Arch. Rational Mech. Anal. 220 (2016) 193–241
Digital Object Identifier (DOI) 10.1007/s00205-015-0929-z

Continuum Limit of Total Variation on Point Clouds

Nicolás García Trillos & Dejan Slepčev

Communicated by F. Otto

Abstract

We consider point clouds obtained as random samples of a measure on a Euclidean domain. A graph representing the point cloud is obtained by assigning weights to edges based on the distance between the points they connect. Our goal is to develop mathematical tools needed to study the consistency, as the number of available data points increases, of graph-based machine learning algorithms for tasks such as clustering. In particular, we study when the cut capacity, and more generally total variation, on these graphs is a good approximation of the perimeter (total variation) in the continuum setting. We address this question in the setting of Γ-convergence. We obtain almost optimal conditions on the scaling, as the number of points increases, of the size of the neighborhood over which the points are connected by an edge for the Γ-convergence to hold. Taking of the limit is enabled by a transportation based metric which allows us to suitably compare functionals defined on different point clouds.

1. Introduction

Our goal is to develop mathematical tools to rigorously study limits of variational problems defined on random samples of a measure as the number of data points goes to infinity. The main application is to establish the consistency of machine learning algorithms for tasks such as clustering and classification. These tasks are of fundamental importance for statistical analysis of randomly sampled data, yet few results on their consistency are available. In particular, a largely open question is determining when the minimizers of graph-based tasks converge, as the amount of available data increases, to a minimizer of a limiting functional in the continuum setting. Here we introduce the mathematical setup needed to address such questions.

To analyze the structure of a data cloud one defines a weighted graph to represent it. Points become vertices and are connected by edges if sufficiently close. The edges are assigned weights based on the distances between points. How the graph is constructed is important: for lower computational complexity one seeks to have fewer edges, but below some threshold the graph no longer contains the desired information on the geometry of the point cloud. The machine learning tasks, such as classification and clustering, can often be given in terms of minimizing a functional on the graph representing the point cloud. Some of the fundamental approaches are based on minimizing graph cuts (graph perimeter) and related functionals (normalized cut, ratio cut, balanced cut), and more generally total variation on graphs [8,13,15,18–20,22,36,37,41,47,49,52,53]. We focus on total variation on graphs (of which graph cuts are a special case). The techniques we introduce are applicable to a rather broad range of functionals, in particular those where total variation is combined with lower-order terms, or those where total variation is replaced by Dirichlet energy.

The graph perimeter (a.k.a. cut size, cut capacity) of a set of vertices is the sum of the weights of edges between the set and its complement. Our goal is to understand for what constructions of graphs from data is the cut capacity a good notion of a perimeter. We pose this question in terms of consistency as the number of data points increases: $n \to \infty$. We assume that the data points are random independent samples of an underlying measure ν with density ρ supported in a set D in $\mathbb{R}^d$. The question is whether or not the graph perimeter on the point cloud is a good approximation of the perimeter on D (weighted by ρ^2). Since machine learning tasks involve minimizing appropriate functionals on graphs, the most relevant question is if the minimizers of functionals on graphs involving graph cuts converge to minimizers of corresponding limiting functionals in a continuum setting, as $n \to \infty$. Such convergence is implied by the variational notion of convergence called the Γ-convergence, which we focus on. The notion of Γ-convergence has been used extensively in the calculus of variations, in particular in homogenization theory, phase transitions, image processing, and materials science. We show how the Γ-convergence can be applied to establishing consistency of data-analysis algorithms.

1.1. Setting and the Main Results

Consider a point cloud $V = \{X_1, \dots, X_n\}$. Let η be a kernel, that is, let $\eta : \mathbb{R}^d \to [0, \infty)$ 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

$$W_{i,j} = \eta_\varepsilon(X_i - X_j) \tag{1.1}$$

and define the graph perimeter of $A \subset V$ to be

$$\mathrm{GPer}(A) = 2 \sum_{X_i \in A} \sum_{X_j \in V \setminus A} W_{i,j}. \tag{1.2}$$

The graph perimeter (that is cut size, cut capacity), can be effectively used as a term in functionals which give a variational description to classification and clustering [13,15,18–22,36,37,41,47,52,53].

The total variation of a function u defined on the point cloud is typically given as

$$\sum_{i,j} W_{i,j} |u(X_i) - u(X_j)|. \tag{1.3}$$

We note that the total variation is a generalization of perimeter since the perimeter of a set of vertices $A \subset V$ is the total variation of the characteristic function of A.

In this paper we focus on point clouds that are obtained as samples from a given distribution ν. Specifically, consider an open, bounded, and connected set $D \subset \mathbb{R}^d$ with Lipschitz boundary and a probability measure ν supported on $\overline{D}$. Suppose that ν has density ρ, which is continuous and bounded above and below by positive constants on D. Assume n data points $X_1, \dots, X_n$ (i.i.d. random points) are chosen according to the distribution ν. We consider a graph with vertices $V = \{X_1, \dots, X_n\}$ and edge weights $W_{i,j}$ given by (1.1), where η_ε to be defined by $\eta_\varepsilon(z) := \frac{1}{\varepsilon^d} \eta\left(\frac{z}{\varepsilon}\right)$. Note that significant weight is given to edges connecting points up to distance of order ε.

Having limits as $n \to \infty$ in mind, we define the *graph total variation* to be a rescaled form of (1.3):

$$GTV_{n,\varepsilon}(u) := \frac{1}{\varepsilon} \frac{1}{n^2} \sum_{i,j} W_{i,j} |u(X_i) - u(X_j)|. \tag{1.4}$$

For a given scaling of ε with respect to n, we study the limiting behavior of $GTV_{n,\varepsilon(n)}$ as the number of points $n \to \infty$. The limit is considered in the variational sense of Γ-convergence.

A key contribution of our work is in identifying the proper topology with respect to which the Γ-convergence takes place. As one is considering functions supported on the graphs, the issue is how to compare them with functions in the continuum setting, and how to compare functions defined on different graphs. Let us denote by ν_n the empirical measure associated with the n data points:

$$\nu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}. \tag{1.5}$$

The issue is then how to compare functions in $L^1(\nu_n)$ with those in $L^1(\nu)$. More generally we consider how to compare functions in $L^p(\mu)$ with those in $L^p(\theta)$ for arbitrary probability measures μ, θ on D and arbitrary $p \in [1, \infty)$. We set

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

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

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

where $\Gamma(\mu, \theta)$ is the set of all *couplings* (or *transportation plans*) between μ and θ, that is, the set of all Borel probability measures on $D \times D$ for which the marginal on the first variable is μ and the marginal on the second variable is θ. As discussed in Section 3, d_{TL^p} is a transportation distance between graphs of functions.

The TL^p topology provides a general and versatile way to compare functions in a discrete setting with functions in a continuum setting. It is a generalization of the weak convergence of measures and of L^p convergence of functions. By this we mean that $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathscr{P}(D)$ converges weakly to $\mu \in \mathscr{P}(D)$ if and only if $(\mu_n, 1) \xrightarrow{TL^p} (\mu, 1)$ as $n \to \infty$, and that for $\mu \in \mathscr{P}(D)$ a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $L^p(\mu)$ converges in $L^p(\mu)$ to f if and only if $(\mu, f_n) \xrightarrow{TL^p} (\mu, f)$ as $n \to \infty$. This fact is established in Proposition 3.12.

Furthermore, if one considers functions defined on a regular grid, then the standard way [17,24] to compare them is to identify them with piecewise constant functions, whose value on the grid cells is equal to the value at the appropriate grid point, and then compare the extended functions using the L^p metric. A TL^p metric restricted to regular grids gives the same topology.

The kernels η we consider are assumed to be isotropic, and thus can be defined as $\eta(x) := \boldsymbol{\eta}(|x|)$ where $\boldsymbol{\eta} : [0, \infty) \to [0, \infty)$ is the radial profile. We assume:

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

We note that the class of admissible kernels is broad and includes both Gaussian kernels and discontinuous kernels like one defined by $\boldsymbol{\eta}$ of the form $\boldsymbol{\eta} = 1$ for $r \leqq 1$ and $\boldsymbol{\eta} = 0$ for $r > 1$. We remark that the assumption (K3) is equivalent to imposing that the surface tension

$$\sigma_\eta = \int_{\mathbb{R}^d} \eta(h)|h_1| \,\mathrm{d}h, \tag{1.6}$$

where h_1 is the first coordinate of vector h, is finite, and also that one can replace h_1 in the above expression by $h \cdot e$ for any fixed $e \in \mathbb{R}^d$ with norm one; this, given that η is radially symmetric.

The weighted total variation in continuum setting (with weight ρ^2), $TV(\cdot, \rho^2) : L^1(D, \nu) \to [0, \infty]$, is given by

$$TV(u; \rho^2) = \sup\left\{\int_D u \,\mathrm{div}(\phi)\,\mathrm{d}x \;:\; |\phi(x)| \leqq \rho^2(x) \;\; \forall x \in D\,,\; \phi \in C_c^\infty(D, \mathbb{R}^d)\right\} \tag{1.7}$$

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 weighted total variation can be written as $TV(u; \rho^2) = \int_D |\nabla u|\rho^2(x)\mathrm{d}x$.

The main result of the paper is:

Theorem 1.1. (Γ-convergence) *Let $D \subset \mathbb{R}^d$, $d \geqq 2$ be an open, bounded, connected set with Lipschitz boundary. Let ν be a probability measure on D with continuous density ρ, which is bounded from below and above by positive constants. Let $\{X_n\}_{n=1,\dots}$ be a sequence of i.i.d. random points chosen according to* *distribution ν on D. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to* 0 *and satisfying*

$$\lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} = 0 \quad \text{if } d = 2,$$
$$\lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} = 0 \quad \text{if } d \geqq 3. \tag{1.8}$$

Assume the kernel η satisfies conditions (K1)–(K3). Then, GTV_{n,ε_n}, defined by (1.4)*, Γ-converge to $\sigma_\eta TV(\cdot, \rho^2)$ as $n \to \infty$ in the TL^1 sense, where σ_η is given by* (1.6) *and $TV(\cdot, \rho^2)$ is the weighted total variation functional defined in* (1.7)*.*

The notion of Γ-convergence in deterministic setting is recalled in Subsection 2.4, where we also extend it to the probabilistic setting in Definition 2.11. The fact that the density in the limit is ρ^2 essentially follows from the fact that graph total variation is a double sum [and becomes more apparent in Section 5 when we write the graph total variation in form (5.1)].

The following compactness result shows that the TL^1 topology is indeed a good topology for the Γ-convergence (in the light of Proposition 2.10).

Theorem 1.2. (Compactness) *Under the assumptions of the theorem above, consider a sequence of functions $u_n \in L^1(D, \nu_n)$, where ν_n is given by* (1.5)*. If $\{u_n\}_{n\in\mathbb{N}}$ have uniformly bounded $L^1(D, \nu_n)$ norms and graph total variations, GTV_{n,ε_n}, then the sequence is relatively compact in TL^1. More precisely, if*

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

and

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

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

If A_n is a subset of $\{X_1, \ldots, X_n\}$ then $GTV_{n,\varepsilon_n}(\chi_{A_n}) = \frac{1}{n^2\epsilon_n} \mathrm{GPer}(A_n)$, where $\mathrm{GPer}(A_n)$ was defined in (1.2). The proof of Theorem 1.1 allows us to show the variational convergence of the perimeter on graphs to the weighted perimeter in domain D, defined by $\mathrm{Per}(E : D, \rho^2) = TV(\chi_E, \rho^2)$.

Corollary 1.3. (Γ-convergence of perimeter) *The conclusions of Theorem 1.1 hold, under the same assumptions, when the functionals are restricted to characteristic functions of sets. That is, the (scaled) graph perimeters, $\frac{1}{n^2\epsilon_n} \mathrm{GPer}(\cdot)$, Γ-converge to the continuum (weighted) perimeter* $\mathrm{Per}(\cdot : D, \rho^2)$.

The proofs of the theorems and of the corollary are presented in Section 5. We remark that the Corollary 1.3 is not an immediate consequence of Theorem 1.1, since in general Γ-convergence may not carry over when a (closed) subspace of a metric space is considered. The proof of Corollary 1.3 is nevertheless straightforward.

Remark 1.4. When one considers ρ to be constant in Theorem 1.1, the points $X_1, \dots, X_n$ are uniformly distributed on D. In this particular case, the theorem implies that the graph total variation converges to the usual total variation on D (appropriately scaled by $1/\operatorname{Vol}(D)^2$). Corollary 1.3 implies that the graph perimeter converges to the usual perimeter (appropriately scaled).

Remark 1.5. The notion of Γ-convergence is different from the notion of pointwise convergence, but often the proof of Γ-convergence implies the pointwise convergence. The pointwise convergence of the graph perimeter to continuum perimeter is the statement that for any set $A \subset D$ of finite perimeter, with probability one:

$$\lim_{n\to\infty} GTV_{n,\varepsilon_n}(\chi_A) = \operatorname{Per}(A : D, \rho^2).$$

In the case that D is smooth, the points $X_1, \dots, X_n$ are uniformly distributed on D and A is smooth, the pointwise convergence of the graph perimeter can be obtained from the results in [7,40] when ε_n is converging to zero so that $\frac{(\log n)^{1/(d+1)}}{n^{1/(d+1)}}\frac{1}{\varepsilon_n} \to 0$ as $n \to \infty$. In Remark 5.1 we point out that our proof of Γ-convergence implies that pointwise convergence also holds, with same scaling for ε_n as in Theorem 1.1, which slightly improves the rate of pointwise convergence in [7]. Note that pointwise convergence does not follow directly from the Γ-convergence.

Remark 1.6. 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 $\sharp A_n \leqq n/2$. Define function $u_n = \frac{n}{\sharp A_n}\chi_{A_n}$. Note that $\|u_n\|_{L^1(\nu_n)} = 1$ and that $GTV_{n,\varepsilon_n}(u_n) = 0$. By compactness, along a subsequence (not relabeled), u_n converges in TL^1 to a function $u \in L^1(\nu)$. Thus $\|u\|_{L^1(\nu)} = 1$. By lower-semicontinuity, which follows from the 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 $d \geqq 3$ then the rate presented in (1.8) is sharp in terms of scaling. To illustrate, this suppose that the data points are uniformly distributed on D and η has compact support. It is known from graph theory (see [33,34,45]) 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 1.6, the compactness property cannot hold if $\varepsilon_n < \lambda\frac{(\log n)^{1/d}}{n^{1/d}}$. It is, of course, not surprising that if the graph is disconnected, the functionals describing clustering tasks may have minimizers which are rather different than the minimizers of the continuum functional.

While the above example shows the optimality of our results in some sense, we caution that there still may be settings relevant to machine learning in which the convergence of minimizers of appropriate functionals may hold even when $\frac{1}{n^{1/d}} \ll \varepsilon_n < \lambda \frac{(\log n)^{1/d}}{n^{1/d}}$.

Finally, we remark that in the case $d = 2$, the rate presented in (1.8) is different from the connectivity rate in dimension $d = 2$ which is $\lambda \frac{(\log n)^{1/2}}{n^{1/2}}$. An interesting open problem is determining what happens to the graph total variation as $n \to \infty$, when one considers $\lambda \frac{(\log n)^{1/2}}{n^{1/2}} \ll \varepsilon_n \leqq \frac{(\log n)^{3/4}}{n^{1/2}}$.

1.3. Related Work

Background on Γ-convergence of functionals related to perimeter. The notion of Γ-convergence was introduced by De Giorgi in the 70's and represents a 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. For extensive exposition of the properties of Γ-convergence see the books by Braides [16] and Dal Maso [25].

A classical example of Γ-convergence of functionals to perimeter is the Modica and Mortola theorem [42] that shows the Γ-convergence of Allen–Cahn (Cahn–Hilliard) free energy to perimeter.

There are a number of results considering nonlocal functionals converging to the perimeter or to total variation. 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. In [48], Savin and Valdinoci consider a related energy involving more general kernels. Esedoḡlu and Otto [27] consider nonlocal total-variation based functionals in multiphase systems and show their Γ-convergence to perimeter. Brezis, Bourgain, and Mironescu [14] consider nonlocal functionals in order to give new characterizations of Sobolev and BV spaces. Ponce [46] extended their work and showed the Γ-convergence of the nonlocal functionals studied to local ones. In our work we adopt the approach of Ponce to show Γ-convergence as it is conceptually clear and efficient.

We also note the works of Gobbino [31] and Gobbino and Mora [32] where elegant nonlocal approximations were considered for more complicated functionals, like the Mumford-Shah functional.

In the discrete setting, works related to the Γ-convergence of functionals to continuous functionals involving perimeter include [17,24,59]. The results by Braides and Yip [17], 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 and the gradient term gets replaced by a finite-difference approximation that depends on the mesh size δ. Van Gennip and Bertozzi [59] consider a similar problem and obtain analogous results. In [24], 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.

Most of the available consistency results are about pointwise consistency. Among them are works of BELKIN and NIYOGI [12], GINÉ and KOLTCHINSKII [30], HEIN, AUDIBERT, VON LUXBURG [35], SINGER [51] and TING, HUANG, and JORDAN [58]. The works of VON LUXBURG, BELKIN and BOUSQUET on consistency of spectral clustering [61] and BELKIN and NIYOGI [11] on the convergence of Laplacian Eigenmaps consider spectral convergence and thus convergence of eigenvalues and eigenvectors, which are relevant for machine learning. An important difference between our work and the spectral convergence works is that in them, there is no explicit rate at which ε_n is allowed to converge to 0 as $n \to \infty$. ARIAS-CASTRO, PELLETIER, and PUDLO [7] considered pointwise convergence of Cheeger energy and consequently of total variation, as well as variational convergence when the discrete functional is considered over an admissible set of characteristic functions which satisfy a "regularity" requirement. For the variational problem they show that the convergence holds essentially when $n^{-\frac{1}{2d+1}} \ll \varepsilon_n \ll 1$. MAIER, VON LUXBURG and HEIN [40] considered pointwise convergence for Cheeger and normalized cuts, both for the geometric and kNN graphs and obtained an analogous range of scalings of graph construction on n for the convergence to hold.

1.4. Example: An Application to Clustering

Many algorithms involving graph cuts, total variation and related functionals on graphs are in use in data analysis. Here we present an illustration of how the Γ-convergence results can be applied in that context. In particular we show the consistency of *minimal bisection* considered for example in [6,26,28]. The example we choose is simple and its primary goal is to give a hint of the possibilities. We intend to investigate the functionals relevant to data analysis in future works.

Let D be domain satisfying the assumptions of Theorem 1.1, for example the one depicted on Fig. 1. 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 $\text{Vol}(D) = 2\,\text{Vol}(A)$. For the domain of Fig. 1 there are exactly two minimizers (A_{min} and its complement); illustrated on Fig. 2.

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

$$F_n(A_n) = GTV_{n,\varepsilon_n}(\chi_{A_n})$$

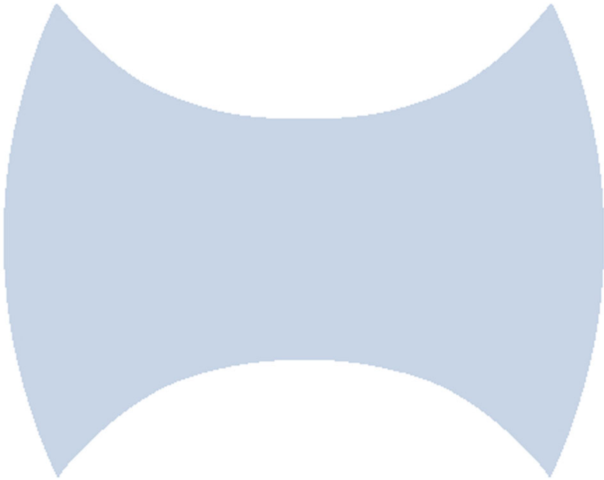

Fig. 1. Domain D

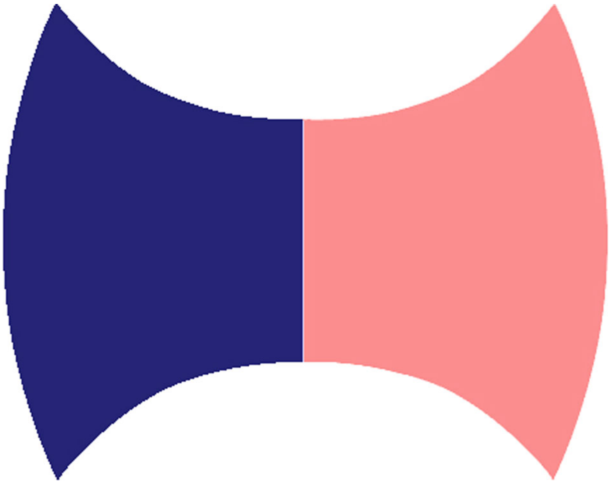

Fig. 2. Energy minimizers

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 constraints in Theorem 1.1, that extension is of a technical nature. In particular, the liminf inequality of the definition of Γ-convergence of Definition 2.6 in the constraint case follows directly, while the limsup inequality follows using the Remark 5.1.

The compactness result implies that if $\varepsilon(n)$ satisfies (1.8), then the minimizers $\bar{A}_n$ of F_n converge along a subsequence to $\bar{A}$ which minimizes F. Thus our results

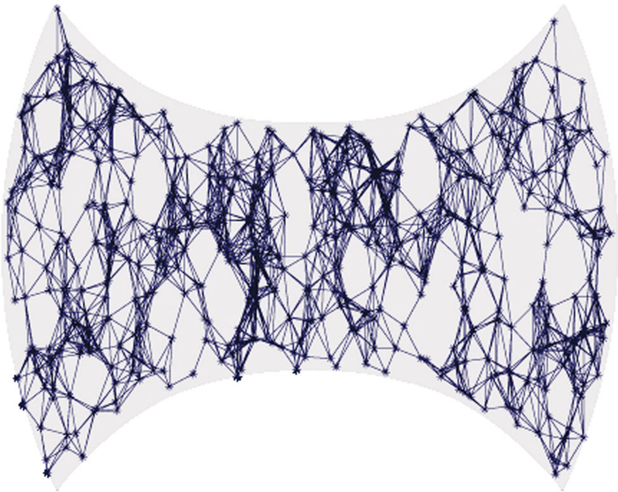

Fig. 3. Graph with n=500, $\varepsilon = 0.18$

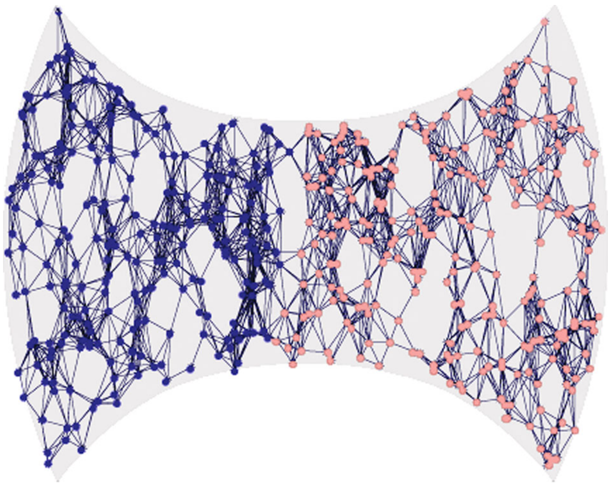

Fig. 4. Minimizers when $\varepsilon = 0.18$

provide sufficient conditions which guarantee the consistency (convergence) of the scheme as the number of data points increases to infinity.

Here we illustrate the minimizers corresponding to different ε on a fixed dataset. Figures 3 and 4 depict the graph and a minimizer when ε is taken large enough. Note that this minimizer resembles the one in the continuous setting in Fig. 2. In contrast, on Figs. 5 and 6 we present the graph and a minimizer when ε is taken too small. Note that in this case the energy of such a minimizer is zero. The solutions are computed using the code of [21].

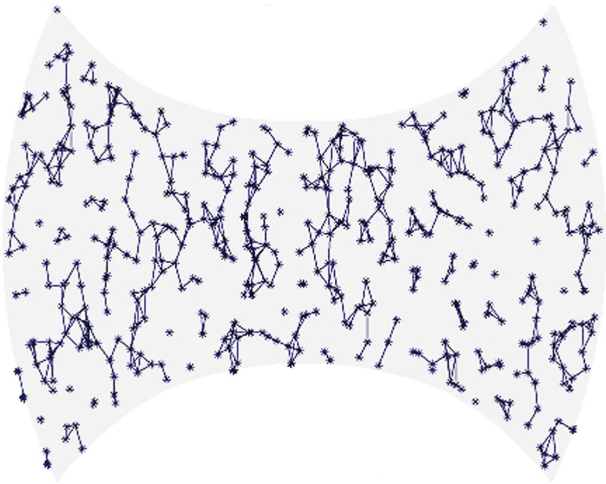

Fig. 5. Graph with n=500, $\varepsilon = 0.1$

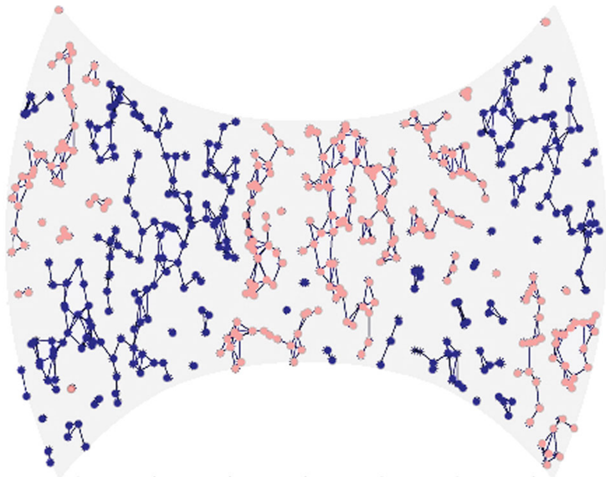

Fig. 6. A minimizer when $\varepsilon = 0.1$

1.5. Outline of the Approach

The proof of Γ-convergence of the graph total variation GTV_{n,ε_n} to weighted total variation $TV(\cdot, \rho^2)$ relies on an intermediate object, the nonlocal functional $TV_\varepsilon(\cdot, \rho) : L^1(D, \nu) \to [0, \infty]$ given by:

$$TV_\varepsilon(u; \rho) := \frac{1}{\varepsilon} \int_D \int_D \eta_\varepsilon(x - y)|u(x) - u(y)|\rho(x)\rho(y)\, \mathrm{d}x\, \mathrm{d}y. \tag{1.9}$$

Note that the argument of GTV_{n,ε_n} is a function u_n supported on the data points, while the argument of $TV_\varepsilon(\cdot; \rho)$ is an $L^1(D, \nu)$ function; in particular a function defined on D. Having defined the TL^1-metric, the proof of Γ-convergence has two main steps: the first step is to compare the graph total variation GTV_{n,ε_n}, with the nonlocal continuum functional $TV_\varepsilon(\cdot, \rho)$. To compare the functionals one needs an $L^1(D, \nu)$ function which, in TL^1 sense, approximates u_n. We use transportation maps (that is measure preserving maps) between the measure ν and ν_n to define $\tilde{u}_n \in L^1(D, \nu)$. More precisely we set $\tilde{u}_n = u_n \circ T_n$ where T_n is the transportation map between ν and ν_n constructed in Section 2.3. Comparing $GTV_{n,\varepsilon_n}(u_n)$ with $TV_\varepsilon(\tilde{u}_n; \rho)$ relies on the fact that T_n is chosen in such a way that it transports mass as little as possible. The estimates on how far the mass needs to be moved were known in the literature when ρ is constant. We extended the results to the case when ρ is bounded from below and from above by positive constants.

The second step consists of comparing the continuum nonlocal total variation functionals (1.9) with the weighted total variation (1.7).

The proof on compactness for GTV_{n,ε_n} depends on an analogous compactness result for the nonlocal continuum functional $TV_\varepsilon(\cdot, \rho)$.

The paper is organized as follows. Section 2 contains the notation and preliminary results from the weighted total variation, transportation theory and Γ-convergence of functionals on metric spaces. More specifically, in Section 2.1 we introduce and present basic facts about weighted total variation. In Section 2.2 we introduce the optimal transportation problem and list some of its basic properties. In Section 2.3 we review results on optimal matching between the empirical measure ν_n and ν. In Section 2.4 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 Section 5.2 we discuss the extension of the main result to the case when $X_1, \ldots, X_n$ are not necessarily independently distributed points.

2. Preliminaries

2.1. Weighted Total Variation

Let D be an open and bounded subset of $\mathbb{R}^d$ and let $\psi : D \to (0, \infty)$ be a continuous function. Consider the measure $\mathrm{d}\nu(x) = \psi(x)\mathrm{d}x$. We denote by $L^1(D, \nu)$ the L^1-space with respect to ν and by $\|\cdot\|_{L^1(D,\nu)}$ its corresponding norm; we use $L^1(D)$ in the special case $\psi \equiv 1$ and $\|\cdot\|_{L^1(D)}$ for its corresponding norm. If the context is clear, we omit the set D and write $L^1(\nu)$ and $\|\cdot\|_{L^1(\nu)}$. Also, with a slight abuse of notation, we often replace ν by ψ in the previous expressions; for example we use $L^1(D, \psi)$ to represent $L^1(D, \nu)$.

Following BALDI, [9], for $u \in L^1(D, \psi)$ define

$$TV(u; \psi) = \sup\left\{ \int_D u \,\mathrm{div}(\phi)\,\mathrm{d}x \ : \ (\forall x \in D) \ \ |\phi(x)| \leqq \psi(x), \ \ \phi \in C_c^\infty(D, \mathbb{R}^d) \right\} \tag{2.1}$$

the *weighted total variation of u in D with respect to the weight* ψ. We denote by $BV(D;\psi)$ the set of functions $u \in L^1(D,\psi)$ for which $TV(u;\psi) < +\infty$. When $\psi \equiv 1$ we omit it and write $BV(D)$ and $TV(u)$. Finally, for measurable subsets $E \subset D$, we define the weighted perimeter in D as the weighted total variation of the characteristic function of the set: $\mathrm{Per}(E;\psi) = TV(\chi_E;\psi)$.

Throughout the paper we restrict our attention to the case where ψ is bounded from below and from above by positive constants. Indeed, in applications we consider $\psi = \rho^2$, where ρ is continuous and bounded below and above by positive constants.

Remark 2.1. Since D is a bounded open set and ψ is bounded from above and below by positive constants, the sets $L^1(D)$ and $L^1(D,\psi)$ are equal and the norms $||\cdot||_{L^1(D)}$ and $||\cdot||_{L^1(D,\psi)}$ are equivalent. Also, it is straightforward to see from the definitions that in this case $BV(D) = BV(D;\psi)$.

Remark 2.2. If $u \in BV(D;\psi)$ is smooth enough (say for example $u \in C^1(D)$) then the weighted total variation $TV(u;\psi)$ can be written as

$$\int_D |\nabla u(x)|\psi(x)\,\mathrm{d}x.$$

If E is a regular subset of D, then $\mathrm{Per}(E;\psi)$ can be written as the following surface integral,

$$\mathrm{Per}(E;\psi) = \int_{\partial E \cap D} \psi(x)\,\mathrm{d}S(x).$$

One useful characterization of $BV(D;\psi)$ is provided in the next proposition whose proof can be found in [9].

Proposition 2.3. *Let $u \in L^1(D,\psi)$, u belongs to $BV(D;\psi)$ if and only if there exists a finite positive Radon measure $|Du|_\psi$ and a $|Du|_\psi$-measurable function $\sigma : D \to \mathbb{R}^d$ with $|\sigma(x)| = 1$ for $|Du|_\psi$-almost everywhere $x \in D$ and such that* $\forall \phi \in C_c^\infty(D,\mathbb{R}^d)$

$$\int_D u\,\mathrm{div}(\phi)\,dx = -\int_D \frac{\phi(x)\cdot\sigma(x)}{\psi(x)}\,d|Du|_\psi(x).$$

The measure $|Du|_\psi$ and the function σ are uniquely determined by the previous conditions and the weighted total variation $TV(u;\psi)$ is equal to $|Du|_\psi(D)$.

We refer to $|Du|_\psi$ as the *weighted total variation measure* (with respect to ψ) associated with u. In the case $\psi \equiv 1$, we denote $|Du|_\psi$ by $|Du|$ and we call it the *total variation measure* associated with u.

Using the previous definitions one can check that σ does not depend on ψ and that the following relation between $|Du|_\psi$ and $|Du|$ holds:

$$\mathrm{d}|Du|_\psi(x) = \psi(x)\,\mathrm{d}|Du|(x). \tag{2.2}$$

In particular,

$$TV(u;\psi) = \int_D \psi(x)\,\mathrm{d}|Du|(x). \tag{2.3}$$

The function $\sigma(x)$ is the Radon–Nikodym derivative of the distributional derivative of u (denoted by Du) with respect to the total variation measure $|Du|$.

Since the functional $TV(\cdot;\psi)$ is defined as a supremum of linear continuous functionals in $L^1(D,\psi)$, we conclude that $TV(\cdot;\psi)$ is lower semicontinuous with respect to the $L^1(D,\psi)$-metric [and thus $L^1(D)$-metric given the assumptions on ψ]. That is, if $u_n \to_{L^1(D,\psi)} u$ as $n\to\infty$, then

$$\liminf_{n\to\infty} TV(u_n;\psi) \geqq TV(u;\psi). \tag{2.4}$$

We finish this section with the following approximation result that we use in the proof of the main theorem of this paper. We give a proof of this result in Appendix A.

Proposition 2.4. *Let D be an open and bounded set with Lipschitz boundary and let $\psi : D \to \mathbb{R}$ be a continuous function which is bounded from below and from above by positive constants. Then, for every function $u \in BV(D,\psi)$ there exists a sequence $\{u_n\}_{n\in\mathbb{N}}$ with $u_n \in C_c^\infty(\mathbb{R}^d)$ such that $u_n \to_{L^1(D)} u$ and $\int_D |\nabla u_n|\psi(x)\,dx \to TV(u;\psi)$ as $n\to\infty$.*

2.2. 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 $\mathscr{P}(D)$ the set of all Borel probability measures on D. Given $1 \leqq p < \infty$, the p-OT distance between $\mu, \tilde{\mu} \in \mathscr{P}(D)$ [denoted by $d_p(\mu,\tilde{\mu})$] is defined by:

$$d_p(\mu,\tilde{\mu}) := \min\left\{\left(\int_{D\times D} |x-y|^p \,\mathrm{d}\pi(x,y)\right)^{1/p} \;:\; \pi \in \Gamma(\mu,\tilde{\mu})\right\}, \tag{2.5}$$

where $\Gamma(\mu,\tilde{\mu})$ is the set of all *couplings* between μ and $\tilde{\mu}$, that is, the set of all Borel probability measures on $D\times D$ for which the marginal on the first variable is μ and the marginal on the second variable is $\tilde{\mu}$. The elements $\pi \in \Gamma(\mu,\tilde{\mu})$ are also referred as *transportation plans* between μ and $\tilde{\mu}$. 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 [60]. When $p=\infty$

$$d_\infty(\mu,\tilde{\mu}) := \inf\left\{\operatorname{esssup}_\pi\{|x-y| \;:\; (x,y)\in D\times D\} \;:\; \pi\in\Gamma(\mu,\tilde{\mu})\right\}, \tag{2.6}$$

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

Since D is bounded the convergence in OT metric is equivalent to weak convergence of probability measures. For details see for instance [5,60] and the references therein. In particular, $\mu_n \xrightarrow{w} \mu$ (to be read μ_n converges weakly to μ) if and only if for any $1\leqq p<\infty$ there is a sequence of transportation plans between μ_n and μ, $\{\pi_n\}_{n\in\mathbb{N}}$, for which:

$$\lim_{n\to\infty} \iint_{D\times D} |x-y|^p \,\mathrm{d}\pi_n(x,y) = 0. \tag{2.7}$$

Since D is bounded, (2.7) is equivalent to $\lim_{n\to\infty} \iint_{D\times D} |x-y| \mathrm{d}\pi_n(x,y) = 0$. 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 (2.7). 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 $\mu \in \mathscr{P}(D)$ the *push-forward* of μ by T, denoted by $T_\sharp\mu \in \mathscr{P}(D)$ is given by:

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

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

$$\int_D \varphi(x)\ \mathrm{d}(T_\sharp\mu)(x) = \int_D \varphi(T(x))\ \mathrm{d}\mu(x). \tag{2.8}$$

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

$$\pi_T := (\mathrm{Id} \times T)_\sharp \mu, \tag{2.9}$$

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

$$\int_{D\times D} c(x,y)\,\mathrm{d}\pi_T(x,y) = \int_D c\,(x,T(x))\ \mathrm{d}\mu(x). \tag{2.10}$$

It is well known that when the measure $\mu \in \mathscr{P}(D)$ is absolutely continuous with respect to the Lebesgue measure, the problem on the right hand side of (2.5) is equivalent to:

$$\min\left\{\left(\int_D |x-T(x)|^p \,\mathrm{d}\mu(x)\right)^{1/p} \ :\ T_\sharp\mu = \tilde{\mu}\right\}, \tag{2.11}$$

and when p is strictly greater than 1, the problem (2.5) has a unique solution which is induced [via (2.9)] by a transportation map T solving (2.11) (see [60]). 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:

$$\int_D |x - T_n(x)|\,\mathrm{d}\mu(x) \to 0, \ \text{ as } n \to \infty. \tag{2.12}$$

We say that a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ is *stagnating* if it satisfies (2.12).

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

$$\pi^{-1} := s_{\sharp}\pi, \tag{2.13}$$

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

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

Let $\mu, \tilde{\mu}, \hat{\mu} \in \mathscr{P}(D)$. The *composition of plans* $\pi_{12} \in \Gamma(\mu, \tilde{\mu})$ and $\pi_{23} \in \Gamma(\tilde{\mu}, \hat{\mu})$ was discussed in [5, Remark5.3.3]. 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 $\pi_{13} = \pi_{23} \circ \pi_{12}$. Note $\pi_{13} \in \Gamma(\mu, \hat{\mu})$.

2.3. 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 measure ν, out of which the data points are sampled, with the empirical measure of data points while moving the mass as little as possible.

Let D be an open, bounded, connected domain on $\mathbb{R}^d$ with Lipschitz boundary. Let ν be a measure on D with density ρ which is bounded from below and from above by positive constants. Consider $(\Omega, \mathscr{F}, \mathbb{P})$ a probability space that we assume to be rich enough to support a sequence of independent random points $X_1, \dots, X_n, \dots$ distributed on D according to measure ν. We seek upper bounds on the transportation distance between ν and the empirical measures $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. It turned out that in the proof of Γ-convergence it was most useful to have estimates on the infinity transportation distance

$$d_\infty(\nu, \nu_n) = \inf\{\|Id - T_n\|_\infty \ : \ T_n : D \to D,\ T_{n\sharp}\nu = \nu_n\},$$

which measures what is the least maximal distance that a transportation map T_n between ν and ν_n has to move the mass.

If ν were a discrete measure with n particles, then the infinity transportation distance is the min–max matching distance. There is a rich history of discrete matching results (see [2,38,50,54–57] and references therein). In fact, let us first consider the case where $D = (0, 1)^d$ and ρ is constant, that is, assume the data points are uniformly distributed on $(0, 1)^d$. Also, assume, for simplicity, that n is 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 [38] we call the points in P grid points and the cubes generated by the points in P grid cubes.

For dimension $d = 2$, LEIGHTON and SHOR [38] showed that, when ρ is constant, 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 $c_1 > 0$):

$$\frac{c(\log n)^{3/4}}{n^{1/2}} \leqq \min_{\pi} \max_{i} |p_i - X_{\pi(i)}| \leqq \frac{C(\log n)^{3/4}}{n^{1/2}}, \tag{2.14}$$

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

For $d \geqq 3$, SHOR and YUKICH [50] proved the analogous result to (2.14). They showed that, when ρ is constant, there exist $c > 0$ and $C > 0$ such that with very high probability

$$\frac{c(\log n)^{1/d}}{n^{1/d}} \leqq \min_{\pi} \max_{i} |p_i - X_{\pi(i)}| \leqq \frac{C(\log n)^{1/d}}{n^{1/d}}. \tag{2.15}$$

The result in dimension $d \geqq 3$ is based on the matching algorithm introduced by AJTAI, KOMLÓS, and TUSNÁDY [2]. It relies on a dyadic decomposition of $(0, 1)^d$ 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 $d = 2$ the AKT algorithm still gives an upper bound, but not a sharp one. As remarked in [50], there is a crossover in the nature of the matching when $d = 2$: for $d \geqq 3$, 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 $d \geqq 3$ 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). The sharp result in dimension $d = 2$ by Leighton and Shor required a more sophisticated matching procedure. An alternative proof in $d = 2$ was provided by Talagrand [54] who also provided more streamlined and conceptually clear proofs in [55,56]. These results can be used to obtain bounds on the transportation distance in the continuum setting.

The results above were extended in [29] to the case of general domains and general measures with densities bounded from above and below by positive constants. Combined with Borel-Cantelli lemma they imply the following:

Theorem 2.5. *Let D be an open, connected and bounded subset of $\mathbb{R}^d$ which has Lipschitz boundary. Let ν be a probability measure on D with density ρ which is bounded from below and from above by positive constants. Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent random points distributed on D according to measure ν and let ν_n be the associated empirical measures* (1.5). *Then there is a constant $C > 0$ such that for $\mathbb{P}$-almost everywhere $\omega \in \Omega$ there exists a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from ν to ν_n ($T_{n\sharp}\nu = \nu_n$) and such that:*

$$\text{if } \; d = 2 \text{ then } \quad \limsup_{n\to\infty} \frac{n^{1/2}\|Id - T_n\|_\infty}{(\log n)^{3/4}} \leqq C \tag{2.16}$$

$$\text{and if } \; d \geqq 3 \text{ then } \quad \limsup_{n\to\infty} \frac{n^{1/d}\|Id - T_n\|_\infty}{(\log n)^{1/d}} \leqq C. \tag{2.17}$$

2.4. 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 2.6. 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,

$$\liminf_{n\to\infty} F_n(x_n) \geqq F(x);$$

2. **Limsup inequality:** For every $x \in X$ there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ converging to x satisfying

$$\limsup_{n\to\infty} F_n(x_n) \leqq F(x).$$

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 2.7. 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 2.8. 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

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

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

Remark 2.9. Note that the boundedness assumption of $\{x_k\}_{k\in\mathbb{N}}$ in the previous definition is a necessary condition for relative compactness 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 2.10. *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,*

$$\lim_{n\to\infty} \inf_{x\in X} F_n(x) = \min_{x\in X} F(x). \tag{2.18}$$

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

$$\lim_{n\to\infty} \left(F_n(x_n) - \inf_{x\in X} F_n(x) \right) = 0 \tag{2.19}$$

is relatively compact 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 (2.19) 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 $\left\{F_{h_n}\right\}_{n\in\mathbb{N}}$ 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 [16] or [25].

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 2.11. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. 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 2.6. 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 2.8.

We do not explicitly write the dependence of F_n on ω understanding that we are always working with a fixed value $\omega \in \Omega$, and hence with a deterministic functional.

3. The Space TL^p

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

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

For (μ, f) and (ν, g) in TL^p we define $d_{TL^p}((\mu, f), (\nu, g))$ by

$$d_{TL^p}((\mu, f), (\nu, g)) = \inf_{\pi\in\Gamma(\mu,\nu)} \left(\iint_{D\times D} |x-y|^p + |f(x)-g(y)|^p \, \mathrm{d}\pi(x, y) \right)^{1/p}. \tag{3.1}$$

Remark 3.1. We remark that formally TL^p is a fiber bundle over $\mathscr{P}(D)$; namely, if one considers the Finsler (Riemannian for $p = 2$) manifold structure on $\mathscr{P}(D)$ provided by the $p - OT$ metric (see [1] for general p and [5,43] for $p = 2$) then TL^p is, formally, a fiber bundle.

In order to prove that d_{TL^p} is a metric, we remark that d_{TL^p} is equal to a transportation distance between graphs of functions. To make this idea precise, let $\mathscr{P}_p(D\times\mathbb{R})$ be the space of Borel probability measures on the product space $D\times\mathbb{R}$ whose p-moment is finite. We consider the map

$$(\mu, f) \in TL^p \longmapsto (Id \times f)_\sharp \mu \in \mathscr{P}_p(D \times \mathbb{R}),$$

which allows us to identify an element $(\mu, f) \in TL^p$ with a measure in the product space $D \times \mathbb{R}$ whose support is contained in the graph of f.

For $\gamma, \tilde{\gamma} \in \mathscr{P}_p(D \times \mathbb{R})$ let $\mathbf{d}_p(\gamma, \tilde{\gamma})$ be given by

$$\left(\mathbf{d}_p(\gamma, \tilde{\gamma})\right)^p = \inf_{\pi\in\Gamma(\gamma,\tilde{\gamma})} \iint_{(D\times\mathbb{R})\times(D\times\mathbb{R})} |x-y|^p + |s-t|^p \,\mathrm{d}\pi((x,s),(y,t)).$$

Remark 3.2. We remark that $\mathbf{d}_p$ is a distance on $\mathscr{P}_p(D\times\mathbb{R})$ and that it is equivalent to the p-OT distance d_p introduced in Section 2.2 (the domain being $D \times \mathbb{R}$). Moreover, when $p = 2$ these two distances are actually equal.

Using the identification of elements in TL^p with probability measures in the product space $D \times \mathbb{R}$ we have the following:

Proposition 3.3. *Let* $(\mu, f), (\nu, g) \in TL^p$. *Then,* $d_{TL^p}((\mu, f), (\nu, g)) = \boldsymbol{d}_p((\mu, f), (\nu, g))$.

Proof. To see this, note that for every $\pi \in \Gamma((\mu, f), (\nu, g))$, it is true that the support of π is contained in the product of the graphs of f and g. In particular, we can write

$$\begin{aligned}\iint_{(D\times\mathbb{R})\times(D\times\mathbb{R})} & |x-y|^p + |s-t|^p \,\mathrm{d}\pi((x,s),(y,t)) \\ &= \iint_{D\times D} |x-y|^p + |f(x)-g(y)|^p \,\mathrm{d}\tilde{\pi}(x,y), \end{aligned} \tag{3.2}$$

where $\tilde{\pi} \in \Gamma(\mu, \nu)$. The right hand side of (3.2) is greater than $d_{TL^p}((\mu, f), (\nu, g))$, which together with the fact that π was arbitrary implies that $\mathbf{d}_p((\mu, f), (\nu, g)) \geqq d_{TL^p}((\mu, f), (\nu, g))$. To obtain the opposite inequality, it is enough to notice that for an arbitrary coupling $\tilde{\pi} \in \Gamma(\mu, \nu)$ we can consider the measure $\pi := ((Id \times f) \times (Id \times g))_\sharp \tilde{\pi}$ which belongs to $\Gamma((\mu, f), (\nu, g))$. Then, Equation (3.2) holds and its left hand side is greater than $d_{TL^p}((\mu, f), (\nu, g))$. The fact that $\tilde{\pi}$ was arbitrary allows us to conclude the opposite inequality. □

Remark 3.4. Proposition 3.3 and Remark 3.2 imply that (TL^p, d_{TL^p}) is a metric space.

Remark 3.5. We remark that the metric space (TL^p, d_{TL^p}) is not complete. To illustrate this, let us consider $D = (0, 1)$. Let μ be the Lebesgue measure on D and define $f_{n+1}(x) := \operatorname{sign}\sin(2^n \pi x)$ for $x \in (0, 1)$. Then, it can be shown that $d_{TL^p}((\mu, f_n), (\mu, f_{n+1})) \leqq 1/2^n$. This implies that the sequence $\{(\mu, f_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in (TL^p, d_{TL^p}). However, if this was a convergent sequence, in particular it would have to converge to an element of the form (μ, f) (see Proposition 3.12 below), but then, by Remark 3.9, it would be true that $f_n \overset{L^p(\mu)}{\longrightarrow} f$. This is impossible because $\{f_n\}_{n\in\mathbb{N}}$ is not a convergent sequence in $L^p(\mu)$.

Remark 3.6. The completion of the metric space (TL^p, d_{TL^p}) is the space $(\mathscr{P}_p(D\times \mathbb{R}), \mathbf{d}_p)$. In fact, in order to show this, it is enough to show that TL^p is dense in $(\mathscr{P}_p(D \times \mathbb{R}), \mathbf{d}_p)$. Since the class of convex combinations of Dirac delta masses is dense in $(\mathscr{P}_p(D \times \mathbb{R}), \mathbf{d}_p)$, it is enough to show that every convex combination of Dirac deltas can be approximated by elements in TL^p. So let us consider $\delta \in \mathscr{P}_p(D \times \mathbb{R})$ of the form

$$\delta = \sum_{i=1}^{m}\sum_{j=1}^{l_i} a_{ij}\delta_{(x_i,t_j^i)},$$

where $x_1, \ldots, x_n$ are n points in D; $t_i^j \in \mathbb{R}$; $a_{ij} > 0$ and $\sum_{i=1}^m \sum_{j=1}^{l_i} a_{ij} = 1$. Now, for every $n \in \mathbb{N}$ and for every $i = 1, \ldots, m$ choose $r_i^n > 0$ such that for all i: $B(x_i, r_i^n) \subseteq D$ and for all $k \neq i$, $B(x_i, r_i^n) \cap B(x_k, r_k^n) = \emptyset$ and such that $(\forall i)$ $r_i^n \leqq \frac{1}{n}$.

For $i = 1, \ldots, m$ consider $y_1^{i,n}, \ldots, y_{l_i}^{i,n}$ a collection of l_i points in $B(x_i, r_i^n)$. We define the function $f_n : D \to \mathbb{R}$ given by $f^n(x) = t_i^j$ if $x = y_j^{i,n}$ for some i, j and $f_n(x) = 0$ if not.

Finally, we define the measure $\mu_n \in \mathscr{P}(D)$ by

$$\mu_n = \sum_{i=1}^{m}\sum_{j=1}^{l_i} a_{ij}\delta_{y_j^{i,n}}.$$

It is straightforward to check that $(\mu_n, f_n) \overset{\mathbf{d}_p}{\longrightarrow} \delta$.

Remark 3.7. Here we make a connection between TL^p spaces and Young measures. Consider a fiber of TL^p over $\mu \in \mathscr{P}(D)$, that is, consider

$$TL_{p\llcorner\mu} := \left\{(\mu, f) \ : \ f \in L^p(\mu)\right\}.$$

Let $\text{Proj}_1 : D \times \mathbb{R} \mapsto D$ be defined by $\text{Proj}_1(x, t) = x$ and let

$$\mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu} := \left\{\gamma \in \mathscr{P}_p(D \times \mathbb{R}) \ : \ \text{Proj}_{1\sharp}\gamma = \mu\right\}.$$

Thanks to the disintegration theorem (see Theorem 5.3.1 in [5]), the set $\mathscr{P}_p(D\times \mathbb{R})_{\llcorner\mu}$ can be identified with the set of Young measures (or parametrized measures) with finite p-moment which have μ as base distribution (see [23,44]). It is straightforward to check that $\mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$ is a closed subset (in the $\mathbf{d}_p$ sense) of $\mathscr{P}_p(D \times \mathbb{R})$. Hence, the closure of $TL_{p\llcorner\mu}$ in $\mathscr{P}_p(D \times \mathbb{R})$ is contained in $\mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$, that is,

$$\overline{TL^p{}_{\llcorner\mu}} \subseteq \mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}.$$

In general, the inclusion may be strict. For example, if we let $D = (-1, 1)$ and consider $\mu = \delta_0$ to be the Dirac delta measure at zero, then it is straightforward to check that $TL^p{}_{\llcorner\mu}$ is actually a closed subset of $\mathscr{P}_p(D \times \mathbb{R})$ and that $TL^p{}_{\llcorner\mu} \subsetneq \mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$. On the other hand, if the measure μ is absolutely continuous with respect to the Lebesgue measure, then the closure of $TL_{p\llcorner\mu}$ is indeed $\mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$. This fact follows from Theorem 2.4.3 in [23]. Here we present a simple proof of this fact using the ideas introduced in the preliminaries. Note that it is enough to show that $TL^p{}_{\llcorner\mu}$ is dense in $\mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$. So let $\gamma \in \mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$. By Remark 3.6, there exists a sequence $\{((\mu_n, f_n)\}_{n\in\mathbb{N}} \subseteq TL^p$ such that

$$(\mu_n, f_n) \xrightarrow{\mathbf{d}_p} \gamma.$$

In particular,

$$\mu_n \xrightarrow{d_p} \mu.$$

Since μ is absolutely continuous with respect to the Lebesgue measure, for every $n \in \mathbb{N}$ there exists a transportation map $T_n : D \to D$ with $T_{n\sharp}\mu = \mu_n$, such that

$$\int_D |x - T_n(x)|^p d\mu(x) = (d_p(\mu, \mu_n))^p \to 0, \quad \text{as } n \to \infty.$$

On the other hand, the transportation map T_n induces the transportation plan $\pi_{T_n} \in \Gamma(\mu, \mu_n)$ defined in (2.9). Hence,

$$\begin{aligned}
(\mathbf{d}_p((\mu, f_n \circ T_n), (\mu_n, f_n)))^p &= (d_{TL^p}((\mu, f_n \circ T_n), (\mu_n, f_n)))^p \\
&\leqq \int_{D\times D} |x - y|^p \, \mathrm{d}\pi_{T_n}(x, y) \\
&\quad + \int_{D\times D} |f_n \circ T_n(x) - f_n(y)|^p \, \mathrm{d}\pi_{T_n}(x, y) \\
&= \int_D |x - T_n(x)|^p \, \mathrm{d}\mu(x).
\end{aligned}$$

From the previous computations, we deduce that $(\mathbf{d}_p((\mu, f_n \circ T_n), (\mu_n, f_n)) \to 0$ as $n \to \infty$, and thus $(\mu, f_n \circ T_n) \xrightarrow{\mathbf{d}_p} \gamma$. This shows that $TL^p{}_{\llcorner\mu}$ is dense in $\mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$, and given that $\mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$ is a closed subset of $\mathscr{P}_p(D \times \mathbb{R})$, we conclude that $\overline{TL^p{}_{\llcorner\mu}} = \mathscr{P}_p(D \times \mathbb{R})_{\llcorner\mu}$.

Remark 3.8. If one restricts the attention to measures $\mu, \nu \in \mathscr{P}(D)$ which are absolutely continuous with respect to the Lebesgue measure then

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

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 3.12.

Remark 3.9. One can think of the convergence in TL^p as a generalization of weak convergence of measures and of L^p convergence of functions. That is $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathscr{P}(D)$ converges weakly to $\mu \in \mathscr{P}(D)$ if and only if $(\mu_n, 1) \xrightarrow{TL^p} (\mu, 1)$ as $n \to \infty$ (which follows from the fact that on bounded sets the p-OT metric metrizes the weak convergence of measures [5]), and that for $\mu \in \mathscr{P}(D)$ a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $L^p(\mu)$ converges in $L^p(\mu)$ to f if and only if $(\mu, f_n) \xrightarrow{TL^p} (\mu, f)$ as $n \to \infty$. The last fact is established in Proposition 3.12.

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

Lemma 3.10. *Let $\mu \in \mathscr{P}(D)$ and let $\pi_n \in \Gamma(\mu, \mu)$ for all $n \in \mathbb{N}$. If $\{\pi_n\}_{n\in\mathbb{N}}$, is a stagnating sequence of transportation plans, then 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:

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

Note that

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

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

$$\begin{aligned}\iint_{D\times D} |u(x)-u(y)| \,\mathrm{d}\pi_n(x, y) \leqq & \iint_{D\times D} |u(x)-v(x)| \,\mathrm{d}\pi_n(x, y)\\ &+ \iint_{D\times D} |v(x)-v(y)| \,\mathrm{d}\pi_n(x, y)\\ &+ \iint_{D\times D} |v(y)-u(y)| \,\mathrm{d}\pi_n(x, y)\end{aligned}$$

$$= 2\int_D |v(x) - u(x)|\,\mathrm{d}\mu(x) + \iint_{D\times D} |v(x) - v(y)|\,\mathrm{d}\pi_n(x, y) < \varepsilon.$$

This proves the result. □

Lemma 3.11. *Suppose that the sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathscr{P}(D)$ converges weakly to $\mu \in \mathscr{P}(D)$. 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 $\left\{\hat{\pi}_n\right\}_{n\in\mathbb{N}}$ [with $\pi_n, \hat{\pi}_n \in \Gamma(\mu, \mu_n)$]. Then:*

$$\lim_{n\to\infty} \iint_{D\times D} |u(x) - u_n(y)|^p\, d\pi_n(x, y) = 0 \Leftrightarrow \lim_{n\to\infty} \iint_{D\times D} |u(x) - u_n(y)|^p\, d\hat{\pi}_n(x, y) = 0 \tag{3.3}$$

Proof. We present the details for $p = 1$, as the other cases are similar. Take $\hat{\pi}_n^{-1} \in \Gamma(\mu_n, \mu)$ as the inverse of $\hat{\pi}_n$ defined in (2.13). We can consider $\boldsymbol{\pi_n} \in \mathscr{P}(D \times D \times D)$ as the measure mentioned at the end of Section 2.2 (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

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

and

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

which implies, after using the triangle inequality:

$$\left|\iint_{D\times D} |u_n(y) - u(x)|\,\mathrm{d}\pi_n(x, y) - \iint_{D\times D} |u(z) - u_n(y)|\,\mathrm{d}\hat{\pi}_n(y, z)\right| \leqq \iiint_{D\times D\times D} |u(z) - u(x)|\,\mathrm{d}\boldsymbol{\pi}_n(x, y, z) = \iint_{D\times D} |u(z) - u(x)|\,\mathrm{d}\hat{\pi}_n^{-1} \circ \pi_n(x, z).$$

Finally, note that

$$\iint_{D\times D} |x - z|\,\mathrm{d}\hat{\pi}_n^{-1} \circ \pi_n(x, z) \leqq \iint_{D\times D} |x - y|\,\mathrm{d}\pi_n(x, y) + \iint_{D\times D} |y - z|\,\mathrm{d}\hat{\pi}_n(z, y) \to 0,$$

as $n \to \infty$. The sequence $\left\{\hat{\pi}_n^{-1} \circ \pi_n\right\}_{n\in\mathbb{N}}$ satisfies the assumptions of Lemma 3.10, so we can deduce that $\iint_{D\times D} |u(z) - u(x)| \, \mathrm{d}\hat{\pi}_n^{-1} \circ \pi_n(x, z) \to 0$ as $n \to \infty$. By (3.4) we get that:

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

This implies the result. □

Proposition 3.12. *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* $\pi_n \in \Gamma(\mu, \mu_n)$*)*

$$\iint_{D\times D} |f(x) - f_n(y)|^p \, d\pi_n(x, y) \to 0, \quad \text{as } n \to \infty. \tag{3.4}$$

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 (3.4) 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:*

$$\int_D |f(x) - f_n(T_n(x))|^p \, d\mu(x) \to 0, \quad \text{as } n \to \infty; \tag{3.5}$$

5. $\mu_n \xrightarrow{w} \mu$ *and for any stagnating sequence of transportation maps* $\{T_n\}_{n\in\mathbb{N}}$ *(with* $T_{n\sharp}\mu = \mu_n$*) (3.5) holds.*

Proof. By Lemma 3.11, 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 $\left\{\pi_{T_n}\right\}_{n\in\mathbb{N}}$ [as defined in (2.9)] and using (2.10) we see that 2., 3., 4., and 5. are all equivalent. We prove the equivalence of 1. and 3.
(1. $\Rightarrow$ 3.) Note that $d_p(\mu, \mu_n) \leqq d_{TL^p}\left((\mu, f), (\mu_n, f_n)\right)$ for every n. Consequently $d_p(\mu, \mu_n) \to 0$ as $n \to \infty$ and in particular $\mu_n \xrightarrow{w} \mu$ as $n \to \infty$. Furthermore, since $d_{TL^p}\left((\mu, f), (\mu_n, f_n)\right) \to 0$ as $n \to \infty$, there exists a sequence $\left\{\pi_n^*\right\}_{n\in\mathbb{N}}$ of transportation plans (with $\pi_n^* \in \Gamma(\mu, \mu_n)$) such that:

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

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

Then, $\{\pi_n^*\}_{n\in\mathbb{N}}$ is a stagnating sequence of transportation plans for which (3.4) 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

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

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

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

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

Definition 3.13. Suppose $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathscr{P}(D)$ converges weakly to $\mu \in \mathscr{P}(D)$. 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 relatively compact in TL^p if the sequence $\{(\mu_n, u_n)\}_{n\in\mathbb{N}}$ is relatively compact in TL^p.

Remark 3.14. Thanks to Proposition 3.12 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 $n \to \infty$ (this in particular implies the last part of Remark 3.9). Also $\{u_n\}_{n\in\mathbb{N}}$ is relatively compact 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 relatively compact in $L^p(\mu)$.

In the light of Proposition 3.12 and Remark 3.7, we finish this section by illustrating a further connection between Young measures and the TL^p space and also we provide a geometric characterization of L^p-convergence. These connections follow from Theorem 2.4.3 in [23], nevertheless, we decided to present them in the context of the tools and results presented in this section. Let us consider μ to be the Lebesgue measure. The set $L^p(\mu)$ can be identified with the fiber $TL_{p \llcorner \mu}$ in a canonical way:

$$f \in L^p(\mu) \mapsto (\mu, f) \in TL^p{\llcorner}_\mu.$$

Thus, we can endow $L^p(\mu)$ with the distance d_{TL^p}. Note that by Remark 3.9, the topologies in $L^p(\mu)$ generated by d_{TL^p} and $\|\cdot\|_{L^p(\mu)}$ are the same. However, Remark 3.5 implies that d_{TL^p} and the distance generated by the norm $\|\cdot\|_{L^p(\mu)}$ are not equivalent. Note that the space $L^p(\mu)$ endowed with the norm $\|\cdot\|_{L^p(\mu)}$ is a complete metric space. On the other hand, by Remark 3.7, the completion of $L^p(\mu)$ endowed with the metric d_{TL^p} is $\mathscr{P}_p(D\times\mathbb{R})_{\llcorner\mu}$ with $\mathbf{d}_p$ as distance. This is a characterization for the class of Young measures with finite p-moment, namely, they can be interpreted as the completion of the space $L^p(\mu)$ endowed with the metric d_{TL^p}. Regarding the geometric interpretation of L^p-convergence, we have the following:

Corollary 3.15. *Let μ be the Lebesgue measure on D. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $L^p(\mu)$ and let $f \in L^p(\mu)$. Then, $\{f_n\}_{n\in\mathbb{N}}$ converges to f in $L^p(\mu)$ if and only if the graphs of f_n converge to the graph of f in the p-OT sense.*

Proof. From Remark 3.9, the sequence $\{f_n\}_{n\in\mathbb{N}}$ converges to f in $L^p(\mu)$ if and only if the sequence $\{(\mu, f_n)\}_{n\in\mathbb{N}}$ converges to (μ, f) in TL^p. This implies the result, because TL^p distance is equivalent to the p-OT distance defined on $\mathscr{P}_p(D \times \mathbb{R})$ (see Proposition 3.3 and Remark 3.2). □

4. Γ-Convergence of $TV_\varepsilon(\cdot, \rho)$

In this section we prove the Γ-convergence of the nonlocal functionals $TV_\varepsilon(\cdot, \rho)$ to the weighted total variation with weight ρ^2.

Theorem 4.1. *Consider an open, bounded domain D in $\mathbb{R}^d$ with Lipschitz boundary. Let $\rho : D \to \mathbb{R}$ be continuous and bounded below and above by positive constants. Then, $\{TV_\varepsilon(\cdot; \rho)\}_{\varepsilon>0}$ [defined in (1.9)] Γ-converges with respect to the $L^1(D, \rho)$-metric to $\sigma_\eta TV(\cdot, \rho^2)$. Moreover, the functionals $\{TV_\varepsilon(\cdot; \rho)\}_{\varepsilon>0}$ satisfy the compactness property (Definition 2.8) with respect to the $L^1(D, \rho)$-metric.*

Part of the proof of this result follows ideas present in the work of Ponce [46]. Specifically, Lemma 4.2 below and the first part of the proof of the liminf inequality are adaptations of results by Ponce. The first part of the proof of the limsup inequality is a careful adaptation of the appendix of a paper by Alberti and Bellettini [3].

We also prove compactness of the functionals $\{TV_\varepsilon(\cdot; \rho)\}_{\varepsilon>0}$.This part required new arguments, due to the presence of domain boundary and lack of L^∞-control. Part of the proof on compactness in [3] is used. 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 the regularity of its boundary.

Since the definition of Γ-convergence for a family of functionals indexed by real numbers is given in terms of sequences, in this section we adopt the following notation: ε is a short-hand notation for ε_n where $\{\varepsilon_n\}_{n\in\mathbb{N}}$ is an arbitrary sequence of positive real numbers converging to zero as $n \to \infty$. Limits as $\varepsilon \to 0$ simply mean limits as $n \to \infty$ for every such sequence.

Lemma 4.2. *Let D be a bounded open subset of $\mathbb{R}^d$ and let $\rho : D \to \mathbb{R}$ be a Lipschitz function that is bounded from below and from above by positive constants. Suppose that $\{u_\varepsilon\}_{\varepsilon>0}$ is a sequence of C^2 functions such that*

$$\sup_{\varepsilon>0} \left\{ ||\nabla u_\varepsilon||_{L^\infty(\mathbb{R}^d)} + ||D^2 u_\varepsilon||_{L^\infty(\mathbb{R}^d)} \right\} < \infty. \tag{4.1}$$

If $\nabla u_\varepsilon \overset{L^1(D)}{\longrightarrow} \nabla u$ *for some* $u \in C^2(\mathbb{R}^d)$*, then*

$$\lim_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho) = \sigma_\eta \int_D |\nabla u(x)|(\rho(x))^2\,dx. \tag{4.2}$$

Proof. **Step 1:** For an arbitrary function $v \in C^2(\mathbb{R}^d)$ we define

$$H_\varepsilon(v) = \frac{1}{\varepsilon}\int_D\int_D \eta_\varepsilon(x-y)|\nabla v(x)\cdot(y-x)|\rho(x)\rho(y)\,\mathrm{d}y\,\mathrm{d}x.$$

First we show that

$$\lim_{\varepsilon\to 0} |TV_\varepsilon(u_\varepsilon;\rho) - H_\varepsilon(u_\varepsilon)| = 0. \tag{4.3}$$

For this purpose, note that by Taylor's theorem and by (4.1), for $x, y \in D$ $x \neq y$ and $\varepsilon > 0$

$$\left|\frac{u_\varepsilon(x)-u_\varepsilon(y)}{|x-y|} - \frac{\nabla u_\varepsilon(x)\cdot(y-x)}{|x-y|}\right| \leqq ||D^2u_\varepsilon||_{L^\infty(\mathbb{R}^d)}|x-y| \leqq C|x-y|,$$

where $||D^2u_\varepsilon||_{L^\infty(\mathbb{R}^d)}$ denotes the L^∞ norm of the Hessian matrix of the function u_ε and C is a positive constant independent of ε. Using this inequality and a simple change of variables we deduce

$$\begin{aligned}|TV_\varepsilon(u_\varepsilon;\rho) - H_\varepsilon(u_\varepsilon)| &\leqq \frac{C\,\mathrm{Vol}(D)||\rho||^2_{L^\infty(D)}}{\varepsilon}\int_{|h|\leqq\gamma}\eta_\varepsilon(h)|h|^2\,\mathrm{d}h \\ &= C\,\mathrm{Vol}(D)||\rho||^2_{L^\infty(D)}\int_{|\hat{h}|\leqq\frac{\gamma}{\varepsilon}}\varepsilon\eta(\hat{h})|\hat{h}|^2\,\mathrm{d}\hat{h},\end{aligned}$$

where γ denotes the diameter of the set D. Finally, using assumption (K3) on the kernel η, it is straightforward to deduce that the last term in the previous expression goes to zero as ε goes to zero, and thus we obtain (4.3).

Step 2: Now, for $v \in C^2(\mathbb{R}^d)$ consider

$$\tilde{H}_\varepsilon(v) = \frac{1}{\varepsilon}\int_D\int_{x+h\in D}\eta_\varepsilon(h)\,|\nabla v(x)\cdot h|\,(\rho(x))^2\,\mathrm{d}h\,\mathrm{d}x. \tag{4.4}$$

We claim that

$$\lim_{\varepsilon\to 0}\left|H_\varepsilon(u_\varepsilon) - \tilde{H}_\varepsilon(u_\varepsilon)\right| = 0. \tag{4.5}$$

Indeed, using the fact that ρ is Lipschitz,

$$\begin{aligned}&\left|H_\varepsilon(u_\varepsilon) - \tilde{H}_\varepsilon(u_\varepsilon)\right| \\ &\leqq \frac{1}{\varepsilon}\int_D\int_{x+h\in D}\eta_\varepsilon(h)\,|\nabla u_\varepsilon(x)\cdot h|\,|\rho(x+h)-\rho(x)|\,\rho(x)\,\mathrm{d}h\,\mathrm{d}x \\ &\leqq \frac{||\nabla u_\varepsilon||_{L^\infty(\mathbb{R}^d)}\,\mathrm{Lip}(\rho)||\rho||_{L^\infty(D)}}{\varepsilon}\int_D\int_{x+h\in D}\eta_\varepsilon(h)|h|^2\,\mathrm{d}h\,\mathrm{d}x \\ &\leqq \frac{||\nabla u_\varepsilon||_{L^\infty(\mathbb{R}^d)}\,\mathrm{Lip}(\rho)||\rho||_{L^\infty(D)}\,\mathrm{Vol}(D)}{\varepsilon}\int_{|h|<\gamma}\eta_\varepsilon(h)|h|^2\,\mathrm{d}h,\end{aligned}$$

where as in Step 1 γ denotes the diameter of the set D. The last term in the previous expression goes to zero as ε goes to zero (as in Step 1).

Step 3: We claim that

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_D\int_{x+h\in\mathbb{R}^d\setminus D}\eta_\varepsilon(h)\,|\nabla u_\varepsilon(x)\cdot h|\,(\rho(x))^2\,\mathrm{d}h\,\mathrm{d}x=0. \tag{4.6}$$

Note that,

$$\begin{aligned}&\frac{1}{\varepsilon}\int_D\int_{x+h\in\mathbb{R}^d\setminus D}\eta_\varepsilon(h)\,|\nabla u_\varepsilon(x)\cdot h|\,(\rho(x))^2\,\mathrm{d}h\,\mathrm{d}x\\&\quad\leqq ||\nabla u_\varepsilon||_{L^\infty(\mathbb{R}^d)}||\rho||^2_{L^\infty(D)}\int_D\int_{x+\varepsilon\hat{h}\in\mathbb{R}^d\setminus D}\eta(\hat{h})|\hat{h}|\,\mathrm{d}\hat{h}\,\mathrm{d}x.\end{aligned}$$

Using (4.1) and assumption (K3) on η, we deduce that the right hand side of the previous inequality goes to zero as ε goes to zero, thus implying (4.6).

Step 4: Using steps 1, 2, and 3 in order to obtain (4.2) it is enough to prove that

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_D\int_{\mathbb{R}^d}\eta_\varepsilon(h)|\nabla u_\varepsilon(x)\cdot h|(\rho(x))^2\,\mathrm{d}h\,\mathrm{d}x=\sigma_\eta\int_D|\nabla u|(\rho(x))^2\,\mathrm{d}x. \tag{4.7}$$

Note that using the change of variables $\hat{h}=\frac{h}{\varepsilon}$ and the isotropy of the kernel η, imply

$$\begin{aligned}&\frac{1}{\varepsilon}\int_D\int_{\mathbb{R}^d}\eta_\varepsilon(h)|\nabla u_\varepsilon(x)\cdot h|(\rho(x))^2\,\mathrm{d}h\,\mathrm{d}x\\&\quad=\int_D\left(\int_{\mathbb{R}^d}\eta(\hat{h})|\nabla u_\varepsilon(x)\cdot\hat{h}|\,\mathrm{d}\hat{h}\right)(\rho(x))^2\,\mathrm{d}x\\&\quad=\sigma_\eta\int_D|\nabla u_\varepsilon(x)|(\rho(x))^2\,\mathrm{d}x.\end{aligned}$$

Taking ε to zero in the previous expression we obtain (4.7), and consequently (4.2). □

4.1. Proof of Theorem 4.1: the Liminf Inequality

Proof. **Case 1: ρ is Lipschitz.** Consider an arbitrary $u\in L^1(\rho)$ and suppose that $u_\varepsilon\xrightarrow{L^1(\rho)}u$ as $\varepsilon\to 0$. Recall that given the assumptions on ρ this is equivalent to $u_\varepsilon\xrightarrow{L^1(D)}u$ as $\varepsilon\to 0$. We want to show that $\liminf_{\varepsilon\to 0}TV_\varepsilon(u_\varepsilon;\rho)\geqq\sigma_\eta TV(u;\rho^2)$. Without the loss of generality we can assume that $\{TV_\varepsilon(u_\varepsilon;\rho)\}_{\varepsilon>0}$ is bounded.

The idea is to reduce the problem to a setting where we can use Lemma 4.2. The plan is to first regularize the functions u_ε to obtain a new sequence of functions $\{u_{\varepsilon,\delta}\}_{\varepsilon>0}$ ($\delta>0$ is a parameter that controls the smoothness of the regularized functions). The point is that regularizing does not increase the energy in the limit, while it gains the regularity needed to use Lemma 4.2.

To make this idea precise, consider $J : \mathbb{R}^d \to [0, \infty)$ a standard mollifier. That is, J is a smooth radially symmetric function, supported in the closed unit ball $\overline{B(0,1)}$ and is such that $\int_{\mathbb{R}^d} J(z)\,\mathrm{d}z = 1$. We set J_δ to be $J_\delta(z) = \frac{1}{\delta^d} J\left(\frac{z}{\delta}\right)$. Note that $\int_{\mathbb{R}^d} J_\delta(z)\,\mathrm{d}z = 1$ for every $\delta > 0$.

Fix D' an open set compactly contained in D. There exists $\delta' > 0$ such that $D'' = \bigcup_{x \in D'} B(x, \delta')$ is contained in D. For $0 < \delta < \delta'$ and for a given function $v \in L^1(D)$ we define the mollified function $v_\delta \in L^1(\mathbb{R}^d)$ by setting $v_\delta(x) = \int_{\mathbb{R}^d} J_\delta(x-z)v(z)\,\mathrm{d}z = \int_{\mathbb{R}^d} J_\delta(z)v(x-z)\,\mathrm{d}z$ where we have extended v to be zero outside of D. The functions v_δ are smooth, and satisfy $v_\delta \overset{L^1(D')}{\longrightarrow} v$ as $\delta \to 0$, see for example [39]. Furthermore

$$\nabla v_\delta(x) = \int_{\mathbb{R}^d} \nabla J_\delta(z) v(x-z)\,\mathrm{d}z = \frac{1}{\delta}\int_{\mathbb{R}^d} \frac{1}{\delta^d} \nabla J\left(\frac{z}{\delta}\right) v(x-z)\,\mathrm{d}z. \tag{4.8}$$

By taking the second derivative, it follows that there is a constant $C > 0$ (only depending on the mollifier J) such that

$$||\nabla v_\delta||_{L^\infty(\mathbb{R}^d)} \leqq \frac{C}{\delta}||v||_{L^1(D)} \quad \text{and} \quad ||D^2 v_\delta||_{L^\infty(\mathbb{R}^d)} \leqq \frac{C}{\delta^2}||v||_{L^1(D)}. \tag{4.9}$$

Since $u_\varepsilon \overset{L^1(D)}{\longrightarrow} u$ as $\varepsilon \to 0$ the norms $||u_\varepsilon||_{L^1(D)}$ are uniformly bounded. Therefore, taking $v = u_\varepsilon$ in inequalities (4.9) and setting $u_{\varepsilon,\delta} = (u_\varepsilon)_\delta$, implies

$$\sup_{\varepsilon > 0}\left\{||\nabla u_{\varepsilon,\delta}||_{L^\infty(\mathbb{R}^d)} + ||D^2 u_{\varepsilon,\delta}||_{L^\infty(\mathbb{R}^d)}\right\} < \infty.$$

Moreover, using (4.8) to express $\nabla u_{\varepsilon,\delta}$ and ∇u_δ, it is straightforward to deduce that

$$\int_{D'} \left|\nabla u_{\varepsilon,\delta}(x) - \nabla u_\delta(x)\right|\,\mathrm{d}x \leqq \frac{C}{\delta}\int_D |u_\varepsilon(x) - u(x)|\,\mathrm{d}x$$

for some constant C independent of ε. In particular, $\int_{D'} \left|\nabla u_{\varepsilon,\delta}(x) - \nabla u_\delta(x)\right|\,\mathrm{d}x \to 0$ as $\varepsilon \to 0$ and hence we can apply Lemma 4.2 taking D to be D') to infer that

$$\begin{aligned}&\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\int_{D'}\int_{D'} \eta_\varepsilon(x-y)|u_{\varepsilon,\delta}(x) - u_{\varepsilon,\delta}(y)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \sigma_\eta \int_{D'} |\nabla u_\delta(x)|(\rho(x))^2\,\mathrm{d}x\,\mathrm{d}y.\end{aligned} \tag{4.10}$$

To measure the approximation error in the energy, we set

$$\begin{aligned}a_{\varepsilon,\delta} = \frac{1}{\varepsilon}\int_{D''}\int_{D''}\int_{\mathbb{R}^d} & J_\delta(z)\eta_\varepsilon(x-y)|u_\varepsilon(x) \\ & - u_\varepsilon(y)|\left(\rho(x)\rho(y) - \rho(x+z)\rho(y+z)\right)\,\mathrm{d}z\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

and estimate

$$
\begin{aligned}
TV_\varepsilon(u_\varepsilon;\rho) &\geqq \frac{1}{\varepsilon}\int_{D''}\int_{D''}\eta_\varepsilon(x-y)|u_\varepsilon(x)-u_\varepsilon(y)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y\\
&= \frac{1}{\varepsilon}\int_{D''}\int_{D''}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(x-y)|u_\varepsilon(x)-u_\varepsilon(y)|\rho(x)\rho(y)\,\mathrm{d}z\,\mathrm{d}x\,\mathrm{d}y\\
&= a_{\varepsilon,\delta}+\frac{1}{\varepsilon}\int_{D''}\int_{D''}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(x-y)|u_\varepsilon(x)-u_\varepsilon(y)|\rho(x+z)\rho(y+z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x\\
&\geqq a_{\varepsilon,\delta}+\frac{1}{\varepsilon}\int_{D'}\int_{D'}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(\hat{x}-\hat{y})|u_\varepsilon(\hat{x}-z)-u_\varepsilon(\hat{y}-z)|\rho(\hat{x})\rho(\hat{y})\,\mathrm{d}z\,\mathrm{d}\hat{y}\,\mathrm{d}\hat{x}\\
&\geqq a_{\varepsilon,\delta}+\frac{1}{\varepsilon}\int_{D'}\int_{D'}\eta_\varepsilon(\hat{x}-\hat{y})\left|\int_{\mathbb{R}^d} J_\delta(z)\left(u_\varepsilon(\hat{x}-z)-u_\varepsilon(\hat{y}-z)\right)\mathrm{d}z\right|\rho(\hat{x})\rho(\hat{y})\,\mathrm{d}\hat{y}\,\mathrm{d}\hat{x}\\
&= a_{\varepsilon,\delta}+\frac{1}{\varepsilon}\int_{D'}\int_{D'}\eta_\varepsilon(\hat{x}-\hat{y})|u_{\varepsilon,\delta}(\hat{x})-u_{\varepsilon,\delta}(\hat{y})|\rho(\hat{x})\rho(\hat{y})\,\mathrm{d}\hat{y}\,\mathrm{d}\hat{x},
\end{aligned}
$$

where the second inequality is obtained using the change of variables $\hat{x}=x+z$, $\hat{y}=y+z$, $z=z$ together with the choice of δ and δ'; Jensen's inequality justifies the third one. This chain of inequalities and (4.10) imply that

$$
\liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho) \geqq \liminf_{\varepsilon\to 0} a_{\varepsilon,\delta}+\sigma_\eta\int_{D'}|\nabla u_\delta(x)|(\rho(x))^2\,\mathrm{d}x. \tag{4.11}
$$

We estimate $a_{\varepsilon,\delta}$ as follows:

$$
\begin{aligned}
|a_{\varepsilon,\delta}| &\leqq \frac{2||\rho||_{L^\infty}}{\varepsilon}\int_{D''}\int_{D''}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(x-y)\,|u_\varepsilon(x)-u_\varepsilon(y)|\\
&\qquad \times|\rho(x)-\rho(x+z)|\,\mathrm{d}z\,\mathrm{d}x\,\mathrm{d}y\\
&\leqq \frac{2\delta||\rho||_{L^\infty}\,\mathrm{Lip}(\rho)}{\varepsilon}\int_{D''}\int_{D''}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(x-y)\,|u_\varepsilon(x)-u_\varepsilon(y)|\,\mathrm{d}z\,\mathrm{d}x\,\mathrm{d}y\\
&= \frac{2\delta||\rho||_{L^\infty}\,\mathrm{Lip}(\rho)}{\varepsilon}\int_{D''}\int_{D''}\eta_\varepsilon(x-y)\,|u_\varepsilon(x)-u_\varepsilon(y)|\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}
$$

Since we had assumed that $\{TV_\varepsilon(u_\varepsilon;\rho)\}_{\varepsilon>0}$ is bounded, and also that ρ is bounded from below by a positive constant, we conclude from the previous inequalities that $\liminf_{\delta\to 0}\liminf_{\varepsilon\to 0} a_{\varepsilon,\delta}=0$ and thus, by (4.11),

$$
\liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho)\geqq \sigma_\eta \liminf_{\delta\to 0}\int_{D'}|\nabla u_\delta|(\rho(x))^2\,\mathrm{d}x.
$$

Given that $u_\delta \to_{L^1(D')} u$ as $\delta\to 0$, we can use the lower semicontinuity of the weighted total variation, (2.4), to obtain

$$
\liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho)\geqq \sigma_\eta \liminf_{\delta\to 0}\int_{D'}|\nabla u_\delta|(\rho(x))^2\,\mathrm{d}x \geqq \sigma_\eta |Du|_{\rho^2}(D'). \tag{4.12}
$$

Given that D' was an arbitrary open set compactly contained in D, we can take $D'\nearrow D$ in the previous inequality to obtain the desired result.

Case 2: ρ is continuous but not necessarily Lipschitz. The idea is to approximate ρ from below by a family of Lipschitz functions $\{\rho_k\}_{k\in\mathbb{N}}$. Indeed, consider $\rho_k : D \to \mathbb{R}$ given by

$$\rho_k(x) := \inf_{y\in D} \rho(y) + k|x-y|. \tag{4.13}$$

The functions ρ_k are Lipschitz functions which are bounded from below and from above by the same constants bounding ρ from below and from above. Moreover, given that ρ is continuous, for every $x \in D$, $\rho_k(x) \nearrow \rho(x)$ as $k \to \infty$.

Let $u \in L^1(D)$ and suppose that $u_\varepsilon \xrightarrow{L^1(D)} u$. Since ρ_k is Lipschitz, we can use Case 1 and the fact that $\rho_k \leqq \rho$ to conclude that

$$\liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho) \geqq \liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho_k) \geqq \sigma_\eta TV(u;\rho_k^2). \tag{4.14}$$

Using (2.3) and the monotone convergence theorem, we see that:

$$\lim_{k\to\infty} TV(u;\rho_k^2) = \lim_{k\to\infty}\int_D \rho_k^2(x)\,\mathrm{d}|Du|(x) = \int_D \rho^2(x)\,\mathrm{d}|Du|(x) = TV(u;\rho^2).$$

Combining with (4.14) yields the desired result. □

4.2. *Proof of Theorem 4.1: The Limsup Inequality*

Proof. Case 1: ρ is Lipschitz. We start by noting that since $\rho : D \to \mathbb{R}^d$ is a Lipschitz function, there exists an extension (that we denote by ρ as well) to the entire $\mathbb{R}^d$ which has the same Lipschitz constant as the original ρ and is bounded below by the same positive constant. Indeed, the extended function $\rho : \mathbb{R}^d \to \mathbb{R}$ can be defined by $\rho(x) = \inf_{y\in D}\rho(y) + \mathrm{Lip}(\rho)|x-y|$, where $\mathrm{Lip}(\rho)$ is the Lipschitz constant of ρ.

To prove the limsup inequality we show that for every $u \in L^1(\rho)$:

$$\limsup_{\varepsilon\to 0} TV_\varepsilon(u;\rho) \leqq \sigma_\eta TV(u;\rho^2). \tag{4.15}$$

It suffices to show (4.15) for functions $u \in BV(D)$ [if the right hand side of (4.15) is $+\infty$ there is nothing to prove]. Since D has Lipschitz boundary, for a given $u \in BV(D)$ we use Proposition 3.21 in [4] to obtain an extension $\hat{u} \in BV(\mathbb{R}^d)$ of u to the entire space $\mathbb{R}^d$ with $\left|D\hat{u}\right|(\partial D) = 0$. In particular from (2.2) we obtain

$$\left|D\hat{u}\right|_{\rho^2}(\partial D) = 0. \tag{4.16}$$

We split the proof of (4.15) in two cases:

Step 1: Suppose that η has compact support, that is assume there is $\alpha > 0$ such that if $|h| \geqq \alpha$ then $\eta(h) = 0$. Let $D_\varepsilon := \left\{x \in \mathbb{R}^d : \mathrm{dist}(x,D) < \alpha\varepsilon\right\}$. For $u \in BV(D)$, Theorem 3.4 in [9] and our assumptions on ρ provide a sequence of functions $\{v_k\}_{k\in\mathbb{N}} \in C^\infty(D_\varepsilon) \cap BV(D_\varepsilon)$ such that as $k \to \infty$

$$v_k \xrightarrow{L^1(D_\varepsilon)} \hat{u} \quad \text{and} \quad \int_{D_\varepsilon} |\nabla v_k(x)|\rho^2(x)\,\mathrm{d}x \to |D\hat{u}|_{\rho^2}(D_\varepsilon). \tag{4.17}$$

For every $k \in \mathbb{N}$

$$
\begin{aligned}
TV_\varepsilon(v_k;\rho) &= \frac{1}{\varepsilon}\int_D \int_{D\cap B(y,\alpha\varepsilon)} \eta_\varepsilon(x-y)|v_k(x)-v_k(y)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{\varepsilon}\int_D \int_{B(y,\alpha\varepsilon)} \eta_\varepsilon(x-y)\left|\int_0^1 \nabla v_k(y+t(x-y))\cdot(x-y)\,\mathrm{d}t\right| \\
&\quad \rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\
&\leqq \frac{1}{\varepsilon}\int_D \int_{B(y,\alpha\varepsilon)} \int_0^1 \eta_\varepsilon(x-y)|\nabla v_k(y+t(x-y))\cdot(x-y)| \\
&\quad \rho(x)\rho(y)\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y \\
&\leqq \int_{D_\varepsilon}\int_{|h|<\alpha}\int_0^1 \eta(h)|\nabla v_k(z)\cdot h|\rho(z-t\varepsilon h)\rho(z+(1-t)\varepsilon h)\,\mathrm{d}t\,\mathrm{d}h\,\mathrm{d}z \\
&= \int_{D_\varepsilon}\int_{|h|<\alpha} \eta(h)|\nabla v_k(z)\cdot h|\rho(z)^2\,\mathrm{d}h\,\mathrm{d}z + a_{\varepsilon,k} \\
&= \sigma_\eta \int_{D_\varepsilon} |\nabla v_k(z)|\,(\rho(z))^2\,\mathrm{d}z + a_{\varepsilon,k},
\end{aligned}
$$

where the last inequality is obtained after using the change of variables $(t,y,x)\mapsto(t,h,z)$, $h=\frac{x-y}{\varepsilon}$ and $z=y+t(x-y)$, noting that the Jacobian of this transformation is equal to ε^d and that the transformed set D is contained in D_ε. The last equality is obtained thanks to the fact that η is radially symmetric. Finally the $a_{\varepsilon,k}$ are given by

$$
\begin{aligned}
a_{\varepsilon,k} = \int_{D_\varepsilon}\int_{|h|<\alpha}\int_0^1 \eta(h)|\nabla v_k(z) \\
\cdot h|\left(\rho(z-t\varepsilon h)\rho(z+(1-t)\varepsilon h)-\rho(z)^2\right)\,\mathrm{d}t\,\mathrm{d}h\,\mathrm{d}z.
\end{aligned}
$$

Since $\rho:\mathbb{R}^d\to\mathbb{R}$ is Lipschitz and since it is bounded below by a positive constant, it is straightforward to show that there exists a constant $C>0$ independent of ε and k for which

$$
a_{\varepsilon,k} \leqq C\varepsilon\int_{D_\varepsilon} |\nabla v_k(x)|\rho^2(x)\,\mathrm{d}x.
$$

Using (4.17) in particular, we obtain that $v_k \overset{L^1(D)}{\longrightarrow} u$ as $k\to\infty$. This, together with the continuity of $TV_\varepsilon(\cdot;\rho)$ with respect to L^1-convergence implies that $TV_\varepsilon(v_k;\rho) \to TV_\varepsilon(u;\rho)$ as $k\to\infty$. Therefore, from the previous chain of inequalities and from (4.17) we conclude that

$$
TV_\varepsilon(u;\rho) \leqq \sigma_\eta|D\hat{u}|_{\rho^2}(D_\varepsilon) + \limsup_{k\to\infty} a_{\varepsilon,k} \leqq \sigma_\eta|D\hat{u}|_{\rho^2}(D_\varepsilon) + C\varepsilon|D\hat{u}|_{\rho^2}(D_\varepsilon). \tag{4.18}
$$

Using (4.16), we deduce $\lim_{\varepsilon\to 0}|D\hat{u}|_{\rho^2}(D_\varepsilon) = |D\hat{u}|_{\rho^2}(\overline{D}) = |D\hat{u}|_{\rho^2}(D) = TV(u;\rho^2) < \infty$. Combining with (4.18) implies the desired estimate, (4.15).

Step 2: Consider η whose support is not compact. The needed control of η at infinity 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 $TV_\varepsilon^\alpha(\cdot,\rho)$ the nonlocal total variation using the kernel η^α. For a given $u \in BV(D)$

$$TV_\varepsilon(u;\rho) = TV_\varepsilon^\alpha(u;\rho) + \frac{1}{\varepsilon}\int_D \int_{\{x\in D\,:\,|x-y|>\alpha\varepsilon\}} \eta_\varepsilon(x-y)|u(x) \\ -u(y)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y.$$

The second term on the right-hand side satisfies:

$$\begin{aligned}
\frac{1}{\varepsilon}\int_D \int_{\{x\in D\,:\,|x-y|>\alpha\varepsilon\}} & \eta_\varepsilon(x-y)|u(x)-u(y)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{\varepsilon}\int_D \int_{\{x\in D\,:\,|x-y|>\alpha\varepsilon\}} \eta_\varepsilon(x-y)|\hat{u}(x) \\
&\quad - \hat{u}(y)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\
&\leqq ||\rho||^2_{L^\infty(D)} \int_{|h|>\alpha} \eta(h)|h| \int_{\mathbb{R}^d} \frac{|\hat{u}(y)-\hat{u}(y+\varepsilon h)|}{\varepsilon|h|}\,\mathrm{d}y\,\mathrm{d}h \\
&\leqq ||\rho||^2_{L^\infty(D)} |D\hat{u}|(\mathbb{R}^d) \int_{|h|>\alpha} \eta(h)|h|\,\mathrm{d}h,
\end{aligned}$$

where the first inequality is obtained using the change of variables $h = \frac{x-y}{\varepsilon}$ and the second inequality obtained using Lemma 13.33 in [39]. By Step 1 we conclude that:

$$\begin{aligned}
\limsup_{\varepsilon\to\infty} TV_\varepsilon(u;\rho) &\leqq \limsup_{\varepsilon\to\infty} TV_\varepsilon^\alpha(u;\rho) + ||\rho||^2_{L^\infty(\mathbb{R}^d)} |D\hat{u}|(\mathbb{R}^d) \int_{|h|>\alpha} \eta(h)|h|\,\mathrm{d}h \\
&\leqq \sigma_{\eta^\alpha} TV(u;\rho^2) + ||\rho||^2_{L^\infty(\mathbb{R}^d)} |D\hat{u}|(\mathbb{R}^d) \int_{|h|>\alpha} \eta(h)|h|\,\mathrm{d}h.
\end{aligned}$$

Taking α to infinity and using condition (K3) on $\boldsymbol{\eta}$ implies (4.15).
Case 2: ρ is continuous but not necessarily Lipschitz. The idea is to approximate ρ from above by a family of Lipschitz functions $\{\rho_k\}_{k\in\mathbb{N}}$. Consider $\rho_k : D \to \mathbb{R}$ given by

$$\rho_k(x) := \sup_{y\in D} \rho(y) - k|x-y|. \tag{4.19}$$

The functions ρ_k are Lipschitz functions which are bounded from below from and above by the same constants bounding ρ from below and from above. Moreover, given that ρ is continuous, it is simple to verify that for every $x \in D$, $\rho_k(x) \searrow \rho(x)$ as $k \to \infty$.

As in Step 1, it is enough to consider $u \in BV(D)$ and prove that:

$$\limsup_{\varepsilon\to 0} TV_\varepsilon(u;\rho) \leqq \sigma_\eta TV(u;\rho^2).$$

The proof of the limsup inequality in Case 1 and the fact that $\rho \leqq \rho_k$ imply that

$$\limsup_{\varepsilon\to 0} TV_\varepsilon(u;\rho) \leqq \limsup_{\varepsilon\to 0} TV_\varepsilon(u;\rho_k) \leqq \sigma_\eta TV(u;\rho_k^2). \tag{4.20}$$

By the dominated convergence theorem,

$$\lim_{k\to\infty} TV(u;\rho_k^2) = \lim_{k\to\infty}\int_D \rho_k^2(x)\,\mathrm{d}|Du|(x) = \int_D \rho^2(x)\,\mathrm{d}|Du|(x) = TV(u;\rho^2).$$

Combining with (4.20) provides the desired result. □

Remark 4.3. Note that using the liminf inequality and the proof of the limsup inequality we deduce the pointwise convergence of the functionals $TV_\varepsilon(\cdot;\rho)$; namely, for every $u \in L^1(D,\rho)$:

$$\lim_{\varepsilon\to 0} TV_\varepsilon(u;\rho) = \sigma_\eta TV(u;\rho^2).$$

4.3. Proof of Theorem 4.1: Compactness

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

Lemma 4.4. *Let D be a bounded, open, and connected set in $\mathbb{R}^d$, with C^2-boundary. Let $\{v_\varepsilon\}_{\varepsilon>0}$ be a sequence in $L^1(D,\rho)$ such that:*

$$\sup_{\varepsilon>0} \|v_\varepsilon\|_{L^1(D,\rho)} < \infty,$$

and

$$\sup_{\varepsilon>0} TV_\varepsilon(v_\varepsilon;\rho) < \infty. \tag{4.21}$$

Then, $\{v_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $L^1(D,\rho)$.

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, (4.21) holds when changing η for $\tilde{\eta}$ and so there is no loss of generality in assuming that $\boldsymbol{\eta}$ has the form of $\tilde{\boldsymbol{\eta}}$. Also, since ρ is bounded below and above by positive constants, it is enough to consider $\rho \equiv 1$.

We first extend each function v_ε to $\mathbb{R}^d$ in a suitable way. Since ∂D is a compact C^2 manifold, there exists $\delta > 0$ such that for every $x \in \mathbb{R}^d$ for which $d(x,\partial D) \leqq \delta$ 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 $s \leqq \delta/8$ and $\xi(s) = 0$ if $s \geqq \delta/4$. We define an auxiliary function $\hat{v}_\varepsilon$ on U, by $\hat{v}_\varepsilon(x) := v_\varepsilon(\hat{x})$ and the desired extended function $\tilde{v}_\varepsilon$ on $\mathbb{R}^d$ by $\tilde{v}_\varepsilon(x) = \xi(|x - Px|)v_\varepsilon(\hat{x})$.

We claim that:

$$\sup_{\varepsilon>0} \frac{1}{\varepsilon}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \eta_\varepsilon(x-y)|\tilde{v}_\varepsilon(x) - \tilde{v}_\varepsilon(y)| < \infty. \tag{4.22}$$

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 $y \in D$

$$|\hat{x} - y| < 2|x - y|. \tag{4.23}$$

Since the mapping $x \mapsto \hat{x}$ 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 $x, y \in W$

$$\frac{1}{4}|x - y| < |\hat{x} - \hat{y}| < 4|x - y|. \tag{4.24}$$

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

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

Squaring and straightforward algebra yield

$$|z - Px|^2 \geqq 2\delta(z - Px) \cdot \frac{x - Px}{|x - Px|}. \tag{4.25}$$

For $x \in W$ and $y \in D$, using (4.25) we obtain

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

Therefore $|y - \hat{x}|^2 \leqq 3|y - x|^2$, which establishes (4.23).

For distinct $x, y \in W$ using (4.25), with $z = Py$ and with $z = Px$, follows

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

Therefore

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

Since the roles on x, y and $\hat{x}, \hat{y}$ can be reversed it follows that $|x - y| \leqq 4|\hat{x} - \hat{y}|$. These estimates establish (4.24).

We now return to proving (4.22). For ε small enough,

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_{\mathbb{R}^n\setminus D}\int_D \eta_\varepsilon(x-y)|\tilde{v}_\varepsilon(x)-\tilde{v}_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y\\
&\quad=\frac{1}{\varepsilon}\int_V\int_D \eta_\varepsilon(x-y)|\hat{v}_\varepsilon(x)-\hat{v}_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y\\
&\quad=\frac{1}{\varepsilon}\int_V\int_D \eta_\varepsilon(x-y)|v_\varepsilon(\hat{x})-v_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq\frac{4^d}{\varepsilon}\int_V\int_D \eta_{4\varepsilon}(\hat{x}-y)|v_\varepsilon(x)-v_\varepsilon(\hat{y})|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq\frac{16^d}{\varepsilon}\int_D\int_D \eta_{4\varepsilon}(z-y)|v_\varepsilon(x)-v_\varepsilon(z)|\,\mathrm{d}z\mathrm{d}y,
\end{aligned}$$

where the first inequality follows from (4.23) and the second follows from the fact that the change of variables $x\mapsto\hat{x}$ is bi-Lipschitz as shown in (4.24). Also,

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_{\mathbb{R}^d\setminus D}\int_{\mathbb{R}^d\setminus D} \eta_\varepsilon(x-y)|\tilde{v}_\varepsilon(x)-\tilde{v}_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y\\
&\quad=\frac{1}{\varepsilon}\int_W\int_W \eta_\varepsilon(x-y)|\xi(x)\hat{v}_\varepsilon(x)-\xi(y)\hat{v}_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq\frac{1}{\varepsilon}\int_W\int_W \eta_\varepsilon(x-y)|\xi(x)-\xi(y)||\hat{v}_\varepsilon(x)|\,\mathrm{d}x\mathrm{d}y\\
&\qquad+\frac{1}{\varepsilon}\int_W\int_W \eta_\varepsilon(x-y)|\hat{v}_\varepsilon(x)-\hat{v}_\varepsilon(y)||\xi(y)|\,\mathrm{d}x\mathrm{d}y.
\end{aligned}$$

Note that for all $x\neq y$, $\frac{\eta_\varepsilon(x-y)}{\varepsilon}\leqq\frac{b}{|x-y|}\eta_\varepsilon(x-y)$. Therefore:

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_W\int_W \eta_\varepsilon(x-y)|\xi(x)-\xi(y)||\hat{v}_\varepsilon(x)|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq b\int_W\int_W \eta_\varepsilon(x-y)\frac{|\xi(x)-\xi(y)|}{|x-y|}|\hat{v}_\varepsilon(x)|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq b\,\mathrm{Lip}(\xi)\int_W\int_W \eta_\varepsilon(x-y)|\hat{v}_\varepsilon(x)|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq 4^d\,b\,\mathrm{Lip}(\xi)\|v_\varepsilon\|_{L^1(D)},
\end{aligned}$$

where we used (4.24) and change of variables to establish the last inequality. Also,

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_W\int_W \eta_\varepsilon(x-y)|\hat{v}_\varepsilon(x)-\hat{v}_\varepsilon(y)||\xi(y)|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq\frac{4^d}{\varepsilon}\int_W\int_W \eta_{4\varepsilon}(\hat{x}-\hat{y})|\hat{v}_\varepsilon(x)-\hat{v}_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y\\
&\quad\leqq\frac{4^{3d}}{\varepsilon}\int_D\int_D \eta_{4\varepsilon}(x-y)|v_\varepsilon(x)-v_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y.
\end{aligned}$$

The first inequality is obtained thanks to the fact that $|\xi(y)|\leqq 1$ and (4.24), while the second inequality is obtained by a change of variables.

Using that

$$\int_D \int_D \eta_{4\varepsilon}(x-y)|v_\varepsilon(x)-v_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y \leqq 4^d \int_D \int_D \eta_\varepsilon(x-y)|v_\varepsilon(x)-v_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y$$

by combining the above inequalities we conclude that

$$\sup_{\varepsilon>0} \frac{1}{\varepsilon} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \eta_\varepsilon(x-y)|\tilde{v}_\varepsilon(x)-\tilde{v}_\varepsilon(y)|\,\mathrm{d}x\,\mathrm{d}y$$
$$\leqq C \sup_{\varepsilon>0} \left(\int_D \int_D \eta_\varepsilon(x-y)|v_\varepsilon(x)-v_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y + \|v_\varepsilon\|_{L^1(D)} \right) < \infty.$$

Using the proof of Proposition 3.1 in [3] we deduce that the sequence $\{\tilde{v}_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $L^1(\mathbb{R}^d)$ which implies that the sequence $\{v_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $L^1(D)$. □

Remark 4.5. 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 4.6. *Let D be a bounded, open, and connected set in $\mathbb{R}^d$, with Lipschitz boundary. Suppose that the sequence of functions $\{u_\varepsilon\}_{\varepsilon>0} \subseteq L^1(D,\rho)$ satisfies:*

$$\sup_{\varepsilon>0} \|u_\varepsilon\|_{L^1(D,\rho)} < \infty,$$
$$\sup_{\varepsilon>0} TV_\varepsilon(u_\varepsilon;\rho) < \infty.$$

Then, $\{u_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $L^1(D,\rho)$.

Proof. Suppose $\{u_\varepsilon\}_{\varepsilon>0} \subseteq L^1(D)$ is as in the statement. As in Lemma 4.4, we can assume that $\rho \equiv 1$. By Remark 5.3 in [10], there exists a bi-Lipschitz map $\Theta : \tilde{D} \to D$ where $\tilde{D}$ is a domain with smooth boundary. For every $\varepsilon > 0$ consider the function $v_\varepsilon := u_\varepsilon \circ \Theta$ and set $\hat{\boldsymbol{\eta}}(s) := \boldsymbol{\eta}\,(\mathrm{Lip}(\Theta)\, s)$, $s \in \mathbb{R}$.

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:

$$\int_{\tilde{D}} |v_\varepsilon(x)|\,\mathrm{d}x \leqq C \int_D |u_\varepsilon(y)|\mathrm{d}y,$$

and

$$C \int_D \int_D \eta_\varepsilon(x-y)\,|u_\varepsilon(x)-u_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y$$
$$\geqq \int_{\tilde{D}} \int_{\tilde{D}} \eta_\varepsilon\,(\Theta(x)-\Theta(y))\,|v_\varepsilon(x)-v_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y$$
$$\geqq \int_{\tilde{D}} \int_{\tilde{D}} \hat{\eta}_\varepsilon(x-y)\,|v_\varepsilon(x)-v_\varepsilon(y)|\,\mathrm{d}x\mathrm{d}y.$$

The second inequality using the fact that η is non-increasing (assumption (K2)). We conclude that the sequence $\{v_\varepsilon\}_{\varepsilon>0} \subseteq L^1(\tilde{D})$ satisfies the assumptions of Lemma 4.4 (taking $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}$). Therefore, $\{v_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $L^1(\tilde{D})$, which implies that $\{u_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $L^1(D)$. □

Corollary 4.7. *Let D be a bounded, open, and connected set in $\mathbb{R}^d$. Suppose that the sequence of functions $\{u_\varepsilon\}_{\varepsilon>0} \subseteq L^1(D,\rho)$ satisfies:*

$$\sup_{\varepsilon>0} \|u_\varepsilon\|_{L^1(D,\rho)} < \infty,$$

$$\sup_{\varepsilon>0} TV_\varepsilon(u_\varepsilon;\rho) < \infty.$$

Then, $\{u_\varepsilon\}_{\varepsilon>0}$ is locally relatively compact in $L^1(D,\rho)$.

In particular if

$$\sup_{\varepsilon>0} \|u_\varepsilon\|_{L^\infty(D)} < \infty,$$

then, $\{u_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $L^1(D,\rho)$.

Proof. If B is a ball compactly contained in D then the relative compactness of $\{u_\varepsilon\}_{\varepsilon>0}$ in $L^1(B,\rho)$ follows from Lemma 4.4. 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 $L^1(D,\rho)$ 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 the fact that $\lim_{k\to\infty} \sup_{\varepsilon>0} \|u_\varepsilon\|_{L^1(D\setminus D_k,\rho)} = 0$. □

5. Γ-Convergence of Total Variation on Graphs

5.1. Proof of Theorems 1.1 and 1.2

Let $D \subset \mathbb{R}^d$, $d \geqq 2$ be an open, bounded and connected set with Lipschitz boundary. Assume ν is a probability measure on D with continuous density ρ, which is bounded from below and above by positive constants. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 satisfying assumption (1.8).

Proof (Proof of Theorem 1.1). We use the sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ considered in Section 2.3. Let $\omega \in \Omega$ be such that (2.16) and (2.17) hold in cases $d = 2$ and $d \geqq 3$ respectively. By Theorem 2.5 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. The key idea in the proof is that the estimates of the Section 2.3 on transportation maps imply that the transportation happens on a length scale which is small compared to ε_n. By taking a kernel with slightly smaller 'radius' than ε_n

we can then obtain a lower bound, and by taking a slightly larger radius a matching upper bound on the graph total variation.

Liminf inequality: Assume that $u_n \xrightarrow{TL^1} u$ as $n \to \infty$. Since $T_{n\sharp}\nu = \nu_n$, using the change of variables (2.8) it follows that

$$GTV_{n,\varepsilon_n}(u_n) = \frac{1}{\varepsilon_n}\int_{D\times D} \eta_{\varepsilon_n}\left(T_n(x) - T_n(y)\right)\left|u_n \circ T_n(x) - u_n \circ T_n(y)\right| \times \rho(x)\rho(y)\,\mathrm{d}x\mathrm{d}y. \tag{5.1}$$

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

$$|T_n(x) - T_n(y)| > b\varepsilon_n \Rightarrow |x - y| > b\varepsilon_n - 2\|Id - T_n\|_\infty. \tag{5.2}$$

Thanks to the assumptions on $\{\varepsilon_n\}_{n\in\mathbb{N}}$ ((2.17) and (2.17) in cases $d = 2$ and $d \geqq 3$ respectively), for large enough $n \in \mathbb{N}$:

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

By (5.2), for large enough n and for almost every $(x, y) \in D \times D$,

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

Let $\tilde{u}_n = u_n \circ T_n$. Thanks to the previous inequality and (5.1), for large enough n

$$\begin{aligned} GTV_{n,\varepsilon_n}(u_n) &\geqq \frac{1}{\varepsilon_n^{d+1}}\int_{D\times D} \boldsymbol{\eta}\left(\frac{|x-y|}{\tilde{\varepsilon}_n}\right)|\tilde{u}_n(x) - \tilde{u}_n(y)|\,\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{d+1} TV_{\tilde{\varepsilon}_n}(\tilde{u}_n;\rho). \end{aligned}$$

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 $\liminf_{n\to\infty} TV_{\tilde{\varepsilon}_n}(\tilde{u}_n;\rho) \geqq \sigma_\eta TV(u;\rho^2)$ and hence:

$$\liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) \geqq \sigma_\eta TV(u;\rho^2).$$

Limsup inequality: By Remark 2.7 and Proposition 2.4, it is enough to prove the limsup inequality for Lipschitz continuous functions $u : D \to \mathbb{R}$. Define u_n to be the restriction of u to the first n data points $X_1, \ldots, X_n$. 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 $(x, y) \in D \times D$

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

Then for all n

$$\begin{aligned} &\frac{1}{\tilde{\varepsilon}_n^{d+1}}\int_{D\times D} \boldsymbol{\eta}\left(\frac{|T_n(x) - T_n(y)|}{\varepsilon_n}\right)|\tilde{u}_n(x) - \tilde{u}_n(y)|\,\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\ &\quad\leqq \frac{1}{\tilde{\varepsilon}_n}\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)\,|\tilde{u}_n(x) - \tilde{u}_n(y)|\,\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned} \tag{5.3}$$

Also

$$\frac{1}{\tilde{\varepsilon}_n}\left|\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)(|u(x)-u(y)|-|u\circ T_n(x)-u\circ T_n(y)|)\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y\right|$$
$$\leqq \frac{2}{\tilde{\varepsilon}_n}\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)|u(x)-u\circ T_n(x)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y$$
$$\leqq \frac{2C\,\mathrm{Lip}(u)\|\rho\|^2_{L^\infty(D)}}{\tilde{\varepsilon}_n}\int_D |x-T_n(x)|\,\mathrm{d}x, \tag{5.4}$$

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

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

Since $\frac{\varepsilon_n}{\tilde{\varepsilon}_n}\to 1$ as $n\to\infty$, using (5.3) we deduce :

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) = \limsup_{n\to\infty}\frac{1}{\tilde{\varepsilon}_n^{d+1}}\int_{D\times D}\boldsymbol{\eta}\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon_n}\right)|u\circ T_n(x)$$
$$-u\circ T_n(y)|\,\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y$$
$$\leqq \limsup_{n\to\infty}\frac{1}{\tilde{\varepsilon}_n}\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)\,|u\circ T_n(x)$$
$$-u\circ T_n(y)|\,\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \limsup_{n\to\infty} TV_{\tilde{\varepsilon}_n}(u;\rho) \leqq \sigma_\eta TV(u;\rho^2),$$

where the last inequality follows from the proof of Theorem 4.1, specifically inequality (4.15).

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

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

$$\liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) = \liminf_{n\to\infty}\sum_{k=1}^{l} GTV^k_{n,\varepsilon_n}(u_n)$$
$$\geqq \sum_{k=1}^{l}\liminf_{n\to\infty} GTV^k_{n,\varepsilon_n}(u_n) \geqq \sum_{k=1}^{l}\sigma_{\eta_k}TV(u;\rho^2)$$
$$= \sigma_\eta TV(u;\rho^2).$$

Limsup inequality: By Remark 2.7 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

$$\begin{aligned}\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) &= \limsup_{n\to\infty} \sum_{k=1}^{l} GTV^k_{n,\varepsilon_n}(u_n) \\ &\leqq \sum_{k=1}^{l} \limsup_{n\to\infty} GTV^k_{n,\varepsilon_n}(u_n) \leqq \sum_{k=1}^{l} \sigma_{\eta_k} TV(u;\rho^2) \\ &= \sigma_\eta TV(u;\rho^2).\end{aligned}$$

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

Liminf Inequality: Note that there exists 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 $k \to \infty$ almost everywhere Denote by GTV^k_{n,ε_n} the graph TV corresponding to $\boldsymbol{\eta}_k$. If $u_n \xrightarrow{TL^1} u$ as $n \to \infty$, by Step 2 $\sigma_{\eta_k} TV(u;\rho^2) \leqq \liminf_{n\to\infty} GTV^k_{n,\varepsilon_n}(u_n) \leqq \liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n)$ for every $k \in \mathbb{N}$. The monotone convergence theorem implies that $\lim_{k\to\infty} \sigma_{\eta_k} = \sigma_\eta$ and thus $\sigma_\eta TV(u;\rho^2) \leqq \liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n)$.

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 $k \to \infty$ almost everywhere. Proceeding in an analogous way to the way we proceeded in the proof of the liminf inequality we can conclude that $\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) \leqq \sigma_\eta TV(u;\rho^2)$.

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 $t \leqq \alpha$ and $\boldsymbol{\eta}_\alpha(t) = 0$ for $t > \alpha$. We denote by $GTV^\alpha_{n,\varepsilon_n}$ the graph TV using $\boldsymbol{\eta}_\alpha$. Then

$$\begin{aligned}GTV_{n,\varepsilon_n}(u_n) = GTV^\alpha_{n,\varepsilon_n}(u_n) + \frac{1}{\varepsilon_n^{d+1}} \int_{|T_n(x)-T_n(y)|>\alpha\varepsilon_n} \boldsymbol{\eta}\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon_n}\right) \\ |u\circ T_n(x) - u\circ T_n(y)|\,\rho(x)\rho(y)\,dx\,dy.\end{aligned} \tag{5.5}$$

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 $|x-y| \leqq |T_n(x)-T_n(y)| + 2\|Id - T_n\|_\infty$ and $|T_n(x)-T_n(y)| \leqq |x-y| + 2\|Id - T_n\|_\infty$ 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 $|x-y| \leqq 2|T_n(x)-T_n(y)|$ and $|T_n(x)-T_n(y)| \leqq 2|x-y|$. We conclude that for large enough n

$$\begin{aligned}
\frac{1}{\varepsilon_n^{d+1}} \int_{|T_n(x)-T_n(y)|>\alpha\varepsilon_n} & \boldsymbol{\eta}\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon_n}\right) |u \circ T_n(x) - u \circ T_n(y)|\, \rho(x)\rho(y)\, \mathrm{d}x\, \mathrm{d}y \\
&\leqq \frac{||\rho||^2_{L^\infty(D)}}{\varepsilon_n^{d+1}} \int_{|x-y|>\alpha\varepsilon_n/2} \boldsymbol{\eta}\left(\frac{|x-y|}{2\varepsilon_n}\right) \\
&\quad \times |u \circ T_n(x) - u \circ T_n(y)|\, \mathrm{d}x\, \mathrm{d}y \\
&\leqq \frac{2\,\mathrm{Lip}(u)||\rho||^2_{L^\infty(D)}}{\varepsilon_n^{d+1}} \int_{|x-y|>\alpha\varepsilon_n/2} \boldsymbol{\eta}\left(\frac{|x-y|}{2\varepsilon_n}\right) |x-y|\, \mathrm{d}x\mathrm{d}y.
\end{aligned}$$

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:

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

where C does not depend on n or α. The previous inequalities, (5.5) and Step 3 imply that

$$\begin{aligned}
\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) &\leqq \limsup_{n\to\infty} GTV^{\alpha}_{n,\varepsilon_n}(u_n) + \mathrm{Lip}(u)||\rho||^2_{L^\infty(D)} C \int_{|h|>\frac{\alpha}{4}} \eta(h)|h|\, \mathrm{d}h \\
&\leqq \sigma_{\eta_\alpha} TV(u;\rho^2) + \mathrm{Lip}(u)||\rho||^2_{L^\infty(D)} C \int_{|h|>\frac{\alpha}{4}} \eta(h)|h|\, \mathrm{d}h.
\end{aligned}$$

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

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) \leqq \sigma_\eta TV(u;\rho^2).$$

□

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 4.4 and Proposition 4.6 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 $t \geqq b$, for some a and b positive constants.

Consider the sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from Section 2.3. Since $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfies (1.8), estimates (2.17) and (2.17) imply that for Lebesgue almost everywhere $z, y \in D$ with $|T_n(z) - T_n(y)| > b\varepsilon_n$ it holds 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 almost everywhere $z, y \in D$:

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

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

$$\begin{aligned}\frac{1}{\varepsilon_n^{d+1}}\int_D\int_D \boldsymbol{\eta}\left(\frac{|z-y|}{\tilde{\varepsilon}_n}\right)|u_n\circ T_n(z)-u_n\circ T_n(y)|\,\rho(z)\rho(y)\,\mathrm{d}z\,\mathrm{d}y \\ \leqq \frac{1}{\varepsilon_n^{d+1}}\int_D\int_D \boldsymbol{\eta}\left(\frac{|T_n(z)-T_n(y)|}{\tilde{\varepsilon}_n}\right)|u_n\circ T_n(z)-u_n\circ T_n(y)| \\ \times\,\rho(z)\rho(y)\,\mathrm{d}z\,\mathrm{d}y \\ = GTV_{n,\varepsilon_n}(u_n).\end{aligned}$$

Thus

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

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

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

By Proposition 4.6 we conclude that $\{u_n\circ T_n\}_{n\in\mathbb{N}}$ is relatively compact in $L^1(D)$ and hence $\{u_n\}_{n\in\mathbb{N}}$ is relatively compact in TL^1. □

We now prove Corollary 1.3 on the Γ convergence of perimeter.

Proof. 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 [24] and take advantage of the coarea formula of the energies GTV_{n,ε_n}. 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 $u_n\in L^1(D,\nu_n)$) such that $\limsup_{n\to\infty}GTV_{n,\varepsilon_n}(u_n)\leqq\sigma_\eta TV(\chi_A,\rho^2)$. It is straightforward to verify that the functionals GTV_{n,ε_n} satisfy the coarea formula:

$$GTV_{n,\varepsilon_n}(u_n)=\int_{-\infty}^{\infty}GTV_{n,\varepsilon_n}(\chi_{\{u_n>s\}})\,\mathrm{d}s.$$

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

$$\int_\delta^{1-\delta}GTV_{n,\varepsilon_n}(\chi_{\{u_n>s\}})ds\leqq GTV_{n,\varepsilon_n}(u_n).$$

For every n there is $s_n\in(\delta,1-\delta)$ such that $GTV_{n,\varepsilon_n}(\chi_{\{u_n>s_n\}})\leqq\frac{1}{1-2\delta}GTV_{n,\varepsilon_n}(u_n)$. Let $A_n^\delta:=\{u_n>s_n\}$. It is direct to show that $\chi_{A_n^\delta}\xrightarrow{TL^1}\chi_A$ as $n\to\infty$ and that $\limsup_{n\to\infty}GTV_{n,\varepsilon_n}(A_n^\delta)\leqq\frac{1}{1-2\delta}\sigma_\eta TV(\chi_A;\rho^2)$. Taking $\delta\to 0$ and using a diagonal argument provides sets $\{A_n\}_{n\in\mathbb{N}}$ such that $\chi_{A_n}\xrightarrow{TL^1}\chi_A$ as $n\to\infty$ and $\limsup_{n\to\infty}GTV_{n,\varepsilon_n}(\chi_{A_n})\leqq\sigma_\eta TV(\chi_A,\rho^2)$. □

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 approximate χ_A in TL^1 topology by characteristic functions of sets of the form $G = E \cap D$ where E is a subset of $\mathbb{R}^d$ with a smooth boundary. As in the proof of Theorem 1.1, the key is to show that for step kernels ($\boldsymbol{\eta}(r) = b$ if $r < a$ and zero otherwise)

$$\lim_{n\to\infty} GTV_{n,\varepsilon_n}(\chi_G) = TV(\chi_G, \rho^2).$$

To do so one needs a substitute for estimate (5.4). The needed estimate follows from the following estimate: for all G as above, there exists δ_0 such that for all n for which $||Id - T_n||_\infty \leqq \delta_0$,

$$\int_D |\chi_G(x) - \chi_G(T_n(x))|\, \mathrm{d}x \leqq 4\,\mathrm{Per}(E)\,||Id - T_n||_\infty.$$

This estimate follows from the fact that if $\chi_G(x) \neq \chi_G(T_n(x))$ then $d(x, \partial E) \leqq |x - T_n(x)|$ and the fact that, for δ small enough, $|\{x \in \mathbb{R}^d \;:\; d(x, \partial E) < \delta\}| \leqq 4\,\mathrm{Per}(E)\delta$, which follows form Weyl's formula [62] for the volume of the tubular neighborhood. Noting that the perimeter of any set can be approximated by smooth sets (see Remark 3.42 in [4]) and using Remark 2.7 we obtain the limsup inequality for the characteristic function of any measurable set.

We remark that if one restricts the functional to the class of sets with specified volume (as in Example 1.4) then each set in the class can be approximated by smooth sets satisfying the volume constraint. This follows by a careful modification to the density argument of Remark 3.43 in [4].

5.2. Extension to Different Sets of Points

Consider the setting of Theorem 1.1. The only information about the points X_i that the proof requires is the upper bound on the ∞-transportation distance between ν and the empirical measure ν_n. Theorem 2.5 provides such bounds when X_i are i.i.d. distributed according to ν. Such randomness assumption is reasonable when modeling randomly obtained data points, but in other settings points may be more regularly distributed and/or given deterministically. In such setting, if one is able to obtain tighter bounds on transportation distance this would translate into better bounds on $\varepsilon(n)$ in Theorem 1.1 for which the Γ-convergence holds.

That is, if $X_1, \dots, X_n, \dots$ are the given points, let ν_n still be $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. If one can find transportation maps T_n from ν to ν_n such that

$$\limsup_{n\to\infty} \frac{n^{1/d}\|Id - T_n\|_\infty}{f(n)} \leqq C \tag{5.6}$$

for some nonnegative function $f : \mathbb{N} \to (0, \infty)$ then Theorem 1.1 would hold if

$$\lim_{n\to\infty} \frac{f(n)}{n^{1/d}} \frac{1}{\varepsilon_n} = 0.$$

We remark that f must be bounded from below, since for any collection $V = \{X_1, \ldots, X_n\}$ in D, $\sup_{y \in D} \operatorname{dist}(y, V) \geqq cn^{-1/d}$ and thus $n^{1/d}\|Id - T_n\|_\infty \geqq c$.

One special case is when $D = (0,1)^d$, ν is the Lebesgue measure and $X_1, \ldots, X_n, \ldots$ is a sequence of grid points on diadicaly refining grids. In this case, (5.6) holds with $f(n) = 1$ for all n and thus Γ-convergence holds for $\varepsilon_n \to 0$ 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 [17,24] 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 3.12.

Acknowledgments. The authors are grateful to Thomas Laurent for many valuable discussions and careful reading of an early version of the manuscript. They are thankful to Giovanni Leoni for valuable advice and pointing out the paper of PONCE [46]. The authors are grateful to Michel Talagrand for letting them know of the elegant proofs of matching results in [55] and generously sharing the chapters of his upcoming book [56]. The authors are thankful to Bob Pego for valuable advice and to ANTONIN CHAMBOLLE, ALAN FRIEZE, JAMES NOLEN, and FELIX OTTO for enlightening discussions. DS is grateful to NSF (Grants DMS-1211760 and DMS-1516677). 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. Proof of Proposition 2.4

Proof. Using the fact that D has Lipschitz boundary and the fact that ψ is bounded above and below by positive constants, Theorem 10.29 in [39] implies that for any $u \in C^\infty(D) \cap BV(D)$ there exists a sequence $\{u_n\}_{n\in\mathbb{N}} \subseteq C_c^\infty(\mathbb{R}^d)$ with $u_n \to_{L^1(D)} u$ and with $\int_D |\nabla u - \nabla u_n| \psi(x)\, dx \to 0$ as $n \to \infty$. Using a diagonal argument we conclude that in order to prove Proposition 2.4 it is enough to prove that for every $u \in BV(D)$ there exists a sequence $\{u_n\}_{n\in\mathbb{N}} \subseteq C^\infty(D) \cap BV(D)$ with $u_n \to_{L^1(D)} u$ and with $\int_D |\nabla u_n| \psi(x)\, dx \to TV(u;\psi)$ as $n \to \infty$.

Step 1: If ψ is Lipschitz this is precisely the content of Theorem 3.4 in [9].

Step 2: If ψ is not necessarily Lispchitz we can find a sequence $\{\psi_k\}_{k\in\mathbb{N}}$ of Lipschitz functions bounded above and below by the same constants bounding ψ and with $\psi_k \searrow \psi$. The functions ψ_k can be defined as in (4.19) (replacing ρ with ψ).

Using Step 1, for a given $u \in BV(D)$ and for every $k \in \mathbb{N}$ we can find a sequence $\{u_{n,k}\}_{n\in\mathbb{N}}$ with $u_{n,k} \to_{L^1(D)} u$ and with $\int_D |\nabla u_{n,k}| \psi_k(x)\, dx \to TV(u;\psi_k)$ as $n \to \infty$. By 2.3 and by the dominated convergence theorem we know that $TV(u;\psi_k) = \int_D \psi_k(x)|Du|(x) \to \int_D \psi(x)|Du|(x) = TV(u;\psi)$ as $k \to \infty$. Therefore, a diagonal argument allows us to conclude that there exists a sequence $\{k_n\}_{n\in\mathbb{N}}$ with the property that, $u_{n,k_n} \to_{L^1(D)} u$ and $\int_D |\nabla u_n| \psi_{k_n}(x)\, dx \to TV(u;\psi)$ as $n \to \infty$. Taking $u_n := u_{n,k_n}$ and using the fact that that $\psi \leqq \psi_{k_n}$ we obtain:

$$\limsup_{n\to\infty} \int_D |\nabla u_n(x)| \psi(x)\, dx \leqq \lim_{n\to\infty} \int_D |\nabla u_n(x)| \psi_{k_n}(x)\, dx = TV(u;\psi).$$

Since $u_n \to_{L^1(D)} u$, the lower semicontinuity of $TV(\cdot, \psi)$ implies that $\liminf_{n\to\infty} \int_D |\nabla u_n(x)|\psi(x)\,\mathrm{d}x \geqq TV(u;\psi)$. The desired result follows. □

References

1. Agueh, M.: Finsler structure in the p-Wasserstein space and gradient flows. *C. R. Math. Acad. Sci. Paris.* **350**, 35–40 (2012)
2. Ajtai, M., Komlós, J., Tusnády, G.: On optimal matchings. *Combinatorica.* **4**, 259–264 (1984)
3. Alberti, G., Bellettini, G.: A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies. *European J. Appl. Math.* **9**, 261–284 (1998)
4. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs*. The Clarendon Press Oxford University Press, New York (2000)
5. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows: In: Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics, Birkhäuser Basel (2008)
6. Andreev, K., Racke, H.: Balanced graph partitioning. *Theory Comput. Syst.* **39**, 929–939 (2006)
7. Arias-Castro, E., Pelletier, B., Pudlo, P.: The normalized graph cut and Cheeger constant: from discrete to continuous. *Adv. Appl. Probab.* **44**, 907–937 (2012)
8. Arora, S., Rao, S., Vazirani, U.: Expander flows, geometric embeddings and graph partitioning. *J. ACM (JACM).* **56**, 5 (2009)
9. Baldi, A.: Weighted BV functions. Houston J. Math. **27**, 683–705 (2001)
10. Ball, J.M., Zarnescu, A.: Partial regularity and smooth topology-preserving approximations of rough domains. arXiv:1312.5156 (2013) (arXiv preprint)
11. Belkin, M., Niyogi, P.: Convergence of Laplacian eigenmaps. *Adv. Neural Inf. Process. Syst. (NIPS).* **19**, 129 (2007)
12. Belkin, M., Niyogi, P.: Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. Syst. Sci.* **74**, 1289–1308 (2008)
13. Bertozzi, A.L., Flenner, A.: Diffuse interface models on graphs for classification of high dimensional data. *Multiscale Model. Simul.* **10**, 1090–1118 (2012)
14. Bourgain, J., Brezis, H., Mironescu, P., et al.: Another look at sobolev spaces. In: *Optimal control and partial differential equations, A volume in honour of A. Benssoussans 60th birthday*, p. 439455. IOS Press, Amsterdam (2001)
15. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *Pattern Anal. Mach. Intell. IEEE Trans.* **23**, 1222–1239 (2001)
16. Braides, A.: *Gamma-Convergence for Beginners. Oxford Lecture Series in Mathematics and Its Applications Series*. Oxford University Press, Incorporated, Oxford (2002)
17. Braides, A., Yip, N.K.: A quantitative description of mesh dependence for the discretization of singularly perturbed nonconvex problems. *SIAM J. Numer. Anal.* **50**, 1883–1898 (2012)
18. Bresson, X., Laurent, T.: Asymmetric cheeger cut and application to multi-class unsupervised clustering. *CAM Report*, pp. 1–8 (2012)
19. Bresson, X., Laurent, T., Uminsky, D., von Brecht, J.: Multiclass total variation clustering. In: Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. (eds.) *Advances in Neural Information Processing Systems*, vol. 26, pp. 1421–1429 (2013)
20. Bresson, X., Laurent, T., Uminsky, D., von Brecht, J. H.: Convergence and energy landscape for cheeger cut clustering. In: P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, (eds.) *Advances in Neural Information Processing Systems (NIPS)*, pp. 1394–1402 (2012)

21. Bresson, X., Laurent, T., Uminsky, D., von Brecht, J. H.: An adaptive total variation algorithm for computing the balanced cut of a graph. arXiv preprint arXiv:1302.2717. (2013)
22. Bresson, X., Tai, X.-C., Chan, T. F., Szlam, A.: Multi-class transductive learning based on l1 relaxations of cheeger cut and mumford-shah-potts model. *UCLA CAM Rep.* 12–03 (2012)
23. Castaing, C., Raynaud de Fitte, P., Valadier, M.: Young measures on topological spaces of Mathematics and its Applications, vol. 571, Kluwer Academic Publishers, Dordrecht, 2004
24. Chambolle, A., Giacomini, A., Lussardi, L.: Continuous limits of discrete perimeters. *M2AN Math. Model. Numer. Anal.* **44**, 207–230 (2010)
25. Dal Maso, G.: *An Introduction to Γ-convergence*. Springer, New York, 1993
26. Delling, D., Fleischman, D., Goldberg, A., Razenshteyn, I., Werneck, R.: An exact combinatorial algorithm for minimum graph bisection. *Math. Program.* **152**(2), 417–458 (2014)
27. Esedoḡlu, S., Otto, F.: Threshold dynamics for networks with arbitrary surface tensions. *Commun. Pure Appl. Math.* **68**, 808–864 (2015)
28. Feige, U., Krauthgamer, R.: A polylogarithmic approximation of the minimum bisection. *SIAM Rev.* **48**, 99–130 (2006)
29. García Trillos, N., Slepčev, D.: On the rate of convergence of empirical measures in ∞-transportation distance. *Can. J. Math.* (2015) **(online first)**
30. Giné, E., Koltchinskii, V.: Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In: *High dimensional probability, of IMS Lecture Notes Monogr. Ser., Inst. Math. Statist.*, Beachwood, OH, vol. 51, pp. 238–259 (2006)
31. Gobbino, M.: Finite difference approximation of the Mumford-Shah functional. *Comm. Pure Appl. Math.* **51**, 197–228 (1998)
32. Gobbino, M., Mora, M.G.: Finite-difference approximation of free-discontinuity problems. *Proc. Roy. Soc. Edinburgh Sect. A.* **131**, 567–595 (2001)
33. Goel, A., Rai, S., Krishnamachari, B.: Sharp thresholds for monotone properties in random geometric graphs. In: *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, New York, ACM, pp. 580–586 (2004)
34. Gupta, P., Kumar, P.R.: Critical power for asymptotic connectivity in wireless networks. In: *Stochastic analysis, control, optimization and applications, Systems Control Found. Appl., Birkhäuser Boston,* Boston, MA, pp. 547–566 (1999)
35. Hein, M., Audibert, J.-Y., Von Luxburg, U.: From graphs to manifolds–weak and strong pointwise consistency of graph Laplacians. In: *Learning theory*, pp. 470–485. Springer, New York, 2005
36. Hein, M., Bühler, T.: An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 847–855, 2010
37. Hein, M., Setzer, S.: Beyond spectral clustering - tight relaxations of balanced graph cuts. In: *Advances in Neural Information Processing Systems (NIPS)* (Eds. J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger) pp. 2366–2374, 2011
38. Leighton, T., Shor, P.: Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica.* **9**, 161–187 (1989)
39. Leoni, G.: *A first course in Sobolev spaces of Graduate Studies in Mathematics*. vol. 105, American Mathematical Society, Providence, 2009
40. Maier, M., von Luxburg, U., Hein, M.: How the result of graph clustering methods depends on the construction of the graph. *ESAIM Probab. Stat.*, **17**, 370–418 (2013)
41. Merkurjev, E., Kostić, T., Bertozzi, A.L.: An mbo scheme on graphs for classification and image processing. *SIAM J. Imaging Sci.* **6**, 1903–1930 (2013)
42. Modica, L., Mortola, S.: Un esempio di Γ-convergenza. *Boll. Un. Mat. Ital. B (5).* **14**, 285–299 (1977)

43. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differ. Equ.* **26**, 101–174 (2001)
44. Pedregal, P.: Parametrized measures and variational principles. Progress in Nonlinear Differential Equations and their Applications, 30, Birkhäuser Verlag, Basel (1997)
45. Penrose, M.: A strong law for the longest edge of the minimal spanning tree. *Ann. Probab.* **27**, 246–260 (1999)
46. Ponce, A.C.: A new approach to Sobolev spaces and connections to Γ-convergence. *Calc. Var. Partial Differ. Equ.* **19**, 229–255 (2004)
47. Rangapuram, S.S., Hein, M.: Constrained 1-spectral clustering. In: *International conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1143–1151, 2012
48. Savin, O., Valdinoci, E.: Γ-convergence for nonlocal phase transitions. *Ann. Inst. H. Poincaré Anal. Non Linéaire.* **29**, 479–500 (2012)
49. Shi, J., Malik, J.: Normalized cuts and image segmentation. *Pattern Anal. Mach. Intell. IEEE Trans.* **22**, 888–905 (2000)
50. Shor, P.W., Yukich, J.E.: Minimax grid matching and empirical measures. *Ann. Probab*. **19**, 1338–1348 (1991)
51. Singer, A.: From graph to manifold Laplacian: the convergence rate. *Appl. Comput. Harmon. Anal*. **21**, 128–134 (2006)
52. Szlam, A., Bresson, X.: A total variation-based graph clustering algorithm for cheeger ratio cuts. *UCLA CAM Report*, pp. 1–12 (2009)
53. Szlam, A., Bresson, X.: Total variation and cheeger cuts., In: *ICML* (Eds. J. Frnkranz and T. Joachims) Omnipress pp. 1039–1046 (2010)
54. Talagrand, M.: The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab*. **22**, 919–959 (1994)
55. Talagrand, M.: *The generic chaining. Springer Monographs in Mathematics* Springer-Verlag, Berlin, 2005
56. Talagrand, M.: *Upper and lower bounds of stochastic processes of Modern Surveys in Mathematics*. vol. 60. Springer-Verlag, Berlin Heidelberg, 2014
57. Talagrand, M., Yukich, J.E.: The integrability of the square exponential transportation cost. *Ann. Appl. Probab*. **3**, 1100–1111 (1993)
58. Ting, D., Huang, L., Jordan, M.I.: An analysis of the convergence of graph Laplacians. In: *Proceedings of the 27th International Conference on Machine Learning*, 2010
59. van Gennip, Y., Bertozzi, A.L.: Γ-convergence of graph Ginzburg-Landau functionals. *Adv. Differential Equations*, **17**, 1115–1180 (2012)
60. Villani, C.: *Topics in Optimal Transportation, Graduate Studies in Mathematics*, vol. 58. American Mathematical Society, Providence, US (2003)
61. von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. *Ann. Statist*. **36**, 555–586 (2008)
62. Weyl, H.: On the Volume of Tubes. *Amer. J. Math*. **61**, 461–472 (1939)

Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh,
PA,15213,
USA
e-mail: ngarciat@andrew.cmu.edu
e-mail: slepcev@math.cmu.edu

(*Received September 18, 2014 / Accepted September 15, 2015*)
Published online September 29, 2015 –