Journal of Machine Learning Research 16 (2015) 1949-2002 Submitted 6/15; Published 9/15

Towards an Axiomatic Approach to Hierarchical Clustering of Measures

Philipp Thomann PHILIPP.THOMANN@MATHEMATIK.UNI-STUTTGART.DE
Ingo Steinwart INGO.STEINWART@MATHEMATIK.UNI-STUTTGART.DE
Nico Schmid NICO.SCHMID@MATHEMATIK.UNI-STUTTGART.DE
Institute for Stochastics and Applications
University of Stuttgart, Germany

Editor: Alex Gammerman and Vladimir Vovk

Abstract

We propose some axioms for hierarchical clustering of probability measures and investigate their ramifications. The basic idea is to let the user stipulate the clusters for some elementary measures. This is done without the need of any notion of metric, similarity or dissimilarity. Our main results then show that for each suitable choice of user-defined clustering on elementary measures we obtain a unique notion of clustering on a large set of distributions satisfying a set of additivity and continuity axioms. We illustrate the developed theory by numerous examples including some with and some without a density.

Keywords: axiomatic clustering, hierarchical clustering, infinite samples clustering, density level set clustering, mixed Hausdorff-dimensions

1. Introduction

Clustering is one of the most basic tools to investigate unsupervised data: finding groups in data. Its applications reach from categorization of news articles over medical imaging to crime analysis. For this reason, a wealth of algorithms have been proposed, among the best-known being: k-means (MacQueen, 1967), linkage (Ward, 1963; Sibson, 1973; Defays, 1977), cluster tree (Stuetzle, 2003), DBSCAN (Ester et al., 1996), spectral clustering (Donath and Hoffman, 1973; von Luxburg, 2007), and expectation-maximization for generative models (Dempster et al., 1977). For more information and research on clustering we refer the reader to Jardine and Sibson (1971); Hartigan (1975); Kaufman and Rousseeuw (1990); Mirkin (2005); Gan et al. (2007); Kogan (2007); Ben-David (2015); Menardi (2015) and the references therein.

However, each *ansatz* has its own implicit or explicit definition of what clustering is. Indeed for k-means it is a particular Voronoi partition, for Hartigan (1975, Section 11.13) it is the collection of connected components of a density level set, and for generative models it is the decomposition of mixed measures into the parts. Stuetzle (2003) stipulates a grouping around the modes of a density, while Chacón (2014) uses gradient-flows. Thus, there is no universally accepted definition.

A good notion of clustering certainly needs to address the inherent random variability in data. This can be achieved by notions of clusterings for infinite sample regimes or complete knowledge scenarios—as von Luxburg and Ben-David (2005) put it. Such an approach has various advantages: one can talk about ground-truth, can compare alternative clustering algorithms (empirically, theoretically, or in a combination of both by using artificial data), and can define and establish consistency and learning rates. Defining clusters as the connected components of density level sets satisfies all of these requirements. Yet it seems to be slightly *ad-hoc* and it will always be debatable, whether thin bridges should connect components, and whether close components should really be separated. Similar concerns may be raised for other infinite sample notions of clusterings such as Stuetzle (2003) and Chacón (2014).

In this work we address these and other issues by asking ourselves: *What does the set of clustering functions look like? What can defining properties—or axioms—of clustering functions be and what are their ramifications? Given such defining properties, are there functions fulfilling these? How many are there? Can a fruitful theory be developed? And finally, for which distributions do we obtain a clustering and for which not?*

These questions have led us to an axiomatic approach. The basic idea is to let the user stipulate the clusters for some elementary measures. Here, his choice does not need to rely on a metric or another pointwise notion of similarity though—only basic shapes for geometry and a separation relation have to be specified. Our main results then show that for each suitable choice we obtain a unique notion of clustering satisfying a set of additivity and continuity axioms on a large set of measures. These will be motivated in Section 1.2 and are defined in Axioms 1, 2, and 3. The major technical achievement of this work is Theorem 20: it establishes criteria (c.f. Definition 18) to ensure a unique limit structure, which in turn makes it possible to define a unique additive and continuous clustering in Theorem 21. Furthermore in Section 3.5 we explain how this framework is linked to density based clustering, and in the examples of Section 4.3 we investigate the consequences in the setting of mixed Hausdorff dimensions.

1.1 Related Work

Some axioms for clustering have been proposed and investigated, but to our knowledge, all approaches concern clustering of finite data. Jardine and Sibson (1971) were probably the first to consider axioms for hierarchical clusterings: these are maps of sets of dissimilarity matrices to sets of e.g. ultrametric matrices. Given such sets they obtain continuity and uniqueness of such a map using several axioms. This setting was used by Janowitz and Wille (1995) to classify clusterings that are equivariant for all monotone transformations of the values of the distance matrix. Later, Puzicha et al. (1999) investigate axioms for cost functions of data-partitionings and then obtain clustering functions as optimizers of such cost functions. They consider as well a hierarchical version, marking the last axiomatic treatment of that case until today. More recently, Kleinberg (2003) put forward an impossibility result. He gives three axioms and shows that any (non-hierarchical) clustering of distance matrices can fulfill at most two of them. Zadeh and Ben-David (2009) remedy the impossibility by restricting to k-partitions, and they use minimum spanning trees to characterize different clustering functions. A completely different setting is Meilă (2005) where an arsenal of axioms is given for distances of clustering partitions. They characterize some distances (variation of information, classification error metric) using different subsets of their axioms.

One of the reviewers brought clustering of discrete data to our attention. As far as we understand, consensus clustering (Mirkin, 1975; Day and McMorris, 2003) and additive clustering (Shepard and Arabie, 1979; Mirkin, 1987) are popular in social studies clustering communities. What we call additive clustering in this work is something completely different though. Still, application of our notions to clustering of discrete structures warrants further research.

1.2 Spirit of Our Ansatz

Let us now give a brief description of our approach. To this end assume for simplicity that we wish to find a hierarchical clustering for certain distributions on $\mathbb{R}^d$. We denote the set of such distributions by $\mathcal{P}$. Then a clustering is simply a map c that assigns every $P \in \mathcal{P}$ to a collection $c(P)$ of non-empty events. Since we are interested in hierarchical clustering, $c(P)$ will always be a forest, i.e. we have

$$A, A' \in c(P) \implies A \perp A' \text{ or } A \subset A' \text{ or } A \supset A'. \tag{1}$$

Here $A \perp A'$ means *sufficiently distinct*, i.e. $A \cap A' = \emptyset$ or something stronger (cf. Definition 1. Following the idea that eventually one needs to store and process the clustering $c(P)$ on a computer, our first axiom assumes that $c(P)$ is *finite*. For a distribution with a continuous density the level set forest, i.e. the collection of all connected components of density level sets, will therefore *not* be viewed as a clustering. For densities with finitely many modes, however, this level set forest consists of long chains interrupted by finitely many branchings. In this case, the most relevant information for clustering is certainly represented at the branchings and not in the intermediate chains. Based on this observation, our second clustering axiom postulates that $c(P)$ does not contain chains. More precisely, if $s(F)$ denotes the forest that is obtained by replacing each chain in the forest F by the maximal element of the chain, our **structured forest axiom** demands that

$$s(c(P)) = c(P)\,. \tag{2}$$

To simplify notations we further extend the clustering to the cone defined by $\mathcal{P}$ by setting

$$c(\alpha P) := c(P) \tag{3}$$

for all $\alpha > 0$ and $P \in \mathcal{P}$. Equivalently we can view $\mathcal{P}$ as a collection of finite non-trivial measures and c as a map on $\mathcal{P}$ such that for $\alpha > 0$ and $P \in \mathcal{P}$ we have $\alpha P \in \mathcal{P}$ and $c(\alpha P) = c(P)$. It is needless to say that this extended view on clusterings does not change the nature of a clustering.

Our next two axioms are based on the observation that there do not only exist distributions for which the “right notion” of a clustering is debatable but there are also distributions for which everybody would agree about the clustering. For example, if P is the uniform distribution on a Euclidean ball B, then certainly everybody would set $c(P) = \{B\}$. Clearly, other such examples are possible, too, and therefore we view the determination of distributions with such simple clusterings as a *design decision*. More precisely, we assume that we have a collection $\mathcal{A}$ of closed sets, called **base sets** and a family $\mathcal{Q} = \{Q_A\}_{A \in \mathcal{A}} \subset \mathcal{P}$ called **base measures** with the property $A = \operatorname{supp} Q_A$ for all $A \in \mathcal{A}$. Now, our **base measure axiom** stipulates

$$c(Q_A) = \{A\}. \tag{4}$$

It is not surprising that different choices of $\mathcal{A}$, $\mathcal{Q}$, and $\perp$ may lead to different clusterings. In particular we will see that larger classes $\mathcal{A}$ usually result in more distributions for which we can construct a clustering satisfying all our clustering axioms. On the other hand, taking a larger class $\mathcal{A}$ means that more agreement needs to be sought about the distributions having a trivial clustering (4). For this reason the choice of $\mathcal{A}$ can be viewed as a trade-off.

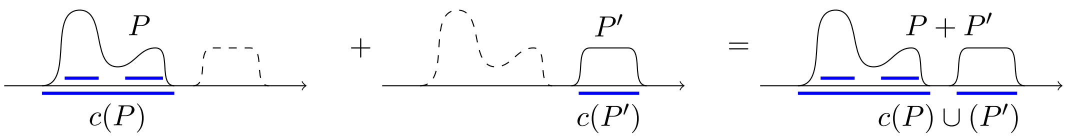


Figure 1: Example of disjoint additivity for two distributions having a density.

Axiom (4) only describes distributions that have a trivial clustering. However, there are also distributions for which everybody would agree on a non-trivial clustering. For example, if P is the uniform distribution on two well separated Euclidean balls B_1 and B_2, then the "natural" clustering would be $c(P) = \{B_1, B_2\}$. Our **disjoint additivity axiom** generalizes this observation by postulating

$$\operatorname{supp} P_1 \perp \operatorname{supp} P_2 \implies c(P_1 + P_2) = c(P_1) \cup c(P_2)\,. \tag{5}$$

In other words, if P consists of two spatially well separated sources P_1 and P_2, the clustering of P should reflect this spatial separation, see also Figure 1. Moreover note this axiom formalizes the vague term "spatially well separated" with the help of the relation $\perp$, which, like $\mathcal{A}$ and $\mathcal{Q}$ is a design parameter that usually influences the nature of the clustering.

The axioms (4) and (5) only described the horizontal behaviour of clusterings, i.e. the depth of the clustering forest is not affected by (4) and (5). Our second additivity axiom addresses this. To motivate it, assume that we have a $P \in \mathcal{P}$ and a base measure Q_A, e.g. a uniform distribution on A, such that $\operatorname{supp} P \subset A$. Then adding Q_A to P can be viewed as pouring uniform noise over P. Intuitively, this uniform noise should not affect the internal and possibly delicate clustering of P but only its roots, see also Figure 2. Our **base additivity axiom** formalizes this intuition by stipulating

$$\operatorname{supp} P \subset A \implies c(P + Q_A) = s\big(c(P) \cup \{A\}\big). \tag{6}$$

Here the structure operation $s(\,\cdot\,)$ is applied on the right-hand side to avoid a conflict with the structured forest axiom (2). Also note that it is this very axiom that directs our theory towards hierarchical clustering, since it controls the vertical growth of clusterings under a simple operation.

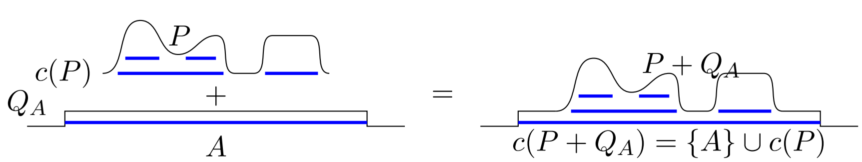


Figure 2: Example of base additivity.

Any clustering satisfying the axioms (1) to (6) will be called an **additive clustering**. Now the first, and rather simple part of our theory shows that under some mild technical assumptions there is a *unique* additive clustering on the set of **simple measures on forests**

$$\mathcal{S}(\mathcal{A}) := \left\{ \sum_{A\in F} \alpha_A Q_A \mid F \subset \mathcal{A} \text{ is a forest and } \alpha_A > 0 \text{ for all } A \in F \right\}.$$

Moreover, for $P \in \mathcal{S}(\mathcal{A})$ there is a unique representation $P = \sum_{A\in F} \alpha_A Q_A$ and the additive clustering is given by $c(P) = s(F)$.

Unfortunately, the set $\mathcal{S}(\mathcal{A})$ of simple measures, on which the uniqueness holds, is usually rather small. Consequently, additive clusterings on large collections $\mathcal{P}$ are far from being uniquely determined. Intuitively, we may hope to address this issue if we additionally impose some sort of continuity on the clusterings, i.e. an implication of the form

$$P_n \to P \implies c(P_n) \to c(P) . \tag{7}$$

Indeed, having an implication of the form (7), it is straightforward to show that the clustering is not only uniquely determined on $\mathcal{S}(\mathcal{A})$ but actually on the "closure" of $\mathcal{S}(\mathcal{A})$. To find a formalization of (7), we first note that from a user perspective, $c(P_n) \to c(P)$ usually describes a *desired* type of convergence. Following this idea, $P_n \to P$ then describes a *sufficient* condition for (7) to hold. In the remainder of this section we thus begin by presenting desirable properties $c(P_n) \to c(P)$ and resulting *necessary* conditions on $P_n \to P$.

Let us begin by assuming that all P_n are contained in $\mathcal{S}(\mathcal{A})$ and let us further denote the corresponding forests in the unique representation of P_n by F_n. Then we already know that $c(P_n) = s(F_n)$, so that the convergence on the right hand side of (7) becomes

$$s(F_n) \to c(P) . \tag{8}$$

Now, every $s(F_n)$, as well as $c(P)$, is a finite forest, and so a minimal requirement for (8) is that $s(F_n)$ and $c(P)$ are graph isomorphic, at least for all sufficiently large n. Moreover, we certainly also need to demand that every node in $s(F_n)$ converges to the corresponding node in $c(P)$. To describe the latter postulation more formally, we fix graph isomorphisms $\zeta_n : s(F_1) \to s(F_n)$ and $\zeta : s(F_1) \to c(P)$. Then our postulate reads as

$$\zeta_n(A) \to \zeta(A), \tag{9}$$

for all $A \in s(F_1)$. Of course, there do exist various notions for describing convergence of sets, e.g. in terms of the symmetric difference or the Hausdorff metric, so at this stage we need to make a decision. To motivate our choice, we first note that (9) actually contains two statements, namely, that $\zeta_n(A)$ converges for $n \to \infty$, and that its limit equals $\zeta(A)$. Now recall from various branches of mathematics that definitions of continuous extensions typically separate these two statements by considering approximating sequences that automatically converge. Based on this observation, we decided to consider monotone sequences in (9), i.e. we assume that $A \subset \zeta_1(A) \subset \zeta_2(A) \subset \dots$ for all $A \in s(F_1)$. Let us denote the resulting limit forest by F_∞, i.e.

$$F_\infty := \left\{ \bigcup_n \zeta_n(A) \mid A \in s(F_1) \right\},$$

which is indeed a forest under some mild assumptions on $\mathcal{A}$ and $\perp$. Moreover, $\zeta_\infty : s(F_1) \to F_\infty$ defined by $\zeta_\infty(A) := \bigcup_n \zeta_n(A)$ becomes a graph isomorphism, and hence (9) reduces to

$$\zeta_\infty(A) = \zeta(A) \qquad\qquad P\text{-almost surely for all } A \in s(F_1). \tag{10}$$

Summing up our considerations so far, we have seen that our demands on $c(P_n) \to c(P)$ imply some conditions on the forests associated to the sequence (P_n), namely $\zeta_n(A) \nearrow$ for all $A \in s(F_1)$. Without a formalization of $P_n \to P$, however, there is clearly no hope that this monotone convergence alone can guarantee (7). Like for (9), there are again various ways for formalizing a convergence of $P_n \to P$. To motivate our decision, we first note that a weak continuity axiom is certainly more desirable since this would potentially lead to more instances of clusterings. Furthermore, observe that (7) becomes weaker the stronger the notion of $P_n \to P$ is chosen. Now, if P_n and P had densities f_n and f, then one of the strongest notions of convergence would be $f_n \nearrow f$. In the absence of densities such a convergence can be expressed by $P_n \nearrow P$, i.e. by

$$P_n(B) \nearrow P(B) \qquad\qquad \text{for all measurable } B.$$

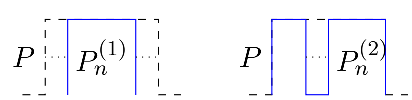
Combining these ideas we write $(P_n, F_n) \nearrow P$ iff $P_n \nearrow P$ and there are graph isomorphisms $\zeta_n : s(F_1) \to s(F_n)$ with $\zeta_n(A) \nearrow$ for all $A \in s(F_1)$. Our formalization of (7) then becomes

$$(P_n, F_n) \nearrow P \implies F_\infty = c(P) \text{ in the sense of (10)}, \tag{11}$$

which should hold for all $P_n \in \mathcal{S}(\mathcal{A})$ and their representing forests F_n.

While it seems tempting to stipulate such a continuity axiom it is unfortunately *inconsistent*. To illustrate this inconsistency, consider, for example, the uniform distribution P on $[0, 1]$. Then P can be approximated by the following two sequences

$$P_n^{(1)} := \mathbf{1}_{[1/n,1-1/n]}P$$
$$P_n^{(2)} := \mathbf{1}_{[0,1/2-1/n]}P + \mathbf{1}_{[1/2,1]}P$$



By (11) the first approximation would then lead to the clustering $c(P) = \{[0, 1]\}$ while the second approximation would give $c(P) = \{[0, 1/2), [1/2, 1]\}$.

Interestingly, this example not only shows that (11) is inconsistent but it also gives a hint how to resolve the inconsistency. Indeed the first sequence seems to be "adapted" to the limiting distribution, whereas the second sequence $(P_n^{(2)})$ is intuitively too complicated since its members have two clusters rather than the anticipated one cluster. Therefore, the idea to find a consistent alternative to (11) is to restrict the left-hand side of (11) to "adapted sequences", so that our **continuity axiom** becomes

$$(P_n, F_n) \nearrow P \text{ and } P_n \text{ is } P\text{-adapted for all } n \implies F_\infty = c(P) \text{ in the sense of (10)}.$$

In simple words, our main result then states that there exists exactly one such continuous clustering on the closure of $\mathcal{S}(\mathcal{A})$. The main message of this paper thus is:
Starting with very simple building blocks $\mathcal{Q} = (Q_A)_{A\in\mathcal{A}}$ for which we (need to) agree that they only have one trivial cluster $\{A\}$, we can construct a unique additive and continuous clustering on a rich set of distributions. Or, in other words, as soon as we have fixed $(\mathcal{A}, \mathcal{Q})$ and a separation relation $\perp$, there is no ambiguity left what a clustering is.

What is left is to explore how the choice of the **clustering base** $(\mathcal{A}, \mathcal{Q}, \perp)$ influences the resulting clustering. To this end, we first present various clustering bases, which, e.g. describe minimal thickness of clusters, their shape, and how far clusters need to be apart from each other. For distributions having a Lebesgue density we then illustrate how different clustering bases lead to different clusterings. Finally, we show that our approach goes beyond density-based clusterings by considering distributions consisting of several lower dimensional, overlapping parts.

2. Additive Clustering

In this section we introduce base sets, separation relations, and simple measures, as well as the corresponding axioms for clustering. Finally, we show that there exists a unique additive clustering on the set of simple measures.

Throughout this work let $\Omega = (\Omega, \mathcal{T})$ be a Hausdorff space and let $\mathcal{B} \supset \sigma(\mathcal{T})$ be a σ-algebra that contains the Borel sets. Furthermore we assume that $\mathcal{M} = \mathcal{M}_\Omega$ is the set of finite, non-zero, inner regular measures P on Ω. Similarly $\mathcal{M}_\Omega^\infty$ denotes the set of non-zero measures on Ω if Ω is a Radon space and else of non-zero, inner regular measures on Ω. In this respect, recall that any Polish space—i.e. a completely metrizable separable space—is Radon. In particular all open and closed subsets of $\mathbb{R}^d$ are Polish spaces and thus Radon. For inner regular measures the support is well-defined and satisfies the usual properties, see Appendix A for details. The set $\mathcal{M}_\Omega$ forms a cone: for all $P, P' \in \mathcal{M}_\Omega$ and all $\alpha > 0$ we have $P + P' \in \mathcal{M}_\Omega$ and $\alpha P \in \mathcal{M}_\Omega$.

2.1 Base Sets, Base Measures, and Separation Relations

Intuitively, any notion of a clustering should combine aspects of concentration and contiguousness. What is a possible core of this? On one hand clustering should be *local* in the sense of disjoint additivity, which was presented in the introduction: If a measure P is understood on two parts of its support and these parts are *nicely separated* then the clustering should be just a union of the two local ones. Observe that in this case $\operatorname{supp} P$ is not connected! On the other hand—in view of base clustering—base sets need to be impossible to partition into nicely separated components. Therefore they ought to be *nicely connected*. Of course, the meaning of *nicely connected* and *nicely separated* are interdependent, and highly disputable. For this reason, our notion of clustering assumes that both meanings are specified in front, e.g. by the user. Provided that both meanings satisfy certain technical criteria, we then show, that there exists exactly one clustering. To motivate how these technical criteria may look like, let us recall that for all connected sets A and all closed sets $B_1, \ldots, B_k$ we have

$$A \subset B_1 \dot\cup \ldots \dot\cup B_k \implies \exists! i \le k \colon A \subset B_i. \tag{12}$$

The left hand side here contains the condition that the $B_1, \ldots, B_k$ are pairwise disjoint, for which we already introduced the following notation:

$$B \perp_\emptyset B' \;:\Longleftrightarrow\; B \cap B' = \emptyset\,.$$

In order to transfer the notion of connectedness to other relations it is handy to generalize the notation $B_1 \dot\cup \ldots \dot\cup B_k$. To this end, let $\perp$ be a relation on subsets of Ω. Then we denote the union $B_1 \cup \ldots \cup B_k$ of some $B_1, \ldots, B_k \subset \Omega$ by

$$B_1 \overset{\perp}{\cup} \ldots \overset{\perp}{\cup} B_k \,,$$

iff we have $B_i \perp B_j$ for all $i \neq j$. Now the key idea of the next definition is to generalize the notion of connectivity and separation by replacing $\perp_\emptyset$ in (12) by another suitable relation.

Definition 1 *Let $\mathcal{A} \subset \mathcal{B}$ be a collection of closed, non-empty sets. A symmetric relation $\perp$ defined on $\mathcal{B}$ is called a $\mathcal{A}$-**separation relation** iff the following holds:*

*(a) **Reflexivity**: For all $B \in \mathcal{B}$: $B \perp B \implies B = \emptyset$.*

*(b) **Monotonicity**: For all $A, A', B \in \mathcal{B}$:*

$$A \subset A' \text{ and } A' \perp B \implies A \perp B.$$

*(c) $\mathcal{A}$-**Connectedness**: For all $A \in \mathcal{A}$ and all closed $B_1, \ldots, B_k \in \mathcal{B}$:*

$$A \subset B_1 \overset{\perp}{\cup} \ldots \overset{\perp}{\cup} B_k \implies \exists i \leq k \colon A \subset B_i.$$

*Moreover, an $\mathcal{A}$-separation relation $\perp$ is called **stable**, iff for all $A_1 \subset A_2 \subset \ldots$ with $A_n \in \mathcal{A}$, all $n \geq 1$, and all $B \in \mathcal{B}$:*

$$A_n \perp B \quad \text{for all } n \geq 1 \implies \bigcup_{n \geq 1} A_n \perp B. \tag{13}$$

*Finally, given a separation relation $\perp$ then we say that B, B' are $\perp$-**separated**, if $B \perp B'$. We write $B \not\perp B'$ iff not $B \perp B'$, and say in this case that B, B' are $\perp$-**connected**.*

It is not hard to check that the disjointness relation $\perp_\emptyset$ is a stable $\mathcal{A}$-separation relation, whenever all $A \in \mathcal{A}$ are topologically connected. To present another example of a separation relation, we fix a metric d on Ω and some $\tau > 0$. Moreover, for $B, B' \subset \Omega$ we write

$$B \perp_\tau B' \;:\iff d(B, B') \geq \tau \,.$$

In addition, recall that a $B \subset \Omega$ is τ-connected, if, for all $x, x' \in B$, there exists $x_0, \ldots, x_n \in B$ with $x_0 = x$, $x_n = x'$, and $d(x_{i-1}, x_i) < \tau$ for all $i = 1, \ldots, n$. Then it is easy to show that $\perp_\tau$ is an stable $\mathcal{A}$-separation relation if all $A \in \mathcal{A}$ are τ-connected. For more examples of separation relations we refer to Section 4.1.

It can be shown that $\perp_\emptyset$ is the weakest separation relation, i.e. for every $\mathcal{A}$-separation relation $\perp$ we have $A \perp A' \implies A \perp_\emptyset A'$ for all $A, A' \in \mathcal{A}$. We refer to Lemma 30, also showing that $\perp$-unions are unique, i.e., for all $A_1, \ldots, A_k$ and all $A'_1, \ldots, A'_{k'}$ in $\mathcal{A}$ we have

$$A_1 \overset{\perp}{\cup} \ldots \overset{\perp}{\cup} A_k = A'_1 \overset{\perp}{\cup} \ldots \overset{\perp}{\cup} A'_{k'} \implies \{A_1, \ldots, A_k\} = \{A'_1, \ldots, A'_{k'}\}.$$

Finally, the stability implication (13) is trivially satisfied for *finite* sequences $A_1 \subset \cdots \subset A_m$ in $\mathcal{A}$, since in this case we have $A_1 \cup \cdots \cup A_m = A_m$. For this reason stability will only become important when we will consider limits in Section 3.

We can now describe the properties a clustering base should satisfy.

Definition 2 *A (stable)* ***clustering base*** *is a triple* $(\mathcal{A}, \mathcal{Q}, \perp)$ *where* $\mathcal{A} \subset \mathcal{B} \setminus \{\emptyset\}$ *is a class of non-empty sets,* $\perp$ *is a (stable)* $\mathcal{A}$*-separation relation, and* $\mathcal{Q} = \{Q_A\}_{A \in \mathcal{A}} \subset \mathcal{M}$ *is a family of probability measures on* Ω *with the following properties:*

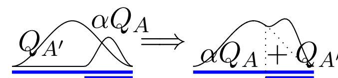
(a) ***Flatness****: For all* $A, A' \in \mathcal{A}$ *with* $A \subset A'$ *we either have* $Q_{A'}(A) = 0$ *or*

$$Q_A(\,\cdot\,) = \frac{Q_{A'}(\,\cdot \cap A)}{Q_{A'}(A)}\,.$$

(b) ***Fittedness****: For all* $A \in \mathcal{A}$ *we have* $A = \operatorname{supp} Q_A$*.*

We call a set A *a* ***base set*** *iff* $A \in \mathcal{A}$ *and a measure* $\mathfrak{a} \in \mathcal{M}$ *a* ***base measure on*** A *iff* $A \in \mathcal{A}$ *and there is an* $\alpha > 0$ *with* $\mathfrak{a} = \alpha Q_A$*.*

Let us motivate the two conditions of clustering bases. Flatness concerns nesting of base sets: Let $A \subset A'$ be base sets and consider the sum of their base measures $Q_A + Q_{A'}$. If the clustering base is not flat, weird things can happen—see the right. The way we defined flatness excludes such cases without taking densities into account. As a result we will be able to handle aggregations of measures of different Hausdorff-dimension in Section 4.3. Fittedness, on the other hand, establishes a link between the sets $A \in \mathcal{A}$ and their associated base measures.



Probably, the easiest example of a clustering base has measures of the form

$$Q_A(\,\cdot\,) = \frac{\mu(\,\cdot \cap A)}{\mu(A)} = \frac{1_A\, d\mu}{\mu(A)}\,, \tag{14}$$

where μ is some reference measure independent of Q_A. The next proposition shows that under mild technical assumptions such distributions do indeed provide a clustering base.

Proposition 3 *Let* $\mu \in \mathcal{M}_\Omega^\infty$ *and* $\perp$ *be a (stable)* $\mathcal{A}$*-separation relation for some* $\mathcal{A} \subset \mathcal{K}(\mu)$*, where*

$$\mathcal{K}(\mu) := \big\{\, C \in \mathcal{B} \mid 0 < \mu(C) < \infty \text{ and } C = \operatorname{supp}\ \mu(\cdot \cap C) \,\big\}$$

denotes the set of μ***-support sets****. We write* $\mathcal{Q}^{\mu,\mathcal{A}} := \big\{\, Q_A \mid A \in \mathcal{A} \,\big\}$*, where* Q_A *is defined by* (14)*. Then* $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ *is a (stable) clustering base.*

Interestingly, distributions of the form (14) are not the only examples for clustering bases. For further details we refer to Section 4.3, where we discuss distributions supported by sets of different Hausdorff dimension.

2.2 Forests, Structure, and Clustering

As outlined in the introduction we are interested in hierarchical clusterings, i.e. in clustering that map a finite measure to a forest of sets. In this section we therefore recall some fundamental definitions and notations for such forests.

Definition 4 *Let $\mathcal{A}$ be a class of closed, non-empty sets, $\perp$ be an $\mathcal{A}$-separation relation, and $\mathcal{C}$ be a class with $\mathcal{A} \subset \mathcal{C} \subset \mathcal{B} \setminus \{\emptyset\}$. We say that a non-empty $F \subset \mathcal{C}$ is a ($\mathcal{C}$-valued) $\perp$-**forest** iff*

$$A, A' \in F \implies A \perp A' \text{ or } A \subset A' \text{ or } A' \subset A.$$

We denote the set of all such finite forests by $\mathcal{F}_{\mathcal{C}}$ and write $\mathcal{F} := \mathcal{F}_{\mathcal{B}\setminus\{\emptyset\}}$.

A finite $\perp$-forest $F \in \mathcal{F}$ is partially ordered by the inclusion relation. The maximal elements $\max F := \{A \in F : \nexists A' \in F \text{ s.t. } A \subsetneqq A'\}$ are called **roots** and the minimal elements $\min F := \{A \in F : \nexists A' \in F \text{ s.t. } A' \subsetneqq A\}$ are called **leaves**. It is not hard to see that $A \perp A'$, whenever $A, A' \in F$ is a pair of roots or leaves. Moreover, the **ground** of F is

$$\mathbb{G}(F) := \bigcup_{A \in F} A\,,$$

that is, $\mathbb{G}(F)$ equals the union over the roots of F. Finally, F is a **tree**, iff it has only a single root, or equivalently, $\mathbb{G}(F) \in F$, and F is a **chain** iff it has a single leaf, or equivalently, iff it is totally ordered.

In addition to these standard notions, we often need a notation for describing certain sub-forests. Namely, for a finite forest $F \in \mathcal{F}$ with $A \in F$ we write

$$F\big|_{\supsetneqq A} := \{A' \in F \mid A' \supsetneqq A\}$$

for the chain of strict ancestors of A. Analogously, we will use the notations $F\big|_{\supset A}$, $F\big|_{\subset A}$, and $F\big|_{\subsetneqq A}$ for the chain of ancestors of A (including A), the tree of descendants of A (including A), and the finite forest of strict descendants of A, respectively. We refer to Figure 3 for an example of these notations.

Definition 5 *Let F be a finite forest. Then we call $A_1, A_2 \in F$ **direct siblings** iff $A_1 \neq A_2$ and they have the same strict ancestors, i.e. $F\big|_{\supsetneqq A_1} = F\big|_{\supsetneqq A_2}$. In this case, any element*

$$A' \in \min F\big|_{\supsetneqq A_1} = \min F\big|_{\supsetneqq A_2}$$

*is called a **direct parent** of A_1 and A_2. On the other hand for $A, A' \in F$ we denote A' as a **direct child** of A, iff*

$$A' \in \max F\big|_{\subsetneqq A}.$$

*Moreover, the **structure** of F is defined by*

$$s(F) := \Big\{ A \in F \,\Big|\, A \text{ is a root or it has a direct sibling } A' \in F \Big\}$$

*and F is a **structured forest** iff $F = s(F)$.*

For later use we note that direct siblings A_1, A_2 in a $\perp$-forest F always satisfy $A_1 \perp A_2$. Moreover, the structure of a forest is obtained by pruning all sub-chains in F, see Figure 3. We further note that $s(s(F)) = s(F)$ for all forests, and if F, F' are structured $\perp$-forests with $\mathbb{G}(F) \perp \mathbb{G}(F')$ then we have $s(F \cup F') = F \cup F'$.

Let us now present our first set of axioms for (hierarchical) clustering.

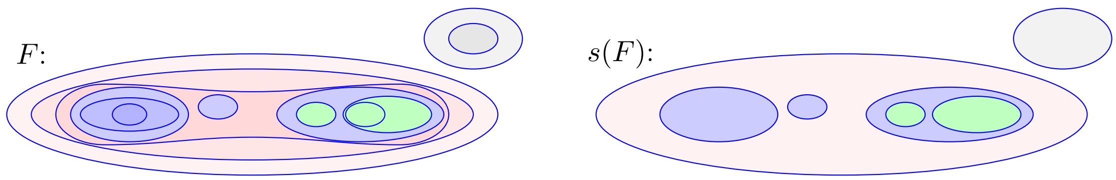


Figure 3: Illustrations of a forest F and of its structure $s(F)$.

Axiom 1 (Clustering) *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $\mathcal{P} \subset \mathcal{M}_\Omega$ be a set of measures with $\mathcal{Q} \subset \mathcal{P}$. A map $c\colon \mathcal{P} \to \mathcal{F}$ is called an **$\mathcal{A}$-clustering** if it satisfies:*

*(a) **Structured**: For all $P \in \mathcal{P}$ the forest $c(P)$ is structured, i.e. $c(P) = s(c(P))$.*

*(b) **ScaleInvariance**: For all $P \in \mathcal{P}$ and $\alpha > 0$ we have $\alpha P \in \mathcal{P}$ and $c(\alpha P) = c(P)$.*

*(c) **BaseMeasureClustering**: For all $A \in \mathcal{A}$ we have $c(Q_A) = \{A\}$.*

Note that the scale invariance is solely for notational convenience. Indeed, we could have defined clusterings for distributions only, in which case the scale invariance would have been obsolete. Moreover, assuming that a clustering produces structured forests essentially means that the clustering is only interested in the skeleton of the cluster forest. Finally, the axiom of base measure clustering means that we have a set of elementary measures, namely the base measures, for which we already agreed upon that they can only be clustered in a trivial way. In Section 4 we will present a couple of examples of $(\mathcal{A}, \mathcal{Q}, \perp)$ for which such an agreement is possible. Finally note that these axioms guarantee that if $c\colon \mathcal{P} \to \mathcal{F}$ is a clustering and $\mathfrak{a}$ is a base measure on A then $\mathfrak{a} \in \mathcal{P}$ and $c(\mathfrak{a}) = \{A\}$.

2.3 Additive Clustering

So far our axioms only determine the clusterings for base measures. Therefore, the goal of this subsection is to describe the behaviour of clusterings on certain combinations of measures. Furthermore, we will show that the axioms describing this behaviour are consistent and uniquely determine a hierarchical clustering on a certain set of measures induced by $\mathcal{Q}$.

Let us begin by introducing the axioms of additivity which we have already described and motivated in the introduction.

Axiom 2 (Additive Clustering) *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $\mathcal{P} \subset \mathcal{M}_\Omega$ be a set of measures with $\mathcal{Q} \subset \mathcal{P}$. A clustering $c\colon \mathcal{P} \to \mathcal{F}$ is called **additive** iff the following conditions are satisfied:*

*(a) **DisjointAdditivity**: For all $P_1, \ldots, P_k \in \mathcal{P}$ with mutually $\perp$-separated supports, i.e. $\operatorname{supp} P_i \perp \operatorname{supp} P_j$ for all $i \neq j$, we have $P_1 + \ldots + P_k \in \mathcal{P}$ and*

$$c(P_1 + \ldots + P_n) = c(P_1) \cup \cdots \cup c(P_n)\,.$$

*(b) **BaseAdditivity**: For all $P \in \mathcal{P}$ and all base measures $\mathfrak{a}$ with $\operatorname{supp} P \subset \operatorname{supp} \mathfrak{a}$ we have $\mathfrak{a} + P \in \mathcal{P}$ and*

$$c(\mathfrak{a} + P) = s\big(\{\operatorname{supp} \mathfrak{a}\} \cup c(P)\big).$$

Our next goal is to show that there exist additive clusterings and that these are uniquely on a set $\mathcal{S}$ of measures that, in some sense, is spanned by $\mathcal{Q}$. The following definition introduces this set.

Definition 6 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $F \in \mathcal{F}_{\mathcal{A}}$ be an $\mathcal{A}$-valued finite $\perp$-forest. A measure Q is **simple on F** iff there exist base measures $\mathfrak{a}_A$ on $A \in F$ such that*

$$Q = \sum_{A \in F} \mathfrak{a}_A. \tag{15}$$

We denote the set of all simple measures with respect to $(\mathcal{A}, \mathcal{Q}, \perp)$ by $\mathcal{S} := \mathcal{S}(\mathcal{A})$.

Figure 4 provides an example of a simple measure. The next lemma shows that the representation 15 of simple measures is actually unique.

Lemma 7 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $Q \in \mathcal{S}(\mathcal{A})$. Then there exists exactly one $F_Q \in \mathcal{F}_{\mathcal{A}}$ such that Q is simple on F_Q. Moreover, the representing base measures $\mathfrak{a}_A$ in (15) are also unique and we have $\operatorname{supp} Q = \mathbb{G} F_Q$.*

Using Lemma 7 we can now define certain restrictions of simple measures $Q \in \mathcal{S}(\mathcal{A})$ with representation (15). Namely, any subset $F' \subset F$ gives a measure

$$Q\big|_{F'} := \sum_{A \in F'} \mathfrak{a}_A \,.$$

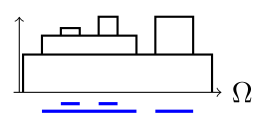


Figure 4: Simple measure.

We write $Q\big|_{\supset A} := Q\big|_{F|_{\supset A}}$ and similarly $Q\big|_{\supsetneq A}, Q\big|_{\subset A}, Q\big|_{\subsetneq A}$.

With the help of Lemma 7 it is now easy to explain how a possible additive clustering could look like on $\mathcal{S}(\mathcal{A})$. Indeed, for a $Q \in \mathcal{S}(\mathcal{A})$, Lemma 7 provides a unique finite forest $F_Q \in \mathcal{F}_{\mathcal{A}}$ that represents Q and therefore the structure $s(F_Q)$ is a natural candidate for a clustering of Q. The next theorem shows that this idea indeed leads to an additive clustering and that every additive clustering on $\mathcal{S}(\mathcal{A})$ retrieves the structure of the underlying forest of a simple measure.

Theorem 8 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $\mathcal{S}(\mathcal{A})$ the set of simple measures. Then we can define an additive $\mathcal{A}$-clustering $c : \mathcal{S}(\mathcal{A}) \to \mathcal{F}_{\mathcal{A}}$ by*

$$c(Q) := s(F_Q)\,, \qquad\qquad Q \in \mathcal{S}(\mathcal{A})\,. \tag{16}$$

Moreover, every additive $\mathcal{A}$-clustering $c : \mathcal{P} \to \mathcal{F}$ satisfies both $\mathcal{S}(\mathcal{A}) \subset \mathcal{P}$ and (16).

3. Continuous Clustering

As described in the introduction, we typically need, besides additivity, also some notion of continuity for clusterings. The goal of this section is to introduce such a notion and to show that, similarly to Theorem 8, this continuity uniquely defines a clustering on a suitably defined extension of $\mathcal{S}(\mathcal{A})$.

To this end, we first introduce a notion of monotone convergence for sequences of simple measures that does not change the graph structure of the corresponding clusterings given by Theorem 8. We then discuss a richness property of the clustering base, which essentially ensures that we can approximate the non-disjoint union of two base sets by another base set. In the next step we describe monotone sequences of simple measures that are in some sense adapted to the limiting distribution. In the final part of this section we then axiomatically describe continuous clusterings and show both their existence and their uniqueness.

3.1 Isomonotone Limits

The goal of this section is to introduce a notion of monotone convergence for simple measures that preserves the graph structure of the corresponding clusterings.

Our first step in this direction is done in the following definition that introduces a sort of monotonicity for set-valued isomorphic forests.

Definition 9 *Let $F, F' \in \mathcal{F}$ be two finite forests. Then F and F' are* ***isomorphic****, denoted by $F \cong F'$, iff there is a bijection $\zeta\colon F \to F'$ such that for all $A, A' \in F$ we have:*

$$A \subset A' \iff \zeta(A) \subset \zeta(A'). \tag{17}$$

Moreover, we write $F \leq F'$ iff $F \cong F'$ and there is a map $\zeta : F \to F'$ satisfying 17 and

$$A \subset \zeta(A)\,, \qquad\qquad A \in F. \tag{18}$$

In this case, the map ζ, which is uniquely determined by (17), (18) *and the fact that F and F' are finite, is called the* ***forest relating map*** *(FRM) between F and F'.*

Forests can be viewed as directed acyclic graphs: There is an edge between A and A' in F iff $A \subset A'$ and no other node is in between. Then $F \cong F'$ holds iff F and F' are isomorphic as directed graphs. From this it becomes clear that $\cong$ is an equivalence relation. Moreover, the relation $F \leq F'$ means that each node A of F can be graph isomorphically mapped to a node of F' that contains A, see Figure 5 for an illustration. Note that $\leq$ is a partial order on $\mathcal{F}$ and in particular it is transitive. Consequently, if we have finite forests $F_1 \leq \cdots \leq F_k$ then $F_1 \leq F_k$ and there is an FRM $\zeta_k\colon F_1 \to F_k$. This observation is used in the following definition, which introduces monotone sequences of forests and their limit.

Definition 10
An ***isomonotone sequence of forests*** *is a sequence of finite forests $(F_n)_n \subset \mathcal{F}$ such that $s(F_n) \leq s(F_{n+1})$ for all $n \geq 1$. If this is the case, we define the limit by*

$$F_\infty := \lim_{n\to\infty} s(F_n) := \Big\{ \bigcup_{n\geq 1} \zeta_n(A) \mid A \in s(F_1) \Big\},$$

where $\zeta_n\colon s(F_1) \to s(F_n)$ is the FRM obtained from $s(F_1) \leq s(F_n)$.

It is easy to see that in general, the limit forest F_∞ of an isomonotone sequence of $\mathcal{A}$-valued forests is not $\mathcal{A}$-valued. To describe the values of F_∞ we define the **monotone closure** of an $\mathcal{A} \subset \mathcal{B}$ by

$$\bar{\mathcal{A}} := \left\{ \bigcup_{n\geq 1} A_n \mid A_n \in \mathcal{A} \text{ and } A_1 \subset A_2 \subset \dots \right\}.$$

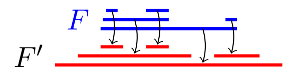


Figure 5: $F \leq F'$ and the arrows indicate ζ.

The next lemma states some useful properties of $\bar{\mathcal{A}}$ and F_∞.

Lemma 11 *Let $\perp$ be an $\mathcal{A}$-separation relation. Then $\perp$ is actually an $\bar{\mathcal{A}}$-separation relation. Moreover, if $\perp$ is stable and $(F_n) \subset \mathcal{F}_\mathcal{A}$ is an isomonotone sequence then $F_\infty := \lim_n s(F_n)$ is an $\bar{\mathcal{A}}$-valued $\perp$-forest and we have $s(F_n) \leq F_\infty$ for all $n \geq 1$.*

Unlike forests, it is straightforward to compare two measures Q_1 and Q_2 on $\mathcal{B}$. Indeed, we say that Q_2 **majorizes** Q_1, in symbols $Q_1 \leq Q_2$, iff

$$Q_1(B) \leq Q_2(B), \qquad \text{for all } B \in \mathcal{B}.$$

For $(Q_n) \subset \mathcal{M}$ and $P \in \mathcal{M}$, we similarly speak of **monotone convergence** $Q_n \uparrow P$ iff $Q_1 \leq Q_2 \leq \dots \leq P$ and

$$\lim_{n\to\infty} Q_n(B) = P(B), \qquad \text{for all } B \in \mathcal{B}.$$

Clearly, $Q \leq Q'$ implies $\operatorname{supp} Q \subset \operatorname{supp} Q'$ and it is easy to show, that $Q_n \uparrow P$ implies

$$P\big(\operatorname{supp} P \setminus \bigcup_{n\geq 1} \operatorname{supp} Q_n\big) = 0.$$

We will use such arguments throughout this section. For example, if $\mathfrak{a}, \mathfrak{a}'$ are base measures on A, A' with $\mathfrak{a} \leq \mathfrak{a}'$ then $A \subset A'$. With the help of these preparations we can now define isomonotone convergence of simple measures.

Definition 12 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $(Q_n) \subset \mathcal{S}(\mathcal{A})$ be a sequence of simple measures on finite forests $(F_n) \subset \mathcal{F}_\mathcal{A}$. Then **isomonotone convergence**, denoted by $(Q_n, F_n) \uparrow P$, means that both*

$$Q_n \uparrow P \quad \text{and} \quad s(F_1) \leq s(F_2) \leq \dots .$$

In addition, $\bar{\mathcal{S}} := \bar{\mathcal{S}}(\mathcal{A})$ denotes the set of all isomonotone limits, i.e.

$$\bar{\mathcal{S}}(\mathcal{A}) = \big\{ P \in \mathcal{M} \mid (Q_n, F_n) \uparrow P \text{ for some } (Q_n) \subset \mathcal{S}(\mathcal{A}) \text{ on } (F_n) \subset \mathcal{F}_\mathcal{A} \big\}.$$

For a measure $P \in \bar{\mathcal{S}}(\mathcal{A})$ it is probably tempting to define its clustering by $c(P) := \lim_n s(F_n)$, where $(Q_n, F_n) \uparrow P$ is some isomonotone sequence. Unfortunately, such an approach does not yield a well-defined clustering as we have discussed in the introduction. For this reason, we need to develop some tools that help us to distinguish between “good” and “bad” isomonotone approximations. This is the goal of the following two subsections.

3.2 Kinship and Subadditivity

In this subsection we present and discuss a technical assumption on a clustering base that will make it possible to obtain unique continuous clusterings.

Let us begin by introducing a notation that will be frequently used in the following. To this end, we fix a clustering base $(\mathcal{A}, \mathcal{Q}, \perp)$ and a $P \in \mathcal{M}$. For $B \in \mathcal{B}$ we then define

$$\mathcal{Q}_P(B) := \{\alpha Q_A \mid \alpha > 0, A \in \mathcal{A}, B \subset A, \alpha Q_A \le P\},$$

i.e. $\mathcal{Q}_P(B)$ denotes the set of all basic measures below P whose support contains B. Now, our first definition describes events that can be combined in a base set:

Definition 13 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $P \in \mathcal{M}$. Two non-empty $B, B' \in \mathcal{B}$ are called* ***kin below*** *P, denoted as $B \sim_P B'$, iff $\mathcal{Q}_P(B \cup B') \neq \emptyset$, i.e., iff there is a base measure $\mathfrak{a} \in \mathcal{Q}$ such that the following holds:*

(a) $B \cup B' \subset \operatorname{supp} \mathfrak{a}$ *(b)* $\mathfrak{a} \le P$.

Moreover, we say that every such $\mathfrak{a} \in \mathcal{Q}_P(B \cup B')$ ***supports*** *B and B' below P.*

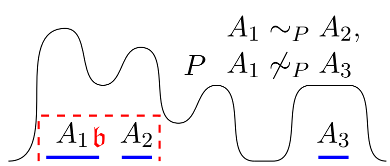


Figure 6: Kinship.

Kinship of two events can be used to test whether they belong to the same root in the cluster forest. To illustrate this we consider two events B and B' with $B \not\sim_P B'$. Moreover, assume that there is an $A \in \mathcal{A}$ with $B \cup B' \subset A$. Then $B \not\sim_P B'$ implies that for all such A there is no $\alpha > 0$ with $\alpha Q_A \le P$. This situation is displayed on the right-hand side of Figure 6. Now assume that we have two base measures $\mathfrak{a}, \mathfrak{a}' \le P$ on $A, A' \in \mathcal{A}$ that satisfy $A \sim_P A'$ and $P(A \cap A') > 0$. If $\mathcal{A}$ is rich in the sense of $A \cup A' \in \mathcal{A}$, then we can find a base measure $\mathfrak{b}$ on $B := A \cup A'$ with $\mathfrak{a} \le \mathfrak{b} \le P$ or $\mathfrak{a}' \le \mathfrak{b} \le P$. The next definition relaxes the requirement $A \cup A' \in \mathcal{A}$, see also Figure 7 for an illustration.

Definition 14 *Let $P \in \mathcal{M}_\Omega^\infty$ be a measure. For $B, B' \in \mathcal{B}$ we write*

$$B \perp\!\!\!\perp_P B' \;:\Longleftrightarrow\; P(B \cap B') = 0 \text{ and}$$
$$B \between_P B' \;:\Longleftrightarrow\; P(B \cap B') > 0\,.$$

Moreover, a clustering base $(\mathcal{A}, \mathcal{Q}, \perp)$ is called ***P-subadditive*** *iff for all base measures $\mathfrak{a}, \mathfrak{a}' \le P$ on $A, A' \in \mathcal{A}$ we have*

$$A \between_P A' \implies \exists \mathfrak{b} \in \mathcal{Q}_P(A \cup A') \colon \mathfrak{b} \ge \mathfrak{a} \text{ or } \mathfrak{b} \ge \mathfrak{a}'. \tag{19}$$

Note that the implication (19) in particular ensures $\mathcal{Q}_P(A \cup A') \neq \emptyset$, i.e. $A \sim_P A'$. Moreover, the relation $\perp\!\!\!\perp_P$ is weaker than any separation relation $\perp$ since we obviously have $A \between_P A' \implies A \between_\emptyset A' \implies A \between A'$, where the second implication is shown in Lemma 30. The following definition introduces a stronger notion of additivity.

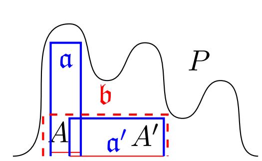


Figure 7: P-subadditivity.

Definition 15 *Let ∞ be a relation on $\mathcal{B}$. An $\mathcal{A} \subset \mathcal{B}$ is ∞-**additive** iff for all $A, A' \in \mathcal{A}$*

$$A \infty A' \implies A \cup A' \in \mathcal{A}.$$

The next proposition compares the several notions of (sub)-additivity. In particular it implies that if $\mathcal{A}$ is $\infty_\emptyset$-additive then $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ is P-subadditive for all $P \in \mathcal{M}$.

Proposition 16 *Let $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ be a clustering base as in Proposition 3. If $\mathcal{A}$ is ∞_P-additive for some $P \in \mathcal{M}$, then $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ is P-subadditive. Conversely, if $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ is P-subadditive for all $P \ll \mu$ then $\mathcal{A}$ is ∞_μ-additive and thus also ∞_P-additive.*

3.3 Adapted Simple Measures

We have already seen that isomonotone approximations by simple measures are not structurally unique. In this subsection we will therefore identify the most economical structure needed to approximate a distribution by simple measures. Such most parsimonious structures will then be used to define continuous clusterings.

Let us begin by introducing a different view on simple measures.

Definition 17 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and Q be a simple measure on $F \in \mathcal{F}_\mathcal{A}$ with the unique representation $Q = \sum_{A \in F} \alpha_A Q_A$. We define the map $\lambda_Q \colon F \to \mathcal{Q}$ by*

$$\lambda_Q(A) := \left(\sum_{A' \in F:\ A' \supset A} \alpha_{A'} Q_{A'}(A) \right) \cdot Q_A \,, \qquad\qquad A \in F.$$

*Moreover, we call the base measure $\lambda_Q(A) \in \mathcal{Q}$ the **level of A in Q**.*

In some sense, the level of an A in Q combines all ancestor measures including Q_A and then restricts this combination to A, see Figure 8 for an illustration of the level of a node. With the help of levels we can now describe structurally economical approximations of measures by simple measures.

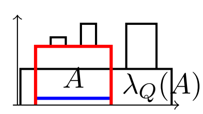


Figure 8: Level.

Definition 18 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $P \in \mathcal{M}_\Omega$ a finite measure. Then a simple measure Q on a forest $F \in \mathcal{F}_\mathcal{A}$ is P-**adapted** iff all direct siblings A_1, A_2 in F are:*

*(a) P-**grounded**: if they are kin below P, i.e. $\mathcal{Q}_P(A_1 \cup A_2) \neq \emptyset$, then there is a parent around them in F.*

*(b) P-**fine**: every $\mathfrak{b} \in \mathcal{Q}_P(A_1 \cup A_2)$ can be majorized by a base measure $\tilde{\mathfrak{b}}$ that supports all direct siblings $A_1, \ldots, A_k$ of A_1 and A_2, i.e.*

$$\mathfrak{b} \in \mathcal{Q}_P(A_1 \cup A_2) \implies \exists \tilde{\mathfrak{b}} \in \mathcal{Q}_P(A_1 \cup \ldots \cup A_k) \text{ with } \tilde{\mathfrak{b}} \geq \mathfrak{b}.$$

*(c) **strictly motivated**: for their levels $\mathfrak{a}_1 := \lambda_Q(A_1)$ and $\mathfrak{a}_2 := \lambda_Q(A_2)$ in Q there is an $\alpha \in (0,1)$ such that every base measure $\mathfrak{b}$ that supports them below P is not larger than $\alpha\mathfrak{a}_1$ or $\alpha\mathfrak{a}_2$, i.e.*

$$\forall \mathfrak{b} \in \mathcal{Q} : \mathfrak{b} \geq \alpha\mathfrak{a}_1 \text{ or } \mathfrak{b} \geq \alpha\mathfrak{a}_2 \implies \mathfrak{b} \notin \mathcal{Q}_P(A_1 \cup A_2). \tag{20}$$

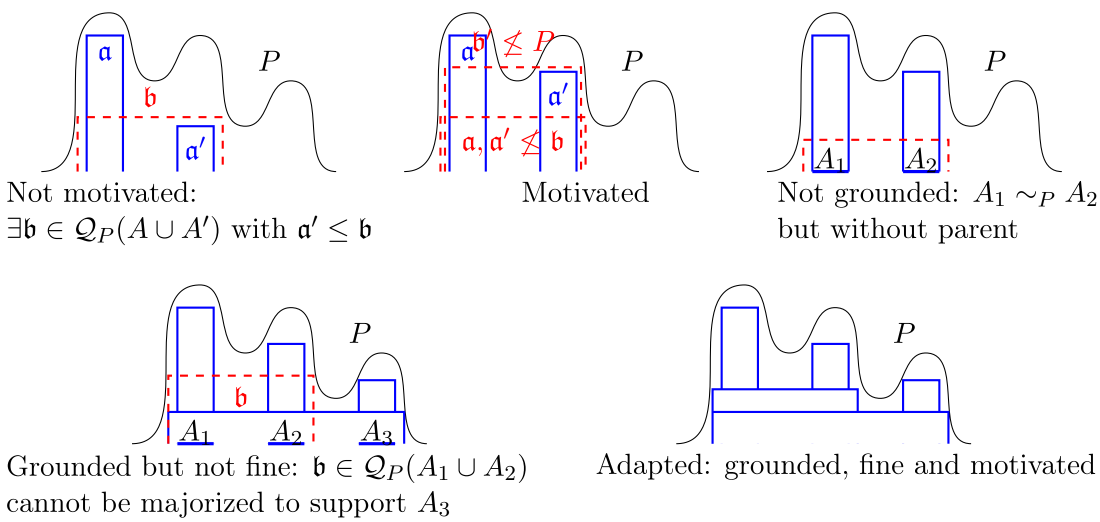


Figure 9: Illustrations for motivated, grounded, fine, and therefore adapted.

Finally, an isomonotone sequence $(Q_n, F_n) \uparrow P$ *is adapted, if* Q_n *is* P*-adapted for all* $n \geq 1$.

Since siblings are $\perp$-separated, they are $\perp\!\!\!\perp_P$-separated, so strict motivation is no contradiction to P-subadditivity. Levels are called **motivated** iff they satisfy condition (20) for $\alpha = 1$. Figure 9 illustrates the three conditions describing adapted measures. It can be shown that if $\mathcal{A}$ is ∞-additive, then any isomonotone sequence can be made adapted.

The following self-consistency result shows that every simple measure is adapted to itself. This result will guarantee that the extension of the clustering from $\mathcal{S}$ to $\bar{\mathcal{S}}$ is indeed an extension.

Proposition 19 *Let* $(\mathcal{A}, \mathcal{Q}, \perp)$ *be a clustering base. Then every* $Q \in \mathcal{S}(\mathcal{A})$ *is* Q*-adapted.*

3.4 Continuous Clustering

In this subsection we finally introduce continuous clusterings with the help of adapted, isomonotone sequences. Furthermore, we will show the existence and uniqueness of such clusterings.

Let us begin by introducing a notation that will be used to identify two clusterings as identical. To this end let $F_1, F_2 \in \mathcal{F}$ be two forests and $P \in \mathcal{M}_\Omega$ be finite measure. Then we write

$$F_1 =_P F_2 \,,$$

if there exists a graph isomorphism $\zeta : F_1 \to F_2$ such that $P(A \triangle \zeta(A)) = 0$ for all $A \in F_1$. Now our first result shows that adapted isomonotone limits of two different sequences coincide in this sense.

Theorem 20 *Let* $(\mathcal{A}, \mathcal{Q}, \perp)$ *be a stable clustering base and* $P \in \mathcal{M}_\Omega$ *be a finite measure such that* $\mathcal{A}$ *is* P*-subadditive. If* $(Q_n, F_n), (Q'_n, F'_n) \uparrow P$ *are adapted isomonotone sequences then we have*

$$\lim_n s(F_\infty) =_P \lim_n s(F'_\infty) \,.$$

Theorem 20 shows that different adapted sequences approximating a measure P necessarily have isomorphic forests and that the corresponding limit nodes of the forests coincide up to P-null sets. This result makes the following axiom possible.

Axiom 3 (Continuous Clustering) *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base, $\mathcal{P} \subset \mathcal{M}_\Omega$ be a set of measures. We say that $c\colon \mathcal{P} \to \mathcal{F}$ is a* ***continuous clustering****, if it is an additive clustering and for all $P \in \mathcal{P}$ and all adapted isomonotone sequences $(Q_n, F_n) \uparrow P$ we have*

$$c(P) =_P \lim_n s(F_\infty)\,.$$

The following, main result of this section shows that there exist continuous clusterings and that they are uniquely determined on a large subset of $\bar{\mathcal{S}}(\mathcal{A})$.

Theorem 21 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a stable clustering base and set*

$$\mathcal{P}_\mathcal{A} := \big\{\, P \in \bar{\mathcal{S}}(\mathcal{A}) \mid \mathcal{A} \text{ is } P\text{-subadditive and there is } (Q_n, F_n) \nearrow P \text{ adapted} \big\}.$$

Then there exists a continuous clustering $c_\mathcal{A}\colon \mathcal{P}_\mathcal{A} \to \mathcal{F}_{\bar{\mathcal{A}}}$. Moreover, $c_\mathcal{A}$ is unique on $\mathcal{P}_\mathcal{A}$, that is, for all continuous clusterings $c\colon \tilde{\mathcal{P}} \to \mathcal{F}$ we have

$$c_\mathcal{A}(P) =_P c(P)\,, \qquad\qquad P \in \mathcal{P}_\mathcal{A}\,.$$

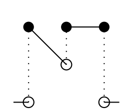
Recall from Proposition 16 that $\mathcal{A}$ is P-subadditive for all $P \in \mathcal{M}_\Omega$ if $\mathcal{A}$ is $\infty_\emptyset$-additive. It can be shown that if $\mathcal{A}$ is $\infty_\mathcal{A}$-additive, then any isomonotone sequence can be made adapted. In this case we thus have $\mathcal{P}_\mathcal{A} = \bar{\mathcal{S}}(\mathcal{A})$ and Theorem 21 shows that there exists a unique continuous clustering on $\bar{\mathcal{S}}(\mathcal{A})$.

3.5 Density Based Clustering

Let us recall from Proposition 3 that a simple way to define a set of base measures $\mathcal{Q}$ was with the help of a reference measure μ. Given a stable separation relation $\perp$, we denoted the resulting stable clustering base by $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$. Now observe that for this clustering base every $Q \in \mathcal{S}(\mathcal{A})$ is μ-absolutely continuous and its unique representation yields the μ-density $f = \sum_{A \in F} \alpha_A 1_A$ for suitable coefficients $\alpha_A > 0$. Consequently, each level set $\{f > \lambda\}$ consists of some elements $A \in F$, and if all elements in $\mathcal{A}$ are connected, the additive clustering $c(Q)$ of Q thus coincides with the "classical" cluster tree obtained from the level sets. It is therefore natural to ask, whether such a relation still holds for continuous clusterings on distributions $P \in \mathcal{P}_\mathcal{A}$.

Clearly, the first answer to this question needs to be negative, since in general the cluster tree is an infinite forest whereas our clusterings are always finite. To illustrate this, let us consider the **Factory** density on $[0,1]$, which is defined by

$$f(x) := \begin{cases} 1 - x, & \text{if } x \in [0, \frac{1}{2}) \\ 1, & \text{if } x \in [\frac{1}{2}, 1] \end{cases}$$



Clearly, this gives the following $\perp_\emptyset$-decomposition of the level sets:

$$\{f > \lambda\} = \begin{cases} [0,1], & \text{if } \lambda < \frac{1}{2}, \\ [0, 1-\lambda) \overset{\perp_\emptyset}{\cup} [\frac{1}{2}, 1], & \text{if } \frac{1}{2} \le \lambda < 1, \end{cases}$$

which leads to the clustering forest $F_f = \{ [0,1], [\frac{1}{2}, 1] \} \cup \{ [0, 1-\lambda) \mid \frac{1}{2} \le \lambda < 1 \}$. Now observe that even though F_f is infinite, it is as graph somehow simple: there is a root $[0,1]$, a node $[\frac{1}{2}, 1]$, and an infinite chain $[0, 1-\lambda)$, $\frac{1}{2} \le \lambda < 1$. Replacing this chain by its supremum $[0, \frac{1}{2})$ we obtain the structured forest

$$\{[0,1],\ [0,\tfrac{1}{2}), [\tfrac{1}{2},1]\},$$

for which we can then ask whether it coincides with the continuous clustering obtained from $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp_\emptyset)$ if $\mathcal{A}$ consists of all closed intervals in $[0,1]$ and μ is the Lebesgue measure.

To answer this question we first need to formalize the operation that assigns a structured to an infinite forest. To this end, let F be an arbitrary $\perp$-forest. We say that $\mathcal{C} \subset F$ is a pure chain, iff for all $C, C' \in \mathcal{C}$ and $A \in F \setminus \mathcal{C}$ the following two implications hold:

$$\begin{aligned} A \subset C &\implies A \subset C', \\ C \subset A &\implies C' \subset A. \end{aligned}$$

Roughly speaking, the first implication ensures that no node above a bifurcation is contained in the chain, while the second implication ensures that no node below a bifurcation is contained in the chain. With this interpretation in mind it is not surprising that we can define the structure of the forest F with the help of the maximal pure chains by setting

$$s(F) := \left\{ \bigcup \mathcal{C} \mid \mathcal{C} \subset F \text{ is a maximal pure chain} \right\}.$$

Note that for infinite forests the structure $s(F)$ may or may not be finite. For example, for the factory density it is finite as we have already seen above.

We have seen in Lemma 11 that the nodes of a continuous clustering are $\perp$-separated elements of $\bar{\mathcal{A}}$. Consequently, it only makes sense to compare continuous clustering with the structure of a level set forest, if this forest shares this property. This is ensured in the following definition.

Definition 22 *Let $f\colon \Omega \to [0,\infty]$ be a measurable function and $(\mathcal{A}, \mathcal{Q}, \perp)$ be a stable clustering base. Then f is of $(\mathcal{A}, \mathcal{Q}, \perp)$-**type** iff there is a dense subset $\Lambda \subset [0, \sup f)$ such that for all $\lambda \in \Lambda$ the level set $\{f > \lambda\}$ is a finite union of pairwise $\perp$-separated events $B_1(\lambda), \dots, B_{k(\lambda)}(\lambda) \in \bar{\mathcal{A}}$. If this is the case the **level set $\perp$-forest** is given by*

$$F_{f,\Lambda} := \{B_i(\lambda) \mid i \le k(\lambda) \text{ and } \lambda \in \Lambda\}.$$

Note that for given f and Λ the forest $F_{f,\Lambda}$ is indeed well-defined since $\perp$ is an $\bar{\mathcal{A}}$-separation relation by Lemma 11 and therefore the decomposition of $\{f > \lambda\}$ into the sets $B_1(\lambda), \dots, B_{k(\lambda)}(\lambda) \in \bar{\mathcal{A}}$ is unique by Lemma 30.

With the help of these preparations we can now formulate the main result of this subsection, which compares continuous clusterings with the structure of level set $\perp$-forests:

Theorem 23 *Let $\mu \in \mathcal{M}_\Omega$, $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ the stable clustering based described in Proposition 3, and $P \in \mathcal{M}_\Omega$ such that $\mathcal{A}$ is P-subadditive. Assume that P has a μ-density f that is of $(\mathcal{A}, \mathcal{Q}, \perp)$-type with a dense subset Λ such that $s(F_{f,\Lambda})$ is finite and for all $\lambda \in \Lambda$ and all $i < j \le k(\lambda)$ we have $\overline{B}_i(\lambda) \perp \overline{B}_j(\lambda)$. Then we have $P \in \bar{\mathcal{S}}(\mathcal{A})$ and*

$$c(P) =_\mu s(F_{f,\Lambda}).$$

On the other hand, it is not difficult to show that if $P \in \bar{\mathcal{S}}(\mathcal{A})$ then P has a density of $(\mathcal{A}, \mathcal{Q}, \perp)$-type. We do not know though whether there has to a density of $(\mathcal{A}, \mathcal{Q}, \perp)$-type for that even the closure of siblings are separated.

If $\operatorname{supp} \mu \neq \Omega$ one might think that this is not true since on the complement of the support anything goes. To be more precise—if μ is not inner regular and hence no support is defined—assume there is an open set $O \subset \Omega$ with $\mu(O) = 0$. This then means that there is no base set $A \subset O$, because base sets are support sets. Hence anything that would happen on O is determined by what happens in $\operatorname{supp} P$!

In the literature density based clustering is only considered for continuous densities since they may serve as a canonical version of the density. The following result investigates such densities.

Proposition 24 *For a compact $\Omega \subset \mathbb{R}^d$ and a measure $\mu \in \mathcal{M}_\Omega$ we consider the stable clustering base $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp_\emptyset)$. We assume that all open, connected sets are contained in $\bar{\mathcal{A}}$ and that $P \in \mathcal{M}_\Omega$ is a finite measure such that $\mathcal{A}$ is P-subadditive. If P has a continuous density f that has only finitely many local maxima $x_1^*, \ldots, x_k^*$ then $P \in \mathcal{P}_\mathcal{A}$ and there a bijection $\psi \colon \{x_1^*, \ldots, x_k^*\} \to \min c(P)$ such that*

$$x_i^* \in \psi(x_i^*)\,.$$

In this case $c(P) =_\mu \left\{ B_i\lambda \mid i \leq k(\lambda) \text{ and } \lambda \in \Lambda_0 \right\}$ where $\Lambda_0 = \left\{ 0 = \lambda_0 < \ldots < \lambda_m < \sup f \right\}$ is the finite set of levels at which the splits occur.

4. Examples

After having given the skeleton of this theory we now give more examples of how to use it. This should as well motivate some of the design decisions. It will also become clear in what way the choice of a clustering base $(\mathcal{A}, \mathcal{Q}, \perp)$ influences the clustering.

4.1 Base Sets and Separation Relations

In this subsection we present several examples of clustering bases. Our first three examples consider different separation relations.

Example 1 (Separation relations) *The following define stable $\mathcal{A}$-separation relations:*

(a) ***Disjointness****: If $\mathcal{A} \subset \mathcal{B}$ is a collection of non-empty, closed, and topologically connected sets then*

$$B \perp_\emptyset B' \iff B \cap B' = \emptyset.$$

(b) ***τ-separation****: Let (Ω, d) be a metric space, $\tau > 0$, and $\mathcal{A} \subset \mathcal{B}$ be a collection of non-empty, closed, and τ-connected sets then*

$$B \perp_\tau B' \;:\!\iff d(B, B') \geq \tau.$$

(c) ***Linear separation****: Let H be a Hilbert space with inner product $\langle \cdot \mid \cdot \rangle$ and $\Omega \subset H$. Then non-empty events $A, B \subset \Omega$ are* ***linearly separated****, $A \perp_\ell B$, iff $A \perp_\emptyset B$ and*

$$\exists v \in H \setminus \{0\}, \alpha \in \mathbb{R}\, \forall a \in A, b \in B \colon \langle a \mid v \rangle \leq \alpha \text{ and } \langle b \mid v \rangle \geq \alpha.$$

The latter means there is an affine hyperplane $U \subset \Omega$ such that A and B are on different sides. Then $\perp_\ell$ is a $\mathcal{A}$ separation relation if no base set $A \in \mathcal{A}$ can be written as a finite union of pairwise $\perp_\ell$-disjoint closed sets. It is stable if H is finite-dimensional.

Our next goal is to present some examples of base set collections $\mathcal{A}$. Since these describe the sets we need to agree upon that their can only be trivially clustered, smaller collections $\mathcal{A}$ are generally preferred. Let μ be the Lebesgue measure on $\mathbb{R}^d$. To define possible collections $\mathcal{A}$ we will consider the following building blocks in $\mathbb{R}^d$:

$$\begin{aligned}\mathcal{C}_{\text{Dyad}} &:= \big\{\text{axis-parallel boxes with dyadic coordinates}\big\},\\ \mathcal{C}_p &:= \big\{\text{closed } \ell_p^d\text{-balls }\big\}, \quad p \in [1,\infty]\,,\\ \mathcal{C}_{\text{Conv}} &:= \big\{\text{convex and compact } \mu\text{-support sets}\big\}.\end{aligned}$$

$\mathcal{C}_{\text{Dyad}}$ corresponds to the cells of a histogram whereas $\mathcal{C}_p$ has connections to moving-window density estimation. When combined with $\perp_\emptyset$ or $\perp_\tau$ and base measures of the form (14) these collections may already serve as clustering bases. However, $\bar{\mathcal{C}}_\bullet$ and $\bar{\mathcal{S}}_{\mathcal{C}}$ are not very rich since monotone increasing sequences in $\mathcal{C}_\bullet$ converge to sets of the same shape, and hence the sets in $\bar{\mathcal{C}}_\bullet$ have the same shape constraint as those in $\mathcal{C}_\bullet$. As a result the sets of measures $\bar{\mathcal{S}}_{\mathcal{C}_\bullet}$ for which we can determine the unique continuous clustering are rather small. However, more interesting collections can be obtained by considering finite, connected unions built of such sets. To describe such unions in general we need the following definition.

Definition 25 *Let $\perp\!\!\!\perp$ be a relation on $\mathcal{B}$, ∞ be its negation, and $\mathcal{C} \subset \mathcal{B}$ be a class of non-empty events. The $\perp\!\!\!\perp$-**intersection graph** on $\mathcal{C}$, $\mathcal{G}_{\perp\!\!\!\perp}(\mathcal{C})$, has $\mathcal{C}$ as nodes and there is an edge between $A, B \in \mathcal{C}$ iff $A \infty B$. We define:*

$$\mathbb{C}_{\perp\!\!\!\perp}(\mathcal{C}) := \{\, C_1 \cup \ldots \cup C_k \mid C_1, \ldots, C_k \in \mathcal{C} \textit{ and the graph } \mathcal{G}_{\perp\!\!\!\perp}(\{C_1, \ldots, C_k\}) \textit{ is connected}\}.$$

Obviously any separation relation can be used. But one can also consider weaker relations like $\perp\!\!\!\perp_P$, or e.g. $A \perp\!\!\!\perp A'$ if $A \cap A'$ has empty interior, or if it contains no ball of size τ. Such examples yield smaller $\mathcal{A}$ and indeed in these cases $\bar{\mathcal{S}}$ is much smaller.

The following example provides stable clustering bases.

Example 2 (Clustering bases) *The following examples are $\infty_\emptyset$-additive:*

$$\begin{aligned}\mathcal{A}_{Dyad} := \mathbb{C}_{\perp_\emptyset}(\mathcal{C}_{Dyad}) \quad &= \big\{\, \textit{finite connected unions of boxes with dyadic coordinates}\,\big\},\\ \mathcal{A}_p := \mathbb{C}_{\perp_\emptyset}(\mathcal{C}_p) \quad &= \big\{\, \textit{finite connected unions of closed } L^p\textit{-balls}\,\big\},\\ \mathcal{A}_{Conv} := \mathbb{C}_{\perp_\emptyset}(\mathcal{C}_{Conv}) \quad &= \big\{\, \textit{finite connected unions of convex } \mu\textit{-support sets}\,\big\}.\end{aligned}$$

Then $\mathcal{A}_{Dyad}, \mathcal{A}_p, \mathcal{A}_{Conv} \subset \mathcal{K}(\mu)$. Furthermore the following examples are ∞_τ-additive:

$$\mathcal{A}^\tau_{Dyad} := \mathbb{C}_{\perp_\tau}(\mathcal{C}_{Dyad}), \qquad \mathcal{A}^\tau_p := \mathbb{C}_{\perp_\tau}(\mathcal{C}_p), \qquad \mathcal{A}^\tau_{Conv} := \mathbb{C}_{\perp_\tau}(\mathcal{C}_{Conv}).$$

This leads to the following examples of stable clustering bases:

$$\begin{aligned}&(\mathcal{A}_{Dyad}, \mathcal{Q}^{\mu,\mathcal{A}_{Dyad}}, \perp_\emptyset), && (\mathcal{A}_p, \mathcal{Q}^{\mu,\mathcal{A}_p}, \perp_\emptyset), && (\mathcal{A}_{Conv}, \mathcal{Q}^{\mu,\mathcal{A}_{Conv}}, \perp_\emptyset),\\ &(\mathcal{A}^\tau_{Dyad}, \mathcal{Q}^{\mu,\mathcal{A}^\tau_{Dyad}}, \perp_\tau), && (\mathcal{A}^\tau_p, \mathcal{Q}^{\mu,\mathcal{A}^\tau_p}, \perp_\tau), && (\mathcal{A}^\tau_{Conv}, \mathcal{Q}^{\mu,\mathcal{A}^\tau_{Conv}}, \perp_\tau),\\ &(\mathcal{A}_{Dyad}, \mathcal{Q}^{\mu,\mathcal{A}_{Dyad}}, \perp_\tau), && (\mathcal{A}_p, \mathcal{Q}^{\mu,\mathcal{A}_p}, \perp_\tau), && (\mathcal{A}_{Conv}, \mathcal{Q}^{\mu,\mathcal{A}_{Conv}}, \perp_\tau).\end{aligned}$$

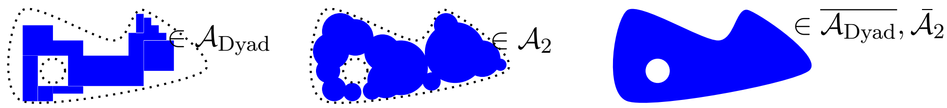


Figure 10: Some examples of sets in $\mathcal{A}_{\text{Box}}$, $\mathcal{A}_{\text{Conv}}$ and their closure.

The first row is the most common case, using connected sets and their natural separation relation. The second row is the τ-connected case. The third row shows how the fine tuning can be handled: We consider connected base sets, but siblings need to be τ-separated, hence e.g. saddle points cannot be approximated.

The larger the extended class $\bar{\mathcal{A}}$ is, the more measures we can cluster. The following proposition provides a sufficient condition for $\bar{\mathcal{A}}$ being rich.

Proposition 26 *Assume all $A \in \mathcal{A}$ are path-connected. Then all $B \in \bar{\mathcal{A}}$ are path-connected. Furthermore assume that $\mathcal{A}$ is intersection-additive and that it contains a countable neighbourhood base. Then $\bar{\mathcal{A}}$ contains all open, path-connected sets.*

One can show that the first statement also holds for topological connectedness. Furthermore note that $\mathcal{C}_{\text{Dyad}}$ is a countable neighbourhood base, and therefore $\mathcal{A}_{\text{Dyad}}$, $\mathcal{A}_p$, and $\mathcal{A}_{\text{Conv}}$ fulfill the conditions of Proposition 26.

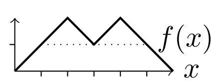
4.2 Clustering of Densities

Following the manual to cluster densities given in Theorem 23 by decomposing the density level sets into $\perp$-disjoint components, one first needs to understand the $\perp$-disjoint components of general events. In this subsection we investigate such decompositions and the resulting clusterings. We assume μ to be the Lebesgue measure on some suitable $\Omega \subset \mathbb{R}^d$ and let the base measures be the ones considered in Proposition 3. For visualization purposes we further restrict our considerations to the one- and two-dimensional case, only.

4.2.1 Dimension $d = 1$

In the one-dimensional case, in which Ω is an interval, the examples $\mathcal{A}_p = \mathcal{A}_{\text{Conv}}$ simply consist of compact intervals, and their monotone closures consist of all intervals. To understand the resulting clusters let us first consider the **twin peaks** density:

$$f(x) := \tfrac{1}{3} - \min\left\{ |x - \tfrac{1}{3}|, |x - \tfrac{2}{3}| \right\}.$$



Clearly, this gives the following $\perp_\emptyset$-decomposition of the level sets:

$$H_f(\lambda) = (\lambda, 1-\lambda) \text{ for } \lambda < \tfrac{1}{6}, \qquad H_f(\lambda) = (\lambda, \tfrac{1}{2} - \lambda) \overset{\perp_\emptyset}{\cup} (\tfrac{1}{2} + \lambda, \lambda) \text{ for } \tfrac{1}{6} \le \lambda < \tfrac{1}{3}$$

and hence the $\perp_\emptyset$-clustering forest is $\left\{ (0,1), (\frac{1}{6}, \frac{1}{2}), (\frac{1}{2}, \frac{5}{6}) \right\}$. Since, none of the boundary points can be reached, any isomonotone, adapted sequence yields this result. However, the clustering changes, if the separation relation $\perp_\tau$ is considered. We obtain

$$H_f(\lambda) = (\lambda, 1-\lambda), \text{ for } \lambda < \tfrac{1}{6} + \tfrac{\tau}{2}, \quad H_f(\lambda) = (\lambda, \tfrac{1}{2} - \lambda) \overset{\perp_\tau}{\cup} (\tfrac{1}{2} + \lambda, \lambda), \text{ for } \tfrac{1}{6} + \tfrac{\tau}{2} \le \lambda < \tfrac{1}{3}$$

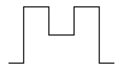
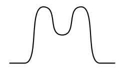
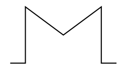
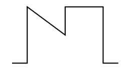
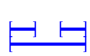
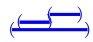
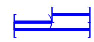
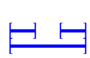
Name	Merlon	Camel	M	Factory
Density				
$(\mathcal{A}_p, \perp_\emptyset)$				
$(\mathcal{A}_p, \perp_\tau)$ with τ small				
$(\mathcal{A}_p, \perp_\tau)$ with τ large				

Table 1: Examples of clustering in dimension $d = 1$ using $\mathcal{A}_p$ and three separation relations.

if $\tau \in (0, \frac{1}{3})$ and the resulting $\perp_\tau$-clustering is $\{\, (0,1), (\frac{1}{6} + \frac{\tau}{2}, \frac{1}{2} - \frac{\tau}{2}), (\frac{1}{2} + \frac{\tau}{2}, \frac{5}{6} - \frac{\tau}{2}) \,\}$. Finally, if $\tau \geq \frac{1}{3}$ then all level sets are τ-connected and the forest is simply $\{(0,1)\}$. In Table 1 more examples of clustering of densities can be found.

4.2.2 Dimension $d = 2$

Our goal in this subsection is to understand the $\perp$-separated decomposition of closed events. We further present the resulting clusterings for some densities that are indicator functions and illustrate clusterings for continuous densities having a saddle point.

Let us begin by assuming that P has a Lebesgue density of the form 1_B, where B is some μ-support set. Then one can show, see Lemma 50 for details, that adapted, isomonotone sequences (F_n) of forests $F_n \uparrow B$ are of the form $F_n = \{A_1^n, \ldots, A_k^n\}$, where the elements of each forest F_n are mutually disjoint and can be ordered in such a way that $A_i^1 \subset A_i^2 \subset \ldots$. The limit forest F_∞ then consists of the k pairwise $\perp$-separated sets:

$$B_i := \bigcup_{n \geq 1} A_i^n \,,$$

and there is a μ-null set $N \in \mathcal{B}$ with

$$B = B_1 \overset{\perp}{\cup} \ldots \overset{\perp}{\cup} B_k \overset{\perp}{\cup} N. \tag{21}$$

Let us now consider the base sets $\mathcal{A}_p$ in Example 2. By Proposition 26 we know that $\bar{\mathcal{A}}_p$ contains all open, path-connected sets and therefore all open L^q-balls. Moreover, all closed L^q-balls B are μ-support sets with $\mu(\partial B) = 0$. Our initial consideration shows that 1_B can be approximated by an adapted, isomonotone sequence (F_n) of forests of the form $F_n = \{A^n\}$ with $A^n \in \mathcal{A}_p$. However, depending on p and q the μ-null set N in (21) may differ.

Now that we have an understanding of $\bar{\mathcal{A}}_p$ and adapted, isomonotone approximations we can investigate some more interesting cases and appreciate the influence of the choice of $\mathcal{A}$ on the outcome of the clustering in the following example.

Example 3 (Clustering of indicators) *We consider 6 examples of μ-support sets $B \in \mathbb{R}^2$. The first 4 have two parts that only intersect at one point, the second to last has two*

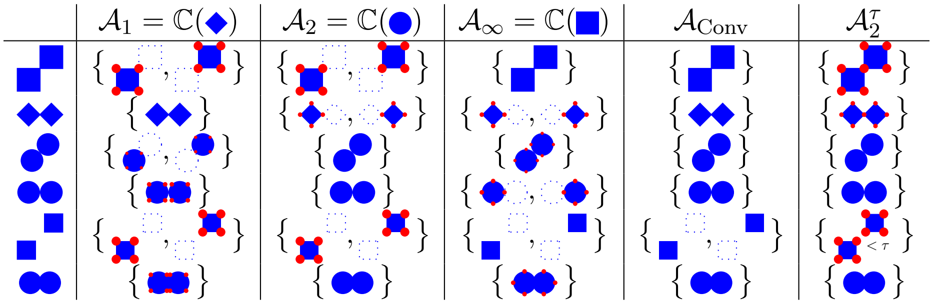


Table 2: Clusterings of indicators.

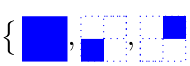
topological components, and the last one is connected in a fat way. By natural approximations we get the clusterings of Table 2. The red dots indicate points which never are achieved by any approximation. Observe how the geometry encoded in $\mathcal{A}$ shapes the clustering. Since $\mathcal{A}_{Conv}$ and $\mathcal{A}_2$ are invariant under rotation, they yield the same structure of clustering for rotated sets. The classes $\mathcal{A}_1$ and $\mathcal{A}_\infty$ on the other hand are not rotation-invariant and therefore the clustering depends on the orientation of B.

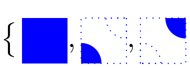
After having familiarized ourselves with the clustering of indicator functions we finally consider a continuous density that has a saddle point.

Example 4 *On $\Omega := [-1,1]^2$ consider the density $f : \Omega \to [0,2]$ given by $f(x,y) := x \cdot y + 1$. Then we have the following $\perp_\emptyset$-decomposition of the level sets $H_f(\lambda)$ of f:*

$$H_f(\lambda) = \begin{cases} \{(x,y) : xy > \lambda - 1\} & \text{if } \lambda \in [0,1), \\ [-1,0)^2 \,\dot\cup\, (0,1]^2 & \text{if } \lambda = 1, \\ \{(x,y) : x < 0 \text{ and } xy > \lambda - 1\} \,\dot\cup \{(x,y) : x > 0 \text{ and } xy > \lambda - 1\} & \text{if } \lambda \in (1,2). \end{cases}$$

For $(\mathcal{A}_p, \mathcal{Q}^{\mu,\mathcal{A}_p}, \perp_\emptyset)$ the clustering forest is therefore given by:

$$\{\, [-1,1]^2, [-1,0)^2, (0,1]^2 \,\} = \{$$  $$\}.$$

Moreover, for $(\mathcal{A}_2^\tau, \mathcal{Q}^{\mu,\mathcal{A}_2^\tau}, \perp_\tau)$ the clustering forest looks like $\{$  $\}$.

4.3 Hausdorff Measures

So far we have only considered clusterings of Lebesgue absolutely continuous distributions. In this subsection we provide some examples indicating that the developed theory goes far beyond this standard example. At first, lower dimensional base sets and their resulting clusterings are investigated. Afterwards we discuss collections of base sets of different dimensions and provide clusterings for some measures that are not absolutely continuous to any Hausdorff measure. For the sake of simplicity we will restrict our considerations to $\perp_\emptyset$-clusterings, but generalizations along the lines of the previous subsections are straightforward.

4.3.1 Lower Dimensional Base Sets

Let us begin by recalling that the s-dimensional Hausdorff-measure on $\mathcal{B}$ is defined by

$$\mathcal{H}^s(B) := \lim_{\varepsilon\to 0} \inf \Big\{ \sum_{i=1}^{\infty} (\operatorname{diam}(B_i))^s \mid B \subset \bigcup_i B_i \text{ and } \forall i \in \mathbb{N}\colon \operatorname{diam}(B_i) \le \varepsilon \Big\}.$$

Moreover, the Hausdorff-dimension of a $B \in \mathcal{B}$ is the value $s \in [0, d]$ at which $s \mapsto \mathcal{H}^s(B)$ jumps from ∞ to 0. If B has Hausdorff-dimension s, then $\mathcal{H}^s(B)$ can be either zero, finite, or infinite. Hausdorff-measures are inner regular (Federer, 1969, Cor. 2.10.23) and $\mathcal{H}^d$ equals the Lebesgue-measure up to a normalization factor. For a reference on Hausdorff-dimensions and -measures we refer to Falconer (1993) and Federer (1969). Recall that given a Borel set $C \subset \mathbb{R}^s$ a map $\varphi\colon C \to \Omega$ is **bi-Lipschitz** iff there are constants $0 < c_1, c_2 < \infty$ s.t.

$$c_1 d(x, y) \le d(\varphi(x), \varphi(y)) \le c_2 d(x, y).$$

Lemma 27 *If C is a Lebesgue-support set in $\mathbb{R}^s$ and $\varphi\colon C \to \Omega$ is bi-Lipschitz then $C' := \varphi(C)$ has Hausdorff-dimension s and it is an $\mathcal{H}^s$-support set in Ω.*

Motivated by Lemma 27, consider the following collection of s-dimensional base sets in Ω:

$$\mathcal{C}_{p,s} := \Big\{ \varphi(C) \subset \Omega \mid C \text{ is the closed unit } p\text{-ball in } \mathbb{R}^s \text{ and } \varphi\colon C \to \Omega \text{ is bi-Lipschitz } \Big\}.$$

Using the notation of Definition 25 and Proposition 3 we further write

$$\mathcal{A}_{p,s} := \mathbb{C}_{\perp_\emptyset}(\mathcal{C}_{p,s}) \qquad \text{and} \qquad \mathcal{Q}^{p,s} := \mathcal{Q}^{\mathcal{H}^s, \mathcal{A}_{p,s}}.$$

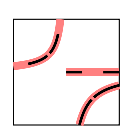
By $\mathcal{A}_0 := \big\{ \{x\} \mid x \in \Omega \big\}$ we denote the singletons and $\mathcal{Q}_0$ the collection of Dirac measures. Since continuous mappings of connected sets are connected, $(\mathcal{A}_{p,s}, \mathcal{Q}^{p,s}, \perp_\emptyset)$ is a stable $\perp_\emptyset$-additive clustering base. Remark that we take the union after embedding into $\mathbb{R}^d$ and therefore also crossings do happen, e.g. the cross $[-1, 1] \times \{0\} \cup \{0\} \times [-1, 1] \in \mathcal{A}_{p,1}$. Another possibility would be to embed $\mathcal{A}_p$ via a set of transformations into $\mathbb{R}^d$. Finally we confine the examples here only to integer Hausdorff-dimensions—it would be interesting though to consider e.g. the Cantor set or the Sierpinski triangle. The following example presents a resulting clustering of an $\mathcal{H}^1$-absolutely continuous measure on $\mathbb{R}^2$.

Example 5 (Measures supported on curves in the plane)
On $\Omega := [-1, 1]^2$ consider the measure $P_1 := f\, d\mathcal{H}^1$ whose density is given by

$$f(x, y) := \begin{cases} f_{Merlon}(x) & \text{if } x \ge 0 \text{ and } y = 0, \\ f_{Camel}(t) & \text{if } x = -3^{2t-2} \text{ and } y = 3^{-2t}, \\ f_M(t) & \text{if } x = 2^{2t-2} \text{ and } y = -2^{-2t}. \end{cases}$$

Here the densities and clusterings for the Merlon, the Camel and the M can be seen in Table 1. So for $(\mathcal{A}_{p,1}, \mathcal{Q}^{p,1}, \perp_\emptyset)$ with any fixed $p \ge 1$ the clustering forest of P_1 is given by:

$$c(P_1) = \left\{ \begin{array}{c} [0, 1] \times \{0\}, [0, \tfrac{1}{3}] \times \{0\}, [\tfrac{2}{3}, 1] \times \{0\}, \\ g_1\big((0, 1)\big), g_1\big((0.2, 0.5)\big), g_1\big((0.5, 0.8)\big), \\ g_2\big([0, 1]\big), g_2\big([0, 0.5)\big), g_2\big((0.5, 1]\big) \end{array} \right\}$$



where $g_i\colon [0, 1] \to \Omega$ are given by $g_1(t) = (-3^{2t-2}, 3^{-2t})$ and $g_2(t) = (2^{2t-2}, -2^{-2t})$.

4.3.2 Heterogeneous Hausdorff-Dimensions

In this subsection we consider measures that can be decomposed into measures that are absolutely continuous with respect to Hausdorff measures of different dimensions. To this end, we write $\mu \prec \mu'$ for two measures μ and μ' on $\mathcal{B}$, iff for all $B \in \mathcal{B}$ with $B \subset \operatorname{supp} \mu \cap \operatorname{supp} \mu'$ we have

$$\mu(B) < \infty \implies \mu'(B) = 0.$$

For $\mathcal{Q}, \mathcal{Q}' \subset \mathcal{M}_\Omega$ we further write $\mathcal{Q} \prec \mathcal{Q}'$ if $\mu \prec \mu'$ for all $\mu \in \mathcal{Q}$ and $\mu' \in \mathcal{Q}'$. Clearly, the relation $\prec$ is transitive. Moreover, we have $\mathcal{H}^s \prec \mathcal{H}^t$ whenever $s < t$. The next proposition shows that clustering bases whose base measures dominate each other in the sense of $\prec$ can be merged.

Proposition 28 *Let $(\mathcal{A}^1, \mathcal{Q}^1, \perp), \dots, (\mathcal{A}^m, \mathcal{Q}^m, \perp)$ be stable clustering bases sharing the same separation relation $\perp$ and assume $\mathcal{Q}^1 \prec \cdots \prec \mathcal{Q}^m$. We define*

$$\mathcal{A} := \bigcup_i \mathcal{A}^i \qquad \text{and} \qquad \mathcal{Q} := \bigcup_i \mathcal{Q}^i.$$

Then $(\mathcal{A}, \mathcal{Q}, \perp)$ is a stable clustering base.

Proposition 28 shows that the $\perp_\emptyset$-additive, stable bases $(\mathcal{A}_{p,s}, \mathcal{Q}^{p,s}, \perp_\emptyset)$ on $\mathbb{R}^d$ can be merged. Unfortunately, however, its union is no longer $\perp_\emptyset$-additive, and therefore we need to investigate P-subadditivity in order to describe distributions for which our theory provides a clustering. This is done in the next proposition.

Proposition 29 *Let $(\mathcal{A}^1, \mathcal{Q}^1, \perp)$ and $(\mathcal{A}^2, \mathcal{Q}^2, \perp)$ be clustering bases with $\mathcal{Q}^1 \prec \mathcal{Q}^2$ and P_1 and P_2 be finite measures with $P_1 \prec \mathcal{A}^2$ and $\mathcal{A}^1 \prec P_2$. Furthermore, assume that $\mathcal{A}^i$ is P_i-subadditive for both $i = 1, 2$ and let $P := P_1 + P_2$. Then we have*

- *(a) For $i = 1, 2$ and all base measures $\mathfrak{a} \in \mathcal{Q}^i_P$ we have $\mathfrak{a} \le P_i$,*
- *(b) If for all base measures $\mathfrak{a} \in \mathcal{Q}^2_{P_2}$ and $\operatorname{supp} P_1 \between \operatorname{supp} \mathfrak{a}$ there exists a base measure $\tilde{\mathfrak{a}} \in \mathcal{Q}^2_{P_2}(\operatorname{supp} P_1)$ with $\mathfrak{a} \le \tilde{\mathfrak{a}}$ then $\mathcal{A}^1 \cup \mathcal{A}^2$ is P-subadditive.*

To illustrate condition *(b)* consider clustering bases $(\mathcal{A}_{p,s}, \mathcal{Q}^{p,s}, \perp_\emptyset)$ and $(\mathcal{A}_{p,t}, \mathcal{Q}^{p,t}, \perp_\emptyset)$ for some $s < t$. The condition specifies that any such base measure $\mathfrak{a}$ intersecting $\operatorname{supp} P_1$ can be majorized by one which supports $\operatorname{supp} P_1$. Then all parts of $\operatorname{supp} P_1$ intersecting at least one component of $\operatorname{supp} P_2$ have to be on the same niveau line of P_2. Note that this is trivially satisfied if the $\operatorname{supp} P_1 \cap \operatorname{supp} P_1 = \emptyset$. Recall that mixtures of the latter form have already been clustered in Rinaldo and Wasserman (2010) by a kernel smoothing approach. Clearly, our axiomatic approach makes it possible to define clusterings for significantly more involved distributions as the following two examples demonstrate.

Example 6 (Mixture of atoms and Lebesgue measure)
Consider $\Omega = \mathbb{R}$. Let $(\mathcal{A}_0, \mathcal{Q}_0, \perp_\emptyset)$ be the singletons with Dirac measures and consider for

any fixed $p \geq 1$ *the clustering base* $(\mathcal{A}_{p,s}, \mathcal{Q}_{p,s}, \perp_\emptyset)$. *Both are* $\infty_\emptyset$*-additive and stable and we have* $\mathcal{Q}_0 \prec \mathcal{Q}_{p,1}$. *Now consider the measures*

$$P_0 := \delta_0 + 2\delta_1 + \delta_2 \qquad \text{and} \qquad P_1(dx) := \sin^2(\tfrac{2x}{\pi})\,\mathcal{H}_1(dx).$$

Then the assumptions of Proposition 29 are satisfied and the clustering of $P := P_0 + P_1$ *is given by*

$$c(P) = c(P_0) \cup c(P_1) = \left\{\, \{0\}, (0, \tfrac{1}{2}), (\tfrac{1}{2}, 1), \{1\}, \{2\} \,\right\}.$$

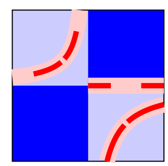
Our last example combines Examples 4 and 5.

Example 7 (Mixtures in dimension 2) *Consider* $\Omega := [-1, 1]^2$ *and the densities* f_1 *and* f_2 *introduced in Examples 5 and 4, respectively. Furthermore, consider the measures*

$$P_2 := f_2\, d\mathcal{H}^2, \qquad\qquad P_1 := f_1\, d\mathcal{H}^1$$

and the clustering bases $(\mathcal{A}_{p,1}, \mathcal{Q}^{p,1}, \perp_\emptyset)$ *and* $(\mathcal{A}_{p',2}, \mathcal{Q}^{p',2}, \perp_\emptyset)$ *for some fixed* $p, p' \geq 1$. *As above* $\mathcal{Q}^{p,1} \prec \mathcal{Q}^{p',2}$. *And by Proposition 29 the clustering forest of* $P = P_1 + P_2$ *is given by*

$$c_1(P_1) \cup c_2(P_2) = \left\{ \begin{array}{c} [0,1] \times \{0\}, [0, \frac{1}{3}] \times \{0\}, [\frac{2}{3}, 1] \times \{0\}, \\ g_1\big((0,1)\big), g_1\big((0.2, 0.5)\big), g_1\big((0.5, 0.8)\big), \\ g_2\big([0,1]\big), g_2\big([0, 0.5)\big), g_2\big((0.5, 1]\big), \\ [-1,1]^2, [-1,0)^2, (0,1]^2 \end{array} \right\}$$



where $g_i \colon [0,1] \to \Omega$ *are given by* $g_1(t) = (-3^{2t-2}, 3^{-2t})$ *and* $g_2(t) = (2^{2t-2}, -2^{-2t})$. *Observe that* g_1 *and* g_2 *lie on niveau lines of* f_2.

5. Proofs

5.1 Proofs for Section 2

We begin with some simple properties of separation relations.

Lemma 30 *Let* $\perp$ *be an* $\mathcal{A}$*-separation relation. Then the following statements are true:*

(a) *For all* $B, B' \in \mathcal{B}$ *with* $B \perp B'$ *we have* $B \cap B' = \emptyset$.

(b) *Suppose that* $\perp$ *is stable and* $(A_i)_{i \geq 1} \subset \mathcal{A}$ *is increasing. For* $A := \bigcup_n A_n$ *and all* $B \in \mathcal{B}$ *we then have*

$$A_n \perp B \quad \text{for all } n \geq 1 \iff A \perp B$$

(c) *Let* $A \in \mathcal{A}$ *and* $B_1, \dots, B_k \in \mathcal{B}$ *be closed. Then:*

$$A \subset B_1 \overset{\perp}{\cup} \dots \overset{\perp}{\cup} B_k \implies \exists! i \leq k \colon A \subset B_i$$

(d) *For all* $A_1, \dots, A_k \in \mathcal{A}$ *and all* $A'_1, \dots, A'_{k'} \in \mathcal{A}$, *we have*

$$A_1 \overset{\perp}{\cup} \dots \overset{\perp}{\cup} A_k = A'_1 \overset{\perp}{\cup} \dots \overset{\perp}{\cup} A'_{k'} \implies \{A_1, \dots, A_k\} = \{A'_1, \dots, A'_{k'}\}.$$

Proof of Lemma 30: *(a).* Let us write $B_0 := B \cap B'$. Monotonicity and $B \perp B'$ implies $B_0 \perp B'$ and thus $B' \perp B_0$ by symmetry. Another application of the monotonicity gives $B_0 \perp B_0$ and the reflexivity thus shows $B \cap B' = B_0 = \emptyset$.
(b). "$\Rightarrow$" is stability and "$\Leftarrow$" follows from monotonicity.
(c). Existence of such an i is $\mathcal{A}$-connectedness. Now assume that there is an $j \neq i$ with $A \subset B_j$. Then $\emptyset \neq A \subset B_i \cap B_j$ contradicting $B_i \perp B_j$ by (a).
(d). We write $F := \{A_1, \ldots, A_k\}$ and $F' := \{A'_1, \ldots, A'_{k'}\}$. By (c) we find an injection $I\colon F \to F'$ such that $A \subset I(A)$ and hence $k \leq k'$. Analogously, we find an injection $J\colon F' \to F$ such that $A \subset J(A)$, and we get $k = k'$. Consequently, I and J are bijections. Let us now fix an $A_i \in F$. For $A_j := J \circ I(A_i) \in F$ we then find $A_i \subset I(A_i) \subset J(I(A_i)) = A_j$. This implies $i = j$, since otherwise $A_i \subset A_j$ would contradict $A_i \perp A_j$ by (a). Therefore we find $A_i = I(A_i)$ and the bijectivity of I thus yields the assertion. ∎

Proof of Proposition 3: We first need to check that the support is defined for all restrictions $\mu_{|C} := \mu(\cdot \cap C)$ to sets $C \in \mathcal{B}$ that satisfy $0 < \mu(C) < \infty$. To this end, we check that $\mu_{|C}$ is inner regular: If Ω is a Radon space then there is nothing to prove since $\mu_{|C}$ is a finite measure. If Ω is not a Radon space, then the definition of $\mathcal{M}_\Omega^\infty$ guarantees that μ is inner regular and hence $\mu_{|C}$ is inner regular by Lemma 51.

Let us now verify that $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ is a (stable) clustering base. To this end, we first observe that each $Q_A \in \mathcal{Q}^{\mu,\mathcal{A}}$ is a probability measure by construction and since we have already seen that $\mu_{|C}$ is inner regular for all $C \in \mathcal{K}(\mu)$ we conclude that $\mathcal{Q}^{\mu,\mathcal{A}} \subset \mathcal{M}$. Moreover, fittedness follows from $\mathcal{A} \subset \mathcal{K}(\mu)$. For flatness let $A, A' \in \mathcal{A}$ with $A \subset A'$ and $Q_{A'}(A) \neq 0$. Then for all $B \in \mathcal{B}$ we have

$$Q_A(B) = \frac{\mu(B \cap A)}{\mu(A)} = \frac{\mu(B \cap A \cap A')}{\mu(A \mid A') \cdot \mu(A')} = \frac{\mu(B \cap A \mid A')}{\mu(A \mid A')} = \frac{Q_{A'}(B \cap A)}{Q_{A'}(A)}. \qquad ∎$$

Proof of Lemma 7: Let $Q = \sum_{A \in F} \alpha_A Q_A$ and $Q = \sum_{A' \in F'} \alpha'_{A'} Q_{A'}$ be two representations of $Q \in \mathcal{Q}$. By part (d) of Lemma 51 we then obtain

$$\operatorname{supp} Q = \operatorname{supp}\left(\sum_{A \in F} \alpha_A Q_A\right) = \bigcup_{A \in F} \operatorname{supp} Q_A = \bigcup_{A \in F} A\,, = \mathbb{G}F$$

and since we analogously find $\operatorname{supp} Q = \mathbb{G}F'$, we conclude that $\mathbb{G}F = \mathbb{G}F'$. The latter together with Lemma 30 gives $\max F = \max F'$. To show that $\alpha_A = \alpha'_A$ for all roots $A \in \max\ F = \max F'$, we pick a root $A \in \max\ F$ and assume that $\alpha_A < \alpha'_A$. Now, if A has no direct child, we set $B := A$. Otherwise we define $B := A \setminus (A_1 \cup \ldots \cup A_k)$, where the A_k are the direct children of A in F. Because of the definition of a direct child and part (d) of Lemma 30 we find $A_1 \cup \ldots \cup A_k \subsetneq A$ in the second case. In both cases we conclude that B is non-empty and relatively open in $A = \operatorname{supp} Q_A$ and by Lemma 51 we obtain $Q_A(B) > 0$. Consequently, our assumption $\alpha_A < \alpha'_A$ yields $\alpha_A Q_A(B) < \alpha'_A Q_A(B) \leq Q(B)$. However, our construction also gives

$$Q(B) = \sum_{A'' \in F} \alpha_{A''} Q_{A''}(B) = \alpha_A Q_A(B) + \sum_{A'' \subsetneq A} \alpha_{A''} Q_{A''}(B) + \sum_{A'' \perp A} \alpha_{A''} Q_{A''}(B) = \alpha_A Q_A(B)\,,$$

i.e. we have found a contradiction. Summing up, we already know that $\max\ F = \max F'$ and $\alpha_A = \alpha'_A$ for all $A \in \max\ F$. This yields

$$\sum_{A\in\max F} \alpha_A Q_A = \sum_{A'\in\max F'} \alpha'_{A'} Q_{A'}\,.$$

Eliminating the roots gives the forests $F_1 := F \setminus \max F$ and $F'_1 := F' \setminus \max F'$ and

$$Q_1 := \sum_{A\in F_1} \alpha_A Q_A = Q - \sum_{A\in\max F} \alpha_A Q_A = Q - \sum_{A'\in\max F'} \alpha'_{A'} Q_{A'} = \sum_{A'\in F'_1} \alpha'_{A'} Q_{A'}\,,$$

i.e. Q_1 has two representations based upon the reduced forests F_1 and F'_1. Applying the argument above recursively thus yields $F = F'$ and $\alpha'_A = \alpha'_A$ for all $A \in F$. ■

Proof of Theorem 8: We first show that (16) defines an additive clustering. Since Axiom 1 is obviously satisfied, it suffices to check the two additivity axioms for $\mathcal{P} := \mathcal{S}(\mathcal{A})$. We begin by establishing DisjointAdditivity. To this end, we pick $Q_1, \dots, Q_k \in \mathcal{S}(\mathcal{A})$ with representing $\perp$-forests F_i such that $\operatorname{supp} Q_i = \mathbb{G}F_i$ are mutually $\perp$-separated. For $A \in \max F_i$ and $A' \in \max F_j$ with $i \neq j$, we then have $A \perp A'$, and therefore

$$F := F_1 \cup \ldots \cup F_k$$

is the representing $\perp$-forest of $Q := Q_1 + \ldots + Q_k$. This gives $Q \in \mathcal{S}(\mathcal{A})$ and

$$c(Q) = s(F) = s(F_1) \cup \cdots \cup s(F_k) = c(Q_1) \cup \cdots \cup c(Q_k).$$

To check BaseAdditivity we fix a $Q \in \mathcal{S}(\mathcal{A})$ with representing $\perp$-forest F and a base measure $\mathfrak{a} = \alpha Q_A$ with $\operatorname{supp} Q \subset \operatorname{supp} \mathfrak{a}$. For all $A' \in F$ we then have $A' \subset \mathbb{G}F = \operatorname{supp} Q \subset A$ and therefore $F' := \{A\} \cup F$ is the representing $\perp$-forest of $\mathfrak{a} + Q$. This yields $\mathfrak{a} + Q \in \mathcal{S}(\mathcal{A})$ and

$$c(\mathfrak{a} + Q) = s(F') = s\big(\{A\} \cup F\big) = s\big(\operatorname{supp} \mathfrak{a} \cup c(Q)\big).$$

Let us now show that every additive $\mathcal{A}$-clustering $c : \mathcal{P} \to \mathcal{F}$ satisfies both $\mathcal{S}(\mathcal{A}) \subset \mathcal{P}$ and (16). To this end we pick a $Q \in \mathcal{S}(\mathcal{A})$ with representing forest F and show by induction over $|F| = n$ that both $Q \in \mathcal{P}$ and $c(Q) = s(F)$. Clearly, for $n = 1$ this immediately follows from Axiom 1. For the induction step we assume that for some $n \geq 2$ we have already proved $Q' \in \mathcal{P}$ and $c(Q') = s(F')$ for all $Q' \in \mathcal{S}(\mathcal{A})$ with representing forest F' of size $|F'| < n$.

Let us first consider the case in which F is a tree. Let A be its root and α_A be corresponding coefficient in the representation of Q. Then $Q' := Q - \alpha_A Q_A$ is a simple measure with representing forest $F' := F \setminus A$ and since $|F'| = n-1$ we know $Q' \in \mathcal{P}$ and $c(Q') = s(F')$ by the induction assumption. By the axiom of BaseAdditivity we conclude that

$$c(Q) = c(\alpha_A Q_A + Q') = s(\{A\} \cup c(Q')) = s(\{A\} \cup F') = s(F)\,,$$

where the last equality follows from the assumption that F is a tree with root A.

Now consider the case where F is a forest with $k \geq 2$ roots $A_1, \dots, A_k$. For $i \leq k$ we define $Q_i := Q\big|_{\subset A_i}$. Then all Q_i are simple measures with representing forests $F_i := F\big|_{\subset A_i}$ and we have $Q = Q_1 + \cdots + Q_k$. Therefore, the induction assumption guarantees $Q_i \in \mathcal{P}$ and $c(Q_i) = s(F_i)$. Since $\operatorname{supp} Q_i = A_i$ and $A_i \perp A_j$ whenever $i \neq j$, the axiom of DisjointAdditivity then shows $Q \in \mathcal{P}$ and

$$c(Q) = c(Q_1) \cup \cdots \cup c(Q_k) = s(F_1) \cup \cdots \cup s(F_k) = s(F)\,. \quad ■$$

5.2 Proofs for Section 3

Proof of Lemma 11: For the first assertion it suffices to check $\bar{\mathcal{A}}$-connectedness. To this end, we fix an $A \in \bar{\mathcal{A}}$ and closed sets $B_1, \dots, B_k$ with $A \subset B_1 \overset{\perp}{\cup} \dots \overset{\perp}{\cup} B_k$. Let $(A_n) \subset \mathcal{A}$ with $A_n \nearrow A$. For all $n \geq 1$ part (c) of Lemma 30 then gives exactly one $i(n)$ with $A_n \subset B_{i(n)}$. This uniqueness together with $A_n \subset A_{n+1}$ yields $i(1) = i(2) = \dots$ and hence $A_n \subset B_{i(1)}$ for all n. We conclude that $A \subset B_{i(1)}$ by part (b) of Lemma 30.

For the second assertion we pick an isomonotone sequence $(F_n) \subset \mathcal{F}_{\mathcal{A}}$ and define $F_\infty := \lim_n s(F_n)$. Let us first show that F_∞ is a $\perp$-forest. To this end, we pick $A, A' \in F_\infty$. By the construction of F_∞ there then exist $A_1, A_1' \in s(F_1)$ such that for $A_n := \zeta_n(A_1)$ and $A_n' := \zeta_n(A_1')$ we have $A_n \nearrow A$ and $A_n' \nearrow A'$ Now, if $A_1 \perp A_1'$ then $A_n \perp A_n'$ and thus $A_m \perp A_n'$ for all m, n by isomonotonicity. Using the stability of $\perp$ twice we first obtain $A \perp A_n'$ for all n and then $A \perp A'$. If $A_1 \not\perp A_1'$, we may assume $A_1 \subset A_1'$ since $s(F_1)$ is a $\perp$-forest. Isomonotonicity implies $A_n \subset A_n' \subset A'$ for all n and hence $A \subset A'$. Finally, $s(F_n) \leq F_\infty$ is trivial. ■

Proof of Proposition 16: We first show that $\mathcal{A}$ is P-subadditive if $\mathcal{A}$ is ∞_P-additive. To this end we fix $A, A' \in \mathcal{A}$ with $A \infty_P A'$. Since $\mathcal{A}$ is ∞_P-additive we find $B := A \cup A' \in \mathcal{A}$. This yields

$$Q_B(A) = \frac{\mu(A \cap B)}{\mu(B)} = \frac{\mu(A)}{\mu(B)} > 0$$

and analogously we obtain $Q_B(A') > 0$. For $\alpha Q_A, \alpha' Q_{A'} \leq P$ we can therefore assume that $\beta := \frac{\alpha}{Q_B(A)} < \frac{\alpha'}{Q_B(A')}$. Setting $\mathfrak{b} := \beta Q_B$ we now obtain by the flatness assumption

$$\alpha Q_A(\cdot) = \alpha \cdot \frac{Q_B(\cdot \cap A)}{Q_B(A)} = \mathfrak{b}(\cdot \cap A) \leq \mathfrak{b}(\cdot).$$

Now assume that $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$ is P-subadditive for all $P \ll \mu$. Let $A, A' \in \mathcal{A}$ with $A \infty_\mu A'$. Then we have $P := Q_A + Q_A' \ll \mu$ and $Q_A, Q_{A'} \leq P$. Since $\mathcal{A}$ is P-subadditive there is a base measure $\mathfrak{b} \leq P$ with $A \cup A' \subset \operatorname{supp} \mathfrak{b} \subset \operatorname{supp} P = A \cup A'$ by Lemma 51. Consequently we obtain $A \cup A' = \operatorname{supp} \mathfrak{b} \in \mathcal{A}$. ■

Lemma 31 *Let $P \in \mathcal{M}$ and $(\mathcal{A}, \mathcal{Q}, \perp)$ be a P-subadditive clustering base. Then the kinship relation $\sim_P$ is a symmetric and transitive relation on $\{B \in \mathcal{B} \mid P(B) > 0\}$ and an equivalence relation on the set $\{A \in \mathcal{A} \mid \exists \alpha > 0$ such that $\alpha Q_A \leq P\}$. Finally, for all finite sequences $A_1, \dots, A_k \in \mathcal{A}$ of sets that are pairwise kin below P there is $\mathfrak{b} \in \mathcal{Q}_P(A_1 \cup \dots \cup A_k)$.*

Proof of Lemma 31: Symmetry is clear. Let $B_1 \sim_P B_2$ and $B_2 \sim_P B_3$ be events with $P(B_i) > 0$. Then there are base measures $\mathfrak{c} = \gamma Q_C \in \mathcal{Q}_P(B_1 \cup B_2)$ and $\mathfrak{c}' = \gamma' Q_{C'} \in \mathcal{Q}_P(B_2 \cup B_3)$ supporting them. This yields $B_2 \subset C \cap C'$ and thus $P(C \cap C') \geq P(B_2) > 0$. In other words, we have $C \infty_P C'$, and by subadditivity we conclude that there is a $\mathfrak{b} \in \mathcal{Q}_P(C \cup C')$. This gives $B_1 \cup B_3 \subset C \cup C' \subset \operatorname{supp} \mathfrak{b}$, and therefore $B_1 \sim_P B_3$ at $\mathfrak{b}$. To show reflexivity on the specified subset of $\mathcal{A}$, we fix an $A \in \mathcal{A}$ and an $\alpha > 0$ such that $\mathfrak{a} := \alpha Q_A \leq P$. Then we have $\mathfrak{a} \in \alpha Q_P(\mathcal{A})$ and hence we obtain $A \sim_P A$.

The last statement follows by induction over k, where the initial step $k = 2$ is simply the definition of kinship. Let us therefore assume the statement is true for some $k \geq 2$. Let

$A_1, \ldots, A_{k+1} \in \mathcal{A}$ be pairwise kin. By assumption there is a $\mathfrak{b} \in \mathcal{Q}_P(A_1 \cup \ldots \cup A_k)$. Since this latter yields $A_1 \subset \operatorname{supp} \mathfrak{b}$ we find $A_1 \sim_P \operatorname{supp} \mathfrak{b}$ and by transitivity of $\sim_P$ we hence have $A_{k+1} \sim_P \operatorname{supp} \mathfrak{b}$. By definition there is thus a $\tilde{\mathfrak{b}} \in \mathcal{Q}_P(A_{k+1} \cup \operatorname{supp} \mathfrak{b})$ and since this gives $A_1 \cup \ldots \cup A_{k+1} \subset A_{k+1} \cup \operatorname{supp} \mathfrak{b} \subset \operatorname{supp} \tilde{\mathfrak{b}}$ we find $\tilde{\mathfrak{b}} \in \mathcal{Q}_P(A_1 \cup \ldots \cup A_{k+1})$. ■

Lemma 32 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $Q \in \mathcal{S}(\mathcal{A})$ with representing forest $F \in \mathcal{F}_{\mathcal{A}}$. Then for all $A \in F$ we have*

$$Q(\,\cdot\, \cap A) = \lambda_Q(A) + Q\big|_{\subsetneq A} .$$

Proof of Lemma 32: Let $A_0 \in \max F$ be the root with $A \subset A_0$. Then we can decompose F into $F = \{A' \in F : A' \supset A\} \,\dot\cup\, \{A' \in F : A' \subsetneq A\} \,\dot\cup\, \{A' \in F : A' \perp A\}$. Moreover, flatness of $\mathcal{Q}$ gives $Q_{A'}(\cdot \cap A) = Q_{A'}(A) \cdot Q_A(\cdot)$ for all $A' \in \mathcal{A}$ with $A \subset A'$ while fittedness gives $Q_{A'}(A) = 0$ for all $A' \in \mathcal{A}$ with $A' \perp A_0$ by the monotonicity of $\perp$, part (a) of Lemma 30, and part (b) of Lemma 51. For $B \in \mathcal{B}$ we thus have

$$\begin{aligned} Q(B \cap A) &= \sum_{A' \supset A} \alpha_{A'} Q_{A'}(B \cap A) + \sum_{A' \subsetneq A} \alpha_{A'} Q_{A'}(B \cap A) + \sum_{A' \perp A_0} \alpha_{A'} Q_{A'}(B \cap A) \\ &= \sum_{A' \supset A} \alpha_{A'} Q_{A'}(A) Q_A(B) + \sum_{A' \subsetneq A} \alpha_{A'} Q_{A'}(B \cap A) \\ &= \lambda_Q(A)(B) + Q\big|_{\subsetneq A}(B) , \end{aligned}$$

where the last step uses $Q_{A'}(B \cap A) = Q_{A'}(B)$ for $A' \subset A$, which follows from fittedness. ■

Lemma 33 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $\mathfrak{a}, \mathfrak{b}$ be base measures on $A, B \in \mathcal{A}$ with $A \subset B$. Then for all $C_0 \in \mathcal{B}$ with $\mathfrak{a}(C_0 \cap A) > 0$ we have*

$$\mathfrak{b}(\cdot \cap A) = \frac{\mathfrak{b}(C_0 \cap A)}{\mathfrak{a}(C_0 \cap A)} \cdot \mathfrak{a}(\cdot \cap A) .$$

Proof of Lemma 33: By assumption there are $\alpha, \beta > 0$ with $\mathfrak{a} = \alpha Q_A$ and $\mathfrak{b} = \beta Q_B$. Moreover, flatness guarantees $Q_B(\cdot \cap A) = Q_B(A) \cdot Q_A(\cdot)$. For all $C \in \mathcal{B}$ we thus obtain

$$\mathfrak{b}(C \cap A) = \beta Q_B(C \cap A) = \beta Q_B(A) \cdot Q_A(C) = \beta Q_B(A) \cdot Q_A(C \cap A) = \frac{\beta Q_B(A)}{\alpha} \mathfrak{a}(C \cap A) .$$

where in the second to last step we used $Q_A(\cdot) = Q_A(\cdot \cap A)$, which follows from $A = \operatorname{supp} Q_A$. For $C_0 \in \mathcal{B}$ with $\mathfrak{a}(C_0 \cap A) > 0$ we thus find $\frac{\beta Q_B(A)}{\alpha} = \frac{\mathfrak{b}(C_0 \cap A)}{\mathfrak{a}(C_0 \cap A)}$ and inserting this in the previous formula gives the assertion. ■

Lemma 34 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $Q \in \mathcal{S}(\mathcal{A})$ be a simple measure, $\mathfrak{a}$ be a base measures on some $A \in \mathcal{A}$, and $C \in \mathcal{B}$. Then the following statements are true:*

(a) If $\mathfrak{a} \le Q$ then there is a level $\mathfrak{b}$ in Q with $\mathfrak{a} \le \mathfrak{b}$.

(b) If $\mathfrak{b}(\cdot \cap C) \le \mathfrak{a}(\cdot \cap C)$ for all levels $\mathfrak{b}$ of Q then $Q(C) \le \mathfrak{a}(C)$.

(c) For all $P \in \mathcal{M}$ we have $Q \leq P$ if and only if $\mathfrak{b} \leq P$ for all levels $\mathfrak{b}$ in Q.

Proof of Lemma 34: In the following we denote the representing forest of Q by F.
(a). By $\mathfrak{a} \leq Q$ we find $A \subset \operatorname{supp} Q = \mathbb{G}F$. Since the roots $\max F$ form a finite $\perp$-disjoint union of closed sets of $\mathbb{G}F$, the $\mathcal{A}$-connectedness shows that A is already contained in one of the roots, say $A_0 \in \max F$. For $F' := \{A' \in F \mid A \subset A'\}$ we thus have $A_0 \in F'$. Moreover, F' is a chain, since if there were $\perp$-disjoint $A', A'' \in F'$ then A would only be contained in one of them by Lemma 30. Therefore there is a unique leaf $B := \min F' \in F$ and thus $A \subset B$. We denote the level of B in Q by $\mathfrak{b}$. Then it suffices to show $\mathfrak{a} \leq \mathfrak{b}$. To this end, let $\{C_1, \dots, C_k\} = \max F\big|_{\subsetneqq B}$ be the direct children of B in F. By construction we know $A \not\subset C_i$ for all $i = 1, \dots, k$ and hence $\mathcal{A}$-connectedness yields $A \not\subset C_1 \overset{\perp}{\cup} \dots \overset{\perp}{\cup} C_k$. Therefore $C_0 := A \setminus \bigcup_i C_i$ is non-empty and relatively open in $A = \operatorname{supp} Q_A$. This gives $\mathfrak{a}(C_0 \cap A) > 0$ by Lemma 51. Let us write $\mathfrak{b} := \lambda_Q(B)$ for the level of B in Q. Lemma 32 applied to the node $B \in F$ then gives

$$Q(C_0) = \mathfrak{b}(C_0) + Q\big|_{\subsetneqq B}(C) = \mathfrak{b}(C_0) + \sum_{A' \in F : A' \subsetneqq B} \alpha_{A'} Q_{A'}(C_0) = \mathfrak{b}(C_0)$$

since for $A' \in F$ with $A' \subsetneqq B$ we have $A' \subset \bigcup_i C_i$ and thus $\operatorname{supp} Q_{A'} \cap C_0 = A' \cap C_0 = \emptyset$. Therefore, we find $\mathfrak{a}(C_0 \cap A) = \mathfrak{a}(C_0) \leq Q(C_0) = \mathfrak{b}(C_0) = \mathfrak{b}(C_0 \cap B)$. By Lemma 33 we conclude that $\mathfrak{b}(\cdot \cap A) \geq \mathfrak{a}(\cdot \cap A)$. For $B' \in \mathcal{B}$ the decomposition $B' = (B' \setminus A) \dot\cup (B' \cap A)$ and the fact that $A = \operatorname{supp} \mathfrak{a} \subset \operatorname{supp} \mathfrak{b}$ then yields the assertion.
(b). For $A \in F$ we define

$$B_A := A \setminus \bigcup_{A' \in F :\, A' \subsetneqq A} A',$$

i.e. B_A is obtained by removing the strict descendants from A. From this description it is easy to see that $\{B_A : A \in F\}$ is a partition of $\mathbb{G}(F) = \operatorname{supp} Q$. Hence we obtain

$$\begin{aligned} Q(C) = \sum_{A \in F} Q(C \cap B_A) &= \sum_{A \in F} \sum_{A' \in F} \alpha_{A'} Q_{A'}(C \cap B_A) \\ &= \sum_{A \in F} \sum_{A' \supset A} \alpha_{A'} Q_{A'}(C \cap B_A) + \sum_{A \in F} \sum_{A' \subsetneqq A} \alpha_{A'} Q_{A'}(C \cap B_A) \\ &= \sum_{A \in F} \lambda_Q(A)(C \cap B_A), \end{aligned} \tag{22}$$

where we used $Q_{A'}(C \cap B_A) = Q_{A'}(C \cap B_A \cap A)$ together with flatness applied to pairs $A \subset A'$ as well as $A' \cap B_A = \emptyset$ applied to pairs $A' \subsetneqq A$. Our assumption now yields

$$Q(C) \leq \sum_{A \in F} \mathfrak{a}(C \cap B_A) = \mathfrak{a}(C \cap \operatorname{supp} Q) \leq \mathfrak{a}(C).$$

(c). Let $\mathfrak{b} := \lambda_Q(B)$ be a level of B in Q with $\mathfrak{b} \not\leq P$. Then there is a $B' \in \mathcal{B}$ with $\mathfrak{b}(B') > P(B)$ and for $B'' := B' \cap \operatorname{supp} \mathfrak{b} = B' \cap B$ we find $Q(B'') \geq \mathfrak{a}(B'') = \mathfrak{a}(B') > P(B') \geq P(B'')$. Conversely, assume $\mathfrak{b} \leq P$ for all levels $\mathfrak{b}$ in Q. By the decomposition (22) we then obtain

$$Q(C) = \sum_{A \in F} \lambda_Q(A)(C \cap B_A) \leq \sum_{A \in F} P(C \cap B_A) = P(C \cap \operatorname{supp} Q) \leq P(C). \qquad \blacksquare$$

Corollary 35 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base, $Q \in \mathcal{S}(\mathcal{A})$ a simple measure with representing forest F and $A_1, A_2 \in F$. Then for all $\mathfrak{a} \in \mathcal{Q}_Q(A_1 \cup A_2)$ there exists a level $\mathfrak{b}$ in Q such that $A_1 \cup A_2 \subset B$ and $\mathfrak{a} \leq \mathfrak{b}$.*

Proof of Corollary 35: Let us fix an $\mathfrak{a} \in \mathcal{Q}_Q(A_1 \cup A_2)$. Since $\mathfrak{a} \leq Q$, Lemma 34 gives a level $\mathfrak{b}$ in Q with $\mathfrak{a} \leq \mathfrak{b}$. Setting $B := \operatorname{supp} \mathfrak{b} \in F$ then gives $A_1 \cup A_2 \subset \operatorname{supp} \mathfrak{a} \subset B$. ■

Proof of Proposition 19: Let Q be a simple measure and $Q = \sum_{A \in F} \alpha_A Q_A$ be its unique representation. Moreover, let A_1, A_2 be direct siblings in F and $\mathfrak{a}_1, \mathfrak{a}_2$ be the corresponding levels in Q. Then Q-groundedness follows directly from Corollary 35. To show that A_1, A_2 are Q-motivated and Q-fine, we fix an $\mathfrak{a} \in \mathcal{Q}_Q(A_1 \cup A_2)$. Furthermore, let $\mathfrak{b}$ be the level in Q found by Corollary 35, i.e. we have $A_1 \cup A_2 \subset \operatorname{supp} \mathfrak{b} =: B$ and $\mathfrak{a} \leq \mathfrak{b} \leq Q$. Now let $A_3, \ldots, A_k \in F$ be the remaining direct siblings of A_1 and A_2. Since B is an ancestor of A_1 and A_2 it is also an ancestor of $A_3, \ldots, A_k$ and hence $A_1 \cup \cdots \cup A_k \subset B$. This immediately gives $\mathfrak{b} \in \mathcal{Q}_Q(A_1 \cup \cdots \cup A_k)$ and we already know $\mathfrak{b} \geq \mathfrak{a}$. In other words, A_1, A_2 are Q-fine. Finally, observe that for $B \subset A'$ flatness gives $Q_{A'}(B)Q_B(\cdot) = Q_{A'}(\cdot \cap B)$. Since $A_1 \subset B$ we hence obtain

$$\mathfrak{a}(A_1) \leq \mathfrak{b}(A_1) = \sum_{A' \supset B} \alpha_{A'} Q_{A'}(B) Q_B(A_1) = \sum_{A' \supset B} \alpha_{A'} Q_{A'}(A_1)$$

and since $Q_{A_1}(A_1) = 1$ we also find

$$\mathfrak{a}_1(A_1) = \sum_{A' \supset A_1} \alpha_{A'} Q_{A'}(A_1) Q_{A_1}(A_1) = \sum_{A' \supset A_1} \alpha_{A'} Q_{A'}(A_1) = \sum_{A' \supset B} \alpha_{A'} Q_{A'}(A_1) + \alpha_{A_1} .$$

Since $\alpha_{A_1} > 0$ we conclude that $\mathfrak{a}(A_1) < (1 - \varepsilon_1)\mathfrak{a}_1(A_1)$ for a suitable $\varepsilon_1 > 0$. Analogously, we find an $\varepsilon_2 > 0$ with $\mathfrak{a}(A_2) < (1 - \varepsilon_2)\mathfrak{a}_2(A_2)$ and taking $\alpha := 1 - \min\{\varepsilon_1, \varepsilon_2\}$ thus yields Q-motivation. ■

5.2.1 Proof of Theorem 20

Lemma 36 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base, $P \in \mathcal{M}_\Omega$, and $Q, Q' \leq P$ be simple measures on finite forests F and F'. If all roots in both F and F' are P-grounded, then any root in one tree can only be kin below P to at most one root in the other tree.*

Proof of Lemma 36: Let us assume the converse, i.e. we have an $A \in \max F$ and $B, B' \in \max F'$ such that $A \sim_P B$ and $A \sim_P B'$. Let $\mathfrak{a}, \mathfrak{b}, \mathfrak{b}'$ be the respective summands in the simple measures Q and Q'. Then $0 < \mathfrak{a}(A) \leq Q(A) \leq P(A)$ and analogously $P(B), P(B') > 0$. Then by transitivity of $\sim_P$ established in Lemma 31 we have $B \sim_P B'$ and by groundedness there has to be a parent for both in F', so they would not be roots. ■

Proposition 37 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a stable clustering base and $P \in \mathcal{M}$ such that $\mathcal{A}$ is P-subadditive. Let $(Q_n, F_n) \uparrow P$, where all forests F_n have k roots $A_n^1, \ldots, A_n^k$, which, in addition, are assumed to be P-grounded. Then $A^i := \bigcup_n A_n^i$ are unique under all such approximations up to a P-null set.*

Proof of Proposition 37: The $A^1, \ldots, A^k$ are pairwise $\perp$-disjoint by Lemma 11 and by Lemma 53 they partition $\operatorname{supp} P$ up to a P-null set, i.e. $P(\operatorname{supp} P \setminus \bigcup_i A^i) = 0$. Therefore any $B \in \mathcal{B}$ with $P(B) > 0$ intersects at least one of the A_i. Moreover, we have $0 < Q_1(A_1^i) \leq P(A_1^i) \leq P(A^i)$, i.e. $P(A^i) > 0$. Now let $(Q'_n, F'_n) \uparrow P$ be another approximation of the assumed type with roots B_n^i and limit roots $B^1, \ldots, B^{k'}$. Clearly, our preliminary considerations also hold for these limit roots. Now consider the binary relation $i \sim j$, which is defined to hold iff $A^i \between_P B^j$.

Since $P(A^i) > 0$ there has to be a B^j with $P(A^i \cap B^j) > 0$, so for all $i \leq k$ there is a $j \leq k'$ with $i \sim j$. Then, since $A_n^i \cap B_n^j \uparrow A^i \cap B^j$, there is an $n \geq 1$ with $P(A_n^i \cap B_n^j) > 0$. By P-subadditivity of $\mathcal{A}$ we conclude that A_n^i and B_n^j are kin below P, and Lemma 36 shows that this can only happen for at most one $j \leq k'$. Consequently, we have $k \leq k'$ and $\sim$ defines an injection $i \mapsto j(i)$. The same argument also holds in the other direction and we see that $k = k'$ and that $i \sim j$ defines a bijection. Clearly, we may assume that $i \sim j$ iff $i = j$. Then $P(A^i \cap B^j) > 0$ if and only if $i = j$, and since both sets of roots partition $\operatorname{supp} P$ up to a P-null set, we conclude that $P(A^i \triangle B^i) = 0$. ∎

Lemma 38 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $P \in \mathcal{M}$ such that $\mathcal{A}$ is P-subadditive. Moreover, let $\mathfrak{a}_1, \ldots, \mathfrak{a}_k \leq P$ be base measures on $A_1, \ldots, A_k \in \mathcal{A}$ such that $A_1 \between_P A_i$ for all $2 \leq i \leq k$. Then there is $\mathfrak{b} \in \mathcal{Q}_P(A_1 \cup \ldots \cup A_k)$ and an $\mathfrak{a}_i$ such that $\mathfrak{b} \geq \mathfrak{a}_i$, and if $k \geq 3$ and the $\mathfrak{a}_2, \ldots, \mathfrak{a}_k$ satisfy the motivation implication* (20) *pairwise, then $\mathfrak{b} \geq \mathfrak{a}_1$.*

Proof of Lemma 38: The proof of the first assertion is based on induction. For $k = 2$ the assertion is P-subadditivity. Now assume that the statement is true for k. Then there is a $\mathfrak{b} \in \mathcal{Q}_P(A_1 \cup \ldots \cup A_k)$ and an $i_0 \leq k$ with $\mathfrak{b} \geq \mathfrak{a}_{i_0}$. The assumed $A_1 \between_P A_{k+1}$ thus yields

$$P(A_{k+1} \cap \operatorname{supp} \mathfrak{b}) \geq P(A_{k+1} \cap A_1) > 0\,,$$

and hence P-subadditivity gives a $\tilde{b} \in \mathcal{Q}_P(A_{k+1} \cup \operatorname{supp} \mathfrak{b})$ with $\tilde{\mathfrak{b}} \geq \mathfrak{a}_{k+1}$ or $\tilde{\mathfrak{b}} \geq \mathfrak{b} \geq \mathfrak{a}_{i_0}$. For the second assertion observe that $\mathfrak{b} \in \mathcal{Q}_P(A_i \cap A_j)$ for all i, j and hence (20) implies $\mathfrak{b} \not\geq \mathfrak{a}_i$ for $i \geq 2$. ∎

Lemma 39 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $Q \leq P$ be a simple and P-adapted measure with representing forest F. Let $C^1, \ldots, C^k \in F$ be direct siblings for some $k \geq 2$. Then there exists an $\varepsilon > 0$ such that:*

(a) For all $\mathfrak{a} \in \mathcal{Q}_P(C^1 \cup \ldots \cup C^k)$ and $i \leq k$ we have $\mathfrak{a}(C^i) \leq (1 - \varepsilon) \cdot Q(C^i)$.

(b) Assume that $\mathcal{A}$ is P-subadditive and that $\mathfrak{a} \leq P$ is a simple measure with $\operatorname{supp} \mathfrak{a} \between_P C^i$ *for at least two $i \leq k$. Then for all $i \leq k$ we have $\mathfrak{a}(C^i) \leq (1 - \varepsilon) \cdot Q(C^i)$.*

(c) If $\mathcal{A}$ is P-subadditive and $Q' \leq P$ is a simple measure with representing forest F' such that there is an $i \leq k$ with the property that for all $B \in F'$ we have

$$B \between_P C^i \implies \exists j \neq i \colon B \between_P C^j.$$

Then $Q'(\cdot \cap C^i) \leq (1 - \varepsilon)\, Q(\cdot \cap C^i)$ holds true.

Proof of Lemma 39: Let $\mathfrak{c}_1, \ldots, \mathfrak{c}_k$ be the levels of $C^1, \ldots, C^k$ in Q. Since Q is adapted, (20) holds for some $\alpha \in (0,1)$. We define $\varepsilon := 1 - \alpha$.
(a). We fix an $\mathfrak{a} \in \mathcal{Q}_P(C^1 \cup \ldots \cup C^k)$, an $i \le k$, and a $j \le k$ with $j \neq i$. Let $\mathfrak{c}_i, \mathfrak{c}_j$ be the levels of C^i and C^j in Q. Since $\alpha\mathfrak{c}_i$ and $\alpha\mathfrak{c}_j$ are motivated, we have $\mathfrak{a} \not\ge \alpha\mathfrak{c}_i$ and $\mathfrak{a} \not\ge \alpha\mathfrak{c}_j$. Hence, there is a $C_0 \in \mathcal{B}$ with $\mathfrak{a}(C_0) < \alpha\mathfrak{c}_i(C_0)$ and thus also $\mathfrak{a}(C_0 \cap C^i) < \alpha\mathfrak{c}_i(C_0 \cap C^i)$. Lemma 33 then yields $\mathfrak{a}(\cdot \cap C^i) \le \alpha\mathfrak{c}_i(\cdot \cap C^i)$ and the definition of levels gives

$$\mathfrak{a}(C^i) \le \alpha\mathfrak{c}_i(C^i) = \alpha Q(C^i) = (1-\varepsilon)Q(C^i).$$

(b). We may assume $\operatorname{supp}\mathfrak{a} \propto_P C^1$ and $\operatorname{supp}\mathfrak{a} \propto_P C^2$. By the second part of Lemma 38 applied to $\operatorname{supp}\mathfrak{a}, C^1, C^2$ there is an $\mathfrak{a}' \in \mathcal{Q}_P(\operatorname{supp}\mathfrak{a} \cup C^1 \cup C^2) \subset \mathcal{Q}_P(C^1 \cup C^2)$ with $\mathfrak{a}' \ge \mathfrak{a}$, and since Q is P-fine, we may actually assume that $\mathfrak{a}' \in \mathcal{Q}_P(C^1 \cup \ldots \cup C^k)$. Now part (a) yields $\mathfrak{a}'(C^i) \le (1-\varepsilon) \cdot Q(C^i)$ for all $i = 1, \ldots, k$.
(c). We may assume $i = 1$. Our first goal is to show

$$\mathfrak{b}(\cdot \cap C^1) \le (1-\varepsilon)\mathfrak{c}_1(\cdot \cap C^1) \tag{23}$$

for all levels $\mathfrak{b}$ in Q', To this end, we fix a level $\mathfrak{b}$ in Q' and write $B := \operatorname{supp}\mathfrak{b}$. If $P(B \cap C^1) = 0$, then (23) follows from

$$\mathfrak{b}(C^1) = \mathfrak{b}(B \cap C^1) \le P(B \cap C^1) = 0\,.$$

In the other case we have $B \propto_P C^1$ and our assumption gives a $j \neq 1$ with $B \propto_P C^j$. By the second part of Lemma 38 we find an $\mathfrak{a} \in \mathcal{Q}_P(B \cup C^1 \cup C^j) \subset \mathcal{Q}_P(C^1 \cup C^j)$ with $\mathfrak{a} \ge \mathfrak{b}$, and by (a) we thus obtain $\mathfrak{a}(C^1) \le (1-\varepsilon)\,Q(C^1) = (1-\varepsilon)\mathfrak{c}_1(C^1)$. Now, Lemma 33 gives $\mathfrak{a}(\cdot \cap C^1) \le (1-\varepsilon)\mathfrak{c}_1(\cdot \cap C^1)$ and hence (23) follows.

With the help of (23) we now conclude by part (b) of Lemma 34 that $Q'(\cdot \cap C^1) \le (1-\varepsilon)\mathfrak{c}_1(\cdot \cap C^1)$ and using $\mathfrak{c}_1(\cdot \cap C^1) \le Q(\cdot \cap C^1)$ we thus obtain the assertion. ■

Lemma 40 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base and $P \in \mathcal{M}$ such that $\mathcal{A}$ is P-subadditive. Moreover, let $Q, Q' \le P$ be simple P-adapted measures on F, F', and $S \in s(F)$ and $S' \in s(F')$ be two nodes that have children in $s(F)$ and $s(F')$, respectively. Let*

$$\{C^1, \ldots, C^k\} = \max s(F)\big|_{\subsetneq S} \qquad \textit{and} \qquad \{D^1, \ldots, D^{k'}\} = \max s(F')\big|_{\subsetneq S'}$$

be their direct children and consider the relation $i \sim j :\Leftrightarrow C^i \propto_P D^j$. Then we have $k, k' \ge 2$ and if $\sim$ is left-total, i.e. for every $i \le k$ there is a $j \le k'$ with $i \sim j$, then it is right-unique, i.e. for every $i \le k$ there is at most one $j \le k'$ with $i \sim j$.

Proof of Lemma 40: The definition of the structure of a forest gives $k, k' \ge 2$. Moreover, we note that $P(A) \ge Q(A) > 0$ for all $A \in F$ and $P(A) \ge Q'(A) > 0$ for all $A \in F'$. Now assume that $\sim$ is not right-unique, say $1 \sim j$ and $1 \sim j'$ for some $j \neq j'$. Applying P-subadditivity twice we then find a $\mathfrak{b} \in \mathcal{Q}_P(C^1 \cup D^j \cup D^{j'})$ with $\mathfrak{b} \ge \mathfrak{c}_1$ or $\mathfrak{b} \ge \mathfrak{d}_j$ or $\mathfrak{b} \ge \mathfrak{d}_{j'}$, where $\mathfrak{c}_1$, $\mathfrak{d}_j$, and $\mathfrak{d}_{j'}$ are the corresponding levels. Since $\mathfrak{d}^j, \mathfrak{d}^{j'}$ are motivated we conclude that $\mathfrak{b} \ge \mathfrak{c}_1$. Now, because of $\mathcal{Q}_P(C^1 \cup D^j \cup D^{j'}) \subset \mathcal{Q}_P(D^j \cup D^{j'})$ and P-fineness of Q' there is a $\mathfrak{b}' \in \mathcal{Q}_P(D_1 \cup \ldots \cup D_{k'})$ with $\mathfrak{b}' \ge \mathfrak{b}$. Now pick a direct sibling of C^1, say C^2. Then there is a j'' with $2 \sim j''$, and since $B' := \operatorname{supp}\mathfrak{b}' \supset D_1 \cup \ldots \cup D_{k'}$ this implies

$P(B' \cap C^2) \geq P(D^{j''} \cap C^i) > 0$. By P-subadditivity we hence find a $\mathfrak{b}'' \in \mathcal{Q}_P(B' \cup C^2) \subset \mathcal{Q}_P(C^1 \cup C^2)$ with $\mathfrak{b}'' \geq \mathfrak{b}'$ or $\mathfrak{b}'' \geq \mathfrak{c}_2$. Clearly, $\mathfrak{b}'' \geq \mathfrak{c}_2$ violates the fact that C^1, C^2 are motivated, and thus $\mathfrak{b}'' \geq \mathfrak{b}'$. However, we have shown $\mathfrak{b}' \geq \mathfrak{b} \geq \mathfrak{c}_1$, and thus $\mathfrak{b}'' \geq \mathfrak{c}_1$. Since this again violates the fact that C^1, C^2 are motivated, we have found a contradiction. ■

Proof of Theorem 20: We prove the theorem by induction over the generations in the forests. For a finite forest F, we define $s_0(F) := \max F$ and

$$s_{N+1}(F) := s_N(F) \cup \big\{ A \in s(F) \mid A \text{ is a direct child of a leaf in } s_N(F) \big\}.$$

We will now show by induction over N that there is a graph-isomorphism $\zeta_N \colon s_N(F_\infty) \to s_N(F'_\infty)$ with $P(A \triangle \zeta_N(A)) = 0$ for all $A \in s_N(F_\infty)$. For $N = 0$ this has already been shown in Proposition 37. Let us therefore assume that the statement is true for some $N \geq 0$. Let us fix an $S \in \min s_N(F_\infty)$ and let $S' := \zeta_N(S) \in \min s_N(F'_\infty)$ be the corresponding node. We have to show that both have the same number of direct children in $s_{N+1}(\cdot)$ and that these children are equal up to P-null sets. By induction this then finishes the proof.

Since $S \in s_N(F_\infty) \subset s(F_\infty)$, the node S has either no children or at least 2. Now, if both S and S' have no direct children then we are finished. Hence we can assume that S has direct children $C^1, \ldots, C^k$ for some $k \geq 2$, i.e.

$$\max(F_\infty\big|_{\subsetneq S}) = \{C^1, \ldots, C^k\}.$$

Let $S_n, C_n^1, \ldots, C_n^k \in s(F_n)$ and $S'_n \in s(F'_n)$ be the nodes that correspond to $S, C^1, \ldots, C^k$, and S', respectively. Since $P(S \triangle S') = 0$ we then obtain for all $i \leq k$

$$P(S' \cap C^i) = P(S \cap C^i) = P(C^i) \geq Q_1(C^i) \geq Q_1(C_1^i) > 0\,,$$

that is $S' \varpropto_P C^i$ for all $i \leq k$. Since $S' = \bigcup_n S'_n$ and $C^i = \bigcup_n C_n^i$ this can only happen if $S'_n \varpropto_P C_n^i$ for all sufficiently large n. We therefore may assume without loss of generality that

$$S'_1 \varpropto_P C_n^i \qquad\qquad \text{for all } i \leq k \text{ and all } n \geq 1. \tag{24}$$

Let us now investigate the structure of $F'_n\big|_{\subset S'_n}$. To this end, we will seek a kind of anchor $B'_n \in F'_n\big|_{\subset S'_n}$, which will turn out later to be the direct parent of the yet to find $\zeta_{N+1}(C^i) \in F'_\infty$. We define this anchor by

$$B'_n := \min\{B \in F'_n \mid B \varpropto_P C_1^i \text{ for all } i = 1, \ldots, k\}.$$

This minimum is unique. Indeed, let $\tilde{B}'_n$ be any other minimum with $\tilde{B}'_n \varpropto_P C_1^i$ for all $i \leq k$. Since both are minima, none is contained in the other and because F'_n is a forest this means $B'_n \perp \tilde{B}'_n$. Let $\mathfrak{b}'_n$ and $\tilde{\mathfrak{b}}'_n$ be their levels in Q'_n. Since Q'_n is P-adapted, these two levels are motivated. This means that there can be no base measure majorizing one of them and supporting $B'_n \cup \tilde{B}'_n$. On the other hand, by the second part of Lemma 38 there exists a $\mathfrak{b}''_n \in \mathcal{Q}_p(B'_n \cup C_1^1 \cup \cdots \cup C_1^k)$ with $\mathfrak{b}''_n \geq \mathfrak{b}'_n$. Now because of $P(\tilde{B}'_n \cap \operatorname{supp} \mathfrak{b}''_n) \geq P(\tilde{B}'_n \cap C_1^1) > 0$ and P-subadditivity there exists a base measure majorizing $\tilde{\mathfrak{b}}'_n \geq \mathfrak{b}'_n$ or $\mathfrak{b}''_n$ and supporting $\tilde{B}'_n \cap \operatorname{supp} \mathfrak{b}''_n$. This contradicts the motivatedness of $\mathfrak{b}'_n$ and $\tilde{\mathfrak{b}}'_n$ and hence the minimum B'_n is unique.

Since B'_n is the unique minimum among all $B \in F'_n$ with $B \between_P C_1^i$ for all i, we also have $B'_n \subset B$ for all such B and hence $B'_n \subset S'_n$ by (24). The major difficulty in handling B'_n though is that it may jump around as a function of n: Indeed we may have $B'_n \in F'_n \setminus s(F'_n)$ and therefore the monotonicity $s(F'_n) \leq s(F'_{n+1})$ says nothing about B'_n. In particular, we have in general $B'_n \not\subset B'_{n+1}$.

Let us now enumerate the set $\min F'_n\big|_{\subsetneq B'_n}$ of direct children of B'_n by $D_n^1, \ldots, D_n^{k_n}$, where $k_n \geq 0$. Again these D_n^i can jump around as a function of n. The number k_n specifies different cases: we have $B'_n \in \min F'_n$, i.e. B'_n is a leaf, iff $k_n = 0$; on the other hand $D_n^i \in s(F'_n)$ iff $k_n \geq 2$. Next we show that for all $i \leq k$ and all sufficiently large n there is an index $j(i,n) \in \{1, \ldots, k_n\}$ with

$$C_1^i \between_P D_n^{j(i,n)} . \tag{25}$$

Note that this in particular implies $k_n \geq 1$ for sufficiently large n. To this end we fix an $i \leq k$. Suppose that $C_1^i \perp\!\!\!\perp_P (D_{n_m}^1 \cup \cdots \cup D_{n_m}^{k_{n_m}})$ for infinitely many $n_1, n_2, \ldots$. By construction B'_{n_m} is the smallest element of F'_{n_m} that $\perp\!\!\!\perp_P$-intersects C_1^i. More precisely, for any $A \in F'_{n_m}$ with $A \between_P C_1^i$ we have $A \supset B'_{n_m}$ and therefore $A \between_P C_1^{i'}$ for all such A and all $i' \leq k$. Hence, all Q'_{n_m} in this subsequence fulfill the conditions of the last statement in Lemma 39 and we get an $\varepsilon > 0$ such that for all such n_m

$$Q'_{n_m}(C_1^i) \leq (1-\varepsilon) Q_1(C_1^i) \leq (1-\varepsilon) P(C_1^i) \tag{26}$$

which contradicts $Q'_{n_m}(C_1^i) \uparrow P(C_1^i)$ since $P(C_1^i) > 0$.

Therefore for all $i \leq k$ and all sufficiently large n there is an index $j(i,n)$ such that (25) holds. Clearly, we may thus assume that there is such an $j(i,n)$ for all $n \geq 1$. Since $j(i,n) \in \{1, \ldots, k_n\}$ we conclude that $k_n \geq 1$ for all $n \geq 1$. Moreover, $k_n = 1$ is impossible, since $k_n = 1$ yields $j(i,n) = 1$, and this would mean, that $C_1^i \between_P D_n^1$ for all $i \leq k$ contradicting that B'_n is the minimal set in F'_n having this property. Consequently B'_n has the direct children $D_n^1, \ldots, D_n^{k_n}$ where $k_n \geq 2$ for all $n \geq 1$.

So far we have seen that $D_n^1, \ldots, D_n^{k_n} \in s(F'_n)$ are inside S'_n. Therefore S'_n is not a leaf, and hence $S' \notin \min F'_\infty$ as well. But still for infinitely many n these D_n^j might not be the direct children of S'_n. Let us therefore denote the direct children of $S'_n \in s(F'_n)$ by $E_n^1, \ldots, E_n^{k'} \in s(F'_n)$, where we pick a numbering such that $E_n^i \subset E_{n+1}^i$ and by the definition of the structure of a forest we have $k' \geq 2$.

For an arbitrary but fixed n we now show $\{D_n^1, \ldots, D_n^{k_n}\} = \{E_n^1, \ldots, E_n^{k'}\}$. To this let us assume the converse. Since the E_n^j are the direct children of S'_n in the structure $s(F'_n)$ there is a $j_n \leq k'$ with $D_n^j \subset E_n^{j_n}$ for all j, and since B'_n is the direct parent of the D_n^j we conclude that $B'_n \subset E_n^{j_n}$. Therefore we have $C_1^i \between_P E_n^{j_n}$ for all $i \leq k$. Since Q_1 and Q'_n are adapted we can use Lemma 40 to see that for all $i \leq k$ we have $C_1^i \perp\!\!\!\perp_P E_n^j$ for all $j \neq j_n$. Let us fix a $j \neq j_n$. Our goal is to show

$$Q_m(E_n^j) < (1-\varepsilon) Q'_n(E_n^j) ,$$

for all sufficiently large $m \geq n$, since this inequality contradicts the assumed convergence of $Q_m(E_n^j)$ to $P(E_n^j) \geq Q'_n(E_n^j) > 0$. By part (c) of Lemma 39 with Q'_n as Q and Q_m as Q' it suffices to show that for all $A \in F_m$ and all sufficiently large $m \geq n$ we have

$$A \between_P E_n^j \implies A \between_P E_n^{j_n} . \tag{27}$$

To this end, we fix an $A \in F_m$ with $A \perp\!\!\!\perp_P E_n^j$. Then we first observe that for all $m \geq n$ we have $P(A \cap S'_m) \geq P(A \cap S'_n) \geq P(A \cap E_n^j) > 0$. Moreover, the induction assumption ensures $P(S \triangle S') = 0$ and since $S_m \nearrow S$ and $S'_m \nearrow S'$, we conclude that $P(A \cap S_m) > 0$ for all sufficiently large m. Now, $C_m^1 \cup \cdots \cup C_m^k$ are direct siblings and hence we either have $C_m^1 \cup \cdots \cup C_m^k \subset A$ or $A \subset C_m^{i_0}$ for exactly one $i_0 \leq k$. In the first case we get

$$P(A \cap E_n^{j_n}) \geq P(C_m^1 \cap E_n^{j_n}) \geq P(C_1^1 \cap E_n^{j_n}) > 0$$

by the already established $C_1^i \perp\!\!\!\perp_P E_n^{j_n}$ for all $i \leq k$. The second case is impossible, since it contradicts adaptedness. Indeed, $A \subset C_m^{i_0}$ implies $C_m^{i_0} \perp\!\!\!\perp_P E_n^j$ and by the already established $C_1^i \perp\!\!\!\perp_P E_n^{j_n}$ for all $i \leq k$, we also know $C_m^{i_0} \perp\!\!\!\perp_P E_n^{j_n}$. By the second part of Lemma 38 we therefore find a $\tilde{\mathfrak{c}} \in \mathcal{Q}_P(C_m^{i_0} \cup E_n^j \cup E_n^{j_n})$ with $\tilde{\mathfrak{c}} \geq \mathfrak{c}_m^{i_0}$, where $\mathfrak{c}_m^{i_0}$ is the level of $C_m^{i_0}$ in Q_m. Now fix any $i \leq k$ with $i \neq i_0$ and observe that we have $P(C_m^i \cap \operatorname{supp} \tilde{\mathfrak{c}}) \geq P(C_m^i \cap E_n^{j_n}) \geq P(C_1^i \cap E_n^{j_n}) > 0$, and hence P-subadditivity yields a $\mathfrak{c}'' \in \mathcal{Q}_P(C_m^i \cup \operatorname{supp} \tilde{\mathfrak{c}})$ with $\mathfrak{c}'' \geq \mathfrak{c}_m^i$ or $\mathfrak{c}'' \geq \tilde{\mathfrak{c}} \geq \mathfrak{c}_m^{i_0}$, where $\mathfrak{c}_m^i$ is the level of C_m^i in Q_m. Since $\mathfrak{c}'' \in \mathcal{Q}_P(C_m^i \cup \operatorname{supp} \tilde{\mathfrak{c}}) \subset \mathcal{Q}_P(C_m^i \cup C_m^{i_0})$, we have thus found a contradiction to the fact that the direct siblings C_m^i and $C_m^{i_0}$ are P-motivated.

So far we have shown $\{D_n^1, \ldots, D_n^{k_n}\} = \{E_n^1, \ldots, E_n^{k'}\}$ and $k_n = k'$ for all n. Without loss of generality we may thus assume that $D_n^j = E_n^j$ for all n and all $j \leq k'$. In particular, this means that the direct children of S'_n in $s(F'_n)$ equal the direct children of B'_n in F'_n. Let us write

$$D^j := \bigcup_{n \geq 1} D_n^j, \qquad\qquad j = 1, \ldots, k'$$

and $i \sim j$ iff $C_1^i \perp\!\!\!\perp_P D_1^j$. We have seen around (25) that for all $i \leq k$ there is at least one $j \leq k_1 = k'$ with $i \sim j$, namely $j(i,1)$. By Lemma 40 we then conclude that $j(i,1)$ is the only index $j \leq k'$ satisfying $i \sim j$. By reversing the roles of C_1^i and D_1^j, which is possible since $D_1^j = E_1^j$ is a direct children of S'_n in $s(F'_n)$, we can further see that for all j there is an index i with $i \sim j$ and again by Lemma 40 we conclude that there is at most one i with $i \sim j$. Consequently, $i \sim j$ defines a bijection between $\{C_1^1, \ldots, C_1^k\}$ and $\{D_1^1, \ldots, D_1^{k'}\}$ and hence we have $k = k'$. Moreover, we may assume without loss of generality that $i \sim j$ iff $i = j$. From the latter we obtain $C_1^i \perp\!\!\!\perp_P D_1^j$ iff $i = j$.

To generalize the latter, we fix $n, m \geq 1$ and write $i \sim j$ iff $C_n^i \perp\!\!\!\perp_P D_m^j$. Since we have $P(C_n^i \cap D_m^i) \geq P(C_1^i \cap D_1^i) > 0$, we conclude that $i \sim i$, and by Lemma 40 we again see that $i \sim j$ is false for $i \neq j$. This yields $C_n^i \perp\!\!\!\perp_P D_m^j$ iff $i = j$ and by taking the limits, we find $C^i \perp\!\!\!\perp_P D^j$ iff $i = j$.

Next we show that $P(C^i \triangle D^i) = 0$ for all $i \leq k$. Clearly, it suffices to consider the case $i = 1$. To this end assume that $R := C^1 \setminus D^1$ satisfies $P(R) > 0$. For $R_n := R \cap C_n^1 = C_n^1 \setminus D^1$, we then have $R_n \uparrow R$ since $C_n^1 \uparrow C^1$ and $R \subset C^1$. Consequently, $0 < P(R) = P(R \cap C^1)$ implies $P(R_n) > 0$ for all sufficiently large n. On the other hand, we have $P(R \cap D^1) = 0$ by the definition of R and $P(R \cap D^j) \leq P(C^1 \cap D^j) = 0$ for all $j \neq 1$ as we have shown above.

We next show that $Q'_m(R_n) = Q'_m|_{\supset B'_m}(R_n)$. To this end it suffices to show that for any $A \in F'_m$ with $A \notin F'_m|_{\supset B'_m}$ we have $Q'_m(A \cap R_n) \leq P(A \cap R_n) = 0$. Let us thus fix an $A \in F'_m$ with $A \notin F'_m|_{\supset B'_m}$. Then we either have $A \subsetneq B'_m$ or $A \perp B'_m$. In the first case there is $j \le k$ with $A \subset D^j_m$ which means, as shown above, that $P(A \cap R_n) \le P(D^j_m \cap R_n) = 0$. In the second case, by definition of structure, we even have $A \perp S'_m$. So there is a $A'_m \in s(F'_m)$ with $A \subset A'_m$ and $A'_m \perp S'_m$ and by isomonotonicity of the structure there is $A' \in F'_\infty$ with $A'_m \subset A'$ and $A' \perp S'$. Hence by induction assumption $P(A \cap R_n) \le P(A \cap S_n) \le P(A \cap S) \le P(A' \cap S) = P(A' \cap S') = 0$.

Using $P(C^i \cap D^i) > 0$ we now observe that $Q'_m\big|_{\supset B'_m}$ fulfills the conditions of part (c) of Lemma 39 for C^1 and C^2 and by $R_n \subset C^1_n$ we thus obtain

$$Q'_m(R_n) = Q'_m\big|_{\supset B'_m}(R_n) \le (1-\varepsilon)Q_n(R_n) \le (1-\varepsilon)P(R_n).$$

This contradicts $0 < P(R_n) = \lim_{m\to\infty} Q'_m(R_n)$. So we can assume $P(R_n) = 0$ for all n and therefore $P(R) = \lim_{n\to\infty} P(R_n) = 0$. By reversing roles we thus find $P(D^1 \triangle C^1) = P(C^1 \setminus D^1) + P(D^1 \setminus C^1) = 0$ and therefore the children are indeed the same up to P-null sets.

Finally, we are able to finish the induction: To this end we extend ζ_N to the map $\zeta_{N+1} \colon s_{N+1}(F_\infty) \to s_{N+1}(F'_\infty)$ by setting, for every leaf $S \in \min s_N(F_\infty)$,

$$\zeta_{N+1}(C^i) := D^i$$

where $C^1, \ldots, C^k \in s_{N+1}(F_\infty)$ are the direct children of S and $D^1, \ldots, D^k \in s_{N+1}(F'_\infty)$ are the nodes we have found during our above construction. Clearly, our construction shows that ζ_{N+1} is a graph isomorphism satisfying $P(A \triangle \zeta_{N+1}(A)) = 0$ for all $A \in s_{N+1}(F_\infty)$. ■

5.2.2 Proof of Theorem 21

Lemma 41 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base, $P_1, \ldots, P_k \in \mathcal{M}$ with $\operatorname{supp} P_i \perp \operatorname{supp} P_j$ for all $i \ne j$, and $Q_i \le P_i$ be simple measures with representing forests F_i. We define $P := P_1 + \ldots + P_k$, $Q := Q_1 + \ldots + Q_k$, and $F := F_1 \cup \cdots \cup F_k$. Then we have:*

- *(a) The measure Q is simple and F is its representing $\perp$-forest.*
- *(b) For all base measures $\mathfrak{a} \le P$ there exists exactly one i with $\mathfrak{a} \le P_i$.*
- *(c) If $\mathcal{A}$ is P_i-subadditive for all $i \le k$, then $\mathcal{A}$ is P-subadditive.*
- *(d) if Q_i is P_i-adapted for all $i \le k$, then Q is adapted to P.*

Proof of Lemma 41: *(a).* Since $Q_i \le P_i \le P$ we have $\mathbb{G}F_i = \operatorname{supp} Q_i \subset \operatorname{supp} P_i$. By the monotonicity of $\perp$ we then obtain $\mathbb{G}F_i \perp \mathbb{G}F_j$ for $i \ne j$. From this we obtain the assertion.
(b). Let $\mathfrak{a} \le P$ be a base measure on $A \in \mathcal{A}$. Then we have $A = \operatorname{supp} \mathfrak{a} \subset \operatorname{supp} P = \bigcup_i \operatorname{supp} P_i$. By $\mathcal{A}$-connectedness there thus exists a i with $A \subset \operatorname{supp} P_i$. For $B \in \mathcal{B}$ we then find $\mathfrak{a}(B) = \mathfrak{a}(B \cap \operatorname{supp} P_i) \le P(B \cap \operatorname{supp} P_i) = P_i(B \cap \operatorname{supp} P_i) = P_i(B)$. Moreover, for $j \ne i$ we have $\mathfrak{a}(A) > 0$ and $P_j(A) = 0$ and thus i is unique.
(c). Let $\mathfrak{a}, \mathfrak{a}' \le P$ be base measures on base sets A, A' with $A \varpropto_P A'$. Since $A \perp A'$ implies $A \perp_\emptyset A'$, we have $A \varpropto A'$. By (b) we find unique indices i, i' with $\mathfrak{a} \le P_i$ and $\mathfrak{a}' \le P_{i'}$. This implies $A \subset \operatorname{supp} P_i$ and $A' \subset \operatorname{supp} P_j$, and hence we have $\operatorname{supp} P_i \varpropto \operatorname{supp} P_{i'}$ by monotonicity. This gives $i = i'$, i.e. $\mathfrak{a}, \mathfrak{a}' \le P_i$. Since $\mathcal{A}$ is P_i-subadditive there now is an $\tilde{\mathfrak{a}} \in \mathcal{Q}_{P_i}(A \cup A')$ with $\tilde{\mathfrak{a}} \ge \mathfrak{a}$ or $\tilde{\mathfrak{a}} \ge \mathfrak{a}'$, and since $\tilde{\mathfrak{a}} \le P_i \le P$ we obtain the assertion.

(d). From (b) we conclude $\mathcal{Q}_P(A_1 \cup A_2) = \emptyset$ for all roots $A_1 \in F_i$ and $A_2 \in F_j$ and all $i \neq j$. This can be used to infer the groundedness and fineness of Q from the groundedness and fineness of the Q_i. Now let $\mathfrak{a}, \mathfrak{a}' \leq P$ be the levels of some direct siblings $A, A' \in F$ in Q and $\mathfrak{b} \in \mathcal{Q}_P(A \cup A')$ be any base measure. By (b) there is a unique i with $\mathfrak{b} \leq P_i$, and hence $\mathfrak{a}, \mathfrak{a}' \leq P_i$ as well. Therefore Q inherits strict motivation from Q_i. ■

Lemma 42 *Let $(\mathcal{A}, \mathcal{Q}, \perp)$ be a clustering base, $P \in \mathcal{M}$, $\mathfrak{a}$ be a base measure on $A \in \mathcal{A}$ with $\operatorname{supp} P \subset A$, and $Q \leq P$ be a simple measure with representing forest F. We define $P' := \mathfrak{a} + P$, $Q' := \mathfrak{a} + Q$, and $F' := \{A\} \cup F$. Then the following statements hold:*

(a) The measure Q' is simple and F' is its representing $\perp$-forest.

(b) Let $\mathfrak{a}' \leq P'$ be a base measure on A'. Then either $\mathfrak{a}' \leq \mathfrak{a}$ or there is an $\alpha \in (0,1)$ such that $\mathfrak{a}'(\cdot \cap A') = \mathfrak{a}(\cdot \cap A') + \alpha \mathfrak{a}'(\cdot \cap A')$.

(c) If $\mathcal{A}$ is P-subadditive then $\mathcal{A}$ is P'-subadditive.

(d) If Q is P-adapted, then Q' is P'-adapted.

Proof of Lemma 42: *(a).* We have $\mathbb{G}F = \operatorname{supp} Q \subset \operatorname{supp} P \subset A$ and hence F' is a $\perp$-forest, which is obviously representing Q.
(b). Let us assume that $\mathfrak{a}' \not\leq \mathfrak{a}$, i.e. there is a $C_0 \in \mathcal{B}$ with $\mathfrak{a}'(C_0) > \mathfrak{a}(C_0)$ and thus we find $\mathfrak{a}'(C_0 \cap A') = \mathfrak{a}'(C_0) > \mathfrak{a}(C_0) \geq \mathfrak{a}(C_0 \cap A')$. In addition, we have $A' = \operatorname{supp} \mathfrak{a}' \subset \operatorname{supp} \mathfrak{a} = A$, and therefore Lemma 33 shows $\mathfrak{a}(\cdot \cap A') = \gamma \mathfrak{a}'(\cdot \cap A')$, where $\gamma := \frac{\mathfrak{a}(C_0 \cap A')}{\mathfrak{a}'(C_0 \cap A')} < 1$. Setting $\alpha := 1 - \gamma$ yields the assertion.
(c). Let $\mathfrak{a}_1, \mathfrak{a}_2 \leq P'$ be base measures on sets $A_1, A_2 \in \mathcal{A}$ with $A_1 \infty_{P'} A_2$. Since $\operatorname{supp} P' = A$, we have $A_1 \cup A_2 \subset A$, and thus $\mathfrak{a} \in \mathcal{Q}_{P'}(A_1 \cup A_2)$. Clearly, if $\mathfrak{a} \geq \mathfrak{a}_1$ or $\mathfrak{a} \geq \mathfrak{a}_2$, there is nothing left to prove, and hence we assume $\mathfrak{a}_1 \not\leq \mathfrak{a}$ and $\mathfrak{a}_2 \not\leq \mathfrak{a}$. Then (b) gives $\alpha_i \in (0,1)$ with $\mathfrak{a}_i(\cdot \cap A_i) = \mathfrak{a}(\cdot \cap A_i) + \alpha_i \mathfrak{a}_i(\cdot \cap A_i)$. We conclude that $\mathfrak{a}(\cdot \cap A_i) + \alpha_i \mathfrak{a}_i(\cdot \cap A_i) = \mathfrak{a}_i(\cdot \cap A_i) \leq P'(\cdot \cap A_i) = \mathfrak{a}(\cdot \cap A_i) + P(\cdot \cap A_i)$, and thus $\alpha_i \mathfrak{a}_i = \alpha_i \mathfrak{a}_i(\cdot \cap A_i) \leq P(\cdot \cap A_i) \leq P$. Since $\mathcal{A}$ is P-subadditive, we thus find an $\tilde{\mathfrak{a}} \in \mathcal{Q}_P(A_1 \cup A_2)$ with say $\tilde{\mathfrak{a}} \geq \alpha_1 \mathfrak{a}_1$. For $\tilde{A} := \operatorname{supp} \tilde{\mathfrak{a}}$ we then have

$$\tilde{\mathfrak{a}}' := \mathfrak{a}(\cdot \cap \tilde{A}) + \tilde{\mathfrak{a}}(\cdot \cap \tilde{A}) \geq \mathfrak{a}(\cdot \cap \tilde{A}) + \alpha_1 \mathfrak{a}_1(\cdot \cap \tilde{A}) \geq \mathfrak{a}(\cdot \cap A_1) + \alpha_1 \mathfrak{a}_1(\cdot \cap A_1) = \mathfrak{a}_1 ,$$

where we used $\operatorname{supp} \mathfrak{a}_1 = A_1 \subset \tilde{A}$. Moreover $\tilde{A} = \operatorname{supp} \tilde{\mathfrak{a}} \subset \operatorname{supp} P \subset A$, together with flatness of $\mathcal{Q}$ shows that $\tilde{\mathfrak{a}}'$ is a base measure, and we also have $\tilde{\mathfrak{a}}' \leq \mathfrak{a} + \tilde{\mathfrak{a}} \leq \mathfrak{a} + P = P'$. Finally we observe that $A_1 \cup A_2 \subset \tilde{A} = \operatorname{supp} \tilde{\mathfrak{a}}'$, and hence $\tilde{\mathfrak{a}}' \in \mathcal{Q}_{P'}(A_1 \cup A_2)$.
(d). Clearly, F' is grounded because it is a tree. Now let $A_1, \ldots, A_k \in F'$, $k \geq 2$ be direct siblings and $\mathfrak{a}_i'$ be their levels in Q'. Since A is the only root it has no siblings, so for all i we have $A_i \in F$. Moreover, the levels $\mathfrak{a}_i$ of A_i in Q are P-motivated and P-fine since Q is P-adapted. Now let $\mathfrak{b} \in \mathcal{Q}_{P'}(A_1 \cup A_2)$ and $B := \operatorname{supp} \mathfrak{b}$.

To check that Q' is P'-fine, we first observe that in the case $\mathfrak{b} \leq \mathfrak{a}$ there is nothing to prove since $\mathfrak{a} \in \mathcal{Q}_{P'}(A_1 \cup \ldots \cup A_k)$ by construction. In the remaining case $\mathfrak{b} \not\leq \mathfrak{a}$ we find a $\beta > 0$ with $\mathfrak{b}(\cdot \cap B) = \mathfrak{a}(\cdot \cap B) + \beta \mathfrak{b}(\cdot \cap B)$ by (b), and by P-fineness of Q, there exists a $\tilde{\mathfrak{b}} \in \mathcal{Q}_P(A_1 \cup \ldots \cup A_k)$ with $\tilde{\mathfrak{b}} \geq \beta \mathfrak{b}$. Since $\operatorname{supp} \tilde{\mathfrak{b}} \subset \operatorname{supp} P \subset \operatorname{supp} \mathfrak{a}$ we see that

$\mathfrak{a} + \tilde{\mathfrak{b}}$ is a simple measure, and hence we can consider the level $\tilde{\mathfrak{b}}'$ of $\operatorname{supp} \tilde{\mathfrak{b}}$ in $\mathfrak{a} + \tilde{\mathfrak{b}}$. Since $\tilde{\mathfrak{b}}' \leq \mathfrak{a} + \tilde{\mathfrak{b}} \leq \mathfrak{a} + P \leq P'$, we then obtain $\tilde{\mathfrak{b}}' \in \mathcal{Q}_{P'}(A_1 \cup \ldots \cup A_k)$ and for $C \in \mathcal{B}$ we also have

$$\mathfrak{b}(C) = \mathfrak{b}(C \cap B) = \mathfrak{a}(C \cap B) + \beta\mathfrak{b}(C \cap B) \leq \mathfrak{a}(C \cap B) + \tilde{\mathfrak{b}}(C \cap B) = \tilde{\mathfrak{b}}'(C \cap B) \leq \tilde{\mathfrak{b}}'(C).$$

To check that Q' is strictly P'-motivated we fix the constant $\alpha \in (0,1)$ appearing in the strict P-motivation of Q. Then there are $\tilde{\alpha}_i \in (0,1)$ such that $\mathfrak{a}(\cdot \cap A_i) + \alpha\mathfrak{a}_i = \tilde{\alpha}_i\mathfrak{a}'_i$. We set $\tilde{\alpha} := \max\{\tilde{\alpha}_1, \tilde{\alpha}_2\} \in (0,1)$ and obtain $\mathfrak{a}(\cdot \cap A_i) + \alpha\mathfrak{a}_i \leq \tilde{\alpha}\mathfrak{a}'_i$ for both $i = 1, 2$. Let us first consider the case $\mathfrak{b} \leq \mathfrak{a}$. Since our construction yields $\mathfrak{a}'_i = \mathfrak{a}(\cdot \cap A_i) + \tilde{\mathfrak{a}}_i \not\leq \mathfrak{a}$, there is a $C_0 \in \mathcal{B}$ with $\mathfrak{a}'_i(C_0) > \mathfrak{a}(C_0)$. This implies $\tilde{\alpha}\mathfrak{a}'_i(C_0) \geq \mathfrak{a}(C_0 \cap A_i) + \alpha\mathfrak{a}_i(C_0) > \mathfrak{a}(C_0 \cap A_i) \geq \mathfrak{b}(C_0 \cap A_i)$, i.e. $\mathfrak{b} \not\geq \tilde{\alpha}\mathfrak{a}'_i$. Consequently, it remains to consider the case $\mathfrak{b} \not\leq \mathfrak{a}$. By (b) and $\operatorname{supp} \mathfrak{b} \subset \operatorname{supp} P' = A$ there is a $\beta \in (0,1]$ with $\mathfrak{b}(\cdot \cap B) = \mathfrak{a}(\cdot \cap B) + \beta\mathfrak{b}(\cdot \cap B)$. Then

$$\beta\mathfrak{b} = \beta\mathfrak{b}(\cdot \cap B) = \mathfrak{b}(\cdot \cap B) - \mathfrak{a}(\cdot \cap B) \leq P'(\cdot \cap B) - \mathfrak{a}(\cdot \cap B) = P(\cdot \cap B) \leq P\,,$$

and since $\beta\mathfrak{b} \in \mathcal{Q}_P(A_1 \cup A_2)$ we obtain $\beta\mathfrak{b} \not\geq \alpha\mathfrak{a}_i$ for $i = 1, 2$. Hence there is an event $C_0 \subset \operatorname{supp} \mathfrak{b}$ with $\beta\mathfrak{b}(C_0) < \alpha\mathfrak{a}_i(C_0)$, which yields $\mathfrak{b}(C_0 \cap A_i) = \mathfrak{a}(C_0 \cap A_i \cap B) + \beta\mathfrak{b}(C_0 \cap A_i) < \mathfrak{a}(C_0 \cap A) + \alpha\mathfrak{a}_i(C_0 \cap A_i) \leq \tilde{\alpha}\mathfrak{a}'_i(C_0 \cap A_i)$, i.e. $\mathfrak{b} \not\geq \tilde{\alpha}\mathfrak{a}'_i$. ■

Proof of Theorem 21: For a $P \in \bar{\mathcal{S}}(\mathcal{A})$ and a P-adapted isomonotone sequence $(Q_n, F_n) \nearrow P$ we define $c_{\mathcal{A}}(P) :=_P \lim_{n\to\infty} s(F_n)$, which is possible by Theorem 20. By Proposition 19 we then now that $c_{\mathcal{A}}(Q) = c(Q)$ for all $Q \in \mathcal{Q}$, and hence $c_{\mathcal{A}}$ satisfies the Axiom of BaseMeasureClustering. Furthermore, $c_{\mathcal{A}}$ is obviously structured and scale-invariant, and continuity follows from Theorem 20.

To check that $c_{\mathcal{A}}$ is disjoint-additive, we fix $P_1, \ldots, P_k \in \mathcal{P}_{\mathcal{A}}$ with pairwise $\perp$-disjoint supports and let $(Q_n^i, F_n^i) \nearrow P_i$ be P_i-adapted isomonotone sequences of simple measures. We set $Q_n := Q_n^1 + \cdots + Q_n^k$ and $P := P_1 + \ldots + P_k$. By Lemma 41 Q_n is simple on $F_n := F_n^1 \cup \cdots \cup F_n^k$ and P-adapted, and $\mathcal{A}$ is P-subadditive. Moreover, we have $Q_n \nearrow P$ and $s(F_n) = \bigcup_i s(F_n^i)$ inherits monotonicity as well. Therefore $(Q_n, F_n) \nearrow P$ is P-adapted and $\lim s(F_n) = \bigcup_i \lim s(F_n^i)$ implies disjoint-additive.

To check BaseAdditivity we fix a $P \in \mathcal{P}_{\mathcal{A}}$ and a base measure $\mathfrak{a}$ with $\operatorname{supp} P \subset \mathfrak{a}$. Moreover, let $(Q_n, F_n) \nearrow P$ be a P-adapted sequence. Let $Q'_n := \mathfrak{a} + Q_n$ and $P' := \mathfrak{a} + P$. Then by Lemma 42 Q'_n is simple on $F'_n := \{A\} \cup F_n$ and P'-adapted, and $\mathcal{A}$ is P'-subadditive. Furthermore we have $(Q'_n, F'_n) \nearrow P'$ and therefore we find $P' \in \mathcal{P}_{\mathcal{A}}$ and $\lim s(F'_n) = s(\{A\} \cup \lim s(F_n))$.

For the uniqueness we finally observe that Theorem 8 together with the Axioms of Additivity shows equality on $\mathcal{S}(\mathcal{A})$ and the Axiom of Continuity in combination with Theorem 20 extends this equality to $\mathcal{P}_{\mathcal{A}}$. ■

5.2.3 Proof of Theorem 23

Lemma 43 *Let $\mu \in \mathcal{M}_\Omega^\infty$, and consider $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp)$.*

(a) If $A, A' \in \mathcal{A}$ with $A \subset A'$ μ-a.s. then $A \subset A'$.

(b) Let $P \in \mathcal{M}_\Omega$ such that $\mathcal{A}$ is P-subadditive and P has a μ-density f that is of $(\mathcal{A}, \mathcal{Q}, \perp)$-type with a dense subset Λ such that $s(F_{f,\Lambda})$ is finite. For all $\lambda \in \Lambda$ and all $A_1, \ldots, A_k \in \mathcal{A}$ with $A_1 \cup \ldots \cup A_k \subset \{f > \lambda\}$ μ-a.s. there is $B \in \mathcal{A}$ with $A_1 \cup \ldots \cup A_k \subset B$ pointwise and $B \subset \{f > \lambda\}$ μ-a.s.

Proof of Lemma 43: *(a).* Let $A, A' \in \mathcal{A}$ with $A \subset A'$ μ-a.s. and let $x \in A$. Now $B := A \setminus A'$ is relative open in A and if it is non-empty then $\mu(B) > 0$ since A is a support set. Since by assumption $\mu(B) = 0$ we have $B = \emptyset$.
(b). Since $H := \{f > \lambda\} \in \bar{\mathcal{A}}$ there is an increasing sequence $C_n \nearrow H$ of base sets. Let $d\mathfrak{b}_n := \lambda 1_{B_n} d\mu \in \mathcal{Q}_P$. For all $i \le k$ eventually $B_n \infty_\mu A_i$, so there is a n s.t. B_n is connected to all of them. By P-subadditivity between $\mathfrak{b}_n$ and $\lambda 1_{A_1}\, d\mu, \dots, \lambda 1_{A_k}\, d\mu$ there is $d\mathfrak{c} = \lambda' 1_C\, d\mu \in \mathcal{Q}_P$ that supports all of them and majorizes at least one of them. Hence $\lambda \le \lambda'$ and thus $A_1 \cup \ldots \cup A_k \subset C \subset \{f > \lambda'\} \subset \{f > \lambda\}$ μ-a.s. By (a) we are finished. ■

Lemma 44 *Let f be a density of $(\mathcal{A}, \mathcal{Q}, \perp)$-type, set $P := f\,d\mu$ and assume $\mathcal{A}$ is P-subadditive and $F_{f,\Lambda}$ is a chain. For all $k \ge 0$ and all $n \in \mathbb{N}$ let $B_n = C_1 \cup \ldots \cup C_k$ be a (possibly empty) union of base sets $C_1, \dots, C_k \in \mathcal{A}$ with $B_n \subset \{\, f > \lambda \,\}$ for all $\lambda \in \Lambda$. Then $P := f\,d\mu \in \mathcal{P}$ and there is $(Q_n, F_n) \nearrow P$ adapted where for all n F_n is a chain and $B_n \subset \min F_n$.*

Proof of Lemma 44: Let $(\lambda_n)_n \subset \Lambda$ be a dense countable subset with $\lambda_n < \rho$ and set $\Lambda_n := \{\lambda_1, \dots, \lambda_n\}$, $\Lambda_\infty := \bigcup_n \Lambda_n$. Remark that $\max \Lambda_n < \rho$ for all n, $|\Lambda_n| = n$ and $\Lambda_1 \subset \Lambda_2 \subset \ldots$. For very n we enumerate the n elements of Λ_n by $\lambda(1,n) < \ldots < \lambda(n,n)$. For every $\lambda \in \Lambda_\infty$ we let $n_\lambda := \min\{n \mid \lambda \in \Lambda_n\} \in \mathbb{N}$.

Since f is of $(\mathcal{A}, \mathcal{Q}, \perp)$-type, $H(\lambda) := \{f > \lambda\} \in \bar{\mathcal{A}}$ for $\lambda \in \Lambda$. Therefore there is $A_{\lambda,n} \in \mathcal{A}$ s.t. $A_{\lambda,n} \uparrow H(\lambda)$, where $n \ge 0$. We would like to use these $A_{\lambda,n}$ to construct Q_n, but they need to be made compatible in order that $(Q_n, F_n)_n$ becomes isomonotone. Hence we construct by induction a family of sets $A(\lambda, n) \in \mathcal{A}$, $\lambda \in \Lambda_n$, $n \in \mathbb{N}$ with the following properties:

$$A_{\lambda,n} \cup A(\lambda(i+1,n),n) \cup A(\lambda, n-1) \cup B_n \subset A(\lambda, n) \subset H(\lambda) \,\dot\cup\, N(\lambda, n), \qquad \mu(N(\lambda,n)) = 0.$$

Here $A(\lambda(i+1,n),n)$ is thought as empty if $i = n$ and similarly $A(\lambda, n-1) = \emptyset$ if $n = 1$ or $\lambda \notin \Lambda_{n-1}$. All of these involved sets C are base sets with $C \subset H(\lambda)$ and hence by Lemma 43 there is such an $A(n, \lambda)$. Since $A_{\lambda,n} \nearrow_n H(\lambda)$ we then also have $A(\lambda, n_\lambda + n) \uparrow H(\lambda)$.

Now for all n consider the chain $F_n := \{A(\lambda, n) \mid \lambda \in \Lambda_n\} \subset \mathcal{A}$ and the simple measure Q_n on F_n given by:

$$h_n := \sum_{i=1}^{n} \big(\lambda(i,n) - \lambda(i-1,n)\big) \cdot 1_{A(\lambda(i,n),n)} = \sum_{\lambda \in \Lambda_n} \lambda \cdot 1_{A(\lambda,n) \setminus \bigcup_{\lambda' > \lambda} A(\lambda',n)} \qquad (\lambda(0,n) := 0)$$

Let $x \in B$. Let

$$\Lambda_n(x) := \{\lambda \in \Lambda_n \mid x \in A(\lambda, n)\}$$

Then $h_n(x) = \max \Lambda_n(x)$. And if $x \in A(\lambda, n)$ then $x \in A(\lambda, n+1)$ so $\Lambda_n(x) \subset \Lambda_{n+1}(x)$ and we have:

$$h_n(x) = \max \Lambda_n(x) \le \max \Lambda_{n+1}(x) = h_{n+1}(x)$$

Furthermore if $\lambda \in \Lambda_n(x)$ then $x \in A(\lambda, n) \subset H(\lambda)$ implying $h(x) > \lambda$. Therefore $h_1 \le h_2 \le \cdots \le h$.

On the other hand for all $\varepsilon > 0$, since Λ_∞ is dense, there is a n and $\lambda \in \Lambda_n$ with $h(x) - \varepsilon \le \lambda < h(x)$). Then $x \in H(\lambda)$ and therefore for n big enough $x \in A(\lambda, n)$ and then:

$$h(x) \ge h_n(x) \ge \lambda \ge h(x) - \varepsilon.$$

This means $h_n(x) \uparrow h(x)$ for all $x \in B$ so we have $h_n \uparrow h$ pointwise and by monotone convergence $(Q_n, F_n) \uparrow P_0$. ■

Proof of Theorem 23: Let f be a density as supposed and set $F := s(F_{f,\Lambda})$. By assumption F is finite. If $|F| = 1$ then $F_{f,\Lambda}$ is a chain and the Theorem follows from Lemma 44 using $B_n = \emptyset$, $n \in \mathbb{N}$, in the notation of the lemma. Hence we can now assume $|F| > 1$. We prove by induction over $|F|$ that $f\,d\mu \in \bar{\mathcal{S}}(\mathcal{A})$ and $c(f\,d\mu) =_\mu s(F_{f,\Lambda})$ and assume that this is true for all f' with level forests $|s(F_{f',\Lambda'}| < |F|$. For readability we first handle the case that F is not a tree.

Assume that F has two or more roots $A_1, \dots, A_k$ with $k = k(0)$. Denote by $f_i := f\big|_{A_i}$ the corresponding densities, hence $f = f_1 + \ldots + f_k$, and set $F_i := s(F_{f_i,\Lambda}) = F\big|_{\subset A_i}$ and $P_i := f_i\,d\mu$. We cannot use DisjointAdditivity, because separation of the A_i does not imply separation of the supports. Hence we have to construct a P-adapted isomonotone sequence $(Q_n, F_n) \nearrow P$. Since $F = F_1 \dot\cup \ldots \dot\cup F_k$ we have $|F_i| < |F|$ and hence by induction assumption for all $i \le k$ we have $c(P_i) = F_i$, and there is an isomonotone P_i-adapted sequence $(Q_{i,n}, F_{i,n}) \nearrow P_i$. For $Q_n := Q_{1,n} + \ldots + Q_{k,n}$ and $F_n := F_{1,n} \cup \ldots \cup F_{k,n}$ it is clear that $(Q_n, F_n) \nearrow P$ is isomonotone. Let $\mathfrak{b} \in \mathcal{Q}_P$ and $B := \operatorname{supp} \mathfrak{b}$. We show that this is ∞_μ-connected to exactly one A_i. There is $\beta > 0$ s.t. $d\mathfrak{b} = \beta 1_B d\mu$ and $\beta 1_B \le f$ μ-a.s. Now let $\lambda \in \Lambda$ with $\lambda < \beta$ and $\lambda < \inf \{ \lambda' \in \Lambda \mid k(\lambda') \ne k(0) \}$. Because for all $\lambda \in \Lambda$ also the closures of clusters are $\perp$-separated we have

$$B \subset \overline{H_f(\lambda)} = \overline{B_1(\lambda))} \overset{\perp}{\cup} \ldots \overset{\perp}{\cup} \overline{B_k(\lambda))}.$$

By connectedness there is a unique $i \le k$ with $B \subset \overline{B_i(\lambda)}$ and by monotonicity $B \perp \overline{B_j(\lambda)}$ for all $i \ne j$. Since this holds for all $\lambda \in \Lambda$ small enough and Λ is dense, this means that B is ∞_μ-connected to exactly i. Using this, P-adaptedness of Q_n is inherited from P_i-adaptedness of $Q_{i,n}$. Therefore $P = \lim_n Q_n \in \mathcal{P}$ and $c(P) = F$.

Now assume that F is a tree. Since $|F| > 1$ there are direct children $A_1, \dots, A_k$ of the root in the structured forest F with $k \ge 2$. Let $\rho := \inf\{\lambda \in \Lambda \mid k(\lambda) \ne 1\}$. Since F is a tree, $\rho > 0$. Let $f_0(\omega) := \min\{\rho, f(\omega)\}$ and $f'(\omega) := \max\{0, f(\omega) - \rho\}$ for all $\omega \in \Omega$, and set $dP_0 := f_0\,d\mu$ and $dP' := f'\,d\mu$. Then $P = P_0 + P'$ is split into a *podest* corresponding to the root and its chain and the density corresponding to the children. We set $\Lambda' := \{ \lambda - \rho \mid \lambda \in \Lambda, \lambda > \rho \}$. Then $|F_{f',\Lambda'}| = |F| - 1$ and by induction assumption there is $(Q'_n, F'_n) \uparrow P'$ adapted. Set $B_n := \mathbb{G}F'_n$ and $B := \bigcup B_n$. Then by Lemma 44 there is $(Q_n, F_n) \nearrow P_0$ adapted, which is given by a density h_n.

Now there might be a gap $\varepsilon_n := \rho - \sup h_n > 0$. By construction $\varepsilon_n \to 0$ but to be precise we let

$$\tilde{Q}_n := Q'_n + \sum_{A \in \max F'_n} \varepsilon_n \cdot 1_A\,d\mu.$$

This is still a simple measure on F'_n and therefore $(Q_n + \tilde{Q}_n, F_n \cup F'_n) \nearrow P$. We have to show P-adapted:

Grounded: Is fulfilled, since we consider trees at the moment.

Fine: Let $C_1, \dots, C_k \in F_n \cup F'_n$ be direct siblings. Then $C_1, \dots, C_k \in F'_n$ because F_n is a chain. If they are contained in one of the roots of F'_n fineness is inherited from adaptedness of Q'_n. Else they are the roots of F'_n. Let $\mathfrak{a} = \alpha 1_A d\mu \in \mathcal{Q}_P$ be a basic measure that $\perp\!\!\!\perp_P$-intersects say C_1 and C_2. Then is clear that $\alpha \le \rho$ and by P-subadditivity fineness is granted.

Motivated Let $C, C' \in F_n \cup F'_n$ be direct siblings. Then again $C, C' \in F'_n$. If they are contained in one of the roots of F'_n motivatedness is inherited from adaptedness of Q'_n. Else they are the roots of F'_n. Let $\mathfrak{a} = \alpha 1_A d\mu \in \mathcal{Q}_P$ be a base measure that supports $C_1 \cup C_2$. Again it is clear that $\alpha \le \rho$ and hence it cannot majorize neither the level of C nor the one of C'. ■

Proof of Proposition 24: Since f is continuous, all $H_f(\lambda)$ are open and it is the disjoint union of its open connected components. We show any connected component contains at least one of the $\hat{x}_1, \dots, \hat{x}_k$. To this end let $\lambda_0 \ge 0$ and B_0 be a connected component of $H_f(\lambda_0)$ (then $B_0 \ne \emptyset$). Because Ω is compact, so is the closure $\bar{B}_0$, and hence the maximum of f on $\bar{B}_0$ is attained at some $y_0 \in \bar{B}_0$. Since there is $y_1 \in B_0$ we have $f(y_0) \ge f(y_1) > \lambda$ we have $y_0 \in H_f(\lambda)$. Now $H_f(\lambda)$ is an open set, so y_0 is an inner point of this open set, and we know $y_0 \in \bar{B}_0$, therefore $y_0 \in B_0$. Therefore $y_0 \in B_0$ is a local maximum.

Hence for all λ there are at most k components and f is of $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp_\emptyset)$-type. The generalized structure $\tilde{s}(F_f)$ is finite, since there are only k leaves.

Now, fix for the moment a local maximum $\hat{x}_i$. Since $\hat{x}_i$ is a local maximum, there is ε_0 s.t. $f(y) \le f(\hat{x}_i)$ for all y with $d(y, \hat{x}_i) < \varepsilon_0$. For all $\varepsilon \in (0, \varepsilon_0)$ consider the sphere

$$S_\varepsilon(\lambda) := \left\{ y \in \Omega \colon f(y) \ge \lambda \text{ and } d(y, \hat{x}_i) = \varepsilon \right\}.$$

Since Ω is compact and $S_\varepsilon(\lambda)$ is closed, it is also compact. So as $\lambda \uparrow f(\hat{x}_i)$ the $S_\varepsilon(\lambda)$ is a monotone decreasing sequence of compact sets. Assume that all $S_\varepsilon(\lambda)$ were non-empty: Let $y_n \in S_{\varepsilon_0/(n+1)}(\lambda)$ then $(y_n)_n$ is a sequence in the compact set $S_{\varepsilon_0/2}(\lambda)$, hence there would be a subsequence converging to some y_ε. This subsequence eventually is in every $S_{\varepsilon_0/(n+1)}$ and hence $y_\varepsilon \in \bigcap_{\lambda < f(\lambda)} S_\varepsilon(\lambda)$, so this would be non-empty. This means that $f(y_\varepsilon) \ge f(\hat{x}_i)$. On the other hand, since $\varepsilon < \varepsilon_0$ we have $f(y_\varepsilon) \ge f(\hat{x}_i)$. Therefore all y_ε are local maxima, yielding a contradiction to the assumption that there are only finitely many. Hence for all ε, $S_\varepsilon(\lambda) = \emptyset$ for all $\lambda \in (\lambda_\varepsilon, f(\hat{x}_i)$. From this follows, that all local maxima have from some point on their own leaf in F_f. Therefore there is a bijection $\psi \colon \{\hat{x}_1, \dots, \hat{x}_k\} \to \min c(P)$ s.t. $\hat{x}_i \in \psi(\hat{x}_i)$.

Lastly, we need to show that also the closures of the connected components are separated, to verify the conditions of Theorem 23. We are allowed to exclude a finite set of levels, in this case the levels $\lambda_1, \dots, \lambda_m$ at which $\lambda \mapsto k(\lambda) \in \mathbb{N}$ changes. Consider $0 \le \lambda_0 < \lambda_1$ s.t. for all $\lambda \in (\lambda_0, \lambda_1)$ $k(\lambda)$ stays constant. Set $\tilde{\lambda} := \frac{\lambda_0 + \lambda}{2} \in (\lambda_0, \lambda)$. Now let A, A' be connected components of $H_f(\lambda)$ and let B, B' be the connected components of $H_f(\tilde{\lambda})$ with $A \subset B$ and $A' \subset B'$. First we show $\bar{A} \subset B$: let $y_0 \in \bar{A}$. Then there is $(y_n) \subset A$ with $y_n \to y_0$. Because f is continuous we have

$$\lambda < f(y_n) \to f(y_0) \ge \lambda > \tilde{\lambda}$$

and hence $y_0 \in B$. Similarly we have $\bar{A}' \subset B'$ and $B \perp_\emptyset B'$ implies $\bar{A} \perp_\emptyset \bar{A}'$. ■

5.3 Proofs for Section 4

Lemma 45 *Let A, A' be closed, non-empty, and (path-)connected. Then:*

$$A \cup A' \text{ is (path-)connected} \iff A \infty_\emptyset A'.$$

Therefore any finite or countable union $A_1 \cup \ldots \cup A_k$, $k \le \infty$ of such sets is connected iff the graph induced by the intersection relation is connected.

Proof of Lemma 45: Topological connectivity means that $A \cup A'$ cannot be written as disjoint union of closed non-empty sets. Hence, if $A \cup A'$ is connected, then this union cannot be disjoint. On the other hand if $x \in A \cap A' \neq \emptyset$ and $A \cup A' = B \cup B'$ with non-empty closed sets then $x \in B$ or $x \in B'$. Say $x \in B$, then still B' has to intersect A or A', say $B' \cap A \neq \emptyset$. Then both B, B' intersect A and both $C := B \cap A$ and $C' := B' \cap A$ are closed and non-empty. But since $A = C \cup C'$ is connected there is $y \in C \cap C' \subset B \cap B'$ and therefore $B \cup B'$ is not a disjoint union.

For path-connectivity: If $x \in A \cap A' \neq \emptyset$ then for all $y \in A \cup A'$ there is a path connecting x to y, so $A \cup A'$ is path-connected. On the other hand, if $A \cup A'$ is path connected then for any $x \in A$ and $x' \in A'$ there is a continuous path $f\colon [0,1] \to A \cup A'$ connecting x to x'. Then $B := f^{-1}(A)$ and $B' := f^{-1}(A')$ are closed and non-empty, and $B \cup B' = [0,1]$. Since $[0,1]$ is topologically connected there is $y \in B \cap B'$ and so $f(y) \in A \cap A'$. ■

Proof of Example 1: Reflexivity and monotonicity are trivial for all the three relations. *Disjointness*: Stability is trivial and connectedness follows from Lemma 45 and from the observation:

$$A \subset B_1 \overset{\perp_\emptyset}{\cup} \ldots \overset{\perp_\emptyset}{\cup} B_k \qquad \Rightarrow \qquad A = (A \cap B_1) \overset{\perp_\emptyset}{\cup} \ldots \overset{\perp_\emptyset}{\cup} (A \cap B_k)$$

τ-separation: Connectedness follows from the definition of τ-connectedness. For stability let $A_n \uparrow_n A$ and $A_n \perp_\tau B$ for $n \in \mathbb{N}$ and observe

$$d(A,B) = \sup_{x \in A} d(x,B) = \sup_{n \in \mathbb{N}} \sup_{x \in A_n} d(x,B) = \sup_{n \in \mathbb{N}} d(A_n,B) \geq \tau.$$

Linear Separation: Connectedness follows from the condition on $\mathcal{A}$ since $A \subset B_1 \overset{\perp_\ell}{\cup} \ldots \overset{\perp_\ell}{\cup} B_k$ implies $A = A \cap B_1 \overset{\perp_\ell}{\cup} \ldots \overset{\perp_\ell}{\cup} A \cap B_k$. To prove stability let $A_n \uparrow_n A$ and $A_n \perp_\ell B$ for $n \in \mathbb{N}$. Observe that

$$v \mapsto \sup\{\alpha \in \mathbb{R} \mid \langle\, v \mid a \,\rangle \leq \alpha \forall a \in A\}$$

is continuous and the same holds for the upper bound for the α. Hence for each n and any vector $v \in H$ with $\langle\, v \mid v \,\rangle = 1$ there is a compact, possibly empty interval $I_n(v)$ of α fulfilling the separation along v. Since by assumption the unit sphere is compact so is the semi-direct product $I_n := \{(v, \alpha) \mid \alpha \in I_n(v)\}$. Since $I_n \neq \emptyset$ and $I_n \supset I_{n+1}$ is a monotone limit of non-empty compact sets, the limit $\bigcap_n I_n$ is non-empty. ■

Lemma 46 *Let $\mu \in \mathcal{M}_\Omega^\infty$. If $\mathcal{C} \subset \mathcal{K}(\mu)$ then $\mathbb{C}_{\perp\!\!\perp}(\mathcal{C}) \subset \mathcal{K}(\mu)$.*

Proof of Lemma 46: Let $A = C_1 \cup \ldots \cup C_k \in \mathbb{C}_{\perp\!\!\!\perp}(\mathcal{C})$ then:

$$\operatorname{supp} 1_A \, d\mu = \operatorname{supp}(1_{C_1} + \cdots + 1_{C_k}) \, d\mu = C_1 \cup \ldots \cup C_k = A. \qquad \blacksquare$$

Lemma 47 *Let $\mathcal{C} \subset \mathcal{B}$ be a class of non-empty closed sets. We assume the following generalized stability: If $B \in \mathcal{B}$ and $A_1, \ldots, A_k \in \mathcal{C}$ form a connected subgraph of $\mathcal{G}_{\perp\!\!\!\perp}(\mathcal{C})$:*

$$A_i \perp\!\!\!\perp B \quad \forall i \le k \implies A \perp\!\!\!\perp B.$$

Then $\mathbb{C}_{\perp\!\!\!\perp}(\mathcal{C})$ is $\perp\!\!\!\perp$-intersection additive. Furthermore the monotone closure $\overline{\mathbb{C}_{\perp\!\!\!\perp}(\mathcal{C})}$ is

$$\bar{\mathbb{C}}_{\perp\!\!\!\perp}(\mathcal{C}) := \{\, C_1 \cup C_2 \cup \ldots \mid C_1, C_2, \ldots \in \mathcal{C} \text{ and the graph } \mathcal{G}_{\perp\!\!\!\perp}(\{C_1, C_2, \ldots\}) \text{ is connected} \}$$

Proof of Lemma 47: Let $A = C_1 \cup \ldots \cup C_n, A' = C'_1 \cup \ldots \cup C'_{n'} \in \mathbb{C}(\mathcal{C})$ with $A \not\perp\!\!\!\perp A'$. If for all $j \le n'$ we had $C'_j \perp\!\!\!\perp A$ then by assumption $A' \perp\!\!\!\perp A$ and therefore there has to be $j \le n'$ with $C'_j \not\perp\!\!\!\perp A$. By the same argument there then is $i \le n$ with $C_i \not\perp\!\!\!\perp C_j$. Therefore the intersection graph on $C_1, \ldots, C_n, C'_1, \ldots, C'_{n'}$ is connected and

$$A \cup A' = C_1 \cup \ldots \cup C_n \cup C'_1 \cup \ldots \cup C'_{n'} \in \mathbb{C}(\mathcal{C}).$$

Let $B \in \overline{\mathbb{C}(\mathcal{C})}$ and $A_1, A_2, \ldots \in \mathbb{C}(\mathcal{C})$ with $A_n \uparrow B$. Then for all n we have $A_n = C_{n1} \cup \ldots \cup C_{nk(n)}$ with $C_{nj} \in \mathcal{C}$ and their intersection graph is connected. Since $A_n \subset A_{n+1}$ for all C_{nj} there is j' with $Cnj \subset C_{(n+1),j'}$ which even gives $C_{nj} \not\perp\!\!\!\perp C_{(n+1)j'}$. Hence, the family $\{C_{nj}\}_{n,j}$ being countable can be enumerated $\tilde{C}_1, \tilde{C}_2, \ldots$ s.t. for all m there is $i(m) < m$ with $C_m \not\perp\!\!\!\perp C_{i(m)}$. Therefore for all m, the intersection graph on $\tilde{C}_1, \ldots, \tilde{C}_m$ is connected and hence

$$\tilde{A}_m := \tilde{C}_1 \cup \ldots \cup \tilde{C}_m \in \mathbb{C}(\mathcal{C}).$$

And we see that $\bigcup_m \tilde{A}_m \in \bar{\mathbb{C}}(\mathcal{C})$ and therefore

$$B = \bigcup_n A_n = \bigcup_{nj} C_{nj} = \bigcup_m \tilde{C}_m \in \bar{\mathbb{C}}(\mathcal{C}).$$

Now let $B \in \bar{\mathbb{C}}(\mathcal{C})$ and $B = \bigcup_n C_n$ with $C_n \in \mathcal{C}$ and s.t. the intersection graph on $C_1, C_2, \ldots$ is connected. By Zorn's Lemma it has a spanning tree. Since there are at most countable many nodes, one can assume that this tree is locally countable and therefore there is an enumeration of the nodes $C_{n(1)}, C_{n(2)}, \ldots$ s.t. they form a connected subgraph for all m. Then the intersection graph on $C_{n(1)}, \ldots, C_{n(m)}$ is connected for all m and therefore $A_m := C_{n(1)} \cup \ldots \cup C_{n(m)} \in \mathbb{C}(\mathcal{C})$. $A_m \in \mathbb{C}_i(\mathcal{C}) \uparrow B$ is monotone and we have $B = \bigcup A_m \in \overline{\mathbb{C}_i(\mathcal{C})}$. $\blacksquare$

Proposition 48 *Let $\mathcal{C} \subset \mathcal{B}$ be a class of non-empty, closed events and $\perp$ a $\mathcal{C}$-separation relation. We assume the following generalized countable stability: If $B \in \mathcal{B}$ and $A_1, A_2, \ldots \in \mathcal{C}$ form a connected subgraph of $\mathcal{G}_\perp(\mathcal{C})$:*

$$A_n \perp B \quad \forall n \implies \bigcup_n A_n \perp B.$$

Then $\perp$ is a $\mathbb{C}_\perp(\mathcal{C})$-separation relation.

Proof of Proposition 48: Set $\tilde{\mathcal{A}} := \mathbb{C}_\perp$. The assumption assures $\tilde{\mathcal{A}}$-stability. We have to show $\tilde{\mathcal{A}}$-connectedness. So let $A \in \tilde{\mathcal{A}}$ and $B_1, \ldots, B_k \in \mathcal{B}$ closed with:

$$A \subset B_1 \overset{\perp}{\cup} \ldots \overset{\perp}{\cup} B_k.$$

By definition of $\mathbb{C}$ there are $C_1, \ldots, C_n \in \mathcal{C}$ with $A = C_1 \cup \ldots \cup C_n$ and s.t. the $\perp$-intersection graph on $\{C_1, \ldots, C_n\}$ is connected. For all $j \leq n$ we have $C_j \subset A \subset B_1 \cup \ldots \cup B_k$ and by $\mathcal{C}$-connectedness there is $i(j) \leq k$ with $C_j \subset B_{i(j)}$. Now, whenever $i(j) \neq i(j')$ since $B_{i(j)} \varpropto B_{i(j')}$ we have by monotonicity $C_j \varpropto C_{j'}$. So whenever there is an edge between C_j and $C_{j'}$ then $i(j) = i(j')$. This means that $i(\cdot)$ is constant on connected components of the graph, and hence on the whole graph. ■

Proposition 49 *Let $\mathcal{C} \subset \mathcal{B}$ be a class of non-empty, closed events and $\perp$ a $\mathcal{C}$-separation relation with the following alternative $\mathbb{C}_\perp(\mathcal{C})$-stability: For all $A_1, A_2, \ldots \in \mathcal{C}$ and $B \in \mathcal{B}$:*

$$\mathcal{G}_{\perp\!\!\perp}(\{A_1, A_2, \ldots\}) \text{ is connected and for all } n\text{: } A_n \perp B \implies \bigcup_n A_n \perp B. \tag{28}$$

Then $\perp$ is a $\mathbb{C}_\perp(\mathcal{C})$-separation relation and $\mathbb{C}_\perp(\mathcal{C})$ is $\perp$-intersection additive.

Assume furthermore $\perp\!\!\perp$ is a weaker relation ($B \perp B' \implies B \perp\!\!\perp B'$). Then $\perp$ is a $\mathbb{C}_{\perp\!\!\perp}(\mathcal{C})$-separation relation and $\mathbb{C}_{\perp\!\!\perp}(\mathcal{C})$ is $\perp\!\!\perp$-intersection additive.

Proof of Proposition 49: The first part is a corollary of Lemma 47 and Proposition 48. For the second part observe $\mathbb{C}_{\perp\!\!\perp}(\mathcal{C}) \subset \mathbb{C}_\perp(\mathcal{C})$. hence $\perp$ is also a $\mathbb{C}_{\perp\!\!\perp}(\mathcal{C})$-separation relation. But now $\mathbb{C}_{\perp\!\!\perp}(\mathcal{C})$ is only $\perp\!\!\perp$-intersection additive. ■

Proof of Proposition 26: First if $A_n \uparrow B \in \bar{\mathcal{A}}$ then for all $x, x' \in B$ there is n with $x, x' \in A_n$ and since A_n is path-connected there is a path connecting x and x' in $A_n \subset B$, so they are connected also in B.

Let O be open and path-connected. Let $(A_n)_n \subset \mathcal{A}'$ be the subsequence of all $A \in \mathcal{A}'$ with $A \subset O$. Since O is open and $\mathcal{A}'$ a neighborhood base $O = \bigcup_n A_n$. Consider the graph on the $(A_n)_n$ given by the intersection relation. Then by Zorn's Lemma there is a spanning tree, and we can assume that it is locally at most countable. Therefore there is an enumeration $A'_1, A'_2, \ldots$ such that $\{A'_1, \ldots, A'_n\}$ is a connected sub-graph for all n. By intersection-additivity hence $\tilde{A}_n := A'_1 \cup \ldots \cup A'_n \in \mathcal{A}$ and $\tilde{A}_n \uparrow O$. ■

Lemma 50 *Let $\mu \in \mathcal{M}_\Omega^\infty$ and assume there is a $B \in \mathcal{K}(\mu)$ with $dP = 1_B \, d\mu$. Assume that $(\mathcal{A}, \mathcal{Q}^{\mu,\mathcal{A}}, \perp_\mathcal{A})$ is a P-subadditive stable clustering base and $(Q_n, F_n) \uparrow P$ is adapted. Then $s(F_n) = \{A_1^n, \ldots, A_k^n\}$ consists only of roots and can be ordered in such a way that $A_i^1 \subset A_i^2 \subset \ldots$. The limit forest F_∞ then consists of the k pairwise $\perp_\mathcal{A}$-separated sets:*

$$B_i := \bigcup_{n \geq 1} A_i^n \,,$$

there is a μ-null set $N \in \mathcal{B}$ with

$$B = B_1 \overset{\perp_\mathcal{A}}{\cup} \ldots \overset{\perp_\mathcal{A}}{\cup} B_k \overset{\perp_\emptyset}{\cup} N. \tag{29}$$

Proof of Lemma 50: Once we have shown that all $s(F_n)$ only consists of their roots, the rest is a direct consequence of the isomonotonicity, and the fact that there is a μ-null set N s.t.:

$$B = \operatorname{supp} P = N \overset{\perp_\emptyset}{\cup} \bigcup_n \operatorname{supp} Q_n = B_1 \overset{\perp_\mathcal{A}}{\cup} \dots \overset{\perp_\mathcal{A}}{\cup} B_k \overset{\perp}{\cup} N.$$

Now let $A, A' \in F_n$ be direct siblings and denote by $\mathfrak{a} =, \mathfrak{a}' \leq P$ their levels in Q_n. Then there are $\alpha, \alpha' > 0$ with $\mathfrak{a} = \alpha 1_A \, d\mu$ and $\mathfrak{a}' = \alpha' 1_{A'} \, d\mu$. Now, $\mathfrak{a}, \mathfrak{a}' \leq P$ implies $\alpha 1_A, \alpha' 1_{A'} \leq 1_B$ (μ-a.s.) and hence $\alpha, \alpha' \leq 1$. Assume they have a common root $A_0 \in \max F_n$, i.e. $A \cup A' \subset A_0 \subset B$. Then $\alpha 1_A, \alpha' 1_{a'} \leq 1_{A_0} \leq 1_B$ (μ-a.s.) and hence they cannot be motivated. ■

Proof of Lemma 27: The Hausdorff-dimension is calculated in (Falconer, 1993, Corollary 2.4). Proposition 2.2 therein gives for all events $B \subset C$ and $B' \subset C'$:

$$\mathcal{H}^s(\varphi(B)) \leq c_2^s \mathcal{H}^s(B) \text{ and } \mathcal{H}^s(\varphi^{-1}(B')) \leq c_1^s \mathcal{H}^s(B').$$

We show that C' is a $\mathcal{H}^s$-support set. Let $B' \subset C'$ be any relatively open set and set $B := \varphi^{-1}(B') \subset C$. Then $B \subset C$ is open because φ is a homeomorphism. And since C is a support set we have $0 < \mathcal{H}^s(B) < \infty$. This gives

$$0 < \mathcal{H}^s(B) = \mathcal{H}^s(\varphi^{-1}(B')) \leq c_1^s \mathcal{H}^s(B') \text{ and } \mathcal{H}^s(B') = \mathcal{H}^s(\varphi(B)) \leq c_2^s \mathcal{H}^s(B) < \infty.$$

Therefore C' is a $\mathcal{H}^s$-support set. ■

Proof of Proposition 28: The proof is split into four steps: *(a).* We first show that for all $A \in \mathcal{A}$ there is a unique index $i(A)$ with $A \in \mathcal{A}^{i(A)}$. To this end, we fix an $A \in \mathcal{A}$. Then there is $i \leq m$ with $A \in \mathcal{A}^i$. Let $\mu \in \mathcal{Q}^i$ be the corresponding base measure with $\operatorname{supp} \mu = A$. Let $j \leq m$ and $\mu' \in \mathcal{Q}^j$ be another measure with $\operatorname{supp} \mu' = A$. Then $\mu(A) = 1$ and $\mu'(A) = 1$. If $j > i$ then by assumption $\mu \prec \mu'$ and this would give $\mu'(A) = 0$. If $j < i$ we have $\mu' \prec \mu$ and this would give $\mu(A) = 0$. So $i = j$.
(b). Next we show that for all $A, A' \in \mathcal{A}$ with $A \subset A'$ we have $i(A) \leq i(A')$. To this end we first observe that $A = A \cap A' = \operatorname{supp} Q_A \cap \operatorname{supp} Q_{A'}$. If we had $i > j$ then $Q_{A'} \in \mathcal{Q}^j \prec \mathcal{Q}^i \ni Q_A$ and since $Q_{A'}(A) \leq Q_{A'}(A') = 1 < \infty$ we would have $Q_A(A) = 0$. Therefore $i \leq j$.
(c). Now we show that $\perp$ is a stable $\mathcal{A}$-separation relation. Clearly, it suffices to show $\mathcal{A}$-stability and $\mathcal{A}$-connectedness. The former follows since $i(A_n)$ is monotone if $A_1 \subset A_2 \subset \dots$ by (b) and hence eventually is constant. For the latter let $A \in \mathcal{A}^i$ and $B_1, \dots, B_k \in \mathcal{B}$ closed with $A \subset B_1 \overset{\perp}{\cup} \dots \overset{\perp}{\cup} B_k$. Then since $\perp$ is an $\mathcal{A}^i$-separation relation there is $j \leq k$ with $A \subset B_j$.
(d). Finally, we show that $(\mathcal{A}, \mathcal{Q}, \perp)$ is a stable clustering base. To this end observe that fittedness is inherited from the individual clustering bases. Let $A \in \mathcal{A}^i$ and $A' \in \mathcal{A}^j$ with $A \subset A'$. Then $i \leq j$ by (b). If $i = j$ then flatness follows from flatness of $\mathcal{A}^i$. If $i < j$ then by assumption $Q_A \prec Q_{A'}$ and because $Q_A(A) = 1 < \infty$ we have $Q_{A'}(A) = 0$. ■

Proof of Proposition 29: *(a).* Let $\mathfrak{a} \leq P$ be a base measure on $A \in \mathcal{A}^i$. If $i = 1$ then $Q_A(A \cap \operatorname{supp} P_2) \leq Q_A(A) = 1$ and by $\mathcal{A}^1 \prec P_2$ we have $Q_A \prec P_2$ and hence $P_2(A \cap \operatorname{supp} P_2) = P_2(A) = 0$. Now for all events $C \in A^c$ therefore $\mathfrak{a}(C) = 0 \leq P_1(C)$ and for all $C \subset A$:

$$\mathfrak{a}(C) \leq P(C) = \alpha_1 P_1(C) + \alpha_2 P_2(C) = \alpha_1 P_1(C).$$

Now if $i = 2$ then by assumption $P_1 \prec \mathfrak{a}$ and since $0 < P_1(A \cap \operatorname{supp} P_1) < \infty$ we therefore have $\mathfrak{a}(A \cap \operatorname{supp} P_1) \leq \mathfrak{a}(\operatorname{supp} P_1) = 0$ and for all events $C \subset \Omega \setminus \operatorname{supp} P_1$ we have $\mathfrak{a}(C) \leq P(C) = \alpha_2 P_2(C)$ and for all events $C \subset \operatorname{supp} P_1$:

$$\mathfrak{a}(C) \leq \mathfrak{a}(\operatorname{supp} P_1) = 0 \leq P_1(C).$$

(b). Let $\mathfrak{a}, \mathfrak{a}' \leq P$ be base measures on $A \in \mathcal{A}^i$ and $A \in \mathcal{A}^j$ with $A \infty_{\mathcal{A}} A'$. By the previous statement we then already have $\mathfrak{a} \leq \alpha_i P_i$ and $\mathfrak{a}' \leq \alpha_j P_j$. Now, if $i = j$ then by P_i-subadditivity of $\mathcal{A}^i$ there is a base measure $\mathfrak{b} \leq P_i \leq P$ on $B \in \mathcal{A}^i$ with $B \supset A \cup A'$.

Now if $i \neq j$ consider say $i = 2$ and $j = 1$. Since $A \cap \operatorname{supp} P_2 \supset A \cap A' \neq \emptyset$ by assumption $\mathfrak{a}$ can be majorized by a base measure $\tilde{\mathfrak{a}} \leq P_2$ on $\tilde{A} \in \mathcal{A}^2$ with $\operatorname{supp} P_1 \subset \tilde{A}$ and $\tilde{\mathfrak{a}} \geq \mathfrak{a}$. The latter also gives $A \subset \tilde{A}$ and hence $\tilde{a}$ supports A and $\operatorname{supp} P_1 \supset \operatorname{supp} \mathfrak{a}'$ and $\tilde{\mathfrak{a}} \geq \mathfrak{a}$. ■

Acknowledgments

This work has been supported by DFG Grant STE 1074/2-1. We thank the reviewers and editors for their helpful comments.

Appendix A. Appendix: Measure and Integration Theoretic Tools

Throughout this subsection, Ω is a Hausdorff space and $\mathcal{B}$ is its Borel σ-algebra. Recall that a measure μ on $\mathcal{B}$ is inner regular iff for all $A \in \mathcal{B}$ we have

$$\mu(A) = \sup \left\{ \mu(K) \mid K \subset A \text{ is compact} \right\}.$$

A Radon space is a topological space such that all finite measures are inner regular. Cohn (2013, Theorem 8.6.14) gives several examples of such spaces such as *a)* Polish spaces, i.e. separable spaces whose topology can be described by a complete metric, *b)* open and closed subsets of Polish spaces, and *c)* Banach spaces equipped with their weak topology. In particular all separable Banach spaces equipped with their norm topology are Polish spaces and infinite dimensional spaces equipped with the weak topology are not Polish spaces but still they are Radon spaces. Furthermore Hausdorff measures, which are considered in Section 4.3, are inner regular (Federer, 1969, Cor. 2.10.23). For any inner regular measure μ we define the **support** by

$$\operatorname{supp} \mu := \Omega \setminus \bigcup \left\{ O \subset \Omega \mid O \text{ is open and } \mu(O) = 0 \right\}.$$

By definition the support is closed and hence measurable. The following lemma collects some more basic facts about the support that are used throughout this paper.

Lemma 51 *Let μ be an inner regular measure and $A \in \mathcal{B}$. Then we have:*

(a) If $A \perp_{\emptyset} \operatorname{supp} \mu$, then we have $\mu(A) = 0$.

(b) If $\emptyset \neq A \subset \operatorname{supp} \mu$ is relatively open in $\operatorname{supp} \mu$, then $\mu(A) > 0$.

(c) If μ' is another inner regular measures and $\alpha, \alpha' > 0$ then

$$\operatorname{supp}(\alpha\mu + \alpha'\mu') = \operatorname{supp}(\mu) \cup \operatorname{supp}(\mu')$$

(d) The restriction $\mu_{|A}$ of μ to A defined by $\mu_{|A}(B) = \mu(B \cap A)$ is an inner regular measure and $\operatorname{supp} \mu_{|A} \subset \overline{A \cap \operatorname{supp} \mu}$.

If μ is not inner regular, (d) also holds provided that Ω is a Radon space and $\mu(A) < \infty$.

Proof of Lemma 51: *(a).* We show that $A := \Omega \setminus \operatorname{supp} \mu$ is a μ-null set. Let $K \subset A$ be any compact set. By definition A is the union of all open sets $O \subset \Omega$ with $\mu(O) = 0$. So those sets form an open cover of A and therefore of K. Since K is compact there exists a finite sub-cover $\{O_1, \dots, O_n\}$ of K. By σ-subadditivity of μ we find

$$\mu(K) \le \sum_{i=1}^{n} \mu(O_i) = 0,$$

and since this holds for all such compact $K \subset A$ we have by inner regularity

$$\mu(A) = \sup_{K \subset A} \mu(K) = 0.$$

(b). By assumption there an open $O \subset \Omega$ with $\emptyset \neq A = O \cap \operatorname{supp} \mu$. Now $O \cap \operatorname{supp} \mu \neq \emptyset$ implies $\mu(O) > 0$. Moreover, we have the partition $O = A \cup (O \setminus \operatorname{supp} \mu)$ and since $O \setminus \operatorname{supp} \mu$ is open, we know $\mu(O \setminus \operatorname{supp} \mu) = 0$, and hence we conclude that $\mu(O) = \mu(A)$.
(c). This follows from the fact that for all open $O \subset \Omega$ we have

$$(\alpha\mu + \alpha'\mu')(O) = \alpha\mu(O) + \alpha'\mu'(O) = 0 \iff \mu(O) = 0 \text{ and } \mu'(O) = 0.$$

(d). The measure $\mu_{|A}$ is inner regular since for $B \in \mathcal{B}$ we have

$$\begin{aligned}\mu'(B) = \sup \{ \mu(K') \mid K' \subset B \cap A \text{ is compact} \} &\le \sup \{ \mu'(K') \mid K' \subset B \text{ is compact} \} \\ &\le \mu'(B).\end{aligned}$$

Now observe that $X \setminus \overline{A \cap \operatorname{supp} \mu} \subset X \setminus (A \cap \operatorname{supp} \mu) = (X \setminus A) \cup (X \setminus \operatorname{supp} \mu)$. For the open set $O := X \setminus \overline{A \cap \operatorname{supp} \mu}$ we thus find

$$\mu_{|A}(O) \le \mu_{|A}(X \setminus A) + \mu_{|A}(X \setminus \operatorname{supp} \mu) \le \mu(X \setminus \operatorname{supp} \mu) = 0. \qquad \blacksquare$$

Lemma 52 *Let Q, Q' be σ-finite measures.*

(a) If Q and Q' have densities h, h' with respect to some measure μ then

$$Q \le Q' \iff h \le h' \quad \mu\text{-a.s.}$$

(b) If $Q \leq Q'$ then Q is absolutely continuous with respect to Q', i.e. Q has a density function h with respect to Q', $dQ = h\,dQ'$ such that:

$$h(x) = \begin{cases} \in [0,1] & \text{if } x \in \operatorname{supp} Q' \\ 0 & \text{else} \end{cases}$$

Proof of Lemma 52: *(a).* "‘⇐"’ a direct calculation gives

$$Q(B) = \int_B h\,d\mu \leq \int_B h'\,d\mu = Q'(B).$$

and monotonicity of the integral.
For "‘⇒"’ assume $\mu(\{x : h(x) > h'(x)\}) > 0$, then

$$\int_{h>h'} h d\mu = Q(\{h > h'\}) \leq Q'(\{h > h'\}) = \int_{h>h'} h' d\mu < \int_{h>h'} h d\mu,$$

where the last inequality holds since we assume $\mu(\{x : h(x) > h'(x)\} > 0$ and again the monotonicity of the integral. Through this contradiction implies the statement.
(b). $Q \leq Q'$ means every Q'-null set is a Q-null set. Furthermore since Q' is σ-finite Q is σ-finite as well. So we can use Radon-Nikodym theorem and there is a $h \geq 0$ s.t. $dQ = h\,dQ'$. Since the complement of $\operatorname{supp} Q'$ is a Q'-null set, we can assume $h(x) = 0$ on this complement.

We have to show that $h \leq 1$ a.s. Let

$$E_n := \{h \geq 1 + \tfrac{1}{n}\} \qquad \text{and} \qquad E := \{h > 1\}.$$

Then $E_n \uparrow E$ and we have

$$Q'(E_n) \geq Q(E_n) = \int_{E_n} h\,dQ' \geq (1 + \tfrac{1}{n}) \cdot Q'(E_n),$$

which implies $Q'(E_n) = 0$ for all n. Therefore $Q'(E) = \lim_n Q'(E_n) = 0$. ■

Lemma 53 *(a) Let $Q_n \uparrow P$, $A := \operatorname{supp} P$ and $B := \bigcup_n \operatorname{supp} Q_n$. Then $B \subset A$ and $P(B \setminus A) = 0$.*

(b) Assume Q is a finite measure and $Q_1 \leq Q_2 \leq \ldots \leq Q$ and let the densities $h_n := \frac{dQ_n}{dQ}$. Then $h_1 \leq h_2 \leq \ldots \leq 1$ Q-a.s. Furthermore, the following are equivalent:

(i) $Q_n \uparrow Q$

(ii) $h_n \uparrow 1$ Q-a.s.

(iii) $h_n \uparrow 1$ in L^1.

Proof of Lemma 53:

(a) Since $Q_n \leq P$ we have $\operatorname{supp} Q_n \subset A$ and therefore $B \subset A$. Because of $(A \setminus B) \cap \operatorname{supp} Q_n = \emptyset$ and the convergence we have for all n

$$P(A \setminus B) = \lim_{n\to\infty} Q_n(A \setminus B) = 0.$$

(b) By the previous lemma we have $h_1 \leq h_2 \leq \cdot \leq 1$ Q-a.s.

(i) ⇒ (ii): Since $(h_n)_n$ is monotone Q-a.s. it converges Q-a.s. to a limit $h \leq 1$. Let

$$E_n := \{h \leq 1 - \tfrac{1}{n}\} \qquad \text{and} \qquad E := \{h < 1\}.$$

Then $E_n \uparrow E$ and we have by the monotone convergence theorem:

$$Q_m(E_n) = \int_{E_n} h_m \, dQ \xrightarrow[m\to\infty]{} \int_{E_n} h \, dQ \leq (1 - \tfrac{1}{n})\, Q(E_n)$$

But since $Q_m(E_n) \uparrow_m Q(E_n)$ this means that $Q(E_n) = 0$ for all n and therefore $Q(E) = \lim_n Q(E_n) = 0$.

(ii) ⇒ (iii): This follows from monotone convergence, because $1 \in L^1(Q)$.

(iii) ⇒ (i): For all $B \in \mathcal{B}$:

$$Q(B) - Q_n(B) = \int_B |1 - h_n| \, dQ \leq \int |1 - h_n| \, dQ \to 0$$

because of $h_n \to 1$ in L^1. ■

References

S. Ben-David. Computational Feasibility of Clustering under Clusterability Assumptions. *ArXiv e-prints*, January 2015.

J. E. Chacón. A population background for nonparametric density-based clustering. *ArXiv e-prints*, August 2014. URL `http://arxiv.org/abs/1408.1381`.

D. L. Cohn. *Measure Theory*. Birkhäuser, 2nd ed. edition, 2013.

W. Day and F. McMorris. *Axiomatic Consensus Theory in Group Choice and Biomathematics*. Society for Industrial and Applied Mathematics, 2003.

D. Defays. An efficient algorithm for a complete link method. *The Computer Journal*, 20 (4):364–366, 1977.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39(1):1–38, 1977.

W.E. Donath and A.J. Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, Sept 1973.

M. Ester, H. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *International Conference on Knowledge Discovery and Data Mining*, pages 226–231. AAAI Press, 1996.

K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, 1993.

H. Federer. *Geometric Measure Theory*. Springer, 1969.

G. Gan, C. Ma, and J. Wu. *Data clustering. Theory, algorithms, and applications*. SIAM, 2007.

John A. Hartigan. *Clustering Algorithms*. Wiley, 1975.

M. F. Janowitz and R. Wille. On the classification of monotone-equivariant cluster methods. In Cox, Hansen, and Julesz, editors, *Partitioning Data Sets: DIMACS Workshop 1993*, pages 117–142. AMS, 1995.

N. Jardine and R. Sibson. *Mathematical Taxonomy*. Wiley, 1971.

L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, 1990.

J. M. Kleinberg. An impossibility theorem for clustering. In Becker, Thrun, and Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 463–470. MIT Press, 2003.

J. Kogan. *Introduction to Clustering Large and High-Dimensional Data*. Cambridge University Press, 2007.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297. University of California Press, 1967.

M. Meilă. Comparing clusterings: An axiomatic view. In *International Conference on Machine Learning*, ICML '05, pages 577–584. ACM, 2005.

G. Menardi. A review on modal clustering. *International Statistical Review*, 2015.

B. Mirkin. *Clustering for Data Mining. A Data Recovery Approach*. Chapman & Hall/CRC, 2005.

B.G. Mirkin. On the problem of reconciling partitions. *Quantitative Sociology, International Perspectives on Mathematical and Statistical Modelling*, pages 441–449, 1975.

B.G. Mirkin. Additive clustering and qualitative factor analysis methods for similarity matrices. *Psychological Review*, 4(1):7–31, 1987.

J. Puzicha, T. Hofmann, and J. M. Buhmann. A theory of proximity based clustering: Structure detection by optimization. *Pattern Recognition*, 33:617–634, 1999.

A. Rinaldo and L. Wasserman. Generalized density clustering. *Ann. of Stat.*, 38(5):2678–2722, 2010.

R. N. Shepard and P. Arabie. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86(2):87–123, 1979.

R. Sibson. SLINK: An optimally efficient algorithm for the single-link cluster method. *The Computer Journal*, 16(1):30–34, 1973.

W. Stuetzle. Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample. *Journal of Classification*, 20(1):025–047, 2003.

U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

U. von Luxburg and S. Ben-David. Towards a statistical theory of clustering. In *PASCAL workshop on Statistics and Optimization of Clustering*, 2005.

J. H. Ward, Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

R. B. Zadeh and S. Ben-David. A uniqueness theorem for clustering. In Bilmes and Ng, editors, *Conference on Uncertainty in Artificial Intelligence, 2009*. AUAI Press, 2009.