Journal of Machine Learning Research 25 (2024) 1-40 Submitted 5/21; Revised 3/24; Published 4/24

# Sum-of-norms clustering does not separate nearby balls

**Alexander Dunlap** DUNLAP@MATH.DUKE.EDU
*Courant Institute of Mathematical Sciences*
*New York University*
*New York, NY 10012, USA*
Current address: *Duke University, Durham, NC 27708, USA*

**Jean-Christophe Mourrat** JEAN-CHRISTOPHE.MOURRAT@ENS-LYON.FR
*Ecole Normale Supérieure de Lyon and CNRS*
*Lyon, France,*
*and Courant Institute of Mathematical Sciences*
*New York University*
*New York, NY 10012, USA*

**Editor:** Ingo Steinwart

## Abstract

Sum-of-norms clustering is a popular convexification of $K$-means clustering. We show that, if the dataset is made of a large number of independent random variables distributed according to the uniform measure on the union of two disjoint balls of unit radius, and if the balls are sufficiently close to one another, then sum-of-norms clustering will typically fail to recover the decomposition of the dataset into two clusters. As the dimension tends to infinity, this happens even when the distance between the centers of the two balls is taken to be as large as $2\sqrt{2}$. In order to show this, we introduce and analyze a continuous version of sum-of-norms clustering, where the dataset is replaced by a general measure. In particular, we state and prove a local-global characterization of the clustering that seems to be new even in the case of discrete datapoints.

**Keywords:** Sum-of-norms clustering, Clusterpath, convex clustering, stochastic ball model, unsupervised learning

## 1. Introduction

### 1.1 Sum-of-norms clustering

Clustering is the task of partitioning a dataset with the aim to optimize a measure of similarity between objects in each element of the partition. Given datapoints $x_1, \dots, x_N \in \mathbf{R}^d$, one may seek to find $K$ "centers" so as to minimize the sum of the distances between each datapoint and its nearest center. This is the $K$-means problem, which can be formulated as follows: find $y_1, \dots, y_N \in \mathbf{R}^d$ that minimize

$$\sum_{n=1}^{N} |y_n - x_n|^2,$$

subject to the constraint that the set $\{y_1, \dots, y_N\}$ has cardinality $K$ (or at most $K$). Here and throughout, $|\cdot|$ denotes the Euclidean norm. However, the $K$-means problem is NP-hard in general, even when we restrict to $K = 2$ (Aloise et al., 2009) or to $d = 2$ (Mahajan et al., 2009). In this article, we focus on a particular convex relaxation of $K$-means, introduced by Pelckmans et al. (2005); Hocking et al. (2011); Lindsten et al. (2011) and called "convex clustering shrinkage," "clusterpath," or "sum-of-norms (SON) clustering," which consists in finding the points $y_1, \dots, y_N \in \mathbf{R}^d$ that minimize

$$\frac{1}{N}\sum_{n=1}^{N} |y_n - x_n|^2 + \frac{\lambda}{N^2}\sum_{k,n=1}^{N} |y_k - y_n|, \tag{1.1}$$

where $\lambda \geq 0$ is a tunable parameter. Two datapoints $x_k$ and $x_n$ are then declared to belong to the same cluster if $y_k = y_n$. In principle, varying the parameter $\lambda$ allows one to tune the number of clusters, as illustrated in Figure 1.1. One of the attractive features of SON clustering is that it produces an ordered path of partitions as we vary $\lambda$. In other words, its natural output is a hierarchy of nested partitions of the dataset (see Hocking et al., 2011; Chiquet et al., 2017, or Theorem 1.4 below).

In the last decade, rigorous guarantees on the behavior of SON clustering have been studied by several authors, including Zhu et al. (2014); Tan and Witten (2015); Chiquet et al. (2017); Panahi et al. (2017); Radchenko and Mukherjee (2017); Jiang et al. (2020); Chi and Steinerberger (2019); Jiang and Vavasis (Preprint, 2020); Sun et al. (2021); Nguyen and Mamitsuka (Preprint, 2021). Most of these works aim at the identification of sufficient conditions for SON clustering to succeed in separating clusters. Our main goal here, stated precisely in Theorem 1.1, is rather to present a seemingly simple clustering problem in which the SON clustering algorithm will typically fail. This requires us to establish necessary *and* sufficient conditions for the success of SON clustering, which we present in Subsection 1.3. We anticipate that these conditions will be useful in future studies of sum-of-norms clustering, and thus are interesting results in their own right.

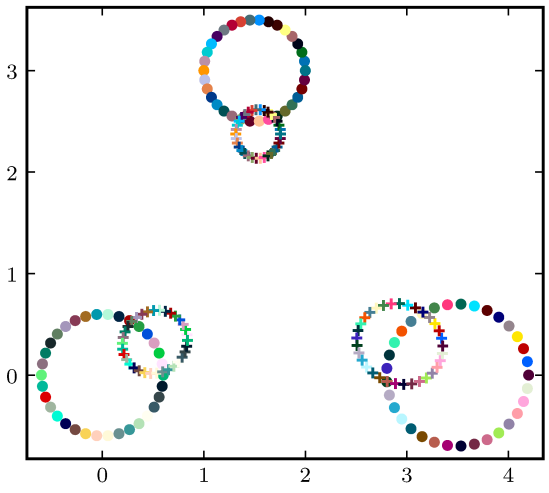
Most of our attention will be towards the analysis of the following generalization of SON clustering: given a nonzero finite Borel measure $\mu$ on $\mathbf{R}^d$ of compact support and $\lambda \geq 0$, we seek to minimize the functional $J_{\mu,\lambda} \colon L^2(\mu; \mathbf{R}^d) \to \mathbf{R}$ given by

$$J_{\mu,\lambda}(u) := \int |u(x) - x|^2 \, \mathrm{d}\mu(x) + \lambda \iint |u(x) - u(y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y). \tag{1.2}$$

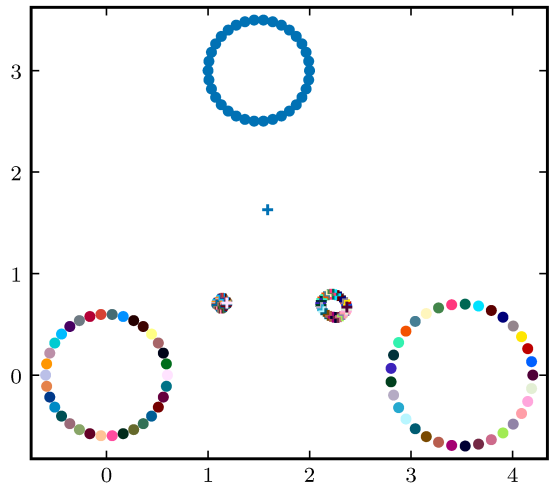
As will be explained at the beginning of Section 4, the functional $J_{\mu,\lambda}$ has a unique minimizer, which we denote by $u_{\mu,\lambda} \in L^2(\mu; \mathbf{R}^d)$. The level sets of $u_{\mu,\lambda}$ yield a partition of $\mathbf{R}^d$, up to modifications by $\mu$-null sets. One of the main general results of our paper, which seems to be new even in the discrete setting, is a local-global characterization of this minimizer, see Theorem 1.7 below. The correspondence between (1.1) and (1.2) is obtained by setting $\mu = \frac{1}{N}\sum_{n=1}^{N} \delta_{x_n}$ and $y_n = u(x_n)$.

### 1.2 The stochastic ball model

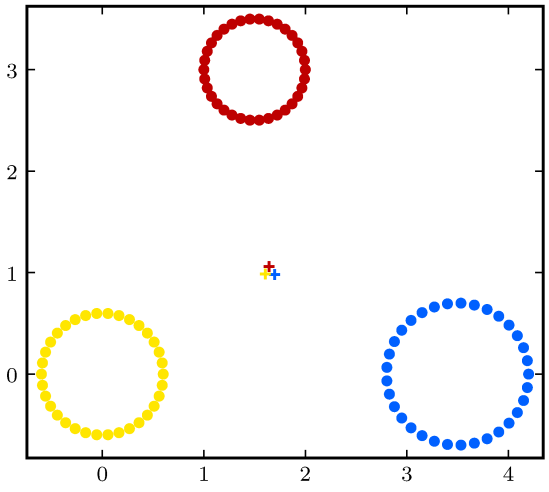
The main motivation for introducing the continuous version of SON clustering is that it allows us to uncover the asymptotic behavior of the discrete problem in (1.1) when the



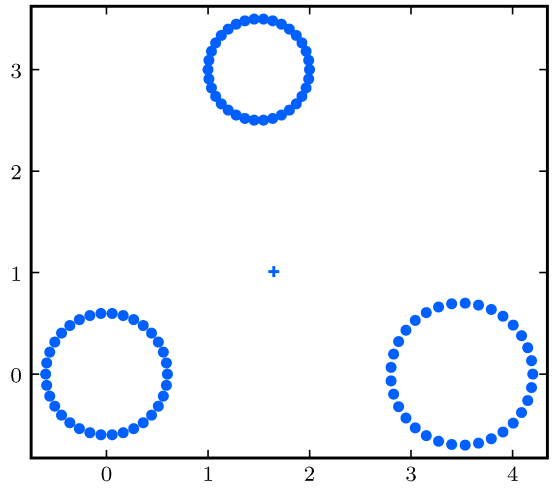

(a) $\lambda = 1.1$. Each point is in its own cluster.




(b) $\lambda = 2.4$. The points in the upper circle have merged into a single cluster, but each point in the lower two circles remains in its own cluster.




(c) $\lambda = 3.4$. Each of the circles now forms a single cluster.




(d) $\lambda = 3.6$. There is now a single cluster comprising all of the points.

Figure 1.1: The output of the clustering algorithm on $N = 100$ datapoints divided between the boundaries of three balls, for four values of $\lambda$. The filled circles represent the datapoints $x_n$, and the crosses represent the cluster representatives $y_n$. Each color represents a cluster. All figures in this paper were generated using an implementation (by the present authors) of the algorithm described in Jiang and Vavasis (Preprint, 2020). The code is available at https://github.com/ajdunlap/son-clustering-experiments.

number of datapoints $N$ becomes very large. In particular, we will study the "stochastic ball model," which has become a common testbed in the analysis of clustering algorithms, see for instance Nellore and Ward (2015); Awasthi et al. (2015); Iguchi et al. (2017); Li et al. (2020); De Rosa and Khajavirad (2022). That is, we suppose that we are given a large number of points sampled independently at random, each being distributed according to the uniform measure on the union of two disjoint balls of unit radius, and ask whether SON clustering allows us to identify the presence of the two balls. Surprisingly, we find that if $d \geq 2$ and the balls are too close to each other, then the algorithm will typically fail to do so.

In order to state this result more precisely, we need to introduce some notation. We write

$$\gamma_d \coloneqq \frac{2d+1}{2d+4} \cdot \begin{cases} \frac{(d+1)(2d)!\pi}{2^{3d}((d/2)!)^2 d!} & \text{if } d \text{ is even,} \\ \frac{(d+1)(((d-1)/2)!)^2(2d)!}{2^d (d!)^3} & \text{if } d \text{ is odd,} \end{cases} \tag{1.3}$$

so that

$$\gamma_1 = 1, \qquad \gamma_2 = \frac{45\pi}{128} \simeq 1.104\ldots, \qquad \gamma_3 = \frac{7}{6}, \tag{1.4}$$

and

$$\frac{\gamma_{d+2}}{\gamma_d} = 1 + \frac{7d+13}{(d+1)(2d+4)(2d+8)} > 1.$$

In particular, for every $d \geq 2$, we have $\gamma_d > 1$, and using Stirling's approximation, one can check that $\gamma_d$ tends to $\sqrt{2}$ as $d$ tends to infinity. We also write $B_r(x)$ for the open Euclidean ball or radius $r > 0$ centered at $x \in \mathbf{R}^d$, and $(\mathrm{e}_1, \ldots, \mathrm{e}_d)$ for the canonical basis of $\mathbf{R}^d$. We use the phrase "with high probability" as shorthand for "with probability tending to 1 as $N$ tends to infinity".

**Theorem 1.1.** *There exists a $\lambda_{\mathrm{c}} \in (0, \infty)$ such that the following holds. Let $r \in [1, \gamma_d)$, $\mu$ be the uniform probability measure on $B_1(-r\mathrm{e}_1) \cup B_1(r\mathrm{e}_1) \subseteq \mathbf{R}^d$, $(X_n)_{n \in \mathbf{N}}$ be independent random variables with law $\mu$, and for every integer $N \geq 1$, define the empirical measure*

$$\mu_N \coloneqq \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n}. \tag{1.5}$$

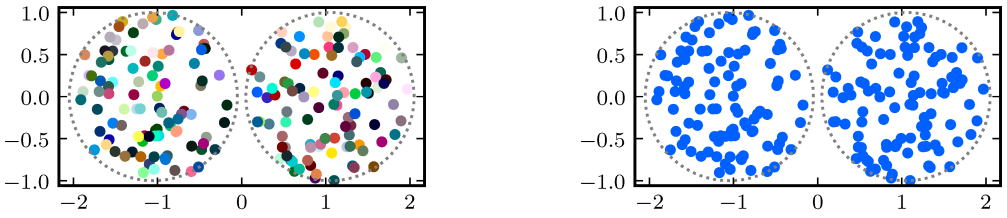
1. *If $\lambda > \lambda_{\mathrm{c}}$, then with high probability, the range of $u_{\mu_N,\lambda}$ is a singleton.*
2. *If $\lambda < \lambda_{\mathrm{c}}$, then there exist $\xi, \eta > 0$ (not depending on $N$) such that, with high probability, one can find $A_N^{(1)}, A_N^{(2)}, A_N^{(3)} \subseteq \{1, \ldots, N\}$, each of cardinality at least $\xi N$ and satisfying, for every $i \neq j \in \{1, 2, 3\}$,*

$$\forall k \in A_N^{(i)},\ \forall \ell \in A_N^{(j)}, \quad |u_{\mu_N,\lambda}(X_k) - u_{\mu_N,\lambda}(X_\ell)| \geq \eta.$$

*In particular, with high probability, the range of $u_{\mu_N,\lambda}$ contains at least three points.*

*In fact, we can take $\lambda_{\mathrm{c}} = \lambda_1(\mu)$, with the latter quantity defined in (1.9) below.*

Theorem 1.1 does not describe the behavior of $u_{\mu_N,\lambda}$ when $\lambda = \lambda_{\mathrm{c}}$, or when $\lambda = \lambda_{\mathrm{c}} + o(1)$ as $N \to \infty$. But at the very least, Theorem 1.1 shows that the detection of two nearby balls by means of SON clustering will be particularly brittle. In contrast, we show in




(a) $\lambda = 2.0$ (b) $\lambda = 2.15$

Figure 1.2: Sum-of-norms clustering of the stochastic ball model with $N = 200$ datapoints drawn from $B(-1.05\mathrm{e}_1, 1) \cup B(1.05\mathrm{e}_1, 1)$. The balls from which the points are drawn are outlined in dotted grey lines. When $\lambda = 2.0$, there are many clusters, but when $\lambda$ is slightly larger ($\lambda = 2.15$), there is just one large cluster. Theorem 1.1 tells us that (since $1.05 < \gamma_2$), in the limit as $N \to \infty$, there will be *no* open interval of values of $\lambda$ for which there are exactly two clusters.

Proposition 6.5 that, using the notation of Theorem 1.1, if $r > 2^{1-\frac{1}{d}}$ and $\lambda \in (2^{2-\frac{1}{d}}, 2r)$, then with high probability, the level sets of $u_{\mu_N,\lambda}$ are the sets $\{X_n, n \leq N\} \cap B_1(-r\mathrm{e}_1)$ and $\{X_n, n \leq N\} \cap B_1(r\mathrm{e}_1)$.

In a nutshell, SON clustering fails to separate balls if $r < \gamma_d$, while it succeeds if $r > 2^{1-\frac{1}{d}}$; see Figure 1.2 for an illustration of this failure when $r < \gamma_d$. We expect neither of these two bounds to be sharp. In view of Corollary 2.4 and of the fact that points in a high-dimensional ball tend to concentrate near the boundary, we conjecture that in the limit of high dimensions, the threshold separating these two regimes converges to $\sqrt{2}$. Since $\lim_{d\to\infty} \gamma_d = \sqrt{2}$, this would indicate that the lower bound on this threshold provided by Theorem 1.1 is asymptotically sharp.

Theorem 1.1 demonstrates in particular that the cardinality of the partition produced by the SON clustering algorithm can be very sensitive to small changes in the parameter $\lambda$. While Theorem 1.1 only asserts that the cardinality of the partition quickly moves from 1 to at least 3 as we only slightly vary $\lambda$, we expect that the partition quickly shatters into many more than just three pieces. This is also what we observe in simulations, see Figure 1.2. We view this phenomenon as a possible theoretical confirmation of the empirical observations of Chiquet et al. (2017) and Nguyen and Mamitsuka (Preprint, 2021). We refer in particular to Figure 1(b) of Chiquet et al. (2017) and the general observation that the tree structures produced by the (unweighted) SON clustering algorithm are often difficult to interpret ("unbalanced"), since the root of the tree very quickly splits into way too many components. (Chiquet et al., 2017, also underline that among these many components, some will be much larger than others.) See also Figure 4 of Nguyen and Mamitsuka (Preprint, 2021).

### 1.3 The structure of clusters

Theorem 1.1 will be proved as a consequence of more general structural results on the clusters obtained by the sum-of-norms clustering algorithm. We foresee these results being useful in more general circumstances as well, and proceed to describe them now.

There are two special cases of clustering that will be particularly important in our discussion. We record them in the following definition.

**Definition 1.2.** *Let $\mu$ be a finite Borel measure of compact support and $\lambda \geq 0$.*

1. *We say that $\mu$ is* $\lambda$-cohesive *if there is a constant $c$ such that $u_{\mu,\lambda} \equiv c$, $\mu$-a.e.*
2. *We say that $\mu$ is* $\lambda$-shattered *if there is a measurable injection $u \colon \mathbf{R}^d \to \mathbf{R}^d$ such that $u_{\mu,\lambda} = u$, $\mu$-a.e.*

Note that if $\operatorname{supp} \mu$ consists of a single point (or if $\mu$ is the zero measure), then $\mu$ is both $\lambda$-shattered and $\lambda$-cohesive for all $\lambda \geq 0$.

Recall that the level sets of $u_{\mu,\lambda}$ define a partition of $\mathbf{R}^d$ up to a $\mu$-null modification. We think of this partition as a clustering of the support of $\mu$. To discuss these clusters, we will often use the notation

$$V_{u,x} \coloneqq u^{-1}(u(x))$$

for the cluster containing $x$. The set $V_{u,x}$ is a Borel subset of $\mathbf{R}^d$ defined up to a $\mu$-null modification. Thus, saying that $\mu$ is $\lambda$-cohesive is equivalent to saying that $V_{u_{\mu,\lambda},x} = \mathbf{R}^d$ (up to a $\mu$-null modification) for $\mu$-a.e. $x \in \mathbf{R}^d$. If $\mu$ is $\lambda$-shattered, then $\mu(V_{u_{\mu,\lambda},x} \setminus \{x\}) = 0$ for $\mu$-a.e. $x \in \mathbf{R}^d$, and in fact, by Proposition 1.6 below, the converse holds as well.

The following theorem, proved in Section 5, extends to the continuous setting results proved in the discrete case by Chiquet et al. (2017); see also Theorem 1 of Jiang et al. (2020).

**Theorem 1.3.** *For $\mu$-a.e. $x \in \mathbf{R}^d$, the measure $\mu|_{V_{u_{\mu,\lambda},x}}$ is $\lambda$-cohesive, and if $A \ni x$ is such that $\mu|_A$ is $\lambda$-cohesive, then $\mu(A \setminus V_{u_{\mu,\lambda},x}) = 0$.*

It is not difficult to see, directly from (1.2), that if $\mu$ is $\lambda$-cohesive, then it is also $\lambda'$-cohesive for any $\lambda' \geq \lambda$. As explained in more details in Section 5, Theorem 1.3 therefore implies the following theorem, referred to in the literature as the *agglomeration conjecture* of Hocking et al. (2011), and also proved in the discrete case by Chiquet et al. (2017).

**Theorem 1.4.** *If $\lambda \leq \lambda'$ then for $\mu$-a.e. $x$ we have $\mu(V_{u_{\mu,\lambda},x} \setminus V_{u_{\mu,\lambda'},x}) = 0$. In words, for $\mu$-almost every $x$, the $\lambda'$-cluster of $x$ is a subset of the $\lambda$-cluster of $x$.*

The discrete case of Theorem 1.3 (in combination with a condition for $\lambda$-cohesivity described in Theorem 1.9 below) is described by Jiang et al. (2020) as an "almost exact characterization" of the clusters. Our first main theoretical contribution is an "exact" characterization of the minimizer $u_{\mu,\lambda}$. This characterization (Theorem 1.7 below) seems to be new even in the discrete case. We need a few definitions and notations. We call a Borel set $V \subseteq \mathbf{R}^d$ *$\mu$-regular* if either $V$ is a singleton or $\mu(V) > 0$. For a $\mu$-regular set $V \subseteq \mathbf{R}^d$, let

$$\mathcal{C}_\mu(V) \coloneqq \begin{cases} \fint_V x \, \mathrm{d}\mu(x) & \text{if } \mu(V) > 0; \\ x & \text{if } V = \{x\} \end{cases} \tag{1.6}$$

be the $\mu$-centroid of $V$. (Here and henceforth we write $\fint_V f \, \mathrm{d}\mu \coloneqq \frac{1}{\mu(V)} \int_V f \, \mathrm{d}\mu$.) Note that when $V$ is a singleton with $\mu(V) > 0$ the two cases of (1.6) agree.

**Definition 1.5.** *We say that a measurable function $u \in L^2(\mu; \mathbf{R}^d)$ is* $\mu$-regular *if there is a measurable representative of $u$ and a Borel set $A \subseteq \mathbf{R}^d$ such that $\mu(\mathbf{R}^d \setminus A) = 0$, $V_{u,x} \cap A$ is $\mu$-regular for $\mu$-a.e. $x$, and $\mathcal{C}_\mu(V_{u,x} \cap A) \neq \mathcal{C}_\mu(V_{u,z} \cap A)$ for $\mu$-a.e. $x, z$ with $u(x) \neq u(z)$. If $u$ is $\mu$-regular, we define $\mathcal{E}_{\mu,u}(x) \coloneqq \mathcal{C}_\mu(V_{u,x} \cap A)$, and we note that $\mathcal{E}_{\mu,u}$ is a well-defined element of $L^\infty(\mu; \mathbf{R}^d)$, independent of the choice of $A$ or the choice of representative of $u$. (See Lemma 7.1 below.) In this case, we let*

$$\mathcal{M}_u(\mu) \coloneqq (\mathcal{E}_{\mu,u})_*(\mu) = \int \delta_{\mathcal{E}_{\mu,u}(x)} \, \mathrm{d}\mu(x)$$

*be the image of the measure $\mu$ under $\mathcal{E}_{\mu,u}$. By this we mean that for any Borel set $B$, we have*

$$\mathcal{M}_u(\mu)(B) = \mu(\mathcal{E}_{\mu,u}^{-1}(B)).$$

*In words, the measure $\mathcal{M}(u)$ is derived from $\mu$ by concentrating all of the $\mu$-mass in each level set of $u$ at the $\mu$-centroid of the level set.*

When the support of $\mu$ is finite, a function $u \colon \operatorname{supp} \mu \to \mathbf{R}^d$ is $\mu$-regular if and only if $\mathcal{C}_\mu(V_{u,x}) \neq \mathcal{C}_\mu(V_{u,z})$ for every $x, z \in \operatorname{supp} \mu$ with $u(x) \neq u(z)$. In words, we ask that different level sets of $u$ have different centroids, and in this case, we have $\mathcal{M}_u(\mu) = \int \delta_{\mathcal{C}_\mu(V_{u,x})} \, \mathrm{d}\mu(x)$. The phrasing of Definition 1.5 is more complicated due to some measure-theoretic technical difficulties that arise when the support of $\mu$ is uncountable. We will prove the following preliminary proposition in Section 4 below.

**Proposition 1.6.** *The function $u_{\mu,\lambda}$ is $\mu$-regular.*

Now we can state our exact characterization of the minimizer $u_{\mu,\lambda}$.

**Theorem 1.7.** *Let $u$ be a $\mu$-regular function and $\lambda \geq 0$. The following are equivalent.*

1. *For $\mu$-a.e. $x$, we have $V_{u,x} = V_{u_{\mu,\lambda},x}$ up to a $\mu$-null set.*
2. *The measure $\mathcal{M}_u(\mu)$ is $\lambda$-shattered and, for $\mu$-a.e. $x$, the restriction $\mu|_{V_{u,x}}$ is $\lambda$-cohesive.*

Shortly after we posted the first version of this article, Nguyen and Mamitsuka (Preprint, 2021) derived several results on the properties of the optimal clusters. Our framework allows us to recover one of their main results in the measure-valued setting. The following proposition, which is analogous to Theorem 3 of Nguyen and Mamitsuka (Preprint, 2021), states that each cluster is contained in a ball centered at the centroid of the cluster and of radius $\lambda$ times the total mass of the cluster; and that the centroids of the different clusters are sufficiently far apart from one another that these balls do not intersect. We denote by $\overline{B}_r(x)$ the closed Euclidean ball of radius $r \geq 0$ centered at $x \in \mathbf{R}^d$.

**Proposition 1.8.** *For $\mu$-a.e. $x, z \in \mathbf{R}^d$, we have*

$$V_{u_{\mu,\lambda},x} \subseteq \overline{B}_{\lambda\mu(V_{u_{\mu,\lambda},x})} \left( \mathcal{E}_{\mu,u_{\mu,\lambda}}(x) \right), \tag{1.7}$$

*and whenever $u_{\mu,\lambda}(x) \neq u_{\mu,\lambda}(z)$,*

$$|\mathcal{E}_{\mu,u_{\mu,\lambda}}(x) - \mathcal{E}_{\mu,u_{\mu,\lambda}}(z)| > \lambda[\mu(V_{u_{\mu,\lambda},x}) + \mu(V_{u_{\mu,\lambda},z})]. \tag{1.8}$$

We will prove Theorem 1.7 and Proposition 1.8 in Section 5 below.

Theorem 1.7 motivates taking particular interest in the properties of $\lambda$-cohesive and $\lambda$-shattered sets. We are mostly interested in situations in which a dataset can be partitioned into a bounded number of clusters in the presence of a large number of datapoints. In light of Theorem 1.7, this means that there should be a $\lambda$ such that the centroids of the clusters, weighted by the fraction of datapoints in the cluster, form a $\lambda$-shattered set, while the datapoints in each cluster form a $\lambda$-cohesive set. In the regime where there is a bounded number of clusters but the number of datapoints tends to infinity, the question of the $\lambda$-shattering of the set of centroids is a bounded-size optimization problem. In this paper we only address it in the simplest case. On the other hand, the question of $\lambda$-cohesion of each cluster lends itself to asymptotic analysis, so this will interest us in the sequel. We will consider the "continuum limit" of situations with continuous measures, and also provide "law of large numbers" results for atomic measures drawn from the corresponding continuous distributions.

We noted above that if $\mu$ is $\lambda$-cohesive, then it is also $\lambda'$-cohesive for any $\lambda' \geq \lambda$. By Theorem 1.3, this means that if $\mu$ is $\lambda$-shattered (which Theorem 1.3 and Proposition 1.6 tell us happens if and only if there are no $\lambda$-cohesive sets of positive $\mu$-measure), then it is also $\lambda'$-shattered for any $\lambda' \leq \lambda$. Thus we define

$$\lambda_1(\mu) := \inf\{\lambda \geq 0 \mid \mu \text{ is } \lambda\text{-cohesive}\} \tag{1.9}$$

and

$$\lambda_*(\mu) := \sup\{\lambda \geq 0 \mid \mu \text{ is } \lambda\text{-shattered}\}. \tag{1.10}$$

We then say that the level sets of a $\mu$-regular function $u$ are *detectable for* $\mu$ if

$$\lambda_*(\mathcal{M}_u(\mu)) > \operatorname*{ess\,sup}_{x\sim\mu} \lambda_1(\mu|_{V_{u,x}}). \tag{1.11}$$

By Theorem 1.7, this is equivalent to there existing some $\lambda$ such that the level sets of $u$ are the same (up to $\mu$-null modifications) as those of $u_{\mu,\lambda}$. We define the *detection parameter set* to be the (possibly empty) interval

$$\Lambda(\mu, u) := \left( \operatorname*{ess\,sup}_{x\sim\mu} \lambda_1(\mu|_{V_{u,x}}), \lambda_*(\mathcal{M}_u(\mu)) \right). \tag{1.12}$$

The parameter $\lambda_1(\mu)$ can be characterized up to a factor of 2 by simple geometric properties of $\mu$. Define the "radius" of the measure $\mu$ by

$$R(\mu) := \operatorname*{ess\,sup}_{x\sim\mu} \left| x - \mathcal{C}_\mu(\mathbf{R}^d) \right|, \tag{1.13}$$

and for $V \subseteq \mathbf{R}^d$, let $\operatorname{diam} V$ denote the Euclidean diameter of $V$. It turns out (see Proposition 4.4 below) that, if $\mu(\mathbf{R}^d) > 0$,

$$\frac{R(\mu)}{\mu(\mathbf{R}^d)} \leq \lambda_1(\mu) \leq \frac{\operatorname{diam}(\operatorname{supp}\mu)}{\mu(\mathbf{R}^d)}. \tag{1.14}$$

Since $R(\mu) \le \operatorname{diam}(\operatorname{supp}\mu) \le 2R(\mu)$, this characterizes $\lambda_1(\mu)$ up to a factor of 2 in terms of only the radius and the diameter of $\operatorname{supp}\mu$. On the other hand, we will compute in Proposition 2.1 below that, for $a_0, a_1 > 0$ and $x_0, x_1 \in \mathbf{R}^d$, we have

$$\lambda_*(a_0\delta_{x_0} + a_1\delta_{x_1}) = \frac{|x_1 - x_0|}{a_0 + a_1}.$$

Therefore, by Theorem 1.7, if equality holds in the first inequality in (1.14), then the partition of $\mu + \tau_x\mu$—the sum of $\mu$ and its translation by $x$—into $\operatorname{supp}\mu$ and $\tau_x \operatorname{supp}\mu$ is detectable as long as $|x| > 2R(\mu)$. We could certainly hope for no better since if $|x| \le R(\mu)$ then the supports of $\mu$ and its translation may overlap (cf. Proposition 1.8). On the other hand, if $\lambda_1(\mu) > \frac{R(\mu)}{\mu(\mathbf{R}^d)}$ then for this partition to be detectable we actually need greater separation than the obvious condition for the supports to not overlap would suggest. For this reason we are motivated to resolve the value of $\lambda_1(\mu)$ more precisely than is done by (1.14). Of particular interest are measures $\mu$ for which $\lambda_1(\mu) = \frac{R(\mu)}{\mu(\mathbf{R}^d)}$, which are such that combinations with any translation by at least twice the radius are detectable.

We now state a characterization of $\lambda_1(\mu)$, which will follow from a more general theorem (Theorem 4.1 below) giving the KKT characterization of the minimizer of $J_{\mu,\lambda}$. (Theorem 4.1 will also be crucial for the proof of Theorem 1.7.) In the discrete setting this result follows from the work of Chiquet et al. (2017); see also Theorem 1 of Jiang et al. (2020).

**Theorem 1.9.** *We have*

$$\lambda_1(\mu) = \mu(\mathbf{R}^d)^{-1} \min_{q \in \mathcal{Q}(\mu)} \|q\|_\infty, \tag{1.15}$$

*where $\mathcal{Q}(\mu)$ is the set of all $q \in L^\infty(\mu^{\otimes 2}; \mathbf{R}^d)$ satisfying, for $\mu$-a.e. $x, y \in \mathbf{R}^d$,*

$$q(x, y) = -q(y, x) \tag{1.16}$$

*and*

$$x - \mathcal{C}_\mu(\mathbf{R}^d) = \fint q(x, z)\,\mathrm{d}\mu(z). \tag{1.17}$$

We will prove Theorem 1.9 as a consequence of the KKT conditions in Section 4.

In Section 2, we use our tools to estimate or compute $\lambda_1(\mu)$ for $\mu$ the uniform measures on the $d$-sphere, the $d$-ball, and the vertices of the cross-polytope. In $d \ge 2$, these examples do not yield equality in the first inequality of (1.14). Thus we also give an explicit example of a nontrivial measure in $d \ge 2$ (a ball with density given by a power of the distance from the origin) for which equality does indeed hold.

In Section 3, we show the results of some additional numerical experiments regarding the examples considered in Section 2.

### 1.4 Stability of the clusters

We now turn our attention to the stability of the splittings. As the quantities in Theorem 1.9 are often more analytically tractable in the presence of symmetries, it can be easier to reason about the detectability of partitions in the case when measures have a nice symmetry property or a continuous density. On the other hand, in applications one is ultimately interested in atomic measures, often with some amount of randomness. In Section 6 we prove

several stability results showing that the clustering properties of these models approach the clustering properties of their limits. As example applications of these results, we prove Theorem 1.1 as well as the following theorem.

**Theorem 1.10.** *Let $\mu$ be a probability measure on $\mathbf{R}^d$ such that*

$$\operatorname{supp}\mu = \bigcup_{i=1}^{I} \overline{U_i} \tag{1.18}$$

*for some bounded connected open sets $U_1, \dots, U_I$, each with a Lipschitz boundary. Assume that the measure $\mu$ is absolutely continuous with respect to the Lebesgue measure, with Radon–Nikodym derivative bounded above and away from zero on each $U_i$. Let $u$ be an arbitrary function that is constant on each $\overline{U_i}$, and suppose that $u$ is detectable for $\mu$. Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables, each with law $\mu$, and define*

$$\mu_N \coloneqq \frac{1}{N}\sum_{n=1}^{N} \delta_{X_n}.$$

*Then the endpoints of $\Lambda(\mu_N, u)$ converge to those of $\Lambda(\mu, u)$ in probability as $N \to \infty$.*

Theorem 1.10 is proved in Section 6 as a consequence of quantitative continuity estimates for the clustering algorithm with respect to perturbations of $\mu$. Both absolutely continuous and Wasserstein perturbations of $\mu$ are considered; see Propositions 6.1, 6.2 and 6.4. These propositions can be applied directly to attain stability results analogous to Theorem 1.10 for other random configurations, or to obtain quantitative results for finite numbers of datapoints.

Several variants of the clustering method discussed in this paper can also be considered. For instance, in the fusion term $\iint |u(x) - u(y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)$ appearing in (1.2), one can consider replacing the Euclidean norm $|\cdot|$ by another norm, such as the $\ell^1$ norm. While this modification may be interesting from a computational perspective, it will also destroy the rotational invariance of the functional $J_{\mu,\lambda}$, and in general, we expect that these modified methods will also fail to correctly resolve the stochastic ball model with nearby balls. Another possibility is to introduce weights in the fusion term, such as

$$\iint_{x\neq y} |x-y|^{-\alpha} |u(x) - u(y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y),$$

for some exponent $\alpha \in (0, d)$ to be decided. The choice of a power-law weight can be motivated by the desire to ensure that the set of partitions discovered by the algorithm as we vary $\lambda$ is only rescaled under a rescaling of the measure; if one has in mind possibly complex datasets involving multiple scales, this seems like a natural requirement. Alternative possibilities that do not satisfy this property include replacing $|x-y|^{-\alpha}$ by $\exp(-c|x-y|)$, or other decreasing functions of the distance $|x-y|$. In the discrete setting, one can enforce stronger locality by restricting the sum to connected pairs in the $k$-nearest-neighbor graph. The latter possibility offers significant computational benefits, see Chi and Lange (2015). After posting the first ArXiv version of this paper, we showed in Dunlap and Mourrat (2022) that the introduction of suitably adjusted exponential weights allows us to recover very

general cluster shapes. In particular, the SON clustering algorithm with suitably adjusted weights succeeds in identifying disjoint balls in stochastic ball models, no matter how close they are; and it can also recover clusters whose convex hulls interesect. This contrasts with the results stated in Theorem 1.1 and Proposition 1.8 for the unweighted SON clustering algorithm. On the other hand, the addition of weights breaks the symmetries that allow us to prove the theoretical results in the present work.

## 2. Examples

In this section we compute $\lambda_1(\mu)$ for several choices of $\mu$.

**Proposition 2.1** (Two points)**.** *Let $x_0, x_1 \in \mathbf{R}^d$, $a_0, a_1 > 0$, and let $\mu = a_0\delta_{x_0} + a_1\delta_{x_1}$. Then*

$$\lambda_1(\mu) = \lambda_*(\mu) = \frac{|x_1 - x_0|}{a_0 + a_1}. \tag{2.1}$$

*Proof.* Since the support of $\mu$ has only two points, it is clear that $\lambda_1(\mu) = \lambda_*(\mu)$. (For a given $\lambda$, either $\mu$ is $\lambda$-cohesive or it is $\lambda$-shattered.) We observe that

$$\mathcal{C}_\mu(\mathbf{R}^d) = \frac{a_0x_0 + a_1x_1}{a_0 + a_1}.$$

For a function $q$ to satisfy (1.16)–(1.17), we must have that

$$x_0 - \frac{a_0x_0 + a_1x_1}{a_0 + a_1} = \frac{a_1}{a_0 + a_1}[x_0 - x_1] = \fint q(x_0, y)\,\mathrm{d}\mu(y) = \frac{a_1}{a_0 + a_1}q(x_0, x_1)$$

and

$$x_1 - \frac{a_0x_0 + a_1x_1}{a_0 + a_1} = \frac{a_0}{a_0 + a_1}[x_1 - x_0] = \fint q(x_1, y)\,\mathrm{d}\mu(y) = \frac{a_0}{a_0 + a_1}q(x_1, x_0).$$

The only function $q$ that satisfies the conditions (1.16)–(1.17) is therefore the function $q(x, y) \coloneqq x - y$. Then (2.1) follows from Theorem 1.9. □

**Proposition 2.2** (Interval)**.** *Let $d = 1$ and let $\mu$ be the Lebesgue measure on $[-1/2, 1/2]$ (with total mass 1). Then $\lambda_1(\mu) = 1/2$.*

*Proof.* Note that $\mathcal{C}_\mu(\mathbf{R}^d) = 0$. Letting $q(x, y) \coloneqq \frac{1}{2}\operatorname{sgn}(x - y)$, we have

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{2}\operatorname{sgn}(x - y)\,\mathrm{d}y = \frac{1}{2}\left[(x - (-1/2)) - (1/2 - x)\right] = x,$$

so (1.17) holds, and $\|q\|_\infty = 1/2$, which means that $\lambda_1 \le 1/2$ by Theorem 1.9. On the other hand, (1.14) shows that $\lambda_1(\mu) \ge 1/2$, so in fact $\lambda_1(\mu) = 1/2$. □

The next proposition is a characterization of $\lambda_1(\mu)$ for measures $\mu$ with support in the unit sphere that satisfy certain symmetry properties. We will next apply this result to several concrete examples.

**Proposition 2.3** (Symmetric measures)**.** *Suppose that $\mu$ is supported on $S^{d-1} = \partial B_1(0) \subseteq \mathbf{R}^d$, the support of $\mu$ comprises at least two points, and there is a subgroup $G \subseteq \mathrm{O}(d)$ (the group of Euclidean isometries of $\mathbf{R}^d$ preserving the origin) preserving $\mu$, acting transitively on $\operatorname{supp}\mu$, and such that for each $x \in \operatorname{supp}\mu$ and each $y \in S^{d-1} \setminus \{x, -x\}$, there is a $g \in G$ such that $g \cdot x = x$ but $g \cdot y \neq y$. Then for every $y \in \operatorname{supp}\mu$ we have*

$$\lambda_1(\mu) = \frac{2}{\int |x-y| \, \mathrm{d}\mu(x)} \tag{2.2}$$

*and*

$$\lambda_1(\mu)\mu(\mathbf{R}^d) \geq \sqrt{2}. \tag{2.3}$$

*Proof.* The strict convexity of $J_{\mu,\lambda}$ noted in the introduction implies that the minimizer $u_{\mu,\lambda}$ is unique. Since the measure $\mu$ is invariant under the action of $G$, the minimizer $u_{\mu,\lambda}$ must also be invariant under the action of $G$, in the sense that, for every $g \in G$ and $\mu$-a.e. $x \in \mathbf{R}^d$, we have

$$u_{\mu,\lambda}(g \cdot x) = g \cdot u_{\mu,\lambda}(x). \tag{2.4}$$

For each $x \in \operatorname{supp}\mu$, if $u_{\mu,\lambda}(x) \notin \mathbf{R}x$, then by assumption there is a $g \in G$ such that $g \cdot x = x$ and $g \cdot u_{\mu,\lambda}(x) \neq u_{\mu,\lambda}(x)$; but this would imply that $u_{\mu,\lambda}(x) = u_{\mu,\lambda}(g \cdot x) = g \cdot u_{\mu,\lambda}(x) \neq u_{\mu,\lambda}(x)$, a contradiction. Therefore, $u_{\mu,\lambda}(x) \in \mathbf{R}x$ for $\mu$-a.e. $x \in \mathbf{R}^d$. In other words, for $\mu$-a.e. $x \in \mathbf{R}^d$, we can find some $a_{\lambda,x} \in \mathbf{R}$ such that $u_{\mu,\lambda}(x) = a_{\lambda,x}x$. Using again (2.4), we deduce that for every $g \in G$, we must have $u_{\mu,\lambda}(g \cdot x) = g \cdot u_{\mu,\lambda}(x) = a_{\lambda,x} g \cdot x$. By the transitivity of the action of $G$ on $\operatorname{supp}\mu$, we must thus therefore have a fixed $a_\lambda \in \mathbf{R}$, depending only on $\lambda$ and not on $x$, such that $u_{\mu,\lambda}(x) = a_\lambda x$ for $\mu$-a.e. $x \in \mathbf{R}^d$. Since $\mu$ is invariant under the action of $G$, which acts transitively on $\operatorname{supp}\mu$, we have that the integral $\int |x-y| \, \mathrm{d}\mu(x)$ does not depend on the choice of $y \in \operatorname{supp}\mu$. Recalling also that $\operatorname{supp}\mu \subseteq S^{d-1}$, we see that, for every $a \in \mathbf{R}$ and an arbitrary $y \in \operatorname{supp}\mu$,

$$J_{\mu,\lambda}(x \mapsto ax) = \int |ax - x|^2 \, \mathrm{d}\mu(x) + \lambda \iint |ax - az| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(z) \tag{2.5}$$

$$= \mu(\mathbf{R}^d) \left[ a^2 + \lambda |a| \int |x-y| \, \mathrm{d}\mu(x) - 2a + 1 \right]. \tag{2.6}$$

The function $u_{\mu,\lambda}$ is constant if and only if the quantity in (2.6) is minimized for $a = 0$. This occurs exactly when

$$\lambda \geq \frac{2}{\int |x-y| \, \mathrm{d}\mu(x)},$$

and we have therefore shown (2.2).

We now argue that $\mathcal{C}_\mu(\mathbf{R}^d) = 0$. Integrating the identity (2.4) in $x$, we see that $\mathcal{C}_\mu(\mathbf{R}^d)$ must be a fixed point of the action of the group $G$. If $\operatorname{supp}\mu$ is of the form $\{x, -x\}$ for some $x \in \mathbf{R}^d$, then by transitivity the measure $\mu$ places the same mass on $x$ and $-x$, so $\mathcal{C}_\mu(\mathbf{R}^d) = 0$. Otherwise, we observe that the group $G$ has no other fixed point than the origin. Indeed, if $G$ had another fixed point, then by scaling we could obtain a fixed point $y \in S^{d-1}$. Since $\operatorname{supp}\mu$ is not of the form $\{y, -y\}$, we can find some $x \in \operatorname{supp}\mu \setminus \{y, -y\}$. The assumption on $G$ then guarantees the existence of some $g \in G$ with $g \cdot y \neq y$, a contradiction.

Now that $\mathcal{C}_\mu(\mathbf{R}^d) = 0$ is established, we apply Jensen's inequality to get that

$$\begin{aligned}\frac{1}{\mu(\mathbf{R}^d)}\int |x-y|\,\mathrm{d}\mu(x) &\le \left(\frac{1}{\mu(\mathbf{R}^d)}\int |x-y|^2\,\mathrm{d}\mu(x)\right)^{1/2}\\ &= \left(\frac{1}{\mu(\mathbf{R}^d)}\int 2(1-x\cdot y)\,\mathrm{d}\mu(x)\right)^{1/2}\\ &= \Big(2-\mathcal{C}_\mu(\mathbf{R}^d)\cdot y\Big)^{1/2} = \sqrt{2}.\end{aligned}$$

Combining this with (2.2) yields that

$$\lambda_1(\mu)\mu(\mathbf{R}^d) \ge \frac{2\mu(\mathbf{R}^d)}{\int |x-y|\,\mathrm{d}\mu(x)} \ge \sqrt{2}. \qquad \square$$

**Corollary 2.4** ($d$-sphere)**.** *Suppose that $d \ge 2$ and let $\mu$ be the uniform measure on the unit sphere $S^{d-1} = \partial B_1(0)$. Then*

$$\lambda_1(\mu)\mu(\mathbf{R}^d) = \frac{\Gamma(d-1/2)\Gamma((d-1)/2)}{\Gamma(d-1)\Gamma(d/2)}, \tag{2.7}$$

*where $\Gamma(z) = \int_0^\infty t^{z-1}\mathrm{e}^{-t}\,\mathrm{d}t$ denotes the standard gamma function. In particular,*

$$\lim_{d\to\infty} \lambda_1(\mu)\mu(\mathbf{R}^d) = \sqrt{2}. \tag{2.8}$$

*Proof.* Assume without loss of generality that $\mu(\mathbf{R}^d)$ is the area of $S^{d-1}$, that is,

$$\mu(\mathbf{R}^d) = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

We also have

$$\begin{aligned}\int |\mathbf{e}_1 - x|\,\mathrm{d}\mu(x) &= \frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)}\int_0^\pi (1-\cos^2\theta)^{\frac{d-2}{2}}\sqrt{(\cos\theta-1)^2+\sin^2\theta}\,\mathrm{d}\theta\\ &= \frac{2^d\pi^{(d-1)/2}}{\Gamma((d-1)/2)}\int_0^\pi \sin^{d-1}(\theta/2)\cos^{d-2}(\theta/2)\,\mathrm{d}\theta\\ &= \frac{2^d\pi^{(d-1)/2}}{\Gamma((d-1)/2)}\int_0^1 t^{d/2-1}(1-t)^{(d-3)/2}\,\mathrm{d}t\\ &= \frac{2^d\pi^{(d-1)/2}\Gamma(d/2)}{\Gamma(d-1/2)}\\ &= \frac{4\pi^{d/2}\Gamma(d-1)}{\Gamma((d-1)/2)\Gamma(d-1/2)}.\end{aligned}$$

The second identity is by the half-angle formulas for sine and cosine, the third is by making the substitution $t = \sin^2(\theta/2)$, the fourth is by the standard formula for the beta integral, and the last is by the Legendre duplication formula. Hence (2.7) follows from Proposition 2.3, noting that the group $G$ can be taken to be all of $\mathrm{O}(d)$, which clearly satisfies the hypotheses. The limit (2.8) is then a simple computation using Stirling's approximation. $\square$

**Corollary 2.5** (Vertices of the $n$-gon)**.** *Let $d = 2$, $n \geq 2$, and let $\mu$ be a uniform measure on the vertices of the regular $n$-gon inscribed in the unit circle, namely*

$$\mu = \frac{1}{n} \sum_{j=1}^{n} \delta_{\mathrm{e}^{2\pi \mathrm{i} j/n}},$$

*where we identify $\mathbf{R}^2$ with $\mathbf{C}$. Then we have*

$$\lambda_1(\mu)\mu(\mathbf{R}^d) = n \tan\left(\frac{\pi}{2n}\right).$$

*Proof.* We have

$$\frac{1}{\mu(\mathbf{R}^d))} \int |x-y| \, \mathrm{d}\mu(x) = \frac{1}{n} \sum_{j=1}^{n} |1 - \mathrm{e}^{2\pi \mathrm{i} j/n}| = \frac{2}{n} \sum_{j=1}^{n} \sin(\pi j/n) = \frac{2}{n} \cot\left(\frac{\pi}{2n}\right),$$

and the result follows from Proposition 2.3. □

**Corollary 2.6** (Vertices of the cross-polytope)**.** *Consider the measure on $\mathbf{R}^d$ given by*

$$\mu = \sum_{i=1}^{d} [\delta_{\mathbf{e}_i} + \delta_{-\mathbf{e}_i}].$$

*Then*

$$\lambda_1(\mu)\mu(\mathbf{R}^d) = \frac{2d}{(d-1)\sqrt{2}+1}$$

*and in particular*

$$\lim_{d \to \infty} \lambda_1(\mu)\mu(\mathbf{R}^d) = \sqrt{2}.$$

*Proof.* We have

$$\int |\mathbf{e}_1 - x| \, \mathrm{d}\mu(x) = 2(d-1)\sqrt{2} + 2$$

and the result follows from Proposition 2.3. □

**Proposition 2.7** ($d$-ball)**.** *Let $\gamma_d$ be as defined in* (1.3)*, and $\mu$ be a uniform measure on the unit ball $B_1(0) \subseteq \mathbf{R}^d$. Then*

$$\gamma_d \leq \lambda_1(\mu)\mu(\mathbf{R}^d) \leq 2^{1-\frac{1}{d}}. \tag{2.9}$$

*Proof.* Similarly to the proof of Proposition 2.3, we start by computing, for every $a \geq 0$,

$$J_{\mu,\lambda}(x \mapsto ax) = (1-a)^2 \int |x|^2 \, \mathrm{d}\mu(x) + \lambda a \iint |x-y| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y).$$

If the ball is $\lambda$-cohesive, then the quantity above must be minimal when $a = 0$. In such a case, we must have

$$\lambda \geq \frac{2 \int |x|^2 \, \mathrm{d}\mu(x)}{\iint |x-y| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)}.$$

In other words, we have

$$\lambda_1(\mu) \geq \frac{2\int |x|^2 \,\mathrm{d}\mu(x)}{\iint |x-y| \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)}. \tag{2.10}$$

The numerator in (2.10) is

$$\mu(\mathbf{R}^d)\frac{\int_0^1 r^{2+d-1}\,\mathrm{d}r}{\int_0^1 r^{d-1}\,\mathrm{d}r} = \mu(\mathbf{R}^d)\frac{d}{d+2}. \tag{2.11}$$

Denoting

$$\beta_d \coloneqq \fint\!\!\fint |x-y| \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y),$$

we have that

$$\beta_d = \frac{2d}{2d+1} \cdot \begin{cases} \frac{2^{3d+1}((d/2)!)^2 d!}{(d+1)(2d)!\pi} & \text{if } d \text{ is even,} \\ \frac{2^{d+1}(d!)^3}{(d+1)(((d-1)/2)!)^2(2d)!} & \text{if } d \text{ is odd.} \end{cases}$$

For $d = 2$, the proof of this identity can be found in Dunbar (1997), Grimmett and Stirzaker (2020, Exercise 4.13.4), or Santaló (1976, Section 4.2). In higher dimension, the computation is only sketched in Dunbar (1997), but does not pose additional difficulties (the high-dimensional integral splits into a product of Wallis integrals). One can verify that, for every $d \geq 1$,

$$\frac{\beta_{d+2}}{\beta_d} = \frac{(2d+2)(2d+4)^3}{2d(2d+3)(2d+5)(2d+6)} = 1 + \frac{9d^2+35d+32}{d(2d+3)(2d+5)(d+3)}.$$

Combining this with (2.10) and (2.11), we obtain the first inequality in (2.9).

For the second inequality in (2.9), if $d = 1$ then the inequality follows from Proposition 2.2, so assume that $d \geq 2$. Fix $\alpha \in \mathbf{R}$ to be chosen later and set

$$q_1(x,y) = \begin{cases} \alpha \operatorname{sgn}(x) & \text{if } |x| > |y|; \\ -\alpha \operatorname{sgn}(y) & \text{if } |x| < |y|; \\ 0 & \text{if } |x| = |y|. \end{cases}$$

Then we have

$$\fint q_1(x,y)\,\mathrm{d}\mu(y) = \alpha\frac{\mu\{y : |y| < |x|\}}{\mu(B_1(0))}\operatorname{sgn}(x) = \alpha|x|^d \operatorname{sgn}(x) = \alpha|x|^{d-1}x.$$

Let $x, y \in B_1(0)$ with $|x| > |y|$. We have

$$\begin{aligned}
&\left| q_1(x,y) + x - y - \fint q_1(x,z)\,\mathrm{d}\mu(z) + \fint q_1(y,z)\,\mathrm{d}\mu(z) \right| \\
&\quad = \left| \alpha \operatorname{sgn}(x) + x - y - \alpha|x|^{d-1}x + \alpha|y|^{d-1}y \right| \\
&\quad = \left| \operatorname{sgn}(x)[\alpha + |x| - \alpha|x|^d] - \operatorname{sgn}(y)[|y| - \alpha|y|^d] \right| \\
&\quad \le \alpha + \left| |x| - \alpha|x|^d \right| + \left| |y| - \alpha|y|^d \right| \\
&\quad \le \alpha + 2\left( \frac{1}{(\alpha d)^{\frac{1}{d-1}}} - \frac{\alpha}{(\alpha d)^{\frac{d}{d-1}}} \right) \\
&\quad = \alpha + \frac{2}{(\alpha d)^{\frac{1}{d-1}}}\left(1 - \frac{1}{d}\right).
\end{aligned}$$

Now taking $\alpha = 2^{\frac{d-1}{d}}/d$, we get

$$\left| q_1(x,y) + x - y - \fint q_1(x,z)\,\mathrm{d}\mu(z) + \fint q_1(y,z)\,\mathrm{d}\mu(z) \right| \le \frac{2^{\frac{d-1}{d}}}{d} + 2^{1-1/d}\left(1 - \frac{1}{d}\right) = 2^{1-1/d}.$$

Thus by Proposition 4.5 we have

$$\lambda_1(B_1(0)) \le \mu(B_1(0))^{-1} 2^{1-1/d}. \qquad \square$$

**Proposition 2.8** (Power-law weighted ball)**.** *Let $R \in (0,\infty)$ and $\mu$ be the measure given by*

$$\mathrm{d}\mu(x) = |x|^{-(d-1)}\mathbf{1}\{|x| \le R\}\mathrm{d}x.$$

*Then*

$$\lambda_1(\mu) = \frac{R(\mu)}{\mu(\mathbf{R}^d)} = \frac{2}{\alpha_{d-1}},$$

*where $\alpha_{d-1} = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the area of the unit $(d-1)$-sphere.*

*Proof.* We first note that, for any $s \in [0, R]$, we have using spherical coordinates that

$$\mu(B_s(0)) = \int_0^s \int_{S^{d-1}} \mathrm{d}\mathcal{H}^{d-1}(\boldsymbol{\theta})\,\mathrm{d}r = \frac{1}{2}\alpha_{d-1}s,$$

Define

$$q(x,y) = \begin{cases} R\operatorname{sgn}(x) & \text{if } |x| > |y|; \\ -R\operatorname{sgn}(y) & \text{if } |x| < |y|; \\ 0 & |x| = 1. \end{cases}$$

Then $q$ is evidently antisymmetric and $\|q\|_\infty = R$, and we have, using spherical coordinates and symmetry, that

$$\fint_{\mathbf{R}^2} q(x,y)\,\mathrm{d}\mu(y) = \frac{1}{\mu(B_R(0))}\int_0^R \int_{S^{d-1}} q(x, r\boldsymbol{\theta})\,\mathrm{d}\mathcal{H}^{d-1}(\boldsymbol{\theta})\,\mathrm{d}r = R\operatorname{sgn}(x)\frac{\mu(B_{|x|}(0))}{\mu(B_R(0))} = x.$$

By Theorem 1.9 this implies that

$$\lambda_1 \leq \frac{R}{\frac{1}{2}\alpha_{d-1}R} = \frac{2}{\alpha_{d-1}}.$$

On the other hand, we have by Proposition 4.4 that

$$\lambda_1 \geq \frac{R}{\mu(\mathbf{R}^d)} = \frac{2}{\alpha_{d-1}}. \qquad \square$$

## 3. Numerical experiments

In this section we supplement our theoretical results with some numerical experiments; see also Figures 1.1 and 1.2. The code is available at

https://github.com/ajdunlap/son-clustering-experiments.

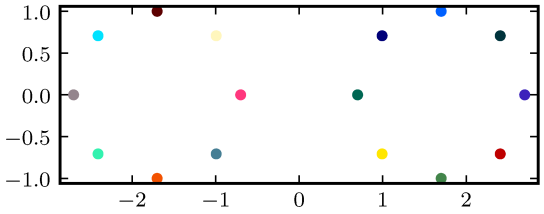
Our experiments were performed using the algorithm of Jiang and Vavasis (Preprint, 2020). This algorithm provides a certificate that the ouput clustering is correct. When $\lambda$ is very close to a value at which the number of clusters changes, limitations on computer time and numerical accuracy make it difficult to perform the calculations to sufficient accuracy to obtain the certificate. In particular, for situations such as that described by Theorem 1.1, the SON clustering algorithm becomes numerically very challenging to resolve for $\lambda$ close to $\lambda_c$, while the clustering structures that are produced for other values of $\lambda$ are not the expected partition into two parts. This further clarifies how the SON algorithm fails to resolve this clustering problem successfully in practice. Further work would be required to numerically probe the behavior of the algorithm very close to these critical values of $\lambda$.

### 3.1 Polygons

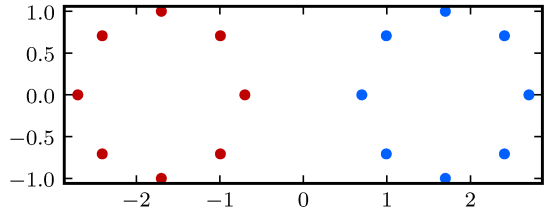
We begin with a case in which we can theoretically compute everything exactly. Fix some integer $n$ and let $\mu$ be a probability measure given by a Dirac mass at each vertex of two regular $n$-gons (each inscribed in a circle of radius 1) whose centers are separated by a distance $2r$. Our clustering characterization Theorem 1.7, combined with Proposition 2.1 and Corollary 2.5, tell us that sum-of-norms clustering makes exactly one cluster from each $n$-gon exactly when $2n\tan\left(\frac{\pi}{2n}\right) < \lambda < 2r$. We test this numerically with $n = 8$ and $r = 1.7$. In this case, $2n\tan\left(\frac{\pi}{2n}\right) \simeq 3.18$. We perform simulations with $\lambda = 3.1, 3.3, 3.5$ (noting that $3.1 < 2n\tan\left(\frac{\pi}{2n}\right) < 3.3 < 2r < 3.5$) and show the results in Figure 3.1. We see that our theoretical results are matched by the experiments.

### 3.2 $\lambda_1$ for a ball

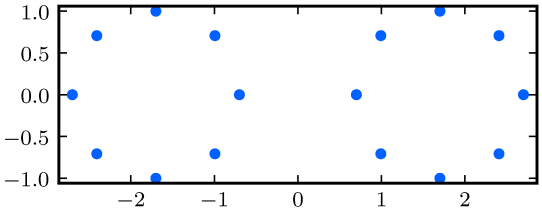
Proposition 2.7 does not precisely determine $\lambda_1(\mu)$ where $\mu$ is the indicator function of the unit ball. Here we perform a numerical experiment to estimate $\lambda_1(\mu)$ in dimension $d = 2$. We approximate the interior of the ball by the set of all points on a rectangular lattice with spacing $\delta$ lying inside the ball, i.e. $\{x \in \delta\mathbf{Z}^2 \mid |x| \leq 1\}$, and compute the number of clusters for varying choices of $\lambda$. The results are shown in Figure 3.2. In view of Corollary 2.4, (1.4), and Proposition 6.2 below, we know that the limit as $\delta \searrow 0$ of $\lambda_1$ is between $1.104\ldots$




(a) $\lambda = 3.1$. Each point is in its own cluster.




(b) $\lambda = 3.3$. Two clusters.




(c) $\lambda = 3.5$. One cluster.

Figure 3.1: Clustering results for the vertices of two octagons. Vertices assigned to the same cluster are drawn in the same color.

and 1.414.... The results of Figure 3.2 are roughly consistent with this, and suggest that the true limit is closer to the lower end of the theoretically proved range. The numerical results also suggest that $\lambda_1$ and $\lambda_*$ may be equal for the ball, which has not been studied theoretically, and thus is an interesting conjecture.

In Figure 3.2, the scheme of Jiang and Vavasis (Preprint, 2020) is again used to compute the clusterings. When $\lambda$ is close to a value at which the number of clusters changes, the certification procedure of Jiang and Vavasis (Preprint, 2020) may fail even when the duality gap in the clustering algorithm is close to machine precision. This is the reason for the missing values in the figure. Using a more sophisticated algorithm to more precisely estimate the values of $\lambda_1$ and $\lambda_*$ for the ball is an interesting topic for future work.

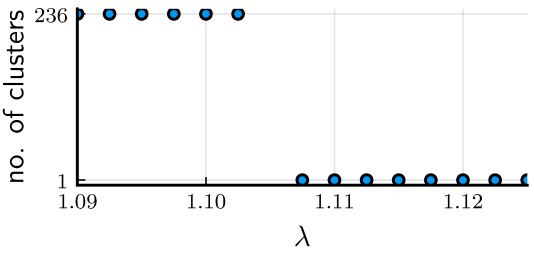
## 4. KKT characterization of the minimizer

Recall that, for convenience, we assume throughout the paper that the measure $\mu$ is finite (meaning that $\mu(\mathbf{R}^d) < \infty$) and has compact support. We start by justifying the existence and uniqueness of a minimizer for $J_{\mu,\lambda}$. It is clear (or see Lemma 7.2 below) that the functional $J_{\mu,\lambda}$ is continuous on $L^2(\mu;\mathbf{R}^d)$. Moreover, $J_{\mu,\lambda}$ is uniformly convex: for every $u, v \in L^2(\mu;\mathbf{R}^d)$, we have

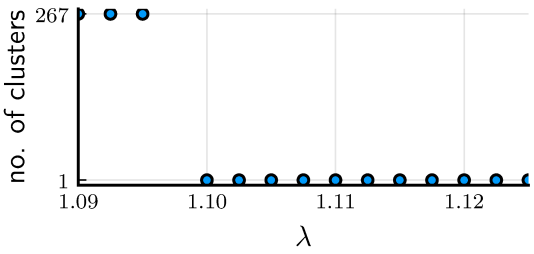
$$\frac{1}{2}\left(J_{\mu,\lambda}(u+v) + J_{\mu,\lambda}(u-v)\right) - J_{\mu,\lambda}(u) \geq \int v^2 \,\mathrm{d}\mu. \tag{4.1}$$

Finally, it is clear that the functional $J_{\mu,\lambda}$ is coercive, i.e. that there exist $c_1 > 0$ and $c_2 \geq 0$ such that $J_{\mu,\lambda}(u) \geq c_1 \|u\|^2_{L^2(\mu;\mathbf{R}^d)} - c_2$ for all $u \in L^2(\mu;\mathbf{R}^d)$. Thus there exists a unique minimizer $u_{\mu,\lambda} \in L^2(\mu;\mathbf{R}^d)$ for $J_{\mu,\lambda}$ (Evans, 2010, Section 8.2).

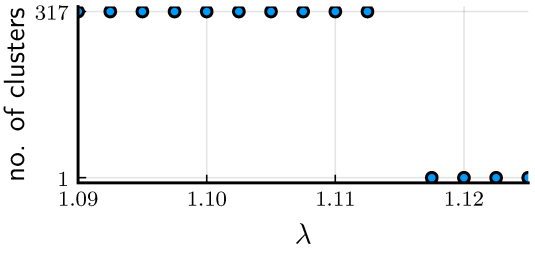
The key to most of our analysis is the following theorem, which evaluates the subdifferential of $J_{\mu,\lambda}$ and derives the resulting KKT characterization of the minimizer. For each



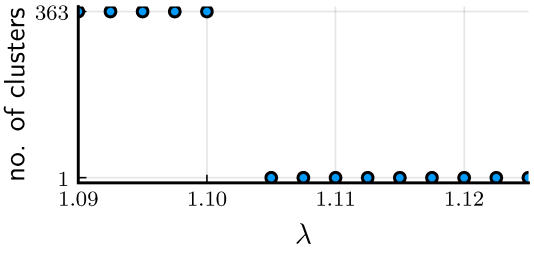

(a) $\delta = 0.1150$



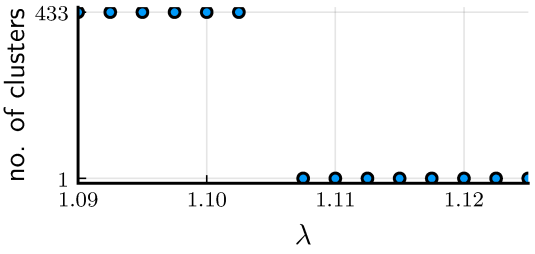

(b) $\delta = 0.1075$



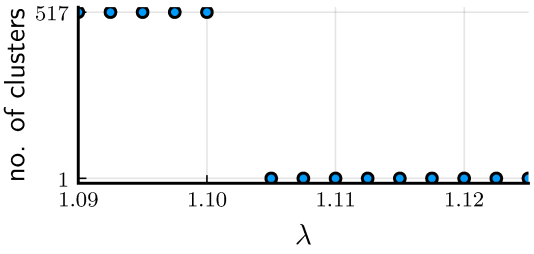

(c) $\delta = 0.1000$



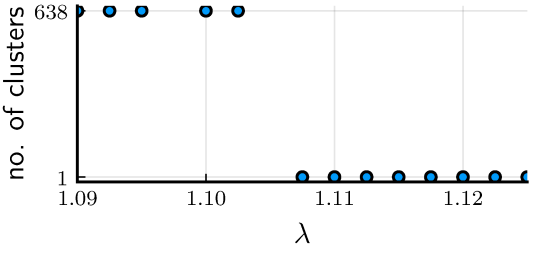

(d) $\delta = 0.0925$



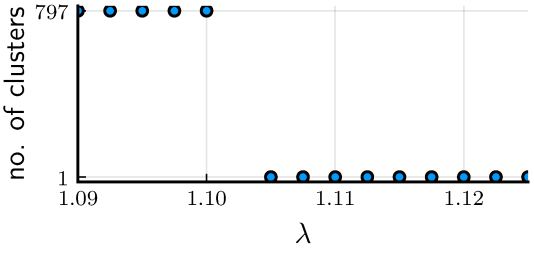

(e) $\delta = 0.0850$




(f) $\delta = 0.0775$




(g) $\delta = 0.0700$




(h) $\delta = 0.0625$

Figure 3.2: The number of clusters produced by sum-of-norms clustering run on the measure $\mu$ given by the uniform distribution on $\{x \in \delta\mathbf{Z}^2 \mid |x| \leq 1\}$, for varying choices of $\lambda$ and $\delta$. Missing values correspond to failures to certify the clustering using the procedure of Jiang and Vavasis (Preprint, 2020).

$z \in \mathbf{R}^d \setminus \{0\}$, we write

$$\operatorname{sgn}(z) := \frac{z}{|z|}. \tag{4.2}$$

**Theorem 4.1.** *Let $u \in L^2(\mu; \mathbf{R}^d)$. We have $u = u_{\mu,\lambda}$ ($\mu$-a.e.) if and only if there exists $w \in L^\infty(\mu^{\otimes 2}; \mathbf{R}^d)$ such that, for $\mu$-a.e. $x, y \in \mathbf{R}^d$, we have*

$$w(x,y) = -w(y,x), \tag{4.3}$$

$$u(x) \neq u(y) \implies w(x,y) = \operatorname{sgn}(u(x) - u(y)), \tag{4.4}$$

$$|w(x,y)| \leq 1, \tag{4.5}$$

*and*

$$x - u(x) = \lambda \int w(x,z)\, \mathrm{d}\mu(z). \tag{4.6}$$

*Proof.* For every measure $\nu$ and functional $F \colon L^2(\nu; \mathbf{R}^d) \to \mathbf{R}$, we define the subdifferential (Ekeland and Temam, 1976, Section I.5) of $F$ at $u \in L^2(\nu; \mathbf{R}^d)$ by

$$\partial F(u) := \left\{ p \in L^2(\nu; \mathbf{R}^d) \ : \ \forall v \in L^2(\nu; \mathbf{R}^d),\ F(u+v) \geq F(u) + \int p \cdot v \, \mathrm{d}\nu \right\}. \tag{4.7}$$

*Step 1.* In this step, for every probability measure $\nu$ on $\mathbf{R}^d$ with compact support, we identify the subdifferential of the functional

$$F(u) := \int |u| \, \mathrm{d}\nu \tag{4.8}$$

at $u \in L^2(\nu; \mathbf{R}^d)$ as

$$\partial F(u) = \left\{ w \in L^\infty(\nu; \mathbf{R}^d) \ : \ \|w\|_{L^\infty} \leq 1 \text{ and for } \nu\text{-a.e. } x \in \mathbf{R}^d,\ u(x) \neq 0 \implies w(x) = \operatorname{sgn}(u(x)) \right\}. \tag{4.9}$$

We denote by $K_1(u)$ the set on the right side of (4.9). Note that for every $a, b, w \in \mathbf{R}^d$, if $|w| \leq 1$ satisfies

$$a \neq 0 \implies w = \operatorname{sgn}(a),$$

then

$$|a + b| \geq |a| + w \cdot b.$$

From this observation, we can verify that $K_1(u) \subseteq \partial F(u)$ directly from (4.7) and (4.9). In order to show the opposite inclusion, we argue by contradiction and suppose that there exists $p \in \partial F(u) \setminus K_1(u)$. Since $K_1(u)$ is convex and closed in the Hilbert space $L^2(\nu; \mathbf{R}^d)$, the hyperplane separation theorem (Ekeland and Temam, 1976, Section I.1) guarantees the existence of a function $v \in L^2(\nu; \mathbf{R}^d)$ such that

$$\int p \cdot v \, \mathrm{d}\nu > \sup_{w \in K_1(u)} \int w \cdot v \, \mathrm{d}\nu. \tag{4.10}$$

Defining $w \in L^\infty(\nu; \mathbf{R}^d)$ by

$$w(x) = \begin{cases} \operatorname{sgn}(u(x)) & \text{if } u(x) \neq 0; \\ \operatorname{sgn}(v(x)) & \text{otherwise,} \end{cases}$$

we have for every $\varepsilon > 0$ that

$$\varepsilon^{-1}(F(u + \varepsilon v) - F(u)) = \int w \cdot v \, \mathrm{d}\nu + \int r_\varepsilon \, \mathrm{d}\nu,$$

where

$$r_\varepsilon = \varepsilon^{-1}(|u + \varepsilon v| - |u| - \varepsilon w \cdot v).$$

At a point where $u = 0$, we have $r_\varepsilon = |v| - \operatorname{sgn}(v) \cdot v = 0$ by the definitions, while at a point where $u \neq 0$, we have $r_\varepsilon = 0$ for $\varepsilon$ sufficiently small by the local linearity of $|\cdot|$, so the function $r_\varepsilon$ tends to 0 $\nu$-a.e. as $\varepsilon \downarrow 0$. Moreover, by the Cauchy–Schwarz and triangle inequalities we see that $|r_\varepsilon| \leq \varepsilon^{-1}(|u| + \varepsilon|v| - |u| + \varepsilon|w||v|) = 2|v|$. It thus follows from dominated convergence that

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-1}(F(u + \varepsilon v) - F(v)) = \int w \cdot v \, \mathrm{d}\nu.$$

On the other hand, recalling that $p \in \partial F(u)$, we must also have for every $\varepsilon > 0$ that

$$\varepsilon^{-1}(F(u + \varepsilon v) - F(u)) \geq \int p \cdot v \, \mathrm{d}\nu.$$

But the two previous displays contradict (4.10).

*Step 2.* In this step, we show that the subdifferential of the functional

$$G(u) := \int |u(x) - u(y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)$$

at $u \in L^2(\mu; \mathbf{R}^d)$ is given by

$$\partial G(u) = \left\{ x \mapsto 2 \int w(x, y) \, \mathrm{d}\mu(y) \ : \ w \text{ satisfies (4.3)–(4.5)} \right\}. \tag{4.11}$$

We denote by $K_2(u)$ the set on the right side of (4.11). Similarly to the previous step, one can check that $K_2(u) \subseteq \partial G(u)$. To show the opposite inclusion, we first introduce some notation. For every $v \in L^2(\mu; \mathbf{R}^d)$, define $\widetilde{v} \in L^2(\mu^{\otimes 2}; \mathbf{R}^d)$ by $\widetilde{v}(x, y) = v(x) - v(y)$, and by $F$ we denote the functional (4.8) with the measure $\nu = \mu^{\otimes 2}$. By definition, we have for every $v \in L^2(\mu; \mathbf{R}^d)$ that $G(v) = F(\widetilde{v})$. We fix $p \in \partial G(u)$, so that for every $v \in L^2(\mu; \mathbf{R}^d)$, we have

$$F(\widetilde{u} + \widetilde{v}) \geq F(\widetilde{u}) + \int p \cdot v \, \mathrm{d}\mu.$$

Since $G$ does not change if we add a constant to its argument, it must be that $\int p \, \mathrm{d}\mu = 0$. As a consequence, we can rewrite the last inequality as

$$F(\widetilde{u} + \widetilde{v}) \geq F(\widetilde{u}) + \frac{1}{2} \int \widetilde{p} \cdot \widetilde{v} \, \mathrm{d}\mu^{\otimes 2}.$$

This implies that the sets

$$\left\{ \left( \widetilde{v}, F(\widetilde{u}) + \frac{1}{2} \int \widetilde{p} \cdot \widetilde{v} \, \mathrm{d}\mu^{\otimes 2} \right) \ : \ v \in L^2(\mu; \mathbf{R}^d) \right\} \tag{4.12}$$

and

$$\left\{ \left( v', \lambda \right) \ : \ v' \in L^2(\mu^{\otimes 2}; \mathbf{R}^d) \text{ and } \lambda > F(\widetilde{u} + v') \right\} \tag{4.13}$$

are disjoint and convex. Moreover, the set in (4.13) is open in $L^2(\mu^{\otimes 2}; \mathbf{R}^d) \times \mathbf{R}$. Therefore, there is a hyperplane that separates the two sets. This means that there exists a $w \in L^2(\mu^{\otimes 2}; \mathbf{R}^d)$ such that for every $v \in L^2(\mu; \mathbf{R}^d)$ and $v' \in L^2(\mu^{\otimes 2}; \mathbf{R}^d)$, we have

$$F(\widetilde{u} + v') - \int w \cdot v' \, \mathrm{d}\mu^{\otimes 2} \geq F(\widetilde{u}) + \frac{1}{2} \int \widetilde{p} \cdot \widetilde{v} \, \mathrm{d}\mu^{\otimes 2} - \int w \cdot \widetilde{v} \, \mathrm{d}\mu^{\otimes 2}.$$

Taking $\widetilde{v} = 0$, we see that $w \in \partial F(\widetilde{u})$, and taking $v' = 0$, we see that

$$\int (p(x) - p(y) - 2w(x,y)) \cdot (v(x) - v(y)) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) = 0$$

for all $v \in L^2(\mu; \mathbf{R}^d)$. Recalling that $\int p \, \mathrm{d}\mu = 0$, we obtain that, for $\mu$-a.e. $x \in \mathbf{R}^d$,

$$p(x) = \int (w(x,y) - w(y,x)) \, \mathrm{d}\mu(y).$$

Since $w \in \partial F(\widetilde{\mu})$, the result of Step 1 gives us that $\|w\|_{L^\infty} \leq 1$ and, for $\mu$-a.e. $x, y \in \mathbf{R}^d$,

$$u(x) \neq u(y) \implies w(x,y) = \operatorname{sgn}(u(x) - u(y)).$$

We have thus completed the verification of the fact that $p \in K_2(u)$.

*Step 3.* It follows from the result of Step 2 that, for every $u \in L^2(\mu; \mathbf{R}^d)$, we have

$$\partial J_{\mu,\lambda}(u) = \left\{ x \mapsto 2(u(x) - x) + 2\lambda \int w(x,y) \, \mathrm{d}\mu(y) \ : \ w \text{ satisfies (4.3)–(4.5)} \right\}.$$

In particular, since $J_{\mu,\lambda}$ is convex, a function $u \in L^2(\mu; \mathbf{R}^d)$ is a minimizer of $J_{\mu,\lambda}$ if and only if $0 \in \partial J_{\mu,\lambda}(u)$. Equivalently,

$$J_{\mu,\lambda}(u) = \inf_{v \in L^2(\mu; \mathbf{R}^d)} J_{\mu,\lambda}(v) \iff \exists w \in L^\infty(\mu; \mathbf{R}^d) \text{ satisfying (4.3)–(4.6).}$$

This completes the proof of the theorem. □

From Theorem 4.1, we can prove Theorem 1.9 as a simple corollary.

*Proof of Theorem 1.9.* By integrating (4.6) in $x$ with respect to the measure $\mu$, we see that $\mu$ is $\lambda$-cohesive if and only if the minimizer of $J_{\mu,\lambda}$ is given by $u(x) = \mathcal{C}_\mu(\mathbf{R}^d)$, which happens if and only if there is a $w$ satisfying (4.3) and (4.5) such that

$$x - \mathcal{C}_\mu(\mathbf{R}^d) = \lambda \int w(x,y) \, \mathrm{d}\mu(y). \tag{4.14}$$

Taking $q \coloneqq \mu(\mathbf{R}^d)\lambda w$ completes the proof. □

We now state a couple of lemmas which we will use to prove Proposition 1.6. For every $V \subseteq \mathbf{R}^d$, we write $V^{\mathrm{c}} \coloneqq \mathbf{R}^d \setminus V$ to denote the complement of $V$.

**Lemma 4.2.** *There is a Borel set $A \subseteq \mathbf{R}^d$ such that $\mu(\mathbf{R}^d \setminus A) = 0$ and, for $\mu$-a.e. $x$, we have that $V_{u_{\mu,\lambda},x} \cap A$ is $\mu$-regular and*

$$\mathcal{C}_\mu(V_{u_{\mu,\lambda},x} \cap A) - u_{\mu,\lambda}(x) = \lambda \int_{V^{\mathrm{c}}_{u_{\mu,\lambda},x}} \operatorname{sgn}(u_{\mu,\lambda}(x) - u_{\mu,\lambda}(y)) \,\mathrm{d}\mu(y). \tag{4.15}$$

*In particular, $\mathcal{E}_{\mu,u_{\mu,\lambda}}(x) \coloneqq \mathcal{C}_\mu(V_{u_{\mu,\lambda},x} \cap A)$ (as in Definition 1.5) is well-defined as an element of $L^\infty(\mu; \mathbf{R}^d)$, independently of the choice of $A$ (up to a $\mu$-null modification).*

*Proof.* For typographical convenience, we write $u = u_{\mu,\lambda}$. Define

$$\mathcal{E}(x) \coloneqq u(x) + \int_{V^{\mathrm{c}}_{u,x}} \operatorname{sgn}(u(x) - u(y)) \,\mathrm{d}\mu(y).$$

Let $A \coloneqq \{x \in \mathbf{R}^d \mid \mu(V_{u,x}) > 0 \text{ or } \mathcal{E}(x) = x\}$, and $w$ be as in the statement of Theorem 4.1. Using (4.4), we can rewrite (4.6) as, for $\mu$-a.e. $x$,

$$x - u(x) = \lambda \int_{V_{u,x}} w(x,y) \,\mathrm{d}\mu(y) + \lambda \int_{V^{\mathrm{c}}_{u,x}} \operatorname{sgn}(u(x) - u(y)) \,\mathrm{d}\mu(y). \tag{4.16}$$

Since $\mathcal{E}$ is constant on each $V_{u,x}$ by definition, if $x \in A$ and $\mu(V_{u,x}) = 0$, then $V_{u,x} \cap A = \{x\}$ and thus (4.15) holds. Moreover, (4.16) implies that $\mu(\mathbf{R}^d \setminus A) = 0$. On the other hand, if $\mu(V_{u,x}) > 0$, then averaging (4.16) over $x \sim \mu|_{V_{u,x}}$, we have

$$\begin{aligned} \mathcal{C}_\mu(V_{u,x}) - u(x) &= \frac{\lambda}{\mu(V_{u,x})} \iint_{V^2_{u,x}} w(z,y) \,\mathrm{d}\mu(y) \,\mathrm{d}\mu(z) \\ &\qquad + \frac{\lambda}{\mu(V_{u,x})} \int_{V_{u,x}} \int_{V^{\mathrm{c}}_{u,x}} \operatorname{sgn}(u(z) - u(y)) \,\mathrm{d}\mu(y) \,\mathrm{d}\mu(z) \\ &= \lambda \int_{V^{\mathrm{c}}_{u,x}} \operatorname{sgn}(u(x) - u(y)) \,\mathrm{d}\mu(y), \end{aligned} \tag{4.17}$$

with the second identity by (4.3) (to eliminate the first term) and the fact that $u(z) = x$ for all $z \in V_{u,x}$ (to simplify the second term). □

Roughly speaking, the next lemma states that the vector formed by the centroids of two clusters and the vector formed by the values taken by the mapping $u$ on these clusters must be positively correlated. One could also say that the mapping sending each cluster centroid to the image under $u$ of any point in this cluster is a monotone operator.

**Lemma 4.3.** *For $\mu$-a.e. $x, z$ we have*

$$\begin{aligned} &(u_{\mu,\lambda}(x) - u_{\mu,\lambda}(z)) \cdot \big(\mathcal{E}_{\mu,u_{\mu,\lambda}}(x) - \mathcal{E}_{\mu,u_{\mu,\lambda}}(z)\big) \\ &\qquad \geq \lambda[\mu(V_{u_{\mu,\lambda},x}) + \mu(V_{u_{\mu,\lambda},z})]|u_{\mu,\lambda}(x) - u_{\mu,\lambda}(z)| + |u_{\mu,\lambda}(x) - u_{\mu,\lambda}(z)|^2. \end{aligned} \tag{4.18}$$

*Proof.* For typographical convenience, let $u = u_{\mu,\lambda}$ and $\mathcal{E} = \mathcal{E}_{\mu,u_{\mu,\lambda}}$. By Lemma 4.2, for $\mu$-a.e. $x$ we have

$$\mathcal{E}(x) - u(x) = \lambda \int_{V_{u,x}^{\mathrm{c}}} \operatorname{sgn}(u(x) - u(y)) \, \mathrm{d}\mu(y).$$

Therefore, we have for $\mu$-a.e. $x, z$ that

$$\begin{aligned}
\mathcal{E}(x) - \mathcal{E}(z) &= u(x) - u(z) + \lambda \int_{V_{u,x}^{\mathrm{c}}} \operatorname{sgn}(u(x) - u(y)) \, \mathrm{d}\mu(y) \\
&\quad - \lambda \int_{V_{u,z}^{\mathrm{c}}} \operatorname{sgn}(u(z) - u(y)) \, \mathrm{d}\mu(y) \\
&= u(x) - u(z) + \lambda[\mu(V_{u,x}) + \mu(V_{u,z})] \operatorname{sgn}(u(x) - u(z)) \\
&\quad + \lambda \int_{(V_{u,z} \cup V_{u,z})^{\mathrm{c}}} [\operatorname{sgn}(u(x) - u(y)) - \operatorname{sgn}(u(z) - u(y))] \, \mathrm{d}\mu(y).
\end{aligned}$$

Taking the dot product of each side with $u(x) - u(z)$, we obtain

$$\begin{aligned}
&(u(x) - u(z)) \cdot (\mathcal{E}(x) - \mathcal{E}(z)) \\
&\qquad = |u(x) - u(z)|^2 + \lambda[\mu(V_{u,x}) + \mu(V_{u,z})]|u(x) - u(z)| \\
&\qquad\quad + \lambda \int_{(V_{u,z} \cup V_{u,z})^{\mathrm{c}}} (u(x) - u(z)) \cdot [\operatorname{sgn}(u(x) - u(z)) - \operatorname{sgn}(u(z) - u(y))] \, \mathrm{d}\mu(y).
\end{aligned} \tag{4.19}$$

We note that for any vectors $a, b, c \in \mathbf{R}^d$, we have

$$\begin{aligned}
(a-b) \cdot (\operatorname{sgn}(a-c) - \operatorname{sgn}(b-c)) &= ((a-c) - (b-c)) \cdot \left( \frac{a-c}{|a-c|} - \frac{b-c}{|b-c|} \right) \\
&= |a-c| + |b-c| - \left( \frac{1}{|a-c|} + \frac{1}{|b-c|} \right) (a-c) \cdot (b-c) \\
&\geq |a-c| + |b-c| - \left( \frac{1}{|a-c|} + \frac{1}{|b-c|} \right) |a-c||b-c| = 0,
\end{aligned}$$

by the Cauchy–Schwarz inequality. (If $a - c = 0$ or $b - c = 0$ then the inequality is still clear.) This means that the integral on the right side of (4.19) is nonnegative, which implies (4.18). □

*Proof of Proposition 1.6.* Theorem 4.1 gives us a $w$ and a set $A \subseteq \mathbf{R}^d$ with $\mu(\mathbf{R}^d \setminus A) = 0$ so that for all $x \in A$ such that $\mu(V_{u_{\mu,\lambda},x}) = 0$ we have

$$\begin{aligned}
x - u_{\mu,\lambda}(x) &= \lambda \int_{V_{u_{\mu,\lambda},x}} w(x,z) \, \mathrm{d}\mu(z) + \lambda \int_{V_{u_{\mu,\lambda},x}^{\mathrm{c}}} \operatorname{sgn}(u_{\mu,\lambda}(x) - u_{\mu,\lambda}(z)) \, \mathrm{d}\mu(z) \\
&= \lambda \int \operatorname{sgn}(u_{\mu,\lambda}(x) - u_{\mu,\lambda}(z)) \, \mathrm{d}\mu(z).
\end{aligned}$$

This implies that for all $y \in V_{u_{\mu,\lambda},x} \cap A$ we must have

$$\begin{aligned}
y &= u_{\mu,\lambda}(y) + \lambda \int \operatorname{sgn}(u_{\mu,\lambda}(y) - u_{\mu,\lambda}(z)) \, \mathrm{d}\mu(z) \\
&= u_{\mu,\lambda}(x) + \lambda \int \operatorname{sgn}(u_{\mu,\lambda}(x) - u_{\mu,\lambda}(z)) \, \mathrm{d}\mu(z) = x.
\end{aligned}$$

This proves the first condition in the definition of $\mu$-regularity. The second condition follows immediately from Lemma 4.3. □

As a simple consequence of Theorem 1.9, we can prove the bound (1.14) mentioned in the introduction.

**Proposition 4.4.** *For any $\mu$ we have*

$$\frac{R(\mu)}{\mu(\mathbf{R}^d)} \le \lambda_1(\mu) \le \frac{\operatorname{diam}_{|\cdot|}(\operatorname{supp}\mu)}{\mu(\mathbf{R}^d)}. \tag{4.20}$$

*Proof.* First we show the lower bound. From Theorem 1.9, we have a $q\colon \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d$ such that (1.16)–(1.17) hold and $\|q\|_\infty = \lambda_1(\mu)\mu(\mathbf{R}^d)$. Therefore, we have for $\mu$-a.e. $x$ that

$$\left|x - \mathcal{C}_\mu(\mathbf{R}^d)\right| \le \fint |q(x,y)|\,\mathrm{d}\mu(y) \le \|q\|_\infty = \lambda_1(\mu)\mu(\mathbf{R}^d),$$

which implies the lower bound in (4.20). To prove the upper bound, let

$$q(x,y) \coloneqq x - y. \tag{4.21}$$

It is obvious that $q$ satisfies (1.16)–(1.17), and that $\|q\|_\infty = \operatorname{diam}_{|\cdot|}(\operatorname{supp}\mu)$. Therefore, Theorem 1.9 implies the upper bound in (4.20). □

We conclude this section with the following simple proposition that allows us to replace the exact equality in (1.17) with an approximation.

**Proposition 4.5.** *For any antisymmetric function $q_1\colon \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d$, we have*

$$\lambda_1(\mu) \le \mu(\mathbf{R}^d)^{-1} \operatorname*{ess\,sup}_{x,y\sim\mu} \left| q_1(x,y) + x - y - \fint q_1(x,z)\,\mathrm{d}\mu(z) + \fint q_1(y,z)\,\mathrm{d}\mu(z) \right|. \tag{4.22}$$

*Proof.* Let

$$q(x,y) \coloneqq q_1(x,y) + x - y - \fint q_1(x,z)\,\mathrm{d}\mu(z) + \fint q_1(y,z)\,\mathrm{d}\mu(z).$$

We have

$$\begin{aligned} q(y,x) &= q_1(y,x) + y - x - \fint q_1(y,z)\,\mathrm{d}\mu(z) + \fint q_1(x,z)\,\mathrm{d}\mu(z) \\ &= -q_1(x,y) + y - x + \fint q_1(x,z)\,\mathrm{d}\mu(z) - \fint q_1(z,y)\,\mathrm{d}\mu(z) = -q(x,y), \end{aligned}$$

so $q$ satisfies (1.16), and moreover

$$\begin{aligned} \fint q(x,y)\,\mathrm{d}\mu(y) &= \fint q(x,y)\,\mathrm{d}\mu(y) + \fint x\,\mathrm{d}\mu(y) - \fint y\,\mathrm{d}\mu(y) - \fint q_1(x,z)\,\mathrm{d}\mu(z) \\ &\qquad + \fint\fint q_1(y,z)\,\mathrm{d}\mu(z)\,\mathrm{d}\mu(y) \\ &= x, \end{aligned}$$

so $q$ satisfies (1.17). Thus Theorem 1.9 implies the result. □

## 5. Exact characterization of the clusters

In this section, we prove Theorems 1.7 and 1.3 and Proposition 1.8.

*Proof of Theorem 1.7.* We first suppose that for $\mu$-a.e. $x$, $V_{u,x} = V_{u_{\mu,\lambda},x}$ up to a $\mu$-null set and try to prove the second statement of the theorem. Since the second statement of the theorem concerns only the level sets of $u$, we can and do assume that $u = u_{\mu,\lambda}$. First we show that $\mu|_{V_{u,x}}$ is cohesive for $\mu$-a.e. $x$.

Subtracting (4.15) from (4.16), we have

$$x - \mathcal{E}_{\mu,u}(x) = \lambda \int_{V_{u,x}} w(x,y)\,\mathrm{d}\mu(y)$$

for $\mu$-a.e. $x$. By Theorem 4.1, this implies that the constant $\mathcal{E}_{\mu,u}(x)$ is a minimizer of $J_{\mu|_{V_{u,x}},\lambda}$, so $\mu|_{V_{u,x}}$ is $\lambda$-cohesive.

To prove that $\mathcal{M}_u(\mu)$ is $\lambda$-shattered, define

$$\widetilde{u}(\mathcal{E}_{\mu,u}(x)) \coloneqq u(x).$$

This is well-defined by Lemma 4.3. Then $\widetilde{u}$ is defined $\mathcal{M}_u(\mu)$-a.e., and it is clear that $\widetilde{u}$ can be extended to an injection on $\mathbf{R}^d$. By (4.15) we have

$$\widetilde{u}(X) = X - \lambda \int \operatorname{sgn}(\widetilde{u}(X) - \widetilde{u}(Y))\,\mathrm{d}\mathcal{M}_u(\mu)(Y)$$

for $\mathcal{M}_u(\mu)$-a.e. $X$. Taking $\widetilde{w}(X,Y) = \operatorname{sgn}(X-Y)$ as the $w$ in Theorem 4.1, we see that $\widetilde{u}$ is in fact a minimizer of $J_{\mathcal{M}_u(\mu),\lambda}$. Thus $\mathcal{M}_u(\mu)$ is $\lambda$-shattered.

Now we prove the other direction, so suppose we have a $\mu$-regular function $u$ such that $\mathcal{M}_u(\mu)$ is $\lambda$-shattered and, for $\mu$-a.e. $x$, the restriction $\mu|_{V_{u,x}}$ is $\lambda$-cohesive. Let $\widetilde{u}$ be the (injective) minimizer of $J_{\mathcal{M}_u(\mu),\lambda}$ and define

$$v(x) = \widetilde{u}(\mathcal{E}_{\mu,u}(x)), \tag{5.1}$$

noting that the assumption that $u$ is $\mu$-regular means that $\mathcal{E}_{\mu,u}$ is defined. Since $\widetilde{u}$ is injective, we see that $v$ has the same level sets as $u$. We want to prove that $v$ is a minimizer of $J_{\mu,\lambda}$. For $\mu$-a.e. $x$, by Theorem 4.1 and the fact that $\mu|_{V_{u,x}}$ is $\lambda$-cohesive, we have an antisymmetric $w_{V_{u,x}}$, bounded in norm by 1, such that

$$x - \mathcal{E}_{\mu,u}(x) = \lambda \int_{V_{u,x}} w_{V_{u,x}}(x,y)\,\mathrm{d}\mu(y). \tag{5.2}$$

Moreover, using (5.1) and (4.6) we have

$$\begin{aligned}\mathcal{E}_{\mu,u}(x) - v(x) &= \lambda \int \operatorname{sgn}(\widetilde{u}(\mathcal{E}_{\mu,u}(x)) - \widetilde{u}(\mathcal{E}_{\mu,u}(y)))\,\mathrm{d}\mathcal{M}_u(\mu)(y) \\ &= \lambda \int_{V_{u,x}^{\mathrm{c}}} \operatorname{sgn}(v(x) - v(y))\,\mathrm{d}\mu(y).\end{aligned} \tag{5.3}$$

So define

$$w(x,y) = \begin{cases} w_{V_{u,x}}(x,y) & \text{if } u(x) = u(y); \\ \operatorname{sgn}(v(x) - v(y)) & \text{if } u(x) \neq u(y). \end{cases}$$

Then we have, using (5.2) and (5.3), that

$$\begin{aligned} x - v(x) &= x - \mathcal{E}_{\mu,u}(x) + \mathcal{E}_{\mu,u}(x) - v(x) \\ &= \lambda \int_{V_{u,x}} w_{V_{u,x}}(x,y)\,\mathrm{d}\mu(y) + \lambda \int_{V_{u,x}^{\mathrm{c}}} \operatorname{sgn}(v(x) - v(y))\,\mathrm{d}\mu(y) \\ &= \lambda \int w(x,y)\,\mathrm{d}\mu(y), \end{aligned}$$

verifying (4.6). Conditions (4.3)–(4.5) are clearly satisfied for $w$, so this proves that $v$ is a minimizer of $J_{\mu,\lambda}$. □

Now we give a proof of Theorem 1.3 in our setting. The key ingredient is the following proposition proved in the discrete case by Chiquet et al. (2017).

**Proposition 5.1.** *Fix a Borel set $A \subseteq \mathbf{R}^d$ with $\mu(A) > 0$ and assume that $\mu|_A$ is $\lambda$-cohesive. Define*

$$u(x) \coloneqq \begin{cases} \mathcal{C}_\mu(A) & \text{if } x \in A; \\ x & \text{if } x \notin A. \end{cases}$$

*Thus $\mathcal{M}_u(\mu)$ is the measure obtained from $\mu$ by consolidating all of the mass in $A$ at $\mathcal{C}_\mu(A)$. Then, for $\mu$-a.e. $x$, we have*

$$u_{\mu,\lambda}(x) = u_{\mathcal{M}_u(\mu),\lambda}(u(x)). \tag{5.4}$$

*Proof.* We follow the argument given by Jiang et al. (2020, proof of Theorem 1(b)). We apply Theorem 4.1 twice. First, by Theorem 4.1 applied to $\mathcal{M}_u(\mu)$, there is an antisymmetric, 1-bounded $w_{\mathrm{out}} \in L^\infty(\mathcal{M}_u(\mu)^{\otimes 2}; \mathbf{R}^d)$ satisfying

$$u_{\mathcal{M}_u(\mu),\lambda}(x) \neq u_{\mathcal{M}_u(\mu),\lambda}(y) \implies w_{\mathrm{out}}(x,y) = \operatorname{sgn}(u_{\mathcal{M}_u(\mu),\lambda}(x) - u_{\mathcal{M}_u(\mu),\lambda}(y)) \tag{5.5}$$

and

$$x - u_{\mathcal{M}_u(\mu),\lambda}(x) = \lambda \int w_{\mathrm{out}}(x,z)\,\mathrm{d}\mathcal{M}_u(\mu)(z)$$

for $\mu$-a.e. $x, y$. Second, by Theorem 4.1 applied to $\mu|_A$, there is an antisymmetric, 1-bounded $w_{\mathrm{in}} \in L^\infty((\mu|_A)^{\otimes 2}; \mathbf{R}^d)$ satisfying

$$x - \mathcal{C}_\mu(A) = \lambda \int_A w_{\mathrm{in}}(x,z)\,\mathrm{d}\mu(z)$$

for $\mu$-a.e. $x \in A$.

Now define

$$w(x,y) = \begin{cases} w_{\mathrm{in}}(x,y) & \text{if } x, y \in A; \\ w_{\mathrm{out}}(u(x), u(y)) & \text{otherwise.} \end{cases}$$

It is clear that $w$ is antisymmetric and 1-bounded since $w_{\mathrm{in}}$ and $w_{\mathrm{out}}$ are. It is also clear from (5.5) that if $u_{\mathcal{M}_u(\mu),\lambda}(u(x)) \neq u_{\mathcal{M}_u(\mu),\lambda}(u(y))$ then $w(x,y) = \mathrm{sgn}(u_{\mathcal{M}_u(\mu),\lambda}(u(x)) - u_{\mathcal{M}_u(\mu),\lambda}(u(y)))$. For $\mu$-a.e. $x \in A$, we have

$$\begin{aligned}
\lambda \int w(x,z)\,\mathrm{d}\mu(z) &= \lambda \int_A w(x,z)\,\mathrm{d}\mu(z) + \lambda \int_{A^{\mathrm{c}}} w(x,z)\,\mathrm{d}\mu(z) \\
&= \lambda \int_A w_{\mathrm{in}}(x,z)\,\mathrm{d}\mu(z) + \lambda \int_{A^{\mathrm{c}}} w_{\mathrm{out}}(\mathcal{C}_\mu(A),z)\,\mathrm{d}\mathcal{M}_u(\mu)(z) \\
&= x - \mathcal{C}_\mu(A) + \mathcal{C}_\mu(A) - u_{\mathcal{M}_u(\mu),\lambda}(\mathcal{C}_\mu(A)) \\
&= x - u_{\mathcal{M}_u(\mu),\lambda}(u(x)),
\end{aligned}$$

while for $\mu$-a.e. $x \notin A$ we have

$$\begin{aligned}
\lambda \int w(x,z)\,\mathrm{d}\mu(z) &= \lambda \int w_{\mathrm{out}}(x,u(z))\,\mathrm{d}\mu(z) \\
&= \lambda \int_A w_{\mathrm{out}}(x,\mathcal{C}_\mu(A))\,\mathrm{d}\mu(z) + \lambda \int_{A^{\mathrm{c}}} w_{\mathrm{out}}(x,z)\,\mathrm{d}\mu(z) \\
&= \lambda \int w_{\mathrm{out}}(x,z)\,\mathrm{d}\mathcal{M}_u(\mu)(z) \\
&= x - u_{\mathcal{M}_u(\mu),\lambda}(u(x)).
\end{aligned}$$

Then (5.4) follows from Theorem 4.1. □

*Proof of Theorem 1.3.* Theorem 1.7 implies that any level set of $u_{\mu,\lambda}$ is $\lambda$-cohesive, and Proposition 5.1 implies that $\mu(A) > 0$ and $\mu|_A$ is $\lambda$-cohesive then $A$ is contained in a single level set of $u_{\mu,\lambda}$. These two facts together imply the statement of the theorem. □

*Proof of Theorem 1.4.* We fix $\lambda \leq \lambda'$. We first confirm that if a measure $\mu$ is $\lambda$-cohesive, then it is $\lambda'$-cohesive. Indeed, if $\mu$ is $\lambda$-cohesive, then there exists a constant $c \in \mathbf{R}^d$ such that for every $u \in L^2(\mu;\mathbf{R}^d)$,

$$J_{\mu,\lambda}(c) \leq J_{\mu,\lambda}(u).$$

Since $\lambda' \geq \lambda$, we deduce that

$$J_{\mu,\lambda'}(c) = J_{\mu,\lambda}(c) \leq J_{\mu,\lambda}(u) \leq J_{\mu,\lambda'}(u).$$

This shows that the constant $c$ minimizes $J_{\mu,\lambda'}$, and by uniqueness of the minimizer, that $\mu$ is $\lambda'$-cohesive.

The proof of Theorem 1.4 is now an application of Theorem 1.3. Outside a set of $\mu$-measure zero, every $x \in \mathbf{R}^d$ satisfies the statement of this theorem both for $\lambda$ and for $\lambda'$. For each such $x \in \mathbf{R}^d$, the measure $\mu|_{V_{u_{\mu,\lambda},x}}$ is $\lambda$-cohesive, so by the previous observation, it is $\lambda'$-cohesive. Applying the second part of Theorem 1.3 with $\lambda'$, we deduce that $\mu(V_{u_{\mu,\lambda},x} \setminus V_{u_{\mu,\lambda'},x}) = 0$, as desired. □

Finally, we prove Proposition 1.8.

*Proof of Proposition 1.8.* By Theorem 1.7, the measure $\mu|_{V_{u_{\mu,\lambda},x}}$ is $\lambda$-cohesive. By Proposition 4.4, we must therefore have that

$$\lambda \geq \lambda_1\left(\mu|_{V_{u_{\mu,\lambda},x}}\right) \geq \frac{R(\mu|_{V_{u_{\mu,\lambda},x}})}{\mu|_{V_{u_{\mu,\lambda},x}}(\mathbf{R}^d)}.$$

Rearranging, we obtain (1.7).

We now turn to (1.8). By Theorem 1.7, the measure $\mathcal{M}_{u_{\mu,\lambda}}(\mu) = (\mathcal{E}_{\mu,u_{\mu,\lambda}})_*(\mu)$ is $\lambda$-shattered. Then Lemma 4.3 implies that, for $\mathcal{M}_{u_{\mu,\lambda}}(\mu)$-a.e. $x, z$ with $x \neq z$, we have

$$|x - z| > \lambda[\mu(V_{u_{\mu,\lambda},x}) + \mu(V_{u_{\mu,\lambda},z})].$$

This yields (1.8) for $\mu$-a.e. $x, z$ with $u_{\mu,\lambda}(x) \neq u_{\mu,\lambda}(z)$. □

## 6. Stability properties

In this section, we prove some stability results for $\lambda_1(\mu)$ and $\lambda_*(\mu)$. For this purpose, we introduce some definitions related to optimal transport. Let $\mu, \widetilde{\mu}$ be finite measures of compact support such that $\mu(\mathbf{R}^d) = \widetilde{\mu}(\mathbf{R}^d)$. We denote by $\Gamma(\mu, \widetilde{\mu})$ the set of Borel measures on $\mathbf{R}^d \times \mathbf{R}^d$ whose first marginal is $\mu(\mathbf{R}^d)\,\mu(\cdot)$ and second marginal is $\mu(\mathbf{R}^d)\,\widetilde{\mu}(\cdot)$. For $p \in [1, \infty)$, the $p$-Wasserstein distance between $\mu$ and $\widetilde{\mu}$ is

$$\mathcal{W}_p(\mu, \widetilde{\mu}) \coloneqq \left(\inf_{\pi \in \Gamma(\mu,\widetilde{\mu})} \int |x - \widetilde{x}|^p \,\mathrm{d}\pi(x, \widetilde{x})\right)^{\frac{1}{p}},$$

while

$$\mathcal{W}_\infty(\mu, \widetilde{\mu}) \coloneqq \inf_{\pi \in \Gamma(\mu,\widetilde{\mu})} \operatorname*{ess\,sup}_{(x,\widetilde{x})\sim\pi} |x - \widetilde{x}|.$$

It is classical to show that for each $p \in [1, \infty]$, this problem admits an optimizer in $\Gamma(\mu, \widetilde{\mu})$. We call any optimizer a $p$-optimal transport plan from $\mu$ to $\widetilde{\mu}$. At least when $p < \infty$ and if the measure $\mu$ is absolutely continuous with respect to the Lebesgue measure, there in fact exists a measurable mapping $T\colon \mathbf{R}^d \to \mathbf{R}^d$ such that the image of the measure $\mu$ by the mapping $(\mathrm{Id}, T)$ is an optimal transport plan from $\mu$ to $\widetilde{\mu}$. In such a case, we call the mapping $T$ an optimal transport map from $\mu$ to $\widetilde{\mu}$. In this paper, we will only make use of optimal transport maps for $p = 1$. In this case, a proof of existence can be found in Ambrosio (2003, Theorem 6.2).

### 6.1 Stability of $\lambda_1$

In this section we prove two stability results for $\lambda_1(\mu)$. The first is that $\lambda_1(\mu)$ is continuous under absolutely continuous perturbations of $\mu$. As is standard in measure theory, for measures $\mu$ and $\widetilde{\mu}$, we write $\widetilde{\mu} \ll \mu$ to mean that $\widetilde{\mu}$ is absolutely continuous with respect to $\mu$.

**Proposition 6.1** (Absolutely continuous perturbations)**.** *Suppose that $\varepsilon < 1$ and $\widetilde{\mu}$ and $\mu$ are finite measures such that $\widetilde{\mu} \ll \mu$,*

$$\left|\frac{\mathrm{d}\widetilde{\mu}}{\mathrm{d}\mu}(z) - 1\right| < \varepsilon,$$

*and*

$$\widetilde{\mu}(\mathbf{R}^d) \geq (1-\varepsilon)\mu(\mathbf{R}^d).$$

*Then*

$$\lambda_1(\widetilde{\mu}) \leq \frac{1+2\varepsilon}{1-\varepsilon}\lambda_1(\mu). \tag{6.1}$$

The second stability result says that $\lambda_1(\mu)$ is continuous under $\mathcal{W}_\infty$ perturbations of $\mu$:

**Proposition 6.2** ($\mathcal{W}_\infty$ perturbations)**.** *Let $\widetilde{\mu}$ and $\mu$ be finite measures of compact support such that $\mu(\mathbf{R}^d) = \widetilde{\mu}(\mathbf{R}^d)$. Then we have*

$$|\lambda_1(\widetilde{\mu}) - \lambda_1(\mu)| \leq \frac{3\mathcal{W}_\infty(\mu,\widetilde{\mu})}{\mu(\mathbf{R}^d)}. \tag{6.2}$$

Now we prove the two preceding propositions.

*Proof of Proposition 6.1.* Let $q$ satisfying (1.16)–(1.17) (for $\mu$) be such that

$$\|q\|_\infty = \lambda_1(\mu)\mu(\mathbf{R}^d).$$

Then by Proposition 4.5 we have

$$\begin{aligned}
\lambda_1(\widetilde{\mu}) &\leq \widetilde{\mu}(\mathbf{R}^d)^{-1} \operatorname*{ess\,sup}_{x,y\sim\mu} \left| q(x,y) + x - y - \fint q(x,z)\,\mathrm{d}\widetilde{\mu}(z) + \fint q(y,z)\,\mathrm{d}\widetilde{\mu}(z) \right| \\
&= \widetilde{\mu}(\mathbf{R}^d)^{-1} \operatorname*{ess\,sup}_{x,y\sim\mu} \left| q(x,y) + x - y - \fint (q(x,z) - q(y,z)) \frac{\mathrm{d}\widetilde{\mu}}{\mathrm{d}\mu}(z)\,\mathrm{d}\mu(z) \right| \\
&= \widetilde{\mu}(\mathbf{R}^d)^{-1} \operatorname*{ess\,sup}_{x,y\sim\mu} \left| q(x,y) - \fint (q(x,z) - q(y,z)) \left( \frac{\mathrm{d}\widetilde{\mu}}{\mathrm{d}\mu}(z) - 1 \right) \mathrm{d}\mu(z) \right| \\
&\leq (1+2\varepsilon)\widetilde{\mu}(\mathbf{R}^d)^{-1}\|q\|_\infty \\
&\leq \frac{1+2\varepsilon}{1-\varepsilon}\lambda_1(\mu),
\end{aligned}$$

as announced. □

*Proof of Proposition 6.2.* Let $q$ satisfying (1.16)–(1.17) (for $\mu$) be such that

$$\|q\|_\infty = \lambda_1(\mu)\mu(\mathbf{R}^d).$$

Let $\pi$ be an $\infty$-optimal transport plan from $\widetilde{\mu}$ to $\mu$. We write the disintegration (Panchenko, Section I.4)

$$\mathrm{d}\pi(x,x') = \mathrm{d}\nu(x' \mid x)\mathrm{d}\widetilde{\mu}(x).$$

Define

$$q_1(x,y) \coloneqq \iint q(w,z)\,\mathrm{d}\nu(z \mid y)\,\mathrm{d}\nu(w \mid x),$$

which is antisymmetric by Fubini's theorem. We note that

$$\|q_1\|_\infty \leq \|q\|_\infty = \lambda_1(\mu)\mu(\mathbf{R}^d).$$

We also have

$$\begin{aligned}\fint q_1(x,y)\,\mathrm{d}\widetilde{\mu}(y) &= \frac{1}{\widetilde{\mu}(\mathbf{R}^d)}\iiint q(w,z)\,\mathrm{d}\nu(z\mid y)\,\mathrm{d}\nu(w\mid x)\,\mathrm{d}\widetilde{\mu}(y)\\ &= \frac{1}{\widetilde{\mu}(\mathbf{R}^d)}\iiint q(w,z)\,\mathrm{d}\nu(w\mid x)\,\mathrm{d}\pi(y,z)\\ &= \frac{1}{\mu(\mathbf{R}^d)}\iint q(w,z)\,\mathrm{d}\mu(z)\,\mathrm{d}\nu(w\mid x)\\ &= \int w\,\mathrm{d}\nu(w\mid x) - \mathcal{C}_\mu(\mathbf{R}^d),\end{aligned}$$

with the last identity by (1.17). Thus we have

$$\left|\fint q_1(x,y)\,\mathrm{d}\widetilde{\mu}(y) - [x-\mathcal{C}_\mu(\mathbf{R}^d)]\right| \le \mathcal{W}_\infty(\mu,\widetilde{\mu}).$$

Therefore, we have by Proposition 4.5 that

$$\begin{aligned}\lambda_1(\widetilde{\mu}) &\le \widetilde{\mu}(\mathbf{R}^d)^{-1}\operatorname*{ess\,sup}_{x,y\sim\mu}\left|q_1(x,y)+x-y-\fint q_1(x,z)\,\mathrm{d}\widetilde{\mu}(z)+\fint q_1(y,z)\,\mathrm{d}\widetilde{\mu}(z)\right|\\ &\le \widetilde{\mu}(\mathbf{R}^d)^{-1}\Bigg[\operatorname*{ess\,sup}_{x,y\sim\mu}\left(|q_1(x,y)|+2\left|\fint q_1(x,z)\,\mathrm{d}\widetilde{\mu}(z)-[x-\mathcal{C}_\mu(\mathbf{R}^d)]\right|\right)\\ &\qquad\qquad +\left|\mathcal{C}_\mu(\mathbf{R}^d)-\mathcal{C}_{\widetilde{\mu}}(\mathbf{R}^d)\right|\Bigg]\\ &\le \widetilde{\mu}(\mathbf{R}^d)^{-1}\left(\lambda_1(\mu)\mu(\mathbf{R}^d)+3\mathcal{W}_\infty(\mu,\widetilde{\mu})\right).\end{aligned}$$

By the symmetry between $\mu$ and $\widetilde{\mu}$, this yields (6.2). □

### 6.2 Stability of $\lambda_*$

We now show that, for atomic measures, $\lambda_*$ is stable under $\mathcal{W}_1$ perturbation of the measures. The key ingredient will be the following continuity property.

**Proposition 6.3.** *Let $\lambda>0$, $M\in(0,\infty)$, and let $\mu,\widetilde{\mu}$ be two Borel probability measures on $\mathbf{R}^d$ such that $\operatorname{supp}\mu,\operatorname{supp}\widetilde{\mu}\subseteq B_M(0)$.*

1. *For every 1-optimal transport plan $\pi$ from $\mu$ to $\widetilde{\mu}$, denoting its disintegration by*

$$\mathrm{d}\pi(x,\widetilde{x}) = \mathrm{d}\nu(\widetilde{x}\mid x)\,\mathrm{d}\mu(x),$$

*we have*

$$\int\left|u_{\mu,\lambda}(x)-\int u_{\widetilde{\mu},\lambda}(\widetilde{x})\,\mathrm{d}\nu(\widetilde{x}\mid x)\right|^2\mathrm{d}\mu(x)\le 16M\mathcal{W}_1(\mu,\widetilde{\mu}). \tag{6.3}$$

2. *There exists a 1-optimal transport plan $\pi$ from $\mu$ to $\widetilde{\mu}$ such that*

$$\int|u_{\mu,\lambda}(x)-u_{\widetilde{\mu},\lambda}(\widetilde{x})|^2\,\mathrm{d}\pi(x,\widetilde{x})\le 16M\mathcal{W}_1(\mu,\widetilde{\mu}).$$

*Proof.* We start with part (1). For $\mu$-a.e. $x \in \mathbf{R}^d$, we put

$$\overline{u}(x) \coloneqq \int u_{\widetilde{\mu},\lambda}(\widetilde{x}) \, \mathrm{d}\nu(\widetilde{x} \mid x).$$

We then observe that

$$\begin{aligned}
\inf J_{\widetilde{\mu},\lambda} &= \int |u_{\widetilde{\mu},\lambda}(\widetilde{x}) - \widetilde{x}|^2 \, \mathrm{d}\widetilde{\mu}(\widetilde{x}) + \lambda \iint |u_{\widetilde{\mu},\lambda}(\widetilde{y}) - u_{\widetilde{\mu},\lambda}(\widetilde{x})| \, \mathrm{d}\widetilde{\mu}(\widetilde{x}) \, \mathrm{d}\widetilde{\mu}(\widetilde{y}) \\
&\geq \int |u_{\widetilde{\mu},\lambda}(\widetilde{x}) - x|^2 \, \mathrm{d}\pi(x, \widetilde{x}) + \lambda \iint |u_{\widetilde{\mu},\lambda}(\widetilde{y}) - u_{\widetilde{\mu},\lambda}(\widetilde{x})| \, \mathrm{d}\widetilde{\mu}(\widetilde{x}) \, \mathrm{d}\widetilde{\mu}(\widetilde{y}) - 4M\mathcal{W}_1(\mu, \widetilde{\mu}) \\
&\geq \int |\overline{u}(x) - x|^2 \, \mathrm{d}\mu(x) + \lambda \iint |\overline{u}(y) - \overline{u}(x)| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) - 4M\mathcal{W}_1(\mu, \widetilde{\mu}),
\end{aligned}$$

where we used the disintegration of $\pi$ and Jensen's inequality in the last step. We can rewrite this as

$$\inf J_{\mu,\lambda} \leq J_{\mu,\lambda}(\overline{u}) \leq \inf J_{\widetilde{\mu},\lambda} + 4M\mathcal{W}_1(\mu, \widetilde{\mu}). \tag{6.4}$$

By symmetry, we conclude that

$$\left|\inf J_{\mu,\lambda} - \inf J_{\widetilde{\mu},\lambda}\right| \leq 4M\mathcal{W}_1(\mu, \widetilde{\mu}). \tag{6.5}$$

Using (4.1) and then (6.4), we thus deduce that

$$\begin{aligned}
\frac{1}{4} \int |u_{\mu,\lambda} - \overline{u}|^2 \, \mathrm{d}\mu &\leq \frac{1}{2}(J_{\mu,\lambda}(u_{\mu,\lambda}) + J_{\mu,\lambda}(\overline{u})) - J_{\mu,\lambda}\left(\frac{u_{\mu,\lambda} + \overline{u}}{2}\right) \\
&\leq \frac{1}{2}\left(\inf J_{\widetilde{\mu},\lambda} - \inf J_{\mu,\lambda}\right) + 2M\mathcal{W}_1(\mu, \widetilde{\mu}).
\end{aligned}$$

Combining this with (6.5), we obtain (6.3).

We now turn to the proof of part (2) of the proposition. We argue by approximation. For every $\varepsilon > 0$, we let $\mu_\varepsilon$ be a measure on $B_M(0)$ that is absolutely continuous with respect to the Lebesgue measure and such that

$$\mathcal{W}_1(\mu, \mu_\varepsilon) \leq \varepsilon. \tag{6.6}$$

We denote by $T_\varepsilon$ and $\widetilde{T}_\varepsilon$ 1-optimal transport maps from $\mu_\varepsilon$ to $\mu$ and from $\mu_\varepsilon$ to $\widetilde{\mu}$, respectively. We have, for every $\delta > 0$, that

$$\begin{aligned}
&\int |u_{\mu,\lambda}(T_\varepsilon(x)) - u_{\widetilde{\mu},\lambda}(\widetilde{T}_\varepsilon(x))|^2 \, \mathrm{d}\mu_\varepsilon(x) \\
&\quad \leq (1 + \delta^{-1}) \int |u_{\mu,\lambda}(T_\varepsilon(x)) - u_{\mu_\varepsilon,\lambda}(x)|^2 \, \mathrm{d}\mu_\varepsilon(x) \\
&\qquad + (1 + \delta) \int |u_{\mu_\varepsilon,\lambda}(x) - u_{\widetilde{\mu},\lambda}(\widetilde{T}_\varepsilon(x))|^2 \, \mathrm{d}\mu_\varepsilon(x).
\end{aligned}$$

Using part (1) of the proposition and (6.6), we deduce that

$$\int |u_{\mu,\lambda}(T_\varepsilon(x)) - u_{\widetilde{\mu},\lambda}(\widetilde{T}_\varepsilon(x))|^2 \, \mathrm{d}\mu_\varepsilon(x) \leq 16M^2(1 + \delta^{-1})\varepsilon + 16M(1 + \delta)\mathcal{W}_1(\mu_\varepsilon, \widetilde{\mu}).$$

The image of the measure $\mu_\varepsilon$ under the mapping $(T_\varepsilon, \widetilde{T}_\varepsilon)$ is a coupling between the measures $\mu$ and $\widetilde{\mu}$. Up to the extraction of a subsequence, we can assume that this image measure converges to a coupling $\pi$ as $\varepsilon \downarrow 0$. Using (6.6) once more, we thus have that

$$\int |u_{\mu,\lambda}(x) - u_{\widetilde{\mu},\lambda}(\widetilde{x})|^2 \,\mathrm{d}\pi(x,\widetilde{x}) \leq 16M(1+\delta)\mathcal{W}_1(\mu,\widetilde{\mu}).$$

Since $\delta > 0$ was arbitrary, the factor $1+\delta$ on the right side can be removed. In order to conclude, we must show that $\pi$ is an optimal transport plan. This follows from a similar line of reasoning: we have

$$\begin{aligned}\int |T_\varepsilon(x) - \widetilde{T}_\varepsilon(\widetilde{x})| \,\mathrm{d}\mu_\varepsilon(x) &\leq \int |T_\varepsilon(x) - x| \,\mathrm{d}\mu_\varepsilon(x) + \int |x - \widetilde{T}_\varepsilon(x)| \,\mathrm{d}\mu_\varepsilon(x) \\ &\leq \varepsilon + \mathcal{W}_1(\mu_\varepsilon, \widetilde{\mu}),\end{aligned}$$

so that, upon passing to the limit $\varepsilon \downarrow 0$, we get

$$\int |x - \widetilde{x}| \,\mathrm{d}\pi(x,\widetilde{x}) \leq \mathcal{W}_1(\mu,\widetilde{\mu}),$$

as desired. □

**Proposition 6.4.** *Let $M \in (0,\infty)$ and suppose that $\mu$ and $\widetilde{\mu}$ are finite, purely atomic probability measures with support in $B_M(0)$. Suppose also that $\mu$ is $\lambda$-shattered, which means that $u_{\mu,\lambda}$ is injective on* $\operatorname{supp}\mu$. *Define*

$$\delta_1 = \operatorname*{ess\,inf}_{\substack{(x,y)\sim\mu^{\otimes 2} \\ x\neq y}} |u_{\mu,\lambda}(x) - u_{\mu,\lambda}(y)| \qquad \text{and} \qquad \delta_2 = \operatorname*{ess\,inf}_{x\sim\mu} \widetilde{\mu}(\{x\}).$$

*If*

$$\mathcal{W}_1(\mu,\widetilde{\mu}) < \frac{\delta_1^2\delta_2}{32M}, \tag{6.7}$$

*then $\widetilde{\mu}$ is also $\lambda$-shattered.*

*Proof.* By Proposition 6.3, there is a 1-optimal transport plan $\pi$ from $\mu$ to $\widetilde{\mu}$ such that

$$\int |u_{\mu,\lambda}(x) - u_{\widetilde{\mu},\lambda}(\widetilde{x})|^2 \,\mathrm{d}\pi(x,\widetilde{x}) \leq 16M\mathcal{W}_1(\mu,\widetilde{\mu}). \tag{6.8}$$

Suppose there are distinct points $\widetilde{x}_1, \widetilde{x}_2 \in \operatorname{supp}\widetilde{\mu}$ (a finite set) such that $u_{\widetilde{\mu},\lambda}(\widetilde{x}_1) = u_{\widetilde{\mu},\lambda}(\widetilde{x}_2)$. Then we have by the triangle inequality that

$$|u_{\mu,\lambda}(x_1) - u_{\widetilde{\mu},\lambda}(\widetilde{x}_1)|^2 + |u_{\mu,\lambda}(x_2) - u_{\widetilde{\mu},\lambda}(\widetilde{x}_2)|^2 \geq \frac{1}{2}|u_{\mu,\lambda}(x_1) - u_{\mu,\lambda}(x_2)|^2 \geq \delta_1^2/2.$$

Denote the disintegration of $\pi$ over the first coordinate by

$$\mathrm{d}\pi(x,\widetilde{x}) = \mathrm{d}\widetilde{\nu}(x \mid \widetilde{x})\mathrm{d}\widetilde{\mu}(\widetilde{x}).$$

Then we have

$$\begin{aligned}\delta_1^2\delta_2 &\le \frac{1}{2}\delta_1^2\left(\widetilde{\mu}(\{x_1\}) + \widetilde{\mu}(\{x_2\})\right)\\ &\le \int_{\widetilde{x}\in\{\widetilde{x}_1,\widetilde{x}_2\}} \iint \left[|u_{\mu,\lambda}(x_1) - u_{\widetilde{\mu},\lambda}(\widetilde{x})|^2 + |u_{\mu,\lambda}(x_2) - u_{\widetilde{\mu},\lambda}(\widetilde{x})|^2\right] \mathrm{d}\widetilde{\nu}(x_1 \mid \widetilde{x})\, \mathrm{d}\widetilde{\nu}(x_2 \mid \widetilde{x})\, \mathrm{d}\widetilde{\mu}(\widetilde{x})\\ &= 2\int_{(x,\widetilde{x})\in\mathbf{R}^d\times\{\widetilde{x}_1,\widetilde{x}_2\}} |u_{\mu,\lambda}(x) - u_{\widetilde{\mu},\lambda}(\widetilde{x})|^2 \,\mathrm{d}\pi(x,\widetilde{x})\\ &\le 32M\mathcal{W}_1(\mu,\widetilde{\mu}),\end{aligned}$$

with the last inequality by (6.8). But this contradicts (6.7). Therefore, $u_{\widetilde{\mu},\lambda}$ must be injective on $\operatorname{supp}\widetilde{\mu}$. This means that $\widetilde{\mu}$ is $\lambda$-shattered. □

### 6.3 Proofs of Theorems 1.10 and 1.1

Now we can prove our main stability results, Theorems 1.10 and 1.1.

*Proof of Theorem 1.10.* For $i \in \{1,\dots,I\}$, define

$$q_{i,N} = \#\{n \in \{1,\dots,N\} \mid X_n \in \overline{U_i}\}.$$

By the law of large numbers, we have with probability 1 that

$$\lim_{N\to\infty} N^{-1} q_{i,N} = \mu(\overline{U_i}). \tag{6.9}$$

Define

$$\widetilde{\mu}_{N,i} = \frac{1}{q_{i,N}}\mu_N|_{\overline{U_i}}.$$

By (6.9) and Theorem 1.1 of García Trillos and Slepčev (2015) for $d \ge 2$, or a similar result using the Glivenko–Cantelli theorem (Durrett, 2010, Theorem 2.4.7) for $d = 1$, we have that

$$\widetilde{\mu}_{N,i} \to \frac{1}{\mu(\overline{U_i})}\mu|_{\overline{U_i}}$$

in probability as $N \to \infty$ with respect to the $\mathcal{W}^\infty$ topology. Therefore, we have that

$$\lim_{N\to\infty} \lambda_1(\widetilde{\mu}_{N,i}) = \lambda_1(\mu|_{\overline{U_i}})$$

in probability by Proposition 6.2. On the other hand, we have that

$$\lim_{N\to\infty} |\lambda_1(\widetilde{\mu}_{N,i}) - \lambda_1(\mu_N|_{\overline{U}_i})| = 0$$

in probability by Proposition 6.1. Combining the last two displays, we see that

$$\lambda_1(\mu_N|_{\overline{U}_i}) \to \lambda_1(\mu|_{\overline{U_i}}) \tag{6.10}$$

as $N \to \infty$. On the other hand, it is clear from the law of large numbers that

$$\lim_{N\to\infty} \mathcal{M}_u(\mu_N) = \mathcal{M}_u(\mu)$$

in probability with respect to the $\mathcal{W}^1$ topology. Therefore, we have from Proposition 6.4 that

$$\lim_{N\to\infty} \lambda_*(\mathcal{M}_u(\mu_N)) = \lambda_*(\mathcal{M}_u(\mu)) \tag{6.11}$$

in probability. Together, (6.10) and (6.11) complete the proof of the theorem. □

*Proof of Theorem 1.1.* We set $\lambda_c := \lambda_1(\mu)$. Using Theorem 1.10 with $u = 0$, we see that $\lambda_1(\mu_N)$ tends to $\lambda_c$ in probability as $N$ tends to infinity. Part (1) of Theorem 1.1 thus follows.

We now turn to the proof of part (2), and fix $\lambda > \lambda_c$. By the definition of $\lambda_c$ and Theorem 1.4, the range of $u_{\mu,\lambda}$ contains at least two points. We decompose the rest of the proof into two steps.

*Step 1.* We show that the range of $u_{\mu,\lambda}$ contains at least three points. We argue by contradiction, assuming that the range of $u_{\mu,\lambda}$ is made of exactly two points. Notice that the measure $\mu$ is symmetric under rotations about the first coordinate axis, and under negations of any of the canonical basis vectors. By the uniqueness of the minimizer, it must be that $u_{\mu,\lambda}$ is invariant under these transformations. As we now argue, the range of $u_{\mu,\lambda}$ must therefore be a subset of the first coordinate axis. Indeed, this is easiest to see if $d \geq 3$, since otherwise the range of $u_{\mu,\lambda}$ would have to contain a circle, and in particular would contain infinitely many points. Suppose now that $d = 2$ and that the range of $u_{\mu,\lambda}$ is made of exactly two points. By the invariance under reflections, the only possibility for the support to not be a subset of the first coordinate axis is that the two points forming the support of $u_{\mu,\lambda}$ are on the second coordinate axis; but in this case, the two level sets of $u_{\mu,\lambda}$ would each have to contain half of each of the balls, and this would contradict Proposition 1.8.

Using again the invariance under reflections, we deduce that there exists $\rho > 0$ such that the range of $u_{\mu,\lambda}$ is the set $\{-\rho e_1, \rho e_1\}$. Let $E := u_{\mu,\lambda}^{-1}(\rho e_1)$. Again by symmetry, it must be that, up to a set of null $\mu$-measure, we have $u_{\mu,\lambda}^{-1}(-\rho e_1) = -E$, and $\mu(E) = \mu(-E) = 1/2$, so that

$$\iint |u_{\mu,\lambda}(x) - u_{\mu,\lambda}(y)| \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \rho. \tag{6.12}$$

Moreover,

$$\begin{aligned}
\int_E |\rho e_1 - x|^2 \,\mathrm{d}\mu(x) &= \int_{E\cap B_1(re_1)} |\rho e_1 - x|^2 \,\mathrm{d}\mu(x) + \int_{E\cap B_1(-re_1)} |\rho e_1 - x|^2 \,\mathrm{d}\mu(x) \\
&= \int_{E\cap B_1(re_1)} |\rho e_1 - x|^2 \,\mathrm{d}\mu(x) + \int_{(-E)\cap B_1(re_1)} |\rho e_1 + x|^2 \,\mathrm{d}\mu(x) \\
&\geq \int_{E\cap B_1(re_1)} |\rho e_1 - x|^2 \,\mathrm{d}\mu(x) + \int_{(-E)\cap B_1(re_1)} |\rho e_1 - x|^2 \,\mathrm{d}\mu(x) \\
&\geq \int_{B_1(re_1)} |\rho e_1 - x|^2 \,\mathrm{d}\mu(x),
\end{aligned}$$

since $E \cap (-E)$ is a $\mu$-null set. This yields that

$$\int |u_{\mu,\lambda} - x|^2 \,\mathrm{d}\mu(x) \geq \int_{B_1(re_1)} |\rho e_1 - x|^2 \,\mathrm{d}\mu(x) + \int_{B_1(-re_1)} |-\rho e_1 - x|^2 \,\mathrm{d}\mu(x).$$

Combining this with (6.12), we see that we must have, up to a $\mu$-null set, that $E = B_1(re_1)$. In other words, the minimizer $u_{\mu,\lambda}$ maps $B_1(re_1)$ to $\rho e_1$ and $B_1(-re_1)$ to $-\rho e_1$.

By Theorem 1.7, we must therefore have that

$$\text{the measure } \tfrac{1}{2}\delta_{-re_1} + \tfrac{1}{2}\delta_{re_1} \text{ is } \lambda\text{-shattered,} \tag{6.13}$$

and

$$\text{the measure } \mu|_{B_1(re_1)} \text{ is } \lambda\text{-cohesive.} \tag{6.14}$$

By Proposition 2.1, the requirement in (6.13) imposes that $\lambda \le 2r$. By Proposition 2.7, the requirement in (6.14) imposes that $\lambda \ge 2\gamma_d$. Since we assume that $r < \gamma_d$, we have reached a contradiction.

*Step 2.* By the result of the previous step, there exist $c_1, c_2, c_3 \in \mathbf{R}^d$ and $\eta > 0$ such that for every $i \neq j \in \{1,2,3\}$, we have $|c_i - c_j| \ge 9\eta$, and

$$\xi := \min\left(\mu[u_{\mu,\lambda}^{-1}(B_\eta(c_1))], \mu[u_{\mu,\lambda}^{-1}(B_\eta(c_2))], \mu[u_{\mu,\lambda}^{-1}(B_\eta(c_3))]\right) > 0. \tag{6.15}$$

Since the measure $\mu$ is absolutely continuous with respect to the Lebesgue measure, there exists a 1-optimal transport map from $\mu$ to $\mu_N$, which we denote by $T_N$. By Proposition 6.3, we have

$$\int |u_{\mu,\lambda}(x) - u_{\mu_N,\lambda}(T_N(x))| \, \mathrm{d}\mu(x) \le 16M\mathcal{W}_1(\mu,\mu_N).$$

In particular, for each $i \in \{1,2,3\}$, we have

$$\int_{u_{\mu,\lambda}^{-1}(B_\eta(c_i))} |c_i - u_{\mu_N,\lambda}(T_N(x))| \, \mathrm{d}\mu(x) \le 16M\mathcal{W}_1(\mu,\mu_N) + \eta\mu[u_{\mu,\lambda}^{-1}(B_\eta(c_i))].$$

Recall that $\mathcal{W}_1(\mu,\mu_N)$ tends to zero in probability as $N$ tends to infinity (see for instance Dudley, 1968). For every $\varepsilon > 0$, we can therefore let $N$ be sufficiently large that with probability at least $1-\varepsilon$, we have

$$\int_{u_{\mu,\lambda}^{-1}(B_\eta(c_i))} |c_i - u_{\mu_N,\lambda}(T_N(x))| \, \mathrm{d}\mu(x) \le 2\eta\mu[u_{\mu,\lambda}^{-1}(B_\eta(c_i))].$$

In particular, by Chebyshev's inequality,

$$\int_{u_{\mu,\lambda}^{-1}(B_\eta(c_i))} \mathbf{1}_{\{|c_i - u_{\mu_N,\lambda}(T_N(x))| \ge 4\eta\}} \, \mathrm{d}\mu(x) \le \frac{1}{2}\mu[u_{\mu,\lambda}^{-1}(B_\eta(c_i))];$$

that is,

$$\int_{u_{\mu,\lambda}^{-1}(B_\eta(c_i))} \mathbf{1}_{\{|c_i - u_{\mu_N,\lambda}(T_N(x))| < 4\eta\}} \, \mathrm{d}\mu(x) \ge \frac{1}{2}\mu[u_{\mu,\lambda}^{-1}(B_\eta(c_i))].$$

Recalling that $T_N$ is an optimal transport map from $\mu$ to $\mu_N$, we see that the left side is bounded from above by

$$\int \mathbf{1}_{\{|c_i - u_{\mu_N,\lambda}(x)| < 4\eta\}} \, \mathrm{d}\mu_N(x) = \frac{1}{N}\left|\{n \le N \ : \ |c_i - u_{\mu_N,\lambda}(X_n)| < 4\eta\}\right|.$$

Recalling also the definition of $\xi$, we have shown that, with probability at least $1-\varepsilon$, the following holds for every $N$ sufficiently large and $i \in \{1,2,3\}$:

$$\frac{1}{N}\left|\{n \le N \ : \ |c_i - u_{\mu_N,\lambda}(X_n)| < 4\eta\}\right| \ge \frac{\xi}{2}.$$

Since $|c_i - c_j| \ge 9\eta$ for every $i \ne j$, this yields the desired result, up to a redefinition of $\xi$. □

To conclude, we give a counterpart to Theorem 1.1 in the case when the two balls are sufficiently far apart.

**Proposition 6.5.** *Let $r > 2^{1-\frac{1}{d}}$, $\mu$ be the uniform measure on $B_1(-r\mathrm{e}_1) \cup B_1(r\mathrm{e}_1) \subseteq \mathbf{R}^d$, $(X_n)_{n\in\mathbf{N}}$ be independent random variables with law $\mu$, and for every integer $N \ge 1$, define the empirical measure*

$$\mu_N \coloneqq \frac{1}{N}\sum_{n=1}^{N} \delta_{X_n}.$$

*If $\lambda \in (2^{2-\frac{1}{d}}, 2r)$, then with high probability, the level sets of $u_{\mu_N,\lambda}$ are the sets*

$$\{X_n,\ n \le N\} \cap B_1(-r\mathrm{e}_1) \quad \text{and} \quad \{X_n,\ n \le N\} \cap B_1(r\mathrm{e}_1).$$

*Proof.* By Theorem 1.7, the level sets of the function $u_{\mu,\lambda}$ are, up to $\mu$-null modifications, the two balls $B_1(-r\mathrm{e}_1)$ and $B_1(r\mathrm{e}_1)$, if and only if (6.13) and (6.14) hold. By Proposition 2.1, the first condition holds whenever $\lambda < 2r$, and by Proposition 2.7, the second condition holds whenever $\lambda > 2\cdot 2^{1-\frac{1}{d}}$. The result then follows by an application of Theorem 1.10. □

## 7. Technical lemmas

In this section we collect a few additional technical lemmas to avoid distracting from the flow of the paper.

**Lemma 7.1.** *Let $\mu$ be a finite Borel measure on $\mathbf{R}^d$. Let $u_1, u_2 \colon \mathbf{R}^d \to \mathbf{R}^d$ be such that $u_1(x) = u_2(x)$ for $\mu$-a.e. $x$, and let $A_1, A_2 \subseteq \mathbf{R}^d$ be Borel sets such that, for each $i = 1,2$, we have $\mu(\mathbf{R}^d \setminus A_i) = 0$ and $V_{u_i,x} \cap A_i$ is $\mu$-regular for $\mu$-a.e. $x$. If we define $\mathcal{E}^{(i)}(x) \coloneqq \mathcal{C}_\mu(V_{u_i,x} \cap A_i)$, then $\mathcal{E}^{(1)}(x) = \mathcal{E}^{(2)}(x)$ for $\mu$-a.e. $x$.*

*Proof.* Let $B$ be the set of all $x \in \mathbf{R}^d$ such that $u_1(x) = u_2(x)$. Note that $\mu(A_1 \cap A_2 \cap B) = \mu(\mathbf{R}^d)$. Let $x \in A_1 \cap A_2 \cap B$. We claim that $\mathcal{E}^{(1)}(x) = \mathcal{E}^{(2)}(x)$. We consider two cases.

First, suppose that there is some $i$ such that $\mu(V_{u_i,x}) > 0$, and assume wlog that $i = 1$. Then we have $V_{u_1,x} \cap B = \{y \in B \ : \ u_1(x) = u_1(y)\} = \{y \in B \ : \ u_2(x) = u_2(x)\} = V_{u_2,x} \cap B$, since $u_1(z) = u_2(z)$ for all $z \in B$. Since $\mu(\mathbf{R}^d \setminus B) = 0$, this implies that $\mathcal{C}_\mu(V_{u_i,x} \cap A_i)$ does not depend on $i$, since changing a positive-measure set by a set of measure zero does not change its centroid.

On the other hand, if $x$ is such that $\mu(V_{u_1,x}) = \mu(V_{u_2,x}) = 0$, then $V_{u_i,x} \cap A_i = \{x\}$ for each $i$ by $\mu$-regularity, and hence $\mathcal{C}_\mu(V_{u,x} \cap A_i) = x$ for each $i$.

Thus we have shown that the set of $x$ such that $\mathcal{E}^{(1)}(x) \ne \mathcal{E}^{(2)}(x)$ is contained in $\mathbf{R}^d \setminus (A_1 \cap A_2 \cap B)$, which has $\mu$-measure 0. □

**Lemma 7.2.** *For any finite Borel measure $\mu$ and any $\lambda \geq 0$, the function $J_{\mu,\lambda} \colon L^2(\mu; \mathbf{R}^d) \to \mathbf{R}$ defined in (1.2) is continuous.*

*Proof.* Let $u_1, u_2 \in L^2(\mu; \mathbf{R}^d)$. We have by the triangle, reverse triangle, and Cauchy–Schwarz inequalities that

$$\begin{aligned}\left| \iint |u_1(x) - u_1(y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) - \iint |u_2(x) - u_2(y)| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \right| \\ \leq \iint (|u_1(x) - u_2(x)| + |u_1(y) - u_2(y)|) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \leq 2\mu(\mathbf{R}^d)^{3/2} \|u_1 - u_2\|_{L^2(\mu;\mathbf{R}^d)}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\left| \int |u_1(x) - x|^2 \, \mathrm{d}\mu(x) - \int |u_2(x) - x|^2 \, \mathrm{d}\mu(x) \right| \\ \leq 2 \int |u_1(x) - u_2(x)|^2 \, \mathrm{d}\mu(x) \leq 2\|u_1 - u_2\|^2_{L^2(\mu;\mathbf{R}^d)}.\end{aligned}$$

Together, the last two displays imply that $J_{\mu,\lambda}$ is continuous. □

**Acknowledgments**

We warmly thank Antonio De Rosa for many interesting discussions. AD was partially supported by the NSF Mathematical Sciences Postdoctoral Fellowship program under grant no. DMS-2002118. JCM was partially supported by the NSF grant DMS-1954357.

## References

Daniel Aloise, Amit Deshpande, Pierre Hansen, and Preyas Popat. NP-hardness of Euclidean sum-of-squares clustering. *Mach. Learn.*, 75(2):245–248, 2009.

Luigi Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, volume 1812 of *Lecture Notes in Mathematics*, page 1–52. Springer, Berlin, 2003.

Pranjal Awasthi, Afonso S Bandeira, Moses Charikar, Ravishankar Krishnaswamy, Soledad Villar, and Rachel Ward. Relax, no need to round: Integrality of clustering formulations. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, page 191–200, 2015.

Eric C. Chi and Kenneth Lange. Splitting methods for convex clustering. *J. Comput. Graph. Statist.*, 24(4):994–1013, 2015.

Eric C. Chi and Stefan Steinerberger. Recovering trees with convex clustering. *SIAM J. Math. Data Sci.*, 1(3):383–407, 2019.

Julien Chiquet, Pierre Gutierrez, and Guillem Rigaill. Fast tree inference with weighted fusion penalties. *J. Comput. Graph. Statist.*, 26(1):205–216, 2017.

Antonio De Rosa and Aida Khajavirad. The ratio-cut polytope and K-means clustering. *SIAM J. Optim.*, 32(1):173–203, 2022.

R. M. Dudley. The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist.*, 40: 40–50, 1968.

Steven R. Dunbar. The average distance between points in geometric figures. *College Math. J.*, 28(3):187–197, 1997.

Alexander Dunlap and Jean-Christophe Mourrat. Local versions of sum-of-norms clustering. *SIAM J. Math. Data Sci.*, 4(4):1250–1271, 2022.

Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010.

Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*. Studies in Mathematics and its Applications, Vol. 1. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, 1976. Translated from the French.

Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

Nicolás García Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in $\infty$-transportation distance. *Canad. J. Math.*, 67(6):1358–1383, 2015.

Geoffrey Grimmett and David Stirzaker. *One Thousand Exercises in Probability*. Oxford University Press, 2020.

Toby Hocking, Jean-Philippe Vert, Francis R. Bach, and Armand Joulin. Clusterpath: an algorithm for clustering using convex fusion penalties. In Lise Getoor and Tobias Scheffer, editors, *Proc. 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, page 745–752. Omnipress, 2011.

Takayuki Iguchi, Dustin G. Mixon, Jesse Peterson, and Soledad Villar. Probably certifiably correct $k$-means clustering. *Math. Prog.*, 165(2, Ser. A):605–642, 2017.

Tao Jiang and Stephen Vavasis. Certifying clusters from sum-of-norms clustering. Preprint, 2020. arXiv:2006.11355.

Tao Jiang, Stephen Vavasis, and Chen Wen Zhai. Recovery of a mixture of Gaussians by sum-of-norms clustering. *J. Mach. Learn. Res.*, 21:Paper No. 225, 16, 2020.

Xiaodong Li, Yang Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. When do birds of a feather flock together? $k$-means, proximity, and conic programming. *Math. Prog.*, 179(1-2, Ser. A):295–341, 2020.

F. Lindsten, H. Ohlsson, and L. Ljung. Clustering using sum-of-norms regularization: With application to particle filter output computation. In *2011 IEEE Statistical Signal Processing Workshop (SSP)*, page 201–204, 2011.

Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. The planar $k$-means problem is NP-hard. In *WALCOM—Algorithms and computation*, volume 5431 of *Lecture Notes in Computer Science*, page 274–285. Springer, Berlin, 2009.

Abhinav Nellore and Rachel Ward. Recovery guarantees for exemplar-based clustering. *Inform. and Comput.*, 245:165–180, 2015.

Canh Hao Nguyen and Hiroshi Mamitsuka. On convex clustering solutions. Preprint, 2021. arXiv:2105.08348.

Ashkan Panahi, Devdatt P. Dubhashi, Fredrik D. Johansson, and Chiranjib Bhattacharyya. Clustering by sum of norms: Stochastic incremental algorithm, convergence and cluster recovery. In Doina Precup and Yee Whye Teh, editors, *Proc. 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proc. Mach. Learn. Res.*, page 2769–2777, 2017.

Dmitry Panchenko. *Lecture notes on probability theory.* URL https://sites.google.com/site/panchenkomath/.

Kristiaan Pelckmans, Joseph De Brabanter, Bart De Moor, and Johan Suykens. Convex clustering shrinkage. In *Workshop on Statistics and optimization of clustering Workshop (PASCAL)*, 2005. URL ftp://ftp.esat.kuleuven.ac.be/sista/kpelckma/ccs_pelckmans2005.pdf.

Peter Radchenko and Gourab Mukherjee. Convex clustering via $l_1$ fusion penalization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(5):1527–1546, 2017.

Luis A. Santaló. *Integral geometry and geometric probability.* Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976.

Defeng Sun, Kim-Chuan Toh, and Yancheng Yuan. Convex clustering: model, theoretical guarantee and efficient algorithm. *J. Mach. Learn. Res.*, 22:Paper No. 9, 32, 2021.

Kean Ming Tan and Daniela Witten. Statistical properties of convex clustering. *Electron. J. Stat.*, 9(2):2324–2347, 2015.

Changbo Zhu, Huan Xu, Chenlei Leng, and Shuicheng Yan. Convex optimization procedure for clustering: Theoretical revisit. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, page 1619–1627, 2014.