

A VARIATIONAL APPROACH TO THE CONSISTENCY OF SPECTRAL CLUSTERING

NICOLÁS GARCÍA TRILLOS AND DEJAN SLEPČEV

ABSTRACT. This paper establishes the consistency of spectral approaches to data clustering. We consider clustering of point clouds obtained as samples of a ground-truth measure. A graph representing the point cloud is obtained by assigning weights to edges based on the distance between the points they connect. We investigate the spectral convergence of both unnormalized and normalized graph Laplacians towards the appropriate operators in the continuum domain. We obtain sharp conditions on how the connectivity radius can be scaled with respect to the number of sample points for the spectral convergence to hold. We also show that the discrete clusters obtained via spectral clustering converge towards a continuum partition of the ground truth measure. Such continuum partition minimizes a functional describing the continuum analogue of the graph-based spectral partitioning. Our approach, based on variational convergence, is general and flexible.

1. INTRODUCTION

Clustering is one of the basic problems of statistics and machine learning: having a collection of n data points and a measure of their pairwise similarity the task is to partition the data into k meaningful groups. There is a variety of criteria for the quality of partitioning and a plethora of clustering algorithms, overviewed in [14, 34, 50, 51]. Among most widely used are centroid based (for example the k -means algorithm), agglomeration based (or hierarchical) and graph based ones. Many graph partitioning approaches are based on dividing the graph representing the data into clusters of balanced sizes which have as few as possible edges between them [4, 5, 24, 35, 39, 40, 49]. Spectral clustering is a relaxation of minimizing graph cuts, which in any of its variants, [29, 35, 47], consists of two steps. The first step is the embedding step where data points are mapped to a euclidean space by using the spectrum of a *graph Laplacian*. In the second step, the actual clustering is obtained by applying a clustering algorithm like k -means to the transformed points.

The input of a spectral clustering algorithm is a weight matrix W which captures the similarity relation between the data points. Typically, the choice of edge weights depends on the distance between the data points and a parameter ε which determines the length scale over which points are connected. We assume that the data set is a random sample of an underlying ground-truth measure. We investigate the convergence of spectral clustering as the number of available data points goes to infinity.

For any given clustering procedure, a natural and important question is whether the procedure is consistent. That is, if it is true that as more data is collected, the partitioning of the data into groups obtained converges to some meaningful partitioning in the limit. Despite the abundance of clustering procedures in the literature, not many results establish their consistency in the nonparametric setting, where the data is assumed to be obtained from a unknown general distribution. Consistency of k -means clustering was established by Pollard [31]. Consistency of k -means clustering for paths

Date: August 7, 2015.

1991 Mathematics Subject Classification. 49J55, 49J45, 60D05, 68R10, 62G20.

Key words and phrases. spectral clustering, graph Laplacian, point cloud, discrete to continuum limit, Gamma-convergence, Dirichlet energy, random geometric graph .

with regularization was recently studied by Thorpe, Theil and Cade [44], using a similar viewpoint to those of this paper. Consistency for a class of single linkage clustering algorithms was shown by Hartigan [22]. Arias-Castro and Pelletier have proved the consistency of maximum variance unfolding [3]. Pointwise estimates between graph Laplacians and the continuum operators were studied by Belkin and Niyogi [8], Coifman and Lafon [11], Giné and Koltchinskii [18], Hein, Audibert and von Luxburg [23], and Singer [37]. Spectral convergence was studied in the works of Ting, Huang, and Jordan [45], Belkin and Niyogi [7] on the convergence of Laplacian eigenmaps, von Luxburg, Belkin and Bousquet on graph Laplacians, and of Singer and Wu [38] on connection graph Laplacian. The convergence of the eigenvalues and eigenvectors these works obtain is of great relevance to machine learning. However obtaining practical and rigorous rates at which the connectivity length scale $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$ remained an open problem. Also relevant to point cloud analysis are studies of Laplacians on discretized manifolds by Burago, Ivanov and Kurylev [10] who obtain precise error estimates for eigenvalues and eigenvectors.

Recently the authors in [15], and together with Laurent, von Brecht and Bresson in [17], introduced a framework for showing the consistency of clustering algorithms based on minimizing an objective functional on graphs. In [17] they applied the technique to Cheeger and Ratio cuts. Here the framework of [15, 17] is used to prove new results on consistency of spectral clustering, which establish the (almost) optimal rate at which the connectivity radius ε can be taken to 0 as $n \rightarrow \infty$. We prove the convergence of the spectrum of the graph Laplacian towards the spectrum of a corresponding continuum operator. An important element of our work is that we establish the convergence of the discrete clusters obtained via spectral clustering to their continuum counterparts. That is, as the number of data points $n \rightarrow \infty$ the discrete clusters (obtained via spectral clustering) are shown to converge towards continuum objects (measures), which themselves are obtained via a clustering procedure in the continuum setting (performed on the ground truth measure). That is, the discrete clusters are shown to converge to continuum clusters obtained via spectral clustering procedure with full information (ground truth measure) available. We obtain results for unnormalized (Theorem 1.2), and normalized (Theorems 1.5 and 1.7) graph Laplacians. The bridge connecting the spectrum of the graph Laplacian and the spectrum of a limiting operator in the continuum is built by using the notion of variational convergence known as Γ -convergence. The setting of Γ -convergence, combined with techniques of optimal transportation, provides an effective viewpoint to address a range of consistency and stability problems based on minimizing objective functionals on a random sample of a measure.

1.1. Description of spectral clustering. Let $V = \{x_1, \dots, x_n\}$ be a set of vertices and let $W \in \mathbb{R}^{n \times n}$ be a symmetric matrix with non-negative entries. We define $\mathcal{D} \in \mathbb{R}^{n \times n}$, the *degree matrix* of the weighted graph (V, W) , to be the diagonal matrix with $\mathcal{D}_{ii} = \sum_j W_{i,j}$ for every i . Also, we define L , the *unnormalized graph Laplacian* matrix of the weighted graph (V, W) , to be

$$(1.1) \quad L := \mathcal{D} - W.$$

We also consider the matrices N^{sym} and N^{rw} given by

$$N^{sym} := \mathcal{D}^{-1/2} L \mathcal{D}^{-1/2}, \quad N^{rw} := \mathcal{D}^{-1} L,$$

both of which we refer to as *normalized graph Laplacians*. The superscript *sym* indicates the fact that N^{sym} is symmetric, whereas the superscript *rw* indicates the fact that N^{rw} is connected to the transition probabilities of a random walk that can be defined on the graph. Each of the matrices L, N^{sym}, N^{rw} is used in a version of spectral clustering. The so called unnormalized spectral clustering uses the spectrum of the unnormalized graph Laplacian to embed the point cloud into a lower dimensional space, typically a method like k -means on the embedded points then provides the desired clusters (see [47]). This is Algorithm 1 below.

Algorithm 1 Unnormalized spectral clustering

- Input:** Number of clusters k and similarity matrix W .
- Construct the unnormalized graph Laplacian L .
 - Compute the eigenvectors u_1, \dots, u_k of L associated to the k smallest (nonzero) eigenvalues of L .
 - Define the matrix $U \in \mathbb{R}^{k \times n}$, where the i -th row of U is the vector u_i .
 - For $i = 1, \dots, n$, let $y_i \in \mathbb{R}^k$ be the i -th column of U .
 - Use the k -means algorithm to partition the set of points $\{y_1, \dots, y_n\}$ into k groups, that we denote by G_1, \dots, G_k .
- Output:** Clusters G_1, \dots, G_k .
-

In the same spirit, the normalized graph Laplacians are used. An algorithm for normalized spectral clustering using N^{sym} was introduced in [29] (see Algorithm 2), and an algorithm using N^{rw} was introduced in [35] (see Algorithm 3).

Algorithm 2 Normalized spectral clustering as defined in [29]

- Input:** Number of clusters k and similarity matrix W .
- Construct the normalized graph Laplacian N^{sym} .
 - Compute the eigenvectors u_1, \dots, u_k of N^{sym} associated to the k smallest (nonzero) eigenvalues of N^{sym} .
 - Define the matrix $U \in \mathbb{R}^{k \times n}$, where the i -th row of U is the vector u_i .
 - Construct the matrix V by normalizing the columns of U so that the columns of V have all euclidean norm equal to one.
 - For $i = 1, \dots, n$, let $y_i \in \mathbb{R}^k$ be the i -th column of V .
 - Use the k -means algorithm to partition the set of points $\{y_1, \dots, y_n\}$ into k groups that we denote by G_1, \dots, G_k .
- Output:** Clusters G_1, \dots, G_k .
-

Algorithm 3 Normalized spectral clustering as defined in [35]

Same as Algorithm 1 but using the normalized graph Laplacian N^{rw} instead of L .

Spectral properties of graph Laplacians have connections to balanced graph cuts. For example, the spectrum of N^{rw} is shown to be connected to the Ncut problem, whereas the spectrum of L is connected to RatioCut (see [47]). A probabilistic interpretation of the spectrum of N^{rw} may be found in [28]. In addition, connections between normalized graph Laplacians, data parametrization and dimensionality reduction via diffusion maps are developed in [25].

We now present some facts about the matrices L , N^{sym} and N^{rw} , all of which may be found in [47]. First of all L is a positive semidefinite symmetric matrix. In fact for every vector $u \in \mathbb{R}^n$

$$(1.2) \quad \langle Lu, u \rangle = \frac{1}{2} \sum_{i,j} W_{i,j} (u_i - u_j)^2,$$

where on the left hand side we are using the usual inner product in \mathbb{R}^n . The smallest eigenvalue of L is equal to zero, and its multiplicity is equal to the number of connected components of the

weighted graph. The matrix N^{sym} is symmetric and positive semidefinite as well. Moreover, for every $u \in \mathbb{R}^n$

$$(1.3) \quad \langle N^{sym}u, u \rangle = \frac{1}{2} \sum_{i,j} W_{i,j} \left(\frac{u_i}{\mathcal{D}_{ii}} - \frac{u_j}{\mathcal{D}_{jj}} \right)^2.$$

In addition, 0 is an eigenvalue of N^{sym} , with multiplicity equal to the number of connected components of the weighted graph. The vector $\mathcal{D}^{1/2}\mathbf{1}$ (where $\mathbf{1}$ is the vector with all entries equal to one) is an eigenvector of N^{sym} with eigenvalue 0.

The two forms of normalized graph Laplacians are closely related due to the correspondence between the spectrums of N^{sym} and N^{rw} . In fact, it is straightforward to show that

$$(1.4) \quad N^{rw}u = \lambda u \quad \text{if and only if} \quad N^{sym}w = \lambda w, \quad \text{where } w = \mathcal{D}^{1/2}u.$$

That is, N^{sym} and N^{rw} have the same eigenvalues, and there is a simple relation between their corresponding eigenvectors.

1.2. Spectral clustering of point clouds. Let $V = \{x_1, \dots, x_n\}$ be a point cloud in \mathbb{R}^d . To give a weighted graph structure to the set V , we consider a kernel η , that is, we consider $\eta : \mathbb{R}^d \rightarrow [0, \infty)$ a radially symmetric, radially decreasing function decaying to zero sufficiently fast. The kernel is appropriately rescaled to take into account data density. In particular, let η_ε depend on the length scale ε where we take $\eta_\varepsilon : \mathbb{R}^d \rightarrow \mathbb{R}$ to be defined by

$$\eta_\varepsilon(z) := \frac{1}{\varepsilon^d} \eta\left(\frac{z}{\varepsilon}\right).$$

In this way we impose that significant weight is given to edges connecting points up to distance ε . We consider the similarity matrix W^ε defined by

$$(1.5) \quad W_{i,j}^\varepsilon = \eta_\varepsilon(x_i - x_j).$$

We denote by $\mathcal{L}_{n,\varepsilon}$ the unnormalized graph Laplacian (1.1) of the weighted graph (V, W^ε) , that is

$$(1.6) \quad \mathcal{L}_{n,\varepsilon} = \mathcal{D}^\varepsilon - W^\varepsilon$$

where \mathcal{D}^ε is the diagonal matrix with $\mathcal{D}_{i,i}^\varepsilon = \sum_j W_{i,j}^\varepsilon$.

We define the *Dirichlet energy on the graph* of a function $u : V \rightarrow \mathbb{R}$ to be

$$(1.7) \quad \sum_{i,j} W_{i,j}^\varepsilon (u(x_i) - u(x_j))^2.$$

The fact that η is a symmetric function guarantees that W is symmetric and thus all the facts presented in Subsection 1.1 apply. In particular, (1.2) can be stated as: for every function $u : V \rightarrow \mathbb{R}$

$$(1.8) \quad \langle \mathcal{L}_{n,\varepsilon}u, u \rangle = \frac{1}{2} \sum_{i,j} W_{i,j}^\varepsilon (u(x_i) - u(x_j))^2,$$

where on the left hand side we have identified the function u with the vector $(u(x_1), \dots, u(x_n))$ in \mathbb{R}^n and where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in \mathbb{R}^n .

The symmetric normalized graph Laplacian $\mathcal{N}_{n,\varepsilon}^{sym}$ is given by

$$\mathcal{N}_{n,\varepsilon}^{sym} := \mathcal{D}^{-1/2} \mathcal{L}_{n,\varepsilon} \mathcal{D}^{-1/2}.$$

Since the kernel η is assumed radially symmetric, it can be defined as $\eta(x) := \boldsymbol{\eta}(|x|)$ for all $x \in \mathbb{R}^d$, where $\boldsymbol{\eta} : [0, \infty) \rightarrow [0, \infty)$ is the radial profile. We assume the following properties on $\boldsymbol{\eta}$:

- (K1) $\boldsymbol{\eta}(0) > 0$ and $\boldsymbol{\eta}$ is continuous at 0.
- (K2) $\boldsymbol{\eta}$ is non-increasing.
- (K3) The integral $\int_0^\infty \boldsymbol{\eta}(r) r^{d+1} dr$ is finite.

Remark 1.1. We remark that the last assumption on η is equivalent to imposing that the *surface tension*

$$(1.9) \quad \sigma_\eta := \int_{\mathbb{R}^d} \eta(h) |h_1|^2 dh$$

is finite, where h_1 represents the first component of h . The second condition implies that more relevance is given to the interactions between points that are close to each other. We notice that the class of acceptable kernels is quite broad and includes both Gaussian kernels and discontinuous kernels like one defined by a function η of the form $\eta = 1$ for $t \leq 1$ and $\eta = 0$ for $t > 1$.

We focus on point clouds that are obtained as independent samples from a given distribution ν . Specifically, consider an open, bounded, and connected set $D \subset \mathbb{R}^d$ with Lipschitz boundary (i.e. locally the graph of a Lipschitz function) and consider a probability measure ν supported on \overline{D} . We assume ν has a continuous density ρ , which is bounded above and below by positive constants on D . We assume that the points x_1, \dots, x_n (i.i.d. random points) are chosen according to the distribution ν . We consider the graph with nodes $V = \{x_1, \dots, x_n\}$ and edge weights $\{W_{i,j}^\varepsilon\}_{i,j}$ defined in (1.5). For an appropriate scaling of $\varepsilon := \varepsilon_n$ with respect to n , we study the limiting behavior of the eigenvalues and eigenvectors of the graph Laplacians as $n \rightarrow \infty$. We now describe the continuum problems which characterize the limit.

1.3. Description of spectral clustering in the continuum setting: the unnormalized case.

Let domain D , "ground-truth" measure ν with density ρ be as above. The object that characterizes the limit of the graph Laplacians $\mathcal{L}_{n,\varepsilon_n}$ as $n \rightarrow \infty$ is the differential operator:

$$(1.10) \quad \mathcal{L} : u \mapsto -\frac{1}{\rho} \operatorname{div}(\rho^2 \nabla u).$$

We consider the pairs $\lambda \in \mathbb{R}$ and $u \in H^1(D)$ (the Sobolev space of $L^2(D)$ functions with distributional derivative ∇u in $L^2(D, \mathbb{R}^d)$), with u not identically equal to zero, such that

$$(1.11) \quad \begin{aligned} \mathcal{L}u &= \lambda u, & \text{in } D, \\ \frac{\partial u}{\partial \mathbf{n}} &= 0, & \text{on } \partial D. \end{aligned}$$

A function u as above is said to be an eigenfunction of \mathcal{L} with corresponding eigenvalue $\lambda \in \mathbb{R}$. In Subsection 2.4 we discuss the precise definition of a solution of (1.11) and present some facts about it. In particular \mathcal{L} is a positive semidefinite self-adjoint operator with respect to the inner product $\langle \cdot, \cdot \rangle_{L^2(D,\nu)}$ and has a discrete spectrum that can be arranged as an increasing sequence converging to infinity

$$0 = \lambda_1 \leq \lambda_2 \leq \dots,$$

where each eigenvalue is repeated according to (finite) multiplicity. Furthermore, there exists an orthonormal basis of $L^2(D)$ (with respect to the inner product $\langle \cdot, \cdot \rangle_{L^2(D,\nu)}$) consisting of eigenfunctions u_i of \mathcal{L} .

Given a mapping $\Phi : D \rightarrow \mathbb{R}^k$ by $\Phi_\# \nu$ we denote the push forward of the measure ν , namely the measure for which $\Phi_\# \nu(A) = \nu(\Phi^{-1}(A))$, for any Borel set A . The continuum spectral clustering analogous to the discrete one of Algorithm 1 is as follows. Let $u_1, \dots, u_k : D \rightarrow \mathbb{R}$ be the orthonormal set of eigenfunctions corresponding to eigenvalues $\lambda_1, \dots, \lambda_k$. Consider the measure $\mu = (u_1, \dots, u_k)_\# \nu$. Let $\tilde{G}_i \subset \mathbb{R}^k$ be the clusters obtained by k-means clustering of μ . Then $G_i = (u_1, \dots, u_k)^{-1}(\tilde{G}_i)$ for $i = 1, \dots, k$ define the *spectral clustering* of ν .

1.4. Description of spectral clustering in the continuum setting: the normalized cases.

The object that characterizes the limit of the symmetric normalized graph Laplacians $\mathcal{N}_{n,\varepsilon_n}^{sym}$ as $n \rightarrow \infty$ is the differential operator

$$\mathcal{N}^{sym} : u \mapsto -\frac{1}{\rho^{3/2}} \operatorname{div} \left(\rho^2 \nabla \left(\frac{u}{\sqrt{\rho}} \right) \right).$$

We consider the space

$$(1.12) \quad H_{\sqrt{\rho}}^1(D) := \left\{ u \in L^2(D) : \frac{u}{\sqrt{\rho}} \in H^1(D) \right\}.$$

The spectrum of \mathcal{N}^{sym} is the set of pairs $\tau \in \mathbb{R}$ and $u \in H_{\sqrt{\rho}}^1(D)$, where u is not identically equal to zero, such that

$$(1.13) \quad \begin{aligned} \mathcal{N}^{sym}(u) &= \tau u, & \text{in } D \\ \frac{\partial(u/\sqrt{\rho})}{\partial \mathbf{n}} &= 0 & \text{on } \partial D. \end{aligned}$$

The sense in which (1.13) holds is made precise in Subsection 2.4. The spectrum of the operator \mathcal{N}^{sym} has similar properties to those of the spectrum of \mathcal{L} . We let

$$0 = \tau_1 \leq \tau_2 \leq \dots,$$

denote the eigenvalues of \mathcal{N}^{sym} , repeated according to multiplicity.

The continuum spectral clustering analogous to the discrete one of Algorithm 2 is as follows. Let $u_1, \dots, u_k : D \rightarrow \mathbb{R}$ be the orthonormal set of eigenfunctions (with respect to the inner product $\langle \cdot, \cdot \rangle_{L^2(D,\nu)}$) corresponding to eigenvalues τ_1, \dots, τ_k . Normalize them by

$$(\tilde{u}_1(x), \dots, \tilde{u}_k(x)) = \frac{(u_1(x), \dots, u_k(x))}{\|(u_1(x), \dots, u_k(x))\|} \text{ for all } x \in D.$$

Consider the measure $\tilde{\mu} = (\tilde{u}_1, \dots, \tilde{u}_k)_\# \nu$. Let $\tilde{G}_i \subset \mathbb{R}^k$ be the clusters obtained by k-means clustering of $\tilde{\mu}$. Then $G_i = (\tilde{u}_1, \dots, \tilde{u}_k)^{-1}(\tilde{G}_i)$ for $i = 1, \dots, k$ define the *spectral clustering* of ν .

Finally, the operator that describes the limit of the graph Laplacians $\mathcal{N}_{n,\varepsilon_n}^{rw} = \mathcal{D}^{\varepsilon_n^{-1}} \mathcal{L}_{n,\varepsilon_n}$ is described by the operator \mathcal{N}^{rw} :

$$\mathcal{N}^{rw}(u) = -\frac{1}{\rho^2} \operatorname{div}(\rho^2 \nabla u).$$

As discussed in Subsection 2.4, the eigenvalues of \mathcal{N}^{rw} are equal to the eigenvalues of \mathcal{N}^{sym} . The continuum clustering, which is analogous to the discrete one of Algorithm 3, is as in Subsection 1.3, where eigenfunctions of \mathcal{N}^{rw} are used.

1.5. Passage from discrete to continuum. We are interested in showing that as $n \rightarrow \infty$ eigenvalues of discrete graph Laplacians and the associated eigenvectors converge towards eigenvalues and eigenfunctions of corresponding differential operators. The issue that arises is how to compare functions on discrete and continuum setting. Typically this is achieved by introducing an interpolation operator that takes discretely defined functions to continuum ones and a restriction operator which restricts the continuum function to the discrete setting. For this setting to work some smoothness of functions considered is required. Furthermore the choice of the interpolation operator and its properties adds an intermediate step that needs to be understood.

We choose a different route and introduce a way to compare the functions between settings directly. This approach is quite general and does not require any regularity assumptions. We use

the TL^p -topologies introduced in [15] and in particular in this paper we focus in the TL^2 -topology that we now recall. Denote by ν_n the empirical measure associated to the n data points, that is

$$(1.14) \quad \nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}.$$

For a given function $u \in L^2(D, \nu)$, the question is how to compare u with a function $v \in L^2(D, \nu_n)$ (a function defined on the set V). More generally, one can consider the problem of how to compare functions in $L^2(D, \mu)$ with those in $L^2(D, \theta)$ for arbitrary probability measures μ, θ on D . We define the set of objects that includes both the functions in discrete setting and those in continuum setting as follows:

$$TL^2(D) := \{(\mu, f) : \mu \in \mathcal{P}(D), f \in L^2(D, \mu)\},$$

where $\mathcal{P}(D)$ denotes the set of Borel probability measures on D . For (μ, f) and (θ, g) in TL^2 we define the distance

$$d_{TL^2}((\mu, f), (\theta, g)) = \inf_{\pi \in \Gamma(\mu, \theta)} \left(\iint_{D \times D} |x - y|^2 + |f(x) - g(y)|^2 d\pi(x, y) \right)^{\frac{1}{2}},$$

where $\Gamma(\mu, \theta)$ is the set of all *couplings* (or *transportation plans*) between μ and θ , that is, the set of all Borel probability measures on $D \times D$ for which the marginal on the first variable is μ and the marginal on the second variable is θ . It was proved in [15] that d_{TL^2} is indeed a metric on TL^2 . As remarked in [15], one of the nice features of the convergence in TL^2 is that it simultaneously generalizes the weak convergence of probability measures and the convergence in L^2 of functions. It also provides us with a way to compare functions which are supported in sets as different as point clouds and continuous domains. In Subsection 2.1 we present more details about this metric.

For a given $\mu \in \mathcal{P}(D)$ we denote by $L^2(\mu)$ the space of L^2 -functions with respect to the measure μ . Also, for $f, g \in L^2(\mu)$ we write

$$\langle f, g \rangle_\mu := \int_D f g d\mu \quad \text{and} \quad \|f\|_\mu^2 = \langle f, f \rangle_\mu.$$

Finally, if the measure μ has a density ρ , that is, if $d\mu = \rho dx$, we may write $\langle f, g \rangle_\rho$ and $\|f\|_\rho$ instead of $\langle f, g \rangle_\mu$ and $\|f\|_\mu$.

1.6. Convergence of eigenvalues, eigenvectors, and of spectral clustering: the unnormalized case. Here we present one of the main results of this paper. We state the conditions on ε_n for the spectrum of the unnormalized graph Laplacian $\mathcal{L}_{n, \varepsilon_n}$, given in (1.6), to converge to the spectrum of \mathcal{L} , given by (1.10) and for the spectral clustering of Algorithm 1 to converge to the clustering of Subsection 1.3. Let $\lambda_1 \leq \lambda_2 \leq \dots$ be the eigenvalues of \mathcal{L} and u_1, u_2, \dots the corresponding orthonormal eigenfunctions, as in Subsection 1.3. We recall that orthogonality is considered with respect to the inner product in $L^2(\nu)$.

To state the results it is convenient to introduce $0 = \bar{\lambda}_1 < \bar{\lambda}_2 < \dots$, the sequence of distinct eigenvalues of \mathcal{L} . For a given $k \in \mathbb{N}$, we denote by $s(k)$ the multiplicity of the eigenvalue $\bar{\lambda}_k$ and we let $\hat{k} \in \mathbb{N}$ be such that $\bar{\lambda}_k = \lambda_{\hat{k}+1} = \dots = \lambda_{\hat{k}+s(k)}$. Also, we denote by $\text{Proj}_k : L^2(\nu) \rightarrow L^2(\nu)$ the projection (with respect to the inner product $\langle \cdot, \cdot \rangle_\nu$) onto the eigenspace of \mathcal{L} associated to the eigenvalue $\bar{\lambda}_k$. For all large enough n , we denote by $\text{Proj}_k^{(n)} : L^2(\nu_n) \rightarrow L^2(\nu_n)$ the projection (with respect to the inner product $\langle \cdot, \cdot \rangle_{\nu_n}$) onto the space generated by all the eigenvectors of $\mathcal{L}_{n, \varepsilon_n}$ associated to the eigenvalues $\lambda_{\hat{k}+1}^{(n)}, \dots, \lambda_{\hat{k}+s(k)}^{(n)}$. Here, as in the rest of the paper, we identify \mathbb{R}^n with the space $L^2(D, \nu_n)$.

Theorem 1.2 (Convergence of the spectra of the unnormalized graph Laplacians). *Let $d \geq 2$ and let $D \subseteq \mathbb{R}^d$, be an open, bounded, connected set with Lipschitz boundary. Let ν be a probability measure on D with continuous density ρ , satisfying*

$$(1.15) \quad (\forall x \in D) \quad m \leq \rho(x) \leq M,$$

for some $0 < m \leq M$. Let x_1, \dots, x_n, \dots be a sequence of i.i.d. random points chosen according to ν . Let $\{\varepsilon_n\}_{n \in \mathbb{N}}$ be a sequence of positive numbers converging to 0 and satisfying

$$(1.16) \quad \lim_{n \rightarrow \infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} = 0 \text{ if } d = 2,$$

$$\lim_{n \rightarrow \infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} = 0 \text{ if } d \geq 3.$$

Assume the kernel η satisfies conditions (K1)-(K3). Then, with probability one, all of the following statements hold true:

1. *Convergence of Eigenvalues: For every $k \in \mathbb{N}$*

$$(1.17) \quad \lim_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} = \sigma_\eta \lambda_k,$$

where σ_η is defined in (1.9).

2. *For every $k \in \mathbb{N}$, every sequence $\{u_k^n\}_{n \in \mathbb{N}}$ with u_k^n an eigenvector of $\mathcal{L}_{n, \varepsilon_n}$ associated to the eigenvalue $\lambda_k^{(n)}$ and with $\|u_k^n\|_{\nu_n} = 1$ is pre-compact in TL^2 . Additionally, whenever $u_k^n \xrightarrow{TL^2} u_k$ along a subsequence as $n \rightarrow \infty$, then $\|u_k\|_\nu = 1$ and u_k is an eigenfunction of \mathcal{L} associated to λ_k .*
3. *Convergence of Eigenprojections: For all $k \in \mathbb{N}$ and for arbitrary sequence $v_n \in L^2(\nu_n)$, if $v_n \xrightarrow{TL^2} v$ as $n \rightarrow \infty$ along some subsequence. Then along that subsequence*

$$\text{Proj}_k^{(n)}(v_n) \xrightarrow{TL^2} \text{Proj}_k(v), \text{ as } n \rightarrow \infty.$$

4. *Consistency of Spectral Clustering. Let G_1^n, \dots, G_k^n be the clusters obtained in Algorithm 1. Let $\nu_i^n = \nu_n \llcorner_{G_i^n}$ (the restriction of ν_n to G_i^n) for $i = 1, \dots, k$. Then $(\nu_1^n, \dots, \nu_k^n)$ is precompact with respect to weak convergence of measures and furthermore if $(\nu_1^n, \dots, \nu_k^n)$ converges along a subsequence to (ν_1, \dots, ν_k) then $(\nu_1, \dots, \nu_k) = (\nu_{\llcorner G_1}, \dots, \nu_{\llcorner G_k})$ where G_1, \dots, G_k is a spectral clustering of ν , described in Subsection 1.3.*

Remark 1.3. We remark that although the choice of the TL^2 -topology used in the previous theorem may seem unusual at first sight, it actually reduces to a more common notion of convergence (like the one used in [48] which we described below) in the presence of regularity assumptions on the density ρ and the domain D . In fact, assume for simplicity that D has smooth boundary and that ρ is a smooth function. Consider $\{u_k^n\}_{n \in \mathbb{N}}$ where u_k^n is an eigenvector of $\mathcal{L}_{n, \varepsilon_n}$ associated to the eigenvalue $\lambda_k^{(n)}$ and satisfying $\|u_k^n\|_{\nu_n} = 1$. The second statement in Theorem 1.2 says that up to subsequence, $u_k^n \xrightarrow{TL^2} u_k$, where u_k is an eigenfunction of \mathcal{L} associated to λ_k . From the regularity theory of elliptic PDEs it follows that u_k is smooth up to the boundary. In particular, it makes sense to define a function \tilde{u}_k^n on the point cloud, by simply taking the restriction of u_k to the points $\{x_1, \dots, x_n\}$. It is straightforward to check that $\tilde{u}_k^n \xrightarrow{TL^2} u_k$ due to the smoothness of u_k . In turn, $u_k^n \xrightarrow{TL^2} u_k$, implies that $d_{TL^2}((\nu_n, \tilde{u}_k^n - u_k^n), (\nu, 0)) \rightarrow 0$. From this and Proposition 2.1, we conclude that

$$\|u_k^n - \tilde{u}_k^n\|_{\nu_n} \rightarrow 0.$$

This is precisely the mode of convergence used in [48].

The proof of Theorem 1.2 relies on the study of the limiting behavior of the following rescaled form of the Dirichlet energy (1.7) on the graph:

$$(1.18) \quad G_{n,\varepsilon}(u) := \frac{1}{\varepsilon^2 \eta^2} \sum_{i,j} W_{i,j} (u(x_i) - u(x_j))^2.$$

The type of limit which is relevant for the problem, is the one given by variational convergence known as the Γ -convergence. The notion of Γ -convergence is recalled in Subsection 2.2. This notion of convergence is particularly suitable in order to study the convergence of minimizers of objective functionals on graphs as $n \rightarrow \infty$, as it is discussed in [17].

The relevant continuum energy is the *weighted Dirichlet* energy $G : L^2(D) \rightarrow [0, \infty]$:

$$(1.19) \quad G(u) := \begin{cases} \int_D |\nabla u(x)|^2 \rho^2(x) dx & \text{if } u \in H^1(D) \\ \infty & \text{if } u \in L^2(D) \setminus H^1(D). \end{cases}$$

Theorem 1.4 (Γ -convergence of Dirichlet energies). *Consider the same setting as in Theorem 1.2 and the same assumptions on η and on $\{\varepsilon_n\}_{n \in \mathbb{N}}$. Then, G_{n,ε_n} , defined by (1.18), Γ -converge to $\sigma_\eta G$ as $n \rightarrow \infty$ in the TL^2 sense, where σ_η is given by (1.9) and G is the weighted Dirichlet energy with weight ρ^2 defined in (1.19). Moreover, the sequence of functionals $\{G_{n,\varepsilon_n}\}_{n \in \mathbb{N}}$ satisfies the compactness property with respect to the TL^2 -metric. That is, every sequence $\{u_n\}_{n \in \mathbb{N}}$ with $u_n \in L^2(\nu_n)$ for which*

$$\sup_{n \in \mathbb{N}} \|u_n\|_{\nu_n} < \infty, \quad \sup_{n \in \mathbb{N}} G_{n,\varepsilon_n}(u_n) < \infty,$$

is precompact in TL^2 .

The fact that the weight in the limiting functional G is ρ^2 (and not ρ) essentially follows from the fact that the graph Dirichlet energy defined in (1.18) is a double sum. This is the same weight that shows up in the study of the continuum limit of the *graph total variation* in [15]. Theorem 1.4 is analogous to Theorems 1.1 and 1.2 in [15] combined.

1.7. Convergence of eigenvalues, eigenvectors, and of spectral clustering: the normalized case. We also study the limit of the spectra of $\mathcal{N}_{n,\varepsilon_n}^{sym}$, the symmetric normalized graph Laplacian which we recall is given by,

$$\mathcal{N}_{n,\varepsilon_n}^{sym} := \mathcal{D}^{-1/2} \mathcal{L}_{n,\varepsilon_n} \mathcal{D}^{-1/2}.$$

For a function $u : V \rightarrow \mathbb{R}$, (1.3) can be written as

$$(1.20) \quad \langle \mathcal{N}_{n,\varepsilon_n}^{sym} u, u \rangle = \frac{1}{2} \sum_{i,j} W_{i,j} \left(\frac{u(x_i)}{\sqrt{\mathcal{D}_{ii}}} - \frac{u(x_j)}{\sqrt{\mathcal{D}_{jj}}} \right)^2.$$

We denote by

$$0 = \tau_1^{(n)} \leq \dots \leq \tau_n^{(n)}$$

the eigenvalues of $\mathcal{N}_{n,\varepsilon_n}^{sym}$ repeated according to multiplicity. Their limit is described by differential operator

$$\mathcal{N}^{sym} : u \mapsto -\frac{1}{\rho^{3/2}} \operatorname{div} \left(\rho^2 \nabla \left(\frac{u}{\sqrt{\rho}} \right) \right).$$

Let

$$0 = \tau_1 \leq \tau_2 \leq \dots,$$

denote the eigenvalues of \mathcal{N}^{sym} , repeated according to multiplicity. We write $0 = \bar{\tau}_1 < \bar{\tau}_2 < \dots$, to denote the distinct eigenvalues of \mathcal{L} . For a given $k \in \mathbb{N}$, we denote by $s(k)$ the multiplicity of the eigenvalue $\bar{\tau}_k$ and we let $\hat{k} \in \mathbb{N}$ be such that $\bar{\tau}_k = \tau_{\hat{k}+1} = \dots = \tau_{\hat{k}+s(k)}$. We define Proj_k and $\operatorname{Proj}_k^{(n)}$

analogously to the way we defined them in the paragraph preceding Theorem 1.2. The following is analogous to Theorem 1.2.

Theorem 1.5 (Convergence of the spectra of the normalized graph Laplacians). *Consider the same setting as in Theorem 1.2 and the same assumptions on η and on $\{\varepsilon_n\}_{n \in \mathbb{N}}$. Then, with probability one, all of the following statements hold*

1. *Convergence of Eigenvalues: For every $k \in \mathbb{N}$*

$$\lim_{n \rightarrow \infty} \frac{2\tau_k^{(n)}}{\varepsilon_n^2} = \frac{\sigma_\eta}{\beta_\eta} \tau_k,$$

where β_η is given by

$$(1.21) \quad \beta_\eta := \int_{\mathbb{R}^d} \eta(h) dh,$$

and where σ_η is given by (1.9).

2. *For every $k \in \mathbb{N}$, every sequence $\{u_k^n\}_{n \in \mathbb{N}}$ with u_k^n being an eigenvector of $\mathcal{N}_{n, \varepsilon_n}^{sym}$ associated to the eigenvalue $\tau_k^{(n)}$ and with $\|u_k^n\|_{\nu_n} = 1$ is pre-compact in TL^2 . Additionally, whenever $u_k^n \xrightarrow{TL^2} u_k$ along a subsequence as $n \rightarrow \infty$, then $\|u_k\|_\nu = 1$ and u_k is an eigenfunction of \mathcal{N} associated to τ_k .*
3. *Convergence of Eigenprojections: For all $k \in \mathbb{N}$ and for arbitrary $v_n \in L^2(\nu_n)$, if $v_n \xrightarrow{TL^2} v$ along a subsequence as $n \rightarrow \infty$ then,*

$$\text{Proj}_k^{(n)}(u_n) \xrightarrow{TL^2} \text{Proj}_k(u), \quad \text{as } n \rightarrow \infty, \text{ along the subsequence.}$$

4. *Consistency of Spectral Clustering. Assume $\rho \in C^1(D)$. Let G_1^n, \dots, G_k^n be the clusters obtained in Algorithm 2. Let $\nu_i^n = \nu_n \llcorner G_i^n$ for $i = 1, \dots, k$. Then $(\nu_1^n, \dots, \nu_k^n)$ is precompact with respect to weak convergence of measures and furthermore if $(\nu_1^n, \dots, \nu_k^n)$ converges along a subsequence to (ν_1, \dots, ν_k) then $(\nu_1, \dots, \nu_k) = (\nu \llcorner G_1, \dots, \nu \llcorner G_k)$ where G_1, \dots, G_k is a spectral clustering of ν , described in Subsection 1.4.*

The proof of the previous theorem is completely analogous to the one of Theorem 1.2 once one has proved the variational convergence of the relevant energies. Indeed, consider $\overline{G}_{n, \varepsilon} : L^2(\nu_n) \rightarrow \mathbb{R}$ defined by

$$(1.22) \quad \overline{G}_{n, \varepsilon}(u) := \frac{1}{n\varepsilon^2} \sum_{i, j} W_{i, j} \left(\frac{u(x_i)}{\sqrt{\mathcal{D}_{ii}}} - \frac{u(x_j)}{\sqrt{\mathcal{D}_{jj}}} \right)^2$$

and $\overline{G} : L^2(D) \rightarrow [0, \infty]$ by

$$(1.23) \quad \overline{G}(u) := \begin{cases} \int_D \left| \nabla \left(\frac{u}{\sqrt{\rho}} \right) \right|^2 \rho^2(x) dx & \text{if } u \in H_{\sqrt{\rho}}^1(D), \\ \infty & \text{if } u \in L^2(D) \setminus H_{\sqrt{\rho}}^1(D), \end{cases}$$

where $H_{\sqrt{\rho}}^1(D)$ is defined in (1.12). The following holds.

Theorem 1.6 (Γ -convergence of normalized Dirichlet energies). *With the same setting as in Theorem 1.2 and the same assumptions on η and on $\{\varepsilon_n\}_{n \in \mathbb{N}}$, $\overline{G}_{n, \varepsilon_n}$, defined by (1.22), Γ -converge to $\frac{\sigma_\eta}{\beta_\eta} \overline{G}$ as $n \rightarrow \infty$ in the TL^2 -sense, where \overline{G} is defined in (1.23), σ_η and β_η are given by (1.9) and (1.21) respectively. Moreover, the sequence of functionals $\{\overline{G}_{n, \varepsilon_n}\}_{n \in \mathbb{N}}$ satisfies the compactness*

property with respect to the TL^2 -metric. That is, every sequence $\{u_n\}_{n \in \mathbb{N}}$ with $u_n \in L^2(\nu_n)$ for which

$$\sup_{n \in \mathbb{N}} \|u_n\|_{\nu_n} < \infty, \quad \sup_{n \in \mathbb{N}} \overline{G}_{n, \varepsilon_n}(u_n) < \infty,$$

is precompact in TL^2 .

Finally, we consider the limit of the spectrum of $\mathcal{N}_{n, \varepsilon_n}^{rw}$, where $\mathcal{N}_{n, \varepsilon_n}^{rw} = \mathcal{D}^{-1} \mathcal{L}_{n, \varepsilon_n}$. Consider the operator \mathcal{N}^{rw} given by

$$\mathcal{N}^{rw}(u) = -\frac{1}{\rho^2} \operatorname{div}(\rho^2 \nabla u).$$

As discussed in Subsection 2.4, the eigenvalues of \mathcal{N}^{rw} are equal to the eigenvalues of \mathcal{N}^{sym} . Thus from (1.4) and from Theorem 1.5, it follows that after appropriate rescaling, the eigenvalues of $\mathcal{N}_{n, \varepsilon_n}^{rw}$ converge to the eigenvalues of \mathcal{N}^{rw} . Moreover, using again (1.4) and Theorem 1.5, we have the following convergence of eigenvectors.

Corollary 1.7. *Consider the same setting as in Theorem 1.2 and the same assumptions on η and on $\{\varepsilon_n\}_{n \in \mathbb{N}}$. Then, with probability one, the following statement holds: For every $k \in \mathbb{N}$, every sequence $\{u_k^n\}_{n \in \mathbb{N}}$ with u_k^n being an eigenvector of $\mathcal{N}_{n, \varepsilon_n}^{rw}$ associated to the eigenvalue $\tau_k^{(n)}$ and with $\|u_k^n\|_{\nu_n} = 1$ is pre-compact in TL^2 . Additionally, all its cluster points are eigenfunctions of \mathcal{N}^{rw} with eigenvalue τ_k . Finally the clusters obtained by Algorithm 3 converge to clusters obtained by spectral clustering corresponding to \mathcal{N}^{rw} described at the end of Subsection 1.3.*

1.8. Stability of k -means clustering. One of the final elements of the proof of the consistency results of spectral clustering (statement 4. in Theorems 1.2 and 1.5) requires new results on stability of k -means clustering with respect to perturbations of the measure being clustered. These results extend the result of Pollard [31] who proved the consistency of k -means clustering. It is important to extend such results because in our setting, at the discrete level, the point set used as input for the k -means algorithm is not a sample from a given distribution and thus one can not apply the results in [31] directly.

Given $k \in \mathbb{N}$ and given a measure μ on \mathbb{R}^N with finite second moments, let $F_{\mu, k} : \mathbb{R}^{N \times k} \rightarrow [0, \infty)$ be defined by

$$(1.24) \quad F_{\mu, k}(z_1, \dots, z_k) := \int d(x, \{z_1, \dots, z_k\})^2 d\mu(x)$$

where $z_i \in \mathbb{R}^N$ for $i = 1, \dots, k$. For brevity we write \mathbf{z} both for (z_1, \dots, z_k) and $\{z_1, \dots, z_k\}$ where the object considered should be clear from the context. The problem of k -means clustering is to minimize $F_{\mu, k}$ over $\mathbb{R}^{N \times k}$. In Subsection 2.3 we show the existence of minimizers of the functional (1.24). The main result is the following.

Theorem 1.8 (Stability of k -means clustering). *Let $k \geq 1$. Let μ be a Borel probability measure on \mathbb{R}^N with finite second moments and whose support has at least k points. Assume $\{\mu_m\}_{m \in \mathbb{N}}$ is a sequence of probability measures on \mathbb{R}^N with finite second moments which converges in the Wasserstein distance (see (2.1)) to μ . Then,*

$$\lim_{m \rightarrow \infty} \min_{\mathbf{z}} F_{\mu_m, k}(\mathbf{z}) = \min_{\mathbf{z}} F_{\mu, k}(\mathbf{z}).$$

Moreover, if \mathbf{z}_m is a minimizer of $F_{\mu_m, k}$ for all m , then the set $\{\mathbf{z}_m, m \in \mathbb{N}\}$ is precompact in $\mathbb{R}^{N \times k}$ and all of its accumulation points are minimizers of $F_{\mu, k}$.

We present the proof of the previous Theorem in Subsection 2.3.

The clusters corresponding to \mathbf{z} minimizing the $F_{\mu, k}$ are the Voronoi cells: $G_i = \{x \in \mathbb{R}^N : d(x, \mathbf{z}) = d(x, z_i)\}$. We prove in Lemma 2.10 that the measure of the boundaries of clusters is zero, that is we show that if $i \neq j$ then $\mu(G_i \cap G_j) = 0$. In other words it is irrelevant to which cluster are

the points on the boundary assigned to and because of this, we are allowed to define the clusters to be either open or closed sets. We furthermore note that the associated measures, $\mu_{\setminus G_i}$ are mutually orthogonal and satisfy $\sum_i \mu_{\setminus G_i} = \mu$.

A consequence of Theorem 1.8 is that as the cluster centers converge so do the measures representing the clusters.

Corollary 1.9. *Under the assumptions of Theorem 1.8, if \mathbf{z}^m converge along a subsequence to \mathbf{z} then the measures $\mu_{m,\setminus G_i^m}$ converge weakly in the sense of measures to $\mu_{\setminus G_i}$ as $m \rightarrow \infty$ for all $i = 1, \dots, k$.*

The corollary follows from Theorem 1.8, since the convergence of centers of Voronoi cells, along with the fact that the boundaries of cells change continuously with respect to cell centers implies that the measures converge in Levy-Prokhorov metric, which characterizes the weak convergence of measures.

1.9. Discussion. Theorems 1.2, 1.5, and 1.7 establish the consistency of spectral clustering. An important difference between our work and the available consistency results is that we provide an explicit range of rates at which ε_n (the length scale used to construct the graph) is allowed to converge to 0 as $n \rightarrow \infty$. In [48] the parameter ε is not allowed to depend on n . As a result, the functional obtained in the limit is a non-local (i.e. integral, rather than differential) operator. Operators with very different spectral properties are obtained in the limit depending on whether one uses a normalized or unnormalized graph Laplacian. In particular, it is argued that normalized spectral clustering is more advantageous than the unnormalized clustering, because in the normalized case the spectrum of the limiting operator is better behaved and the spectral consistency in the unnormalized case is only guaranteed in restrictive settings. We remark that our results show that when the parameter ε_n decays to zero such difference between the normalized and the unnormalized settings disappears and the limiting operators in both cases have a discrete spectrum.

When constructing the graph it is advantageous, from the point of view of computational complexity, to have fewer edges (that is to take ε small). However below some threshold the graph thus constructed does not contain enough information to accurately recover the geometry of the underlying ground-truth distribution. How large ε should be taken depends on n , the number of data points available. As number of data points increases ε converges to zero. We remark that for $d \geq 3$, the results of Theorems 1.2, 1.5, and 1.7 are (almost) optimal in the sense of scaling. Namely, we show that if the kernel η used to construct the graph is compactly supported, then convergence holds if $\varepsilon_n \gg \frac{\log(n)^{1/d}}{n^{1/d}}$, while if $\varepsilon_n \ll \frac{\log(n)^{1/d}}{n^{1/d}}$ the convergence does not hold. This follows from the results on the connectivity of random geometric graphs in [20, 19, 30] which show that with high probability for large n the graph thus obtained is disconnected.

Finally, we remark that our results are essentially independent of the kernel used to construct the weights. For example, when the points are sampled from the uniform distribution on a domain D , our results show that the spectra of the graph Laplacians converge to the spectrum of the Laplacian on the domain D , regardless of the kernel used.

1.10. Outline of the approach. Theorem 1.2 is based on the variational convergence of the energies G_{n,ε_n} towards $\sigma_\eta G$, together with the corresponding compactness result (Theorem 1.4). In order to show Theorem 1.4, we first introduce the functional $G_{\varepsilon_n} : L^2(D, \rho) \rightarrow [0, \infty)$ given by,

$$(1.25) \quad G_{\varepsilon_n}(u) := \frac{1}{\varepsilon_n^2} \int_D \int_D \eta_{\varepsilon_n}(x-y) |u(x) - u(y)|^2 \rho(x) \rho(y) dx dy,$$

which serves as an intermediate object between the functionals G_{n,ε_n} and G . It is important to observe that the argument of G_{n,ε_n} is a function u_n supported on the data points, whereas the argument of G_{ε_n} is a $L^2(D, \rho)$ function; in particular a function defined on D . The functional G_{ε_n} is a *non-local* functional, where the term *non-local* refers to the fact that differences of a given function

on a ε_n -neighborhood are averaged, which contrasts the local approach of averaging derivatives of the given function. Non-local functionals have been of interest in the last decades due to their wide range of applications which includes phase transitions, image processing and PDEs. From a statistical point of view, for a fixed function $u : D \rightarrow \mathbb{R}$, $G_{\varepsilon_n}(u)$ is nothing but the expectation of $G_{n,\varepsilon_n}(u)$. On the other hand, the functional G_{ε_n} is relevant for our purposes because not only it approximates G defined in (1.19) in a pointwise sense, but it also approximates it in a variational sense (as the parameter ε_n goes to zero). More precisely the following holds.

Proposition 1.10. *Consider an open, bounded domain D in \mathbb{R}^d with Lipschitz boundary. Let $\rho : D \rightarrow \mathbb{R}$ be continuous and bounded below and above by positive constants. Let $\{\varepsilon_k\}_{k \in \mathbb{N}}$ be a sequence of positive numbers converging to zero. Then, $\{G_{\varepsilon_k}\}_{k \in \mathbb{N}}$ (defined in (1.25)) Γ -converges with respect to the $L^2(D, \rho)$ -metric to $\sigma_\eta G$, where σ_η is defined in (1.9) and G is defined in (1.19). Moreover, the functionals $\{G_{\varepsilon_k}\}_{k \in \mathbb{N}}$ satisfy the compactness property, with respect to the $L^2(D, \rho)$ -metric. That is, every sequence $\{u_k\}_{k \in \mathbb{N}}$ with $u_k \in L^2(D, \rho)$ for which*

$$\sup_{k \in \mathbb{N}} \|u_k\|_{L^2(D, \rho)} < \infty, \quad \sup_{k \in \mathbb{N}} G_{\varepsilon_k}(u_k) < \infty,$$

is precompact in $L^2(D, \rho)$. Finally, for every $u \in L^2(D, \rho)$

$$(1.26) \quad \lim_{n \rightarrow \infty} G_{\varepsilon_n}(u) = \sigma_\eta G(u).$$

Proof. When ρ is constant, the proof may be found in the Appendix of [1] in case D is a convex set, and in [32] for a general domain D satisfying the assumptions in the statement. In case ρ is not constant the results are obtained in a straightforward way by adapting the arguments presented in [32] just as it is done in Section 4 in [15] when studying the variational limit of the non-local functional

$$TV_\varepsilon(u) := \frac{1}{\varepsilon_n} \int_D \int_D \eta_\varepsilon(x-y) |u(x) - u(y)| \rho(x) \rho(y) dx dy,$$

which is the L^1 analogue of G_ε . □

As observed earlier, the argument of G_{n,ε_n} is a function u_n supported on the data points, while the argument of G_{ε_n} is an $L^2(D)$ function. For a function u_n defined on the set $V = \{x_1, \dots, x_n\}$, the idea is to associate an $L^2(D)$ function \tilde{u}_n which approximates u_n in the TL^2 -sense and is such that $G_{\varepsilon_n}(\tilde{u}_n)$ is comparable to $G_{n,\varepsilon_n}(u_n)$. The purpose of doing this is to use Proposition 1.10. We construct the approximating function \tilde{u}_n by using transportation maps (i.e. measure preserving maps) between the measure ν and ν_n . More precisely, we set $\tilde{u}_n = u_n \circ T_n$ where T_n is a transportation map between ν and ν_n which moves mass as little as possible. The estimates on how far the mass needs to be moved were known in the literature when ρ is constant and when the domain D is the unit cube $(0, 1)^d$ (see [26, 41, 42, 43] for $d = 2$ and [36] for $d \geq 3$). In [16] these estimates are extended to general domains D and densities ρ satisfying (1.15). Indeed, the following is proved.

Proposition 1.11. *Let $D \subseteq \mathbb{R}^d$ be a bounded, connected, open set with Lipschitz boundary. Let ν be a probability measure on D with density $\rho : D \rightarrow (0, \infty)$ satisfying (1.15). Let x_1, \dots, x_n, \dots be i.i.d. samples from ν . Let ν_n be the empirical measure associated to the n data points. Then, for any fixed $\alpha > 2$, except on a set with probability $O(n^{-\alpha/2})$, there exists a transportation map $T_n : D \rightarrow D$ between the measure ν and the measure ν_n (denoted $T_n\# \nu = \nu_n$) such that*

$$\|T_n - Id\|_\infty \leq C \begin{cases} \frac{\ln(n)^{3/4}}{n^{1/2}}, & \text{if } d = 2, \\ \frac{\ln(n)^{1/d}}{n^{1/d}}, & \text{if } d \geq 3, \end{cases}$$

where C depends only on α , D , and the constants m, M .

From the previous result, Chebyshev's inequality and Borel-Cantelli lemma one obtains the following rate of convergence of the ∞ -transportation distance between the empirical measures ν_n and the measure ν (see [16] for details associated to Proposition 1.11).

Proposition 1.12. *Let D be an open, connected and bounded subset of \mathbb{R}^d which has Lipschitz boundary. Let ν be a probability measure on D with density ρ satisfying (1.15). Let x_1, \dots, x_n, \dots be a sequence of independent samples from ν and let ν_n be the associated empirical measures (1.14). Then, there is a constant $C > 0$ such that with probability one, there exists a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ from ν to ν_n ($T_n\# \nu = \nu_n$) and such that:*

$$(1.27) \quad \text{if } d = 2 \text{ then} \quad \limsup_{n \rightarrow \infty} \frac{n^{1/2} \|Id - T_n\|_\infty}{(\log n)^{3/4}} \leq C$$

$$(1.28) \quad \text{and if } d \geq 3 \text{ then} \quad \limsup_{n \rightarrow \infty} \frac{n^{1/d} \|Id - T_n\|_\infty}{(\log n)^{1/d}} \leq C.$$

As shown in Section 3, Proposition 1.10 and Proposition 1.12 are at the backbone of Theorem 1.4. Schematically,

$$G_\varepsilon \xrightarrow{\Gamma} \sigma_\eta G \quad \text{in } L^2 \quad + \quad \text{Proposition 1.12} \implies G_{n, \varepsilon_n} \xrightarrow{\Gamma} \sigma_\eta G \quad \text{in } TL^2.$$

We note that the statement $G_\varepsilon \xrightarrow{\Gamma} \sigma_\eta G$ is a purely analytic, purely deterministic fact. Proposition 1.12, on the other hand contains all the probabilistic estimates needed to establish all the results on this paper. Such estimates in particular provide the constraints on the parameter ε_n in Theorem 1.4. It is worth observing that Proposition 1.12 is a statement that only involves the underlying measure ν and the empirical measure ν_n , and that in particular it does not involve estimates on the difference between the functional $G_{\varepsilon_n}(u)$ and the functional $G_{n, \varepsilon_n}(u)$ for u belonging to a small (in the sense of VC-dimension) class of functions. In other words our estimates are related to the domains where the functions are defined (discrete/continuous) and not to the actual values of functions defined on those domains.

With Theorem 1.4 at hand, the proof of Theorem 1.2 now relies on some spectral properties of the operator \mathcal{L} and analogous properties of $\mathcal{L}_{n, \varepsilon_n}$. As shown in Section 2.4, the space $L^2(D, \rho)$ has a countable orthonormal basis (with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$) formed with eigenfunctions of \mathcal{L} . Additionally, the different eigenvalues of \mathcal{L} can be organized as an increasing sequence of positive numbers converging to infinity. Each of the eigenvalues has finite multiplicity. Moreover, the eigenvalues of \mathcal{L} have a variational characterization, as they can be written as the minimum value of optimization problems over successive subspaces of $L^2(D, \rho)$. This is the content of the Courant-Fisher mini-max principle which states that for every k

$$(1.29) \quad \lambda_k = \sup_{S \in \Sigma_{k-1}} \min_{\|u\|_\rho=1, u \in S^\perp} G(u),$$

where we recall $0 = \lambda_1 \leq \lambda_2 \leq \dots$, denote the eigenvalues of \mathcal{L} repeated according to multiplicity, Σ_{k-1} denotes the set of $(k-1)$ -dimensional subspaces of $L^2(D, \rho)$, and where S^\perp represents the orthogonal complement of S with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$. Moreover, the supremum in (1.29) is attained by the span of the first $(k-1)$ eigenfunctions of \mathcal{L} . In Subsection 2.4 we review the previously mentioned spectral properties of \mathcal{L} . Likewise, we can write the eigenvalues of $\mathcal{L}_{n, \varepsilon_n}$ as

$$(1.30) \quad \lambda_k^{(n)} = \frac{n\varepsilon_n^2}{2} \sup_{S \in \Sigma_{k-1}^{(n)}} \min_{\|u\|_{\nu_n}=1, u \in S^\perp} G_{n, \varepsilon_n}(u),$$

where $\Sigma_{k-1}^{(n)}$ denotes the set of $(k-1)$ -dimensional subspaces of \mathbb{R}^n , and where S^\perp represents the orthogonal complement of S with respect to the inner product $\langle \cdot, \cdot \rangle_{\nu_n}$ in \mathbb{R}^n . Moreover, as in the

continuum setting, the supremum in (1.30) is attained by the span of the first $(k - 1)$ eigenvectors of $\mathcal{L}_{n, \varepsilon_n}$. Theorem 1.4 allows us to exploit expressions (1.30) and (1.29) and in fact in Section 3 we show how 1.30 and 1.29 together with Theorem 1.4 imply Theorem 1.2, thus establishing the spectral convergence in the unnormalized case.

In the normalized case, the same approach used in the unnormalized case can be taken. In fact, the proof of Theorem 1.5 follows from the proof of Theorem 1.2 by *mutatis mutandis* after Theorem 1.6 has been proved.

The paper is organized as follows. Section 2 contains the notation and the background we need in the rest of the paper. In particular in Subsection 2.1 we review some facts about the TL^2 space, in Subsection 2.2 we review the definition of Γ -convergence, in Subsection 2.3 we present some results on stability of k -means clustering, and in Subsection 2.4 some facts about the spectrum of the operators \mathcal{L} , \mathcal{N}^{sym} and \mathcal{N}^{rw} . In Section 3 we prove Theorem 1.4 and Theorem 1.2. Finally, in Section 4 we prove Theorem 1.6, Theorem 1.5 and Corollary 1.7.

2. PRELIMINARIES

2.1. Transportation theory and the TL^2 space. Let D be an open domain in \mathbb{R}^d . We denote by $\mathfrak{B}(D)$ the Borel σ -algebra of D , by $\mathcal{P}(D)$ the set of all Borel probability measures on D and by $\mathcal{P}_2(D)$ the Borel probability measures on D with finite second moments. The Wasserstein distance between $\mu, \tilde{\mu} \in \mathcal{P}_2(D)$ (denoted by $d_2(\mu, \tilde{\mu})$) is defined by:

$$(2.1) \quad d_2(\mu, \tilde{\mu}) := \min \left\{ \left(\int_{D \times D} |x - y|^2 d\pi(x, y) \right)^{1/2} : \pi \in \Gamma(\mu, \tilde{\mu}) \right\},$$

where $\Gamma(\mu, \tilde{\mu})$ is the set of all *couplings* between μ and $\tilde{\mu}$, that is, the set of all Borel probability measures on $D \times D$ for which the marginal on the first variable is μ and the marginal on the second variable is $\tilde{\mu}$. The elements $\pi \in \Gamma(\mu, \tilde{\mu})$ are also referred as *transportation plans* between μ and $\tilde{\mu}$. The existence of minimizers, which justifies the definition above, is straightforward to show, see [46]. It is known that the convergence in Wasserstein metric is equivalent to weak convergence of probability measures and uniform integrability of second moments.

In the remainder, unless otherwise stated, we assume that D is a bounded set. In that setting, we have $\mathcal{P}(D) = \mathcal{P}_2(D)$ and uniform integrability of second moments is immediate. In particular, convergence in the Wasserstein metric is equivalent to weak convergence of measures. For details see for instance [46], [2] and the references within. In particular, $\mu_n \rightharpoonup \mu$ (to be read μ_n converges weakly to μ) if and only if there is a sequence of transportation plans between μ_n and μ , $\{\pi_n\}_{n \in \mathbb{N}}$, for which:

$$(2.2) \quad \lim_{n \rightarrow \infty} \iint_{D \times D} |x - y|^2 d\pi_n(x, y) = 0.$$

Actually, note that if D is bounded, (2.2) is equivalent to $\lim_{n \rightarrow \infty} \iint_{D \times D} |x - y| d\pi_n(x, y) = 0$. We say that a sequence of transportation plans, $\{\pi_n\}_{n \in \mathbb{N}}$ (with $\pi_n \in \Gamma(\mu, \mu_n)$), is *stagnating* if it satisfies (2.2). Given a Borel map $T : D \rightarrow D$ and $\mu \in \mathcal{P}(D)$, the *push-forward* of μ by T , denoted by $T_{\#}\mu \in \mathcal{P}(D)$ is given by:

$$T_{\#}\mu(A) := \mu(T^{-1}(A)), \quad A \in \mathfrak{B}(D).$$

For any bounded Borel function $\varphi : D \rightarrow \mathbb{R}$ the following change of variables in the integral holds:

$$(2.3) \quad \int_D \varphi(x) d(T_{\#}\mu)(x) = \int_D \varphi(T(x)) d\mu(x).$$

We say that a Borel map $T : D \rightarrow D$ is a *transportation map* between the measures $\mu \in \mathcal{P}(D)$ and $\tilde{\mu} \in \mathcal{P}(D)$ if $\tilde{\mu} = T_{\#}\mu$. In this case, we associate a transportation plan $\pi_T \in \Gamma(\mu, \tilde{\mu})$ to T by:

$$(2.4) \quad \pi_T := (\text{Id} \times T)_{\#}\mu,$$

where $(\text{Id} \times T) : D \rightarrow D \times D$ is given by $(\text{Id} \times T)(x) = (x, T(x))$.

It is well known that when the measure $\mu \in \mathcal{P}_2(D)$ is absolutely continuous with respect to the Lebesgue measure, the problem on the right hand side of (2.1) is equivalent to:

$$(2.5) \quad \min \left\{ \left(\int_D |x - T(x)|^2 d\mu(x) \right)^{1/2} : T_{\#}\mu = \tilde{\mu} \right\}.$$

In fact, the problem (2.1) has a unique solution which is induced (via (2.4)) by a transportation map T solving (2.5) (see [46]). In particular, boundedness of D implies that when μ has a density, then $\mu_n \rightharpoonup \mu$ as $n \rightarrow \infty$ is equivalent to the existence of a sequence $\{T_n\}_{n \in \mathbb{N}}$ of transportation maps, $(T_n)_{\#}\mu = \mu_n$ such that:

$$(2.6) \quad \int_D |x - T_n(x)|^2 d\mu(x) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

We say that a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ is *stagnating* if it satisfies (2.6).

We now introduce the space of objects that allows to simultaneously consider the discrete and continuum setting. Let

$$TL^2(D) := \{(\mu, f) : \mu \in \mathcal{P}_2(D), f \in L^2(\mu)\},$$

where $L^2(\mu)$ denotes the space of L^2 functions with respect to measure μ . For $(\mu, f), (\nu, g)$ in TL^2 define

$$(2.7) \quad d_{TL^2}((\mu, f), (\nu, g)) = \inf_{\pi \in \Gamma(\mu, \nu)} \left(\iint_{D \times D} |x - y|^2 + |f(x) - g(y)|^2 d\pi(x, y) \right)^{\frac{1}{2}}.$$

The set TL^2 and d_{TL^2} were introduced in [15], where it was also proved that d_{TL^2} is a metric. Note that if we delete the second term on the right hand side of (2.7) we recover the Wasserstein distance between the measures μ and ν . The idea of introducing the second term on the right hand side of (2.7) is to make it possible to compare functions in spaces as different as point clouds and continuous domains. We have the following characterization of convergence in TL^2 . See [15][Propositions 3.3 and 3.12] for its proof.

Proposition 2.1. *Let $(\mu, f) \in TL^2$ and let $\{(\mu_n, f_n)\}_{n \in \mathbb{N}}$ be a sequence in TL^2 . The following statements are equivalent:*

1. $(\mu_n, f_n) \xrightarrow{TL^2} (\mu, f)$ as $n \rightarrow \infty$.
2. The graphs of functions considered as measures converge in the Wasserstein sense (2.1), that is

$$(I \times f_n)_{\#}\mu_n \xrightarrow{d_2} (I \times f)_{\#}\mu \quad \text{as } n \rightarrow \infty.$$

3. $\mu_n \rightharpoonup \mu$ and for every stagnating sequence of transportation plans $\{\pi_n\}_{n \in \mathbb{N}}$ (with $\pi_n \in \Gamma(\mu, \mu_n)$)

$$(2.8) \quad \iint_{D \times D} |f(x) - f_n(y)|^2 d\pi_n(x, y) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

4. $\mu_n \rightharpoonup \mu$ and there exists a stagnating sequence of transportation plans $\{\pi_n\}_{n \in \mathbb{N}}$ (with $\pi_n \in \Gamma(\mu, \mu_n)$) for which (2.8) holds.

Moreover, if the measure μ is absolutely continuous with respect to the Lebesgue measure, the following are equivalent to the previous statements:

4. $\mu_n \rightharpoonup \mu$ and there exists a stagnating sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ (with $T_n)_{\#}\mu = \mu_n$) such that:

$$(2.9) \quad \int_D |f(x) - f_n(T_n(x))|^2 d\mu(x) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

5. $\mu_n \rightharpoonup \mu$ and for any stagnating sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ (with $T_{n\#}\mu = \mu_n$) (2.9) holds.

Remark 2.2. One can think of the convergence in TL^2 as a generalization of weak convergence of measures and convergence in L^2 of functions. By this we mean that $\{\mu_n\}_{n \in \mathbb{N}}$ in $\mathcal{P}(D)$ converges weakly (and in the Wasserstein sense) to $\mu \in \mathcal{P}(D)$ if and only if $(\mu_n, 1) \xrightarrow{TL^2} (\mu, 1)$ as $n \rightarrow \infty$, and that for $\mu \in \mathcal{P}(D)$ a sequence $\{f_n\}_{n \in \mathbb{N}}$ in $L^2(\mu)$ converges in $L^2(\mu)$ to f if and only if $(\mu, f_n) \xrightarrow{TL^2} (\mu, f)$ as $n \rightarrow \infty$. The last fact is established in Proposition 2.1.

Definition 2.3. Suppose $\{\mu_n\}_{n \in \mathbb{N}}$ in $\mathcal{P}(D)$ converges weakly to $\mu \in \mathcal{P}(D)$. We say that the sequence $\{u_n\}_{n \in \mathbb{N}}$ (with $u_n \in L^2(\mu_n)$) converges in the TL^2 -sense to $u \in L^2(\mu)$, if $\{(\mu_n, u_n)\}_{n \in \mathbb{N}}$ converges to (μ, u) in the TL^2 -metric. In this case we use a slight abuse of notation and write $u_n \xrightarrow{TL^2} u$ as $n \rightarrow \infty$. Also, we say the sequence $\{u_n\}_{n \in \mathbb{N}}$ (with $u_n \in L^2(\mu_n)$) is precompact in TL^2 if the sequence $\{(\mu_n, u_n)\}_{n \in \mathbb{N}}$ is precompact in TL^2 .

Remark 2.4. Thanks to Proposition 2.1 when μ is absolutely continuous with respect to the Lebesgue measure, $u_n \xrightarrow{TL^2} u$ as $n \rightarrow \infty$ if and only if for every (or one) $\{T_n\}_{n \in \mathbb{N}}$ stagnating sequence of transportation maps (with $T_{n\#}\mu = \mu_n$) it is true that $u_n \circ T_n \xrightarrow{L^2(\mu)} u$ as $n \rightarrow \infty$. Also $\{u_n\}_{n \in \mathbb{N}}$ is precompact in TL^2 if and only if for every (or one) $\{T_n\}_{n \in \mathbb{N}}$ stagnating sequence of transportation maps (with $T_{n\#}\mu = \mu_n$) it is true that $\{u_n \circ T_n\}_{n \in \mathbb{N}}$ is pre-compact in $L^2(\mu)$.

Lemma 2.5. Let μ_n be a sequence of Borel probability measures on \mathbb{R}^N with finite second moments, converging to a probability measure μ in the Wasserstein sense. Let A_n be μ_n measurable, and A be μ measurable. Let $\tilde{\mu}_n = \mu_{n \setminus A_n}$ and $\tilde{\mu} = \mu_{\setminus A}$. Then,

$$(2.10) \quad (\mu_n, \chi_{A_n}) \xrightarrow{TL^2} (\mu, \chi_A) \quad \text{if and only if} \quad \tilde{\mu}_n \rightharpoonup \tilde{\mu}$$

as $n \rightarrow \infty$.

Proof. From Proposition 2.1 follows that $(\mu_n, \chi_{A_n}) \xrightarrow{TL^2} (\mu, \chi_A)$ if and only if $\tilde{\mu}_n \times \delta_{\{1\}} + (\mu_n - \tilde{\mu}_n) \times \delta_{\{0\}} \xrightarrow{d_2} \tilde{\mu} \times \delta_{\{1\}} + (\mu - \tilde{\mu}) \times \delta_{\{0\}}$, as $n \rightarrow \infty$.

Since convergence in Wasserstein distance implies weak convergence, we deduce that $\tilde{\mu}_n \times \delta_{\{1\}} + (\mu_n - \tilde{\mu}_n) \times \delta_{\{0\}} \rightharpoonup \tilde{\mu} \times \delta_{\{1\}} + (\mu - \tilde{\mu}) \times \delta_{\{0\}}$, and in particular we conclude that

$$\tilde{\mu}_n \rightharpoonup \tilde{\mu}, \quad \text{as } n \rightarrow \infty.$$

Conversely, the weak convergence $\tilde{\mu}_n \rightharpoonup \tilde{\mu}$, together with the fact that $\mu_n \xrightarrow{d_2} \mu$ (which in particular implies that $\mu_n \rightharpoonup \mu$), imply that

$$\tilde{\mu}_n \times \delta_{\{1\}} + (\mu_n - \tilde{\mu}_n) \times \delta_{\{0\}} \rightharpoonup \tilde{\mu} \times \delta_{\{1\}} + (\mu - \tilde{\mu}) \times \delta_{\{0\}}.$$

In order to conclude that the above convergence also holds in the Wasserstein sense, we simply note that this follows from the the uniform integrability of the second moments of $\{\tilde{\mu}_n\}_{n \in \mathbb{N}}$, which in turn follows from

$$\limsup_{t \rightarrow \infty} \sup_{n \in \mathbb{N}} \int_{\{|x| > t\}} |x|^2 d\tilde{\mu}_n(x) \leq \limsup_{t \rightarrow \infty} \sup_{n \in \mathbb{N}} \int_{\{|x| > t\}} |x|^2 d\mu_n(x) = 0.$$

The equality in the previous expression follows from the fact that $\mu_n \xrightarrow{d_2} \mu$. \square

The following proposition states that inner products are continuous with respect to the TL^2 -convergence.

Proposition 2.6. *Suppose that $(\mu_n, u_n) \xrightarrow{TL^2} (\mu, u)$ and $(\mu_n, v_n) \xrightarrow{TL^2} (\mu, v)$ as $n \rightarrow \infty$. Then,*

$$(2.11) \quad \lim_{n \rightarrow \infty} \langle u_n, v_n \rangle_{\mu_n} = \langle u, v \rangle_{\mu}.$$

Proof. By the polarization identity, it is enough to prove that if $(\mu_n, u_n) \xrightarrow{TL^2} (\mu, u)$ then,

$$(2.12) \quad \lim_{n \rightarrow \infty} \|u_n\|_{\mu_n} = \|u\|_{\mu}.$$

For this purpose, consider a stagnating sequence of transportation plans $\{\pi_n\}_{n \in \mathbb{N}}$ with $\pi_n \in \Gamma(\mu, \mu_n)$.

We can write $\|u_n\|_{\mu_n} = \left(\int_{D \times D} |u_n(y)|^2 d\pi_n(x, y) \right)^{1/2}$ and $\|u\|_{\mu} = \left(\int_{D \times D} |u(x)|^2 d\pi_n(x, y) \right)^{1/2}$. Hence,

$$(2.13) \quad | \|u_n\|_{\mu_n} - \|u\|_{\mu} | \leq \left(\int_{D \times D} |u_n(y) - u(x)|^2 d\pi_n(x, y) \right)^{1/2} \rightarrow 0, \text{ as } n \rightarrow \infty.$$

□

In proving the convergence of k -means clustering (statement 4. in Theorems 1.2 and 1.5)) we also need the following result on TL^2 convergence of a composition of functions.

Lemma 2.7 (Continuity of composition in TL^2). *Let $\{\mu_n\}_{n \in \mathbb{N}}$ and μ be a collection of Borel probability measures on \mathbb{R}^d with finite second moments. Let $F_n \in L^2(\mu_n, \mathbb{R}^d : \mathbb{R}^k)$ for all $n \in \mathbb{N}$ and $F \in L^2(\mu, \mathbb{R}^d : \mathbb{R}^k)$. Consider the measures $\tilde{\mu}_n := F_n \# \mu_n$ for all $n \in \mathbb{N}$ and $\tilde{\mu} := F \# \mu$. Finally, let $\tilde{f}_n \in L^2(\tilde{\mu}_n, \mathbb{R}^k : \mathbb{R})$ for all $n \in \mathbb{N}$ and $\tilde{f} \in L^2(\tilde{\mu}, \mathbb{R}^k : \mathbb{R})$. If*

$$(\mu_n, F_n) \xrightarrow{TL^2} (\mu, F) \quad \text{as } n \rightarrow \infty,$$

and

$$(\tilde{\mu}_n, \tilde{f}_n) \xrightarrow{TL^2} (\tilde{\mu}, \tilde{f}) \quad \text{as } n \rightarrow \infty.$$

Then,

$$(\mu_n, \tilde{f}_n \circ F_n) \xrightarrow{TL^2} (\mu, \tilde{f} \circ F) \quad \text{as } n \rightarrow \infty.$$

Proof. First of all note that the fact that $F_n \in L^2(\mu_n, \mathbb{R}^d : \mathbb{R}^k)$ and $F \in L^2(\mu, \mathbb{R}^d : \mathbb{R}^k)$ guarantees that $\tilde{\mu}_n$ and $\tilde{\mu}$ are probability measures on \mathbb{R}^k with finite second moments. On the other hand, $(\mu_n, F_n) \xrightarrow{TL^2} (\mu, F)$ as $n \rightarrow \infty$ implies the existence of a stagnating sequence of transportation maps $\{\pi_n\}_{n \in \mathbb{N}}$ with $\pi_n \in \Gamma(\mu, \mu_n)$ such that

$$(2.14) \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} |F(x) - F_n(y)|^2 d\pi_n(x, y) = 0.$$

We consider the measures $\tilde{\pi}_n := (F \times F_n) \# \pi_n$ for all $n \in \mathbb{N}$. It is straightforward to check that $\tilde{\pi}_n \in \Gamma(\tilde{\mu}, \tilde{\mu}_n)$ for all $n \in \mathbb{N}$ and by the definition of $\tilde{\pi}_n$ that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^k \times \mathbb{R}^k} |\tilde{x} - \tilde{y}|^2 d\tilde{\pi}_n(\tilde{x}, \tilde{y}) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} |F(x) - F_n(y)|^2 d\pi_n(x, y) = 0$$

In other words, $\{\tilde{\pi}_n\}_{n \in \mathbb{N}}$ is a stagnating sequence of transportation maps with $\tilde{\pi}_n \in \Gamma(\tilde{\mu}, \tilde{\mu}_n)$. From the fact that $(\tilde{\mu}_n, \tilde{f}_n) \xrightarrow{TL^2} (\tilde{\mu}, \tilde{f})$ as $n \rightarrow \infty$ and from Proposition 2.1 it follows that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^k \times \mathbb{R}^k} |\tilde{f}(\tilde{x}) - \tilde{f}_n(\tilde{y})|^2 d\tilde{\pi}_n(\tilde{x}, \tilde{y}) = 0.$$

But again by the definition of $\tilde{\pi}_n$ we deduce

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} |\tilde{f}(F(x)) - \tilde{f}_n(F_n(y))|^2 d\pi_n(x, y) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^k \times \mathbb{R}^k} |\tilde{f}(\tilde{x}) - \tilde{f}_n(\tilde{y})|^2 d\tilde{\pi}_n(\tilde{x}, \tilde{y}) = 0$$

Using again Proposition 2.1 we obtain the desired result. \square

2.2. Γ -convergence. We recall the notion of Γ -convergence in general setting.

Definition 2.8. Let (X, d_X) be a metric space and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{F_n\}_{n \in \mathbb{N}}$ be a sequence of (random) functionals $F_n : X \times \Omega \rightarrow [0, \infty]$ and let F be a (deterministic) functional $F : X \rightarrow [0, \infty]$. We say that the sequence of functionals $\{F_n\}_{n \in \mathbb{N}}$ Γ -converges (in the d_X metric) to F , if for almost every $\omega \in \Omega$, **all** of the following conditions hold:

- (1) **Liminf inequality:** For all $x \in X$ and all sequences $\{x_n\}_{n \in \mathbb{N}}$ converging to x in the metric d_X it is true that

$$\liminf_{n \rightarrow \infty} F_n(x) \geq F(x)$$

- (2) **Limsup inequality:** For all $x \in X$ there exists a sequence $\{x_n\}_{n \in \mathbb{N}}$ converging to x in the metric d_X such that

$$\limsup_{n \rightarrow \infty} F_n(x) \leq F(x)$$

The notion of Γ -convergence is particularly useful when combined with an appropriate notion of compactness. See [9, 12].

Definition 2.9. We say that the sequence of nonnegative random functionals $\{F_n\}_{n \in \mathbb{N}}$ satisfies the compactness property if for almost every $\omega \in \Omega$, it is true that every bounded (with respect to d_X) sequence $\{x_n\}_{n \in \mathbb{N}}$ in X for which

$$\sup_{n \in \mathbb{N}} F_n(x) < \infty,$$

is precompact in X .

Now that we have defined the TL^2 -space, and we have defined the notion of Γ -convergence, we can rephrase the content of Theorem 1.4 in the following way. Under the conditions on the domain D , the density ρ and the parameter ε_n in Theorem 1.4, with probability one, **all** of the following statements hold:

- (1) **Liminf inequality:** For all $u \in L^2(\nu)$, and all sequences $\{u_n\}_{n \in \mathbb{N}}$ with $u_n \in L^2(\nu_n)$ and with $u_n \xrightarrow{TL^2} u$ it is true that

$$\liminf_{n \rightarrow \infty} G_{n, \varepsilon_n}(u_n) \geq \sigma_\eta G(u).$$

- (2) **Limsup inequality:** For all $u \in L^2(\nu)$, there exists a sequence $\{u_n\}_{n \in \mathbb{N}}$ with $u_n \in L^2(\nu_n)$ and with $u_n \xrightarrow{TL^2} u$ for which

$$\limsup_{n \rightarrow \infty} G_{n, \varepsilon_n}(u_n) \leq \sigma_\eta G(u).$$

- (3) **Compactness:** Every sequence $\{u_n\}_{n \in \mathbb{N}}$ with $u_n \in L^2(\nu_n)$, satisfying

$$\sup_{n \in \mathbb{N}} F_n(x) < \infty,$$

is precompact in TL^2 , that is, every subsequence of $\{u_n\}_{n \in \mathbb{N}}$ has a further subsequence, which converges in the TL^2 -sense to an element of $L^2(D)$.

In a similar fashion we can rephrase the content of Theorem 1.6.

2.3. Stability of k -means clustering. Here we prove some basic facts about the functional $F_{\mu,k}$ defined in (1.24) and about Theorem 1.8. Our first observation is that there exist minimizers of $F_{\mu,k}$. We note that $F_{\mu,k}$ is a continuous function which is non-negative and the existence of minimizers can be obtained from a straightforward application of the direct method of the calculus of variations as we now illustrate. If the support of μ has k or fewer points, then including these points in \mathbf{z} provides a minimizer for which $F(\mathbf{z}) = 0$. On the other hand, if the support of μ has more than k points then to show that a minimizer exists it is enough to obtain pre-compactness of a minimizing sequence due to the continuity of $F_{\mu,k}$. Let $\{\mathbf{z}^m\}_{m \in \mathbb{N}}$ be a minimizing sequence of $F_{\mu,k}$.

By considering a subsequence we can assume that for any $i = 1, \dots, k$, z_i^m either converges to some $z_i \in \mathbb{R}^N$ or diverges to $\pm\infty$. Also without the loss of generality we can assume that for some $1 \leq l \leq k+1$, the sequence $\{z_i^m\}_{m \in \mathbb{N}}$ converges for $i < l$ and diverges for $i \geq l$. Our goal is to show that $l = k+1$. Assume for the sake of contradiction that $l \leq k$. First note that if $l = 1$ (when no subsequence converges) then $F_{\mu,k}(\mathbf{z}^m) \rightarrow \infty$ as $m \rightarrow \infty$, which is impossible. So we can assume that z_1^m converges to z_1 as $m \rightarrow \infty$. It is straightforward to show using the finiteness of the second moment of μ , that

$$(2.15) \quad F_{\mu,k}(\mathbf{z}^m) \rightarrow F_{\mu,l-1}(\{z_1, \dots, z_{l-1}\}), \quad \text{as } m \rightarrow \infty.$$

However unless $l = k+1$, adding $k - (l-1)$ points from $\text{supp}(\mu) \setminus \{z_1, \dots, z_{l-1}\}$ to $\{z_1, \dots, z_{l-1}\}$ would result on a value of $F_{\mu,k}$ that is strictly below $F_{\mu,l-1}(\{z_1, \dots, z_{l-1}\})$ and from (2.15), this would contradict the assumption that $\{\mathbf{z}^m\}_{m \in \mathbb{N}}$ is a minimizing sequence. We conclude that $\{\mathbf{z}^m\}_{m \in \mathbb{N}}$ converges up to subsequence.

We now turn to comparing the properties of $F_{\mu,k}$ for different measures μ , the ultimate goal is to prove Theorem 1.8. Let μ and ν be Borel probability measures on \mathbb{R}^N with finite second moments and let $\pi \in \Gamma(\mu, \nu)$ be the optimal transportation plan realizing the Wasserstein distance between μ and ν , that is, assume

$$d_2^2(\mu, \nu) = \iint_{\mathbb{R}^N \times \mathbb{R}^N} |x - y|^2 d\pi(x, y).$$

Then

$$\begin{aligned} |F_{\mu,k}(\mathbf{z}) - F_{\nu,k}(\mathbf{z})| &= \left| \int_{\mathbb{R}^N} d(x, \mathbf{z})^2 d\mu(x) - \int_{\mathbb{R}^N} d(y, \mathbf{z})^2 d\nu(y) \right| \\ &= \left| \iint_{\mathbb{R}^N \times \mathbb{R}^N} (d(x, \mathbf{z})^2 - d(y, \mathbf{z})^2) d\pi(x, y) \right| \\ &\leq \iint_{\mathbb{R}^N \times \mathbb{R}^N} |d(x, \mathbf{z})^2 - d(y, \mathbf{z})^2| d\pi(x, y) \\ &\leq \iint_{\mathbb{R}^N \times \mathbb{R}^N} (|x - y| + d(y, \mathbf{z}))^2 - d(y, \mathbf{z})^2 d\pi(x, y) \\ &\leq d_2^2(\mu, \nu) + 2d_2(\mu, \nu) \sqrt{F_{\nu,k}(\mathbf{z})}, \end{aligned}$$

where the last inequality is obtained after expanding the integrand and using Cauchy-Schwartz inequality. By symmetry, we conclude that

$$(2.16) \quad |F_{\mu,k}(\mathbf{z}) - F_{\nu,k}(\mathbf{z})| \leq d_2(\mu, \nu) \left(2 \min \left\{ \sqrt{F_{\mu,k}(\mathbf{z})}, \sqrt{F_{\nu,k}(\mathbf{z})} \right\} + d_2(\mu, \nu) \right).$$

We also need the following Lemma.

Lemma 2.10. *Let μ be a Borel probability measure on \mathbb{R}^k with finite second moment and at least k points in its support. Let $\mathbf{z} = (z_1, \dots, z_k)$ be a minimizer of the functional $F_{\mu,k}$. Denote by V_1, \dots, V_k*

the (closed) Voronoi cells induced by the points z_1, \dots, z_k , i.e., $V_i := \{x \in \mathbb{R}^k : |x - z_i| = d(x, \mathbf{z})\}$. Then,

$$\mu(V_i \cap V_j) = 0, \quad \forall i \neq j.$$

Proof. We start by recalling that if $k = 1$ then the minimizer z_1 of $F_{\mu,1}$ is the centroid of μ , that is $z_1 = \int x d\mu(x)$. We now consider $k \geq 2$. Since the support of μ has at least k points, the points z_1, \dots, z_k are distinct. Assume that $\mu(V_i \cap V_j) > 0$ for some $i \neq j$. Note that the set $V_i \cap V_j$ is contained in the plane P_{ij} with normal vector $z_i - z_j$, and containing the point $\frac{1}{2}z_i + \frac{1}{2}z_j$. Let $\bar{\mu}_i = \mu_{\perp V_i}$ and $\bar{\theta}_i = \mu - \bar{\mu}_i$. Let $\hat{\mathbf{z}} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n)$. Note that $F_{\mu,k}(\mathbf{z}) = F_{\bar{\mu}_i,1}(z_i) + F_{\bar{\theta}_i,k-1}(\hat{\mathbf{z}})$. Consequently z_i minimizes $F_{\bar{\mu}_i,1}$ and by remark above, z_i is the centroid of $\bar{\mu}_i$, that is

$$z_i = \frac{1}{\bar{\mu}_i(\mathbb{R}^k)} \int_{\mathbb{R}^k} x d\bar{\mu}_i(x).$$

Now, let $\underline{\mu}_i = \mu_{\perp V_i \setminus V_j}$ and $\underline{\theta}_i = \mu - \underline{\mu}_i$. Analogously to above $F_{\mu,k}(\mathbf{z}) = F_{\underline{\mu}_i,1}(z_i) + F_{\underline{\theta}_i,k-1}(\hat{\mathbf{z}})$. Hence z_i minimizes $F_{\underline{\mu}_i,1}$ and thus is the centroid of $\underline{\mu}_i$, i.e.,

$$z_i = \frac{1}{\underline{\mu}_i(\mathbb{R}^k)} \int_{\mathbb{R}^k} x d\underline{\mu}_i(x).$$

But note that $\mu(V_i \cap V_j) > 0$ implies that

$$\frac{1}{\bar{\mu}_i(\mathbb{R}^k)} \int_{\mathbb{R}^k} \langle x, z_j - z_i \rangle d\bar{\mu}_i(x) > \frac{1}{\underline{\mu}_i(\mathbb{R}^k)} \int_{\mathbb{R}^k} \langle x, z_j - z_i \rangle d\underline{\mu}_i(x).$$

This, contradicts the fact that the centroids of $\underline{\mu}_i$ and $\bar{\mu}_i$ are both equal to z_i . \square

The proof of Theorem 1.8 is now a direct consequence of (2.16) and Lemma 2.10.

Proof of Theorem 1.8. Let $a^k = F_{\mu,k}(\mathbf{z}^k)$ and $a_m^k = F_{\mu_m,k}(\mathbf{z}_m^k)$, where \mathbf{z}^k is a minimizer of $F_{\mu,k}$ and where \mathbf{z}_m^k is a minimizer of $F_{\mu_m,k}$. Note that since the support of μ has at least k points, $a_l > a_k$ for all $l < k$. From (2.16) it follows that $a_m^k \rightarrow a^k$ as $m \rightarrow \infty$ for any $k \in \mathbb{N}$. To show that $\{\mathbf{z}_m^k, m \in \mathbb{N}\}$ is precompact it is enough to show that all coordinates are uniformly bounded. If this is not the case then there exists $1 \leq l \leq k$ such that coordinates 1 to l converge, while those between $l+1$ and k diverge to $\pm\infty$. Arguing as in the proof of the existence of minimizers at the beginning of this Section, and using (2.16), one obtains that $F_{\mu_m,k}(\mathbf{z}_m^k)$ converges to $F_{\mu,l}(\{z_1, \dots, z_l\})$ for some $z_1, \dots, z_l \in \mathbb{R}^N$. If $l < k$, then this would imply that $a_l \leq F_{\mu,l}(\{z_1, \dots, z_l\}) = \lim_{m \rightarrow \infty} a_m^k = a_k$, which would contradict the fact that $a_l > a_k$. Thus concluding that along a subsequence $\mathbf{z}_m^k \rightarrow \bar{\mathbf{z}}^k$ for some $\bar{\mathbf{z}}^k$. To show that $\bar{\mathbf{z}}^k$ minimizes $F_{\mu,k}$ simply observe that from (2.16) and the continuity of $F_{\mu,k}$ it follows that

$$F_{\mu,k}(\bar{\mathbf{z}}^k) = \lim_{m \rightarrow \infty} F_{\mu,k}(\mathbf{z}_m^k) = \lim_{n \rightarrow \infty} F_{\mu_m,k}(\mathbf{z}_m^k) = \lim_{m \rightarrow \infty} a_m^k = a^k,$$

which implies that indeed $\bar{\mathbf{z}}^k$ minimizes $F_{\mu,k}$.

Finally, the last part of the Theorem on convergence of clusters, follows from the fact that μ_m converge weakly to μ , that their second moments are uniformly bounded, and that the boundaries of Voronoi cells change continuously when the centers are perturbed. \square

2.4. The Spectra of \mathcal{L} , \mathcal{N}^{sym} and \mathcal{N}^{rw} . The purpose of this section is to present some facts about the spectra of the operators \mathcal{L} , \mathcal{N}^{sym} , and \mathcal{N}^{rw} . These facts are standard (see [13], or Chapter 8 in [6]). We present them for the convenience of the reader.

Let D be an open, bounded, connected domain with Lipschitz boundary and let $\rho : D \rightarrow \mathbb{R}$ be a continuous density function satisfying (1.15). For a given $w \in L^2(D)$ we consider the PDE

$$(2.17) \quad \begin{aligned} \mathcal{L}(u) &= w && \text{in } D, \\ \frac{\partial u}{\partial \mathbf{n}} &= 0 && \text{on } \partial D, \end{aligned}$$

where we recall \mathcal{L} is formally defined as $\mathcal{L}(u) = -\frac{1}{\rho} \operatorname{div}(\rho^2 \nabla u)$. We say that $u \in H^1(D)$ is a *weak solution* of (2.17) if

$$(2.18) \quad \int_D \nabla u \cdot \nabla v \rho^2(x) dx = \int_D v w \rho(x) dx, \quad \forall v \in H^1(D).$$

Remark 2.11. Note that if u is a solution of (2.17) in the classical sense, then integration by parts shows that u is a weak solution of (2.17).

A necessary condition for (2.17) to have a solution in the weak sense, is that w belongs to the space

$$\mathcal{U} := \left\{ w \in L^2(D) : \int_D w \rho(x) dx = 0 \right\}.$$

This can be deduced by considering the test function $v \equiv 1$ in (2.18). We consider the space

$$\mathcal{V} := \left\{ v \in H^1(D) : \int_D v \rho(x) dx = 0 \right\},$$

and consider the bilinear form $a : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ given by

$$(2.19) \quad a(u, v) := \int_D \nabla u \cdot \nabla v \rho^2 dx.$$

One can use the assumptions on ρ in (1.15), and Poincaré's inequality (see Theorem 12.23 in [27]), to show that a is coercive with respect to the H^1 inner product on \mathcal{V} , defined by

$$\langle u, v \rangle_{H^1(D)} := \int_D u v dx + \int_D \nabla u \cdot \nabla v dx.$$

In addition, a is continuous and symmetric.

Therefore by Lax-Milgram theorem [13][Sec. 6.2] for any $w \in \mathcal{U}$ there exists a unique solution $u \in \mathcal{V}$ to (2.17). From (2.18) and the assumption (1.15) on ρ , it follows that

$$(2.20) \quad \int_D |\nabla u|^2 \rho^2(x) dx \leq C \int_D |w|^2 \rho(x) dx,$$

for a constant C . We can then define the inverse $\mathcal{L}^{-1} : \mathcal{U} \rightarrow \mathcal{V}$ of \mathcal{L} , by letting $\mathcal{L}^{-1} : w \mapsto u$, where u is the unique solution of (2.17). From (2.20), it follows that \mathcal{L}^{-1} is a continuous linear function. Rellich–Kondrachov theorem (see Theorem 11.10 in [27]) implies that \mathcal{L}^{-1} is compact.

We say that $\lambda \in \mathbb{R}$ is an *eigenvalue* of the operator \mathcal{L} , if there exists a nontrivial $u \in H^1(D)$ which is a weak solution of (1.11). That is if

$$(2.21) \quad a(u, v) = \int_D \nabla u \cdot \nabla v \rho^2(x) dx = \lambda \int_D u v \rho(x) dx = \lambda \langle u, v \rangle_\rho, \quad \forall v \in H^1(D).$$

Such function u is called an *eigenfunction*.

Remark 2.12. We remark that $\lambda_1 = 0$ is an eigenvalue of \mathcal{L} and that the function u_1 identically equal to one is an eigenfunction associated to λ_1 . Given that D is connected, it follows that the eigenspace associated to $\lambda_1 = 0$ is the space of constant functions on D . We also remark that \mathcal{U} is by definition the orthogonal complement (with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$) of $\operatorname{Span}\{u_1\}$.

Using the definition of \mathcal{L} , the definition of weak solutions to (2.17) it follows that

$$(2.22) \quad u \text{ is an eigenfunction of } \mathcal{L} \text{ with eigenvalue } \lambda \neq 0 \text{ iff } \mathcal{L}^{-1}(u) = \frac{1}{\lambda}u.$$

In other words the non-constant eigenfunctions of \mathcal{L} are the eigenfunctions of \mathcal{L}^{-1} , and the nonzero eigenvalues of \mathcal{L} are the reciprocals of the eigenvalues of \mathcal{L} . Thus, by understanding the structure of the spectrum of \mathcal{L}^{-1} , one can obtain properties of the spectrum of \mathcal{L} .

Proposition 2.13. *The operator $\mathcal{L}^{-1} : \mathcal{V} \rightarrow \mathcal{V}$ is a selfadjoint, positive semidefinite (with respect to the inner product $a(\cdot, \cdot)$) and compact. The eigenvalues of \mathcal{L}^{-1} can be arranged as a decreasing sequence of positive numbers,*

$$\lambda_2^{-1} \geq \lambda_3^{-1} \geq \dots$$

repeated according to (finite) multiplicity and converging to zero. Moreover, there exists an orthonormal basis $\{v_k\}_{k \geq 2}$ of \mathcal{V} , where each of the functions v_k is an eigenfunction of \mathcal{L}^{-1} with corresponding eigenvalue λ_k^{-1} .

Proof. In order to show that $\mathcal{L}^{-1} : \mathcal{V} \rightarrow \mathcal{V}$ is self-adjoint with respect to $a(\cdot, \cdot)$, take $v_1, v_2 \in \mathcal{V}$ and let $u_i = \mathcal{L}^{-1}v_i$ for $i = 1, 2$. We claim that

$$a(\mathcal{L}^{-1}v_1, v_2) = \langle v_1, v_2 \rangle_\rho.$$

In fact, from the definition of \mathcal{L}^{-1} it follows that

$$a(\mathcal{L}^{-1}v_1, v_2) = a(u_1, v_2) = \int_D \nabla u_1 \cdot \nabla v_2 \rho^2(x) dx = \int_D v_1 v_2 \rho(x) dx = \langle v_1, v_2 \rangle_\rho$$

From the previous identity, it immediately follows that \mathcal{L}^{-1} is self-adjoint and positive semidefinite with respect to the inner product $a(\cdot, \cdot)$. The compactness of \mathcal{L}^{-1} follows from Rellich–Kondrachov theorem (see Theorem 11.10 in [27]). The statements about the spectrum of \mathcal{L}^{-1} are a direct consequence of Riesz-Schauder theorem and Hilbert-Schmidt theorem (see [33]). \square

For $k \geq 2$, let v_k be eigenfunctions as in the previous proposition and define u_k by

$$(2.23) \quad u_k := \sqrt{\lambda_k} v_k.$$

We claim that $\{u_k\}_{k \geq 2}$ is an orthonormal base of \mathcal{U} with respect to $\langle \cdot, \cdot \rangle_\rho$. In fact, it follows from the definition of \mathcal{L}^{-1} and (2.28) that

$$\delta_{kl} = a(v_k, v_l) = \lambda_k \langle v_k, v_l \rangle_\rho = \frac{\sqrt{\lambda_k}}{\sqrt{\lambda_l}} \langle u_k, u_l \rangle_\rho,$$

where $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ if $k \neq l$. Hence $\langle u_k, u_l \rangle_\rho = \delta_{kl}$. In other words $\{u_k\}_{k \geq 2}$ is an orthonormal set. Completeness follows from the completeness in Proposition 2.13 and density of $H^1(D)$ in $L^2(D)$.

By setting $u_1 \equiv 1$ and by noticing that $L^2(D) = \text{Span}\{u_1\} \oplus \mathcal{U}$, we conclude that $\{u_k\}_{k \in \mathbb{N}}$ is an orthonormal base for $L^2(D)$ with inner product $\langle \cdot, \cdot \rangle_\rho$. The next proposition is a direct consequence of the previous discussion and (2.22).

Proposition 2.14. *\mathcal{L} has a countable family of eigenvalues $\{\lambda_k\}_{k \in \mathbb{N}}$ which can be written as an increasing sequence of nonnegative numbers which tends to infinity as k goes to infinity, that is,*

$$0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_k \leq \dots$$

Each eigenvalue, is repeated according to (finite) multiplicity. Moreover, there exists $\{u_k\}_{k \in \mathbb{N}}$ an orthonormal basis (with respect to $\langle \cdot, \cdot \rangle_\rho$) of $L^2(D)$, such that for every $k \in \mathbb{N}$, u_k is an eigenfunction of \mathcal{L} associated to λ_k .

Finally we present the Courant-Fisher maxmini principle.

Proposition 2.15. *Consider an orthonormal base $\{u_k\}_{k \in \mathbb{N}}$ for $L^2(D)$ with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$, where for each $k \in \mathbb{N}$, u_k is an eigenfunction of \mathcal{L} with eigenvalue λ_k . Then, for every $k \in \mathbb{N}$*

$$(2.24) \quad \lambda_k = \min_{\|u\|_\rho=1, u \in S^{*\perp}} G(u),$$

where $S^* = \text{Span}\{u_1, \dots, u_{k-1}\}$ and where $S^{*\perp}$ denotes the orthogonal complement of S^* with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$. Additionally,

$$(2.25) \quad \lambda_k = \sup_{S \in \Sigma_{k-1}} \min_{\|u\|_\rho=1, u \in S^\perp} G(u),$$

where Σ_{k-1} denotes the set of $(k-1)$ -dimensional subspaces of $L^2(D)$, and where S^\perp represents the orthogonal complement of S with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$.

The proof (of a similar statement) can be found in Chapter 8.3 in [6].

Remark 2.16. If the density ρ is smooth, then the eigenfunctions of \mathcal{L} are smooth inside D .

We now turn to the spectrum of \mathcal{N}^{sym} . We say that $\tau \in \mathbb{R}$ is an *eigenvalue* of the operator \mathcal{N}^{sym} , if there exists a nontrivial $u \in H^1_{\sqrt{\rho}}(D)$ which solves (1.13). That is if

$$(2.26) \quad \int_D \nabla \left(\frac{u}{\sqrt{\rho}} \right) \cdot \nabla \left(\frac{v}{\sqrt{\rho}} \right) \rho^2(x) dx = \tau \int_D uv \rho(x) dx, \quad \forall v \in H^1_{\sqrt{\rho}}(D).$$

The function u is then called an eigenfunction of \mathcal{N}^{sym} with eigenvalue τ .

Remark 2.17. We remark that $\tau_1 = 0$ is an eigenvalue of \mathcal{N}^{sym} and that the function \bar{u}_1 equal to

$$\bar{u}_1(x) = \frac{\sqrt{\rho(x)}}{\|\sqrt{\rho}\|_\rho}$$

is an eigenfunction of \mathcal{N}^{sym} , with eigenvalue $\tau_1 = 0$. Given that D is connected, it actually follows that $\tau_1 = 0$ has multiplicity one and thus the eigenspace associated to $\tau_1 = 0$ is the space of multiples of $\sqrt{\rho}$.

Following the same ideas used when considering the spectrum of \mathcal{L} , we can establish the following analogous results.

Proposition 2.18. *\mathcal{N}^{sym} has a countable family of eigenvalues $\{\tau_k\}_{k \in \mathbb{N}}$ which can be written as an increasing sequence of nonnegative numbers which tends to infinity as k goes to infinity, that is,*

$$0 = \tau_1 \leq \tau_2 \leq \dots \leq \tau_k \leq \dots$$

Each eigenvalue, is repeated according to (finite) multiplicity. Moreover, there exists $\{\bar{u}_k\}_{k \in \mathbb{N}}$ an orthonormal basis (with respect to $\langle \cdot, \cdot \rangle_\rho$) of $L^2(D)$, such that for every $k \in \mathbb{N}$, \bar{u}_k is an eigenfunction of \mathcal{N}^{sym} associated to τ_k .

Proposition 2.19. *Consider a orthonormal base $\{\bar{u}_k\}_{k \in \mathbb{N}}$ for $L^2(D)$ with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$, where for each $k \in \mathbb{N}$, \bar{u}_k is an eigenfunction of \mathcal{N}^{sym} with eigenvalue τ_k . Then, for every $k \in \mathbb{N}$*

$$(2.27) \quad \tau_k = \min_{\|u\|_\rho=1, u \in S^{*\perp}} \bar{G}(u),$$

where $S^* = \text{Span}\{\bar{u}_1, \dots, \bar{u}_{k-1}\}$. Additionally,

$$\tau_k = \sup_{S \in \Sigma_{m-1}} \min_{\|u\|_\rho=1, u \in S^\perp} \bar{G}(u),$$

where Σ_{m-1} denotes the set of $(m-1)$ -dimensional subspaces of $L^2(D)$, and where S^\perp represents the orthogonal complement of S with respect to the inner product $\langle \cdot, \cdot \rangle_\rho$.

Finally, we consider the spectrum of \mathcal{N}^{rw} . We say that $\tau \in \mathbb{R}$ is an *eigenvalue* of the operator \mathcal{N}^{rw} , if there exists a nontrivial $u \in H^1(D)$ for which

$$(2.28) \quad \int_D \nabla u \cdot \nabla v \rho^2(x) dx = \tau \int_D uv \rho^2(x) dx, \quad \forall v \in H^1(D).$$

The function u is then called an eigenfunction of \mathcal{N}^{rw} with eigenvalue τ . From the definition, it follows that τ is an eigenvalue of \mathcal{N}^{rw} with eigenfunction u if and only if τ is an eigenvalue of \mathcal{N}^{sym} with eigenvector $w := \sqrt{\rho}u$. This is analogous to (1.4) in the discrete case.

3. CONVERGENCE OF THE SPECTRA OF UNNORMALIZED GRAPH LAPLACIANS

We start by establishing Theorem 1.4.

Proof of Theorem 1.4. As done in Section 5 in [15] and due to the assumptions **(K1)** – **(K3)** on $\boldsymbol{\eta}$, we can reduce the problem to that of showing the result for the kernel $\boldsymbol{\eta}$ defined by

$$\boldsymbol{\eta}(t) := \begin{cases} 1, & \text{if } t \in [0, 1], \\ 0, & \text{if } t > 1. \end{cases}$$

We use the sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ from Proposition 1.12. Let $\omega \in \Omega$ be such that (1.27) and (1.28) hold in cases $d = 2$ and $d \geq 3$ respectively. By Proposition 1.12 the complement in Ω of such ω 's is contained in a set of probability zero. The key idea in the proof is that the estimates of Proposition 1.12 imply that the transportation happens on a length scale which is small compared to ε_n . By taking a kernel with slightly smaller radius than ε_n we can then obtain a lower bound, and by taking a slightly larger radius a matching upper bound on the functional G_{n, ε_n} .

Liminf inequality: Assume that $u_n \xrightarrow{TL^1} u$ as $n \rightarrow \infty$. Since $T_{n\#}\nu = \nu_n$, using the change of variables (2.3) it follows that

$$(3.1) \quad G_{n, \varepsilon_n}(u_n) = \frac{1}{\varepsilon_n^2} \int_{D \times D} \eta_{\varepsilon_n}(T_n(x) - T_n(y)) (u_n \circ T_n(x) - u_n \circ T_n(y))^2 \rho(x) \rho(y) dx dy.$$

Note that for $(x, y) \in D \times D$

$$(3.2) \quad |T_n(x) - T_n(y)| > \varepsilon_n \Rightarrow |x - y| > \varepsilon_n - 2\|Id - T_n\|_\infty.$$

Thanks to the assumptions on $\{\varepsilon_n\}_{n \in \mathbb{N}}$ ((1.27) and (1.28) in cases $d = 2$ and $d \geq 3$ respectively), for large enough $n \in \mathbb{N}$:

$$(3.3) \quad \tilde{\varepsilon}_n := \varepsilon_n - 2\|Id - T_n\|_\infty > 0.$$

By (3.2), and our choice of kernel $\boldsymbol{\eta}$, for large enough n and for every $(x, y) \in D \times D$, we obtain

$$\boldsymbol{\eta}\left(\frac{|x - y|}{\tilde{\varepsilon}_n}\right) \leq \boldsymbol{\eta}\left(\frac{|T_n(x) - T_n(y)|}{\varepsilon_n}\right).$$

We now consider $\tilde{u}_n = u_n \circ T_n$. Thanks to the previous inequality and (3.1), for large enough n

$$\begin{aligned} G_{n, \varepsilon_n}(u_n) &\geq \frac{1}{\varepsilon_n^{d+1}} \int_{D \times D} \boldsymbol{\eta}\left(\frac{|x - y|}{\tilde{\varepsilon}_n}\right) (\tilde{u}_n(x) - \tilde{u}_n(y))^2 \rho(x) \rho(y) dx dy \\ &= \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{d+2} G_{\tilde{\varepsilon}_n}(\tilde{u}_n). \end{aligned}$$

Note that $\frac{\tilde{\varepsilon}_n}{\varepsilon_n} \rightarrow 1$ as $n \rightarrow \infty$ and that $u_n \xrightarrow{TL^1} u$ by definition implies $\tilde{u}_n \xrightarrow{L^1(D, \rho)} u$ as $n \rightarrow \infty$. We deduce from the liminf inequality of Proposition 1.10 that $\liminf_{n \rightarrow \infty} G_{\tilde{\varepsilon}_n}(\tilde{u}_n) \geq \sigma_\eta G(u)$ and hence:

$$\liminf_{n \rightarrow \infty} G_{n, \varepsilon_n}(u_n) \geq \sigma_\eta G(u).$$

Limsup inequality: By using a diagonal argument it is enough to establish the limsup inequality for a dense subset of $L^2(D)$ and in particular we consider the set of Lipschitz continuous functions $u : D \rightarrow \mathbb{R}$. That is, we want to show that if $u : D \rightarrow \mathbb{R}$ is a Lipschitz continuous function, then there exists a sequence of functions $\{u_n\}_{n \in \mathbb{N}}$, where $u_n \in L^2(\nu_n)$ and

$$u_n \xrightarrow{TL^2} u \quad \text{as } n \rightarrow \infty, \quad \limsup_{n \rightarrow \infty} G_{n, \varepsilon_n}(u_n) \leq \sigma_\eta G(u).$$

We define u_n to be the restriction of u to the first n data points x_1, \dots, x_n . We note that this operation is well defined due to the fact that u is in particular continuous. It is straightforward to show that given that u is Lipschitz we have $u_n \xrightarrow{TL^2} u$.

Now, consider $\tilde{\varepsilon}_n := \varepsilon_n + 2\|Id - T_n\|_\infty$ and let $\tilde{u}_n = u \circ T_n$. The choice of kernel $\boldsymbol{\eta}$ implies that for every $(x, y) \in D \times D$

$$\boldsymbol{\eta} \left(\frac{|T_n(x) - T_n(y)|}{\varepsilon_n} \right) \leq \boldsymbol{\eta} \left(\frac{|x - y|}{\tilde{\varepsilon}_n} \right).$$

It follows that for all $n \in \mathbb{N}$

$$(3.4) \quad \begin{aligned} & \frac{1}{\tilde{\varepsilon}_n^{d+2}} \int_{D \times D} \boldsymbol{\eta} \left(\frac{|T_n(x) - T_n(y)|}{\varepsilon_n} \right) (\tilde{u}_n(x) - \tilde{u}_n(y))^2 \rho(x) \rho(y) dx dy \\ & \leq \frac{1}{\tilde{\varepsilon}_n^2} \int_{D \times D} \eta_{\tilde{\varepsilon}_n}(x - y) (\tilde{u}_n(x) - \tilde{u}_n(y))^2 \rho(x) \rho(y) dx dy. \end{aligned}$$

Now let A_n and B_n be given by

$$\begin{aligned} A_n &:= \frac{1}{\tilde{\varepsilon}_n^2} \int_{D \times D} \eta_{\tilde{\varepsilon}_n}(x - y) (u(x) - u(y))^2 \rho(x) \rho(y) dx dy \\ B_n &:= \frac{1}{\tilde{\varepsilon}_n^2} \int_{D \times D} \eta_{\tilde{\varepsilon}_n}(x - y) (\tilde{u}_n(x) - \tilde{u}_n(y))^2 \rho(x) \rho(y) dx dy. \end{aligned}$$

Then,

$$(3.5) \quad \begin{aligned} \left(\sqrt{A_n} - \sqrt{B_n} \right)^2 &\leq \frac{1}{\tilde{\varepsilon}_n^2} \int_{D \times D} \eta_{\tilde{\varepsilon}_n}(x - y) (u(x) - \tilde{u}_n(x) + \tilde{u}_n(y) - u(y))^2 \rho(x) \rho(y) dx dy \\ &\leq \frac{4}{\tilde{\varepsilon}_n^2} \int_{D \times D} \eta_{\tilde{\varepsilon}_n}(x - y) (u(x) - \tilde{u}_n(x))^2 \rho(x) \rho(y) dx dy \\ &\leq \frac{4C \text{Lip}(u)^2 \|\rho\|_{L^\infty(D)}^2 \|Id - T_n\|_\infty^2}{\tilde{\varepsilon}_n^2}, \end{aligned}$$

where the first inequality follows using Minkowski's inequality, and where $C = \int_{\mathbb{R}^d} \eta(h) dh$. The last term of the previous expression goes to 0 as $n \rightarrow \infty$, yielding

$$\lim_{n \rightarrow \infty} |\sqrt{A_n} - \sqrt{B_n}| = 0.$$

On the other hand, by (1.26) it follows that A_n is bounded on n and in particular it follows that

$$(3.6) \quad \lim_{n \rightarrow \infty} |A_n - B_n| = 0.$$

We conclude that

$$\begin{aligned} \limsup_{n \rightarrow \infty} G_{n, \varepsilon_n}(u_n) &= \limsup_{n \rightarrow \infty} \frac{1}{\tilde{\varepsilon}_n^{d+2}} \int_{D \times D} \boldsymbol{\eta} \left(\frac{|T_n(x) - T_n(y)|}{\varepsilon_n} \right) (\tilde{u}_n(x) - \tilde{u}_n(y))^2 \rho(x) \rho(y) dx dy \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{\tilde{\varepsilon}_n^2} \int_{D \times D} \eta_{\tilde{\varepsilon}_n}(x - y) (\tilde{u}_n(x) - \tilde{u}_n(y))^2 \rho(x) \rho(y) dx dy \\ &= \limsup_{n \rightarrow \infty} G_{\tilde{\varepsilon}_n}(u) = \sigma_\eta G(u), \end{aligned}$$

where the first equality is obtained from the fact that $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \rightarrow 1$ as $n \rightarrow \infty$, the first inequality is obtained from (3.4), the second equality is obtained from (3.6) and the last equality is obtained from (1.26).

Compactness: Finally, to see that the compactness statement holds suppose that $\{u_n\}_{n \in \mathbb{N}}$ is a sequence with $u_n \in L^2(\nu_n)$ and such that

$$\sup_{n \in \mathbb{N}} \|u_n\|_{L^2(\nu_n)} < \infty, \quad \sup_{n \in \mathbb{N}} G_{n, \varepsilon_n}(u_n) < \infty.$$

Note that in particular $\sup_{n \in \mathbb{N}} \|u_n \circ T_n\|_{L^2(\nu)} < \infty$. We want to show that

$$\sup_{n \in \mathbb{N}} G_{\varepsilon_n}(u_n \circ T_n) < \infty$$

To see this, note that for large enough n , we can set $\tilde{\varepsilon}_n := \varepsilon_n - 2\|Id - T_n\|_\infty$ as in (3.3). Thus, for large enough n :

$$\begin{aligned} & \frac{1}{\varepsilon_n^{d+2}} \int_{D \times D} \boldsymbol{\eta} \left(\frac{|z-y|}{\tilde{\varepsilon}_n} \right) (u_n \circ T_n(z) - u_n \circ T_n(y))^2 \rho(z) \rho(y) dz dy \\ & \leq \frac{1}{\varepsilon_n^{d+2}} \int_{D \times D} \boldsymbol{\eta} \left(\frac{|T_n(z) - T_n(y)|}{\tilde{\varepsilon}_n} \right) (u_n \circ T_n(z) - u_n \circ T_n(y))^2 \rho(z) \rho(y) dz dy \\ & = G_{n, \varepsilon_n}(u_n). \end{aligned}$$

Thus

$$\sup_{n \in \mathbb{N}} \frac{1}{\varepsilon_n^{d+2}} \int_{D \times D} \boldsymbol{\eta} \left(\frac{|z-y|}{\tilde{\varepsilon}_n} \right) (u_n \circ T_n(z) - u_n \circ T_n(y))^2 \rho(z) \rho(y) dz dy < \infty.$$

Finally noting that $\frac{\tilde{\varepsilon}_n}{\varepsilon_n} \rightarrow 1$ as $n \rightarrow \infty$ we deduce that:

$$\sup_{n \in \mathbb{N}} G_{\varepsilon_n}(u_n \circ T_n) < \infty.$$

By Proposition 1.10 we conclude that $\{u_n \circ T_n\}_{n \in \mathbb{N}}$ is relatively compact in $L^2(\nu)$ and hence $\{u_n\}_{n \in \mathbb{N}}$ is relatively compact in TL^2 . \square

Now we prove Theorem 1.2.

3.1. Convergence of Eigenvalues. First of all note that because $\mathcal{L}_{n, \varepsilon_n}$ is self-adjoint with respect to the Euclidean inner product in \mathbb{R}^n , in particular it is also self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_{\nu_n}$ and furthermore, it is positive semi-definite. In particular, we can use the Courant-Fisher maxmini principle to write the eigenvalues $0 = \lambda_1^{(n)} \leq \dots \leq \lambda_n^{(n)}$ of $\mathcal{L}_{n, \varepsilon_n}$ as

$$\lambda_k^{(n)} = \sup_{S \in \Sigma_{k-1}^{(n)}} \min_{\|u\|_{\nu_n}=1, u \in S^\perp} \langle \mathcal{L}_{n, \varepsilon_n} u, u \rangle_{\nu_n},$$

where $\Sigma_{k-1}^{(n)}$ denotes the set of subspaces of \mathbb{R}^n of dimension $k-1$. On the other hand, for any $u_n \in L^2(\nu_n)$, from (1.8) it follows that

$$(3.7) \quad G_{n, \varepsilon_n}(u_n) = \frac{2}{n\varepsilon_n^2} \langle \mathcal{L}_{n, \varepsilon_n} u_n, u_n \rangle_{\nu_n}$$

Therefore,

$$\frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} = \sup_{S \in \Sigma_{k-1}^{(n)}} \min_{\|u\|_{\nu_n}=1, u \in S^\perp} G_{n, \varepsilon_n}(u).$$

Let us first prove the first statement from Theorem 1.2. The proof is by induction on k . For $k=1$, we know that $\lambda_1^{(n)} = 0$ for every n . Also, $\lambda_1 = 0$, so trivially (1) is true when $k=1$. Now, suppose that (1) is true for $i=1, \dots, k-1$. We want to prove that the result holds for k .

Step 1: In this first step we prove that $\sigma_\eta \lambda_k \leq \liminf_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2}$. Let $S \in \Sigma_{k-1}$, we let $\{u_1, \dots, u_{k-1}\}$ be an orthonormal base for S . Then, for every $i = 1, \dots, k-1$, there exists a sequence $\{u_i^n\}_{n \in \mathbb{N}}$ (with $u_i^n \in L^2(\nu_n)$) such that $u_i^n \xrightarrow{TL^2} u_i$ as $n \rightarrow \infty$. The existence of such sequence follows from the limsup inequality of Theorem 1.4. Proposition 2.6 implies that for all $i = 1, \dots, k-1$

$$\lim_{n \rightarrow \infty} \|u_i^n\|_{\nu_n} = \|u_i\|_\rho = 1,$$

and that for $i \neq j$

$$(3.8) \quad \lim_{n \rightarrow \infty} \langle u_i^n, u_j^n \rangle_{\nu_n} = \langle u_i, u_j \rangle_\rho = 0.$$

Thus, for large enough n , the space generated by $\{u_1^n, \dots, u_{k-1}^n\}$ is $k-1$ dimensional. We can use the Gram-Schmidt orthogonalization process, to obtain an orthonormal base $\{\tilde{u}_1^n, \dots, \tilde{u}_{k-1}^n\}$. That is, we define $\tilde{u}_1^n := u_1^n / \|u_1^n\|_{\nu_n}$, and recursively $\tilde{v}_i^n := u_i^n - \sum_{j=1}^{i-1} \langle u_i^n, \tilde{u}_j^n \rangle_{\nu_n} \tilde{u}_j^n$, and $\tilde{u}_i^n := \tilde{v}_i^n / \|\tilde{v}_i^n\|_{\nu_n}$ for $i = 2, \dots, k-1$.

It follows from (3.8) and Proposition 2.6 that $\tilde{u}_i^n \xrightarrow{TL^2} u_i$ as $n \rightarrow \infty$ for every $i = 1, \dots, k-1$. Let $S_n := \text{Span}\{\tilde{u}_1^n, \dots, \tilde{u}_{k-1}^n\}$. We claim that

$$(3.9) \quad \liminf_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} \geq \min_{\|u\|_\rho=1, u \in S^\perp} \sigma_\eta G(u)$$

First, note that if

$$\liminf_{n \rightarrow \infty} \min_{\|u\|_{\nu_n}=1, u \in S_n^\perp} G_{n,\varepsilon_n}(u) = \infty,$$

then in particular

$$\liminf_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} \geq \liminf_{n \rightarrow \infty} \min_{\|u\|_{\nu_n}=1, u \in S_n^\perp} G_{n,\varepsilon_n}(u) = \infty,$$

and in that case (3.9) follows trivially. Let us now assume that $\liminf_{n \rightarrow \infty} \min_{\|u\|_{\nu_n}=1, u \in S_n^\perp} G_{n,\varepsilon_n}(u) < \infty$. Working on a subsequence that we do not relabel, we can assume without the loss of generality that the liminf is actually a limit, that is,

$$\lim_{n \rightarrow \infty} \min_{\|u\|_{\nu_n}=1, u \in S_n^\perp} G_{n,\varepsilon_n}(u) = \liminf_{n \rightarrow \infty} \min_{\|u\|_{\nu_n}=1, u \in S_n^\perp} G_{n,\varepsilon_n}(u) < \infty.$$

Consider now a sequence $\{v_n\}_{n \in \mathbb{N}}$ with $\|v_n\|_{\nu_n} = 1$ and $v_n \in S_n^\perp$ such that

$$\lim_{n \rightarrow \infty} G_{n,\varepsilon_n}(v_n) = \lim_{n \rightarrow \infty} \min_{\|u\|_{\nu_n}=1, u \in S_n^\perp} G_{n,\varepsilon_n}(u) < \infty.$$

Using the compactness from Theorem 1.4 and working on a subsequence that we do not relabel, we may assume that

$$(3.10) \quad v_n \xrightarrow{TL^2} v, \text{ as } n \rightarrow \infty,$$

for some $v \in L^2(D)$. From Proposition 2.6, $\|v\|_\rho = \lim_{n \rightarrow \infty} \|v_n\|_{\nu_n} = 1$ and $\langle v, u_i \rangle_\rho = \lim_{n \rightarrow \infty} \langle v_n, \tilde{u}_i^n \rangle_{\nu_n} = 0$ for every $i = 1, \dots, k-1$. In particular, $\|v\|_\rho = 1$ and $v \in S^\perp$. Moreover, given that $v_n \xrightarrow{TL^2} v$, it

follows from the liminf inequality of Theorem 1.4 that

$$\begin{aligned}
\min_{\|u\|_\rho=1, u \in S^\perp} \sigma_\eta G(u) &\leq \sigma_\eta G(v) \\
&\leq \liminf_{n \rightarrow \infty} G_{n, \varepsilon_n}(v_n) \\
&= \lim_{n \rightarrow \infty} \min_{\|u\|=1, u \in S_n^\perp} G_{n, \varepsilon_n}(u) \\
&\leq \liminf_{n \rightarrow \infty} \sup_{\tilde{S} \in \Sigma_{k-1}^{(n)}} \min_{\|u\|=1, u \in \tilde{S}^\perp} G_{n, \varepsilon_n}(u) \\
&= \liminf_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2}.
\end{aligned}$$

Thus showing (3.9) in all cases. Finally, since $S \in \Sigma_{k-1}$ was arbitrary, taking the supremum over all $S \in \Sigma_{k-1}$ and using the Courant-Fisher maxmini principle we deduce that

$$\sigma_\eta \lambda_k \leq \liminf_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2}.$$

Step 2: Now we prove that $\limsup_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} \leq \lambda_k$. Consider $\{u_1^n, \dots, u_{k-1}^n\}$ an orthonormal set (with respect to $\langle \cdot, \cdot \rangle_{\nu_n}$) with u_i^n an eigenvector of $\mathcal{L}_{n, \varepsilon_n}$ associated to $\lambda_i^{(n)}$ (this is possible because $\mathcal{L}_{n, \varepsilon_n}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_{\nu_n}$). Consider then $S_n^* := \text{Span}\{u_1^n, \dots, u_{k-1}^n\}$. We have:

$$\frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} = \sup_{S \in \Sigma_{k-1}^{(n)}} \min_{\|u\|_{\nu_n}=1, u \in S^\perp} G_{n, \varepsilon_n}(u) = \min_{\|u\|_{\nu_n}=1, u \in S_n^{*\perp}} G_{n, \varepsilon_n}(u).$$

Working along a subsequence that we do not relabel, we can assume without the loss of generality that $\limsup_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} = \lim_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2}$. Note that by the induction hypothesis, for every $i = 1, \dots, k-1$ we have:

$$(3.11) \quad \lim_{n \rightarrow \infty} G_{n, \varepsilon_n}(u_i^n) = \lim_{n \rightarrow \infty} \frac{2\lambda_i^{(n)}}{n\varepsilon_n^2} = \sigma_\eta \lambda_i < \infty.$$

Thanks to this, we can use the compactness from Theorem 1.4 to conclude that for every $i = 1, \dots, k-1$ (working with a subsequence that we do not relabel) :

$$u_i^n \xrightarrow{TL^2} u_i, \text{ as } n \rightarrow \infty,$$

for some $u_i \in L^2(D)$. From Proposition 2.6, $\langle u_i, u_j \rangle_\rho = \lim_{n \rightarrow \infty} \langle u_i^n, u_j^n \rangle_{\nu_n} = 0$ for $i \neq j$ and $\|u_i\|_\rho = \lim_{n \rightarrow \infty} \|u_i^n\|_{\nu_n} = 1$ for every i . Take $S := \text{Span}\{u_1, \dots, u_{k-1}\}$, note that in particular $S \in \Sigma_{k-1}$. Also, take $v \in S^\perp$ with $\|v\|_\rho = 1$ and such that:

$$(3.12) \quad \sigma_\eta G(v) = \min_{\|u\|_\rho=1, u \in S^\perp} \sigma_\eta G(u) \leq \sigma_\eta \lambda_k.$$

The last inequality in the previous expression holds thanks to the Courant-Fisher maxmini principle. By the limsup inequality from Theorem 1.4, we can find $\{v_n\}_{n \in \mathbb{N}}$ with $v_n \xrightarrow{TL^2} v$ as $n \rightarrow \infty$ and such that $\limsup_{n \rightarrow \infty} G_{n, \varepsilon_n}(v_n) \leq \sigma_\eta G(v)$. Let \tilde{v}_n be given by

$$\tilde{v}_n := v_n - \sum_{i=1}^{k-1} \langle v_n, u_i^n \rangle_{\nu_n} u_i^n.$$

Note that $\tilde{v}_n \in S_n^{\perp}$. Also note that from Proposition 2.6, we deduce that $\langle v_n, u_i^n \rangle_{\nu_n} \rightarrow 0$ as $n \rightarrow \infty$ for all $i = 1, \dots, k-1$ and thus $\tilde{v}_n \xrightarrow{TL^2} v$ as $n \rightarrow \infty$. Moreover,

(3.13)

$$\begin{aligned}
G_{n,\varepsilon_n}(\tilde{v}_n) &= \frac{2}{n\varepsilon_n^2} \langle \mathcal{L}_{n,\varepsilon_n} \tilde{v}_n, \tilde{v}_n \rangle \\
&= \frac{2}{n\varepsilon_n^2} \langle \mathcal{L}_{n,\varepsilon_n} v_n, v_n \rangle - \frac{2}{n\varepsilon_n^2} \sum_{i=1}^{k-1} \langle v_n, u_i^n \rangle_{\nu_n} \langle \mathcal{L}_{n,\varepsilon_n} v_n, u_i^n \rangle_{\nu_n} - \frac{2}{n\varepsilon_n^2} \sum_{i=1}^{k-1} \langle v_n, u_i^n \rangle_{\nu_n} \langle \mathcal{L}_{n,\varepsilon_n} u_i^n, \tilde{v}_n \rangle_{\nu_n} \\
&= G_{n,\varepsilon_n}(v_n) - \sum_{i=1}^{k-1} \frac{2\lambda_i^{(n)}}{n\varepsilon_n^2} \langle v_n, u_i^n \rangle_{\nu_n}^2 - \frac{2}{n\varepsilon_n^2} \sum_{i=1}^{k-1} \lambda_i^{(n)} \langle v_n, u_i^n \rangle_{\nu_n} \langle u_i^n, \tilde{v}_n \rangle_{\nu_n} \\
&= G_{n,\varepsilon_n}(v_n) - \sum_{i=1}^{k-1} \frac{2\lambda_i^{(n)}}{n\varepsilon_n^2} \langle v_n, u_i^n \rangle_{\nu_n}^2 \\
&\leq G_{n,\varepsilon_n}(v_n).
\end{aligned}$$

Therefore,

$$(3.14) \quad \limsup_{n \rightarrow \infty} G_{n,\varepsilon_n}(\tilde{v}_n) \leq \limsup_{n \rightarrow \infty} G_{n,\varepsilon_n}(v_n) \leq \sigma_\eta G(v).$$

Since $\tilde{v}_n \xrightarrow{TL^2} v$ and $\|v\|_\rho = 1$, once again from Proposition 2.6 we obtain $\lim_{n \rightarrow \infty} \|\tilde{v}_n\|_{\nu_n} = 1$. In particular we can set $\tilde{u}_n := \frac{\tilde{v}_n}{\|\tilde{v}_n\|_{\nu_n}}$ and use (3.14) together with (3.12) to conclude that:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{2\lambda_k^{(n)}}{n\varepsilon_n^2} &= \lim_{n \rightarrow \infty} \min_{\|u\|_{\nu_n}=1, u \in S_n^{\perp}} G_{n,\varepsilon_n}(u) \\
&\leq \limsup_{n \rightarrow \infty} G_{n,\varepsilon_n}(\tilde{u}_n) \\
&= \limsup_{n \rightarrow \infty} G_{n,\varepsilon_n}(\tilde{v}_n) \\
&\leq \sigma_\eta G(v) \\
&\leq \sigma_\eta \lambda_k,
\end{aligned}$$

which implies the desired result.

3.2. Convergence of Eigenprojections. We prove the second and third part of Theorem 1.2. We recall that the numbers $\bar{\lambda}_1 < \bar{\lambda}_2 < \dots$ denote the distinct eigenvalues of $\mathcal{L}_{n,\varepsilon_n}$. For a given $k \in \mathbb{N}$, we recall that $s(k)$ is the multiplicity of the eigenvalue $\bar{\lambda}_k$ and that $\hat{k} \in \mathbb{N}$ is such that

$$\bar{\lambda}_k = \lambda_{\hat{k}+1} = \dots = \lambda_{\hat{k}+s(k)}.$$

We let E_k be the subspace of $L^2(D)$ of eigenfunctions of \mathcal{L} associated to $\bar{\lambda}_k$, and for large n we let $E_k^{(n)}$ be the subspace of \mathbb{R}^n generated by all the eigenvectors of $\mathcal{L}_{n,\varepsilon_n}$ corresponding to all eigenvalues listed in $\lambda_{\hat{k}+1}^{(n)}, \dots, \lambda_{\hat{k}+s(k)}^{(n)}$. We remark that by the convergence of the eigenvalues proved in Subsection 3.1 we have

$$(3.15) \quad \lim_{n \rightarrow \infty} \dim(E_k^{(n)}) = \dim(E_k) = s(k).$$

We prove simultaneously the second and third statement of Theorem 1.2. The proof is by induction on k .

Base Case: Let $k = 1$. Suppose that $u_n \xrightarrow{TL^2} u$. We need to show that $\text{Proj}_1^{(n)}(u_n) \xrightarrow{TL^2} \text{Proj}_1(u)$. Now, note that since the domain D is connected, the multiplicity of $\bar{\lambda}_1$ is equal to one. In particular,

$\text{Proj}_1(u)$ is the function which is identically equal to

$$\langle u, 1 \rangle_\rho = \int_D u d\nu(x).$$

On the other hand, thanks to (3.15), it follows that for all large enough n , we have $\dim(E_1^{(n)}) = 1$ (note that in particular this means that asymptotically the graphs are connected regardless of what kernel η is being used). Therefore, for large enough n , $\text{Proj}_1^{(n)}(u_n)$ is the function which is identically equal to $\langle u_n, 1 \rangle_{\nu_n}$. Proposition 2.6 implies that $\lim_{n \rightarrow \infty} \langle u_n, 1 \rangle_{\nu_n} = \langle u, 1 \rangle_\rho$ and thus $\text{Proj}_1^{(n)}(u_n) \xrightarrow{TL^2} \text{Proj}_1(u)$ as desired. The second statement of Theorem 1.2 is trivial in this case since for large enough n , the only two eigenvectors of $\mathcal{L}_{n, \varepsilon_n}$ with eigenvalue $\lambda_1^{(n)} = 0$ and with $\|\cdot\|_{\nu_n}$ -norm equal to one is the function which is identically equal to one or the function that is identically equal to -1 .

Inductive Step: Now, suppose that the second and third statements of Theorem 1.2 are true for $1, \dots, k-1$. We want to prove the result for k . Let $j \in \{\hat{k} + 1, \dots, \hat{k} + s(k)\}$. We start by proving the second statement of the theorem. Consider $\{u_j^n\}_{n \in \mathbb{N}}$ as in the statement. From (3.7) it follows that $G_{n, \varepsilon_n}(u_j^n) = \frac{2\lambda_j^{(n)}}{n\varepsilon_n^2}$. Now, from Subsection 3.1, we know that

$$\lim_{n \rightarrow \infty} \frac{2\lambda_j^{(n)}}{n\varepsilon_n^2} = \sigma_\eta \lambda_j$$

and so in particular we have:

$$\sup_{n \in \mathbb{N}} G_{n, \varepsilon_n}(u_j^n) < \infty.$$

Since the norms of the u_j^n are equal to one, the compactness statement from Theorem 1.4, implies that $\{u_j^n\}_{n \in \mathbb{N}}$ is pre-compact. We have to prove now that every cluster point of $\{u_j^n\}_{n \in \mathbb{N}}$ is an eigenfunction of \mathcal{L} with eigenvalue λ_j . So without the loss of generality let us assume that $u_j^n \xrightarrow{TL^2} u_j$ for some u_j . Our goal is to show that u_j is an eigenfunction of \mathcal{L} with eigenvalue λ_j .

By the induction hypothesis, we have $\text{Proj}_i^{(n)}(u_j^n) \xrightarrow{TL^2} \text{Proj}_i(u_j)$ for every $i = 1, \dots, k-1$. On the other hand, since $\text{Proj}_i^{(n)}(u_j^n) = 0$ for every $n \in \mathbb{N}$ and for every $i = 1, \dots, k-1$, we conclude that $\text{Proj}_i(u_j) = 0$ for all $i = 1, \dots, k-1$. A straightforward computation as in the proof of Proposition 2.14 shows that:

$$(3.16) \quad G(u_j) = \sum_{i=k}^{\infty} \bar{\lambda}_i \|\text{Proj}_i(u_j)\|_\rho^2 \geq \bar{\lambda}_k \sum_{i=k}^{\infty} \|\text{Proj}_i(u_j)\|_\rho^2 = \bar{\lambda}_k \|u_j\|_\rho^2.$$

In addition, since $\|u_j^n\|_{\nu_n} = 1$ for all n , we deduce from Proposition 2.6 that $\|u_j\|_\rho = 1$. Thus,

$$G(u_j) \geq \bar{\lambda}_k.$$

On the other hand, the liminf inequality from Theorem 1.4 implies that:

$$\sigma_\eta \bar{\lambda}_k = \sigma_\eta \lambda_j = \lim_{n \rightarrow \infty} \frac{2\lambda_j^{(n)}}{n\varepsilon_n^2} = \lim_{n \rightarrow \infty} G_{n, \varepsilon_n}(u_j^n) \geq \sigma_\eta G(u_j) \geq \sigma_\eta \bar{\lambda}_k.$$

Therefore, $G(u_j) = \bar{\lambda}_k$ and from (3.16) we conclude that $\|\text{Proj}_i(u_j)\|_\rho = 0$ for all $i \neq k$. Thus, u_j is an eigenfunction of \mathcal{L} with corresponding eigenvalue $\lambda_j (= \bar{\lambda}_k)$.

Now we prove the third statement from Theorem 1.2. Suppose that $u^n \xrightarrow{TL^2} u$. We want to show that $\text{Proj}_k^{(n)}(u^n) \xrightarrow{TL^2} \text{Proj}_k(u)$. To achieve this we prove that for a given sequence of natural

numbers there exists a further subsequence for which the convergence holds. We do not relabel subsequences to avoid cumbersome notation.

From (3.15) it follows that for large enough n , $\dim(E_k^{(n)}) = s(k)$. Hence, for large enough n , we can consider $\{u_1^n, \dots, u_{s(k)}^n\}$ an orthonormal basis (with respect to the inner product $\langle \cdot, \cdot \rangle_{\nu_n}$) for $E_k^{(n)}$, where u_j^n is an eigenvector of $\mathcal{L}_{n, \varepsilon_n}$ with corresponding eigenvalue $\lambda_{\hat{k}+j}^{(n)}$. Now, by the first part of the proof, for every $j = 1, \dots, s(k)$, the sequence $\{u_j^n\}_{n \in \mathbb{N}}$ is pre-compact in TL^2 . Therefore, passing to a subsequence that we do not relabel we can assume that for every $j = 1, \dots, s(k)$ we have:

$$(3.17) \quad u_j^n \xrightarrow{TL^2} u_j, \text{ as } n \rightarrow \infty$$

for some $u_j \in L^2(D)$. From (2.6), the u_j satisfy $\|u_j\|_\rho = 1$ for every j and $\langle u_i, u_j \rangle_\rho = 0$ for $i \neq j$. In other words, $\{u_1, \dots, u_{s(k)}\}$ is an orthonormal set in $L^2(D)$ (with respect to $\langle \cdot, \cdot \rangle_\rho$). Furthermore, $u_j \in E_k$ for all j by the first part of the proof. In other words, $\{u_1, \dots, u_{s(k)}\}$ is an orthonormal basis for E_k and in particular

$$\text{Proj}_k(u) = \sum_{j=1}^{s(k)} \langle u, u_j \rangle_\rho u_j.$$

On the other hand, for large enough n , we have

$$\text{Proj}_k^{(n)}(u^n) = \sum_{j=1}^{s(k)} \langle u^n, u_j^n \rangle_{\nu_n} u_j^n.$$

Finally, the fact that $u^n \xrightarrow{TL^2} u$ and (3.17) combined with Proposition 2.6 imply that

$$\text{Proj}_k^{(n)}(u^n) = \sum_{j=1}^{s(k)} \langle u^n, u_j^n \rangle_{\nu_n} u_j^n \xrightarrow{TL^2} \sum_{j=1}^{s(k)} \langle u, u_j \rangle_\rho u_j = \text{Proj}_k(u).$$

3.3. Consistency of spectral clustering.

Here we prove statement 4. of Theorem 1.2.

The procedure in Algorithm 1, can be reformulated as follows. Let $\mu_n = (u_1^n, \dots, u_k^n)_\# \nu_n$, where u_1^n, \dots, u_k^n are orthonormal eigenvectors of $\mathcal{L}_{n, \varepsilon_n}$ corresponding to eigenvalues $\lambda_1^{(n)}, \dots, \lambda_k^{(n)}$, respectively. Consider the functional $F_{\mu_n, k}$. Let \mathbf{z}_n be its minimizer, and let $\tilde{G}_1^n, \dots, \tilde{G}_k^n$ be corresponding clusters. The clusters G_1, \dots, G_k of Algorithm 1 are defined by $G_i = (u_1^n, \dots, u_k^n)^{-1}(\tilde{G}_i^n)$.

By Theorem 1.8 the sequence \mathbf{z}_n is precompact. By Corollary 1.9 the sequence of measures $\mu_n^i = \mu_n \llcorner_{\tilde{G}_i^n}$ is precompact for all $i = 1, \dots, k$. Consider a subsequence along which μ_n^i converges for every $i = 1, \dots, k$, and denote the limit by μ^i . Since $z_n^i = \int y d\mu_n^i(y)$ it follows that z_n^i converge as $n \rightarrow \infty$, along the same subsequence. By statement 2. of Theorem 1.2 along a further subsequence (ν_n, u_i^n) converge to (ν, u_i) in TL^2 sense for all $i = 1, \dots, k$ as $n \rightarrow \infty$. Furthermore from the definition of TL^2 convergence follows that measures μ_n converge in the Wasserstein sense to $\mu := (u_1, \dots, u_k)_\# \nu$. Combined with convergence of μ_n^i to μ^i implies, via Lemma 2.5, that $(\mu_n, \chi_{\tilde{G}_i^n})$ converge in TL^2 topology to $(\mu, \chi_{\tilde{G}_i})$. Consequently, by Lemma 2.7, $(\nu_n, \chi_{\tilde{G}_i^n} \circ (u_1^n, \dots, u_k^n))$ converge to $(\nu, \chi_{\tilde{G}_i} \circ (u_1, \dots, u_k))$ in TL^2 topology. Noting that $\chi_{G_i^n} = \chi_{\tilde{G}_i^n} \circ (u_1^n, \dots, u_k^n)$ and $\chi_{G_i} = \chi_{\tilde{G}_i} \circ (u_1, \dots, u_k)$ implies that $\nu_n \llcorner_{G_i^n}$ converges weakly to $\nu \llcorner_{G_i}$ as desired.

4. CONVERGENCE OF THE SPECTRA OF NORMALIZED GRAPH LAPLACIANS

We start by proving Theorem 1.6. Recall that for given $u_n \in L^2(\nu_n)$

$$\bar{G}_{n,\varepsilon_n}(u_n) = \frac{1}{n\varepsilon_n^2} \sum_{i,j} W_{i,j} \left(\frac{u_n(x_i)}{\sqrt{\mathcal{D}_{ii}}} - \frac{u_n(x_j)}{\sqrt{\mathcal{D}_{jj}}} \right)^2,$$

where $W_{ij} = \eta_{\varepsilon_n}(x_i - x_j)$ and $\mathcal{D}_{ii} = \sum_{k=1}^n \eta_{\varepsilon_n}(x_i - x_k)$. With a slight abuse of notation we set $\mathcal{D}(x_i) := \mathcal{D}_{ii}$.

For $u_n \in L^2(\nu_n)$, define $\bar{u}_n \in L^2(\nu_n)$ by

$$(4.1) \quad \bar{u}_n(x_i) := \frac{u_n(x_i)}{\sqrt{\mathcal{D}(x_i)/n}}, \quad i \in \{1, \dots, n\}.$$

From the definition of G_{n,ε_n} and $\bar{G}_{n,\varepsilon_n}$, it follows that $\bar{G}_{n,\varepsilon_n}(u_n) = G_{n,\varepsilon_n}(\bar{u}_n)$. Similarly, for every $u \in L^2(D)$ it is true that $\bar{G}(u) = G(\frac{u}{\sqrt{\rho}})$. To prove Theorem 1.6 we use the following lemma.

Lemma 4.1. *Assume that the sequence $\{\varepsilon_n\}_{n \in \mathbb{N}}$ satisfies (1.16). With probability one the following statement holds: a sequence $\{u_n\}_{n \in \mathbb{N}}$, with $u_n \in L^2(\nu_n)$, converges to $u \in L^2(\rho)$ in the TL^2 -metric if and only if $\bar{u}_n \xrightarrow{TL^2} \frac{u}{\sqrt{\beta_{\eta\rho}}}$, where \bar{u}_n is defined in (4.1) and where β_{η} is defined in (1.21).*

Proof. We prove that $u_n \xrightarrow{TL^2} u$ implies $\bar{u}_n \xrightarrow{TL^2} \frac{u}{\sqrt{\beta_{\eta\rho}}}$; the converse implication is obtained similarly.

Let $\{T_n\}_{n \in \mathbb{N}}$ be the transportation maps from Proposition 1.12, which we know exist with probability one. Using the change of variables (2.3) we obtain

$$\frac{\mathcal{D}(X_i)}{n} = \int_D \eta_{\varepsilon_n}(x_i - T_n(y)) \rho(y) dy.$$

If $u_n \xrightarrow{TL^2} u$, in particular from Proposition 2.1 we have $u_n \circ T_n \xrightarrow{L^2(\rho)} u$. By Proposition 2.1, in order to prove that $\bar{u}_n \xrightarrow{TL^2} \frac{u}{\sqrt{\beta_{\eta\rho}}}$, it is enough to prove that $\bar{u}_n \circ T_n \xrightarrow{L^2(\rho)} \frac{u}{\sqrt{\beta_{\eta\rho}}}$, which in turn is equivalent to $\bar{u}_n \circ T_n \rightarrow_{L^2(D)} \frac{u}{\sqrt{\beta_{\eta\rho}}}$ due to the fact that ρ satisfies (1.15). To achieve this, we first find an L^∞ -control on the terms $\frac{1}{\sqrt{\mathcal{D} \circ T_n/n}}$ and then prove that $\frac{1}{\sqrt{\mathcal{D} \circ T_n/n}}$ converges point-wise to $\frac{1}{\sqrt{\beta_{\eta\rho}}}$. Since $u_n \circ T_n \xrightarrow{L^2(D)} u$ this is enough to obtain the desired result. For that purpose, we fix an arbitrary $\alpha > 0$ and define $\underline{\eta}^\alpha : [0, \infty) \rightarrow [0, \infty)$ and $\bar{\eta}^\alpha : [0, \infty) \rightarrow [0, \infty)$ to be

$$(4.2) \quad \underline{\eta}^\alpha(t) := \begin{cases} \eta(t), & \text{if } t > 2\alpha \\ \eta(2\alpha), & \text{if } t \leq 2\alpha, \end{cases}$$

and

$$(4.3) \quad \bar{\eta}^\alpha(t) := \begin{cases} \eta(t), & \text{if } t > 2\alpha \\ \eta(0), & \text{if } t \leq 2\alpha, \end{cases}$$

where we recall that η is the radial profile of the kernel η . We let $\underline{\eta}^\alpha$ and $\bar{\eta}^\alpha$ be the isotropic kernels whose radial profiles are $\underline{\eta}^\alpha$ and $\bar{\eta}^\alpha$ respectively. Note that thanks to assumption (K2) on η , we have $\underline{\eta}^\alpha \leq \eta \leq \bar{\eta}^\alpha$. Set

$$\begin{aligned} \hat{\varepsilon}_n &:= \varepsilon_n - \frac{2\|Id - T_n\|_\infty}{\alpha}, \\ \tilde{\varepsilon}_n &:= \varepsilon_n + \frac{2\|Id - T_n\|_\infty}{\alpha}. \end{aligned}$$

Note that thanks to the assumptions on ε_n and the properties of the maps T_n , for large enough n , $\hat{\varepsilon}_n > 0$, $\frac{\hat{\varepsilon}_n}{\varepsilon_n} \rightarrow 1$ and $\frac{\tilde{\varepsilon}_n}{\varepsilon_n} \rightarrow 1$ as $n \rightarrow \infty$. In addition, from assumption (K2) on η and the definitions of $\underline{\eta}^\alpha$, $\overline{\eta}^\alpha$, $\hat{\varepsilon}_n$ and $\tilde{\varepsilon}_n$, it is straightforward to check that for large enough n and for Lebesgue almost every $x, y \in D$,

$$\eta\left(\frac{T_n(x) - T_n(y)}{\varepsilon_n}\right) \geq \underline{\eta}^\alpha\left(\frac{x - y}{\hat{\varepsilon}_n}\right),$$

and

$$\eta\left(\frac{T_n(x) - T_n(y)}{\varepsilon_n}\right) \leq \overline{\eta}^\alpha\left(\frac{x - y}{\tilde{\varepsilon}_n}\right).$$

From these inequalities, we conclude that for large enough n and Lebesgue almost every $x \in D$

$$(4.4) \quad \int_D \eta_{\varepsilon_n}(T_n(x) - T_n(y))\rho(y)dy \geq \left(\frac{\hat{\varepsilon}_n}{\varepsilon_n}\right)^d \int_D \underline{\eta}_{\hat{\varepsilon}_n}^\alpha(x - y)\rho(y)dy$$

and

$$(4.5) \quad \int_D \eta_{\varepsilon_n}(T_n(x) - T_n(y))\rho(y)dy \leq \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^d \int_D \overline{\eta}_{\tilde{\varepsilon}_n}^\alpha(x - y)\rho(y)dy.$$

Given that D is assumed to be a bounded open set with Lipschitz boundary, it is straightforward to check that exists a ball $B(0, \theta)$, a cone C with nonempty interior and a family of rotations $\{R_x\}_{x \in D}$ with the property that for every $x \in D$ it is true that $x + R_x(B(0, \theta) \cap C) \subseteq D$. For large enough n (so that $1 > \hat{\varepsilon}_n > 0$), and for almost every $x \in D$ we have:

$$\begin{aligned} \int_D \underline{\eta}_{\hat{\varepsilon}_n}^\alpha(x - y)\rho(y)dy &\geq m \int_D \underline{\eta}_{\hat{\varepsilon}_n}^\alpha(x - y)dy = m \int_{x + \hat{\varepsilon}_n h \in D} \underline{\eta}^\alpha(h)dh \\ &\geq m \int_{R_x(B(0, \theta) \cap C)} \underline{\eta}^\alpha(h)dh = m \int_{B(0, \theta) \cap C} \underline{\eta}^\alpha(h)dh > 0, \end{aligned}$$

where in the first inequality we used assumption (1.15) on ρ , and we used the change of variables $h = \frac{x - y}{\hat{\varepsilon}_n}$ to deduce the first equality; to obtain the last equality we used the fact that $\underline{\eta}^\alpha$ is radially symmetric. From the previous chain of inequalities and from (4.4) we conclude that for large enough n and for almost every $x \in D$ we have

$$\int_D \eta_{\varepsilon_n}(T_n(x) - T_n(y))\rho(y)dy \geq b > 0$$

for some positive constant b . From the previous inequality we obtain the desired L^∞ -control on the terms $\frac{1}{\sqrt{\mathcal{D} \circ T_n/n}}$. It remains to show that for almost every $x \in D$,

$$(4.6) \quad \lim_{n \rightarrow \infty} \int_D \eta_{\varepsilon_n}(T_n(x) - T_n(y))\rho(y)dy = \beta_\eta \rho(x).$$

For this purpose, we use the continuity of ρ to deduce that for every $x \in D$,

$$(4.7) \quad \lim_{n \rightarrow \infty} \left| \underline{\beta}_\alpha \rho(x) - \int_D \underline{\eta}_{\hat{\varepsilon}_n}^\alpha(x - y)\rho(y)dy \right| = 0,$$

where $\underline{\beta}_\alpha = \int_{\mathbb{R}^d} \underline{\eta}^\alpha(h)dh$. Similarly, for every $x \in D$,

$$(4.8) \quad \lim_{n \rightarrow \infty} \left| \overline{\beta}_\alpha \rho(x) - \int_D \overline{\eta}_{\tilde{\varepsilon}_n}^\alpha(x - y)\rho(y)dy \right| = 0,$$

where $\bar{\beta}_\alpha = \int_{\mathbb{R}^d} \bar{\eta}^\alpha(h) dh$. From (4.4), we deduce that for large enough n , and for almost every $x \in D$,

$$\begin{aligned} \beta_\eta \rho(x) - \int_D \eta_{\varepsilon_n}(T_n(x) - T_n(y)) \rho(y) dy &\leq \beta_\eta \rho(x) - \left(\frac{\hat{\varepsilon}_n}{\varepsilon_n}\right)^d \int_D \eta_{\hat{\varepsilon}_n}^\alpha(x-y) \rho(y) dy \\ &\leq \left(\frac{\hat{\varepsilon}_n}{\varepsilon_n}\right)^d \left(\beta_\eta \rho(x) - \underline{\beta}_\alpha \rho(x) + \underline{\beta}_\alpha \rho(x) - \int_D \eta_{\hat{\varepsilon}_n}^\alpha(x-y) \rho(y) dy \right) \\ &\quad + \left(1 - \left(\frac{\hat{\varepsilon}_n}{\varepsilon_n}\right)^d\right) \beta_\eta \rho(x). \end{aligned}$$

Analogously, from (4.5), for almost every $x \in D$,

$$\begin{aligned} \int_D \eta_{\varepsilon_n}(T_n(x) - T_n(y)) \rho(y) dy - \beta_\eta \rho(x) &\leq \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^d \left(\bar{\beta}_\alpha \rho(x) - \beta_\eta \rho(x) + \int_D \bar{\eta}_{\tilde{\varepsilon}_n}^\alpha(x-y) \rho(y) dy - \bar{\beta}_\alpha \rho(x) \right) \\ &\quad + \left(\left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^d - 1 \right) \beta_\eta \rho(x). \end{aligned}$$

From these previous inequalities, (4.7) and (4.8) we conclude that for almost every $x \in D$,

$$\limsup_{n \rightarrow \infty} \left| \beta_\eta \rho(x) - \int_D \eta_{\varepsilon_n}(T_n(x) - T_n(y)) \rho(y) dy \right| \leq \rho(x) (\bar{\beta}_\alpha - \underline{\beta}_\alpha).$$

Finally, given that α was arbitrary we can take $\alpha \rightarrow 0$ in the previous expression to deduce that the left hand side of the previous expression is actually equal to zero. This establishes (4.6) and thus the desired result. \square

The proof of Theorem 1.6 is now straightforward.

Proof of Theorem 1.6. Liminf inequality: Let $u \in L^2(D)$ and suppose that $\{u_n\}_{n \in \mathbb{N}}$, $u_n \in L^2(\nu_n)$, is such that $u_n \xrightarrow{TL^2} u$. From Lemma 4.1, we know that $\bar{u}_n \xrightarrow{TL^2} \frac{u}{\sqrt{\beta_\eta \rho}}$, where \bar{u}_n was defined in (4.1). From Theorem 1.4 and the discussion at the beginning of this section, we obtain

$$\liminf_{n \rightarrow \infty} \bar{G}_{n, \varepsilon_n}(u_n) = \liminf_{n \rightarrow \infty} G_{n, \varepsilon_n}(\bar{u}_n) \geq \sigma_\eta G\left(\frac{u}{\sqrt{\beta_\eta \rho}}\right) = \frac{\sigma_\eta}{\beta_\eta} \bar{G}(u),$$

where the inequality is obtained using the liminf inequality from Theorem 1.4.

Limsup inequality: Let $u \in L^2(D)$. Since ρ is bounded below by a positive constant, $\frac{u}{\sqrt{\beta_\eta \rho}}$ belongs to $L^2(D)$ as well. From the limsup inequality in Theorem 1.4, there exists a sequence $\{v_n\}_{n \in \mathbb{N}}$, $v_n \in L^2(\nu_n)$, with $v_n \xrightarrow{TL^2} \frac{u}{\sqrt{\beta_\eta \rho}}$ and such that

$$\limsup_{n \rightarrow \infty} G_{n, \varepsilon_n}(v_n) \leq \sigma_\eta G\left(\frac{u}{\sqrt{\beta_\eta \rho}}\right) = \frac{\sigma_\eta}{\beta_\eta} \bar{G}(u).$$

Let us consider the function $u_n \in L^2(\nu_n)$ given by $u_n(x_i) := v_n(x_i) \sqrt{\mathcal{D}(x_i)/n}$ for $i = 1, \dots, n$. Lemma 4.1 implies that $u_n \xrightarrow{TL^2} u$. From the discussion at the beginning of this section we obtain

$$\limsup_{n \rightarrow \infty} \bar{G}_{n, \varepsilon_n}(u_n) = \limsup_{n \rightarrow \infty} G_{n, \varepsilon_n}(v_n) \leq \frac{\sigma_\eta}{\beta_\eta} \bar{G}(u).$$

Compactness: Suppose that $\{u_n\}_{n \in \mathbb{N}}$, $u_n \in L^2(\nu_n)$, is such that

$$\sup_{n \in \mathbb{N}} \|u_n\|_{\nu_n} < \infty, \quad \sup_{n \in \mathbb{N}} \bar{G}_{n, \varepsilon_n}(u_n) < \infty.$$

From the discussion at the beginning of the section, we deduce that $\sup_{n \in \mathbb{N}} G_{n, \varepsilon_n}(\bar{u}_n) < \infty$. Also, from the proof of Lemma 4.1, the terms $\frac{1}{\sqrt{\mathcal{D} \circ T_n/n}}$ are uniformly bounded in L^∞ . This implies that $\sup_{n \in \mathbb{N}} \|\bar{u}_n\|_{L^2(\nu_n)} < \infty$ as well. Hence, we can apply the compactness property from Theorem 1.4 to conclude that $\{\bar{u}_n\}_{n \in \mathbb{N}}$ is precompact in TL^2 . Using Lemma 4.1, this implies that $\{u_n\}_{n \in \mathbb{N}}$ is precompact in TL^2 as well. \square

Proof of Theorem 1.5. Using Theorem 1.6, similar arguments to the ones used in the proof of Theorem 1.2 can be used to establish statements 1., 2., and 3. of Theorem 1.5.

The proof of statement 4. (consistency of spectral clustering) of Theorem 1.5 is analogous to the proof of the statement 4. of Theorem 1.2 which is given in Subsection 3.3. The reason that the normalization step does not create new difficulties is the following: since the eigenvectors $\mathbf{u}^n := (u_1^n, \dots, u_k^n)$ of $\mathcal{N}_{n, \varepsilon_n}^{sym}$ converge in TL^2 to eigenfunctions $\mathbf{u} = (u_1, \dots, u_k)$ of \mathcal{N}^{sym} along a subsequence, it can be shown that the normalized vectors $\mathbf{u}^n / \|\mathbf{u}^n\|$ converge to $\mathbf{u} / \|\mathbf{u}\|$ in TL^2 provided that the set of $x \in D$ for which $\mathbf{u} = 0$ is of ν -measure zero.

In fact, assuming that $\nu(\{x \in D : \mathbf{u}(x) = 0\}) = 0$ let us show the TL^2 convergence. From the assumption on the set of zeroes of \mathbf{u} follows that $\lim_{H \rightarrow 0^+} \nu(\{\|\mathbf{u}(x)\| < H\}) = 0$. Let $U_H = \{(x, y) \in D \times D : \|\mathbf{u}(x)\| < H\}$. Given $n \in \mathbb{N}$ let $\pi_n \in \Pi(\nu_n, \nu)$ be such that

$$\iint |x - y|^2 + \|\mathbf{u}^n(x) - \mathbf{u}(y)\|^2 d\pi_n(x, y) \leq 2d_{TL^2}^2(\mathbf{u}^n, \mathbf{u}).$$

Then for any $H > 0$

$$\begin{aligned} d_{TL^2} \left(\frac{\mathbf{u}^n}{\|\mathbf{u}^n\|}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \right) &\leq \iint |x - y|^2 d\pi_n(x, y) + \iint_{U_H} 2^2 d\pi_n(x, y) \\ &\quad + \iint_{D \times D \setminus U_H} \frac{\|\|\mathbf{u}^n(y)\|\mathbf{u}(x) \pm \mathbf{u}^n(y)\mathbf{u}^n(y) - \|\mathbf{u}(x)\|\mathbf{u}^n(y)\|^2}{\|\mathbf{u}(x)\|^2 \|\mathbf{u}^n(y)\|^2} d\pi_n(x, y) \\ &\leq 4d_{TL^2}^2(\mathbf{u}^n, \mathbf{u}) + o(H) + \frac{16}{H^2} d_{TL^2}^2(\mathbf{u}^n, \mathbf{u}). \end{aligned}$$

The right hand side can be made arbitrarily small by first picking H small enough and then n large enough along the subsequence where \mathbf{u}^n converges to \mathbf{u} . The convergence of normalized eigenvector k -tuples follows.

To show that $\nu(\{x \in D : \mathbf{u}(x) = 0\}) = 0$ it suffices to show that the set of $x \in D$ for which $u^1(x) = 0$ has zero Lebesgue measure. To show this, we need the extra technical condition that $\rho \in C^1(D)$. Because of it and the fact that ρ is bounded away from zero, it follows from the regularity theory of elliptic PDEs, that the function $w_1 := \frac{u_1}{\sqrt{\rho}}$ is of class $C^{1, \alpha}(D)$ (for $\alpha \in (0, 1)$) and is a solution of

$$-\operatorname{div}(\rho^2 \nabla w_1) - \tau_1 \rho^2 w_1 = 0, \quad \forall x \in D.$$

Consider the sets

$$N(w_1) := \{x \in D : w_1(x) = 0\} \quad S(w_1) := \{x \in N(w_1) : \nabla w_1(x) = 0\}.$$

By the implicit function theorem, it follows that $N(w_1) \setminus S(w_1)$ can be covered by at most countable $d-1$ dimensional manifolds and hence it follows that the Lebesgue measure of $N(w_1) \setminus S(w_1)$ is equal to zero. On the other hand, it follows from the results in [21], that $S(w_1)$ is $(d-2)$ -rectifiable, which in particular implies that the Lebesgue measure of $S(w_1)$ is equal to zero. Since $u_1^{-1}(\{0\}) = N(w_1)$, we conclude that the set in which u_1 is equal to zero has zero Lebesgue measure. \square

Proof of Corollary 1.7. Given a sequence $\{u_k^n\}_{n \in \mathbb{N}}$, as in the statement of the corollary, we define

$$w_k^n := \mathcal{D}^{1/2} u_k^n.$$

From (1.4) it follows that w_k^n is an eigenvector of $\mathcal{N}_{n,\varepsilon_n}^{sym}$. We consider a rescaled version of the vectors w_k^n , by setting

$$\tilde{w}_k^n := \frac{w_k^n}{\sqrt{n}} = \frac{1}{\sqrt{n}} \mathcal{D}^{1/2} u_k^n.$$

From the proof of Lemma 4.1, it follows that

$$\sup_{n \in \mathbb{N}} \|\tilde{w}_k^n\|_{\nu_n} < \infty.$$

Thus, from Theorem 1.5, up to subsequence,

$$\tilde{w}_k^n \xrightarrow{TL^2} w,$$

for some $w \in L^2(D)$ which is an eigenfunction of \mathcal{N}^{sym} with eigenvalue τ_k . Hence, up to subsequence, from Lemma 4.1 it follows that

$$u_k^n \xrightarrow{TL^2} \frac{w}{\sqrt{\beta_\eta \rho}}.$$

By discussion of Subsection 2.4, it follows that $\frac{w}{\sqrt{\beta_\eta \rho}}$ is an eigenfunction of \mathcal{N}^{rw} with eigenvalue τ_k .

The proof of convergence of clusters is the same as given in the proof of Theorem 1.2 presented in Subsection 3.3. \square

Acknowledgments. DS is grateful to NSF (grant DMS-1211760) for its support. The authors are thankful to Moritz Gerlach, Matthias Hein, Thomas Laurent, James von Brecht, and Ulrike von Luxburg, for enlightening conversations. The authors would like to thank the Center for Nonlinear Analysis of the Carnegie Mellon University for its support. Furthermore they are thankful to ICERM, where part of the work was done, for hospitality.

REFERENCES

- [1] G. ALBERTI AND G. BELLETTINI, *A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies*, European J. Appl. Math., 9 (1998), pp. 261–284.
- [2] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics, Birkhäuser Basel, 2008.
- [3] E. ARIAS-CASTRO AND B. PELLETIER, *On the convergence of maximum variance unfolding*, Journal of Machine Learning Research, 14 (2013), pp. 1747–1770.
- [4] E. ARIAS-CASTRO, B. PELLETIER, AND P. PUDLO, *The normalized graph cut and Cheeger constant: from discrete to continuous*, Adv. in Appl. Probab., 44 (2012), pp. 907–937.
- [5] S. ARORA, S. RAO, AND U. VAZIRANI, *Expander flows, geometric embeddings and graph partitioning*, Journal of the ACM (JACM), 56 (2009), p. 5.
- [6] H. ATTOUCH, G. BUTTAZZO, AND G. MICHAILLE, *Variational analysis in Sobolev and BV spaces*, vol. 6 of MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2006. Applications to PDEs and optimization.
- [7] M. BELKIN AND P. NIYOGI, *Convergence of Laplacian eigenmaps*, Advances in Neural Information Processing Systems (NIPS), 19 (2007), p. 129.
- [8] ———, *Towards a theoretical foundation for Laplacian-based manifold methods*, J. Comput. System Sci., 74 (2008), pp. 1289–1308.
- [9] A. BRAIDES, *Gamma-Convergence for Beginners*, Oxford Lecture Series in Mathematics and Its Applications Series, Oxford University Press, Incorporated, 2002.
- [10] D. BURAGO, S. IVANOV, AND Y. KURYLEV, *A graph discretization of the Laplace-Beltrami operator*, J. Spectr. Theory, 4 (2014), pp. 675–714.
- [11] R. R. COIFMAN AND S. LAFON, *Diffusion maps*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 5–30.
- [12] G. DAL MASO, *An Introduction to Γ -convergence*, Springer, 1993.
- [13] L. EVANS, *Partial Differential Equations*, Graduate studies in mathematics, American Mathematical Society, 2010.
- [14] S. FORTUNATO, *Community detection in graphs*, Physics Reports, 486 (2010), pp. 75–174.
- [15] N. GARCÍA TRILLOS AND D. SLEPČEV, *Continuum limit of total variation on point clouds*. Preprint, 2014.

- [16] ———, *On the rate of convergence of empirical measures in ∞ -transportation distance*, Canadian J. Math., (2015). online first.
- [17] N. GARCÍA TRILLOS, D. SLEPČEV, J. VON BRECHT, T. LAURENT, AND X. BRESSON, *Consistency of Cheeger and ratio graph cuts*. Preprint, 2014.
- [18] E. GINÉ AND V. KOLTCHINSKII, *Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results*, in High dimensional probability, vol. 51 of IMS Lecture Notes Monogr. Ser., Inst. Math. Statist., Beachwood, OH, 2006, pp. 238–259.
- [19] A. GOEL, S. RAI, AND B. KRISHNAMACHARI, *Sharp thresholds for monotone properties in random geometric graphs*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing, New York, 2004, ACM, pp. 580–586 (electronic).
- [20] P. GUPTA AND P. R. KUMAR, *Critical power for asymptotic connectivity in wireless networks*, in Stochastic analysis, control, optimization and applications, Systems Control Found. Appl., Birkhäuser Boston, Boston, MA, 1999, pp. 547–566.
- [21] Q. HAN, *Singular sets of solutions to elliptic equations*, Indiana Univ. Math. J., 43 (1994), pp. 983–1002.
- [22] J. HARTIGAN, *Consistency of single linkage for high density clusters*, J. Amer. Statist. Assoc., 76 (1981), pp. 388–394.
- [23] M. HEIN, J.-Y. AUDIBERT, AND U. VON LUXBURG, *From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians*, in Learning theory, Springer, 2005, pp. 470–485.
- [24] M. HEIN AND S. SETZER, *Beyond Spectral Clustering - Tight Relaxations of Balanced Graph Cuts*, in Advances in Neural Information Processing Systems (NIPS), 2011.
- [25] S. LAFON AND A. LEE, *Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28 (2006), pp. 1393–1403.
- [26] T. LEIGHTON AND P. SHOR, *Tight bounds for minimax grid matching with applications to the average case analysis of algorithms*, Combinatorica, 9 (1989), pp. 161–187.
- [27] G. LEONI, *A first course in Sobolev spaces*, vol. 105 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2009.
- [28] B. NADLER, S. LAFON, R. COIFMAN, AND I. KEVREKIDIS, *Diffusion Maps - a Probabilistic Interpretation for Spectral Embedding and Clustering Algorithms*, in Principal Manifolds for Data Visualization and Dimension Reduction, A. Gorban, B. Kégl, D. Wunsch, and A. Zinovyev, eds., vol. 58 of Lecture Notes in Computational Science and Enginee, Springer Berlin Heidelberg, 2008, pp. 238–260.
- [29] A. Y. NG, M. I. JORDAN, AND Y. WEISS, *On spectral clustering: Analysis and an algorithm*, in Advances in Neural Information Processing Systems (NIPS), MIT Press, 2001, pp. 849–856.
- [30] M. PENROSE, *Random geometric graphs*, vol. 5 of Oxford Studies in Probability, Oxford University Press, Oxford, 2003.
- [31] D. POLLARD, *Strong consistency of k -means clustering*, The Annals of Statistics, 9 (1981), pp. 135–140.
- [32] A. C. PONCE, *A new approach to Sobolev spaces and connections to Γ -convergence*, Calc. Var. Partial Differential Equations, 19 (2004), pp. 229–255.
- [33] M. REED AND B. SIMON, *I: Functional Analysis*, Methods of Modern Mathematical Physics, Elsevier Science, 1981.
- [34] S. E. SCHAEFFER, *Graph clustering*, Computer Science Review, 1 (2007), pp. 27–64.
- [35] J. SHI AND J. MALIK, *Normalized cuts and image segmentation*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22 (2000), pp. 888–905.
- [36] P. W. SHOR AND J. E. YUKICH, *Minimax grid matching and empirical measures*, Ann. Probab., 19 (1991), pp. 1338–1348.
- [37] A. SINGER, *From graph to manifold Laplacian: the convergence rate*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 128–134.
- [38] A. SINGER AND H.-T. WU, *Spectral convergence of the connection laplacian from random samples*. arXiv preprint arXiv:1306.1587, 2013.
- [39] D. A. SPIELMAN AND S. TENG, *A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning*, SIAM Journal on Computing, 42 (2013), pp. 1–26.
- [40] A. SZLAM AND X. BRESSON, *Total variation and cheeger cuts.*, in ICML, J. Fnkranz and T. Joachims, eds., Omnipress, 2010, pp. 1039–1046.
- [41] M. TALAGRAND, *The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3* , Ann. Probab., 22 (1994), pp. 919–959.
- [42] M. TALAGRAND, *The generic chaining*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2005. Upper and lower bounds of stochastic processes.
- [43] ———, *Upper and lower bounds of stochastic processes*, vol. 60 of Modern Surveys in Mathematics, Springer-Verlag, Berlin Heidelberg, 2014.

- [44] M. THORPE, F. THEIL, A. M. JOHANSEN, AND N. CADE, *Convergence of the k -means minimization problem using γ -convergence*. arXiv preprint arXiv:1501.01320, 2015.
- [45] D. TING, L. HUANG, AND M. I. JORDAN, *An analysis of the convergence of graph Laplacians*, in Proceedings of the 27th International Conference on Machine Learning, 2010.
- [46] C. VILLANI, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, American Mathematical Society, 2003.
- [47] U. VON LUXBURG, *A tutorial on spectral clustering*, Statistics and computing, 17 (2007), pp. 395–416.
- [48] U. VON LUXBURG, M. BELKIN, AND O. BOUSQUET, *Consistency of spectral clustering*, Ann. Statist., 36 (2008), pp. 555–586.
- [49] Y.-C. WEI AND C.-K. CHENG, *Towards efficient hierarchical designs by ratio cut partitioning*, in Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International Conference on, IEEE, 1989, pp. 298–301.
- [50] R. XU AND D. WUNSCH, II, *Survey of clustering algorithms*, Trans. Neur. Netw., 16 (2005), pp. 645–678.
- [51] J. YANG AND J. LESKOVEC, *Defining and evaluating network communities based on ground-truth*, Knowledge and Information Systems, 42 (2015), pp. 181–213.