usual sign function, that is, $\operatorname{sign} t := 1$ if $t \geq 0$ and $\operatorname{sign} t := -1$, otherwise. Since $\bar{h}_{P,\mathcal{A}_\delta}$ can be viewed as an approximation of the μ-densities of P, $\hat{h}$ can also be viewed as such an approximation. Following this intuition, $\{\hat{f}_\rho = -1\}$ and $\{\hat{f}_\rho = 1\}$ can be viewed as approximations of the sets V_ρ and M_ρ, respectively. However, using finitely many samples, we can only expect estimates of the level sets M_ρ that are both horizontically and *vertically* blurry. The following lemma makes this intuition precise with the help of the sets $V_{\rho,\delta}$ and $M_{\rho,\delta}$ defined earlier.

Algorithm 3.1 Estimate clusters with the help of empirical histograms

Require: Some $\delta > 0$, $\tau > 0$, and $\varepsilon > 0$.
A δ-uniform partition $\mathcal{A}_\delta$ of X.
A dataset $D \in X^n$.

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

Compute the empirical histogram $\bar{h}_{D,\mathcal{A}_\delta}$.
$\rho \leftarrow -\varepsilon$
repeat
    $\rho \leftarrow \rho + \varepsilon$
    Compute $\hat{f}_\rho$ by (11).
    Identify the τ-connected components $B_1', \dots, B_M'$ of $\{\hat{f}_\rho = 1\}$ satisfying

$$B_i' \cap \{\hat{f}_{\rho+2\varepsilon} = 1\} \neq \emptyset.$$

until $M \neq 1$
Compute $\hat{f}_{\rho+2\varepsilon}$ by (11).
Identify the τ-connected components $B_1', \dots, B_M'$ of $\{\hat{f}_{\rho+2\varepsilon} = 1\}$ satisfying

$$B_i' \cap \{\hat{f}_{\rho+4\varepsilon} = 1\} \neq \emptyset.$$

return ρ and $B_1', \dots, B_M'$.

Lemma 23 *Let (X, d) be a compact metric space and μ be a finite measure on X such that (X, d, μ) admits uniform d_X-dimensional partitions. Moreover, let P be a μ-absolutely continuous probability measure on X and $\hat{h} : X \to \mathbb{R}$ be a function with $\|\hat{h} - \bar{h}_{P,\mathcal{A}_\delta}\|_\infty \leq \varepsilon$ for some $\varepsilon > 0$. Then, for all $\delta > 0$ and $\rho \geq 0$, and $\hat{f}_\rho$ defined by (11), we have:*

i) If P is regular at the level $\rho + \varepsilon$, then $V_{\rho+\varepsilon,\delta} \subset \{\hat{f}_\rho = 1\}$.

ii) If P is regular at the level $\rho - \varepsilon$, then $\{\hat{f}_\rho = 1\} \subset M_{\rho-\varepsilon,\delta}$.

Motivated by Lemma 23, our next goal is to relate the τ-connected components of our estimate $\{\hat{f}_\rho = 1\}$ to the τ-connected components of $V_{\rho,\delta}$.

Theorem 24 *Let (X, d) be a compact metric space, μ be a finite measure on X such that (X, d, μ) admits uniform d_X-dimensional partitions, and P be a μ-absolutely continuous probability measure on X that has two thick clusters of order $\gamma \in (0, 1]$ between the critical levels ρ^* and ρ^{**}. Let ψ be the corresponding thickness function. Moreover, for some fixed $\varepsilon^* > 0$ with $\varepsilon^* \leq \rho^{**} - \rho^*$ we define $\delta_{\varepsilon^*} > 0$ and $\tau_{\varepsilon^*} > 0$ by (9). Let us further fix some $\varepsilon \in (0, \varepsilon^*]$, $\epsilon \geq 0$, $\delta \in (0, \delta_{\varepsilon^*}]$ with $\delta \leq \tilde{\delta}_0$, and $\rho \in [0, \rho^{**} - 3\varepsilon - \epsilon]$. In addition, let $\hat{h} : X \to \mathbb{R}$ be a uniform ε-approximate of $\bar{h}_{P,\delta}$ and $\hat{f}_\rho$ be the function defined by (11). Then, for all $\tau \in (0, \tau_{\varepsilon^*}]$ with $\psi(\delta) < \tau$, the following disjoint union holds*

$$\mathcal{C}_\tau(\{\hat{f}_\rho = 1\}) = \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})) \cup \left\{B' \in \mathcal{C}_\tau(\{\hat{f}_\rho = 1\}) : B' \cap \{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\} = \emptyset\right\},$$

where $\zeta : \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}) \to \mathcal{C}_\tau(\{\hat{f}_\rho = 1\})$ is the τ-CCRM.

Theorem 24 shows that eventually all τ-connected components B' of our estimate $\{\hat{f}_\rho = 1\}$ of M_ρ are either contained in $\zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$ or satisfy $B' \cap \{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\} = \emptyset$. Now, the latter components are easy to identify and remove, and therefore we have a device that allows us to eventually identify exactly the τ-connected components B' that are contained in $\zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$. This suggests that, for sufficiently small $\delta > 0$, $\tau > 0$, $\varepsilon > 0$, and $\epsilon \geq 0$, we only need to scan through the values of ρ. Algorithm 3.1 formalizes this idea.

Note that Algorithm 3.1 stops if either $M > 1$ or $M = 0$ components are identified for the current level set ρ. Moreover, the latter is eventually satisfied, since

$$\|\bar{h}_{D,\mathcal{A}_\delta}\|_\infty \leq \kappa_X \delta^{-d_X} \sum_{i=1}^{m} D(A_i) = \kappa_X \delta^{-d_X} ,$$

yields $\{\hat{f}_\rho = 1\} = \emptyset$ for all $\rho > \kappa_X \delta^{-d_X}$. In the following, we denote the level returned by Algorithm 3.1 by $\rho^*(D)$. The following theorem shows that $\rho^*(D)$ is close to ρ^*, whenever the empirical histogram approximates the true histogram.

Theorem 25 *Let (X, d) be a compact metric space and μ be a finite measure on X such that (X, d, μ) admits uniform d_X-dimensional partitions. Moreover, let P be a μ-absolutely continuous probability measure on X that has two thick clusters of order $\gamma \in (0, 1]$ between the critical levels ρ^* and ρ^{**} and let ψ be the corresponding thickness function. Moreover, we fix an $\varepsilon^* > 0$ that satisfies $\varepsilon^* < (\rho^{**} - \rho^*)/8$ and define $\delta_{\varepsilon^*} > 0$ and $\tau_{\varepsilon^*} > 0$ by (9). Then, for all fixed $n \geq 1$, $\varepsilon \in (0, \varepsilon^*]$, $\delta \in (0, \delta_{\varepsilon^*}]$, and $\tau \in (0, \tau_{\varepsilon^*}]$ with $\psi(\delta) < \tau$ and $\delta \leq \tilde{\delta}_0$, and all data sets $D \in X^n$ for which $\|\bar{h}_{D,\mathcal{A}_\delta} - \bar{h}_{P,\mathcal{A}_\delta}\|_\infty \leq \varepsilon$ holds, the following statements are true:*

i) $\rho^*(D) \in [\rho^* - \varepsilon, \rho^* + \varepsilon^* + 2\varepsilon]$.

ii) $|\mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta})| = 2$ *and the τ-CCRM $\zeta : \mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta}) \to \mathcal{C}_\tau(\{\hat{f}_{\rho^*(D)+2\varepsilon} = 1\})$ is injective.*

iii) Algorithm 3.1 returns the two τ-connected components of $\zeta(C_\tau(V_{\rho^*(D)+3\varepsilon,\delta}))$.

iv) There exist CCRMs $\zeta_{\rho^{**}} : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}(M_{\rho^{**}})$ and $\zeta_{\rho^*(D)+3\varepsilon} : \mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta}) \to \mathcal{C}(M_{\rho^*(D)+3\varepsilon})$ such that the following diagram

$$\begin{array}{ccc}
\mathcal{C}_\tau(V_{\rho^{**},\delta}) & \xrightarrow{\zeta_{\rho^{**}}} & \mathcal{C}(M_{\rho^{**}}) \\
\Big\downarrow {\scriptstyle \zeta_{\rho^{**},\rho^*(D)+3\varepsilon}} & & \Big\downarrow {\scriptstyle \tilde{\zeta}} \\
\mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta}) & \xrightarrow[\zeta_{\rho^*(D)+3\varepsilon}]{} & \mathcal{C}(M_{\rho^*(D)+3\varepsilon})
\end{array}$$

commutes, where $\zeta_{\rho^{**},\rho^*(D)+3\varepsilon}$ is the τ-CCRM and $\tilde{\zeta}$ is the top-CCRM. Moreover, every map in the diagram is bijective.

To fully appreciate Theorem 25 let us assume that we are in the situation of this theorem. Moreover, let A_1 and A_2 be the topologically connected components of $M_{\rho^{**}}$ and

V_1'' and V_2'' be the τ-connected components of $V_{\rho^{**},\delta}$. In addition, let V_1' and V_2' be the τ-connected components of $\mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta})$ and $B_1(D)$ and $B_2(D)$ be the components returned by Algorithm 3.1. By Theorem 25, we may assume without loss of generality that $V_i'' \subset A_i$, $V_i'' \subset V_i'$, and $V_i' \subset B_i(D)$ for $i = 1, 2$. This yields $V_i'' \subset B_i(D)$ and $V_i'' \subset A_i^*$, that is, $V_i'' \subset B_i(D) \cap A_i^*$. Consequently, the returned components $B_i(D)$ contain a chunk of the desired clusters A_i^*, $i = 1, 2$. Our next and final goal is to show that $B_i(D) \vartriangle A_i^*$ actually becomes arbitrarily small. To this end, we assume in the following that Algorithm 3.1 always returns two components, denoted by $B_1(D)$ and $B_2(D)$. Note that this can be easily enforced by a simple modification of its return statement in line 10 of its pseudo-code.

With these preparations, we are in the position to put all pieces together. This is done in the following main result that establishes a type of clustering consistency for Algorithm 3.1.

Theorem 26 *Let (X, d) be a compact metric space and μ be a finite measure on X such that (X, d, μ) admits uniform d_X-dimensional partitions. Moreover, let P be a μ-absolutely continuous probability measure on X that has two thick clusters of order $\gamma \in (0, 1]$ between the critical levels ρ^* and ρ^{**} and let ψ be the corresponding thickness function. Furthermore, let (ε_n), (δ_n), and (τ_n) be strictly positive sequences converging to zero such that $\psi(\delta_n) < \tau_n$ and*

$$d_X \kappa_X^2 \ln \delta_n + 2\delta_n^{2d_X} \varepsilon_n^2 n \to \infty\,.$$

For $n \geq 1$ consider Algorithm 3.1 with the input parameters ε_n, δ_n, and τ_n. Then, $\rho^*(D) \to \rho^*$ in probability P^∞ for $n \to \infty$ and, for all $\epsilon > 0$, we have

$$\lim_{n\to\infty} P^n\Big(\big\{D \in X^n : \mu(B_1(D) \vartriangle A_1^*) + \mu(B_2(D) \vartriangle A_2^*) \leq \epsilon\big\}\Big) = 1\,.$$

Here we use the numbering convention of $B_1(D)$ and $B_2(D)$ described in the paragraph above.

Theorem 26 shows that Algorithm 3.1 asymptotically recovers the clusters A_1^* and A_2^*, whenever the distribution P has clusters that are thicker than a pre-described order. In other words, as soon as we assume a minimal thickness, we are able to recover the clusters. Moreover, we have already mentioned previously, that, for intervals $X \subset \mathbb{R}$, we automatically have thickness of order $\gamma = 1$, and hence Algorithm 3.1 asymptotically recovers the clusters, e.g., for every distribution P on intervals that can be topologically clustered. Note that it is easy to construct distributions in this class that do not have a continuous density, for example consider the distribution P on $X := [0, 1]$ that has the Lebesgue density $h := \mathbf{1}_{[0,1/4]\cup[3/4,1]} + 0.5 \cdot \mathbf{1}_{[0,1]}$, and whose clusters are given by $A_1^* := [0, 1/4]$ and $A_2^* := [3/4, 1]$. It is obvious, that similar constructions can also be made in higher dimensions, and finally, such examples are, of course, by no means the only examples of distributions for which the clusters can be recovered by Algorithm 3.1.

Furthermore, note that although Theorem 26 presents an asymptotic result, its entire proof uses finite-sample results and estimates, i.e., we could have also stated a result of the form: if the algorithm parameters are smaller than some thresholds determined in the proof of Theorem 26, then $\mu(B_1(D) \vartriangle A_1^*) + \mu(B_2(D) \vartriangle A_2^*) \leq \epsilon$ holds with probability P^n not smaller than some value also determined in the proof. Since the presented algorithm was meant to be a proof-of-concept rather than an algorithm actually used in dimensions greater than, say, 2 or 3, we decided to omit the technically rather cumbersome formulation of such a result.

4. Discussion

The goal of this work was to provide the first density-based clustering algorithm that, under mild assumptions on the density h, can choose the density level in a data-dependent and asymptotically optimal way and completely recovers the corresponding clusters even when they touch each other.

Although the algorithm works in theory, we do not expect it to perform well in most practical situations. Let us therefore briefly describe what should be done to obtain a more interesting algorithm:

- The algorithm should be based on density level set estimators that are better than histograms. A natural first alternative in this direction would be kernel density rules since they have already been successfully considered in the single level clustering problem.
- The algorithm and its analysis should be extended to situations in which P has either only $N = 1$ or $N > 2$ clusters. Note that the first scenario can probably be rather easily analyzed with our techniques, while the second scenario probably needs a refined algorithm, first. In this direction note that our proofs already show that the algorithm recovers at least two of the N clusters. So far, however, we cannot ensure that it accidentally glues some of the N clusters together.
- The algorithm should not only determine the level ρ in a data-dependent way, but also the other algorithm parameters δ, ε, and τ. Analogous, data-dependent choices should be investigated for other underlying density level set estimators.
- Finally, the algorithm does not necessarily need to stop once it leaves the loop. Instead, it could save the found clusters together with the level and reenter the loop recursively for both clusters. This way it seems plausible, that the algorithm is actually able to recover all clusters contained in the cluster tree.

Besides these issues that are more of practical interest, it would also be helpful to investigate the following, more theoretically orientated questions:

- Can we replace the thickness assumption by some other assumption such as Hölder continuity or rectifiable cluster boundaries, which have already been considered in the literature. Or, more challenging, is it possible in dimension $d \geq 2$ to remove such assumptions at all?
- Can we replace density level sets by level sets of generalized densities in the sense of Rinaldo and Wasserman (2010)?

Acknowledgments

I like to thank Michael Eisermann for many fruitful discussions.

References

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 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. *The Canadian Journal of Statistics*, 28:367–382, 2000.

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer, New York, 1996.

J. A. Hartigan. *Clustering Algorithms.* John Wiley & Sons, New York, 1975.

J.A. Hartigan. Consistency of single linkage for high-density clusters. *J. Amer. Statist. Assoc.*, 76:388–394, 1981.

M. Maier, M. Hein, and U. von Luxburg. Generalized density clustering. *Optimal construction of k-nearest neighbor graphs for identifying noisy clusters*, 410:1749–1764, 2009.

P. Rigollet. Generalization error bounds in semi-supervised classification under the cluster assumption. *J. Mach. Learn. Res.*, 8:1369–1392, 2007.

A. Rinaldo and L. Wasserman. Generalized density clustering. *Ann. Statist.*, 38:2678–2722, 2010.

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.

Appendix A. Proofs related to the definition of level sets

Proof of Lemma 1 The second inclusion has already been shown in (2), and hence it suffices to show the first. To show the first inclusion we fix an $x \in \{h \overset{\circ}{\geq} \rho\}$ and an open set U with $x \in U$. Then $\{h \overset{\circ}{\geq} \rho\} \cap U$ is open and non-empty, and hence $\operatorname{supp} \mu = X$ yields

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

By (3) we conclude that $x \in M_\rho$, that is we have shown $\{h \overset{\circ}{\geq} \rho\} \subset M_\rho$. Since M_ρ is closed, we then obtain the first inclusion. ■

Proof of Corollary 2 Clearly, we have $\{h > \rho\} \subset \{h \geq \rho\}$ and since $\{h > \rho\}$ is open, we conclude that $\{h > \rho\} \subset \{h \overset{\circ}{\geq} \rho\} \subset M_\rho$ by Lemma 1. This implies the first and, since $\{h > \rho\}$ is open, also the third inclusion. The second and forth inclusion also follows from Lemma 1 and the fact that $\{h \geq \rho\}$ is closed. ■

Proof of Lemma 3 Obviously, it suffice to show the first inclusion. To this end, 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

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

and hence we obtain $x \in M_{\rho_1}$ by (3). ■

Proof of Lemma 5 *i) ⇒ iii).* Let $\tilde{h} : X \to [0, \infty)$ be an arbitrary μ-density of P. We define

$$h(x) := \begin{cases} \tilde{h}(x) & \text{if } x \notin M_\rho \vartriangle \{\tilde{h} \geq \rho\} \\ \rho & \text{if } x \in M_\rho \setminus \{\tilde{h} \geq \rho\} \\ 0 & \text{if } x \in \{\tilde{h} \geq \rho\} \setminus M_\rho\,. \end{cases}$$

Since $\{h \neq \tilde{h}\} \subset M_\rho \vartriangle \{\tilde{h} \geq \rho\}$, the regularity shows that $\mu(\{h \neq \tilde{h}\}) = 0$, and hence h is a μ-density of P. Furthermore, for $x \in \{h \geq \rho\}$, we either have $x \in M_\rho \setminus \{\tilde{h} \geq \rho\} \subset M_\rho$ or

$$x \in X \setminus \big(M_\rho \vartriangle \{\tilde{h} \geq \rho\}\big) \cap \{\tilde{h} \geq \rho\} \subset M_\rho\,,$$

where in the last step we used that $x \notin A \vartriangle B$ together with $x \in B$ implies $x \in A$. Conversely, if $x \in M_\rho$, then $\tilde{h}(x) < \rho$ implies $h(x) = \rho$ and $\tilde{h}(x) \geq \rho$ implies $h(x) = \tilde{h}(x) \geq \rho$. These considerations show $\{h \geq \rho\} = M_\rho$.

iii) ⇒ ii). Since M_ρ is closed, this implication is trivial.

ii) ⇒ i). The inclusion (2) shows $M_\rho \subset \overline{\{h \geq \rho\}} = \{h \geq \rho\}$, and hence we obtain $\mu(M_\rho \setminus \{h \geq \rho\}) = \mu(\emptyset) = 0$.

Finally, if h is an upper semi-continuous μ-density, then $\{h \geq \rho\}$ is closed for all $\rho \in \mathbb{R}$. Consequently, P is regular at every level by the already proved implication from *ii)* to *i)*. ■

Proof of Lemma 7 The inclusion $\subset$ is trivial. To show the converse, we fix an $\rho > \rho^*$. Then there exists an $\rho' \in (\rho^*, \rho)$, and by Corollary 2 we thus find

$$M_\rho \subset \{h \geq \rho\} \subset \{h > \rho'\} \subset \mathring{M}_{\rho'}\,.$$

From this we easily derive the assertion. ■

Proof of Lemma 8 The regularity follows from (4). To show that that P is normal, we fix a $\rho_0 \geq \rho^*$. Because of the monotonicity of M_ρ in ρ, it then suffices to show that

$\mu(M_\rho \setminus \mathring{M}_\rho) = 0$ for all $\rho > \rho_0$. However, Lemma 1 ensures both $\{h \overset{\circ}{\geq} \rho\} \subset \mathring{M}_\rho$ and $M_\rho \subset \overline{\{h \geq \rho\}}$, and hence we obtain $M_\rho \setminus \mathring{M}_\rho \subset \partial\{h \geq \rho\}$. ■

Appendix B. Proofs related to basic properties of connected components

Lemma 27 *Let (X, d) be a compact metric space and $A, B \subset X$ be two subsets. Then the following statements hold:*

i) If A is closed, then $T_\delta(A) := \{x \in X : \exists x' \in A \text{ with } d(x, x') \leq \delta\}$.

ii) We have $d(A, B) \leq d(T_\delta(A), T_\delta(B)) + 2\delta$.

iii) We have

$$\bigcap_{\delta>0} T_\delta(A) = \overline{A}\,. \tag{12}$$

Proof of Lemma 27 *i).* For fixed $x \in T_\delta(A)$, there exists a sequence $(x_n) \subset A$ with $d(x, x_n) \leq \delta + 1/n$ for all $n \geq 1$. Since X is compact, we may assume without loss of generality that (x_n) converges to some $x' \in X$, and since we assumed that A is closed, we obtain $x' \in A$. Now we easily obtain the assertion from $d(x, x') \leq d(x, x_n) + d(x_n, x')$.

ii). Let us fix an $x \in T_\delta(A)$ and an $y \in T_\delta(B)$. 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

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

and by first letting $n \to \infty$ and then taking the infimum over all $x \in T_\delta(A)$ and $y \in T_\delta(B)$, we obtain the assertion.

iii). To show the inclusion $\supset$, we fix an $x \in \overline{A}$. Then there exists a sequence $(x_n) \subset A$ with $x_n \to x$ for $n \to \infty$. 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 $x \in T_\delta(A)$. To show the converse inclusion $\subset$, we fix an $x \in X$ that satisfies $x \in T_{1/n}(A)$ 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 $x_n \to x$ for $n \to \infty$. This shows $x \in \overline{A}$. ■

Lemma 28 *Let (X, d) be a compact metric space, $A \subset X$ be a non-empty subset and $\tau > 0$. Then every τ-connected component of A is τ-connected.*

Proof of Lemma 28 Let A' be a τ-connected component of A and $x, x' \in A'$. Then x and x' are τ-connected in A, and hence there exist $x_1, \ldots, x_n \in A$ such that $x_1 = x$, $x_n = x'$ and $d(x_i, x_{i+1}) < \tau$ for all $i = 1, \ldots, n-1$. Now, $d(x_1, x_2) < \tau$ shows that x_1 and x_2 are τ-connected in A, and hence they belong to the same τ-connected component, i.e. we have found $x_2 \in A'$. Iterating this argument, we find $x_i \in A'$ for all $i = 1, \ldots, n$. Consequently, x and x' are not only τ-connected in A, but also τ-connected in A'. This shows that A' is τ-connected. ■

Lemma 29 *Let (X, d) be a compact metric space, $A \subset X$ be a non-empty subset and $\tau > 0$. Then there exist only finitely many τ-connected components $A_1, \dots, A_m$ of A. Moreover, we have $d(A_i, A_j) \geq \tau$ for all $i \neq j$. Finally, if A is closed, these components are closed, too.*

Proof of Lemma 29 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 $x_i \to x$ for some $x \in X$. Since A is closed, we have $x \in A$, and hence $x \in A''$ for $A'' \in \mathcal{C}_\tau(A)$. By construction we find $d(A', A'') = 0$, and hence we obtain $A' = A''$ by the assertion that has been shown first. ■

Lemma 30 *Let (X, d) be a compact 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 30 *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. Then Lemma 29 shows that there exist finitely many τ-connected components $A_1, \dots, A_m$ of A. By definition, these components are non-empty, mutually disjoint, and satisfy $A = A_1 \cup \dots \cup A_m$. Moreover, Lemma 28 shows that each component is τ-connected, and since we assumed that A itself is not τ-connected, we conclude that $m \geq 2$. In addition, Lemma 29 shows $d(A_j, A_{j'}) \geq \tau$, whenever $j \neq j'$. Let us define $A^- := A_1$ and $A^+ := A_2 \cup \dots \cup A_m$. Then our previous considerations show that the subsets A^+ and A^- of A are non-empty and satisfy $A^+ \cup A^- = A$, $A^+ \cap A^- = \emptyset$, and $d(A^+, A^-) \geq \tau$. ■

Corollary 31 *Let (X, d) be a compact 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 31 The second assertion is a direct consequence of the first, and hence it suffice to show the first assertion. Now, by Lemma 29, there exist finitely many τ-connected components $B_1, \dots, B_m$ of B. Since we obviously have $A \subset B_1 \cup \dots \cup B_m$ it suffices to show $A \cap B_i = \emptyset$ for all but one index $i \in \{1, \dots, m\}$. Let us assume the converse, that is, there exist two indices $i, j \in \{1, \dots, m\}$ with $i \neq j$, $A \cap B_i \neq \emptyset$, and $A \cap B_j \neq \emptyset$. We write $A^- := A \cap B_i$ and $A^+ := A \cap (B \setminus B_i)$. Since $B_j \subset B \setminus B_i$, we obtain $A^+ \neq \emptyset$, and therefore, Lemma 30 shows $d(A^-, A^+) < \tau$. Consequently, there exist $x^- \in A^-$ and $x^+ \in A^+$ with $d(x^+, x^-) < \tau$. Now we obviously have $x^- \in B_i$, and by construction, we also find an index $i' \neq i$ with $x^+ \in B_{i'}$. Our previous inequality then yields $d(B_i, B_{i'}) < \tau$, while Lemma 29 shows $d(B_i, B_{i'}) \geq \tau$, that is, we have found a contradiction. ■

Lemma 32 *Let (X, d) be a compact 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 32 *i) ⇒ ii).* Follows from Lemma 29.

ii) ⇒ i). Let us fix an $A' \in \mathcal{C}_\tau(A)$. Then, by Corollary 31, every A_i with $A_i \cap A' \neq \emptyset$ satisfies $A_i \subset A'$. Since $A_1, \dots, A_m$ is a partition, 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 30 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. ■

Proof of Lemma 10 For $A' \in \mathcal{C}_\tau(A)$, Corollary 31 shows that there exists exactly $B' \in \mathcal{C}_\tau(B)$ with $A' \subset B'$. Setting $\zeta(A') := B'$ then gives the desired map and this map is uniquely determined since B' is. ■

Proof of Lemma 11 Clearly, $\zeta_{B,C} \circ \zeta_{A,B}$ maps from $C_\tau(A)$ to $C_\tau(C)$. Moreover, for $A' \in \mathcal{C}_\tau(A)$ we have $A' \subset \zeta_{A,B}(A')$ and for $B' := \zeta_{A,B}(A') \in \mathcal{C}_\tau(B)$ we have $B' \subset \zeta_{B,C}(B')$. Combining these inclusions we find $A' \subset \zeta_{B,C}(\zeta_{A,B}(A')) = \zeta_{B,C} \circ \zeta_{A,B}(A')$ for all $A' \in \mathcal{C}_\tau(A)$. By Lemma 10, $\zeta_{A,C}$ is the only map satisfying this property, and hence we conclude that $\zeta_{A,C} = \zeta_{B,C} \circ \zeta_{A,B}$. ■

Lemma 33 *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 33 *i)* $\Rightarrow$ *ii)*. Let us assume that A is not τ-connected for some $\tau > 0$. Then, by Lemma 29, there are finitely many τ-connected components $A_1, \ldots, A_m$ of A with $m > 1$. We write $A' := A_1$ and $A'' := A_2 \cup \cdots \cup A_m$. Then A' and A'' are non-empty, disjoint and $A' \cup A'' = A$ by construction. Moreover, Lemma 29 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 30 then shows that A is not τ-connected. ■

Proof of Lemma 12 *i)*. Let $A' \subset A$ be a topologically connected component of A and $\tau > 0$. Then we have already seen in Lemma 33 that A' is τ-connected, and since $A' \subset A$, Corollary 31 shows that there exists exactly one $A'' \in \mathcal{C}_\tau(A)$ with $A' \subset A''$. Consequently, $\zeta(A') := A''$ is the only possible definition of ζ. Moreover, if $A'' \in \mathcal{C}_\tau(A)$ is an arbitrary τ-connected component, then there exists an $x \in A''$, and to this x, there exists an $A' \in \mathcal{C}(A)$ with $x \in A'$. Corollary 31 then shows that $A' \subset A''$, and hence we obtain $\zeta(A') = A''$. In other words, ζ is surjective.

ii). Let $A_1, \ldots, A_m$ 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 33 shows that each A_i is τ-connected, and therefore Lemma 32 together with $d(A_i, A_j) \geq \tau_A^* \geq \tau$ for all $i \neq j$ yields $\mathcal{C}_\tau(A) = \{A_1, \ldots, A_m\}$. 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 29, 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 $\tau_A^* = \tau^*$. ■

Proof of Lemma 14 Let $A', A'' \in \mathcal{C}(A)$ 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 Lemma 12. Taking the infimum over all A' and A'' with $A' \neq A''$ yields the assertion. ■

Proof of Lemma 15 *i)*. Since $\tau > \delta$, there exist an $\varepsilon > 0$ with $\delta + \varepsilon < \tau$. For $x \in T_\delta(A')$, 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 $x, x'' \in T_\delta(A')$ is τ-connected.

ii). Let us fix an $A' \in \mathcal{C}_\tau(T_\delta(A))$ 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 29 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 $x \in T_\delta(A'')$, i.e. we have $T_\delta(A'') \cap A' \neq \emptyset$. Since $T_\delta(A'') \subset T_\delta(A)$, we then conclude that $T_\delta(A'') \subset A'$ by Corollary 31 and part *i)*. Furthermore, we clearly have $A'' \subset T_\delta(A'')$, and hence $\zeta(A'') = A'$.

iii). We write $A_1, \dots, A_m$ for the τ-connected components of A. For arbitrary $\tau > 0$ and $\delta > 0$, we first show that

$$T_\delta(A) = \bigcup_{i=1}^{m} T_\delta(A_i)\,. \tag{13}$$

Obviously, the inclusion „$\supset$" is trivial. To show the converse inclusion, we fix an $x \in T_\delta(A)$. Since A is compact, there then exists an $x' \in A$ with $d(x, x') \leq \delta$. Obviously, we further have $x' \in A_i$ for some component A_i, and hence we find $x \in T_\delta(A_i)$.

We now choose τ^* and δ^* by (8) and fix some $\tau \in (0, \tau^*]$ and $\delta \in (0, \delta^*]$. Moreover, let $A_1, \dots, A_m$ again be the τ-connected components of A. Since $\tau \leq \tau^* \leq \tau_A^*$, part *ii)* of Lemma 12 shows $\mathcal{C}(A) = \mathcal{C}_\tau(A)$, and consequently we obtain $d(A_i, A_j) \geq \tau_A^* = 3\tau^*$ for all $i \neq j$ by another application of part *ii)* of Lemma 12. Our next goal is to show that

$$d(T_\delta(A_i), T_\delta(A_j)) \geq \tau\,, \qquad\qquad i \neq j\,. \tag{14}$$

To this end, we fix an $x_i \in T_\delta(A_i)$ and an $x_j \in T_\delta(A_j)$. Then there exist $x_i' \in A_i$ and $x_j' \in A_j$ with $d(x_i, x_i') \leq \delta$ and $d(x_j, x_j') \leq \delta$, and therefore using $\delta \leq \delta^* = \tau^*$ we obtain

$$3\tau^* \leq d(x_i', x_j') \leq d(x_i', x_i) + d(x_i, x_j) + d(x_j, x_j') \leq 2\tau^* + d(x_i, x_j)\,.$$

Obviously, the latter together with $\tau^* \geq \tau$ implies (14).

Now part *i)* showed that each $T_\delta(A_i)$, $i = 1, \dots, m$, is τ-connected whenever $\tau > \delta$. Combining this with (13), (14), and Lemma 32, we thus see that $T_\delta(A_1), \dots, T_\delta(A_m)$ are the τ-connected components of $T_\delta(A)$. The bijectivity of ζ then follows from the surjectivity and a simple cardinality argument. ■

Appendix C. Proofs related to the identification of components

Proof of Theorem 18 *i).* Since $\tau > \delta$, part *ii)* of Lemma 15 and part *i)* of Lemma 12 yield

$$|\mathcal{C}_\tau(M_{\rho,\delta})| \leq |\mathcal{C}_\tau(M_\rho)| \leq |\mathcal{C}(M_\rho)| \leq 2\,. \tag{15}$$

ii). Let us fix a $\rho \in [\rho^* + \varepsilon^*, \rho^{**}]$. Then Definition 16 guarantees that both $M_{\rho^*+\varepsilon^*}$ and M_ρ have two topologically connected components and that the top-CCRM $\zeta : \mathcal{C}(M_\rho) \to \mathcal{C}(M_{\rho^*+\varepsilon^*})$ is bijective. From Lemma 14 we thus obtain $\tau_{M_\rho}^* \geq \tau_{M_{\rho^*+\varepsilon^*}}^*$, and consequently, we find $\tau \leq \tau_{M_{\rho^*+\varepsilon^*}}^* \leq \tau_{M_\rho}^*$. This implies $\mathcal{C}(M_\rho) = \mathcal{C}_\tau(M_\rho)$ by part *ii)* of Lemma 12, that is, $|\mathcal{C}_\tau(M_\rho)| = 2$. Furthermore, $\tau_{M_\rho}^* \geq \tau_{M_{\rho^*+\varepsilon^*}}^*$ implies $\delta \leq \tau_{M_\rho}^*/3$ and $\tau \leq \tau_{M_\rho}^*/3$, and hence part *iii)* of Lemma 15 shows that the τ-CCRM $\zeta : \mathcal{C}_\tau(M_\rho) \to \mathcal{C}_\tau(M_{\rho,\delta})$ is bijective. This implies $|\mathcal{C}_\tau(M_{\rho,\delta})| = 2$.

iii). If $|\mathcal{C}_\tau(M_{\rho,\delta})| > 1$, then (15) implies $|\mathcal{C}(M_\rho)| > 1$, and hence Definition 16 yields $\rho \geq \rho^*$. Moreover, for ρ^{**}, we have already seen in part *ii)* that the τ-CCRM ζ_M :

$\mathcal{C}_\tau(M_{\rho^{**}}) \to \mathcal{C}_\tau(M_{\rho^{**},\delta})$ is bijective, and the proof of *ii)* further showed $\mathcal{C}(M_{\rho^{**}}) = \mathcal{C}_\tau(M_{\rho^{**}})$. Consequently, we can identify ζ_M with the CCRM $\mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}_\tau(M_{\rho^{**},\delta})$. Moreover, by (15) we obtain $|\mathcal{C}(M_\rho)| = 2$, and hence Definition 16 ensures that the top-CCRM $\zeta^{**} : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_\rho)$ is bijective. In addition, $\tau > \delta$ together with part *ii)* of Lemma 15 and part *i)* of Lemma 12 shows that the CCRM $\zeta_\rho : \mathcal{C}(M_\rho) \to \mathcal{C}_\tau(M_{\rho,\delta})$ is surjective. Now, by Lemma 11 these maps commute in the sense of the following diagram

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

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). Let us fix an $\rho \in [\rho^* + \varepsilon^*, \rho^{**}]$. By part *ii)* and *i)* we then see that $M_{\rho,\delta}$ has two τ-connected components and part *iii)* thus shows that the τ-CCRM $\zeta_M : \mathcal{C}_\tau(M_{\rho^{**},\delta}) \to \mathcal{C}_\tau(M_{\rho,\delta})$ is bijective. Moreover, Lemma 11 yields the following diagram

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

where ζ_V and $\zeta_{V,M}$ are the corresponding τ-CCRMs. 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 we obtain $2 = |\mathcal{C}_\tau(M_{\rho,\delta})| = |\mathcal{C}_\tau(V_{\rho^{**},\delta})| \leq |\mathcal{C}_\tau(V_{\rho,\delta})|$. ■

Lemma 34 *Let (X, d) be a compact metric space, μ be a finite measure on X, and P be a μ-absolutely continuous probability measure on X that has two thick clusters of order $\gamma \in (0, 1]$ between the critical levels ρ^* and ρ^{**}. We write ψ for the corresponding thickness function. Then, for all $\rho \in [0, \rho^{**}]$, $\delta \in (0, \tilde{\delta}_0]$, and $\tau > \psi(\delta)$, the following statements hold:*

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

ii) *We have $|\mathcal{C}_\tau(V_{\rho,\delta})| \leq |\mathcal{C}(M_\rho)|$.*

iii) *If $|\mathcal{C}_\tau(V_{\rho,\delta})| = |\mathcal{C}(M_\rho)|$, then there exists a unique map $\zeta : \mathcal{C}_\tau(V_{\rho,\delta}) \to \mathcal{C}(M_\rho)$ that satisfies*

$$A' \subset \zeta(A') \,, \qquad\qquad A' \in \mathcal{C}_\tau(V_{\rho,\delta}) \,. \tag{16}$$

Moreover, ζ is bijective.

Proof of Lemma 34 *i).* Let us fix a $\tau' \in (0, \tau^*_{M_\rho}]$ such that $\psi(\delta) + \tau' < \tau$, where $\tau^*_{M_\rho}$ is the constant defined in Lemma 12 and $c \geq 1$ is the constant appearing in Definition

19. Moreover, we fix a $B' \in \mathcal{C}(M_\rho)$. By Lemma 12 we then see that $\mathcal{C}(M_\rho) = \mathcal{C}_{\tau'}(M_\rho)$, and hence B' is τ'-connected. Now let $A_1, \dots, A_m$ be the τ-connected components of $V_{\rho,\delta}$. Clearly, Lemma 29 yields $d(A_i, A_j) \geq \tau$ for all $i \neq j$. Assume that the assertion of the lemma is not true, that is, there exist $i_0 \neq j_0$ with $A_{i_0} \cap B' \neq \emptyset$ and $A_{j_0} \cap B' \neq \emptyset$. Then 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 $x_0, \dots, x_{n+1} \in B' \subset M_\rho$ 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 $d(x_i, V_{\rho,\delta}) < \psi(\delta)/2$ for all $i = 0, \dots, n+1$. For all $i = 0, \dots, n+1$, there thus exists an index ℓ_i such that

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

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

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

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

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

iii). As mentioned in part *ii)*, we have an injective map $\zeta : \mathcal{C}_\tau(V_{\rho,\delta}) \to \mathcal{C}(M_\rho)$ that satisfies

$$A' \cap \zeta(A') \neq \emptyset\,, \qquad\qquad A' \in \mathcal{C}_\tau(V_{\rho,\delta})\,. \tag{17}$$

Now, $|\mathcal{C}_\tau(V_{\rho,\delta})| = |\mathcal{C}(M_\rho)|$ implies that ζ is actually bijective. Let us show that ζ is the only map that satisfies (17). To this end, assume the converse, that is, for some $A' \in \mathcal{C}_\tau(V_{\rho,\delta})$, there exists an $B' \in \mathcal{C}(M_\rho)$ with $B' \neq \zeta(A')$ and $A' \cap B' \neq \emptyset$. Since ζ is bijective, there then exists an $A'' \in \mathcal{C}_\tau(V_{\rho,\delta})$ with $\zeta(A'') = B'$, and hence we have $A'' \cap B' \neq \emptyset$. 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 (17). From this it is easy to conclude, that there exists at most one map that satisfies (16). Let us therefore finally show that ζ satisfies (16). To this end, we pick an $A' \in \mathcal{C}_\tau(V_{\rho,\delta})$ and write $B_1, \dots, B_m$ for the topologically connected components of M_ρ. Since $V_{\rho,\delta} \subset M_\rho$, 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 $A' \subset \zeta(A')$. ■

Proof of Theorem 20 *i).* This follows from $|\mathcal{C}_\tau(V_{\rho,\delta})| \leq |\mathcal{C}(M_\rho)| \leq 2$, where the first inequality was established in part *ii)* of Lemma 34.

ii). Our definition of ε^* yields $\delta_{\varepsilon^*} = \tau_{\varepsilon^*} = \tau^*_{M_{\rho^*+\varepsilon^*}}/3 \leq \tau^*_{M_{\rho^{**}}}/3$. By part *iii)* of Lemma 15 we then conclude that the τ-CCRM $\mathcal{C}_\tau(M_{\rho^{**}}) \to \mathcal{C}_\tau(M_{\rho^{**},\delta})$ is bijective. By Lemma 11 it thus suffices to show that the τ-CCRM $\zeta : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}_\tau(M_{\rho^{**}})$ is bijective. Furthermore, if $|\mathcal{C}_\tau(V_{\rho^{**},\delta})| = 1$, the map ζ is automatically injective, and if $|\mathcal{C}_\tau(V_{\rho^{**},\delta})| = 2$, the injectivity follows from the surjectivity. 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 $d(x, V_{\rho^{**},\delta}) < c\delta^\gamma$, and hence there exists an $A' \in \mathcal{C}_\tau(V_{\rho^{**},\delta})$ with $d(x, A') < c\delta^\gamma$. Therefore, $c\delta^\gamma < \psi(\delta) < \tau$ 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 31 and the definition of ζ, we conclude that $\zeta(A') = B'$.

iii). By part *ii)* of Lemma 34, we conclude that $|\mathcal{C}_\tau(V_{\rho,\delta})| = |\mathcal{C}(M_\rho)| = 2$, and hence the definition of topological clustering ensures $\rho \geq \rho^*$. Furthermore, part *iii)* of Lemma 34 yields a unique map $\zeta_\rho : \mathcal{C}_\tau(V_{\rho,\delta}) \to \mathcal{C}(M_\rho)$ satisfying (16). Moreover, part *ii)* of Theorem 18 shows $|\mathcal{C}_\tau(M_{\rho^{**},\delta})| = 2$ and the already established bijectivity of ζ^{**} then gives $|\mathcal{C}_\tau(V_{\rho^{**},\delta})| = |\mathcal{C}_\tau(M_{\rho^{**},\delta})| = 2$. Consequently, part *iii)* of Lemma 34 yields a unique map $\zeta_{\rho^{**}} : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}(M_{\rho^{**}})$ satisfying (16). Finally, let $\zeta_{\text{top}} : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_\rho)$ be the top-CCRM, which is bijective according to the definition of topological clustering. Then the τ-CCRM $\zeta : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}_\tau(V_{\rho,\delta})$ enjoys the following diagram

$$\begin{array}{ccc} \mathcal{C}_\tau(V_{\rho^{**},\delta}) & \xrightarrow{\;\zeta_{\rho^{**}}\;} & \mathcal{C}(M_{\rho^{**}}) \\ \zeta \big\downarrow & & \big\downarrow \zeta_{\text{top}} \\ \mathcal{C}_\tau(V_{\rho,\delta}) & \xrightarrow[\;\zeta_\rho\;]{} & \mathcal{C}(M_\rho) \end{array}$$

whose commutativity can be checked analogously to the proof of Lemma 11. Then the bijectivity of $\zeta_{\rho^{**}}$, ζ_{top}, and ζ_ρ yields the bijectivity of ζ, which completes the proof. ■

Appendix D. Proofs related to basic properties of histograms

Proof of Theorem 22 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 $\|f\|_\infty \leq \kappa_X \delta^{-d_X}$. Consequently, Hoeffding's inequality, see e.g. (Devroye et al., 1996, Theorem 8.1), yields

$$P^n\biggl(\Bigl\{D \in X^n : \Bigl|\frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbb{E}_P f\Bigr| < \varepsilon\Bigr\}\biggr) \geq 1 - 2e^{-2\kappa_X^{-2}\delta^{2d_X}\varepsilon^2 n}$$

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 $|\mathcal{A}_\delta| \leq \kappa_X \delta^{-d_X}$, we thus obtain

$$P^n\biggl(\Bigl\{D \in X^n : \sup_{A \in \mathcal{A}_\delta}\Bigl|\frac{D(A)}{\mu(A)} - \frac{P(A)}{\mu(A)}\Bigr| < \varepsilon\Bigr\}\biggr) \geq 1 - 2\kappa_X\delta^{-d_X}e^{-2\kappa_X^{-2}\delta^{2d_X}\varepsilon^2 n}\,.$$

For $A \in \mathcal{A}_\delta$ and $x \in A$, we have $\bar{h}_{D,\mathcal{A}_\delta}(x) = \mu(A)^{-1}D(A)$ and $\bar{h}_{P,\mathcal{A}_\delta}(x) = \mu(A)^{-1}P(A)$, and hence 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 $\|f\|_\infty \leq \kappa_X \delta^{-d_X}$ and

$$\mathbb{E}f^2 \leq \mu(A)^{-2}P(A) \leq \mu(A)^{-1}\|h\|_\infty \leq \kappa_X\delta^{-d_X}\|h\|_\infty\,.$$

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

$$P^n\Big(\Big\{D \in X^n : \Big|\frac{1}{n}\sum_{i=1}^n f(x_i)\Big| < \varepsilon\Big\}\Big) \geq 1 - 2e^{-\frac{3\varepsilon^2 \delta^{d_X} n}{\kappa_X (6\|h\|_\infty + 2\varepsilon)}}\,.$$

The rest of the proof follows the lines of the proof of the first inequality. ■

Proof of Lemma 23 *i).* We will show the equivalent inclusion $\{\hat f_\rho = -1\} \subset T_\delta(V_{\rho+\varepsilon})$. To this end, we fix an $x \in X$ with $\hat f_\rho(x) = -1$. If $x \in V_{\rho+\varepsilon}$, we immediately obtain $x \in T_\delta(V_{\rho+\varepsilon})$, and hence we may restrict our considerations to the case $x \in M_{\rho+\varepsilon}$. Then, $\hat f_\rho(x) = -1$ implies $\hat h(x) < \rho$ and from $\|\hat h - \bar h_{P,\mathcal{A}_\delta}\|_\infty \leq \varepsilon$, we thus conclude $\bar h_{P,\mathcal{A}_\delta}(x) \leq \hat h(x) + \varepsilon < \rho + \varepsilon$. Now let A_i be the unique cell of the partition $\mathcal{A}_\delta$ satisfying $x \in A_i$. The definition of $\bar h_{P,\mathcal{A}_\delta}$ together with the assumed $0 < \mu(A_i) < \infty$ then yields

$$\int_{A_i} h\, d\mu = P(A_i) < (\rho+\varepsilon)\mu(A_i)\,, \tag{18}$$

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

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

However, this inequality contradicts (18), and hence there does exist an $x' \in V_{\rho+\varepsilon} \cap A_i$. This implies

$$d(x, V_{\rho+\varepsilon}) \leq d(x,x') \leq \operatorname{diam} A_i \leq \delta\,,$$

i.e. we have shown $x \in T_\delta(V_{\rho+\varepsilon})$.

ii). Let us fix an $x \in X$ with $\hat f_\rho(x) = 1$. If $x \in M_{\rho-\varepsilon}$, we immediately obtain $x \in T_\delta(M_{\rho-\varepsilon})$, and hence it remains to consider the case $x \in V_{\rho-\varepsilon}$. Clearly, if $\rho - \varepsilon \leq 0$, this case is impossible, and hence we may additionally assume $\rho - \varepsilon > 0$. Then, $\hat f_\rho(x) = 1$ implies $\hat h(x) \geq \rho$ and from $\|\hat h - \bar h_{P,\mathcal{A}_\delta}\|_\infty \leq \varepsilon$, we thus conclude $\bar h_{P,\mathcal{A}_\delta}(x) \geq \hat h(x) - \varepsilon \geq \rho - \varepsilon$. Now let A_i be the unique cell of the partition $\mathcal{A}_\delta$ satisfying $x \in A_i$. By the definition of $\bar h_{P,\mathcal{A}_\delta}$ and $\mu(A_i) > 0$ we then obtain

$$\int_{A_i} h\, d\mu = P(A_i) \geq (\rho-\varepsilon)\mu(A_i)\,, \tag{19}$$

where, again, $h : X \to [0,\infty)$ is an arbitrary μ-density of P. Our next goal is to show that there exists an $x' \in M_{\rho-\varepsilon} \cap A_i$. Suppose the converse holds, that is $A_i \subset V_{\rho-\varepsilon}$. Then the assumed regularity of P at the level $\rho - \varepsilon$ yields (6), and hence we conclude that $\mu(A_i \setminus \{h < \rho-\varepsilon\}) \leq \mu(V_{\rho-\varepsilon} \setminus \{h < \rho-\varepsilon\}) = 0$. This implies

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

However, this inequality contradicts (19), and hence there does exist an $x' \in M_{\rho-\varepsilon} \cap A_i$. This yields

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

i.e. we have shown $x \in T_\delta(M_{\rho-\varepsilon})$. ■

Lemma 35 *Let (X,d) be a compact metric space and μ be a finite measure on X such that (X,d,μ) admits uniform d_X-dimensional partitions. Moreover, let P be a μ-absolutely continuous probability measure on X and $\hat{h}: X \to \mathbb{R}$ be a function with $\|\hat{h} - \bar{h}_{P,\mathcal{A}_\delta}\|_\infty \leq \varepsilon$ for some $\varepsilon > 0$. Furthermore, for fixed $\delta > 0$, $\epsilon \geq 0$, $\tau > 0$, and $\rho \geq 0$, let $\zeta : \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}) \to \mathcal{C}_\tau(\{\hat{f}_\rho = 1\})$ be the τ-CCRM. Then the following statements hold:*

i) If, for some $V' \in \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})$, we have $V' \cap V_{\rho+3\varepsilon+\epsilon,\delta} \neq \emptyset$, then we obtain

$$\zeta(V') \cap \{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\} \neq \emptyset\,.$$

ii) For all $B' \in \mathcal{C}_\tau(\{\hat{f}_\rho = 1\})$ with $B' \notin \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$, we have

$$B' \subset T_\delta(X \setminus M_{\rho+\varepsilon}) \cap T_\delta(M_{\rho-\varepsilon})\,.$$

Moreover, every $A \subset B' \cap \{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\}$ satisfies $A \subset T_\delta(X \setminus M_{\rho+\varepsilon}) \cap T_\delta(M_{\rho+\varepsilon+\epsilon})$.

Proof of Lemma 35 *i).* This assertion follows from the τ-CCRM property $V' \subset \zeta(V')$ and the inclusion $V_{\rho+3\varepsilon+\epsilon,\delta} \subset \{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\}$ established in Lemma 23.

ii). For $x \in B'$ we have $x \notin \bigcup_{V' \in \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})} \zeta(V')$, and hence the τ-CCRM property yields

$$x \notin \bigcup_{V' \in \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})} V' = V_{\rho+\varepsilon,\delta}\,.$$

This shows $x \in T_\delta(X \setminus M_{\rho+\varepsilon})$, i.e. we have proved $B' \subset T_\delta(X \setminus M_{\rho+\varepsilon})$. The inclusion $B' \subset T_\delta(M_{\rho-\varepsilon})$ directly follows from $B' \subset \{\hat{f}_\rho = 1\}$ and Lemma 23. The last assertion follows from the inclusion $\{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\} \subset T_\delta(M_{\rho+\varepsilon+\epsilon})$ established in Lemma 23 and the previously shown inclusion. ■

Proof of Theorem 24 Our first goal is to establish the following *disjoint* union:

$$\begin{aligned}\mathcal{C}_\tau(\{\hat{f}_\rho = 1\}) = \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))& \\ \cup \big\{B' \in \mathcal{C}_\tau(\{\hat{f}_\rho = 1\}) \setminus \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})) &: B' \cap \{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\} \neq \emptyset\big\} \\ \cup \big\{B' \in \mathcal{C}_\tau(\{\hat{f}_\rho = 1\}) &: B' \cap \{\hat{f}_{\rho+2\varepsilon+\epsilon} = 1\} = \emptyset\big\}\,. \end{aligned} \tag{20}$$

We begin by showing the auxiliary result

$$V' \cap V_{\rho+3\varepsilon+\epsilon,\delta} \neq \emptyset\,, \qquad V' \in \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}). \tag{21}$$

To this end, we observe that parts *i)* and *ii)* of Theorem 18 yield $|\mathcal{C}_\tau(M_{\rho^{**},\delta})| = 2$, and hence part *ii)* of Theorem 20 implies $|\mathcal{C}_\tau(V_{\rho^{**},\delta})| = 2$, and thus $V_{\rho^{**},\delta} \neq \emptyset$. Let W' and W'' be the two τ-connected components of $V_{\rho^{**},\delta}$. Let us first assume that $V_{\rho+\varepsilon,\delta}$ has exactly one τ-connected component V', i.e. $V' = V_{\rho+\varepsilon,\delta}$. Then $\rho+3\varepsilon+\epsilon \leq \rho^{**}$ and $\rho+\varepsilon \leq \rho+3\varepsilon+\epsilon$ imply

$$\emptyset \neq V_{\rho^{**},\delta} \subset V_{\rho+3\varepsilon+\epsilon,\delta} = V_{\rho+\varepsilon,\delta} \cap V_{\rho+3\varepsilon+\epsilon,\delta} = V' \cap V_{\rho+3\varepsilon+\epsilon,\delta}\,,$$

i.e. we have shown (21). Let us now assume that $V_{\rho+\varepsilon,\delta}$ has more than one τ-component. Then it has exactly two such components V' and V'' by $\rho+\varepsilon < \rho^{**}$ and part *i)* of Theorem 20. By part *iii)* of Theorem 20 we may then assume without loss of generality that we have $W' \subset V'$ and $W'' \subset V''$. Since $\rho + 3\varepsilon + \epsilon \leq \rho^{**}$ implies $V_{\rho^{**},\delta} \subset V_{\rho+3\varepsilon+\epsilon,\delta}$, these inclusions yield $\emptyset \neq W' = W' \cap V_{\rho^{**},\delta} \subset V' \cap V_{\rho+3\varepsilon+\epsilon,\delta}$ and $\emptyset \neq W'' = W'' \cap V_{\rho^{**},\delta} \subset V'' \cap V_{\rho+3\varepsilon+\epsilon,\delta}$. Consequently, we have proved (21) in this case, too.

Now, from (21) we conclude by part *i)* of Lemma 35 that $B' \cap \{\hat f_{\rho+2\varepsilon+\epsilon} = 1\} \neq \emptyset$ for all $B' \in \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$. This yields

$$\begin{aligned}&\bigl\{B' \in \mathcal{C}_\tau(\{\hat f_\rho = 1\}) \setminus \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})) : B' \cap \{\hat f_{\rho+2\varepsilon+\epsilon} = 1\} = \emptyset\bigr\}\\ =&\bigl\{B' \in \mathcal{C}_\tau(\{\hat f_\rho = 1\}) : B' \cap \{\hat f_{\rho+2\varepsilon+\epsilon} = 1\} = \emptyset\bigr\}\,,\end{aligned}$$

and the latter immediately implies (20).

Now, using (20) and $\{\hat f_{\rho+2\varepsilon} = 1\} \supset \{\hat f_{\rho+2\varepsilon+\epsilon} = 1\}$ it remains to show

$$B' \cap \{\hat f_{\rho+2\varepsilon} = 1\} = \emptyset\,,$$

for all $B' \in \mathcal{C}_\tau(\{\hat f_\rho = 1\}) \setminus \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$. Let us assume the converse, that is, there exists some $B' \in \mathcal{C}_\tau(\{\hat f_\rho = 1\})$ with $B' \notin \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$ and $B' \cap \{\hat f_{\rho+2\varepsilon} = 1\} \neq \emptyset$. Since $\{\hat f_{\rho+2\varepsilon} = 1\} \subset T_\delta(M_{\rho+\varepsilon})$ by Lemma 23, there then exists an $x \in B' \cap T_\delta(M_{\rho+\varepsilon})$. The latter yields an $x' \in M_{\rho+\varepsilon}$ with $d(x,x') \leq \delta$, and since P has thick clusters we obtain

$$d(x', V_{\rho+\varepsilon,\delta}) < c\,\delta^\gamma\,.$$

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

$$d(B', B'') \leq d(x, x'') \leq d(x,x') + d(x', x'') < \delta + c\delta^\gamma \leq \tau\,.$$

However, since $B' \notin \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$ and $B'' \in \zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$ we obtain $B' \neq B''$, and therefore, Lemma 29 yields $d(B', B'') \geq \tau$, i.e. we have found a contradiction. ■

Proof of Theorem 25 *i).* Let $D \in X^n$ be a dataset such that $\|\bar h_{D,\mathcal{A}_\delta} - \bar h_{P,\mathcal{A}_\delta}\|_\infty < \varepsilon$. Moreover, let $\epsilon := 0$ and $\rho \geq 0$ be the current level that is considered by Algorithm 3.1. Then, Theorem 24 shows that, for $\rho \in [0, \rho^{**} - 3\varepsilon]$, Algorithm 3.1 identifies exactly the τ-connected components of $\{\hat f_\rho = 1\}$ in its loop that belong to the set $\zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$, where $\zeta : \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}) \to \mathcal{C}_\tau(\{\hat f_\rho = 1\})$ is the τ-CCRM. In the following, we thus consider the set $\zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$ for $\rho \in [0, \rho^{**} - 3\varepsilon]$.

Let us first consider the case $\rho \in [0, \rho^* - \varepsilon)$. Then, part *i)* and *iii)* of Theorem 20 together with the assumed $\rho + \varepsilon < \rho^*$ show $|\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})| = 1$. This yields $|\zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))| = 1$, and hence Algorithm 3.1 does not stop. Consequently, we have $\rho^*(D) \geq \rho^* - \varepsilon$.

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 $\rho = i\varepsilon$, $i = 0, 1, \ldots,$ and the width of the interval above is ε. Moreover, our assumptions on ε^* and ε guarantee $\rho^* + \varepsilon^* + 2\varepsilon \leq \rho^{**} - 3\varepsilon$ and hence we have $\rho \in [\rho^* + \varepsilon^* + \varepsilon, \rho^{**} - 3\varepsilon]$. Let us write $\zeta_V : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta})$, $\zeta_M : \mathcal{C}_\tau(M_{\rho^{**},\delta}) \to \mathcal{C}_\tau(M_{\rho-\varepsilon,\delta})$, and $\zeta_{V,M} : \mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}) \to \mathcal{C}_\tau(M_{\rho-\varepsilon,\delta})$ for the τ-CCRMs between the involved sets. Using Lemma 11 twice, we then obtain the following diagram:

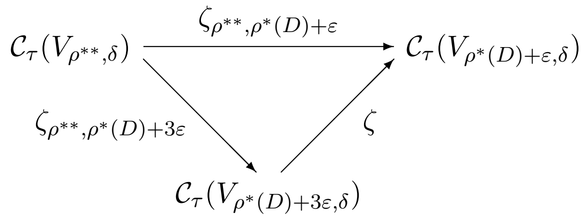
$$
\begin{array}{ccc}
\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}) & \xrightarrow{\ \zeta_{V,M}\ } & \mathcal{C}_\tau(M_{\rho-\varepsilon,\delta}) \\
\uparrow \zeta_V & & \uparrow \zeta_M \\
\mathcal{C}_\tau(V_{\rho^{**},\delta}) & \xrightarrow[\ \zeta^{**}\ ]{} & \mathcal{C}_\tau(M_{\rho^{**},\delta})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

Moreover, we have $\rho - \varepsilon \geq \rho^* + \varepsilon^*$ and $\rho + \varepsilon \geq \rho^* + \varepsilon^*$, and hence part *ii)* and *iv)* of Theorem 18 together with part *i)* and *ii)* of Theorem 20 show that the sets $M_{\rho-\varepsilon,\delta}$ and $V_{\rho+\varepsilon,\delta}$ both have two τ-connected components. Consequently, part *iii)* of Theorem 18 and part *iii)* of Theorem 20 ensure that the maps ζ_V and ζ_M are bijective, and, in addition, part *ii)* of Theorem 20 shows that ζ^{**} is bijective. Consequently, $\zeta_{V,M}$ is bijective. Let us further consider the τ-CCRM $\zeta' : \mathcal{C}_\tau(\{\hat{f}_\rho = 1\}) \to \mathcal{C}_\tau(M_{\rho-\varepsilon,\delta})$. Then Lemma 11 yields another diagram:

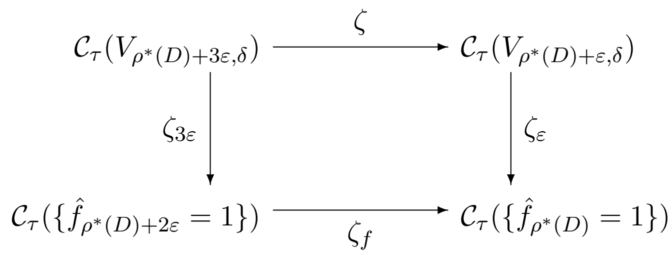
$$
\begin{array}{ccc}
\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}) & \xrightarrow{\ \zeta_{V,M}\ } & \mathcal{C}_\tau(M_{\rho-\varepsilon,\delta}) \\
\zeta \searrow & & \nearrow \zeta' \\
 & \mathcal{C}_\tau(\{\hat{f}_\rho = 1\}) &
\end{array}
$$

Since $\zeta_{V,M}$ is bijective, we then find that ζ is injective, and since we have already seen that $V_{\rho+\varepsilon,\delta}$ has two τ-connected components, we conclude that $\zeta(\mathcal{C}_\tau(V_{\rho+\varepsilon,\delta}))$ contains two elements. Consequently, the stopping criterion of Algorithm 3.1 is satisfied, that is, $\rho^*(D) \leq \rho^* + \varepsilon^* + 2\varepsilon$.

ii). Theorem 24 shows that in its last run through the loop Algorithm 3.1 identifies exactly the τ-connected components of $\{\hat{f}_{\rho^*(D)} = 1\}$ that belong to the set $\zeta_\varepsilon(\mathcal{C}_\tau(V_{\rho^*(D)+\varepsilon,\delta}))$, where $\zeta_\varepsilon : \mathcal{C}_\tau(V_{\rho^*(D)+\varepsilon,\delta}) \to \mathcal{C}_\tau(\{\hat{f}_{\rho^*(D)} = 1\})$ is the τ-CCRM. Moreover, since Algorithm 3.1 stops at $\rho^*(D)$, we have $|\zeta_\varepsilon(\mathcal{C}_\tau(V_{\rho^*(D)+\varepsilon,\delta}))| \neq 1$ and thus $|\mathcal{C}_\tau(V_{\rho^*(D)+\varepsilon,\delta})| \neq 1$. From $\rho^*(D) + \varepsilon \leq \rho^{**}$ and part *i)* of Theorem 20 we thus conclude that $|\mathcal{C}_\tau(V_{\rho^*(D)+\varepsilon,\delta})| = 2$. For later purposes, note that the latter implies the injectivity of ζ_ε. Therefore, *iii)* of Theorem 20 shows that the τ-CCRM $\zeta_{\rho^{**},\rho^*(D)+\varepsilon} : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}_\tau(V_{\rho^*(D)+\varepsilon,\delta})$ is bijective. Let us now consider the following commutative diagram:



where the remaining two maps are the corresponding τ-CCRMs. Now the bijectivity of $\zeta_{\rho^{**},\rho^*(D)+\varepsilon}$ shows that $\zeta_{\rho^{**},\rho^*(D)+3\varepsilon}$ is injective, and since $\rho^*(D)+3\varepsilon \leq \rho^{**}$ implies $|\mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta})| \leq 2 = |\mathcal{C}_\tau(V_{\rho^{**},\delta})|$ by part *i)* of Theorem 20, $\zeta_{\rho^{**},\rho^*(D)+3\varepsilon}$ is actually bijective. This yields $|\mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta})| = 2$ and the bijectivity of ζ. Let us consider yet another commutative diagram



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

iii). Follows from Theorem 24 and $\rho^*(D)+2\varepsilon \leq \rho^{**}-3\varepsilon$.

iv). By part *iii)* of Lemma 34 there exist bijective CCRMs $\zeta_{\rho^{**}} : \mathcal{C}_\tau(V_{\rho^{**},\delta}) \to \mathcal{C}(M_{\rho^{**}})$ and $\zeta_{\rho^*(D)+3\varepsilon} : \mathcal{C}_\tau(V_{\rho^*(D)+3\varepsilon,\delta}) \to \mathcal{C}(M_{\rho^*(D)+3\varepsilon})$. Moreover, in the proof of *ii)* we have already seen that τ-CCRM $\zeta_{\rho^{**},\rho^*(D)+3\varepsilon}$ is bijective. This gives the diagram. ■

Appendix E. Proofs related to large sample sizes

Lemma 36 *Let (X,d) be a complete separable metric space, μ be a finite measure on X, and $(A_\rho)_{\rho\in\mathbb{R}}$ be a family of closed subsets of X with $A_\rho \subset A_{\rho'}$ for all $\rho' \leq \rho$. For $\rho^* \in \mathbb{R}$, we write*

$$\bar{A}_{\rho^*} := \bigcup_{\rho>\rho^*} A_\rho \qquad\qquad \text{and} \qquad\qquad \dot{A}_{\rho^*} := \bigcup_{\rho>\rho^*} \mathring{A}_\rho \,.$$

Then we have

$$\dot{A}_{\rho^*} = \bigcup_{\rho>\rho^*} \bigcup_{\varepsilon>0} \bigcup_{\delta>0} \bigl(X \setminus T_\delta(X \setminus A_{\rho+\varepsilon})\bigr)$$
$$\bar{A}_{\rho^*} = \bigcup_{\rho>\rho^*} \bigcap_{\varepsilon>0} \bigcap_{\delta>0} T_\delta(A_{\rho-\varepsilon})\,.$$

Moreover, the following statements are equivalent:

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

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

Proof of Lemma 36 To show the first equality, we observe that (12) implies

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

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

$$\bigcap_{\rho>\rho^*} \overline{X \setminus A_\rho} = \bigcap_{\rho>\rho^*} (X \setminus \mathring{A}_\rho) = X \setminus \bigcup_{\rho>\rho^*} \mathring{A}_\rho .$$

Combining both equalities and then taking the complement, we find the first assertion. Analogously, (12) shows

$$\bigcup_{\rho>\rho^*} \bigcap_{\varepsilon>0} \bigcap_{\delta>0} T_\delta(A_{\rho-\varepsilon}) = \bigcup_{\rho>\rho^*} \bigcap_{\varepsilon>0} \overline{A_{\rho-\varepsilon}} = \bigcup_{\rho>\rho^*} \bigcap_{\varepsilon>0} A_{\rho-\varepsilon} .$$

Moreover, using the monotonicity of the family (A_ρ), it is straightforward to show that $\bigcup_{\rho>\rho^*} \bigcap_{\varepsilon>0} A_{\rho-\varepsilon} = \bigcup_{\rho>\rho^*} A_\rho$, which finishes the proof.

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

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

Proof of Theorem 26 Let us write $A_{\rho^{**},i}$, $i = 1, 2$, for the two topologically connected components of $M_{\rho^{**}}$. Moreover, for $\rho \in (\rho^*, \rho^{**}]$, we define $A_{\rho,i} := \zeta_\rho(A_{\rho^{**},i})$, where $\zeta_\rho : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_\rho)$ is the top-CCRM. 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

$$\mu(\bar{A}_{\rho^*,i} \setminus \dot{A}_{\rho^*,i}) = 0 \tag{22}$$

for $i = 1, 2$, where we used the notation of Lemma 36. To this end, we fix an $\epsilon > 0$. Since the definition of clusters ensures that P is normal at level ρ^*, we have $\mu(\bar{M}_{\rho^*} \setminus \dot{M}_{\rho^*}) = 0$, Lemma 36 then shows that there exists a $\rho_\epsilon > \rho^*$ such that $\mu(M_\rho \setminus \mathring{M}_\rho) \leq \epsilon$ for all $\rho \in (\rho^*, \rho_\epsilon]$, where we may assume without loss of generality that $\rho_\epsilon \leq \rho^{**}$. Let us fix a $\rho \in (\rho^*, \rho_\epsilon]$. Then the fact that $M_\rho = A_{\rho,1} \cup A_{\rho,2}$ is a disjoint union of closed sets yields $\mathring{M}_\rho = \mathring{A}_{\rho,1} \cup \mathring{A}_{\rho,2}$. Consequently, we obtain

$$M_\rho \setminus \mathring{M}_\rho = \big(A_{\rho,1} \setminus (\mathring{A}_{\rho,1} \cup \mathring{A}_{\rho,2})\big) \cup \big(A_{\rho,2} \setminus (\mathring{A}_{\rho,1} \cup \mathring{A}_{\rho,2})\big) = (A_{\rho,1} \setminus \mathring{A}_{\rho,1}) \cup (A_{\rho,2} \setminus \mathring{A}_{\rho,2}) .$$

This implies $\mu(A_{\rho,i} \setminus \mathring{A}_{\rho,i}) \leq \epsilon$, and hence Lemma 36 shows (22).

Let us now fix an $\epsilon > 0$. We define $V_{\rho,\delta,i} := X \setminus T_\delta(X \setminus A_{\rho,i})$ for all $\delta > 0$, $\rho \in \mathbb{R}$, and $i = 1, 2$. By the first equality of Lemma 36, Equation (22), and the σ-continuity of finite measures there then exist $\delta_\epsilon > 0$, $\varepsilon_\epsilon > 0$, and $\rho_\epsilon > \rho^*$ 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

$$\mu(\bar{A}_{\rho^*,i} \setminus V_{\rho+\varepsilon,\delta,i}) = \mu(\dot{A}_{\rho^*,i} \setminus V_{\rho+\varepsilon,\delta,i}) \leq \epsilon\,. \tag{23}$$

Moreover, the second equality of Lemma 36 shows that, for all $\rho > \rho^*$, we have

$$\bigcap_{\varepsilon>0} \bigcap_{\delta>0} M_{\rho-\varepsilon,\delta} \subset \bar{M}_{\rho^*}\,.$$

Clearly, this implies $\bigcap_{\varepsilon>0} \bigcap_{\delta>0} M_{\rho-\varepsilon,\delta} \setminus \bar{M}_{\rho^*} = \emptyset$. Consequently, we have

$$\mu(M_{\rho-\varepsilon,\delta} \setminus \bar{M}_{\rho^*}) \leq \epsilon \tag{24}$$

for all $\rho > \rho^*$ and all sufficiently small $\varepsilon > 0$, $\delta > 0$. Without loss of generality, we may thus assume that (24) holds for all $\varepsilon \in (0, \varepsilon_\epsilon]$, $\delta \in (0, \delta_\epsilon]$ and all $\rho > \rho^*$. We now define $\varepsilon^* := \min\{\frac{\rho_\epsilon - \rho^*}{5}, \frac{\rho^{**} - \rho^*}{8}\}$, $\varepsilon^\star := \min\{\varepsilon^*, \varepsilon_\epsilon\}$, $\delta^\star := \min\{\delta_\epsilon, \delta_{\varepsilon^*}, \tilde{\delta}_0\}$, and $\tau^\star := \tau_{\varepsilon^*}$. Then, for all sufficiently large n, we have $\varepsilon_n \in (0, \varepsilon^\star]$, $\delta_n \in (0, \delta^\star]$, $\tau_n \in (0, \tau^\star]$, and by Theorem 22 we further know that the probability P^n of $\|\bar{h}_{D,\mathcal{A}_{\delta_n}} - \bar{h}_{P,\mathcal{A}_{\delta_n}}\|_\infty < \varepsilon_n$ converges to 1 for $n \to \infty$. Let us therefore only consider such data sets D and parameters satisfying $\varepsilon_n \in (0, \varepsilon^\star]$, $\delta_n \in (0, \delta^\star]$, $\tau_n \in (0, \tau^\star]$. Then our construction ensures that we can apply Theorem 25. In particular, we have $\rho^* < \rho^*(D) + 2\varepsilon_n \leq \rho^* + \varepsilon^* + 4\varepsilon_n \leq \rho^* + 5\varepsilon^* \leq \rho_\epsilon$, and hence (23) and (24) hold for $\rho := \rho^*(D) + 2\varepsilon_n$. Following the discussion in front of Theorem 26, we further have two τ_n-connected components V_1' and V_2' of $V_{\rho+\varepsilon_n,\delta_n}$ and two τ_n-connected components V_1'' and V_2'' of V_{ρ^{**},δ_n} such that $V_i'' \subset V_i'$, $V_i'' \subset A_{\rho^{**},i}$, and $V_i' \subset B_i(D)$ for $i = 1, 2$. Let us next show that, for $i = 1, 2$, we have

$$V_{\rho+\varepsilon_n,\delta_n,i} \subset V_i'\,. \tag{25}$$

To this end, we fix an $x \in V_{\rho+\varepsilon_n,\delta_n,1} = X \setminus T_{\delta_n}(X \setminus A_{\rho+\varepsilon_n,1})$. Since $V_{\rho+\varepsilon_n,\delta_n,1} \subset A_{\rho+\varepsilon_n,1}$ and $V_{\rho+\varepsilon_n,\delta_n,1} \subset V_{\rho+\varepsilon_n,\delta_n}$, we then have $x \in A_{\rho+\varepsilon_n,1}$ and $x \in V_1' \cup V_2'$. Let us assume that $x \in V_2'$. Then we have $V_2' \cap A_{\rho+\varepsilon_n,1} \neq \emptyset$. Now, the diagram of Theorem 25 shows that $\zeta_{\rho+\varepsilon_n} : \mathcal{C}_{\tau_n}(V_{\rho+\varepsilon_n,\delta_n}) \to \mathcal{C}(M_{\rho+\varepsilon_n})$ satisfies $\zeta_{\rho+\varepsilon_n}(V_i') = A_{\rho+\varepsilon_n,i}$, and hence we have $V_2' \subset A_{\rho+\varepsilon_n,2}$. Consequently, $V_2' \cap A_{\rho+\varepsilon_n,1} \neq \emptyset$ implies $A_{\rho+\varepsilon_n,2} \cap A_{\rho+\varepsilon_n,1} \neq \emptyset$, which is a contradiction. Therefore, we have $x \in V_1'$, that is, we have shown (25) for $i = 1$. The case $i = 2$ can be shown analogously.

Using $A_i^* = \bar{A}_{\rho^*,i}$, $V_i' \subset B_i(D)$, (25), and (23) we now obtain

$$\mu\big(A_i^* \setminus B_i(D)\big) = \mu\big(\bar{A}_{\rho^*,i} \setminus B_i(D)\big) \leq \mu(\bar{A}_{\rho^*,i} \setminus V_i') \leq \mu(\bar{A}_{\rho^*,i} \setminus V_{\rho+\varepsilon_n,\delta_n,i}) \leq \epsilon\,. \tag{26}$$

Conversely, using $\mu(B \setminus A) = \mu(B) - \mu(A \cap B)$ twice, we obtain

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

Since $B_1(D) \cap B_2(D) = \emptyset$ implies $B_1(D) \cap A_2^* \subset A_2^* \setminus B_2(D)$ and Lemma 23 shows $B_1(D) \subset M_{\rho+\varepsilon_n,\delta_n}$, we can thus conclude with the help of the previous estimate that

$$\begin{aligned}\mu(B_1(D) \setminus A_1^*) &\leq \mu\big(B_1(D) \setminus (A_1^* \cup A_2^*)\big) + \mu(A_2^* \setminus B_2(D)) \\ &\leq \mu\big(M_{\rho+\varepsilon_n,\delta_n} \setminus (A_1^* \cup A_2^*)\big) + \mu(A_2^* \setminus B_2(D)) \\ &\leq 2\epsilon\,,\end{aligned}$$

where in the last step we used (24) and (26). Clearly, we can establish $\mu(B_2(D) \setminus A_2^*) \leq 2\epsilon$ analogously, and hence we we finally obtain $\mu(B_i(D) \vartriangle A_i^*) \leq 3\epsilon$ for $i = 1, 2$. ■