Γ-CONVERGENCE OF PERIMETER ON RANDOM GEOMETRIC GRAPHS

NICOLÁS GARCÍA TRILLOS AND DEJAN SLEPČEV

ABSTRACT. We study the perimeter on the sets V_n of n random, uniformly distributed points in a Euclidean domain. To define the notion of perimeter on a finite set of points, the set is considered as a weighted graph, where the weights are assigned to edges connecting pairs of points based on their distances. The perimeter of $A_n \subset V_n$ is defined by summing the weights of edges between A_n and $V_n \backslash A_n$. We investigate under which choice of weights do the functionals which assign the graph perimeter converge to the perimeter in the Euclidean space as the number of points n goes to infinity. In particular, for $\varepsilon(n)$ such that significant weight is given to edges of length up to $\varepsilon(n)$ we investigate under which scaling of ε on n does the convergence hold. We consider this question in the setting of Γ-convergence and consider it for total-variation functional on graphs (which extends the notion of the perimeter).

1. Introduction

Our research is motivated by the study of point clouds and the analysis of the information they contain. By a point cloud we mean a finite set of points in the Euclidean space $\mathbb{R}^N$. One standard approach to exploring the structure of a data cloud is to define a weighted graph to represent it. Points become vertices, while the distances between them are used to assign weights to edges. In a variety of tasks such as classification and clustering it is useful to have information about a perimeter of a subset of the point cloud. To be precise, let $V = \{X_1, \dots, X_n\}$ be a point cloud. Let η be a kernel, that is, let $\eta : \mathbb{R}^N \to [0, \infty)$ be a radially symmetric, radially decreasing, function decaying to zero sufficiently fast. Typically the kernel is appropriately rescaled to take into account data density. In particular, let η_ε depend on a length scale ε so that significant weight is given to edges connecting points up to distance ε. We assign for $i, j \in \{1, \dots, n\}$ the weights by

$$w_{i,j} = \eta_\varepsilon(X_i - X_j)$$

and define the graph perimeter of $A \subset V$ by

$$\mathrm{Per}(A) = 2 \sum_{X_i \in A} \sum_{X_j \in V \backslash A} w_{i,j}. \tag{1}$$

The graph perimeter is also known as the cut capacity, if we thing of cutting the edges between A and its complement. It can be effectively used as a term in functionals which give a variational description to classification and clustering tasks [9, 12, 13, 16, 14, 15, 27, 28, 32, 39, 42, 43].

It is important to understand when is the graph perimeter defined above a good notion of a perimeter in the variational setting, especially for the case that the number of data points is large. To formulate this question mathematically we assume that the data points are random independent samples of an underlying measure ρ. In a practical data-analytic setting, the data considered often have some underlying, lower dimensional, structure. We make the "manifold assumption" that ρ is supported on a d-dimensional manifold in $\mathbb{R}^N$ and that ρ has a continuous density with respect

Date: November 4, 2013.

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

Key words and phrases. gamma convergence, graph perimeter, cut capacity, total variation, random geometric graph, point cloud, clustering.

to volume form on the manifold. The question is if the perimeter computed on the graph based on point cloud is a good (in a variational sense) approximation of the perimeter on the manifold. It is important for determining if the minimizers of functionals involving perimeter defined on the graphs converge, as the number of data points increases ($n \to \infty$), to a minimizer of the appropriate limiting functional in the continuum setting. Further question is to determine for what scaling of ε on n does the convergence hold. For computational considerations one desires to take ε small, but the question is how small ε can one choose when n is large, so that the graph structure still contains enough information for the graph perimeter to keep the needed properties.

In this paper we take the first, but in our opinion crucial step in investigating the above questions rigorously. Our main result is on a localized version of the above question. Namely we assume that data belong to $(0,1)^d$ (which can be thought as the coordinate chart for the manifold supporting ρ), and furthermore that on $(0,1)^d$ the points are drawn with uniform distribution (which is relevant locally given that we assumed that ρ varies continuously). While one can consider $(0,1)^d$ in a periodic setting, and thus entirely sidestep discussing the boundary of the domain, we chose to consider the question on bounded domains D in Euclidean setting. We first consider $D = (0,1)^d$, and then extend the result to more general domains.

1.1. **Setting and the main results.**

Consider $D = (0,1)^d$, the unit cube in $\mathbb{R}^d$, and assume n data points $X_1, \dots, X_n$ (i.i.d. random vectors) are chosen uniformly on D. We construct a graph on the set of these n data points using an isotropic kernel $\eta : \mathbb{R}^d \to [0,\infty)$ and a parameter $\varepsilon > 0$ by setting the weight of the edge between X_i and X_j to be $W_{i,j} := \eta_\varepsilon(X_i - X_j)$, where $\eta_\varepsilon(z) := \frac{1}{\varepsilon^d}\eta\left(\frac{z}{\varepsilon}\right)$. The total variation of a function u defined on a graph is typically defined as

$$\sum_{i,j} W_{i,j}|u(X_i) - u(X_j)|. \tag{2}$$

We note that the total variation is a generalization of perimeter since for a set A its perimeter is nothing but the total variation of χ_A, the characteristic function of A.

Having limits as $n \to \infty$ in mind, we define the *graph total variation* to be the following rescaled form of the functional above:

$$\widetilde{TV}_{n,\varepsilon}(u) := \frac{1}{\varepsilon}\frac{1}{n^2}\sum_{i,j} W_{i,j}|u(X_i) - u(X_j)|. \tag{3}$$

For a given scaling of ε with respect to n, we study the limiting behavior of $\widetilde{TV}_{n,\varepsilon(n)}$ as the number of points $n \to \infty$. The limit is considered in the variational sense of Γ-convergence. Before we state it precisely, we have to define the topology with respect to which the Γ-convergence is considered. To be able to compare functions defined on different sets of random points we define what we call the TL^1 metric space. We discuss this TL^1 metric in detail in Section 3. Here we just give an idea regarding the type of convergence it provides. We denote by ν_0 the Lebesgue measure restricted to the domain D and by ν_n the empirical measure associated to the first n data points, that is:

$$\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}, \tag{4}$$

It is known that except on a set of probability zero the sequence of measures $\{\nu_n\}_{n\in\mathbb{N}}$ converges weakly to ν_0. By the optimal transportation theory (discussed in Subsection 2.1) there exists a sequence of Borel maps $T_n : D \to D$ with $\nu_n = T_{n\sharp}\nu_0$ (i.e. T_n transports ν_0 to ν_n) such that:

$$\lim_{n\to\infty}\int_D |x - T_n(x)|dx = 0. \tag{5}$$

A sequence of functions $\{u_n\}_{n\in\mathbb{N}}$, where u_n belongs to $L^1(D,\nu_n)$, converges to $u \in L^1(D,\nu_0)$ in the TL^1 sense if the sequence $\{u_n \circ T_n\}_{n\in\mathbb{N}}$ converges in $L^1(D,\nu_0)$ to u. It is worth remarking that the notion of TL^1 convergence is independent of the choice of maps $\{T_n\}_{n\in\mathbb{N}}$ above, as long as (5) holds. For further details see Section 3 and in particular Proposition 3.6.

Since the kernel η is assumed isotropic, it can be thought to be defined as $\eta(x) := \boldsymbol{\eta}(|x|)$ for all $x \in \mathbb{R}^d$, where $\boldsymbol{\eta} : [0,\infty) \to [0,\infty)$ 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 $\int_0^\infty \boldsymbol{\eta}(r)\, r^d dr$ is finite.

We remark that the last assumption on $\boldsymbol{\eta}$ is equivalent to imposing that the surface tension

$$\sigma_\eta = \int_{\mathbb{R}^d} \eta(h)|h_1|dh \tag{6}$$

is finite. We note that the class of acceptable kernels is quite broad and includes both Gaussian kernels and discontinuous kernels like one defined by function $\boldsymbol{\eta}$ of the form $\boldsymbol{\eta} = 1$ for $t \le 1$ and $\boldsymbol{\eta} = 0$ for $t > 1$.

The total variation in continuum setting, $TV : L^1(D,\nu_0) \to [0,\infty]$, is given by

$$TV(u) = \sup\left\{ \int_D u \operatorname{div}\phi\, dx \ : \ \phi \in C_c^\infty(D,\mathbb{R}^d) \text{ and } \|\, |\phi|\, \|_{L^\infty} \le 1 \right\} \tag{7}$$

if the right-hand side is finite and is set to equal infinity otherwise. Here and in the rest of the paper we use $|\cdot|$ to denote the euclidean norm in $\mathbb{R}^d$. Note that if u is smooth enough then the total variation can be written as $TV(u) = \int_D |\nabla u| dx$.

The main result of the paper is:

Theorem 1.1 (Γ-convergence)**.** *Let $D = (0,1)^d$ and let $X_1, \dots, X_n, \dots$ be a sequence of i.i.d. random vectors chosen uniformly on D. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 and satisfying*

$$\begin{aligned} \lim_{n\to\infty} \frac{\sqrt{\log(\log n)}}{n^{1/2}} \frac{1}{\varepsilon_n} &= 0 \ \text{ if } d = 1, \\ \lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} &= 0 \ \text{ if } d = 2, \\ \lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} &= 0 \ \text{ if } d \ge 3. \end{aligned} \tag{8}$$

Let $\boldsymbol{\eta}$ be a function satisfying (K1)-(K3) and consider $\eta : \mathbb{R}^d \to [0,\infty)$ given by $\eta(x) := \boldsymbol{\eta}(|x|)$. Then, $\widetilde{TV}_{n,\varepsilon_n}$, defined by (3), Γ-converge to $\sigma_\eta TV$ as $n \to \infty$ in the TL^1 sense, where σ_η is given by (6) and TV is the total variation functional on D.

The notion of Γ-convergence in deterministic setting is recalled in Subsection 2.3, where we also extend it to the probabilistic setting in Definition 2.10. The following compactness result shows that the TL^1 topology is indeed a good topology for the Γ-convergence (see also Proposition 2.9).

Theorem 1.2 (Compactness)**.** *Under the assumptions of the theorem above, consider the sequence of functions $u_n \in L^1(D,\nu_n)$, where ν_n is given by (4). If $\{u_n\}_{n\in\mathbb{N}}$ have uniformly bounded $L^1(D,\nu_n)$ norms and graph total variations, $\widetilde{TV}_{n,\varepsilon_n}$, then the sequence is precompact in TL^1. More precisely if*

$$\sup_{n\in\mathbb{N}} \|u_n\|_{L^1(D,\nu_n)} < \infty,$$

and

$$\sup_{n\in\mathbb{N}} \widetilde{TV}_{n,\varepsilon_n}(u_n) < \infty,$$

then $\{u_n\}_{n\in N}$ *is* TL^1*-precompact.*

When A_n is a subset of $\{X_1,\dots,X_n\}$, it holds that $\widetilde{TV}_{n,\varepsilon_n}(\chi_{A_n}) = \frac{1}{n^2\epsilon_n}\,\mathrm{Per}(A_n)$, where $\mathrm{Per}(A_n)$ was defined in (1). Also, when $A \subseteq D$ we can write $TV(\chi_A) = \mathrm{Per}(A)$, where $\mathrm{Per}(A)$ in this context is the usual perimeter of A in D. Theorem 1.1 allows us to consider the variational convergence of the perimeter on the graphs to the perimeter in the domain D, that is:

Corollary 1.3 (Γ-convergence of Perimeter)**.** *Under the hypothesis of Theorem 1.1 the conclusions hold when all of the functionals are restricted to characteristic functions of sets. That is, the graph perimeters* Γ*-converge to the continuum perimeter.*

The proof of the theorems and of the corollary are presented in Section 5.

Remark 1.4. The notion of Γ-convergence was introduced by De Giorgi in the 70's and since then it has been used as the standard notion of variational convergence. With compactness it ensures that minimizers of approximate functionals converge (along a subsequence) to a minimizer of the limiting functional. This notion is quite different from the pointwise convergence of functionals, which of course does not guarantee convergence of minimizers. For extensive exposition of the properties of Γ-convergence see the books by Braides [10] and Dal Maso [21]. Pointwise convergence of the functionals $\widetilde{TV}_{n,\varepsilon_n}$ to TV (scaled with the appropriate constant) can be obtained from the results in [7]. In particular if one considers ε_n converging to zero, such that $\frac{(\log n)^{1/(d+1)}}{n^{1/(d+1)}}\frac{1}{\varepsilon_n}$ goes to zero as $n\to\infty$, and if A is a fixed closed regular subset of D then, $\widetilde{TV}_{n,\varepsilon_n}(\chi_{A\cap\{X_1,\dots,X_n\}}) \to \sigma_\eta TV(\chi_A)$ as $n\to\infty$ almost surely. We remark that from the techniques we use in our proofs, it is possible to establish that with the same scaling for ε_n as in Theorem 1.1, we get pointwise convergence as described above (which slightly improves the rate of pointwise convergence). Note that this does not follow directly from the Γ-convergence.

Remark 1.5. Theorem 1.2 implies that the probability that the weighted graph, with vertices $X_1,\dots,X_n$ and edge weights $w_{i,j} = \eta_{\varepsilon_n}(X_i - X_j)$, is connected converges to 1 as $n\to\infty$. Otherwise there is a sequence $n_k \nearrow \infty$ as $k\to\infty$ such that with positive probability, the graph above is not connected for all k. We can assume that $n_k = k$ for all k. Consider a connected component $A_n \subset \{X_1,\dots,X_n\}$ such that $\sharp A_n \le n/2$. Define function $u_n = \frac{n}{\sharp A_n}\chi_{A_n}$. Note that $\|u_n\|_{L^1(\nu_n)} = 1$ and that $\widetilde{TV}_{n,\varepsilon_n}(u_n) = 0$. By compactness, along a subsequence u_n converges in TL^1 to a function $u \in L^1(\nu_0)$. Thus $\|u\|_{L^1(\nu_0)} = 1$. By lower-semicontinuity which follows from Γ-convergence of Theorem 1.1 it follows that $TV(u) = 0$ and thus $u = 1$ on D. But since the values of u_n are either 0 or greater or equal to 2, it is not possible that u_n converges to u in TL^1. This is a contradiction.

1.2. **Optimal scaling of** $\varepsilon(n)$**.** If η is compactly supported and $d\ge 3$ then the rate presented in (8) is sharp in terms of scaling. Namely it is known from graph theory (see [35], [24] and [26]) that there exists a constant $\lambda > 0$ such that if $\varepsilon_n < \lambda\frac{(\log n)^{1/d}}{n^{1/d}}$ then the weighted graph associated to $X_1,\dots,X_n$ is disconnected with high probability. Therefore, in the light of Remark 1.5, the compactness property cannot hold if $\varepsilon_n < \lambda\frac{(\log n)^{1/d}}{n^{1/d}}$.

An interesting question arises if one restricts attention to a class of functions which are also uniformly bounded in L^∞ (for example the set of characteristic functions). In fact, the example of Remark 1.5 no longer applies. Thus, it is possible that Γ-convergence and compactness hold even when $\varepsilon_n < \lambda\frac{(\log n)^{1/d}}{n^{1/d}}$. Determining the optimal scaling on $\varepsilon(n)$ for Γ-convergence and compactness to hold in this setting is an important open problem. One could imagine that the (assymptotic) connectivity of the random geometric graph is still necessary condition for the Γ-convergence of

$\widetilde{TV}_{n,\varepsilon_n}$ functionals to TV, but it is worth remarking that this is not as simple as it seems. Indeed, it is known that for a scaling of ε_n between the connectivity threshold and the thermodynamical limit threshold a giant component (one containing a positive fraction of all points) appears in the graph. This component can asymptotically contain all but an asymptotically negligible fraction of the total number of points in the graph. If the scaling for ε_n is such that the fraction of points in the giant component approaches one as $n \to \infty$, then the Γ-convergence may still hold.

Nevertheless, it is important to notice that if ε_n scales like $1/n^{1/d}$ (so that the expected number of points per neighborhood is bounded by a constant, that is, $n\varepsilon_n^d = O(1)$), then even with the L^∞ bound, the Γ-convergence to $\sigma_\eta TV$ together with its compactness result cannot hold, since in this case the ratio of the cardinality of the biggest component in the graph and the total number of points n does not approach one as $n \to \infty$ (see Chapter 13 in [36] and [37]) and hence one can create sets in the discrete setting with zero (discrete) perimeter that approach a set with nonzero perimeter. For a more complete discussion of random geometric graphs, the giant component, the thermodynamical limit and the connectivity threshold for a random geometric graph see [36].

1.3. **Related works.** *Background on Γ-convergence of functionals to perimeter.* A classical example of Γ-convergence of functionals to perimeter is the Modica and Mortola theorem ([33]) that shows the Γ-convergence of Allen-Cahn (Cahn-Hilliard) free energy to perimeter. In particular they consider the energy

$$\varepsilon \int_D |\nabla u|^2 dx + \frac{1}{\varepsilon} \int W(u) dx, \tag{9}$$

where W is a double well potential.

In [3], Alberti and Bellettini study a nonlocal model for phase transitions where the energies do not have a gradient term as in the setting of Modica and Mortola, but a nonlocal term penalizing the spatial inhomogeneity of a given function u:

$$F_\varepsilon^W(u) := \frac{1}{\varepsilon} \int_D \int_D \eta_\varepsilon(x-y)|u(x)-u(y)|^2 dxdy + \frac{1}{\varepsilon}\int_D W(u)dx, \tag{10}$$

In the context of [3] the kernel η is not assumed radially symmetric. However, in the isotropic case, the results in [3] imply that the Γ-limit is the perimeter (multiplied by a surface tension).

In [40], Savin and Valdinoci consider the energy related to (10) with $\eta(z) = z^{-(d+2s)}$ for $s > 0$ and show that for $s \in (0.5, 1)$ the functionals Γ-converge to perimeter. For $s < 0.5$ they show that, under a somewhat different scaling, the nonlocality is preserved in the limit.

In the discrete setting, works related to the Γ convergence of functionals to continuous functionals involving perimeter include [11], [47] and [19]. The results by Braides and Yip [11], can be interpreted as the analogous results in a discrete setting to the ones obtained by Modica and Mortola. They give the description of the limiting functional (in the sense of Γ-convergence) after appropriately rescaling the energies. In the discretized version considered, they work on a regular grid. The gradient term gets replaced by a finite-difference approximation to it that depends on the mesh size δ. Different limiting functionals are obtained (as $\varepsilon \to 0$ and $\delta \to 0$) depending on the regime for the ratio $\frac{\varepsilon}{\delta}$. In particular in the case $\delta \ll \varepsilon$, the limiting functional is, as expected, the perimeter. Van Gennip and Bertozzi [47] consider a similar problem and obtain analogous results.

In [19], Chambolle, Giacomini and Lussardi consider a very general class of anisotropic perimeters defined on discrete subsets of a finite lattice of the form $\delta\mathbb{Z}^N$. They prove the Γ-convergence of the functionals as $\delta \to 0$ to an anisotropic perimeter defined on a given domain in $\mathbb{R}^d$.

Background on analysis of algorithms on point clouds as $n \to \infty$. In the past years a diverse set of geometrically based methods has been developed to solve different tasks of data analysis like classification, regression, dimensionality reduction and clustering. One desirable and important property that one expects from these methods is consistency. That is, it is desirable that as the number of data points tends to infinity the procedure used "converges" to some "limiting" procedure. Usually this "limiting" procedure involves a continuum functional defined on a domain in a Euclidean space or more generally on a manifold.

The available consistency results, include the work of Belkin and Niyogi [8] on the convergence of Laplacian Eigenmaps, the work by Pollard [38] where there are consistency results for the k-means algorithm used for clustering tasks and the work by von Luxburg, Belkin and Bousquet on consistency of spectral clustering [49].

The notion of Γ-convergence has been used extensively in the calculus of variations, in particular in homogenization theory, phase transitions, image processing, and material science. However, to the best of our knowledge it has not been used in data analysis to establish consistency of algorithms. It is our intention to introduce this notion in this context. We intend to use the ideas developed in this paper to explore some consistency problems in subsequent papers.

1.4. **Outline of the approach.** Since the functions we consider are given only at the data points, to be able to compare functions given on different data sets (in particular as number of data points is increasing) we introduce the TL^p topology in Section 3. This topology is used to formulate the Γ-convergence result.

The proof of Γ-convergence has two main steps. One is to compare the the graph total variation, $\widetilde{TV}_{n,\varepsilon_n}$ with a nonlocal continuum functional. In particular we consider the functional $TV_\varepsilon : L^1(D,\nu_0) \to [0,\infty]$ given by:

$$TV_\varepsilon(u) := \frac{1}{\varepsilon}\int_D \int_D \eta_\varepsilon(x-y)|u(x)-u(y)|dxdy, \tag{11}$$

Note that the argument of $\widetilde{TV}_{n,\varepsilon_n}$, is a function, u_n supported on the data points, while the argument of TV_ε is an L^1 function with respect to the Lebesgue measure. Thus to comare the functionals one needs to find a suitable $L^1(\nu_0)$ function which, in appropriate sense, approximates u_n. We use a transportation map (a notion recalled in Subsection 2.1) between ν_0 and ν_n to define such $\tilde{u}_n \in L^1(\nu_0)$. More precisely we set $\tilde{u}_n = u_n \circ T_n$ where T_n is the transportation map which is constructed in Subsection 2.2. We note that $\tilde{u}_n$ is piecewise constant and takes the same values as u_n. The map T_n is what enables going from the discrete to the continuum setting. Once again we contrast the situation here with the previous works which dealt with functionals on grids. When the discrete function is given on a grid it is straightforward to define the corresponding $L^1(\nu_0)$ function by letting it be constant on each grid cell, where it takes the value of a chosen grid point. On the other hand in the random setting finding the right correspondence was one of the main challenges (which also led us to introduce the TL^p topology). Comparing $\widetilde{TV}_{n,\varepsilon_n}(u_n)$ with $TV_{\varepsilon_n}(\tilde{u}_n)$ relies on choosing a transportation plan between the Lebesgue measure and the empirical measure which transports the points as little as possible. The estimates on how far the mass needs to be moved rely on previous works and are discussed in Subsection 2.2. One of the key properties of TL^1 convergence is that if functions u_n defined at data points converge to an $L^1(\nu_0)$ function u, $u_n \xrightarrow{TL^1} u$ as $n \to \infty$, then $\tilde{u}_n \xrightarrow{L^1(\nu_0)} u$ as $n \to \infty$.

This enables us to reduce the problem to comparing the continuum nonlocal total variation functional with the total variation. We prove, using results of Alberti and Bellettini [3] on non-local models for phase transitions, that the energy TV_ε, Γ-converges in the L^1-metric to $\sigma_\eta TV$. In their paper, Alberti and Bellettini consider energies F_ε^W of the form (10). It is proven that the functionals F_ε^W Γ-converge as $\varepsilon \to 0$ (in the L^1 sense) to $\sigma_{W,\eta} TV$, where $\sigma_{W,\eta}$ is the surface tension associated to the energies F_ε^W, and can be written as the solution to an optimal profile problem. Our result can be seen as a shape interface Γ-convergence result which parallels the diffuse-interface result of Alberti and Bellettini.

The compactness result that accompanies the Γ-convergence in [3] differs with the one we need. Indeed, the main difference is the L^∞ boundedness in [3] which is absent in our situation. The absence of L^∞ boundedness creates additional technical difficulties that we solve by reflecting the functions considered near the bounday.

The paper is organized as follows. Section 2 contains the notation and preliminary results from transportation theory and Γ-convergence of functionals on metric spaces. More specifically, in Subsection 2.1 we introduce the optimal transportation problem and list some basic results. In Subsection 2.2 we review the results on optimal matching between the empirical measure, ν_n, and the Lebesgue measure, ν_0. In Subsection 2.3 we recall the notion of Γ-convergence on metric spaces and introduce the appropriate extension to random setting. In Section 3 we define the metric space TL^p and prove some basic results about it. Section 4 contains the proof of the Γ-convergence of the nonlocal continuum total variation functional TV_ε to the TV functional. The main result, the Γ-convergence of the graph TV functionals to the TV functional is proved in Section 5. In Subsection 5.2 we discuss the extension of the main result to the case when $X_1, \dots, X_n$ are not independent, uniformly distributed points, but are still "close" to describing the Lebesgue measure. In Subsection 5.3 we extend the main results to more general domains D and in Subsection 5.4 we present a simple example of an application to data clustering.

2. Preliminaries

2.1. **Transportation theory.** In this section D is an open and bounded domain in $\mathbb{R}^d$. We denote by $\mathfrak{B}(D)$ the Borel σ-algebra of D and by $\mathcal{P}(D)$ the set of all Borel probability measures on D. Given $1 \leq p < \infty$, the p-OT distance between $\mu, \nu \in \mathcal{P}(D)$ (denoted by $d_p(\mu,\nu)$) is defined by:

$$d_p(\mu,\nu) := \min \left\{ \left(\int_{D\times D} |x-y|^p d\pi(x,y) \right)^{1/p} \,:\, \pi \in \Gamma(\mu,\nu) \right\}, \tag{12}$$

where $\Gamma(\mu,\nu)$ is the set of all *couplings* between μ and ν, 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 ν. The elements $\pi \in \Gamma(\mu,\nu)$ are called *transportation plans* between μ and ν. When $p = 2$ the distance is also known as the Wasserstein distance. The existence of minimizers, which justifies the definition above, is straightforward to show, see [48]. When $p = \infty$

$$d_\infty(\mu,\nu) := \inf \left\{ \text{esssup}_\pi \{|x-y| \,:\, (x,y) \in D \times D\} \,:\, \pi \in \Gamma(\mu,\nu) \right\}, \tag{13}$$

defines a metric on $\mathcal{P}(D)$, which is called the ∞-transportation distance.

It is known that for any $1 \leq p < \infty$, the convergence in OT metric is equivalent to weak convergence of probability measures and uniform integrability of p-moments. In our setting, the uniform integrability of p-moments is immediate since the domain D is assumed to be bounded, and hence for our purposes, convergence in OT metric is equivalent to weak convergence. For details see for instance [48], [6] and the references within. In particular, $\mu_n \xrightarrow{w} \mu$ (to be read μ_n converges weakly to μ) if and only if for any $1 \leq p < \infty$ there is a sequence of transportation plans between μ_n and μ, $\{\pi_n\}_{n\in\mathbb{N}}$, for which:

$$\lim_{n\to\infty} \iint_{D\times D} |x-y|^p d\pi_n(x,y) = 0. \tag{14}$$

Actually, note that since D is bounded, (14) is equivalent to $\lim_{n\to\infty} \iint_{D\times D} |x-y| d\pi_n(x,y) = 0$. 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 the condition (14). We remark that, since D is bounded, it is straightforward to show that a sequence of transportation plans is stagnating if and only if π_n converges weakly in the space of probability measures on $D \times D$ to $\pi = (id \times id)_\sharp \mu$.

Given a Borel map $T : D \to D$ and $\mu \in \mathcal{P}(D)$ the *push-forward* of μ by T, denoted by $T_\sharp \mu \in \mathcal{P}(D)$ is given by:

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

Then for any bounded continuous function $\varphi \in C(\overline{D})$ the following change of variables in the integral holds:

$$\int_D \varphi(x)\, d(T_\sharp \mu)(x) = \int \varphi(T(x))\, d\mu(x). \tag{15}$$

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

$$\pi_T := (\mathrm{Id} \times T)_\sharp \mu, \tag{16}$$

where $(\mathrm{Id} \times T) : D \to D \times D$ is given by $(\mathrm{Id} \times T)(x) = (x, T(x))$. Then for any $c \in L^1(D \times D, \mathfrak{B}\left(D \times D\right), \pi)$

$$\int_{D\times D} c(x,y) d\pi_T(x,y) = \int_D c\left(x, T(x)\right) d\mu(x). \tag{17}$$

It is well known that when the measure $\mu \in \mathcal{P}(D)$ is absolutely continuous with respect to the Lebesgue measure, the problem on the right hand side of (12) is equivalent to:

$$\min \left\{ \left(\int_D |x - T(x)|^p d\mu(x) \right)^{1/p} \ : \ T_\sharp \mu = \nu \right\}, \tag{18}$$

and when p is strictly greater than 1, the problem (12) has a unique solution which is induced (via (16)) by a transportation map T solving (18) (see [48]). In particular when the measure μ is absolutely continuous with respect to the Lebesgue measure, $\mu_n \xrightarrow{w} \mu$ as $n \to \infty$ is equivalent to the existence of a sequence $\{T_n\}_{n\in\mathbb{N}}$ of transportation maps, ($T_{n\sharp}\mu = \mu_n$) such that:

$$\int_D |x - T_n(x)| d\mu(x) \to 0, \ \text{ as } \ n \to \infty. \tag{19}$$

We say that a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ is *stagnating* if it satisfies (19).

We consider now the notion of inverse of transportation plans. For $\pi \in \Gamma(\mu, \nu)$, the *inverse plan* $\pi^{-1} \in \Gamma(\nu, \mu)$ of π is given by:

$$\pi^{-1} := s_\sharp \pi, \tag{20}$$

where $s : D \times D \to D \times D$ is defined as $s(x, y) = (y, x)$. Note that for any $c \in L^1(D \times D, \pi)$:

$$\int_{D\times D} c(x,y) d\pi(x,y) = \int_{D\times D} c(y,x) d\pi^{-1}(x,y).$$

Let $\mu, \nu, \rho \in \mathcal{P}(D)$. The *composition of plans* $\pi_{12} \in \Gamma(\mu, \nu)$ and $\pi_{23} \in \Gamma(\nu, \rho)$ was discussed in [6][Remark 5.3.3]. In particular there exists a probability measure $\boldsymbol{\pi}$ on $D \times D \times D$ such that the projection of $\boldsymbol{\pi}$ to first two variables is π_{12}, and to second and third variables is π_{23}. We consider π_{13} to be the projection of $\boldsymbol{\pi}$ to the first and third variables. We will refer π_{13} as *a* composition of π_{12} and π_{23} and write $\pi_{13} = \pi_{23} \circ \pi_{12}$.

2.2. **Optimal matching results.** In this section we discuss how to construct the transportation maps which allow us to make the transition from the functions of the data points to continuum functions. To obtain good estimates we want to match the Lebesgue measure with the empirical measure of data points while moving the mass as little as possible.

Consider $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space that we assume to be rich enough to support a sequence of independent random vectors uniformly distributed on the cube $(0,1)^d$. All the random variables that we use are to be understood as functions defined on Ω. We consider $X_1, \dots, X_n$ i.i.d. random vectors uniformly distributed in the cube $(0,1)^d$.

In the previous section we introduced the notion of stagnating sequence of transportation maps. In this section we present results by Leighton and Shor [30], and Shor and Yukich [41], that imply that not only we can find a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ (with $T_{n\sharp}\nu_0 = \nu_n$) which is stagnating but also satisfies a stronger condition:

$$\lim_{n\to\infty} \|Id - T_n\|_\infty = 0, \tag{21}$$

where Id is the identity mapping and $\|\cdot\|_\infty$ denotes the L^∞ norm with respect to the Lebesgue measure restricted to $(0,1)^d$. In other words $d_\infty(\nu_0, \nu_n) \to 0$ as $n \to \infty$. Moreover, the rates of convergence of $\|Id - T_n\|_\infty$ are obtained. These results play an important role in the proof of the main theorem of this paper.

To construct the transportation map between ν_0 and ν_n we first use the results above to match ν_n with a discrete measure supported on a regular grid with n points and then match the grid points with ν_0 locally. For that reason we first consider n of the form $n = k^d$ for some $k \in \mathbb{N}$. Consider $P = \{p_1, \dots, p_n\}$ the set of n points in $(0,1)^d$ of the form $(\frac{i_1}{2k}, \dots, \frac{i_n}{2k})$ for $i_1, \dots, i_n$ odd integers between 1 and $2k$. The points in P form a regular $k \times \cdots \times k$ array in $(0,1)^d$ and in particular each point in P is the center of a cube with volume $1/n$. As in [30] we call the points in P grid points and the cubes generated by the points in P grid cubes.

First consider the case $d = 2$. Leighton and Shor [30] showed that there exist $c > 0$ and $C > 0$ such that with very high probability (meaning probability greater than $1 - n^{-\alpha}$ where $\alpha = c_1(\log n)^{1/2}$ for some constant $c_1 > 0$):

$$\frac{c(\log n)^{3/4}}{n^{1/2}} \le \min_\pi \max_i |p_i - X_{\pi(i)}| \le \frac{C(\log n)^{3/4}}{n^{1/2}} \tag{22}$$

where π ranges over all permutations of $\{1, \dots, n\}$. In other words, when $d = 2$, with very high probability the ∞-transportation distance between the random points and the grid points is of order $\frac{(\log n)^{3/4}}{n^{1/2}}$.

We use the previous result to construct a transportation map T_n. Indeed, consider π minimizing $\min_\pi \max_i |X_i - p_{\pi(i)}|$. We define the map $T_n : (0,1)^d \to \{X_1, \dots, X_n\}$ as follows. Any $x \in (0,1)^d$ belongs to a grid cube associated to some $p_i \in P$. We set $T_n(x) := X_{\pi(i)}$. Then $T_{n\sharp}\nu_0 = \nu_n$ and (thanks to (22)) there are constants $c > 0$ and $C > 0$ such that with very high probability:

$$\frac{c(\log n)^{3/4}}{n^{1/2}} \le \|Id - T_n\|_\infty \le \frac{C(\log n)^{3/4}}{n^{1/2}}. \tag{23}$$

Remark 2.1. For the previous statement, we considered $n = k^2$ for some $k \in \mathbb{N}$ since the result in [30] is stated under the assumption that we work with a regular grid. Nevertheless, this is not a restriction as we now explain. In fact if n is not a square of an integer, so that $k^2 < n < (k+1)^2$ for some k, one can construct $\{C_1, \dots, C_n\}$, n rectangles covering D with sides parallel to the axis, each of them with volume $1/n$ and length of its diagonal below $\frac{c}{\sqrt{n}}$ where c is a universal constant. We can consider points $p_1, \dots, p_n$ in $C_1, \dots, C_n$ respectively. From the proof in [30] one can deduce (22) in this case again (even if n is not a square). The idea is that all probabilistic estimates in the proof in [30] are independent of the fact whether n is a square or not. The ingredient they needed to utilize their probabilistic estimates on the regular grid was the Hall's marriage lemma. The lemma can be applied to points $p_1, \dots, p_n$ in a similar way if n is not a square, given the properties of the rectangles $C_1, \dots, C_n$. Therefore one can take $x \in D$ and define $T_n(x) = X_{\pi(i)}$ where $x \in C_i$ and π is the optimal matching between the points $X_1, \dots, X_n$ and the points $p_1, \dots, p_n$. It is straightforward to check that $T_{n\sharp}\nu_0 = \nu_n$ and that (23) holds.

Using Borel–Cantelli lemma we deduce the following proposition.

Proposition 2.2. *Consider $d = 2$ and ν_n given by (4). Then there are constants $c > 0$ and $C > 0$ such that for $\mathbb{P}$-a.e. $\omega \in \Omega$ there exists a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from ν_0 to ν_n ($T_{n\sharp}\nu_0 = \nu_n$) and such that:*

$$c \leq \liminf_{n\to\infty} \frac{n^{1/2}\|Id - T_n\|_\infty}{(\log n)^{3/4}} \leq \limsup_{n\to\infty} \frac{n^{1/2}\|Id - T_n\|_\infty}{(\log n)^{3/4}} \leq C. \tag{24}$$

For $d \geq 3$, Shor and Yukich [41] proved the analogous result to (22). As in the case $d = 2$ one is not restricted to considering only $n = k^d$ for some integer k. An immediate consequence of their results is the following:

Proposition 2.3. *Consider $d \geq 3$ and ν_n given by (4). For $\mathbb{P}$-a.e. $\omega \in \Omega$ there exists a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from ν_0 to ν_n ($T_{n\sharp}\nu_0 = \nu_n$) and constants $c > 0$, $C > 0$ (only depending on d) such that:*

$$c \leq \liminf_{n\to\infty} \frac{n^{1/d}\|Id - T_n\|_\infty}{(\log n)^{1/d}} \leq \limsup_{n\to\infty} \frac{n^{1/d}\|Id - T_n\|_\infty}{(\log n)^{1/d}} \leq C. \tag{25}$$

One should note that the powers of the log terms are different when $d = 2$ and when $d \geq 3$. In both cases they are sharp, meaning it is impossible to find transportation maps $\{T_n\}_{n\in\mathbb{N}}$ with $T_{n\sharp}\nu_0 = \nu_n$ and where $||Id - T_n||_\infty$ converges faster to zero than $\frac{(\log(n))^{1/d}}{n^{1/d}}$ in case $d \geq 3$ or $\frac{(\log(n))^{3/4}}{n^{1/2}}$ in case $d = 2$.

The basic matching algorithm involving random uniformly distributed points was introduced by Ajtai, Komlós, and Tusnády in [2]. That algorithm can still be used to obtain results like the ones above (not quite as sharp for $d = 2$). It relies on a dyadic decomposition of the cube $(0, 1)^d$ and transporting step by step between levels of the dyadic decomposition. The final matching is obtained as a composition of the matchings between consecutive levels. For $d \geq 3$ the AKT algorithm is the basic tool used in [41] to prove the analogous result to (22) for $d \geq 3$. For $d = 2$ the AKT algorithm gives an upper bound with extra log factors, which makes the proof of the sharp bounds more complicated.

As remarked in [41], there is a crossover in the nature of the matching when $d = 2$: for $d \geq 3$, the matching length between the random points and the points in the grid is determined by the behavior of the points locally, for $d = 1$ on the other hand, the matching length is determined by the behavior of random points globally, and finally for $d = 2$ the matching length is determined by the behavior of the random points at all scales. At the level of the AKT algorithms this means that for $d \geq 3$ the major source of the transportation distance is at the finest scale, for $d = 1$ at the coarsest scale, while for $d = 2$ distances at all scales are of the same size (in terms of how they scale with n).

Problems related to the ones considered in [30] and [41] were also considered by Talagrand and Yukich [45] and Talagrand [44]. In particular it is possible to deduce (25) from the results in [44] and to obtain an upper bound for the asymptotic behavior of $\|Id - T_n\|_\infty$ in the case $d = 2$ from the results in [45].

So far we have not discussed the result in case $d = 1$. In this case we can take advantage of the classical results by Chung [20] (or Kiefer [29]) on the behaviour of the discrepancy of the measures ν_n and ν_0 as $n \to \infty$. In fact if we consider the points $p_1, \dots, p_n$ with $p_i = i/n$, and consider random points $X_1, \dots, X_n$ that (after reordering) we assume to be in increasing order, it is not hard to see that the permutation π that minimizes $\min_\pi \max_i |p_i - X_{\pi(i)}|$ is precisely the identity, in other words $\min_\pi \max_i |p_i - X_{\pi(i)}| = \max_i |p_i - X_i|$. The results on [20] can then be used to obtain asymptotic (as $n \to \infty$) sharp estimates on the value of $\max_i |p_i - X_i|$, and in particular deduce that with probability one $\max_i |p_i - X_i| = O\left(\sqrt{\frac{\log\log(n)}{n}}\right)$. It is an immediate consequence of these results the following:

Proposition 2.4. *There are constants $c > 0$ and $C > 0$ such that for $\mathbb{P}$-a.e. $\omega \in \Omega$ there exists a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ (with $T_{n\sharp}\nu_0 = \nu_n$) and such that:*

$$c \leq \liminf_{n\to\infty} \frac{n^{1/2}\|Id - T_n\|_\infty}{\sqrt{\log\log(n)}} \leq \limsup_{n\to\infty} \frac{n^{1/2}\|Id - T_n\|_\infty}{\sqrt{\log\log(n)}} \leq C. \tag{26}$$

2.3. **Γ-convergence on metric spaces.** We recall and discuss the notion of Γ-convergence in general setting. Let (X, d_X) be a metric space. Let $F_n : X \to [0,\infty]$ be a sequence of functionals.

Definition 2.5. *The sequence $\{F_n\}_{n\in\mathbb{N}}$ Γ-converges with respect to metric d_X to the functional $F : X \to [0,\infty]$ as $n \to \infty$ if the following inequalities hold:*

1. ***Liminf inequality:*** *For every $x \in X$ and every sequence $\{x_n\}_{n\in\mathbb{N}}$ converging to x,*
$$\liminf_{n\to\infty} F_n(x_n) \geq F(x),$$
2. ***Limsup inequality:*** *For every $x \in X$ there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ converging to x satisfying*
$$\limsup_{n\to\infty} F_n(x_n) \leq F(x).$$

We say that F is the Γ-limit of the sequence of functionals $\{F_n\}_{n\in\mathbb{N}}$ (with respect to the metric d_X).

Remark 2.6. In most situations one does not prove the limsup inequality for all $x \in X$ directly. Instead, one proves the inequality for all x in a dense subset X' of X where it is somewhat easier to prove, and then deduce from this that the inequality holds for all $x \in X$. To be more precise, suppose that the limsup inequality is true for every x in a subset X' of X and the set X' is such that for every $x \in X$ there exists a sequence $\{x_k\}_{k\in\mathbb{N}}$ in X' converging to x and such that $F(x_k) \to F(x)$ as $k \to \infty$, then the limsup inequality is true for every $x \in X$. It is enough to use a diagonal argument to deduce this claim.

Definition 2.7. *We say that the sequence of nonnegative functionals $\{F_n\}_{n\in\mathbb{N}}$ satisfies the compactness property if the following holds: Given $\{n_k\}_{k\in\mathbb{N}}$ an increasing sequence of natural numbers and $\{x_k\}_{k\in\mathbb{N}}$ a bounded sequence in X for which*

$$\sup_{k\in\mathbb{N}} F_{n_k}(x_k) < \infty \tag{27}$$

$\{x_k\}_{k\in\mathbb{N}}$ is precompact in X.

Remark 2.8. Note that the boundedness assumption of $\{x_k\}_{k\in\mathbb{N}}$ in the previous definition is a necessary condition for precompactness and so it is not restrictive.

The notion of Γ-convergence is particularly useful when the functionals $\{F_n\}_{n\in\mathbb{N}}$ satisfy the compactness property. This is because it guarantees convergence of minimizers (or approximate minimizers) of F_n to minimizers of F and it also guarantees convergence of the minimum energy of F_n to the minimum energy of F (this statement is made precise in the next proposition). This is the reason why Γ-convergence is said to be a variational type of convergence.

Proposition 2.9. *Let $F_n : X \to [0,\infty]$ be a sequence of nonnegative functionals which are not identically equal to $+\infty$, satisfying the compactness property and Γ-converging to the functional $F : X \to [0,\infty]$ which is not identically equal to $+\infty$. Then,*

$$\lim_{n\to\infty} \inf_{x\in X} F_n(x) = \min_{x\in X} F(x). \tag{28}$$

Furthermore every bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ *in* X *for which*

$$\lim_{n\to\infty}\left(F_n(x_n) - \inf_{x\in X} F_n(x)\right) = 0 \tag{29}$$

is precompact and each of its cluster points is a minimizer of F.

In particular, if F *has a unique minimizer, then a sequence* $\{x_n\}_{n\in\mathbb{N}}$ *satisfying* (29) *converges to the unique minimizer of* F.

One can extend the concept of Γ-convergence to families of functionals indexed by real numbers in a simple way, namely, the family of functionals $\{F_h\}_{h>0}$ is said to Γ-converge to F as $h\to 0$ if for every sequence $\{h_n\}_{n\in\mathbb{N}}$ with $h_n\to 0$ as $n\to\infty$ the sequence $\{F_{h_n}\}_{n\in\mathbb{N}}$ Γ-converges to the functional F as $n\to\infty$. Similarly one can define the compactness property for the functionals $\{F_h\}_{h>0}$. For more on the notion of Γ-convergence see [10] or [21].

Since the functionals we are most interested in depend on data (and hence are random), we need to define what it means for a sequence of random functionals to Γ-converge to a deterministic functional.

Definition 2.10. *For* $\{F_n\}_{n\in\mathbb{N}}$ *a sequence of (random) functionals* $F_n : X\times\Omega\to[0,\infty]$ *and* F *a (deterministic) functional* $F : X\to[0,\infty]$, *we say that the sequence of functionals* $\{F_n\}_{n\in\mathbb{N}}$ Γ-*converges (in the* d_X *metric) to* F, *if for* $\mathbb{P}$-*almost every* $\omega\in\Omega$ *the sequence* $\{F_n(\cdot,\omega)\}_{n\in\mathbb{N}}$ Γ-*converges to* F *according to Definition 2.5. Similarly, we say that* $\{F_n\}_{n\in\mathbb{N}}$ *satisfies the compactness property if for* $\mathbb{P}$-*almost every* $\omega\in\Omega$, $\{F_n(\cdot,\omega)\}_{n\in\mathbb{N}}$ *satisfies the compactness property according to Definition 2.7.*

We do not explicitly write the dependence of F_n on ω and we simply write $F_n : X\to[0,\infty]$, understanding that we are always working with a fixed value $\omega\in\Omega$, and hence with deterministic functionals.

3. The space TL^p

In this section D denotes an open and bounded domain in $\mathbb{R}^d$. Consider the set

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

For (μ, f) and (ν, g) in TL^p we define

$$d_{TL^p}((\mu,f),(\nu,g)) = \inf_{\pi\in\Pi(\mu,\nu)}\left(\iint_{D\times D}|x-y|^p d\pi(x,y)\right)^{\frac{1}{p}} + \left(\iint_{D\times D}|f(x)-g(y)|^p d\pi(x,y)\right)^{\frac{1}{p}} \tag{30}$$

The next proposition shows that d_{TL^p} is a metric. We remark that formally TL^p is a fiber bundle over $\mathcal{P}(D)$. Namely if one considers the Finsler (Riemannian for $p=2$) manifold structure on $\mathcal{P}(D)$ provided by the $p-OT$ metric (see [1] for general p and [34, 6] for $p=2$) then TL^p is, formally, a fiber bundle. We also remark that one could also change the set and consider a metric where the powers of the terms in (30) would be different (p and q, instead of p and p and the natural name for the space in this case would be $TL^{p,q}$).

Remark 3.1. One can think of the convergence in TL^p as a generalization of weak convergence of measures and convergence in L^p of functions. By this we mean that $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathcal{P}(D)$ converges weakly (and in p-OT sense for any p since D is bounded) to $\mu\in\mathcal{P}(D)$ if and only if $(\mu_n,1)\xrightarrow{TL^p}(\mu,1)$ as $n\to\infty$, and that for $\mu\in\mathcal{P}(D)$ a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $L^p(\mu)$ converges in $L^p(\mu)$ to f if and only if $(\mu,f_n)\to_{TL^p}(\mu,f)$ as $n\to\infty$. The last fact is established in Proposition 3.6.

Remark 3.2. If one restricts the attention to measures μ and ν which are absolutely continuous with respect to the Lebesgue measure then

$$\inf_{T\,:\,T_\sharp\mu=\nu}\left(\int_D |x-T(x)|^p d\mu(x)\right)^{\frac{1}{p}}+\left(\int_D |f(x)-g(T(x))|^p d\mu(x)\right)^{\frac{1}{p}}$$

majorizes $d_{TL^p}((\mu,f),(\nu,g))$ and furthermore provides a metric (on the subset of TL^p) which gives the same topology as d_{TL^p}. The fact that these topologies are the same follows from Proposition 3.6.

Proposition 3.3. *d_{TL^p} defines a metric on TL^p.*

Proof. To prove that $d_{TL^p}\left((\mu,f),(\mu,f)\right)=0$, note that if we consider $\pi=r_\sharp\mu$ where $r:D\to D\times D$ is given by $r(x)=(x,x)$, $\pi\in\Gamma(\mu,\mu)$ and $\iint_{D\times D}|x-y|^p+|f(x)-g(y)|^p d\pi(x,y)=0$.

Suppose now that $d_{TL^p}\left((\mu,f),(\nu,g)\right)=0$. In particular it is true that $d_p(\mu,\nu)=0$ and so $\mu=\nu$. Thus there exists a sequence $\{\pi_n\}_{n\in\mathbb{N}}$ of transportation plans belonging to $\Gamma(\mu,\mu)$ such that:

$$\lim_{n\to\infty}\iint_{D\times D}|x-y|^p d\pi_n(x,y)+\iint_{D\times D}|g(x)-f(y)|^p d\pi_n(x,y)=0 \tag{31}$$

Note that the sequence $\{\pi_n\}_{n\in\mathbb{N}}$ is tight in $\mathcal{P}\left(D\times D\right)$, and hence it has a subsequence (that we do not relabel) that converges weakly to some $\pi\in\Gamma(\mu,\mu)$. Using the boundedness of D we conclude that:

$$\iint_{D\times D}|x-y|^p d\pi(x,y)=\lim_{n\to\infty}\iint_{D\times D}|x-y|^p d\pi_n(x,y)=0$$

Consequently $\pi=r_\sharp\mu$, for r as above. For every $\varepsilon>0$ there exist $\tilde f$, $\tilde g$ functions defined on D which are continuous and bounded and such that $\|f-\tilde f\|_{L^p(\mu)}<\varepsilon/2$ and $\|g-\tilde g\|_{L^p(\mu)}<\varepsilon/2$. Then

$$\begin{aligned}\left(\iint_{D\times D}\left|\tilde f(x)-\tilde g(y)\right|^p d\pi_n(x,y)\right)^{1/p}&\le\left(\iint_{D\times D}\left|\tilde f(x)-f(x)\right|^p d\pi_n(x,y)\right)^{1/p}\\&\quad+\left(\iint_{D\times D}|f(x)-g(y)|^p\, d\pi_n(x,y)\right)^{1/p}\\&\quad+\left(\iint_{D\times D}|g(y)-\tilde g(y)|^p\, d\pi_n(x,y)\right)^{1/p}\\&<\left(\iint_{D\times D}|f(x)-g(y)|^p\, d\pi_n(x,y)\right)^{1/p}+\varepsilon\end{aligned}$$

From (31) and the fact that $\pi_n\xrightarrow{w}\pi$ we obtain:

$$\begin{aligned}\|\tilde f-\tilde g\|_{L^p(\mu)}&=\left(\iint_{D\times D}\left|\tilde f(x)-\tilde g(y)\right|^p d\pi(x,y)\right)^{1/p}\\&=\lim_{n\to\infty}\left(\iint_{D\times D}\left|\tilde f(x)-\tilde g(y)\right|^p d\pi_n(x,y)\right)^{1/p}<\varepsilon\end{aligned}$$

Therefore:

$$\|f-g\|_{L^p(\mu)}\le\|f-\tilde f\|_{L^p(\mu)}+\|\tilde f-\tilde g\|_{L^p(\mu)}+\|\tilde g-g\|_{L^p(\mu)}<3\varepsilon.$$

Since ε was arbitrary, we conclude that $f=g$, and so $d_{TL^p}\left((\mu,f),(\nu,g)\right)=0$ implies $(\mu,f)=(\nu,g)$.

To prove that $d_{TL^p}((\mu,f),(\nu,g)) = d_{TL^p}((\nu,g),(\mu,f))$ simply note that for every $\pi \in \Gamma(\mu,\nu)$

$$\iint_{D\times D} |x-y|^p d\pi(x,y) = \iint_{D\times D} |x-y|^p d\pi^{-1}(x,y)$$

and
$$\iint_{D\times D} |f(x)-g(y)|^p d\pi(x,y) = \iint_{D\times D} |g(x)-f(y)|^p d\pi^{-1}(x,y)$$

where $\pi^{-1} \in \Gamma(\nu,\mu)$ is the inverse of π as defined in (20).

Finally, consider $(\mu,f),(\nu,g),(\sigma,h) \in TL^p$. Take $\pi_1 \in \Gamma(\mu,\nu)$ and $\pi_2 \in \Gamma(\nu,\sigma)$. We now use a measure $\boldsymbol{\pi} \in \mathcal{P}(D\times D\times D)$ to obtain $\pi_2 \circ \pi_1 \in \Gamma(\mu,\sigma)$ as mentioned at the end of Subsection 2.1 . Using Minkowski's inequality for the measure $\boldsymbol{\pi}$, one obtains that:

$$\begin{aligned} d_{TL^p}((\mu,f),(\sigma,h)) \leq &\left(\iint_{D\times D} |x-z|^p d\pi_2\circ\pi_1(x,z)\right)^{\frac{1}{p}} + \left(\iint_{D\times D} |f(x)-h(z)|^p d\pi_2\circ\pi_1(x,z)\right)^{\frac{1}{p}} \\ \leq &\left(\iint_{D\times D} |x-y|^p d\pi_1(x,y)\right)^{\frac{1}{p}} + \left(\iint_{D\times D} |f(x)-g(y)|^p d\pi_1(x,y)\right)^{\frac{1}{p}} \\ &+ \left(\iint_{D\times D} |y-z|^p d\pi_2(y,z)\right)^{\frac{1}{p}} + \left(\iint_{D\times D} |g(y)-h(z)|^p d\pi_2(y,z)\right)^{\frac{1}{p}} \end{aligned}$$

Taking infimum over π_1 and over π_2 on the previous expression we deduce that $d_{TL^p}((\mu,f),(\sigma,h)) \leq d_{TL^p}((\mu,f),(\nu,g)) + d_{TL^p}((\nu,g),(\sigma,h))$. □

We wish to establish a simple characterization for the convergence in the space TL^p. For this, we need first the following two lemmas.

Lemma 3.4. *Fix $\mu \in \mathcal{P}(D)$. For any stagnating sequence of transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ (with $\pi_n \in \Gamma(\mu,\mu)$) and for any $u \in L^p(\mu)$:*

$$\lim_{n\to\infty} \iint_{D\times D} |u(x)-u(y)|^p d\pi_n(x,y) = 0$$

Proof. We prove the case $p=1$ since the other cases are similar. Let $u \in L^1(\mu)$ and let $\{\pi_n\}_{n\in\mathbb{N}}$ be a stagnating sequence of transportation maps with $\pi_n \in \Gamma(\mu,\mu)$. Since the probability measure μ is inner regular, we know that the class of Lipschitz and bounded functions on D is dense in $L^1(\mu)$. Fix $\varepsilon > 0$, we know there exists a function $v : D \to \mathbb{R}$ which is Lipschitz and bounded and for which:

$$\int_D |u(x)-v(x)| d\mu(x) < \frac{\varepsilon}{3}$$

Note that:

$$\iint_{D\times D} |v(x)-v(y)| d\pi_n(x,y) \leq \mathrm{Lip}(v) \iint_{D\times D} |x-y| d\pi_n(x,y) \to 0, \ as\ n\to\infty$$

Hence we can find $N \in \mathbb{N}$ such that if $n \geq N$ then $\iint_{D\times D} |v(x)-v(y)| d\pi_n(x,y) < \frac{\varepsilon}{3}$. Therefore, for $n \geq N$, using the triangle inequality, we obtain

$$\begin{aligned} \iint_{D\times D} |u(x)-u(y)| d\pi_n(x,y) \leq &\iint_{D\times D} |u(x)-v(x)| d\pi_n(x,y) + \iint_{D\times D} |v(x)-v(y)| d\pi_n(x,y) \\ &+ \iint_{D\times D} |v(y)-u(y)| d\pi_n(x,y) \\ =&2\int_D |v(x)-u(x)| d\mu(x) + \iint_{D\times D} |v(x)-v(y)| d\pi_n(x,y) < \varepsilon \end{aligned}$$

This proves the result. □

Lemma 3.5. *Suppose that the sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathcal{P}(D)$ converges weakly to $\mu \in \mathcal{P}(D)$. Let $\{u_n\}_{n\in\mathbb{N}}$ be a sequence with $u_n \in L^p(\mu_n)$ and let $u \in L^p(\mu)$. Consider two sequences of stagnating transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ and $\{\hat\pi_n\}_{n\in\mathbb{N}}$ (with $\pi_n, \hat\pi_n \in \Gamma(\mu,\mu_n)$). Then:*

$$\lim_{n\to\infty}\iint_{D\times D}|u(x)-u_n(y)|^p d\pi_n(x,y) = 0 \Leftrightarrow \lim_{n\to\infty}\iint_{D\times D}|u(x)-u_n(y)|^p d\hat\pi_n(x,y) = 0 \tag{32}$$

Proof. We present the details for $p = 1$, as the other cases are similar. Take $\hat\pi_n^{-1} \in \Gamma(\mu_n,\mu)$ the inverse of $\hat\pi_n$ defined in (20). We can consider $\boldsymbol{\pi_n} \in \mathcal{P}(D\times D\times D)$ as the one mentioned at the end of Subsection 2.1 (taking $\pi_{23} = \hat\pi_n^{-1}$ and $\pi_{12} = \pi_n$). In particular $\hat\pi_n^{-1}\circ\pi_n \in \Gamma(\mu,\mu)$. Then

$$\iint_{D\times D}|u_n(y)-u(x)|d\pi_n(x,y) = \iiint_{D\times D\times D}|u_n(y)-u(x)|d\boldsymbol{\pi}_n(x,y,z),$$

and

$$\begin{aligned}\iint_{D\times D}|u_n(z)-u(y)|d\hat\pi_n(y,z) &= \iint_{D\times D}|u_n(y)-u(z)|d\hat\pi_n^{-1}(y,z)\\ &= \iiint_{D\times D\times D}|u_n(y)-u(z)|d\boldsymbol{\pi}_n(x,y,z),\end{aligned}$$

which imply after using the triangle inequality:

$$\begin{aligned}&\left|\iint_{D\times D}|u_n(y)-u(x)|d\pi_n(x,y) - \iint_{D\times D}|u(z)-u_n(y)|d\hat\pi_n(y,z)\right|\\ &\qquad\le \iiint_{D\times D\times D}|u(z)-u(x)|d\boldsymbol{\pi}_n(x,y,z) = \iint_{D\times D}|u(z)-u(x)|d\hat\pi_n^{-1}\circ\pi_n(x,z)\end{aligned} \tag{33}$$

Finally note that :

$$\iint_{D\times D}|x-z|d\hat\pi_n^{-1}\circ\pi_n(x,z) \le \iint_{D\times D}|x-y|d\pi_n(x,y) + \iint_{D\times D}|y-z|d\hat\pi_n(z,y) \to 0, \text{ as } n\to\infty.$$

The sequence $\left\{\hat\pi_n^{-1}\circ\pi_n\right\}_{n\in\mathbb{N}}$ satisfies the assumptions of Lemma 3.4, so we can deduce that $\iint_{D\times D}|u(z)-u(x)|d\hat\pi_n^{-1}\circ\pi_n(x,z) \to 0$ as $n\to\infty$. By (33) we get that:

$$\lim_{n\to\infty}\left|\iint_{D\times D}|u_n(y)-u(x)|d\pi_n(x,y) - \iint_{D\times D}|u_n(z)-u(y)|d\hat\pi_n(y,z)\right| = 0.$$

This implies the result. □

Proposition 3.6. *Let $(\mu,f)\in TL^p$ and let $\{(\mu_n,f_n)\}_{n\in\mathbb{N}}$ be a sequence in TL^p. The following statements are equivalent:*

1. *$(\mu_n,f_n) \xrightarrow{TL^p} (\mu,f)$ as $n\to\infty$.*
2. *$\mu_n \xrightarrow{w} \mu$ and for every stagnating sequence of transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ (with $\pi_n \in \Gamma(\mu,\mu_n)$)*

$$\iint_{D\times D}|f(x)-f_n(y)|^p\, d\pi_n(x,y) \to 0, \text{ as } n\to\infty. \tag{34}$$

3. *$\mu_n \xrightarrow{w} \mu$ 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 (34) holds.*

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

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

$$\int_D |f(x) - f_n(T_n(x))|^p \, d\mu(x) \to 0, \text{ as } n \to \infty. \tag{35}$$

5. $\mu_n \xrightarrow{w} \mu$ *and for any stagnating sequence of transportation maps* $\{T_n\}_{n\in\mathbb{N}}$ *(with* $T_{n\sharp}\mu = \mu_n$*)* (35) *holds.*

Proof. By Lemma 3.5, claims 2. and 3. are equivalent. In case μ is absolutely continuous with respect to the Lebesgue measure, we know that there exists a stagnating sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ (with $T_{n\sharp}\mu = \mu_n$). Considering the sequence of transportation plans $\{\pi_{T_n}\}_{n\in\mathbb{N}}$ (as defined in (16)) and using (17) we see that 2., 3., 4., and 5. are all equivalent. We prove the equivalence of 1. and 3.

(1. $\Rightarrow$ 3.) Note that $d_p(\mu,\mu_n) \le d_{TL^p}((\mu,f),(\mu_n,f_n))$ for every n. Hence $d_p(\mu,\mu_n) \to 0$ as $n\to\infty$ and in particular $\mu_n \xrightarrow{w} \mu$ as $n \to \infty$. Also, since $d_{TL^p}((\mu,f),(\mu_n,f_n)) \to 0$ as $n\to\infty$, in particular there exists a sequence $\{\pi_n^*\}_{n\in\mathbb{N}}$ of transportation plans (with $\pi_n^* \in \Gamma(\mu,\mu_n)$) such that:

$$\lim_{n\to\infty} \iint_{D\times D} |x-y|^p d\pi_n^*(x,y) = 0,$$

$$\lim_{n\to\infty} \iint_{D\times D} |f(x)-f_n(y)|^p d\pi_n^*(x,y) = 0.$$

$\{\pi_n^*\}_{n\in\mathbb{N}}$ is then a stagnating sequence of transportation plans for which (34) holds.

(3. $\Rightarrow$ 1.) Since $\mu_n \xrightarrow{w} \mu$ as $n\to\infty$ (and since D is bounded), we know that $d_p(\mu_n,\mu)\to 0$ as $n\to\infty$. In particular, we can find a sequence of transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ with $\pi_n \in \Gamma(\mu,\mu_n)$ such that:

$$\lim_{n\to\infty} \iint_{D\times D} |x-y|^p d\pi_n(x,y) = 0$$

$\{\pi_n\}_{n\in\mathbb{N}}$ is then a stagnating sequence of transportation plans. By the hypothesis we conclude that:

$$\lim_{n\to\infty} \iint_{D\times D} |f(x)-f_n(y)|^p d\pi_n(x,y) = 0$$

From this we deduce that $\lim_{n\to\infty} d_{TL^p}((\mu,f),(\mu_n,f_n)) = 0$. □

Definition 3.7. *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* $u_n \in L^p(\mu_n)$*) converges in the* TL^p *sense to* $u\in L^p(\mu)$*, if* $\{(\mu_n,u_n)\}_{n\in\mathbb{N}}$ *converges to* (μ,u) *in the* TL^p *metric. In this case we use a slight abuse of notation and write* $u_n \xrightarrow{TL^p} u$ *as* $n\to\infty$*. Also, we say the sequence* $\{u_n\}_{n\in\mathbb{N}}$ *(with* $u_n\in L^p(\mu_n)$*) is precompact in* TL^p *if the sequence* $\{(\mu_n,u_n)\}_{n\in\mathbb{N}}$ *is precompact in* TL^p*.*

Remark 3.8. Thanks to Proposition 3.6 when μ is absolutely continuous with respect to the Lebesgue measure $u_n \xrightarrow{TL^p} u$ as $n\to\infty$ if and only if for every (or one) $\{T_n\}_{n\in\mathbb{N}}$ stagnating sequence of transportation maps (with $T_{n\sharp}\mu=\mu_n$) it is true that $u_n\circ T_n \xrightarrow{L^p(\mu)} u$ as $n\to\infty$. Also $\{u_n\}_{n\in\mathbb{N}}$ is precompact in TL^p if and only if for every (or one) $\{T_n\}_{n\in\mathbb{N}}$ stagnating sequence of transportation maps (with $T_{n\sharp}\mu=\mu_n$) it is true that $\{u_n\circ T_n\}_{n\in\mathbb{N}}$ is precompact in $L^p(\mu)$.

4. Γ-convergence of TV_ε

Consider an open, bounded domain D in $\mathbb{R}^d$ with Lipschitz boundary. Here we prove that $\{TV_\varepsilon\}_{\varepsilon>0}$ (defined in (11)) Γ-converges with respect to the L^1 metric to $\sigma_\eta TV$. This result and the proof we present rely on the work of Alberti and Bellettini [3] and [4]. While a direct approach, which bypasses using the diffuse interface functional (36) is possible (see for example [22], [19]), we base our argument on results of [3] to make the presentation short. The only part which required substantial new arguments is the proof of compactness, where due to the presence of domain boundary and lack of L^∞ control, new ideas were needed. For this result we impose extra regularity assumptions on the domain D. As a corollary, we show that if one considers only functions uniformly bounded in L^∞, the compactness holds for open and bounded domains D regardless of its boundary.

Let $W(z) = 4z^2(1-z)^2$ and let $W_k = kW$. We define the functionals:

$$F_\varepsilon^{(k)}(u) := \frac{1}{\varepsilon}\int_{D\times D} \eta_\varepsilon(x-y)|u(x)-u(y)|^2 dxdy + \frac{1}{\varepsilon}\int_D W_k(u)dx,\ u \in L^1(D). \tag{36}$$

The double-well potential W_k forces functions with low energy to stay close to the pure states 0 and 1. This tendency becomes stronger as k increases. The first term resembles the functional TV_ε. In fact if u is the characteristic function of a measurable subset of D then for every $k \in \mathbb{N}$ and for every $\varepsilon > 0$

$$F_\varepsilon^{(k)}(u) = TV_\varepsilon(u). \tag{37}$$

One of the consequences of the results in [4] is that for fixed $k \in \mathbb{N}$, $F_\varepsilon^{(k)}$ Γ-converges in the L^1-metric as $\varepsilon \to 0$ to the functional F^k defined by $\sigma_\eta^k TV(u)$ for all functions $u \in BV(D,\{0,1\})$ and extended to $+\infty$ for all functions belonging to $L^1(D)\setminus BV(D,\{0,1\})$, where σ_η^k is the surface tension associated to η and W_k. We show in Appendix A that σ_η^k converge as $k\to\infty$ to σ_η, defined in (6).

We recall that the functional TV_ε is convex and satisfies the generalized coarea formula:

$$TV_\varepsilon(u) = \int_{-\infty}^{\infty} TV_\varepsilon(\chi_{\{u>s\}})ds, \tag{38}$$

for every $u \in L^1(D)$. The coarea formula is obtained by directly computing:

$$\begin{aligned}
\int_{-\infty}^{\infty} TV_\varepsilon(\chi_{\{u>s\}})ds &= \frac{1}{\varepsilon}\int_{-\infty}^{\infty}\int_{D\times D}\eta_\varepsilon(x-y)|\chi_{\{u>s\}}(x)-\chi_{\{u>s\}}(y)|dxdyds\\
&= \frac{1}{\varepsilon}\int_{D\times D}\eta_\varepsilon(x-y)\int_{-\infty}^{\infty}|\chi_{\{u>s\}}(x)-\chi_{\{u>s\}}(y)|dsdxdy\\
&= \frac{1}{\varepsilon}\int_{D\times D}\eta_\varepsilon(x-y)|u(x)-u(y)|dxdy.
\end{aligned}$$

Theorem 4.1. *For D an open and bounded domain in $\mathbb{R}^d$ with Lipschitz boundary and $\boldsymbol{\eta}$ satisfying (K1)-(K3):*

$$TV_\varepsilon \xrightarrow{\Gamma} \sigma_\eta TV \ \textit{as}\ \varepsilon \to 0, \tag{39}$$

where the Γ limit has to be understood in the L^1-sense.

Proof. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0.

Liminf Inequality: Suppose $u_n \xrightarrow{L^1(D)} u$ as $n\to\infty$.

Step 1: Assume first that for every $n\in\mathbb{N}$, $u_n := \chi_{V_n}$ for some measurable subset V_n of D. In particular $u = \chi_V$ for some V. From the Γ-convergence of the functionals $F_{\varepsilon_n}^{(k)}$ and (37), for every $k\in\mathbb{N}$

$$\liminf_{n\to\infty} TV_{\varepsilon_n}(\chi_{V_n}) = \liminf_{n\to\infty} F_{\varepsilon_n}^{(k)}(\chi_{V_n}) \geq \sigma_\eta^k TV(\chi_V). \tag{40}$$

Since this is true for every k, we can let $k\to\infty$ in the previous equation and from (A.2) obtain:

$$\liminf_{n\to\infty} TV_{\varepsilon_n}(\chi_{V_n}) \geq \sigma_\eta TV(\chi_V) \tag{41}$$

So we conclude that the liminf inequality is true when restricting to characteristic functions.

Step 2: In the general case, we follow the proof in [19]. Indeed, note that:

$$\int_D |u_n(x) - u(x)|\, dx = \int_{-\infty}^{\infty} \int_D \left|\chi_{\{u_n > s\}} - \chi_{\{u > s\}}\right| dxds.$$

So if $u_n \overset{L^1(D)}{\longrightarrow} u$ as $n \to \infty$, then (up to a subsequence) we can assume that for a.e. $s \in \mathbb{R}$. $\chi_{\{u_n > s\}} \overset{L^1(D)}{\longrightarrow} \chi_{\{u > s\}}$. Using Step 1, Fatou's lemma, coarea formula (38) and the coarea formula for the total variational functional we deduce that:

$$\begin{aligned}
\liminf_{n\to\infty} TV_{\varepsilon_n}(u_n) &= \liminf_{n\to\infty} \int_{-\infty}^{\infty} TV_{\varepsilon_n}(\chi_{\{u_n > s\}}) ds \\
&\geq \int_{-\infty}^{\infty} \liminf_{n\to\infty} TV_{\varepsilon_n}(\chi_{\{u_n > s\}}) ds \\
&\geq \sigma_\eta \int_{-\infty}^{\infty} TV(\chi_{\{u > s\}}) ds \\
&= \sigma_\eta TV(u).
\end{aligned}$$

Limsup inequality: We show that for every $u \in L^1(D)$:

$$\limsup_{n\to\infty} TV_{\varepsilon_n}(u) \leq \sigma_\eta TV(u). \tag{42}$$

This is a stronger statement than necessary to prove the limsup inequality needed for Γ-convergence (Definition 2.5). We show this statement since we use it is some of the remarks and results that follow.

It suffices to show (42) for functions $u \in BV(D)$, that is, L^1 functions on D with finite total variation. For $u \in BV(D)$ we use Proposition 3.21 in [5] which says that there exists an extension $\hat{u} \in BV(\mathbb{R}^d)$ of u to the entire space $\mathbb{R}^d$ such that:

$$|D\hat{u}|(\partial D) = 0. \tag{43}$$

Here $D\hat{u}$ denotes the distributional derivative of $\hat{u}$. Since $\hat{u}$ is a BV function, $|D\hat{u}|$ is a finite measure. In case that u belongs to the Sobolev space $W^{1,1}(D)$, $\hat{u}$ can be taken to be in $W^{1,1}(\mathbb{R}^d)$. We split the proof of (42) in two cases.

Step 1: Suppose that η has compact support, i.e. assume there is $\alpha > 0$ such that if $|h| \geq \alpha$ then $\eta(h) = 0$. We define $D_{\varepsilon_n} := \left\{x \in \mathbb{R}^d \,:\, \mathrm{dist}(x, D) < \alpha\varepsilon_n\right\}$. For $\hat{v} \in W^{1,1}(D_{\varepsilon_n}) \cap C^\infty(D_{\varepsilon_n})$ let v be its restriction to D. Then

$$\begin{aligned}
TV_{\varepsilon_n}(v) &= \frac{1}{\varepsilon_n} \int_D \int_{D \cap B(y, \alpha\varepsilon_n)} \eta_{\varepsilon_n}(x-y) |v(x) - v(y)| dxdy \\
&= \frac{1}{\varepsilon_n} \int_{D_{\varepsilon_n}} \int_{B(y,\alpha\varepsilon_n)} \eta_{\varepsilon_n}(x-y) \left| \int_0^1 \nabla\hat{v}(y + t(x-y)) \cdot (x-y) dt \right| dxdy \\
&\leq \frac{1}{\varepsilon_n} \int_D \int_{B(y,\alpha\varepsilon_n)} \int_0^1 \eta_{\varepsilon_n}(x-y) |\nabla\hat{v}(y + t(x-y)) \cdot (x-y)| dtdxdy \\
&\leq \int_{D_{\varepsilon_n}} \int_{\mathbb{R}^d} \int_0^1 \eta(h) |\nabla\hat{v}(z) \cdot h| dtdhdz \\
&= \int_{D_{\varepsilon_n}} \int_{\mathbb{R}^d} \eta(h) |\nabla\hat{v}(z) \cdot h| dhdz \\
&\leq \sigma_\eta \int_{D_{\varepsilon_n}} |\nabla\hat{v}(z)|\, dz,
\end{aligned}$$

where the second inequality is obtained after using the change of variables $(t, y, x) \mapsto (t, h, z)$ where $h = \frac{x-y}{\varepsilon_n}$ and $z = y + t(x-y)$, noting that the Jacobian of this transformation is equal to ε_n^d and that the transformed set D is contained in D_{ε_n}; the last equality obtained thanks to the fact that η is radially symmetric.

Now suppose $u \in BV(D)$ (not necessarily in $W^{1,1}(D)$). By Theorem 13.9 in [31], there exists a sequence $\{\hat{v}_k\}_{k\in\mathbb{N}}$ in $W^{1,1}(D_{\varepsilon_n}) \cap C^\infty(D_{\varepsilon_n})$ such that $\hat{v}_k \overset{L^1(D_{\varepsilon_n})}{\longrightarrow} \hat{u}$ and such that $\int_{D_{\varepsilon_n}} |\nabla \hat{v}_k(z)|\, dz \to |D\hat{u}|(D_{\varepsilon_n})$. This fact, together with continuity of TV_{ε_n} with respect to L^1 convergence, and the previous calculations implies

$$TV_{\varepsilon_n}(u) \le \sigma_\eta |D\hat{u}|(D_{\varepsilon_n}).$$

Sending $n \to \infty$ in the previous expression and using (43) provides that

$$\limsup_{n\to\infty} TV_{\varepsilon_n}(u) \le \sigma_\eta |D\hat{u}|(D) = \sigma_\eta TV(u). \tag{44}$$

Step 2: Consider η whose support is not compact. The control of η at infinity, which is still needed for the statement to hold, is provided by the condition (K3). For $\alpha > 0$ define the kernel $\eta^\alpha(h) := \eta(h)\chi_{B(0,\alpha)}(h)$, which satisfies the conditions of Step 1. Denote by $TV^\alpha_{\varepsilon_n}$ the nonlocal total variation using the kernel η^α. For a given $u \in BV(D)$

$$TV_{\varepsilon_n}(u) = TV^\alpha_{\varepsilon_n}(u) + \frac{1}{\varepsilon_n}\int_D \int_{\{x\in D \,:\, |x-y|>\alpha\varepsilon_n\}} \eta_{\varepsilon_n}(x-y)|u(x)-u(y)|dxdy.$$

The second term on the right-hand side satisfies:

$$\begin{aligned}
\frac{1}{\varepsilon_n}\int_D \int_{\{x\in D \,:\, |x-y|>\alpha\varepsilon_n\}} &\eta_{\varepsilon_n}(x-y)|u(x)-u(y)|dxdy \\
&= \frac{1}{\varepsilon_n}\int_D \int_{\{x\in D \,:\, |x-y|>\alpha\varepsilon_n\}} \eta_{\varepsilon_n}(x-y)|\hat{u}(x)-\hat{u}(y)|dxdy \\
&\le \int_{|h|>\alpha} \eta(h)|h| \int_{\mathbb{R}^d} \frac{|\hat{u}(y)-\hat{u}(y+\varepsilon_n h)|}{\varepsilon_n |h|} dydh \\
&\le |D\hat{u}|(\mathbb{R}^d) \int_{|h|>\alpha} \eta(h)|h|dh,
\end{aligned}$$

where the first inequality is obtained using the change of variables $h = \frac{x-y}{\varepsilon_n}$ and the second inequality obtained using Lemma 13.33 in [31]. By Step 1 we conclude that:

$$\begin{aligned}
\limsup_{n\to\infty} TV_{\varepsilon_n}(u) &\le \limsup_{n\to\infty} TV^\alpha_{\varepsilon_n}(u) + |D\hat{u}|(\mathbb{R}^d) \int_{|h|>\alpha} \eta(h)|h|dh \\
&\le \sigma_{\eta^\alpha} TV(u) + |D\hat{u}|(\mathbb{R}^d) \int_{|h|>\alpha} \eta(h)|h|dh
\end{aligned}$$

Since α was arbitrary, by condition (K3) on $\boldsymbol{\eta}$ we can send α to infinity and obtain (42). □

Remark 4.2. Note that from the previous proof, we can deduce the pointwise convergence of the functionals TV_ε; namely, for every $u \in L^1(D)$:

$$\lim_{\varepsilon\to 0} TV_\varepsilon(u) = \sigma_\eta TV(u)$$

4.1. **Compactness.** We first establish compactness for regular domains and then extend it to more general ones.

Lemma 4.3. *Let D be a bounded open set in $\mathbb{R}^d$, with C^2 boundary. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 and let $\{v_n\}_{n\in\mathbb{N}}$ be a sequence in $L^1(D)$ such that:*

$$\sup_{n\in\mathbb{N}} \|v_n\|_{L^1(D)} < \infty,$$

and

$$\sup_{n\in\mathbb{N}} \frac{1}{\varepsilon_n}\int_D\int_D \eta_{\varepsilon_n}(x-y)|v_n(x)-v_n(y)|dxdy < \infty. \tag{45}$$

Then, $\{v_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(D)$.

Proof. Note that thanks to assumption (K1), we can find $a>0$ and $b>0$ such that the function $\tilde{\boldsymbol{\eta}}:[0,\infty)\to\{0,a\}$ defined as $\tilde{\boldsymbol{\eta}}(t)=a$ for $t<b$ and $\tilde{\boldsymbol{\eta}}(t)=0$ otherwise, is bounded above by $\boldsymbol{\eta}$. In particular, (45) holds when changing η for $\tilde{\eta}$ and so there is no loss of generality in assuming that $\boldsymbol{\eta}$ has the form of $\tilde{\boldsymbol{\eta}}$.

We first extend each function v_n to $\mathbb{R}^d$. Since ∂D is a compact C^2 manifold, there exists $\delta>0$ such that for every $x\in\mathbb{R}^d$ for which $d(x,\partial D)\le\delta$ there exists a unique closest point on ∂D. For all $x\in U:=\{x\in\mathbb{R}^d : d(x,D)<\delta\}$ let Px be the closest point to x in $\overline{D}$. We define the local reflection mapping from U to $\overline{D}$ by $\hat{x}=2Px-x$. Let ξ be a smooth cut-off function such that $\xi(s)=1$ if $s\le\delta/8$ and $\xi(s)=0$ if $s\ge\delta/4$. We define an auxiliary function $\hat{v}_n$ on U, by $\hat{v}_n(x):=v_n(\hat{x})$ and the desired extended function $\tilde{v}_n$ on $\mathbb{R}^d$ by $\tilde{v}_n(x)=\xi(|x-Px|)v_n(\hat{x})$.

We claim that:

$$\sup_{n\in\mathbb{N}} \frac{1}{\varepsilon_n}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \eta_{\varepsilon_n}(x-y)|\tilde{v}_n(x)-\tilde{v}_n(y)| < \infty. \tag{46}$$

To show the claim we first establish the following geometric properties: Let $W:=\{x\in\mathbb{R}^d\backslash D : d(x,D)<\delta/4\}$ and $V:=\{x\in\mathbb{R}^d\backslash D : d(x,D)<\delta/8\}$. For all $x\in W$ and all $y\in D$

$$|\hat{x}-y| < 2|x-y|. \tag{47}$$

Since the mapping $x\to\hat{x}$ is smooth and invertible on W it is bi-Lipschitz. While this would be enough for our argument, we present an argument which establishes the value of the Lipschitz constant: for all $x,y\in W$

$$\frac{1}{4}|x-y| < |\hat{x}-\hat{y}| < 4|x-y|. \tag{48}$$

By definition of δ the domain D satisfies the outside and inside ball conditions with radius δ. Therefore if $x\in W$ and $z\in\overline{D}$

$$\left|z-\left(Px+\delta\frac{x-Px}{|x-Px|}\right)\right| \ge \delta.$$

Squaring and straightforward algebra yield

$$|z-Px|^2 \ge 2\delta(z-Px)\cdot\frac{x-Px}{|x-Px|}. \tag{49}$$

For $x\in W$ and $y\in D$, using (49) we obtain

$$\begin{aligned}|y-\hat{x}|^2-|y-x|^2 &= |y-Px+(x-Px)|^2-|y-Px-(x-Px)|^2\\ &= 4(y-Px)\cdot(x-Px) \le \frac{2}{\delta}|y-Px|^2\,|x-Px|\\ &\le \frac{1}{2}|y-Px|^2 \le |y-x|^2+|x-Px|^2 \le 2|y-x|^2.\end{aligned}$$

Therefore $|y-\hat{x}|^2 \leq 3|y-x|^2$, which establishes (47).

For distinct $x, y \in W$ using (49), with $z = Py$ and with $z = Px$, follows

$$\begin{aligned}|x-y| &\geq (x-y)\cdot\frac{Px-Py}{|Px-Py|} = (x-Px-(y-Py)+Px-Py)\cdot\frac{Px-Py}{|Px-Py|}\\ &\geq |Px-Py| - \frac{1}{2\delta}(|x-Px|\,|Py-Px| + |y-Py|\,|Py-Px|)\\ &\geq |Px-Py|\,\frac{3}{4}.\end{aligned}$$

Therefore

$$|\hat{x}-\hat{y}| = |2Px - x + 2Py - y| \leq 2|Px-Py| + |x-y| \leq \left(\frac{8}{3}+1\right)|x-y| \leq 4|x-y|.$$

Since the roles on x, y and $\hat{x}, \hat{y}$ can be reversed it follows that $|x-y| \leq 4|\hat{x}-\hat{y}|$. Combining the estimates establishes (48).

We now return to proving (46). For ε_n large

$$\begin{aligned}\frac{1}{\varepsilon_n}\int_{\mathbb{R}^n\setminus D}\int_D \eta_{\varepsilon_n}(x-y)|\tilde{v}_n(x)-\tilde{v}_n(y)|dxdy &= \frac{1}{\varepsilon_n}\int_V\int_D \eta_{\varepsilon_n}(x-y)|\hat{v}_n(x)-\hat{v}_n(y)|dxdy\\ &= \frac{1}{\varepsilon_n}\int_V\int_D \eta_{\varepsilon_n}(x-y)|v_n(\hat{x})-v_n(y)|dxdy\\ &\leq \frac{4^d}{\varepsilon_n}\int_V\int_D \eta_{4\varepsilon_n}(\hat{x}-y)|v_n(x)-v_n(\hat{y})|dxdy\\ &\leq \frac{16^d}{\varepsilon_n}\int_D\int_D \eta_{4\varepsilon_n}(z-y)|v_n(x)-v_n(z)|dzdy,\end{aligned}$$

where the first inequality follows from (47) and the second follows from the fact that the change of variables $x \mapsto \hat{x}$ is bi-Lipschitz as shown in (48). Also,

$$\begin{aligned}\frac{1}{\varepsilon_n}\int_{\mathbb{R}^d\setminus D}\int_{\mathbb{R}^d\setminus D} \eta_{\varepsilon_n}(x-y)|\tilde{v}_n(x)-\tilde{v}_n(y)|dxdy =& \frac{1}{\varepsilon_n}\int_W\int_W \eta_{\varepsilon_n}(x-y)|\xi(x)\hat{v}_n(x)-\xi(y)\hat{v}_n(y)|dxdy\\ \leq& \frac{1}{\varepsilon_n}\int_W\int_W \eta_{\varepsilon_n}(x-y)|\xi(x)-\xi(y)||\hat{v}_n(x)|dxdy\\ &+\frac{1}{\varepsilon_n}\int_W\int_W \eta_{\varepsilon_n}(x-y)|\hat{v}_n(x)-\hat{v}_n(y)||\xi(y)|dxdy.\end{aligned}$$

Note that for all $x \neq y$, we have $\frac{\eta_{\varepsilon_n}(x-y)}{\varepsilon_n} \leq \frac{b}{|x-y|}\eta_{\varepsilon_n}(x-y)$. Therefore:

$$\begin{aligned}\frac{1}{\varepsilon_n}\int_W\int_W \eta_{\varepsilon_n}(x-y)|\xi(x)-\xi(y)||\hat{v}_n(x)|dxdy &\leq b\int_W\int_W \eta_{\varepsilon_n}(x-y)\frac{|\xi(x)-\xi(y)|}{|x-y|}|\hat{v}_n(x)|dxdy\\ &\leq b\,\mathrm{Lip}(\xi)\int_W\int_W \eta_{\varepsilon_n}(x-y)|\hat{v}_n(x)|dxdy\\ &\leq 4^d\,b\,\mathrm{Lip}(\xi)\|v_n\|_{L^1(D)},\end{aligned}$$

where we used (48) and change of variables to establish the last inequality. Also,

$$\begin{aligned}\frac{1}{\varepsilon_n}\int_W\int_W \eta_{\varepsilon_n}(x-y)|\hat{v}_n(x)-\hat{v}_n(y)||\xi(y)|dxdy &\leq \frac{4^d}{\varepsilon_n}\int_W\int_W \eta_{4\varepsilon_n}(\hat{x}-\hat{y})|\hat{v}_n(x)-\hat{v}_n(y)|dxdy\\ &\leq \frac{4^{3d}}{\varepsilon_n}\int_D\int_D \eta_{4\varepsilon_n}(x-y)|v_n(x)-v_n(y)|dxdy.\end{aligned}$$

The first inequality is obtained thanks to the fact that $|\xi(y)| \leq 1$ and (48), while the second inequality is obtained by a change of variables.

Using that

$$\int_D \int_D \eta_{4\varepsilon_n}(x-y)|v_n(x)-v_n(y)|dxdy \leq 4^d \int_D \int_D \eta_{\varepsilon_n}(x-y)|v_n(x)-v_n(y)|dxdy$$

by combining the above inequalities we conclude that

$$\begin{aligned}\sup_{n\in\mathbb{N}} \frac{1}{\varepsilon_n} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \eta_{\varepsilon_n}(x-y)|\tilde{v}_n(x)-\tilde{v}_n(y)| \\ \leq C \sup_{n\in N}\left(\int_D\int_D \eta_{\varepsilon_n}(x-y)|v_n(x)-v_n(y)|dxdy + \|v_n\|_{L^1(D)}\right) < \infty.\end{aligned}$$

Using the proof of Proposition 3.1 in [3] we deduce that the sequence $\{\tilde{v}_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(\mathbb{R}^d)$ which implies that the sequence $\{v_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(D)$. $\square$

Remark 4.4. We remark that the difference between the compactness result we proved above and the one proved in Proposition 3.1 in [3] is the fact that we consider functions bounded in L^1, instead of bounded in L^∞ as was assumed in [3]. Nevertheless, after extending the functions to the entire $\mathbb{R}^d$ as above, one can directly apply the proof in [3] to obtain the desired compactness result.

Proposition 4.5. *Let D be a bounded open set in $\mathbb{R}^d$, which is homeomorphic to a C^2 domain $\tilde{D}$ via a bi-Lipschits mapping Θ. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 and suppose that the sequence of functions $\{u_n\}_{n\in\mathbb{N}} \subseteq L^1(D)$ satisfies:*

$$\sup_{n\in\mathbb{N}} \|u_n\|_{L^1(D)} < \infty,$$
$$\sup_{n\in\mathbb{N}} TV_{\varepsilon_n}(u_n) < \infty.$$

Then, $\{u_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(D)$.

We remark that the proposition applies to $D = (0,1)^d$, for example by using line-wise linear rescalling.

Proof. Suppose $\{u_n\}_{n\in\mathbb{N}} \subseteq L^1(D)$ is as in the statement. As in Lemma 4.3, we can assume that $\boldsymbol{\eta}$ is of the form $\boldsymbol{\eta}(t) = a$ if $t < b$ and $\boldsymbol{\eta}(t) = 0$ for $t \geq b$, where a and b are two positive constants. For every $n \in \mathbb{N}$ consider the function $v_n := u_n \circ \Theta$ and set $\hat{\boldsymbol{\eta}}(s) := \boldsymbol{\eta}\left(\mathrm{Lip}(\Theta)\, s\right)$, $s \in \mathbb{R}$.

Since Θ is bi-Lipchitz we can use a change of variables, to conclude that there exists a constant $C > 0$ (only depending on Θ) such that:

$$\int_{\tilde{D}} |v_n(x)|dx \leq C \int_D |u_n(y)|dy,$$

and

$$\begin{aligned}\frac{C}{\varepsilon_n}\int_D\int_D \eta_{\varepsilon_n}(x-y)\,|u_n(x)-u_n(y)|\,dxdy &\geq \frac{1}{\varepsilon_n}\int_{\tilde{D}}\int_{\tilde{D}} \eta_{\varepsilon_n}\left(\Theta(x)-\Theta(y)\right)|v_n(x)-v_n(y)|\,dxdy \\ &\geq \frac{1}{\varepsilon_n}\int_{\tilde{D}}\int_{\tilde{D}} \hat{\eta}_{\varepsilon_n}(x-y)\,|v_n(x)-v_n(y)|\,dxdy.\end{aligned}$$

We conclude that the sequence $\{v_n\}_{n\in\mathbb{N}} \subseteq L^1(\tilde{D})$ satisfies the hypothesis of Lemma 4.3 (taking $\boldsymbol{\eta} := \hat{\boldsymbol{\eta}}$). Therefore, $\{v_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(\tilde{D})$, which implies that $\{u_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(D)$. $\square$

Corollary 4.6. *Let D be a bounded open set in $\mathbb{R}^d$. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 and suppose that the sequence of functions $\{u_n\}_{n\in\mathbb{N}} \subseteq L^1(D)$ satisfies:*

$$\sup_{n\in\mathbb{N}} \|u_n\|_{L^1(D)} < \infty,$$
$$\sup_{n\in\mathbb{N}} TV_{\varepsilon_n}(u_n) < \infty.$$

Then, $\{u_n\}_{n\in\mathbb{N}}$ is locally precompact in $L^1(D)$.

In particular if

$$\sup_{n\in\mathbb{N}} \|u_n\|_{L^\infty(D)} < \infty,$$

then, $\{u_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(D)$.

Proof. If B is a ball compactly contained in D then the pre compactness of $\{u_n\}_{n\in\mathbb{N}}$ in $L^1(B)$ follows from Lemma 4.3, using that $TV_{\varepsilon_n}(u_n|_B) \leq TV_{\varepsilon_n}(u_n)$. We note that if compactness holds on two sets D_1 and D_2 compactly contained in D, then it holds on their union. Therefore it holds on any set compactly contained in D, since it can be covered by finitely many balls contained in D.

The compactness in $L^1(D)$ under the L^∞ boundedness follows via a diagonal argument. This can be achieved by approximating D by compact subsets: $\overline{D}_k \subset D$, $D = \cup_k D_k$, and using that $\lim_{k\to\infty} \sup_{n\in\mathbb{N}} \|u_n\|_{L^1(D\setminus D_k)} = 0$. □

5. Gamma Convergence of Total Variation on Graphs

5.1. **Proof Main result.** Let $D = (0,1)^d$ and let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 satisfying (8).

Proof of Theorem 1.1. We use the sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ considered in Section 2.2. We work with $\omega \in \Omega$ for which (26), (24), (25) in case $d = 1$, $d = 2$ and $d \geq 3$ respectively. By the results in Section 2.2. the complement in Ω of such ω's is contained in a set of probability zero.

Step 1: Suppose first that $\boldsymbol{\eta}$ is of the form $\boldsymbol{\eta}(t) = a$ for $t < b$ and $\boldsymbol{\eta} = 0$ for $t > b$, where a, b are two positive constants. Note it does not matter what value we give to $\boldsymbol{\eta}$ at b.

Liminf inequality: Assume that $u_n \xrightarrow{TL^1} u$ as $n \to \infty$. Since $T_{n\sharp}\nu_0 = \nu_n$, using the change of variables (15) it follows that

$$\widetilde{TV}_{n,\varepsilon_n}(u_n) = \frac{1}{\varepsilon_n} \int_{D\times D} \eta_{\varepsilon_n}\left(T_n(x) - T_n(y)\right) |u_n \circ T_n(x) - u_n \circ T_n(y)|\, dxdy. \tag{50}$$

The key idea in the proof is that the estimates of the Section 2.2 on transportation maps provide 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 graph total variation. In particular, note that for Lebesgue almost every $(x, y) \in D \times D$

$$|T_n(x) - T_n(y)| > b\varepsilon_n \Rightarrow |x - y| > b\varepsilon_n - 2\|Id - T_n\|_\infty. \tag{51}$$

Thanks to the assumptions on $\{\varepsilon_n\}_{n\in\mathbb{N}}$ and (26), (24), (25) in case $d = 1$, $d = 2$ and $d \geq 3$ respectively, for large enough $n \in \mathbb{N}$:

$$\tilde{\varepsilon}_n := \varepsilon_n - \frac{2}{b}\|Id - T_n\|_\infty > 0.$$

By (51), for large enough n and for almost every $(x, y) \in D \times D$,

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

Let $\tilde{u}_n = u_n \circ T_n$. Thanks to the previous inequality and (50), for large enough n

$$\begin{aligned}\widetilde{TV}_{n,\varepsilon_n}(u_n) \geq &\frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \boldsymbol{\eta}\left(\frac{|x-y|}{\tilde{\varepsilon}_n}\right) |\tilde{u}_n(x) - \tilde{u}_n(y)|\, dxdy \\ = &\left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{d+1} TV_{\tilde{\varepsilon}_n}(\tilde{u}_n).\end{aligned}$$

Note that $\frac{\tilde{\varepsilon}_n}{\varepsilon_n} \to 1$ as $n \to \infty$ and that $u_n \xrightarrow{TL^1} u$ implies $\tilde{u}_n \xrightarrow{L^1(D)} u$ as $n \to \infty$. We deduce from Theorem 4.1 that $\liminf_{n\to\infty} TV_{\tilde{\varepsilon}_n}(\tilde{u}_n) \geq \sigma_\eta TV(u)$ and hence:

$$\liminf_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n) \geq \sigma_\eta TV(u).$$

Limsup inequality: By Remark 2.6 it is enough to consider Lipschitz continuous functions $u : D \to \mathbb{R}$. Define u_n to be the restriction of u to the first n data points $X_1, \dots, X_n$. Consider $\tilde{\varepsilon}_n := \varepsilon_n + \frac{2}{b}\|Id - T_n\|_\infty$ and let $\tilde{u}_n = u_n \circ T_n$. Then note that for Lebesgue almost every $(x,y) \in D\times D$

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

Then for all n

$$\begin{aligned}&\frac{1}{\tilde{\varepsilon}_n^{d+1}} \int_{D\times D} \boldsymbol{\eta}\left(\frac{|T_n(x) - T_n(y)|}{\varepsilon_n}\right) |\tilde{u}_n(x) - \tilde{u}_n(y)|\, dxdy \\ &\qquad\qquad \leq \frac{1}{\tilde{\varepsilon}_n} \int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)\, |\tilde{u}_n(x) - \tilde{u}_n(y)|\, dxdy.\end{aligned} \tag{52}$$

Also

$$\begin{aligned}\frac{1}{\tilde{\varepsilon}_n}&\left|\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|u(x) - u(y)|\; dxdy - \int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|u\circ T_n(x) - u\circ T_n(y)| dxdy\right| \\ &\leq \frac{2}{\tilde{\varepsilon}_n} \int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|u(x) - u\circ T_n(x)| dxdy \\ &\leq \frac{2C\,\mathrm{Lip}(u)}{\tilde{\varepsilon}_n} \int_D |x - T_n(x)| dx,\end{aligned} \tag{53}$$

where $C = \int_{\mathbb{R}^d} \eta(h)dh$. The last term of the last expression goes to 0 as $n \to \infty$, yielding

$$\lim_{n\to\infty} \frac{1}{\tilde{\varepsilon}_n}\left(\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|u(x) - u(y)| dxdy - \int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|u\circ T_n(x) - u\circ T_n(y)| dxdy\right) = 0$$

Since $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$ as $n \to \infty$, using (52) we deduce :

$$\begin{aligned}\limsup_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n) = &\limsup_{n\to\infty} \frac{1}{\tilde{\varepsilon}_n^{d+1}} \int_{D\times D} \boldsymbol{\eta}\left(\frac{|T_n(x) - T_n(y)|}{\varepsilon_n}\right) |u\circ T_n(x) - u\circ T_n(y)|\, dxdy \\ \leq &\limsup_{n\to\infty} \frac{1}{\tilde{\varepsilon}_n} \int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)\, |u\circ T_n(x) - u\circ T_n(y)|\, dxdy \\ = &\limsup_{n\to\infty} TV_{\tilde{\varepsilon}_n}(u) \leq \sigma_\eta TV(u),\end{aligned}$$

where the last inequality follows from the proof of Theorem 4.1, specifically inequality (42).

Step 2: Now consider $\boldsymbol{\eta}$ to be a piecewise constant function with compact support, satisfying (K1)-(K3). In this case, we can write $\boldsymbol{\eta} = \sum_{k=1}^{l} \boldsymbol{\eta}_k$ for some l and functions $\boldsymbol{\eta}_k$ as in Step 1. For this step of the proof we denote by $\widetilde{TV}^k_{n,\varepsilon_n}$ the total variation function on the graph using the function $\boldsymbol{\eta}_k$.

Liminf inequality: Assume that $u_n \xrightarrow{TL^1} u$ as $n \to \infty$. By Step 1:

$$\liminf_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n) = \liminf_{n\to\infty} \sum_{k=1}^{l} \widetilde{TV}^k_{n,\varepsilon_n}(u_n) \geq \sum_{k=1}^{l} \liminf_{n\to\infty} \widetilde{TV}^k_{n,\varepsilon_n}(u_n) \geq \sum_{k=1}^{l} \sigma_{\eta_k} TV(u) = \sigma_\eta TV(u)$$

Limsup inequality: By Remark 2.6 it is enough to prove the limsup inequality for $u : D \to \mathbb{R}$ Lipschitz. Consider u_n as in the proof of the limsup inequality in Step 1. Then

$$\begin{aligned}
\limsup_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n) &= \limsup_{n\to\infty} \sum_{k=1}^{l} \widetilde{TV}^k_{n,\varepsilon_n}(u_n) \\
&\leq \sum_{k=1}^{l} \limsup_{n\to\infty} \widetilde{TV}^k_{n,\varepsilon_n}(u_n) \\
&\leq \sum_{k=1}^{l} \sigma_{\eta_k} TV(u) = \sigma_\eta TV(u)
\end{aligned}$$

Step 3: Assume $\boldsymbol{\eta}$ is compactly supported and satisfies (K1)-(K3).

Liminf Inequality: Note that we can find an increasing sequence of piecewise constant functions $\boldsymbol{\eta}_k : [0,\infty) \to [0,\infty)$ (η from Step 2 is used as η_k here), with $\boldsymbol{\eta}_k \nearrow \boldsymbol{\eta}$ as $k \to \infty$ a.e. Denote by $\widetilde{TV}^k_{n,\varepsilon_n}$ the graph TV corresponding to $\boldsymbol{\eta}_k$. If $u_n \xrightarrow{TL^1} u$ as $n \to \infty$, by Step 2 $\sigma_{\eta_k} TV(u) \leq \liminf_{n\to\infty} \widetilde{TV}^k_{n,\varepsilon_n}(u_n) \leq \liminf_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n)$ for every $k \in \mathbb{N}$. The monotone convergence theorem implies that $\lim_{k\to\infty} \sigma_{\eta_k} = \sigma_\eta$ and so we conclude that $\sigma_\eta TV(u) \leq \liminf_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n)$.

Limsup inequality: As in Steps 1 and 2 it is enough to prove the limsup inequality for u Lipschitz. Consider u_n as in the proof of the limsup inequality in Steps 1 and 2. Analogously to the proof of the liminf inequality, we can find a decreasing sequence of functions $\boldsymbol{\eta}_k : [0,\infty) \to [0,\infty)$ (of the form considered in Step 2), with $\boldsymbol{\eta}_k \searrow \boldsymbol{\eta}$ as $k \to \infty$ a.e. Proceeding in an analogous way to the way we proceeded in the proof of the liminf inequality we can conclude that $\limsup_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n) \leq \sigma_\eta TV(u)$.

Step 4: Consider general $\boldsymbol{\eta}$, satisfying (K1)-(K3). Note that for the liminf inequality we can use the proof given in Step 3. For the limsup inequality, as in the previous steps we can assume that u is Lipschitz and we take u_n as in the previous steps. Let $\alpha > 0$ and define $\boldsymbol{\eta}_\alpha : [0,\infty) \to [0,\infty)$ by $\boldsymbol{\eta}_\alpha(t) := \boldsymbol{\eta}(t)$ for $t \leq \alpha$ and $\boldsymbol{\eta}_\alpha(t) = 0$ for $t > \alpha$. We denote by $\widetilde{TV}^\alpha_{n,\varepsilon_n}$ the graph TV using $\boldsymbol{\eta}_\alpha$.

Note that:

$$\begin{aligned}
\widetilde{TV}_{n,\varepsilon_n}(u_n) =& \widetilde{TV}^\alpha_{n,\varepsilon_n}(u_n) \\
&+ \frac{1}{\varepsilon_n^{d+1}} \int_{|T_n(x)-T_n(y)|>\alpha\varepsilon_n} \boldsymbol{\eta}\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon_n}\right) |u \circ T_n(x) - u \circ T_n(y)| \, dxdy
\end{aligned} \tag{54}$$

Let us find bounds on the second term on the right hand side of the previous equality for large n. Indeed since for almost every $(x,y) \in D \times D$ it is true that $|x-y| \leq |T_n(x)-T_n(y)| + 2\|Id - T_n\|_\infty$ and $|T_n(x) - T_n(y)| \leq |x-y| + 2\|Id - T_n\|_\infty$ we can use the fact that $\frac{\|Id-T_n\|_\infty}{\varepsilon_n} \to 0$ as $n \to \infty$ to conclude that for large enough n, for almost every $(x,y) \in D \times D$ for which $|T_n(x) - T_n(y)| > \alpha\varepsilon_n$ it holds that $|x-y| \leq 2|T_n(x) - T_n(y)|$ and $|T_n(x) - T_n(y)| \leq 2|x-y|$. From this we can conclude that for large enough n

$$\begin{aligned}\frac{1}{\varepsilon_n^{d+1}}\int_{|T_n(x)-T_n(y)|>\alpha\varepsilon_n}&\boldsymbol{\eta}\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon_n}\right)|u\circ T_n(x)-u\circ T_n(y)|\,dxdy\\ &\leq\frac{1}{\varepsilon_n^{d+1}}\int_{|x-y|>\alpha\varepsilon_n/2}\boldsymbol{\eta}\left(\frac{|x-y|}{2\varepsilon_n}\right)|u\circ T_n(x)-u\circ T_n(y)|\,dxdy\\ &\leq\frac{2\,\mathrm{Lip}(u)}{\varepsilon_n^{d+1}}\int_{|x-y|>\alpha\varepsilon_n/2}\boldsymbol{\eta}\left(\frac{|x-y|}{2\varepsilon_n}\right)|x-y|dxdy\end{aligned}$$

To find bounds on the last term of the previous chain of inequalities, consider the change of variables $(x,y)\in D\times D\mapsto(x,h)$ where $x=x$ and $h=\frac{x-y}{2\varepsilon_n}$, we deduce that:

$$\frac{2}{\varepsilon_n^{d+1}}\int_{|x-y|>\alpha\varepsilon_n/2}\boldsymbol{\eta}\left(\frac{|x-y|}{2\varepsilon_n}\right)|x-y|dxdy\leq C\int_{|h|>\frac{\alpha}{4}}\eta(h)|h|dh,$$

where C does not depend on n or α. From the previous inequalities, (54) and Step 3 we deduce that

$$\begin{aligned}\limsup_{n\to\infty}\widetilde{TV}_{n,\varepsilon_n}(u_n)\leq&\limsup_{n\to\infty}\widetilde{TV}^{\alpha}_{n,\varepsilon_n}(u_n)+\mathrm{Lip}(u)C\int_{|h|>\frac{\alpha}{4}}\eta(h)|h|dh\\ \leq&\sigma_{\eta_\alpha}TV(u)+\mathrm{Lip}(u)C\int_{|h|>\frac{\alpha}{4}}\eta(h)|h|dh\end{aligned}$$

Finally, given the assumptions (K1)-(K3) on η, sending α to infinity we conclude that

$$\limsup_{n\to\infty}\widetilde{TV}_{n,\varepsilon_n}(u_n)\leq\sigma_\eta TV(u)$$

□

We now present the proof of Theorem 1.2 on compactness.

Proof. Assume that $\{u_n\}_{n\in N}$ is a sequence of functions with $u_n\in L^1(D,\nu_n)$ satisfying the assumptions of the theorem. As in Lemma 4.3 and Proposition 4.5 without loss of generality we can assume that $\boldsymbol{\eta}$ is of the form $\boldsymbol{\eta}(t)=a$ if $t<b$ and $\boldsymbol{\eta}(t)=0$ for $t\geq b$, for some a and b positive constants.

Consider the sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from Section 2.2. Since $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfies (8), estimates (24), (25), and (26), imply that for Lebesgue a.e. $z,y\in D$ with $|T_n(z)-T_n(y)|>b\varepsilon_n$ it hold that $|z-y|>b\varepsilon_n-2\|Id-T_n\|_\infty$. For large enough n, we set $\tilde{\varepsilon}_n:=\varepsilon_n-\frac{2\|Id-T_n\|_\infty}{b}>0$. We conclude that for large n and Lebesgue a.e. $z,y\in D$:

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

Using this, we can conclude that for large enough n:

$$\begin{aligned}\frac{1}{\varepsilon_n^{d+1}}\int_D\int_D&\boldsymbol{\eta}\left(\frac{|z-y|}{\tilde{\varepsilon}_n}\right)|u_n\circ T_n(z)-u_n\circ T_n(y)|\,dzdy\\ &\leq\frac{1}{\varepsilon_n^{d+1}}\int_D\int_D\boldsymbol{\eta}\left(\frac{|T_n(z)-T_n(y)|}{\tilde{\varepsilon}_n}\right)|u_n\circ T_n(z)-u_n\circ T_n(y)|\,dzdy\\ &=\widetilde{TV}_{n,\varepsilon_n}(u_n).\end{aligned}$$

Thus

$$\sup_{n\in\mathbb{N}}\frac{1}{\varepsilon_n^{d+1}}\int_D\int_D\boldsymbol{\eta}\left(\frac{|z-y|}{\tilde{\varepsilon}_n}\right)|u_n\circ T_n(z)-u_n\circ T_n(y)|\,dzdy<\infty.$$

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

$$\sup_{n\in\mathbb{N}} \frac{1}{\tilde{\varepsilon}_n} \int_D \int_D \eta_{\tilde{\varepsilon}_n}\left(z-y\right) \left|u_n \circ T_n(z) - u_n \circ T_n(y)\right| dzdy < \infty.$$

By Proposition 4.5 we conclude that $\{u_n \circ T_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(D)$ and hence $\{u_n\}_{n\in\mathbb{N}}$ is precompact in TL^1. □

Finally, to prove Corollary 1.3 on the Γ convergence of perimeter, note that if $\{A_n\}_{n\in\mathbb{N}}$ is such that $A_n \subseteq \{X_1, \ldots, X_n\}_{n\in\mathbb{N}}$ and $\chi_{A_n} \xrightarrow{TL^1} \chi_A$ as $n \to \infty$ for some $A \subseteq D$, then the liminf inequality follows automatically from the liminf inequality in Theorem 1.1. The limsup inequality is not immediate, since we cannot use the density of Lipschitz functions as we did in the proof of Theorem 1.1 given that we restrict our attention to characteristic functions.

We follow the proof of Proposition 3.5 in [19] and take advantage of the coarea formula of the energies $\widetilde{TV}_{n,\varepsilon_n}$. In fact, consider a measurable subset A of D. By the limsup inequality in Theorem 1.1, we know there exists a sequence $\{u_n\}_{n\in\mathbb{N}}$ (with $u_n \in L^1(\nu_n)$) such that $\limsup_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(u_n) \leq \sigma_\eta TV(\chi_A)$. It is easy to check that the functionals $\widetilde{TV}_{n,\varepsilon_n}$ satisfy the coarea formula:

$$\widetilde{TV}_{n,\varepsilon_n}(u_n) = \int_{-\infty}^{\infty} \widetilde{TV}_{n,\varepsilon_n}(\chi_{\{u_n>s\}})ds.$$

Fix $0 < \delta < \frac{1}{2}$. Then in particular we have:

$$\int_{\delta}^{1-\delta} \widetilde{TV}_{n,\varepsilon_n}(\chi_{\{u_n>s\}})ds \leq \widetilde{TV}_{n,\varepsilon_n}(u_n).$$

For every n we can find $s_n \in (\delta, 1-\delta)$ such that $\widetilde{TV}_{n,\varepsilon_n}(\chi_{\{u_n>s_n\}}) \leq \frac{1}{1-2\delta}\widetilde{TV}_{n,\varepsilon_n}(u_n)$. Define $A_n^\delta := \{u_n > s_n\}$. It is straightforward to see that $\chi_{A_n^\delta} \xrightarrow{TL^1} \chi_A$ as $n \to \infty$ and that $\limsup_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(A_n^\delta) \leq \sigma_\eta TV(\chi_A)$. We can now take δ arbitrarily close to 0 and use a diagonal argument to obtain sets $\{A_n\}_{n\in\mathbb{N}}$ such that $\chi_{A_n} \xrightarrow{TL^1} \chi_A$ as $n \to \infty$ and $\limsup_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(\chi_{A_n}) \leq \sigma_\eta TV(\chi_A)$.

Remark 5.1. There is an alternative proof of the limsup inequality above. It is possible to proceed in a similar fashion as in the proof of the limsup inequality in Theorem 1.1. In this case, instead of approximating by Lipschitz functions, one would consider the characteristic functions of $E \cap D$ where E is a subset of $\mathbb{R}^d$ with smooth boundary. As in the proof of Theorem 1.1, the key is to show that for step kernels ($\eta(r) = b$ if $r < a$ and zero otherwise), and sets $G := E \cap D$

$$\lim_{n\to\infty} \widetilde{TV}_{n,\varepsilon_n}(\chi_G) = TV(\chi_G).$$

To do so one needs a substitute for estimate (53). The needed estimate follows from the estimate

$$\int_D |\chi_G(x) - \chi_G(T_n(x))|dx \lesssim \mathrm{Per}(G : D)\, ||Id - T_n||_\infty$$

where $\mathrm{Per}(G : D)$ is the relative perimeter of G in D. We do not prove this estimate, but remark that it has a straightforward proof. Finally, noting that this class of sets is dense with respect to perimeter (see Remark 3.42 in [5]) and using Remark 2.6 we obtain the limsup inequality for the characteristic function of any measurable set.

5.2. **Extension to different sets of points.** Consider $D = (0,1)^d$. Instead of random, uniformly distributed, independently chosen points $X_1, \dots, X_n, \dots$ one can consider different sets of points. The only requirement is that one has estimates on how well the sequence of points approximates the Lebesgue measure on D. In particular one would need to obtain analogues of estimates (24), (25), and (26). That is one would need to show that

$$\limsup_{n\to\infty} \frac{n^{1/d}\|Id - T_n\|_\infty}{f(n)} \leq C \tag{55}$$

for some nonnegative function $f : \mathbb{N} \to (0,\infty)$. The map T_n in this context transports the Lebesgue measure ν_0 to the empirical measure determined by the new set of points (not necessarily randomly chosen with uniform distribution).

We remark that f must be bounded from below, since for any collection $V = \{X_1, \dots, X_n\}$ in D

$$\sup_{y\in D} \operatorname{dist}(y, V) \geq cn^{-1/d}$$

and thus $n^{1/d}\|Id - T_n\|_\infty \geq c$.

If (55) holds and if ε_n is a positive sequence converging to zero such that $\lim_{n\to\infty} \frac{f(n)}{n^{1/d}} \frac{1}{\varepsilon_n} = 0$ then the conclusion of the Theorem 1.1 holds. We note that almost no modifications to the proof are needed.

One special case is when we consider $X_1, \dots, X_n, \dots$ a sequence of grid points on diadicaly refining grids. In this case one can take $f(n) = 1$ for all n and $\varepsilon_n \to 0$ such that $\lim_{n\to\infty} \frac{1}{n^{1/d}\varepsilon_n} = 0$. Note that our results imply Γ-convergence in the TL^1 metric, however in this particular case, this is equivalent to the L^1 -metric considered in [19] and [11] where for a function defined on the grid points we associate a function defined on D by simply setting the function to be constant on the grid cells. This follows from Proposition 3.6.

5.3. **Extension to more general domains.** The purpose of this subsection is to extend the results obtained for the cube $(0,1)^d$ to more general bounded domains D with continuous boundary. To frame our approach consider $X'_1, \dots, X'_n$ to be independent samples of the uniform distribution D. Let ν'_n to be the empirical measure associated to the data points and ν'_0 the Lebesgue measure on D, rescaled by $\mathrm{Vol}(D)$. If one can find a sequence of transportation maps $T'_n : D \to D$ with $T'_{n\sharp}\nu'_0 = \nu'_n$ and obtain an upper bound on how fast $||Id - T'_n||_\infty$ converges to zero, then one can find a scaling of ε with respect to n for which the rescaled total variation on the graphs determined by the points $X'_1, X'_2, \dots$ Γ-converges (in the TL^1 metric) to $\frac{\sigma_\eta}{\mathrm{Vol}(D)} TV$ as $n \to \infty$. Here TV denotes the total variation on D. To obtain this result it is enough to imitate the proof of Theorem 1.1 in Subsection 5.1, noting that almost no modifications to the proof are required.

Thus the goal is to determine for general domains how well does the empirical measure approximate the Lebesgue measure. While this question deserves a careful investigation, here we only present two simple criteria, which still apply to a broad class of domains for which estimates on $||Id - T'_n||_\infty$ are as good as those for the domain $(0,1)^d$.

Proposition 5.2. *Let D be an open and bounded domain with Lipschitz boundary. Let ν'_0 be the Lebesgue measure on D rescaled by* $\mathrm{Vol}(D)$. *Let $X'_1, \dots, X'_n, \dots$ be a sequence of i.i.d. random vectors chosen uniformly on D and let $\nu'_n = \frac{1}{n}\sum_{i=1}^n \delta_{X'_i}$ be the corresponding empirical measures. Assume that*

(i) *there exists a volume preserving bi-Lipschitz mapping $\Theta : (0,L)^d \to D$ for some $L > 0$, or*
(ii) *D is diffeomorphic to $B(0,1)$ via a $C^{1,\alpha}$ mapping $\Theta : D \to B(0,1)$.*

Then, there exists a sequence $\{T'_n\}_{n\in\mathbb{N}}$ of transportation maps $T'_{n\sharp}\nu'_0 = \nu'_n$ and such that $\|Id - T'_n\|_\infty$ satisfies (26), (24), (25) *for $d = 1$, $d = 2$ and $d \geq 3$ respectively. In particular Theorems 1.1 and 1.2 hold for the domain D.*

Remark 5.3. It was shown in [23], and an explicit construction was given in [25], that there exists a bi-Lipschitz volume preserving map between the unit ball and the unit cube (after rescaling by the volume of the unit ball). Thus the case (i) of Lemma above applies to $D = B(0,1)$.

Proof. Consider the case that (i) holds. By rescaling, we can assume that $L = 1$. The points $X_i := \Theta(X_i')$ are independent and uniformly distributed on $(0,1)^d$. Using the maps T_n of Subsection 2.2 we define $T_n' := \Theta^{-1} \circ T_n \circ \Theta$. Since Θ is bi-Lipschitz, the maps T_n' are such that $||Id - T_n'||_\infty$ has the same rate of convergence to zero as $||Id - T_n||_\infty$. This in particular implies that Theorem 1.1 is also true when considering D, with the same scaling for ε_n. The compactness also follows.

The proof in the case that (ii) holds depends on Lemma 5.4 we present below. We use the diffeomorphism Θ to obtain a probability measure $\hat{\mu} := \left(\Theta^{-1}\right)_\sharp \nu_0'$ defined on $B(0,1)$. Since Θ is a $C^{1,\alpha}$ diffeomorphism, $\hat{\mu}$ is absolutely continuous with respect to the Lebesgue measure and $\rho := \frac{d\hat{\mu}}{dx}$ is C^α and bounded below by a positive constant. Note that the points $\hat{X}_i := \left(\Theta^{-1}\right)(X_i')$ are independent and distributed according to $\hat{\mu}$. We can obtain maps $\hat{T}_n$ as in Lemma 5.4 and check that the maps $T_n' := \Theta \circ \hat{T}_n \circ \Theta^{-1}$ satisfy the desired properties. □

Lemma 5.4. *Let $\rho : \overline{B(0,1)} \to (0,\infty)$ be a Hölder continuous function such that $\int_{B(0,1)} \rho(x)dx = 1$. Let $\hat{X}_1, \dots, \hat{X}_n, \dots$ be i.i.d. random vectors on $B(0,1)$ chosen according to the density ρ. Denote by $\hat{\mu}_n := \frac{1}{n}\sum_{i=1}^n \delta_{\hat{X}_i}$ the empirical measure associated to the first n data points and by $\hat{\mu}$ the measure with density ρ: $d\hat{\mu}(x) = \rho(x)dx$. Then, there exists a sequence $\hat{T}_n : B(0,1) \to B(0,1)$ of (random) transportation maps with $\hat{T}_{n\sharp}\hat{\mu} = \hat{\mu}_n$ and such that $\|Id - \hat{T}_n\|_\infty$ satisfy the analogous estimates to* (26)*,* (24)*,* (25) *for $d = 1$, $d = 2$ and $d \geq 3$ respectively.*

Proof. **Step 1:** If ρ is identically equal to $1/\operatorname{Vol}(B(0,1))$, we can use Proposition 5.2 case (i), and Remark 5.3 to deduce the result.

Step 2: Denote by $\hat{\nu}$ the Lebesgue measure on $B(0,1)$ rescaled by $\operatorname{Vol}(B(0,1))$. By Brenier's theorem (see [48]) there exists a convex function Φ on $B(0,1)$ such that $\nabla\Phi_\sharp\hat{\nu} = \hat{\mu}$. Furthermore, $\nabla\Phi$ is the unique optimal transportation map for the quadratic cost, that is $\nabla\Phi$ minimizes (18) with $p = 2$ (where $\hat{\mu}$ plays the role of ν and $\hat{\nu}$ the role of μ). The function Φ, called the transportation potential, satisfies a Monge–Ampere equation. The deep boundary regularity results for solutions of the Monge–Ampere equation by Caffarelli [18] (see also [17], and Urbas [46]) imply that $\Phi \in C^{2,\beta}$ up to the boundary, for some $\beta > 0$. Therefore the transportation map $\nabla\Phi$ is $C^{1,\beta}$ up to the boundary and thus Lipschitz. By a similar reasoning, $(\nabla\Phi)^{-1}$ is Lipschitz as well. Since $\hat{X}_1, \dots, \hat{X}_n$ are i.i.d samples of measure $\hat{\mu}$, it follows that the random variables $Y_1 := (\nabla\Phi)^{-1}(\hat{X}_1), \dots, Y_n := (\nabla\Phi)^{-1}(\hat{X}_n)$ are independent and uniformly distributed on $B(0,1)$. Let $\hat{\nu}_n$ be the empirical measure corresponding to $Y_1, \dots, Y_n$. Let $\tilde{T}_n$ be the sequence of transportation maps between $\hat{\nu}$ and $\hat{\nu}_n$ which satisfy the conclusions of this lemma (which exist by Step 1). Then $\hat{T}_n := \nabla\Phi \circ \tilde{T}_n \circ (\nabla\Phi)^{-1}$ are transportation maps between $\hat{\mu}$ and $\hat{\mu}_n$ and satisfy the desired estimates. □

5.4. **Example: An application to clustering.** Over the last couple of years a number of algorithms involving total-variation and related functionals have been introduced for the purposes of data analysis [9, 12, 13, 16, 14, 15, 27, 28, 39, 42, 43]. Here we present an illustration of how the Γ-convergence results can be applied to functionals in data analysis. The example we choose is simple and its primary goal is to give a hint of the possibilities. We intend to carefully investigate the more relevant functionals in future works.

Let D be the planar domain depicted on Figure 1, which we define as follows. Let Φ_t be the flow map of the vector field $v(x,y) := (\sin(x)\cos(y), -\cos(x)\sin(y))$. Let $\Psi = \Phi_1$ and $D = \Psi([-1,1]^2)$. Since v is divergence free and smooth, Ψ is a bi-Lipschitz volume-preserving homeomorphism between $[-1,1]^2$ and D.

Consider the problem of dividing the domain into two clusters of equal sizes. In the continuum setting the problem can be posed as finding $A_{min} \subset D$ such that $F(A) = TV(\chi_A)$, is minimized over all A such that $\mathrm{Vol}(D) = 2\,\mathrm{Vol}(A)$. It is not hard to see that there are exactly two minimizers (A_{min} and its complement) of the energy, see Figure 2.

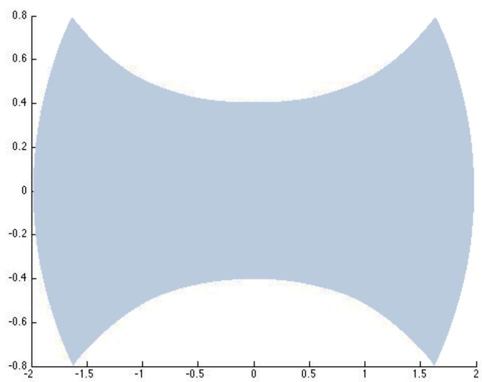

FIGURE 1. Domain D

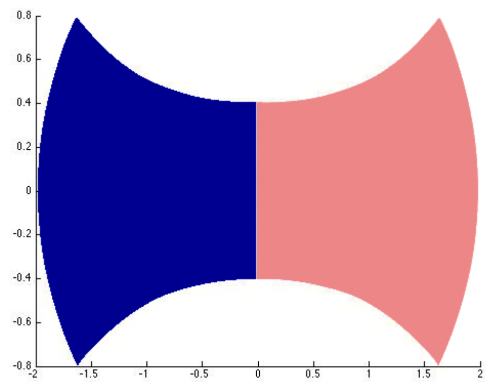

FIGURE 2. Energy minimizers

In the discrete setting assume that n is even and that $V_n = \{X_1, \dots, X_n\}$ are random points in D. The clustering problem can be described as finding $\bar{A}_n \subset V_n$, which minimizes

$$F_n(A_n) = \widetilde{TV}_{n,\varepsilon_n}(\chi_{A_n})$$

among all $A_n \subset V_n$ with $\sharp A_n = n/2$. We can extend the functionals F_n and F to be equal to $+\infty$ for sets which do not satisfy the volume constraint.

The kernel we consider for simplicity is the one given by $\eta(x) = 1$ if $|x| < 1$ and $\eta(x) = 0$ otherwise. While we did not consider the graph total variation with this constraint in the paper thus far, we note that Γ-convergence of graph total variation on domains like D was considered in Subsection 5.3 (Note that the corresponding Γ-convergence for perimeter also holds). The liminf inequality in the constraint case follows directly. To show the limsup inequality, one can use Remark 5.1 to obtain it first for sets G which satisfy the volume constraint and are of the form $G = E \cap D$ where E is a set with smooth boundary. After a careful modification to the density argument in Remark 3.43 in [5] one can impose a volume constraint to the approximating sequences and finally use Remark 2.6 to obtain the limsup inequality in the general case.

The compactness result implies that if ε satisfy (8), then along a subsequence, the minimizers $\bar{A}_n$ of F_n converge to $\bar{A}$ which minimizes F. Thus our results provide sufficient conditions which guarantee the consistency (convergence) of the scheme as the number of data points increases to infinity. That is, they indicate that for $\varepsilon(n)$ sufficiently large the minimizers converge (along a subsequence) to the desired set.

Here we illustrate the minimizers corresponding to different ε on a fixed dataset. On Figure 4 we present the discrete minimizer when ε is taken large enough. Note that this minimizer resembles the one in the continuous setting in Figure 2. In contrast, on Figure 6 we present a minimizer when ε is taken too small. Note that in this case the energy of such minimizer is zero. The solutions are computed by using the code of [14].

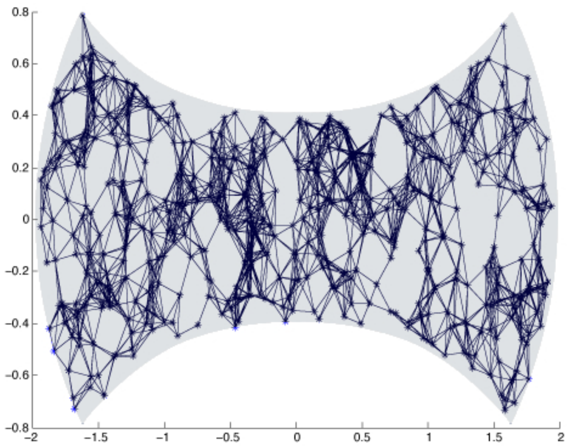

FIGURE 3. Graph with n=500, $\varepsilon = 0.18$

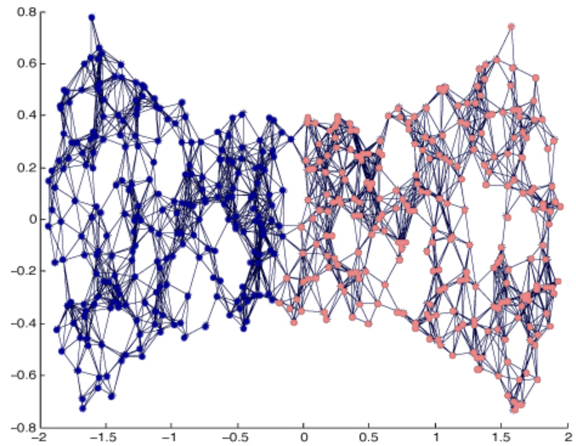

FIGURE 4. Minimizers when $\varepsilon = 0.18$

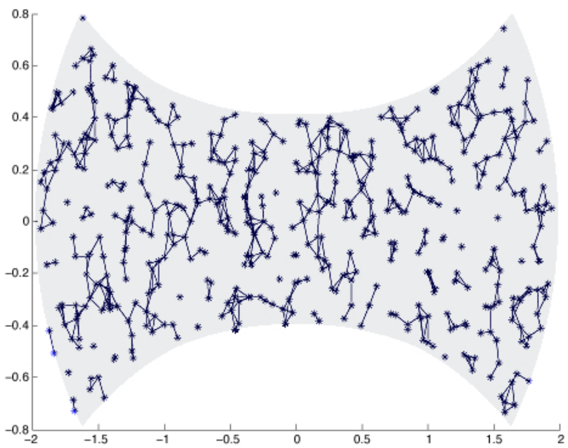

FIGURE 5. Graph with n=500, $\varepsilon = 0.1$

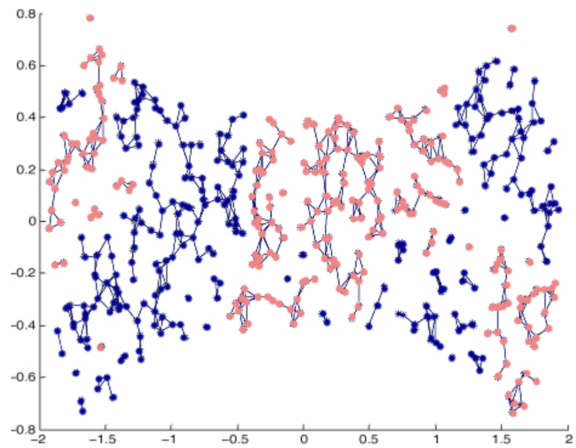

FIGURE 6. A minimizer when $\varepsilon = 0.1$

Acknowledgments. The authors are grateful to Thomas Laurent for many valuable discussions and careful reading of an early version of the manuscript. The authors are thankful to Alan Frieze, Giovanni Leoni, and James Nolen for enlightening discussions. DS is grateful to NSF (grant DMS-0908415). The research was also supported by NSF PIRE grant OISE-0967140. Authors are thankful to the Center for Nonlinear Analysis (NSF grant DMS-0635983) for its support.

Appendix A. Approximating surface tensions

Here we show that surface tensions that correspond to approximating diffuse interface functionals (36) converge to the surface tension of the sharp-interface problem as $k \to \infty$.

Consider a unit vector $e \in \mathbb{R}^d$. Denote by $\mathcal{C}_e$ the set of $d-1$ dimensional cubes, centered at the origin and contained in the orthogonal complement of e. For $C \in \mathcal{C}_e$ we denote by T_C the set of points x whose projection on $e^\perp$ is in C and finally for given $C \in \mathcal{C}_e$ consider X_C to be the class of mesurable functions $u : \mathbb{R}^d \to [0,1]$ which are C-periodic and satisfy $\lim_{\langle x, \varepsilon \rangle \to \infty} u(x) = 1$ and

$\lim_{\langle x,\varepsilon\rangle\to-\infty} u(x)=0$. We define the energy,

$$\mathcal{F}^{(k)}(u,T_C):=\int_{x\in T_C}\int_{\mathbb{R}^d}\eta(h)(u(x+h)-u(x))^2dxdh+\int_{x\in T_C}W_k(u(x))dx,\ u\in L^1(D).$$

We can now specify σ_η^k:

$$\sigma_\eta^k:=\inf\left\{\,\mathcal{F}^{(k)}(u,T_C)/\mid C|: u\in X_C, C\in\mathcal{C}_e\,\right\}, \tag{56}$$

An important remark is that σ_η^k is independent of the choice of unit vector e. In fact this follows from the isotropy assumed for η. From now on we take $e=(1,0,\dots,0)$.

The minimization problem defining σ_η^k in (56) is called the optimal profile problem associated to η and W_k and it basically quantifies the least expensive way to transition from state 0 to state 1, where the cost is measured by the energy $\mathcal{F}^{(k)}(\cdot,T_C)$. In [3] it was proved that the infimum in (56) is actually a minimum. Moreover, it is proved that for any $C\in\mathcal{C}_e$ there exists a function $\hat{u}_k:\mathbb{R}^d\to[0,1]$ which is C-periodic, only depends on the first coordinate, is increasing in the first coordinate and together with C solves (56). Since the energy $\mathcal{F}^{(k)}(\cdot,T_C)$ is invariant under translations in the first coordinate we select the optimal function $\hat{u}_k$ so that $0=\sup\{\,x_1:\hat{u}_k(x)<1/2\,\}$, by first translating if necessary. From now on we take $C=\{0\}\times[-1/2,1/2]\times\cdots\times[-1/2,1/2]$. Intuitively because of the double-well potential term in $\mathcal{F}^{(k)}(\cdot,T_C)$ we should expect that as k gets bigger, the functions $\hat{u}_k$ get closer to the sharp interface $\hat{u}$ given by

$$\hat{u}(x):=\begin{cases}1 & \text{if } x_1\geq 0\\ 0 & \text{if } x_1<0\end{cases}$$

This is precisely the content of the next proposition. Note that a direct computation shows that σ_η defined in (6) satisfies $\sigma_\eta=\int_{T_C}\int_{\mathbb{R}^d}\eta(h)|\hat{u}(x+h)-\hat{u}(x)|dxdh$.

Lemma A.1. *For the functions $\{\hat{u}_k\}_{k\in\mathbb{N}}$ and $\hat{u}$ considered above, $\hat{u}_k(x)\to\hat{u}(x)$ as $k\to\infty$ for almost every $x\in\mathbb{R}^d$.*

Proof. First of all note that it is enough to prove the result for a.e. point in T_C by C-periodicity. Consider $\{a_k\}_{k\in N}$ a decreasing sequence of positive numbers converging to 0 with $a_1<1/2$. Then

$$\int_{T_C}W_k(\hat{u}_k)dx\geq\int_{\{\,x\in T_c|a_k<\hat{u}_k(x)<1-a_k\,\}}W_k(\hat{u}_k)dx\geq kW(a_k)\,\mathrm{Vol}\left(\{x\in T_c\ :\ a_k<\hat{u}_k(x)<1-a_k\}\right)$$

This implies,

$$\sigma_\eta\geq\sigma_\eta^k\geq kW(a_k)\,\mathrm{Vol}\left(\{\,x\in T_c\ :\ a_k<\hat{u}_k(x)<1-a_k\,\}\right)$$

Therefore

$$\sigma_\eta/(kW(a_k))\geq\mathrm{Vol}\left(\{\,x\in T_c\mid a_k<\hat{u}_k(x)<1-a_k\,\}\right)$$

Let us choose the sequence $\{a_k\}_{k\in\mathbb{N}}$ in such a way that $kW(a_k)\to\infty$ as $k\to\infty$. From the fact that $\hat{u}_k$ only depends on x_1 and that it is increasing in x_1, we conclude that if $x_1>\sigma_\eta/(kW(a_k))$ then $\hat{u}_k(x)\geq 1-a_k$ and if $x_1<-\sigma_\eta/(kW(a_k))$ then $\hat{u}_k(x)\leq a_k$. This immediately implies $\hat{u}_k\to\hat{u}$ a.e. in T_c. □

Lemma A.2.

$$\lim_{k\to\infty}\sigma_\eta^k=\sigma_\eta. \tag{57}$$

Proof. Indeed, by the previous lemma and Fatou's lemma

$$\begin{aligned}
\sigma_\eta &= \int_{T_C} \int_{\mathbb{R}^d} \eta(h) |\hat{u}(x+h) - \hat{u}(x)| dh dx \\
&= \int_{T_C} \int_{\mathbb{R}^d} \eta(h) |\hat{u}(x+h) - \hat{u}(x)|^2 dh dx \\
&= \int_{T_C} \int_{\mathbb{R}^d} \eta(h) \lim_{k\to\infty} |\hat{u}_k(x+h) - \hat{u}_k(x)|^2 dh dx \\
&\leq \liminf_{k\to\infty} \int_{T_C} \int_{\mathbb{R}^d} \eta(h) |\hat{u}_k(x+h) - \hat{u}_k(x)|^2 dh dx \\
&\leq \liminf_{k\to\infty} \sigma_\eta^k \leq \limsup_{k\to\infty} \sigma_\eta^k \leq \sigma_\eta.
\end{aligned}$$

□

References

[1] M. Agueh, *Finsler structure in the p-Wasserstein space and gradient flows*, C. R. Math. Acad. Sci. Paris, 350 (2012), pp. 35–40.

[2] M. Ajtai, J. Komlós, and G. Tusnády, *On optimal matchings*, Combinatorica, 4 (1984), pp. 259–264.

[3] 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.

[4] ———, *A nonlocal anisotropic model for phase transitions. I. The optimal profile problem*, Math. Ann., 310 (1998), pp. 527–560.

[5] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.

[6] 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.

[7] 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.

[8] M. Belkin and P. Niyogi, *Convergence of laplacian eigenmaps*, Advances in Neural Information Processing Systems (NIPS), 19 (2007), p. 129.

[9] A. L. Bertozzi and A. Flenner, *Diffuse interface models on graphs for classification of high dimensional data*, Multiscale Modeling and Simulation, 10 (2012), pp. 1090–1118.

[10] A. Braides, *Gamma-Convergence for Beginners*, Oxford Lecture Series in Mathematics and Its Applications Series, Oxford University Press, Incorporated, 2002.

[11] A. Braides and N. K. Yip, *A quantitative description of mesh dependence for the discretization of singularly perturbed nonconvex problems*, SIAM J. Numer. Anal., 50 (2012), pp. 1883–1898.

[12] X. Bresson and T. Laurent, *Asymmetric cheeger cut and application to multi-class unsupervised clustering.* CAM report, 2012.

[13] X. Bresson, T. Laurent, D. Uminsky, and J. H. von Brecht, *Convergence and energy landscape for cheeger cut clustering*, in Advances in Neural Information Processing Systems (NIPS), P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., 2012, pp. 1394–1402.

[14] X. Bresson, T. Laurent, D. Uminsky, and J. H. von Brecht, *An adaptive total variation algorithm for computing the balanced cut of a graph*, arXiv preprint arXiv:1302.2717, (2013).

[15] ———, *Multiclass total variation clustering*, arXiv preprint arXiv:1306.1185, (2013).

[16] X. Bresson, X.-C. Tai, T. F. Chan, and A. Szlam, *Multi-class transductive learning based on l1 relaxations of cheeger cut and mumford-shah-potts model*, UCLA CAM Report, (2012), pp. 12–03.

[17] L. A. Caffarelli, *The regularity of mappings with a convex potential*, J. Amer. Math. Soc., 5 (1992), pp. 99–104.

[18] ———, *Boundary regularity of maps with convex potentials. II*, Ann. of Math. (2), 144 (1996), pp. 453–496.

[19] A. Chambolle, A. Giacomini, and L. Lussardi, *Continuous limits of discrete perimeters*, M2AN Math. Model. Numer. Anal., 44 (2010), pp. 207–230.

[20] K.-L. Chung, *An estimate concerning the Kolmogoroff limit distribution*, Trans. Amer. Math. Soc., 67 (1949), pp. 36–50.

[21] G. Dal Maso, *An Introduction to Γ-convergence*, Springer, 1993.

[22] S. Esedoglu and F. Otto, *Threshold dynamics for networks with arbitrary surface tensions.* Max Planck Institute for Mathematics in the Sciences (Leipzig) Preprint, 2013.

[23] I. Fonseca and G. Parry, *Equilibrium configurations of defective crystals*, Arch. Rational Mech. Anal., 120 (1992), pp. 245–283.

[24] 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).

[25] J. A. Griepentrog, W. Höppner, H.-C. Kaiser, and J. Rehberg, *A bi-Lipschitz continuous, volume preserving map from the unit ball onto a cube*, Note Mat., 28 (2008), pp. 177–193.

[26] 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.

[27] M. Hein and T. Bühler, *An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA*, in Advances in Neural Information Processing Systems (NIPS), 2010, pp. 847–855.

[28] M. Hein and S. Setzer, *Beyond spectral clustering - tight relaxations of balanced graph cuts*, in Advances in Neural Information Processing Systems (NIPS), J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, eds., 2011, pp. 2366–2374.

[29] J. Kiefer, *On large deviations of the empiric D. F. of vector chance variables and a law of the iterated logarithm*, Pacific J. Math., 11 (1961), pp. 649–660.

[30] 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.

[31] G. Leoni, *A first course in Sobolev spaces*, vol. 105 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2009.

[32] E. Merkurjev, T. Kostić, , and A. Bertozzi, *An MBO scheme on graphs for segmentation and image processing*. to appear in SIAM J. Imag. Proc., 2013.

[33] L. Modica and S. Mortola, *Un esempio di* Γ*-convergenza*, Boll. Un. Mat. Ital. B (5), 14 (1977), pp. 285–299.

[34] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[35] M. Penrose, *A strong law for the longest edge of the minimal spanning tree*, Ann. Probab., 27 (1999), pp. 246–260.

[36] ———, *Random geometric graphs*, vol. 5 of Oxford Studies in Probability, Oxford University Press, Oxford, 2003.

[37] M. D. Penrose and J. E. Yukich, *Weak laws of large numbers in geometric probability*, Ann. Appl. Probab., 13 (2003), pp. 277–303.

[38] D. Pollard, *Strong consistency of k-means clustering*, The Annals of Statistics, 9 (1981), pp. 135–140.

[39] S. S. Rangapuram and M. Hein, *Constrained 1-spectral clustering*, in International conference on Artificial Intelligence and Statistics (AISTATS), 2012, pp. 1143–1151.

[40] O. Savin and E. Valdinoci, Γ*-convergence for nonlocal phase transitions*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 29 (2012), pp. 479–500.

[41] P. W. Shor and J. E. Yukich, *Minimax grid matching and empirical measures*, Ann. Probab., 19 (1991), pp. 1338–1348.

[42] A. Szlam and X. Bresson, *A total variation-based graph clustering algorithm for cheeger ratio cuts*, UCLA CAM Report, (2009), pp. 09–68.

[43] ———, *Total variation, cheeger cuts.*, in ICML, J. Frnkranz and T. Joachims, eds., Omnipress, 2010, pp. 1039–1046.

[44] M. Talagrand, *The transportation cost from the uniform measure to the empirical measure in dimension* ≥ 3, Ann. Probab., 22 (1994), pp. 919–959.

[45] M. Talagrand and J. E. Yukich, *The integrability of the square exponential transportation cost*, Ann. Appl. Probab., 3 (1993), pp. 1100–1111.

[46] J. Urbas, *On the second boundary value problem for equations of Monge-Ampère type*, J. Reine Angew. Math., 487 (1997), pp. 115–124.

[47] Y. van Gennip and A. L. Bertozzi, Γ*-convergence of graph Ginzburg-Landau functionals*, Adv. Differential Equations, 17 (2012), pp. 1115–1180.

[48] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, American Mathematical Society, 2003.

[49] U. von Luxburg, M. Belkin, and O. Bousquet, *Consistency of spectral clustering*, Ann. Statist., 36 (2008), pp. 555–586.