Carnegie Mellon University

MELLON COLLEGE OF SCIENCE

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE Variational Limit of Graph Cuts on Point Clouds

PRESENTED BY Nicolas Garcia Trillos

ACCEPTED BY THE DEPARTMENT OF Mathematical Sciences

Dejan Slepcev	May 2015
MAJOR PROFESSOR	DATE
Thomas Bohman	May 2015
DEPARTMENT HEAD	DATE

APPROVED BY THE COLLEGE COUNCIL

Frederick J. Gilman	
DEAN	DATE

VARIATIONAL LIMIT OF GRAPH CUTS ON POINT CLOUDS

Nicolás García Trillos

Submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in
Mathematical Sciences

Carnegie Mellon University
Pittsburgh, PA

May 2015

Para Lelia, Camilo, Luz y Guillo.

Abstract

The main goal of this thesis is to develop tools that enable us to study the convergence of minimizers of functionals defined on point clouds towards minimizers of equivalent functionals in the continuum; the point clouds we consider are samples of a ground-truth distribution. In particular, we investigate approaches to clustering based on minimizing objective functionals defined on proximity graphs of the given sample. Our focus is on functionals based on graph cuts like the Cheeger and ratio cuts. We show that minimizers of these cuts converge as the sample size increases to a minimizer of a corresponding continuum cut (which partitions the ground-truth distribution). Moreover, we obtain sharp conditions on how the connectivity radius can be scaled with respect to the number of sample points for the consistency to hold. We provide results for two-way and for multi-way cuts. The results are obtained by using the notion of Γ-convergence and an appropriate choice of metric which allows us to compare functions defined on point clouds with functions defined on continuous domains.

Acknowledgements

Thank you:

To my Lelia who was always there for me, even when I needed to make graphs and pictures.

To my family (specially my parents and brother) and all my friends who supported me from a distance in my Colombia.

To all my friends and classmates here in Pittsburgh with whom I shared thoughts, feelings, dreams, and imaginary revolutions.

To the staff in the math department at Carnegie Mellon for their help and for allowing me to experience their unique workspace in the Steelers nation.

To the people who opened the doors of Carnegie Mellon to me, especially Professors Bill Hrusa and Kasper Larsen.

To all the faculty members of the Center of Nonlinear Analysis who welcomed me and supported me during my research.

To Thomas Laurent, James von Brecht, and Xavier Bresson for collaborating with me during these past two years.

To my thesis committee: Professors Irene Fonseca, Giovanni Leoni and Larry Wassermann, for valuable advice during the preparation of this thesis and for the time they gave to me during these past years.

To my adviser Professor Dejan Slepĉev for his immense dedication, support, patience and of course *wisdom*.

A good portion of the text and images in this thesis also appears in:

- *Continuum limit of total variation on point clouds.* (with Dejan Slepčev). *Preprint*.
- *On the rate of convergence of empirical measures in ∞-Wasserstein distance.* (with Dejan Slepčev). *To appear in Canad. J. Math. 2015*.
- *Consistency of Cheeger and Ratio graph cuts.* (with Dejan Slepčev, James von Brecht, Thomas Laurent, Xavier Bresson). *Preprint*.
- *Learning Perimeter from graph cuts.* (with Dejan Slepčev, James von Brecht). *Preprint.*

This thesis puts those papers in perspective.

Contents

List of Figures

Chapter 1

Introduction

With the boom of machine learning and data analysis during the last decades, new opportunities for research in fields like statistics, computer science and mathematics have emerged. Strong theoretical results in these areas are important as they support the development of new and better procedures aiming to extract information from data. Many important applications such us automated cancer detection or interpretation of economic data rely on such procedures.

The topic of this thesis lies at the intersection of calculus of variations, geometric measure theory, optimal transportation and the applications of these areas to problems that arise from data analysis, machine learning, and statistics. In particular, this work focuses on the study of the convergence of solutions of minimization problems in random geometric graphs towards solutions of analogous minimization problems in the continuum. The minimization problems of interest in this thesis are those connected to important tasks in machine learning like clustering and specifically graph-based clustering procedures.

In the context of data analysis, the goal of understanding the behavior of graph-based minimization problems in the large sample limit, can be interpreted as studying consistency of the procedures obtained by solving the minimization problems at the discrete level: such procedures should converge as the number of samples goes to infinity to a ground-truth limiting procedure. In my opinion, the most important contribution of this work is that of creating a framework where different tools from mathematical analysis can be combined, thus making it possible to study in a rigorous setting problems arising from statistics and machine learning. This work shows that modern analytical techniques can be used to obtain strong results on classic questions in those fields. The specific results that are presented in this thesis on consistency of graph cuts are one example of the application of such mathematical tools.

Graph cuts

In general, when given a data cloud, the goal is to extract information from it. Mathematically speaking, a data cloud is a point cloud in some high-dimensional Euclidean space, which is obtained by sampling some underlying ground-truth probability distribution. One way to extract information from the data is to study different geometrical and topological structures of the cloud; these structures are thought of as best approximations to those of the ground-truth which one is trying to recognize. For example, one may want to identify meaningful groups among the data points in order to get an idea of the global organization of the ground-truth distribution; this is the goal of one of the fundamental tasks in data analysis and machine learning known as clustering.

Regardless of the approach used to clustering, a partition of a point cloud into clusters should capture the idea that points in each of the groups in the partition are more similar among themselves than with points in other groups. Some of the popular clustering algorithms include those based on graph cuts [59,45,63,9], of centroid type like k-means [44], relaxations of normalized cuts like spectral clustering [72], and of agglomeration type (or hierarchical type) [75].

In this work we consider graph-based procedures to clustering. These procedures rely on creating a graph out of the data cloud by connecting nearby points. This allows one to leverage the geometry of the data set and obtain high quality clustering. So, given a data cloud $V := \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ we construct a similarity graph induced by the distance between the points. Typically, the graphs constructed are an instance of geometric graphs, where a length scale ε is chosen so that points within distance ε are connected by an edge which is given significant weight. With the graph structure at hand, the basic desire to obtain clusters which are well separated leads to the introduction of objective functionals which penalize the size of *cuts* between clusters.

Definition 1.0.1. *Let $G = (V, W)$ be a weighted undirected graph. For a given subset $Y \subseteq V$, the* graph cut *between Y and its complement Y^c is given by*

$$\mathrm{Cut}(Y, Y^c) := \sum_{\boldsymbol{x}_i \in Y} \sum_{\boldsymbol{x}_j \in V \setminus Y} W_{ij}.$$

The cut between Y and its complement measures the level of interaction between both sets. From that point of view, if the task at hand was to partition the point cloud V into two clusters, then one possible approach to obtain such partition could be to consider the minimization problem

$$\text{Minimize} \quad \mathrm{Cut}(Y, Y^c) \quad \text{over all nonempty } Y \subsetneq V. \tag{1.1}$$

This approach is natural, simple and computationally efficient (due to its connection to the max-flow problem). Nevertheless it has important disadvantages, the principal being its sensitivity to outliers. The desire to have clusters of meaningful size and for approaches to be robust to outliers leads to the introduction of ”balance” terms. The generic *balanced graph cut* minimization problem takes the form

$$\text{Minimize} \quad \frac{\mathrm{Cut}(Y, Y^c)}{\mathrm{Bal}(Y, Y^c)} := \frac{\sum_{\mathbf{x}_i \in Y} \sum_{\mathbf{x}_j \in Y^c} W_{ij}}{\mathrm{Bal}(Y, Y^c)} \quad \text{over all nonempty } Y \subsetneq V, \tag{1.2}$$

where $\mathrm{Bal}(Y,Y^c)$ is a *balanced term* which penalizes the asymmetry in size of Y and its complement. Well-known balance terms include

$$\mathrm{Bal}_{\mathrm{R}}(Y,Y^c) = 2|Y||Y^c| \qquad \text{and} \qquad \mathrm{Bal}_{\mathrm{C}}(Y,Y^c) = \min(|Y|,|Y^c|), \tag{1.3}$$

where for $Y \subseteq V$, $|Y|$ represents the ratio between the number of points in Y and the number of points in V. The balance terms in (1.3) correspond to Ratio Cut [42,45,72,73] and Cheeger Cut [9, 30, 32] respectively. For the remainder we will always consider the balance terms associated to either Cheeger or Ratio cut, but we remark that a variety of other balance terms have appeared in the literature in the context of two-class and multi-class clustering [22,45]. We refer to a pair $\{Y,Y^c\}$ that solves (1.2) as an *optimal balanced cut of the graph*. Note that a given graph $G = (V,W)$ may have several optimal balanced cuts (although generically the optimal cut is unique).

We are also interested in multi-class balance cuts. Specifically, in order to partition the set V into $R \geq 3$ clusters, we consider the following ratio cut functional

$$\underset{(Y_1,\dots,Y_R)}{\mathrm{Minimize}} \quad \sum_{r=1}^{R} \frac{\mathrm{Cut}(Y_r,Y_r^c)}{|Y_r|}, \quad Y_r \cap Y_s = \emptyset \quad \text{if} \quad r \neq s, \quad \bigcup_{r=1}^{R} Y_r = V. \tag{1.4}$$

The previous functional is the prototypical multi-class balance cut we consider in the remainder.

Continuum partitioning

Having defined the notion of balanced cuts in a discrete setting, we now consider the analogous notion in the continuum. Let $D \subseteq \mathbb{R}^d$ be a bounded, connected, open domain and let ν be a probability measure on D with positive density $\rho > 0$. A balanced domain-cut problem takes the form

$$\mathrm{Minimize} \quad \frac{\mathrm{Per}(A;\rho^2)}{\mathrm{Bal}_\rho(A,A^c)}, \qquad A \subseteq D \ \text{ with } 0 < \nu(A) < 1. \tag{1.5}$$

where $A^c = D \backslash A$. Just as the graph cut term $\mathrm{Cut}(Y,Y^c)$ in (1.2) provides a weighted (by W) measure of the boundary between Y and Y^c, the cut term $\mathrm{Per}(A;\rho^2)$ for a domain denotes a ρ^2-weighted area of the boundary between the sets A and A^c. If $\partial_D A := \partial A \cap D$ (the boundary between A and A^c) is a smooth curve (in 2d), surface (in 3d) or manifold (in 4d+) then we define

$$\mathrm{Per}(A;\rho^2) \ := \ \int_{\partial_D A} \rho^2(x)\, \mathrm{d}S(x). \tag{1.6}$$

For our results and analysis we need the notion of continuous cut which is defined for sets with less regular boundary. We present the required notions of geometric measure theory in Section 4.1 and the rigorous and mathematically precise formulation of problem (1.5) in Section 5.2 .

If $\rho(x) = 1$ then $\mathrm{Per}(A;\rho^2)$ simply corresponds to arc-length (in 2d) or surface area (in 3d). In the general case, the presence of $\rho^2(x)$ in (1.6) indicates that the regions of low density are easier to cut, so $\partial_D A$ has a tendency to pass through regions in D of low density. As in the graph case, we consider balance terms

$$\mathrm{Bal}_\rho(A,A^c) = 2|A||A^c| \qquad \text{and} \qquad \mathrm{Bal}_\rho(A,A^c) = \min(|A|,|A^c|), \tag{1.7}$$

which correspond to weighted continuous equivalents of the Ratio Cut and the Cheeger Cut. In the continuum setting $|A|$ stands for the total ν-content of the set A, that is,

$$|A| = \nu(A) = \int_A \rho(x)\,\mathrm{d}x. \tag{1.8}$$

We refer to a pair $\{A,A^c\}$ that solves (1.5) as an *optimal balanced cut of the domain*.

In a similar fashion, the continuum equivalent of the multiway cut problem (1.4) reads

$$\underset{(A_1,\dots,A_R)}{\text{Minimize}} \quad \sum_{r=1}^{R} \frac{\mathrm{Per}(A_r;\rho^2)}{|A_r|}, \quad A_r \cap A_s = \emptyset \quad \text{if} \quad r \neq s, \quad \bigcup_{r=1}^{R} A_r = D. \tag{1.9}$$

Consistency of partitioning of data clouds

To introduce the questions we will answer in this thesis and give an idea of the type of theoretical results that are going to be established, let us assume that a given data cloud $V_n = \{\mathbf{x}_1,\dots,\mathbf{x}_n\}$ in $\mathbb{R}^d$ is obtained by sampling some density ρ. In Figure 1.1 below, a data cloud sampled from the uniform distribution on the domain D appearing in the background is illustrated. By selecting the parameter ε, an associated geometric graph is then constructed by connecting two points with an edge provided the points are within distance ε from each other. The resulting graph is illustrated in Figure 1.2. Finally, we consider the corresponding Cheeger cut problem and its solution is presented in Figure 1.3. This solution was obtained by using a recent set of algorithms [25, 62, 26] which perform well in practice.

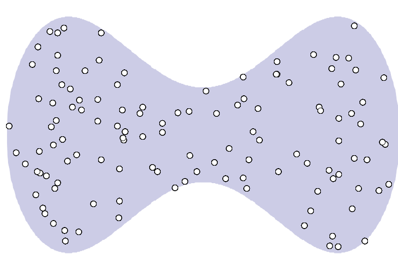


Figure 1.1: A sample of $n = 120$ points.

Figure 1.2: Geometric graph with $\varepsilon = 0.3$.

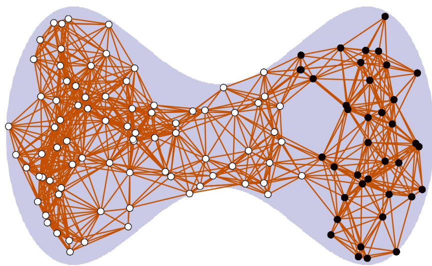


Figure 1.3: Minimizer of Cheeger cut.

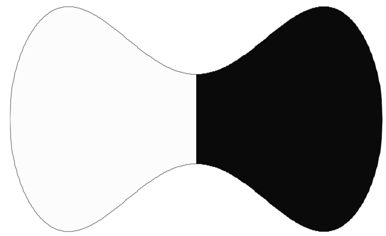


Figure 1.4: Minimizer of continuous Cheeger cut problem.

We also consider the continuous version of the Cheeger cut problem using constant ρ. The solution to the continuous Cheeger cut problem is illustrated in Figure 1.4. Now we observe that despite the fact that the solutions of the discrete and continuous problems are of different nature (one is discrete the other one is continuous), at first glance one can say that they resemble each other. Motivated by this observation, we ask the following questions. First, in what sense (what topology or metric) are the solutions close? Second, can we actually prove that optimal graph cuts converge to optimal domain cuts as the number of data points increases? and related to this last question, how does the construction of the graph, i.e., how does the choice of ε, influence or guarantee such convergence? The previous questions motivated the results that are presented in this manuscript. Note that from a statistical point of view, the previous questions can be interpreted as asking about consistency of graph cut clustering procedures in the large sample limit.

We remark that an important consideration when investigating consistency of optimal cuts of the graph is how the graphs on V_n are actually constructed. In simple terms, when building a graph on V_n one sets the length scale ε_n such that edges between vertices in V_n are given significant weights if the distance of points they connect is ε_n or less. In some way this sets the length scale over which the geometric information is averaged when setting up the graph. Taking smaller ε_n is desirable because it is computationally less expensive and gives better resolution, but there is a price. Taking ε_n small increases the error due to randomness and in fact if ε_n is too small the resulting graph may not represent the geometry of D well and consequently the discrete graph cut may be very far from the desired one. We determine precisely how small ε_n can be taken for the consistency to hold. We obtain consistency results both for two-way and multiway cuts.

Informal statement of (a part of) the main results. *Consider* $d \geq 2$ *and assume the continuum balanced cut* (1.5) *has a unique minimizer* $\{A, A^c\}$*. Consider* $\varepsilon_n > 0$ *such that* $\lim_{n\to\infty} \varepsilon_n = 0$ *no faster than an explicit (to be determined) rate. Then almost surely the minimizers,* $\{Y_n, Y_n^c\}$*, of the balanced cut* (1.2) *of the graph* G_n *, converge to* $\{A, A^c\}$*. Moreover, after appropriate rescaling, almost surely the minimum of problem* (1.2) *converges to the minimum of* (1.5)*. The result also holds for multiway cuts. That is the minimizers of* (1.4) *converge* *towards minimizers of* (1.9).

We remark that, despite the fact that for a general graphs the problems (1.2) and (1.4) are NP hard, in practice when the graph is obtained by sampling from a measure ν as above, such minimization problems can be effectively approached [23, 24]. In fact, by choosing an appropriate initialization, the algorithms (see [23, 24]), give very good results in clustering real-world data.

Background on consistency of clustering algorithms and related problems

Consistency of clustering algorithms has been considered for a number of approaches. Pollard [57] has proved the consistency of k-means clustering. Consistency for a class of single linkage clustering algorithms was shown by Hartigan [43]. Arias-Castro and Pelletier have proved the consistency of maximum variance unfolding [7]. Consistency of spectral clustering was rigorously considered by von Luxburg, Belkin, and Bousquet [71]. These works show the convergence of all eigenfunctions of the graph Laplacian for fixed length scale $\varepsilon_n = \varepsilon$ which results in the limiting (as $n \to \infty$) continuum problem being a nonlocal one. Belkin and Niyogi [13] consider the spectral problem (Laplacian eigenmaps) and show that there exists a sequence $\varepsilon_n \to 0$ such that in the limit the (manifold) Laplacian is recovered, however no rate at which ε_n can go to zero is provided. Consistency of normalized cuts was considered by Arias-Castro, Pelletier, and Pudlo [8] who provide a rate on $\varepsilon_n \to 0$ under which the minimizers of the discrete cut functionals minimized over a specific family of subsets of V_n converge to the continuum Cheeger set. Our work improves on [8] in several ways. We minimize the discrete functionals over all discrete partitions on V_n as it is considered in practice and prove the result for the optimal, in terms of scaling, range of rates at which $\varepsilon_n \to 0$ as $n \to \infty$.

There are also a number of works which investigate how well the discrete functionals approximate the continuum ones for a particular function. Among them are works by Belkin and Niyogi [14], Giné and Koltchinskii [37], Hein, Audibert, von Luxburg [46], Singer [61] and Ting, Huang, and Jordan [68]. Maier, von Luxburg and Hein [51] considered pointwise convergence for Cheeger and normalized cuts, both for the geometric and kNN graphs and obtained a range of scalings of graph construction on n for the convergence to hold. While these results are quite valuable, we point out that they do not imply that the minimizers of discrete objective functionals are close to minimizers of continuum functionals.

Setup and main results

We consider the sample $V_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ consisting of i.i.d. random points drawn from an underlying ground-truth measure ν supported on $\overline{D}$. We assume that D is a bounded, open set with Lipschitz boundary and that ν has continuous density ρ which is bounded below and above by positive constants, that is, we assume that there exists $\lambda \geq 1$ such that

$$\frac{1}{\lambda} \leq \rho(x) \leq \lambda, \quad \forall x \in D. \tag{1.10}$$

In general, a geometric graph induced by the point cloud V_n is constructed as follows. Let $\eta : \mathbb{R}^d \to [0,\infty)$ be a radially symmetric, radially decreasing kernel decaying to zero at a sufficiently fast rate. For a specified $\varepsilon > 0$, for all $i, j \in \{1,\dots,n\}$ we let the weight between $\mathbf{x}_i$ and $\mathbf{x}_j$ be given by

$$W_{ij} := \eta\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{\varepsilon}\right). \tag{1.11}$$

To be more precise on the assumptions we impose on the kernel η, let us denote by $\boldsymbol{\eta} : [0,\infty) \to [0,\infty)$ its radial profile, that is we consider $\boldsymbol{\eta}$ such that $\eta(x) = \boldsymbol{\eta}(|x|)$. 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 $0 \leq t \leq 1$ and $\boldsymbol{\eta} = 0$ for $t > 1$. We remark that the assumption (K3) is equivalent to imposing that the *surface tension*

$$\sigma_\eta := \int_{\mathbb{R}^d} \eta(h)|h_1|\mathrm{d}h, \tag{1.12}$$

where h_1 is the first coordinate of vector h, is finite. We also remark that in (1.12), one can replace h_1 by $h \cdot e$ for any fixed $e \in \mathbb{R}^d$ with norm one; this because η is radially symmetric.

After setting the assumptions on the kernel η, we turn our attention to the following: as one is considering functions supported on the graphs (characteristic functions of subsets of V_n), the issue is how to compare them with functions in the continuum setting, and how to compare functions defined on different graphs. We introduce the TL^p spaces to answer that question.

First, let us denote by ν_n the empirical measure associated to the data points $\mathbf{x}_1,\dots,\mathbf{x}_n$, that is

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{x}_i}. \tag{1.13}$$

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

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

where $\mathscr{P}_p(D)$ denotes the set of Borel probability measures on D with integrable p-moment. We use $\mathscr{P}(D)$ to denote the set of all 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 2.2, 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. To give a more intuitive interpretation of the convergence in TL^p, we make use of its characterization in Proposition 2.2.13. Suppose that $\nu \in \mathscr{P}_p(D)$ is as before, and let ν_n be the empirical measure associated to the sample $\mathbf{x}_1,\dots,\mathbf{x}_n$ drawn from ν. Let $u_n \in L^p(D,\nu_n)$ and $u \in L^p(D,\nu)$. Given a sequence of *transportation maps* $\{T_n\}_{n\in\mathbb{N}}$, where for each $n\in\mathbb{N}$ ν_n is the *push forward* of ν by T_n, such that

$$\int_D |T_n(x)-x|^p \mathrm{d}\nu(x) \to 0, \qquad \text{as } n\to\infty, \tag{1.14}$$

the statements $(\nu_n,u_n) \xrightarrow{TL^p} (\nu,u)$ as $n\to\infty$ and $u_n\circ T_n \xrightarrow{L^p(D,\nu)} u$ as $n\to\infty$ are equivalent. The previous characterization allows us to think of TL^p convergence as convergence of a sort of extrapolation of u_n towards u in the $L^p(D,\nu)$ sense. Indeed, note that for every n, the transportation map T_n induces a partition of the domain D into n regions with the same ν-measure, as well as it induces a matching between those regions and the points $\mathbf{x}_1,\dots,\mathbf{x}_n$. Intuitively, condition (1.14) guarantees that each region is matched to a nearby data point. The function $u_n\circ T_n$ supported on D, can then be thought as an extrapolation of the function u_n in the sense that the value of u_n at some point $\mathbf{x}_i$ is the value of the function $u_n\circ T_n$ on a region which is near the point $\mathbf{x}_i$.

Finally, it is worth remarking that TL^p convergence reduces to more common ways of comparing functions at the discrete level with functions at the continuum level in some settings. For example if the discrete set consisted of points arranged on a regular grid in D, then the standard way [28, 69, 21], to compare functions at the discrete level with functions at the continuum level is to identify functions at the discrete level with piecewise constant functions on the domain D. This is achieved by assigning values on grid cells to be the value at the appropriate grid points. The resulting extensions can then be compared using the usual L^p metric. The notion of convergence induced by such procedure is equivalent to the convergence in TL^p due to Proposition 2.2.13.

Having introduced the TL^p metric, we now have all the necessary ingredients to formulate in a theorem the observations we made about the behavior of optimal balanced cuts of the graph when $n\to\infty$.

Theorem 1.0.2 (Consistency of two-class cuts)**.** *Let $D \subset \mathbb{R}^d$, $d \geq 2$ be an open, bounded, connected set with Lipschitz boundary. Let $\nu \in \mathscr{P}(D)$ be 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 *and satisfying*

$$\begin{aligned} \lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} &= 0 \quad \text{if } d = 2, \\ \lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} &= 0 \quad \text{if } d \geq 3. \end{aligned} \tag{1.15}$$

Let $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\dots$ be a sequence of i.i.d. random points chosen according to distribution ν on D and let $V_n = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$. Let $G_n = (V_n, W_n)$ denote the graph whose edge weights are defined as in (1.11)*, where we assume the kernel η satisfies conditions (K1)-(K3). Finally, let $\{Y_n^*, Y_n^{*c}\}$ denote any optimal balanced cut of G_n (solution of problem* (1.2)*). If $\{A^*, A^{*c}\}$ is the unique optimal balanced cut of the domain D (solution of problem* (5.11)*) then with probability one the sequence $\{Y_n^*, Y_n^{*c}\}$ converges to $\{A^*, A^{*c}\}$ in the TL^1-sense. If there is more than one optimal continuum balanced cut* (5.11) *then $\{Y_n^*, Y_n^{*c}\}$ converges along a subsequence to an optimal continuum balanced cut.*

Additionally, with probability one, $\mathscr{C}_n$, the minimum balance cut of the graph G_n (the minimum of (1.2)*), satisfies*

$$\lim_{n\to\infty} \frac{\mathscr{C}_n}{n^2 \varepsilon_n^{d+1}} = \sigma_\eta \mathscr{C}, \tag{1.16}$$

where σ_η is the surface tension associated to the kernel η defined in (1.12) *and $\mathscr{C}$ is the minimum of* (1.5)*.*

In order to prove Theorem 1.0.2, we have to introduce some ideas and establish several preliminary results. In fact, we will not be able to prove Theorem 1.0.2 until we have introduced and established the required foundations in Chapters 2, 3, 4 and the beginning of Chapter 5. Nevertheless, to get us started in the development of these ideas, let us first observe that the solution to problem (1.2) is unchanged if the functional to be minimized is multiplied by a positive quantity. Having limits as $n \to \infty$ in mind, we first consider a rescaled version of the cut functional as well as we extend the notion of perimeter to functions on the graphs taking real values.

Definition 1.0.3. *Let $G_n = (V_n, W_n)$ be a weighted graph with weights given by* (1.11)*. Given a function $u_n : V_n \to \mathbb{R}$, we define its graph total variation as*

$$GTV_{n,\varepsilon_n}(u_n) := \frac{1}{n^2 \varepsilon_n^{d+1}} \sum_{i,j}^{n} W_{ij} |u_n(\boldsymbol{x}_i) - u_n(\boldsymbol{x}_j)|.$$

Also, we define the graph perimeter of a subset $Y \subseteq V_n$ by

$$\mathrm{Per}_{n,\varepsilon_n}(Y) := GTV_{n,\varepsilon_n}(\mathbf{1}_Y) = \frac{2}{n^2 \varepsilon_n^{d+1}} \mathrm{Cut}(Y, Y^c),$$

which is simply a rescaled version of $\mathrm{Cut}(Y,Y^c)$.

As mentioned previously, the problem of minimizing (1.2) can then be rewritten as

$$\text{Minimize} \quad \frac{\mathrm{Per}_{n,\varepsilon_n}(Y)}{\mathrm{Bal}(Y,Y^c)} \quad \text{over all nonempty } Y \subsetneq V. \tag{1.17}$$

Now, note that roughly speaking Theorem 1.0.2 says that with probability one, the minimizers of (1.17) converge in a suitable sense to minimizers of problem (1.5) and that the minimum in (1.17) converges to a constant multiple of the minimum of problem (1.5). Looking at the functionals to be minimized in (1.17) and (1.5), a connection between $\mathrm{Per}_{n,\varepsilon_n}$ (the graph perimeter) and $\mathrm{Per}(\cdot;\rho^2)$ (the weighted perimeter in the continuum) should be expected. The relevant question is if the functional $\mathrm{Per}_{n,\varepsilon_n}$ approximates $\mathrm{Per}(\cdot;\rho^2)$ in a suitable sense, noting that the sense in which we would like $\mathrm{Per}_{n,\varepsilon_n}$ to approximate $\mathrm{Per}(\cdot;\rho^2)$ should guarantee the convergence of minimizers in Theorem 1.0.2. An example of a notion of convergence of functionals that will not guarantee the consistency in Theorem 1.0.2 is that of *pointwise convergence*. For the sake of simplicity let us assume that $\boldsymbol{\eta}$ is of the form $\boldsymbol{\eta} = 1$ for $0 \leq t \leq 1$ and $\boldsymbol{\eta} = 0$ for $t > 1$, that D is the unit box $(0,1)^d$ and that ρ is constant. As discussed in Chapter 7, for a fixed measurable subset $A \subseteq D$, $\mathrm{Per}_{n,\varepsilon_n}(A \cap V_n)$ is a consistent estimator of the relative perimeter $\mathrm{Per}(A)$ of A, provided that $\frac{1}{n^{2/(d+1)}} \ll \varepsilon_n \ll 1$. In particular, pointwise convergence holds even for ε_n decaying at a faster rate than the connectivity rate for random geometric graphs (see [56, 39, 41]) which is $O\left(\frac{(\log(n))^{1/d}}{n^{1/d}}\right)$.

For the highly disconnected graphs that can be obtained by choosing $\frac{1}{n^{2/(d+1)}} \ll \varepsilon_n \ll \frac{(\log(n))^{1/d}}{n^{1/d}}$, (and for which the pointwise convergence still holds), it is clear that Theorem 1.0.2 will not hold, as in that case the minimum of the graph cut problem will be zero for all large enough n, whereas the minimum of the domain cut problem will be positive.

Thus instead of focusing on the notion of pointwise convergence, we focus on a different type of convergence for functionals which is better suited for studying the convergence of minimizers in Theorem 1.0.2. Introduced by De Giorgi in the 70's, the notion of Γ-convergence is appropriate as it gives precise and sufficient conditions under which minimizers of a sequence of functionals converge to minimizers of a limiting functional. Used in the context of deterministic variational problems in diverse fields like phase transitions [3,53,52] and homogenization [19, 33, 20], this notion of convergence has a natural extension to random functionals like the balance graph cut obtained from random points. In Section 1.1, we review some basic facts about Γ-convergence, as well as we introduce the proper extension to problems where there is randomness involved.

In Section 6.1 we show that the graph total variation approximates in the Γ-convergence sense (with respect to the TL^1-metric) the weighted total variation $TV(\cdot;\rho^2)$, which for smooth functions $u : D \to \mathbb{R}$ is given by

$$TV(u;\rho^2) := \int_D |\nabla u| \rho^2(x) \mathrm{d}x,$$

and in general defined by

$$TV(u;\rho^2) = \sup\left\{ \int_D u \,\mathrm{div}(\phi) \mathrm{d}x \,:\, (\forall x \in D)\ |\phi(x)| \leq \rho^2(x)\,,\ \phi \in C_c^\infty(D:\mathbb{R}^d) \right\}, \tag{1.18}$$

see Section 4.1. We also show that the graph perimeter Γ-converges in the TL^1-metric to the weighted perimeter $\mathrm{Per}(\cdot;\rho^2)$ which for sets $A \subseteq D$ with regular boundary is given by (1.6) and for general measurable sets $A \subseteq D$ is given by

$$\mathrm{Per}(A;\rho^2) = TV(\mathbf{1}_A;\rho^2).$$

More precisely the following holds.

Theorem 1.0.4 (Γ-convergence of GTV_{n,ε_n})**.** *Let $D \subset \mathbb{R}^d$, $d \geq 2$, be an open, bounded, connected set with Lipschitz boundary. Let $\nu \in \mathscr{P}(D)$ be a probability measure on D with continuous density ρ, which is bounded from below and above by positive constants. Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \dots$ 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*

$$\begin{aligned} \lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} &= 0 \quad \text{if } d = 2, \\ \lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} &= 0 \quad \text{if } d \geq 3. \end{aligned} \tag{1.19}$$

Assume the kernel η satisfies conditions (K1)-(K3). Then, GTV_{n,ε_n}, defined in Definition 1.0.3, Γ-converge to $\sigma_\eta TV(\cdot;\rho^2)$ as $n \to \infty$ in the TL^1 sense, where σ_η is given by (1.12) *and $TV(\cdot;\rho^2)$ is the weighted total variation functional defined in* (1.18)*.*

Corollary 1.0.5 (Γ-convergence of $\mathrm{Per}_{n,\varepsilon_n}$)**.** *Under the assumptions of Theorem 1.0.4, with probability one the following statement holds: for every $A \subseteq D$ measurable, there exists a sequence of sets $\{Y_n\}_{n\in\mathbb{N}}$ with $Y_n \subseteq V_n$ such that,*

$$\mathbf{1}_{Y_n} \xrightarrow{TL^1} \mathbf{1}_A$$

and

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(\mathbf{1}_{Y_n}) \leq \sigma_\eta \,\mathrm{Per}(A;\rho^2).$$

We also establish the following compactness result.

Theorem 1.0.6 (Compactness of GTV_{n,ε_n})**.** *Under the assumptions of Theorem 1.0.4, the following statement holds with probability one: every sequence $\{u_n\}_{n\in\mathbb{N}}$ with $u_n \in L^1(D,\nu_n)$ (ν_n*

is given by (1.13)*) and with uniformly bounded* $L^1(D,\nu_n)$ *norms and graph total variations,* GTV_{n,ε_n} *is relatively compact in* TL^1*. More precisely, with probability one, 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.*

In order to understand the role that the parameter ε_n has in the convergence of GTV_{n,ε_n} toward the total variation in the continuum, it is important to observe that in order to recover the geometric information of the underlying ground-truth distribution, the graph (V_n, W_n) should have information on a larger scale than that on which the randomness of the problem operates. This implies understanding the spatial difference between the underlying distribution ν and the empirical measure ν_n associated to the data points. The way we measure such spatial difference between ν and ν_n is by finding the best way to match regions in the domain D with the data points, so that the maximum distance an arbitrary point in the underlying domain D has to travel to meet its matched data point is minimal. Let us outline the proof of Theorem 1.0.4 and make precise the way the spatial difference between ν and ν_n is measured.

The idea is to first introduce the functional $TV_\varepsilon(\cdot;\rho) : L^1(D,\nu) \to [0,\infty)$ given by,

$$TV_\varepsilon(u;\rho) := \frac{1}{\varepsilon}\int_D\int_D \eta_\varepsilon(x-y)|u(x)-u(y)|\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y, \tag{1.20}$$

which serves as an intermediate object between the functionals GTV_{n,ε_n} and $TV(\cdot;\rho^2)$. Here and in the remainder, $\eta_\varepsilon(x-y) = \frac{1}{\varepsilon^d}\eta\left(\frac{x-y}{\varepsilon}\right)$. It is important to observe that the argument of GTV_{n,ε_n} is a function u_n supported on the data points, whereas the argument of $TV_\varepsilon(\cdot;\rho)$ is an $L^1(D,\nu)$ function; in particular a function defined on D. The functional $TV_\varepsilon(\cdot;\rho)$ is a *non-local* functional, where the term *non-local* refers to the fact that differences of a given function on a ε-neighborhood are averaged, which contrasts the local approach of averaging derivatives of the given function. Non-local functionals have been of interest in the last decades due to their wide range of applications which includes phase transitions, image processing and PDEs. From a statistical point of view, for a fixed function $u : D \to \mathbb{R}$, $TV_{\varepsilon_n}(u;\rho)$ is nothing but the expectation of $GTV_{n,\varepsilon_n}(u)$. On the other hand, the functionals $TV_\varepsilon(\cdot;\rho)$ are relevant for our purposes because as $\varepsilon \to 0$ not only they approximate $TV(\cdot;\rho^2)$ in a pointwise sense, but they also approximate it in the Γ-convergence sense. More precisely the following result is established in Section 4.2.

Proposition 1.0.7. *Consider an open, connected, 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 and consider the measure* $\mathrm{d}\nu := \rho\mathrm{d}x$. *Le* $\{\varepsilon_k\}_{k\in\mathbb{N}}$ *be a sequence of positive numbers converging to zero. Then,* $\{TV_{\varepsilon_k}(\cdot;\rho)\}_{k\in\mathbb{N}}$ *(defined in* (1.20)*)* Γ*-converges with respect to the*

$L^1(D,\nu)$-metric to $\sigma_\eta TV(\cdot;\rho^2)$, where σ_η is defined in (1.12) *and $TV(\cdot;\rho^2)$ is defined in* (1.18). *Moreover, the functionals $\{TV_{\varepsilon_k}(\cdot;\rho)\}_{k\in\mathbb{N}}$ satisfy the compactness property, with respect to the $L^1(D,\nu)$-metric. That is, every sequence $\{u_k\}_{k\in\mathbb{N}}$ with $u_k \in L^1(D,\nu)$ for which*

$$\sup_{k\in\mathbb{N}} ||u_k||_{L^1(D,\nu)} < \infty, \quad \sup_{k\in\mathbb{N}} TV_{\varepsilon_k}(u_k;\rho) < \infty,$$

is precompact in $L^1(D,\nu)$. Finally, for every $u \in L^1(D,\nu)$

$$\lim_{n\to\infty} TV_{\varepsilon_k}(u;\rho) = \sigma_\eta TV(u;\rho^2). \tag{1.21}$$

As observed earlier, the argument of GTV_{n,ε_n}, is a function u_n supported on the data points, while the argument of $TV_{\varepsilon_n}(\cdot;\rho)$ is an $L^1(D,\nu)$ functional. For a function u_n defined on V_n, the idea is to associate an $L^1(D,\nu)$ function $\tilde{u}_n$ which approximates u_n in the TL^1-sense and is such that $TV_{\varepsilon_n}(\tilde{u}_n;\rho)$ is comparable to $GTV_{n,\varepsilon_n}(u_n)$. The purpose of doing this is to use Proposition 1.0.7. We construct the approximating function $\tilde{u}_n$ by using transportation maps (i.e. measure preserving maps) between the measure ν and ν_n. More precisely, we set $\tilde{u}_n = u_n \circ T_n$ where T_n is a transportation map between ν and ν_n which moves mass as little as possible. The estimates on how far the mass needs to be moved were known in the literature when ρ is constant and when the domain D is the unit cube $(0,1)^d$ (see [49, 64, 66, 67] for $d = 2$ and [60] for $d \geq 3$). In [36] these estimates are extended to general domains D and densities ρ satisfying (1.10). We present the proofs of the following result in Chapter 3.

Proposition 1.0.8. *Let $D \subseteq \mathbb{R}^d$ be a bounded, connected, open set with Lipschitz boundary. Let ν be a probability measure on D with density $\rho : D \to (0,\infty)$ satisfying* (1.10). *Let $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\dots$ be i.i.d. samples from ν. Let ν_n be the empirical measure associated to the n data points. Then, for any fixed $\alpha > 2$, except on a set with probability $O(n^{-\alpha/2})$, there exists a transportation map $T_n : D \to D$ between the measure ν and the measure ν_n (denoted $T_{n\sharp}\nu = \nu_n$) such that*

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

where C depends only on α, D, and λ (see 1.10).

The rates presented in the previous Theorem are optimal as discussed in Chapter 3. From the previous result, Chebyshev's inequality and the Borel-Cantelli lemma one obtains the following rate of convergence of the ∞-*optimal transportation* distance between the empirical measures ν_n and the measure ν.

Proposition 1.0.9. *Let D be an open, connected and bounded subset of $\mathbb{R}^d$ which has Lipschitz boundary. Let ν be a probability measure on D with density ρ satisfying* (1.10). *Let*

$\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\dots$ be a sequence of independent samples from ν and let ν_n be the associated empirical measures (1.13)*. Then, there is a constant $C>0$ such that with probability one, there exists a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from ν to ν_n ($T_{n\sharp}\nu=\nu_n$) and such that:*

$$\text{if } d=2 \text{ then} \qquad \limsup_{n\to\infty} \frac{n^{1/2}\|Id-T_n\|_\infty}{(\log n)^{3/4}} \le C \tag{1.22}$$

$$\text{and if } d\ge 3 \text{ then} \qquad \limsup_{n\to\infty} \frac{n^{1/d}\|Id-T_n\|_\infty}{(\log n)^{1/d}} \le C. \tag{1.23}$$

As shown in Section 5.2, Proposition 1.0.7 and Proposition 1.0.9 are at the backbone of Theorem 1.0.4. Schematically,

$$TV_\varepsilon(\cdot;\rho) \xrightarrow{\Gamma} \sigma_\eta TV(\cdot;\rho^2) \quad \text{in } L^1 \quad + \quad \text{Proposition 1.0.9} \Longrightarrow GTV_{n,\varepsilon_n} \xrightarrow{\Gamma} \sigma_\eta TV(\cdot;\rho^2) \quad \text{in } TL^1.$$

We note that the statement $TV_{\varepsilon_n}(\cdot;\rho) \xrightarrow{\Gamma} \sigma_\eta TV(\cdot;\rho^2)$ is a purely analytic, purely deterministic fact. Proposition 1.0.9, on the other hand contains all the probabilistic estimates needed to establish all the results in this work. Such estimates in particular provide the constraints on the parameter ε_n in Theorem 1.0.4. It is worth observing that Proposition 1.0.9 is a statement that only involves the underlying measure ν and the empirical measure ν_n, and that in particular it does not involve estimates on the difference between the functional $TV_{\varepsilon_n}(u;\rho)$ and the functional $GTV_{n,\varepsilon_n}(u)$ for u belonging to a small (in the sense of VC-dimension) class of functions. In other words our estimates are related to the domains where the functions are defined (discrete/continuous) and not to the actual values of functions defined on those domains. The results from Theorem 1.0.4, Corollary 1.0.5 and Theorem 1.0.6 can then be put together to prove Theorem 1.0.2 in Section 5.2 thus establishing the consistency of optimal graph cuts in the two-class setting.

We take our analysis of consistency of graph cuts one step further in order to include the multi-way cut case. In Chapter 6 we establish the following analogue to Theorem 1.0.2.

Theorem 1.0.10. *Let domain D, measure ν, kernel η, sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$, sample points $\{\boldsymbol{x}_i\}_{i\in N}$, and graph G_n satisfy the assumptions of Theorem 1.0.2. Let $({Y^*}^n_1,\dots,{Y^*}^n_R)$ denote any optimal balanced cut of G_n, that is a minimizer of* (1.4)*. If $(A^*_1,\dots,A^*_R)$ is the unique optimal balanced cut of D (i.e. minimizer of* (1.9)*) then with probability one the sequence $({Y^*}^n_1,\dots,{Y^*}^n_R)$ converges to $(A^*_1,\dots,A^*_R)$ in the TL^1-sense. If the optimal continuum balanced cut is not unique then the convergence to a minimizer holds along subsequences. Additionally, $\mathscr{C}_n$, the minimum of* (1.4)*, satisfies*

$$\lim_{n\to\infty} \frac{\mathscr{C}_n}{n^2\varepsilon_n^{d+1}} = \sigma_\eta \mathscr{C},$$

where σ_η is the surface tension associated to the kernel η and $\mathscr{C}$ is the minimum of (1.9).

The proof of Theorem 1.0.10 involves modifying the geometric measure theoretical results used to prove Theorem 1.0.2. This leads to a longer and more technical proof than the proof of Theorem 1.0.2, but the overall spirit of the proof remains the same in the sense that the Γ-convergence plays the leading role.

The overall organization of this manuscript is as follows. In Section 1.1 we give a short background on Γ-convergence. In Chapter 2, we give some background on optimal transportation, we introduce the TL^p spaces and prove some of their properties. In Chapter 3 we establish the results that quantify the convergence of empirical measures towards their underlying distribution in the ∞-transportation distance, that is, we establish Proposition 1.0.8. In Chapter 4 we start by defining rigorously the weighted total variation in the continuum and after that we prove Theorem 1.0.7 which establishes the Γ-convergence of the non-local functionals $TV_\varepsilon(\cdot;\rho)$. In Chapter 5 we use the metric defined in Chapter 2, and the results from Chapter 3 and Chapter 4 to establish the variational convergence of the graph total variation towards the weighted total variation in the continuum, that is, we prove Theorem 1.0.4, Corollary 1.0.5 and Theorem 1.0.6; from those results we establish consistency of Cheeger and Ratio graph cuts (Theorem 1.0.2). In Chapter 6 we establish consistency of multiway cuts i.e. Theorem 1.0.10. Finally, in Chapter 7, we investigate the pointwise convergence of the graph perimeter and show that such notion of convergence is inadequate to investigate the consistency of graph based clustering procedures like Cheeger and ratio cuts as explained earlier. This last chapter is independent of the rest of the chapters of this manuscript and is included with the intention of contrasting the notion of Γ-convergence and that of pointwise convergence.

1.1 Background on Γ-convergence

We recall and discuss the notion of Γ-convergence. In the literature Γ-convergence is defined for deterministic functionals (see [33], [19] for extensive exposition of properties and applications of Γ-convergence). Nevertheless, the objects we are interested in are random and thus we decided to introduce this notion of convergence in this non-deterministic setting.

Let (X,d_X) be a metric space and let $(\Omega,\mathfrak{F},\mathbb{P})$ be a probability space. Let $F_n : X\times\Omega \to [0,\infty]$ be a sequence of random functionals.

Definition 1.1.1. *The sequence $\{F_n\}_{n\in\mathbb{N}}$ Γ-converges with respect to the metric d_X to the deterministic functional $F : X \to [0,\infty]$ as $n\to\infty$ if with $\mathbb{P}$-probability one* ***all*** *of the following conditions hold simultaneously:*

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

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

In the previous definition, it is important to highlight the way the quantifiers are considered, as it is in this way that one guarantees that all the relevant properties obtained from the Γ-convergence in the deterministic setting continue to hold in the random setting.

Remark 1.1.2. *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. This property is not related to the randomness of the functionals in any way.*

Definition 1.1.3. *We say that the sequence of nonnegative random functionals $\{F_n\}_{n\in\mathbb{N}}$ satisfies the compactness property if with $\mathbb{P}$-probability one, the following statement holds: any sequence $\{x_n\}_{n\in\mathbb{N}}$ bounded in X and for which*

$$\sup_{k\in\mathbb{N}} F_n(x_n) < +\infty,$$

is relatively compact in X.

Remark 1.1.4. *The boundedness assumption of $\{x_n\}_{n\in\mathbb{N}}$ in the previous definition is a necessary condition for relative compactness and so it is not restrictive. Once again we also remark the way the quantifiers are considered.*

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 that with $\mathbb{P}$-probability one, minimizers (or approximate minimizers) of F_n converge 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. The next proposition can be found in [19,33]. We present its proof for completeness and for the benefit of the reader. We also want to highlight the way this type of convergence works as ultimately this is one of the essential tools used to prove the main theorems of this paper.

Proposition 1.1.5. *Let $F_n : X \times \Omega \to [0,\infty]$ be a sequence of random nonnegative functionals which are not identically equal to $+\infty$, satisfying the compactness property and Γ-converging to the deterministic functional $F : X \to [0,\infty]$ which is not identically equal to $+\infty$. If it is true that with $\mathbb{P}$-probability one, there is a bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfying*

$$\lim_{n\to\infty}\left(F_n(x_n) - \inf_{x\in X} F_n(x)\right) = 0, \tag{1.24}$$

then, with $\mathbb{P}$-probability one the following statement holds

$$\lim_{n\to\infty} \inf_{x\in X} F_n(x) = \min_{x\in X} F(x). \tag{1.25}$$

Furthermore, every bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ in X satisfying (1.24) *is relatively compact and each of its cluster points is a minimizer of F. In particular, if F has a unique minimizer, then a bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfying* (1.24) *converges to the unique minimizer of F.*

Proof. Consider Ω' a set with $\mathbb{P}$-probability one for which all the statements in the definition of Γ-convergence together with the statement of the compactness property hold. We also assume that for every $\omega \in \Omega'$, there exists a bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfying (1.24). We fix such $\omega \in \Omega'$ and in particular we can assume that F_n is deterministic for every $n \in \mathbb{N}$.

Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence as the one described above. Let $\tilde{x} \in X$ be arbitrary. By the limsup inequality we know that there exists a sequence $\{\tilde{x}_n\}_{n\in\mathbb{N}}$ with $\tilde{x}_n \to \tilde{x}$ and such that

$$\limsup_{n\to\infty} F_n(\tilde{x}_n) \leq F(\tilde{x}).$$

By (1.24) we deduce that

$$\limsup_{n\to\infty} F(x_n) = \limsup_{n\to\infty} \inf_{x\in X} F_n(x) \leq \limsup_{n\to\infty} F_n(\tilde{x}_n) \leq F(\tilde{x}), \tag{1.26}$$

and since $\tilde{x}$ was arbitrary we conclude that

$$\limsup_{n\to\infty} F_n(x_n) \leq \inf_{x\in X} F(x). \tag{1.27}$$

The fact that F is not identically equal to $+\infty$ implies that the term on the right hand side of the previous expression is finite and thus $\limsup_{n\to\infty} F_n(x_n) < +\infty$. Since the sequence $\{x_n\}_{n\in\mathbb{N}}$ was assumed bounded, we conclude from the compactness property for the sequence of functionals $\{F_n\}_{n\in\mathbb{N}}$ that $\{x_n\}_{n\in\mathbb{N}}$ is relatively compact.

Now let x^* be any accumulation point of the sequence $\{x_n\}_{n\in\mathbb{N}}$ (we know there exists at least one due to compactness), we want to show that x^* is a minimizer of F. Working along subsequences, we can assume without the loss of generality that $x_n \to x^*$. By the liminf inequality, we deduce that

$$\inf_{x\in X} F(x) \leq F(x^*) \leq \liminf_{n\to\infty} F(x_n). \tag{1.28}$$

The previous inequality and (1.26) imply that

$$F(x^*) \leq F(\tilde{x}),$$

where $\tilde{x}$ is arbitrary. Thus, x^* is a minimizer of F and in particular $\inf_{x\in X} F(x) = \min_{x\in X} F(x)$. Finally, to establish (1.25) note that this follows from (1.27) and (1.28). □

Chapter 2

The TL^p spaces

In this chapter we introduce a metric which allows us to compare functions defined on a discrete set with functions defined on a continuous domain. The idea is to regard these objects as elements of the same metric space. Let us describe how this is achieved. For arbitrary Borel probability measures ν and μ on D, let us denote by $L^1(D,\nu)$ and by $L^1(D,\mu)$ their corresponding L^1-spaces. Given $u \in L^1(D,\nu)$ and $v \in L^1(D,\mu)$, we can regard them as elements (ν,u) and (μ,v). Now, suppose that there is a map $T : D \to D$ that pushes forward the measure μ into the measure ν, in such a way that on average the difference $|T(x)-x|$ is as "small" as possible . Then the map T induces a function $\hat{u} := u \circ T \in L^1(D,\mu)$ which can be compared to v using the $L^1(D,\mu)$ metric. A possible way to compare the functions u and v is then to determine how far is μ from ν (measured by using the "best" T as described above) and how far is $\hat{u}$ from v using the $L^1(D,\mu)$ distance. See Figure 2.1 below.

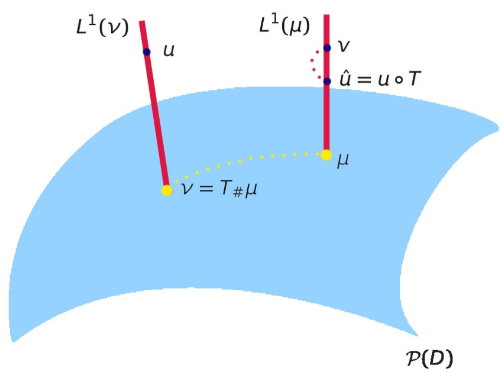


Figure 2.1: TL^1 space seen as a formal fiber bundle

In general a map $T : D \to D$ like the one described above does not exist, but we remark that the notion of *transportation map* is generalized by that of *transportation plan*. In this way one can make the previous discussion rigorous. We first introduce some notions and ideas from optimal transportation and then we introduce the so called TL^p-spaces. Their name comes from the fact that their metric can be regarded as extending the L^p convergence by using transportation maps/plans.

2.1 Background on transportation theory

Let D be an open 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 supported on D. Given $1 \leq p < \infty$, we define $\mathscr{P}_p(D)$ to be the set of Borel probability measures supported on D with finite p-moment, that is, $\mu \in \mathscr{P}(D)$ belongs to $\mathscr{P}_p(D)$ if

$$\int_D |x|^p \mathrm{d}\mu(x) < \infty. \tag{2.1}$$

The p-OT distance between $\mu, \tilde{\mu} \in \mathscr{P}_p(D)$ (denoted by $d_p(\mu, \tilde{\mu})$) is defined by:

$$d_p(\mu, \tilde{\mu}) := \min \left\{ \left(\int_{D \times D} |x-y|^p \mathrm{d}\pi(x,y) \right)^{1/p} \, : \, \pi \in \Gamma(\mu, \tilde{\mu}) \right\}, \tag{2.2}$$

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 deduced using the direct method of the calculus of variations (see [70]).

If D is a bounded open subset of $\mathbb{R}^d$ (as will be assumed often in the remainder) condition (2.1) is automatically satisfied for all $\mu \in \mathscr{P}(D)$ and from this we conclude that $\mathscr{P}_p(D) = \mathscr{P}(D)$. Moreover, the boundedness of D also implies that convergence in OT-metric is equivalent to weak convergence of probability measures. For details see for instance [5, 70] and the references therein. In particular, in that case, $\mu_n \xrightarrow{w} \mu$ (to be read μ_n converges weakly to μ) if and only if for any $1 \leq p < \infty$ there is a sequence of transportation plans between μ_n and μ, $\{\pi_n\}_{n \in \mathbb{N}}$, for which:

$$\lim_{n \to \infty} \iint_{D \times D} |x-y|^p \mathrm{d}\pi_n(x,y) = 0. \tag{2.3}$$

Boundedness of D implies that (2.3) is equivalent to $\lim_{n\to\infty} \iint_{D\times D} |x-y| d\pi_n(x,y) = 0$. We say that a sequence of transportation plans, $\{\pi_n\}_{n\in\mathbb{N}}$ (with $\pi_n \in \Gamma(\mu, \mu_n)$), is *stagnating* if it satisfies the condition (2.3). We remark that if 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:

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

From the previous definition, it follows that for any Borel function $\varphi : D \to \mathbb{R}$ the following change of variables in the integral holds:

$$\int_D \varphi(x)\,\mathrm{d}(T_\sharp\mu)(x) = \int_D \varphi(T(y))\,\mathrm{d}\mu(y). \tag{2.4}$$

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 $\tilde{\mu} = T_\sharp\mu$. We associate a transportation plan $\pi_T \in \Gamma(\mu,\tilde{\mu})$ to the transportation map T by:

$$\pi_T := (\mathrm{Id} \times T)_\sharp\mu, \tag{2.5}$$

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

$$\int_{D\times D} c(x,y)\mathrm{d}\pi_T(x,y) = \int_D c(x,T(x))\,\mathrm{d}\mu(x). \tag{2.6}$$

It is well known that when the measure $\mu \in \mathscr{P}_p(D)$ is absolutely continuous with respect to the Lebesgue measure, the problem on the right hand side of (2.2) is equivalent to:

$$\min\left\{\left(\int_D |x - T(x)|^p \mathrm{d}\mu(x)\right)^{1/p} \,:\, T_\sharp\mu = \tilde{\mu}\right\}, \tag{2.7}$$

and when p is strictly greater than 1, the problem (2.2) has a unique solution which is induced (via (2.5)) by a transportation map T solving (2.7) (see [70]). In particular we conclude that when D is bounded and the measure $\mu \in \mathscr{P}(D)$ is absolutely continuous with respect to the Lebesgue measure, then $\mu_n \xrightarrow{w} \mu$ 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:

$$\int_D |x - T_n(x)|^p \mathrm{d}\mu(x) \to 0, \text{ as } n \to \infty. \tag{2.8}$$

We say that a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ is *stagnating* if it satisfies (2.8).

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:

$$\pi^{-1} := s_\sharp\pi, \tag{2.9}$$

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

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

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

So far we have considered the p-OT distance for $1 \leq p < \infty$, we now consider the case $p = \infty$. Assume that D is a bounded open subset of $\mathbb{R}^d$. Then,

$$d_\infty(\mu, \tilde{\mu}) := \inf\{\operatorname{esssup}_\pi\{|x-y| \,:\, (x,y) \in D \times D\} \,:\, \pi \in \Gamma(\mu, \tilde{\mu})\}, \tag{2.10}$$

defines a metric on $\mathscr{P}(D)$, which is called the ∞-*transportation distance*. A natural question that arises from the connection between transportation maps and transportation plans when $1 \leq p < \infty$ is the following: in the definition of $d_\infty(\mu, \tilde{\mu})$ can we restrict our attention to couplings induced by transportation maps?. The answer to this question is affirmative in case the measure μ is absolutely continuous with respect to the Lebesgue measure. In fact, this is one of the results in [29], where it is proved that there exists solutions to the problem (2.10) which are ∞-*cyclically monotone*, that provided $\mu \ll \mathscr{L}^d$, happen to be induced by transportation maps. In the remainder we consider μ taken to be $\mathrm{d}\mu = \rho \mathrm{d}x$, where ρ is bounded above and below by positive constants and so in this setting the results in [29] can be stated as follows: if $\mu(D) = \tilde{\mu}(D)$, then there exists a transportation map $T^* : D \to D$ with $T^*_\sharp \mu = \tilde{\mu}$ and such that

$$d_\infty(\mu, \tilde{\mu}) = \|T^* - Id\|_{L^\infty(D)}. \tag{2.11}$$

The question of uniqueness of the optimal transportation map T^*, although interesting on its own, is not of importance for the results we present in the remainder. Nevertheless, it is worth mentioning that if $\tilde{\mu}$ is concentrated on finitely many points then, the transportation map T^* for which (2.11) holds is unique; this is the content of Theorem 5.4 in [29]. In particular, if $\tilde{\mu}$ is taken to be μ_n, where μ_n is the empirical measure associated to data points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ sampled from μ, then the uniqueness of the optimal transportation map is guaranteed.

We remark that for any transportation map T_n between μ and μ_n it holds that

$$d_\infty(\mu, \mu_n) \leq \|T_n - Id\|_{L^\infty(D)}.$$

Thus, we can estimate $d_\infty(\mu, \mu_n)$ by estimating the right hand side of the previous expression for some T_n.

2.2 The TL^p spaces and properties

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

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

Note that this definition is consistent with the definition given in the introduction given that boundedness of D implies $\mathscr{P}_p(D) = \mathscr{P}(D)$.

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

$$d_{TL^p}((\mu, f), (\nu, g)) = \inf_{\pi \in \Gamma(\mu,\nu)} \left(\iint_{D\times D} |x-y|^p + |f(x) - g(y)|^p \mathrm{d}\pi(x,y) \right)^{1/p}. \tag{2.12}$$

Here we should understand $L^p(D,\mu)$ as the set of equivalence classes of μ-measurable functions which are identified when they agree except on a set with μ-measure zero. Nevertheless, it is well known that for a μ-measurable function f, we can always find an equivalent function $\tilde{f}$ which is Borel measurable. In the remainder, we implicitly assume that when we take $f \in L^p(D,\mu)$ we are considering f as a Borel measurable function. In this way we can take compositions with other Borel functions without having measurability issues. The only potential issue that may arise from doing this is the ambiguity when choosing the Borel representative. However, in our context such issue will not arise as the next lemma shows.

Lemma 2.2.1. *Let $\mu, \nu \in \mathscr{P}(D)$ and suppose that $T : D \to D$ is a Borel map such that $T_\sharp \mu = \nu$. Let $f, g : D \to \mathbb{R}$ be two Borel measurable functions such that $f = g$ ν-a.e. Then, $f \circ T = g \circ T$, μ a.e.*

Proof. We may assume that $f = g$ except on a set $B \in \mathscr{B}(D)$ with $\nu(B) = 0$. Now, note that $0 = \nu(B) = \mu\left(T^{-1}(B)\right)$. Hence $f \circ T$ and $g \circ T$ are equal except on the set $T^{-1}(B)$ which has μ-measure zero. □

Remark 2.2.2. *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, 54] for $p = 2$) then TL^p is, formally, a fiber bundle. See Figure 3.1 for an illustration.*

In order to prove that d_{TL^p} is a metric, we remark that d_{TL^p} is actually equal to a transportation distance between graphs of functions. To make this idea precise 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

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

Remark 2.2.3. *We remark that $\boldsymbol{d}_p$ 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 2.1 (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 2.2.4. *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

$$\iint_{(D\times\mathbb{R})\times(D\times\mathbb{R})}|x-y|^p+|s-t|^p\mathrm{d}\pi((x,s),(y,t))=\iint_{D\times D}|x-y|^p+|f(x)-g(y)|^p\mathrm{d}\tilde{\pi}(x,y), \tag{2.13}$$

where $\tilde{\pi}\in\Gamma(\mu,\nu)$. The right hand side of the previous expression is greater than $d_{TL^p}((\mu,f),(\nu,g))$, which together with the fact that π was arbitrary allows us to conclude that $\mathbf{d}_p((\mu,f),(\nu,g))\geq d_{TL^p}((\mu,f),(\nu,g))$. 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 (2.13) 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. $\square$

Remark 2.2.5. *Proposition 2.2.4 and Remark 2.2.3 imply that (TL^p,d_{TL^p}) is a metric space.*

Remark 2.2.6. *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 $d_{TL^p}((\mu,f_n),(\mu,f_{n+1}))\leq 1/2^n$. 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 2.2.13 below). But then, by Remark 2.2.10, it would be true that $f_n\overset{L^p(D,\mu)}{\longrightarrow}f$. This is impossible because $\{f_n\}_{n\in\mathbb{N}}$ is not a convergent sequence in $L^p(D,\mu)$.*

Remark 2.2.7. *The completion of the metric space (TL^p,d_{TL^p}) is the space $(\mathscr{P}_p(\overline{D}\times\mathbb{R}),\boldsymbol{d}_p)$. In fact, in order to show this, it is enough to show that TL^p is dense in $(\mathscr{P}_p(\overline{D}\times\mathbb{R}),\boldsymbol{d}_p)$. Since the class of convex combinations of Dirac delta masses at points in D is dense in $(\mathscr{P}_p(\overline{D}\times\mathbb{R}),\boldsymbol{d}_p)$, 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 $x_1,\dots,x_n$ 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 $i=1,\dots,m$ 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 $(\forall i)$ $r_i^n\leq\frac{1}{n}$.

For $i = 1,\dots,m$ consider $y_1^{i,n},\dots,y_{l_i}^{i,n}$ 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 $(\mu_n, f_n) \xrightarrow{\boldsymbol{d}_p} \delta$.

Remark 2.2.8. *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*

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

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

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

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 [27, 55]). It is straightforward to check that $\mathscr{P}_p(D\times\mathbb{R})_{\llcorner\mu}$ is a closed subset (in the $\boldsymbol{d}_p$ 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 [27]. Here we present a simple proof of this fact using the ideas introduced so far. 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 2.2.7, there exists a sequence $\{(\mu_n, f_n)\}_{n\in\mathbb{N}} \subseteq TL^p$ such that

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

In particular,

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

Since μ is absolutely continuous with respect to the Lebesgue measure, it follows from the discussion in Section 2.1 that 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

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

On the other hand, the transportation map T_n induces the transportation plan $\pi_{T_n} \in \Gamma(\mu, \mu_n)$ defined in (2.5)*. Hence,*

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

From the previous computations, we deduce that $(\boldsymbol{d}_p((\mu, f_n \circ T_n), (\mu_n, f_n)) \to 0$ as $n \to \infty$, and thus $(\mu, f_n \circ T_n) \xrightarrow{\boldsymbol{d}_p} \gamma$. 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 2.2.9. *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 2.2.13.

Remark 2.2.10. *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 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 $L^p(D,\mu)$ converges in $L^p(D,\mu)$ 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 2.2.13.*

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

Lemma 2.2.11. *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 $u \in L^p(D,\mu)$*

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

Proof. We prove the case $p = 1$ since the other cases are similar. Let $u \in L^1(D,\mu)$ 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 $L^1(D,\mu)$. 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

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

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

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

This proves the result. □

Lemma 2.2.12. *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 $u_n \in L^p(D,\mu_n)$ and let $u \in L^p(D,\mu)$. Consider two sequences of stagnating transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ and $\{\hat{\pi}_n\}_{n\in\mathbb{N}}$ (with $\pi_n, \hat{\pi}_n \in \Gamma(\mu,\mu_n)$). Then:*

$$\lim_{n\to\infty} \iint_{D\times D} |u(x) - u_n(y)|^p \mathrm{d}\pi_n(x,y) = 0 \Leftrightarrow \lim_{n\to\infty} \iint_{D\times D} |u(x) - u_n(y)|^p \mathrm{d}\hat{\pi}_n(x,y) = 0 \quad (2.14)$$

Proof. We present the details for $p = 1$, as the other cases are similar. Take $\hat{\pi}_n^{-1} \in \Gamma(\mu_n,\mu)$ the inverse of $\hat{\pi}_n$ defined in (2.9). We can consider $\boldsymbol{\pi_n} \in \mathscr{P}(D\times D\times D)$ inducing a composition $\hat{\pi}_n^{-1} \circ \pi_n$ of the transportation plans π_n and $\hat{\pi}_n^{-1}$ (see Section 2.1). In particular $\hat{\pi}_n^{-1} \circ \pi_n \in \Gamma(\mu,\mu)$. From

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

and

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

and after using the triangle inequality, we deduce

$$\begin{aligned}\left| \iint_{D\times D} |u_n(y)-u(x)|\mathrm{d}\pi_n(x,y) - \iint_{D\times D} |u(z)-u_n(y)|\mathrm{d}\hat{\pi}_n(y,z)\right| \\ \leq \iiint_{D\times D\times D} |u(z)-u(x)|\mathrm{d}\boldsymbol{\pi}_n(x,y,z) = \iint_{D\times D} |u(z)-u(x)|\mathrm{d}\hat{\pi}_n^{-1}\circ\pi_n(x,z).\end{aligned} \tag{2.15}$$

Finally note that :

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

as $n\to\infty$. We conclude that the sequence $\left\{\hat{\pi}_n^{-1}\circ\pi_n\right\}_{n\in\mathbb{N}}$ is stagnating and thus from Lemma 2.2.11 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 (2.15) we conclude 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 2.2.13. *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* $\pi_n\in\Gamma(\mu,\mu_n)$*)*
$$\iint_{D\times D}|f(x)-f_n(y)|^p\,\mathrm{d}\pi_n(x,y)\to 0,\ as\, n\to\infty. \tag{2.16}$$
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* (2.16) *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*
$$\int_D|f(x)-f_n(T_n(x))|^p\,\mathrm{d}\mu(x)\to 0,\ as\, n\to\infty. \tag{2.17}$$
5. $\mu_n\xrightarrow{w}\mu$ *and for any stagnating sequence of transportation maps* $\{T_n\}_{n\in\mathbb{N}}$ *(with* $T_{n\sharp}\mu=\mu_n$*)* (2.17) *holds.*

Proof. By Lemma 2.2.12, statements 2. and 3. are equivalent. In case μ is absolutely continuous with respect to the Lebesgue measure, we know that there exists a stagnating sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ (with $T_{n\sharp}\mu = \mu_n$). Considering the sequence of transportation plans $\{\pi_{T_n}\}_{n\in\mathbb{N}}$ (as defined in (2.5)) and using (2.6) we see that 2., 3., 4., and 5. are all equivalent. We prove the equivalence of 1. and 3.

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

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

$\{\pi_n^*\}_{n\in\mathbb{N}}$ is then a stagnating sequence of transportation plans for which (2.16) holds.

(3. $\Rightarrow$ 1.) By hypothesis and using the fact that D is bounded, we can find a sequence of transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ with $\pi_n \in \Gamma(\mu,\mu_n)$ such that:

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

and

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

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

Definition 2.2.14. *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 $u_n\in L^p(D,\mu_n)$) converges in the TL^p sense to $u\in L^p(D,\mu)$, 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 $u_n\in L^p(D,\mu_n)$) is relatively compact in TL^p if the sequence $\{(\mu_n,u_n)\}_{n\in\mathbb{N}}$ is relatively compact in TL^p.*

Remark 2.2.15. *Thanks to Proposition 2.2.13 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 $u_n\circ T_n \xrightarrow{L^p(D,\mu)} u$ as $n\to\infty$ (this in particular implies the last part of Remark 2.2.10). 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 $L^p(D,\mu)$.*

In the light of Proposition 2.2.13 and Remark 2.2.8, 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 [27], 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 uniform distribution on D. The set $L^p(D,\mu)$ can be identified with the fiber $TL_{p \llcorner \mu}$ in a canonical way:

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

Thus, we can endow $L^p(D,\mu)$ with the distance d_{TL^p}. Note that by Remark 2.2.10, the topologies in $L^p(D,\mu)$ generated by d_{TL^p} and $||\cdot||_{L^p(D,\mu)}$ are the same. However, Remark 2.2.6 implies that d_{TL^p} and the distance generated by the norm $||\cdot||_{L^p(D,\mu)}$ are not equivalent. Note that the space $L^p(D,\mu)$ endowed with the norm $||\cdot||_{L^p(D,\mu)}$ is a complete metric space. On the other hand, by Remark 2.2.8, the completion of $L^p(D,\mu)$ 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 $L^p(D,\mu)$ endowed with the metric d_{TL^p}. Regarding the geometric interpretation of L^p-convergence, we have the following.

Corollary 2.2.16. *Let μ be the uniform distribution on D. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $L^p(D,\mu)$ and let $f \in L^p(D,\mu)$. Then, $\{f_n\}_{n\in\mathbb{N}}$ converges to f in $L^p(D,\mu)$ if and only if the graphs of f_n converge to the graph of f in the p-OT sense.*

Proof. From Remark 2.2.10, the sequence $\{f_n\}_{n\in\mathbb{N}}$ converges to f in $L^p(D,\mu)$ 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 2.2.4 and Remark 2.2.3). $\square$

Chapter 3

Rate of convergence of empirical measures in ∞-transportation distance

Let $\nu \in \mathscr{P}(D)$ be absolutely continuous with respect to the Lebesgue measure and let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d. samples from ν. We let ν_n be the empirical measure

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{x}_i}.$$

A classical result in probability theory is that ν_n converges weakly to ν as $n \to \infty$ with probability one. In particular, with probability one, for any sequence of functions $\{u_n\}_{n\in\mathbb{N}}$ with $u_n \in L^p(D, \nu_n)$ and for $u \in L^p(D, \nu)$, the statement $u_n \xrightarrow{TL^p} u$ is equivalent to the statement $u_n \circ T_n \xrightarrow{L^p(D,\nu)} u$ for all (or one) stagnating sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$. In this section we construct a special sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ with $T_{n\sharp}\nu = \nu_n$ that in addition of being stagnating satisfies the stronger condition $||Id - T_n||_{L^\infty(D)} \to 0$ as $n \to \infty$ with an explicit rate of convergence. Based on the discussion in Section 2.1 this is connected to the problem of determining how fast $d_\infty(\nu, \nu_n)$ approaches zero as $n \to \infty$. Convergence in TL^p is then equivalent to convergence in L^p after composing with these special transportation maps. In short, the main goal of this chapter is to establish Proposition 1.0.8, which can be rephrased as finding the rate of convergence of $d_\infty(\nu, \nu_n)$ as $n \to \infty$.

One of the main steps in the proof of Proposition 1.0.8 consists on establishing estimates on the ∞-transportation distance between two measures which are absolutely continuous with respect to the Lebesgue measure and whose densities are bounded from above and below by positive constants. We prove the following result which is of interest on its own.

Proposition 3.0.17. *Let $D \subset \mathbb{R}^d$ be a bounded, connected, open set with Lipschitz boundary. Let ν_1, ν_2 be measures on D of the same total mass: $\nu_1(D) = \nu_2(D)$. Assume the measures are absolutely continuous with respect to the Lebesgue measure and let ρ_1 and ρ_2 be their* *densities. Furthermore assume that for some $\lambda \geq 1$, for all $x \in D$*

$$\frac{1}{\lambda} \leq \rho_i(x) \leq \lambda \quad \text{for } i = 1,2. \tag{3.1}$$

Then, there exists a constant $C(\lambda,D)$ depending only on λ and D such that for all ν_1, ν_2 as above

$$d_\infty(\nu_1,\nu_2) \leq C(\lambda,D)\|\rho_1-\rho_2\|_{L^\infty(D)}.$$

We use this result in proving Proposition 1.0.8.

3.1 Background

Before studying the rate of convergence of $d_\infty(\nu,\nu_n)$, it is worth presenting the rates on the p-OT distance between ν and ν_n for $1 \leq p < \infty$. In fact, from the work of Dudley [35] it follows for $1 \leq p < \infty$ and $d \geq 3$ and under rather general conditions on ν (weaker than the ones assumed in Proposition 1.0.8) that the expected p-transportation distance between a measure ν and the empirical measure ν_n scales as $n^{-1/d}$, that is,

$$d_p(\nu,\nu_n) \sim n^{-1/d} \quad \text{for } d \geq 3.$$

Ajtai, Komlós, and Tusnády in [2] showed optimal bounds on the p-OT distance, for $1 \leq p < \infty$, between two empirical measures sampled from the Lebesgue measure on a square. That is they showed that if $\mathbf{x}_1,\dots,\mathbf{x}_n,\dots \in (0,1)^2$ and $\mathbf{y}_1,\dots,\mathbf{y}_n,\dots \in (0,1)^2$ are two independent samples and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{y}_i}$, while ν_n is as before, then the minimum over all permutations π of $\{1,\dots,n\}$ satisfies

$$\min_\pi \frac{1}{n}\sum_{i=1}^n |\mathbf{x}_{\pi(i)} - \mathbf{y}_i| \leq C\sqrt{\frac{\ln n}{n}}$$

with probability $1-o(1)$. They introduced the technique of obtaining probabilistic estimates by dyadically dividing the cube into 2^k subcubes, obtaining a matching estimate at the fine level and estimating the transformations needed to bridge different scales to obtain an upper bound on the total distance. The proof of Proposition 1.0.8 also relies on a similar decomposition of the domain. Dobrić and Yukich [34], Talagrand [64] and Talagrand, Yukich, [65], Bolley, Guillin, and Villani [17], Boissard [16], and others later refined these results and obtained more precise information on the distribution of the p-OT distance between a measure on a cube and the empirical measure.

For the ∞-transportation distance obtaining estimates is more delicate, since almost all of the mass needs to be matched within the desired distance to obtain the bound. Furthermore the optimal scaling itself has a logarithmic correction compared to the case $1 \leq p < \infty$. The optimal scaling in dimension $d = 2$, for ν being the Lebesgue measure, was obtained by Leighton and

Shor [49]. They consider i.i.d random samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ distributed according to the Lebesgue measure and points $\mathbf{y}_1, \dots, \mathbf{y}_n$ on a regular grid. They showed that there exist $c > 0$ and $C > 0$ such that with very high probability:

$$\frac{c(\ln n)^{3/4}}{n^{1/2}} \leq \min_{\pi} \max_{i} |\mathbf{x}_{\pi(i)} - \mathbf{y}_i| \leq \frac{C(\ln n)^{3/4}}{n^{1/2}}, \tag{3.2}$$

where π ranges over all permutations of $\{1, \dots, n\}$. In other words, when $d = 2$, with high probability the ∞-transportation distance between the measure μ_n and the measure ν_n is of order $\frac{(\ln n)^{3/4}}{n^{1/2}}$. We remark that the ∞-transportation distance in this context is also known as the *min-max matching distance*. Also, it is worth remarking that the discrete matching presented above implies the estimates on the distance between ν_n and ν when $D = (0,1)^2$ and ν is the Lebesgue measure.

For $d \geq 3$, Shor and Yukich [60] proved the analogous result on $(0,1)^d$ with ν being the Lebesgue measure. They showed that there exist $c > 0$ and $C > 0$ such that with very high probability:

$$\frac{c(\ln n)^{1/d}}{n^{1/d}} \leq \min_{\pi} \max_{i} |\mathbf{x}_{\pi(i)} - \mathbf{y}_i| \leq \frac{C(\ln n)^{1/d}}{n^{1/d}}. \tag{3.3}$$

The result in dimension $d \geq 3$ is based on the matching algorithm introduced by Ajtai, Komlós, and Tusnády in [2]. For $d = 2$ the AKT scheme still gives an upper bound, but not a sharp one. As remarked in [60], there is a crossover in the nature of the matching when $d = 2$: for $d \geq 3$, the matching length between the random points and the points in the grid is determined by the behavior of the points locally, for $d = 1$ on the other hand, the matching length is determined by the behavior of random points globally, and finally for $d = 2$ the matching length is determined by the behavior of the random points at all scales. At the level of the AKT scheme this means that for $d \geq 3$ the major source of the transportation distance is at the finest scale, for $d = 1$ at the coarsest scale, while for $d = 2$ distances at all scales are of the same size (in terms of how they scale with n). The sharp result in dimension $d = 2$ by Leighton and Shor required a more sophisticated matching procedure; an alternative proof was provided by Talagrand [64] who also provided more streamlined and conceptually clear proofs in [66, 67].

In this chapter we extend the previous results to general domains and general densities. It is worth remarking that the method used in [66, 67] to prove the matching results in dimension $d = 2$ are adaptable to general domains and general densities. On the other hand the method in [60], although it shares some similarities with the method we consider in this chapter, is not directly adaptable to the case of general domains and general densities. We use a dyadic decomposition similar to the one introduced by Ajtai, Komlós, and Tusnády (also used by Shor and Yukich). However the fact that we adjust the densities and not the geometry of the subdomains makes it easier to handle general densities.

3.2 The matching results for $(0,1)^d$

The first goal of this section is to prove Proposition 3.0.17 for $D=(0,1)^d$. In order to do this we need a few preliminary lemmas.

Lemma 3.2.1. *Let $Q\subseteq\mathbb{R}^d$ be a rectangular box (rectangular parallelepiped). Let Q_1,Q_2 be the rectangular boxes obtained from Q by bisecting one of its sides. Let $\rho: Q\to(0,\infty)$ be given by*

$$\rho(x) := \begin{cases} c_1, \text{ if } x\in Q_1,\\ c_2, \text{ if } x\in Q_2,\end{cases}$$

where $c_1,c_2>0$ are such that $1=\frac{c_1}{2}+\frac{c_2}{2}$. Denote by ν the measure with $\mathrm{d}\nu=\rho(x)\mathrm{d}x$ and let ν_0 be the Lebesgue measure restricted to Q. Then,

$$d_\infty(\nu_0,\nu)\le\frac{L}{2}\left|\frac{\nu(Q_1)}{\nu_0(Q_1)}-1\right|,$$

where L is the length of the side of Q bisected to generate Q_1 and Q_2.

Proof. Without the loss of generality we can assume that $Q=[0,L]\times\hat{Q}$ where $\hat{Q}$ is a $d-1$ dimensional rectangular box. Thus $Q_1=[0,\frac{L}{2}]\times\hat{Q}$ and $Q_2=[\frac{L}{2},L]\times\hat{Q}$. Note that the condition $1=\frac{c_1}{2}+\frac{c_2}{2}$ is equivalent to $\nu(Q)=\nu_0(Q)$. Let us introduce auxiliary functions $h(t)=c_1\mathbf{1}_{[0,\frac{L}{2}]}(t)+c_2\mathbf{1}_{(\frac{L}{2},L]}(t)$ and $f(t)=\mathbf{1}_{[0,L]}(t)$. For $t\in[0,L]$ let $F(t)=\int_0^t\mathrm{d}s=t$ and $H(t)=\int_0^t h(s)\mathrm{d}s$, that is,

$$H(t):=\begin{cases} c_1t & \text{if } 0\le t\le\frac{L}{2},\\ \frac{c_1}{2}L+c_2(t-\frac{L}{2}) & \text{if } \frac{L}{2}\le t\le L.\end{cases}$$

A direct computation shows that

$$H^{-1}\circ F(t):=\begin{cases} \frac{t}{c_1} & \text{if } 0\le t\le\frac{c_1L}{2},\\ \frac{t}{c_2}+\frac{L}{2}(1-\frac{c_1}{c_2}) & \text{if } \frac{c_1L}{2}\le t\le L.\end{cases}$$

Notice that the map $T_1:=H^{-1}\circ F$ is a transportation plan between the measures $\mathrm{d}t$ and $h(t)\mathrm{d}t$. Therefore, $T=T_1\times I_{d-1}$ is a transportation plan between ν_0 and ν.

A direct computation shows that

$$|T(x)-x|=|H^{-1}\circ F(x_1)-x_1|\le\frac{L}{2}|c_1-1|,$$

for all $x\in Q$. Since $c_1=\frac{\nu(Q_1)}{\nu_0(Q_1)}$, we conclude from the previous inequality that:

$$\|T-Id\|_{L^\infty(Q)}\le\frac{L}{2}\left|\frac{\nu(Q_1)}{\nu_0(Q_1)}-1\right|,$$

which implies the result. $\square$

Lemma 3.2.2. *Let $\rho : (0,1)^d \to (0,\infty)$ be integrable and let ν be the measure given by $\mathrm{d}\nu = \rho \mathrm{d}x$. Let $a = \int_{(0,1)^d} \rho(x)\mathrm{d}x$ and denote by ν_0 the measure on $(0,1)^d$ given by $\mathrm{d}\nu_0 = a\mathrm{d}x$. Then,*

$$d_\infty(\nu_0,\nu) \leq \frac{\overline{C}(d)}{a}\|a-\rho\|_{L^\infty((0,1)^d)},$$

where $\overline{C}(d)$ is a constant that depends on d only.

Proof. First, note that although the ∞-transportation distance was defined only for probability measures, it can be defined in the obvious way for finite measures with the same total mass. Moreover, since d_∞ measures the maximum transportation travelled when coupling two measures, the d_∞ distance does not change under rescaling of total mass. In particular, for ν_0 and ν as in the statement, it follows that

$$d_\infty(\nu_0,\nu) = d_\infty\left(\frac{1}{a}\nu_0, \frac{1}{a}\nu\right).$$

From the previous identity, we conclude that it is enough to prove the result for $a = 1$. Consider first the case that $\|1-\rho\|_{L^\infty((0,1)^d)} < 1/2$.

Step 1. For every $k \in \mathbb{N}$ we consider a partition of $[0,1]^d$ into a family $\mathscr{G}_k$ of 2^k rectangular boxes. The boxes are constructed recursively. Let $\mathscr{G}_0 = \{(0,1)^d\}$. Given the collection of boxes $\mathscr{G}_k$, the collection of rectangular boxes $\mathscr{G}_{k+1}$ is obtained by bisecting each of the rectangular boxes belonging to $\mathscr{G}_k$ through one of their longest sides. We note that all boxes in $\mathscr{G}_k$ have volume $\frac{1}{2^k}$ and have the same diameter (which depends only on k and d).

Consider $\rho_0 := 1$ and for all $k > 0$ and all $Q \in \mathscr{G}_k$ let:

$$\rho_k(x) := \frac{1}{\nu_0(Q)}\int_Q \rho(z)\mathrm{d}z = \frac{\nu(Q)}{\nu_0(Q)} \qquad \text{for all } x \in Q. \tag{3.4}$$

Let ν_k be the measure on $(0,1)^d$ with density ρ_k. The assumption $\|1-\rho\|_{L^\infty((0,1)^d)} < \frac{1}{2}$ implies $\frac{1}{2} \leq \rho \leq \frac{3}{2}$ and consequently for all k, $\frac{1}{2} \leq \rho_k \leq \frac{3}{2}$.

Note that for all $Q \in \mathscr{G}_k$ and all $j \geq k$, $\nu_j(Q) = \nu_k(Q) = \nu(Q)$. We denote by $\nu_{k \llcorner Q}$, the restriction of the measure ν_k to Q. The relation of ν to ν_k on Q is similar to the one of ν to ν_0 on $(0,1)^d$, only that the scale is smaller. We show that estimates on ∞-transportation distance on the finer scale lead to the desired estimates on the macroscopic scale. Note that

$$d_\infty(\nu_k,\nu_{k+1}) \leq \max_{Q\in\mathscr{G}_k} d_\infty(\nu_{k\llcorner Q}, \nu_{k+1\llcorner Q}), \tag{3.5}$$

and that

$$d_\infty(\nu_k,\nu) \leq \max_{Q\in\mathscr{G}_k} d_\infty(\nu_{k\llcorner Q}, \nu_{\llcorner Q}) \leq \max_{Q\in\mathscr{G}_k} \mathrm{diam}(Q) \leq \frac{C}{2^{k/d}}, \tag{3.6}$$

where C is a constant only depending on d.

Step 2. Let $Q \in \mathscr{G}_k$ and let $Q_1, Q_2 \in \mathscr{G}_{k+1}$ be the two sub-boxes of Q. Then, $\nu_k(Q_1) = \frac{1}{2}\nu_k(Q)$ and $\nu_0(Q_1) = \frac{1}{2}\nu_0(Q)$. It follows from (3.4) that

$$\begin{aligned}
|\nu(Q_1) - \nu_k(Q_1)| &\le |\nu(Q_1) - \nu_0(Q_1)| + |\nu_0(Q_1) - \nu_k(Q_1)| \\
&= \|\rho - 1\|_{L^\infty((0,1)^d)}\nu_0(Q_1) + \left|\frac{1}{2}\nu_0(Q) - \frac{1}{2}\nu_k(Q)\right| \\
&\le \|\rho - 1\|_{L^\infty((0,1)^d)}\nu_0(Q_1) + \frac{1}{2}\|\rho - 1\|_{L^\infty((0,1)^d)}\nu_0(Q) \\
&= 2\|\rho - 1\|_{L^\infty((0,1)^d)}\nu_0(Q_1).
\end{aligned}$$

Therefore,

$$\frac{|\nu(Q_1) - \nu_k(Q_1)|}{\nu_k(Q_1)} \le \frac{2\|\rho - 1\|_{L^\infty((0,1)^d)}\nu_0(Q_1)}{\nu_0(Q_1)/2} = 4\|\rho - 1\|_{L^\infty((0,1)^d)}. \tag{3.7}$$

Step 3. For a fixed cube $Q \in \mathscr{G}_k$, denote the value of ρ_k in Q by b. Then,

$$d_\infty(\nu_{k\llcorner Q}, \nu_{k+1\llcorner Q}) = d_\infty\left(\frac{1}{b}\nu_{k\llcorner Q}, \frac{1}{b}\nu_{k+1\llcorner Q}\right).$$

By Lemma 3.2.1 and by (3.7) we have

$$\begin{aligned}
d_\infty\left(\frac{1}{b}\nu_{k\llcorner Q}, \frac{1}{b}\nu_{k+1\llcorner Q}\right) &\le \frac{1}{2^{k/d}}\left|\frac{\nu(Q_1)}{\nu_k(Q_1)} - 1\right| \\
&\le \frac{4}{2^{k/d}}\|\rho - 1\|_{L^\infty((0,1)^d)}.
\end{aligned}$$

From (3.5) and the previous inequality it follows that for every $k \in \mathbb{N}$

$$d_\infty(\nu_k, \nu_{k+1}) \le \frac{4}{2^{k/d}}\|\rho - 1\|_{L^\infty((0,1)^d)}.$$

Choose $\tilde{k}$ such that $2^{-\tilde{k}/d} \le \|\rho - 1\|_{L^\infty}$. From the previous inequality and (3.6) we deduce that

$$\begin{aligned}
d_\infty(\nu_0, \nu) &\le \sum_{k=0}^{\tilde{k}-1} d_\infty(\nu_k, \nu_{k+1}) + d_\infty(\nu_{\tilde{k}}, \nu) \\
&\le 4\|\rho - 1\|_{L^\infty((0,1)^d)}\sum_{k=0}^{\tilde{k}-1}\frac{1}{2^{k/d}} + C\frac{1}{2^{\tilde{k}/d}} \\
&\le C(d)\|\rho - 1\|_{L^\infty((0,1)^d)},
\end{aligned} \tag{3.8}$$

which shows the desired result.

We now turn to case $\|\rho - 1\|_{L^\infty((0,1)^d)} \ge 1/2$. The desired estimate follows from

$$d_\infty(\nu_0, \nu) \le \operatorname{diam}((0,1)^d) = \sqrt{d} \le 2\sqrt{d}\|1 - \rho\|_{L^\infty((0,1)^d)}.$$

In conclusion taking the larger of the constants of the cases above, $\overline{C} = \max\{C(d), 2\sqrt{d}\}$, provides the desired estimate. $\square$

Proof of Proposition 3.0.17 for $D=(0,1)^d$. Suppose first that $\|\rho_1-\rho_2\|_{L^\infty((0,1)^d)} \le \frac{1}{2\lambda}$. Let $g(x)=\rho_1(x)-\rho_2(x)+\frac{1}{\lambda}$. Note that $g\ge 0$ and that

$$\rho_1 = \left(\rho_2-\frac{1}{\lambda}\right)+g,$$
$$\rho_2 = \left(\rho_2-\frac{1}{\lambda}\right)+\frac{1}{\lambda}.$$

By (2.11) and by Lemma 3.2.2, there exists a transportation map T between the measures $g\mathrm{d}x$ and $\frac{1}{\lambda}\mathrm{d}x$ such that

$$\|T-Id\|_{L^\infty((0,1)^d)} \le \lambda C(d)\left\|g-\frac{1}{\lambda}\right\|_{L^\infty((0,1)^d)} = \lambda C(d)\|\rho_1-\rho_2\|_{L^\infty((0,1)^d)}.$$

Note that

$$\gamma := (Id\times Id)_\sharp\left(\rho_2-\frac{1}{\lambda}\right)\mathrm{d}x+(Id\times T)_\sharp g\mathrm{d}x \in \Gamma(\nu_1,\nu_2).$$

Moreover for γ-a.e. $(x,y)\in(0,1)^d\times(0,1)^d$,

$$|x-y|\le \lambda C(d)\|\rho_1-\rho_2\|_{L^\infty((0,1)^d)}.$$

Thus,

$$d_\infty(\nu_1,\nu_2)\le \lambda C(d)\|\rho_1-\rho_2\|_{L^\infty((0,1)^d)}.$$

To get our estimate in case $\|\rho_1-\rho_2\|_{L^\infty} > \frac{1}{2\lambda}$ note that:

$$d_\infty(\nu_1,\nu_2)\le \mathrm{diam}((0,1)^d)=\sqrt{d}\le 2\lambda\sqrt{d}\|\rho_1-\rho_2\|_{L^\infty((0,1)^d)}.$$

□

Remark 3.2.3. *Note that from the previous proof, Proposition 3.0.17 is true for any domain* D *of the form* $D=(a_1,b_1)\times\cdots\times(a_d,b_d)$. *To deduce this fact, it is enough to consider a translation and rescaling of the coordinate axes to transform the rectangular box* D *into the unit box* $(0,1)^d$ *and then use Proposition 3.0.17 for the unit cube.*

3.2.1 The matching results for $(0,1)^d$: $d\ge 3$

Now we prove Proposition 1.0.8 for $D=(0,1)^d$ when $d\ge 3$. To achieve this it is useful to consider a partition of the cube $(0,1)^d$ into rectangular boxes analogous to the ones used in the proof of Lemma 3.2.2. The main difference is that we divide rectangular boxes into sub-boxes of the same ν-measure, instead of the same Lebesgue measure.

Let $\rho:(0,1)^d\to(0,\infty)$ be a density function satisfying $1/\lambda\le\rho\le\lambda$. For every $k\in\mathbb{N}$ we construct a family $\mathscr{F}_k$ of 2^k rectangular boxes which partition the cube $(0,1)^d$ with each rectangular box having ν-volume equal to $\frac{1}{2^k}$ and aspect ratio (ratio between its longest side and its shortest side) controlled in terms of λ. We let $\mathscr{F}_0 = \{(0,1)^d\}$. For $k=1$ we construct rectangular boxes Q_1 and Q_2 by bisecting one of the sides (say the one lying on the first coordinate) of the cube $(0,1)^d$ using the measure ν. That is, we define $Q_1 := (0,a)\times(0,1)^{d-1}$ and $Q_2 := [a,1)\times(0,1)^{d-1}$ where $a\in(0,1)$ is such that $\nu_{Q_1} = 1/2\nu(Q)$. Recursively, the collection of rectangular boxes at level $k+1$ is obtained by bisecting, according the measure ν, each rectangular box from level k through one of its **longest** sides.

Lemma 3.2.4. *The aspect ratio of every rectangular box in $\mathscr{F}_k$ is bounded by $2\lambda^2$.*

Proof. We show that for every $k\in\mathbb{N}$, every rectangular box in $\mathscr{F}_k$ has aspect ratio less than $2\lambda^2$. The proof is by induction on k.

Base Case: At level $k=1$ we consider $Q_1 = (0,a)\times(0,1)^{d-1}$, a chosen so that $\nu(Q_1) = 1/2$. Note that the aspect ratio of Q_1 is equal to $1/a$. Notice that,

$$\frac{1}{2} = \int_{Q_1}\rho(x)\mathrm{d}x \le a\lambda.$$

From this we conclude that the aspect ratio of Q_1 is no larger than 2λ and in particular no larger than $2\lambda^2$. By symmetry, the aspect ratio of Q_2 is no larger than $2\lambda^2$.

Inductive Step. Suppose that the aspect ratio of every rectangular box in $\mathscr{F}_k$ is bounded by $2\lambda^2$. Let Q be a rectangular box in $\mathscr{F}_{k+1}$. Note that Q is obtained by bisecting (using the measure ν) the longest side of a rectangular box $Q'\in\mathscr{F}_k$. Without the loss of generality we can assume that $\overline{Q'} = [a_1,b_1]\times[a_2,b_2]\times\cdots\times[a_d,b_d]$ and that $\overline{Q} = [a_1,c]\times[a_2,b_2]\times\cdots\times[a_d,b_d]$, where $a_1<c<b_1$. If (a_1,c) is not the smallest side of Q then the aspect ratio of Q is no greater than the aspect ratio of Q' and hence by the induction hypothesis is less than $2\lambda^2$. If on the other hand (a_1,c) is the smallest side of Q then we let (a_i,b_i) be the longest side of Q; the aspect ratio of Q is then equal to $\frac{b_i-a_i}{c-a_1}$. Since (a_1,b_1) is the longest side of $\tilde{Q}$, we have:

$$\frac{b_i-a_i}{c-a_1} = \frac{b_1-a_1}{c-a_1}\frac{b_i-a_i}{b_1-a_1} \le \frac{b_1-a_1}{c-a_1}.$$

Finally, since

$$\nu(Q) = \frac{1}{2}\nu(\tilde{Q}),$$

we deduce that

$$(c-a_1)\lambda \ge \frac{1}{2\lambda}(b_1-a_1).$$

This implies the desired result. $\square$

The proof of Proposition 1.0.8 requires estimating how many of the sampled points fall in certain rectangles. These estimates rely on two concentration inequalities for binomial random variables, which we now recall. Let $S_m \sim \mathrm{Bin}(m,p)$ be a binomial random variable, with m trials and probability of success for each trial of p. Chernoff's inequality [31] states that

$$\mathbb{P}\left(\left|\frac{S_m}{m}-p\right| \geq t\right) \leq 2\exp(-2mt^2). \tag{3.9}$$

Bernstein's inequality [15], which is sharper for small values of p gives that

$$\mathbb{P}\left(\left|\frac{S_m}{m}-p\right| \geq t\right) \leq 2\exp\left(-\frac{\frac{1}{2}m^2t^2}{mp(1-p)+\frac{1}{3}mt}\right). \tag{3.10}$$

Proof of Proposition 1.0.8 for $D=(0,1)^d$ when $d \geq 3$. *Step 1.* Let $\rho_0 := \rho$ and let $\mu_0 := \nu$. For every $Q \in \mathscr{F}_k$, consider

$$\rho_k(x) := \frac{\nu_n(Q)}{\nu(Q)}\rho(x) = \frac{\nu_n(Q)}{2^{-k}}\rho(x) \qquad \text{for all } x \in Q. \tag{3.11}$$

Let μ_k be the measure with density ρ_k. Note that for all $Q \in \mathscr{F}_k$, and all $j \geq k$, $\mu_j(Q) = \mu_k(Q) = \nu_n(Q)$. Since by construction $\nu(Q) = 2^{-k}$, $n\nu_n(Q)$ is a binomial random variable with n trials and probability of success for each trial of $p = 2^{-k}$. Fix $\alpha > 2$ and let

$$k_n := \log_2\left(\frac{n}{10\alpha \ln n}\right).$$

Consider $k \in \mathbb{N}$ with $k \leq k_n$. Using Bernstein's inequality (3.10) with $t = \frac{p}{2}$ we obtain

$$\begin{aligned}\mathbb{P}\left(\left|\nu_n(Q)-\frac{1}{2^k}\right| \geq \frac{1}{2^{k+1}}\right) &\leq 2\exp\left(-\frac{\frac{1}{2}\cdot\frac{1}{4}n^2p^2}{np(1-p)+\frac{1}{3}\cdot\frac{1}{2}np}\right)\\ &\leq 2\exp\left(-\frac{1}{10}np\right)\\ &\leq 2\exp\left(-\frac{1}{10}n\frac{10\alpha\ln n}{n}\right)\\ &= 2n^{-\alpha}.\end{aligned} \tag{3.12}$$

Since the probability of the union of events is less or equal to the sum of the probability of the events, we obtain

$$\mathbb{P}\left(\max_{Q\in\mathscr{F}_k}\left|\nu_n(Q)-\frac{1}{2^k}\right| \geq \frac{1}{2^{k+1}}\right) \leq 2^k 2n^{-\alpha}.$$

Summing over all $k \leq k_n$, we deduce that with probability at least $1-n^{-\alpha/2}$,

$$\frac{1}{2\lambda} \leq \rho_k \leq \frac{3\lambda}{2} \qquad \text{on } (0,1)^d, \tag{3.13}$$

for every $k \leq k_n$.

Let $Q \in \mathscr{F}_k$ and let $Q_1, Q_2 \in \mathscr{F}_{k+1}$ be the sub-boxes of Q. Let $m = n\nu_n(Q)$. Since $\nu(Q_1) = 2^{-(k+1)} = \frac{1}{2}\nu(Q)$ then, $m\frac{\nu_n(Q_1)}{\nu_n(Q)} \sim \mathrm{Bin}(m, \frac{1}{2})$ given $\nu_n(Q)$. Using Chernoff's bound (3.9) and (3.12), we deduce that

$$\mathbb{P}\left(\left|\frac{\nu_n(Q_1)}{\nu_n(Q)} - \frac{1}{2}\right| \geq \sqrt{\frac{\alpha 2^k \ln n}{n}}\right) \leq 4n^{-\alpha}.$$

Using the previous inequality, (3.11) and a union bound, we conclude that

$$\mathbb{P}\left(\sup_{x\in(0,1)^d}\left|\frac{\rho_{k+1}(x)}{\rho_k(x)} - 1\right| \geq 2\sqrt{\frac{\alpha 2^k \ln n}{n}}\right) \leq 2^k 4n^{-\alpha}.$$

Summing over all $k \leq k_n$, we deduce that with probability at least $1 - n^{-\alpha/2}$,

$$\sup_{x\in(0,1)^d}\left|\frac{\rho_{k+1}(x)}{\rho_k(x)} - 1\right| \leq 2\sqrt{\frac{\alpha 2^k \ln n}{n}} \tag{3.14}$$

for every $k \leq k_n$.

Notice that for all $Q \in \mathscr{F}_k$, and all $j \geq k$, $\mu_j(Q) = \mu_k(Q) = \nu_n(Q)$. Then,

$$d_\infty(\mu_k, \mu_{k+1}) \leq \max_{Q\in\mathscr{F}_k} d_\infty(\mu_k \llcorner_Q, \mu_{k+1}\llcorner_Q), \tag{3.15}$$

and

$$d_\infty(\mu_k, \nu_n) \leq \max_{Q\in\mathscr{F}_k} d_\infty(\mu_k\llcorner_Q, \nu_n\llcorner_Q) \leq \max_{Q\in\mathscr{F}_k} \mathrm{diam}(Q) \leq C(\lambda)\frac{1}{2^{k/d}}, \tag{3.16}$$

where $C(\lambda)$ is a constant only depending on λ; the last inequality in the previous expression obtained from Lemma 3.2.4 and from the fact that $\nu(Q) = 2^{-k}$.

Using estimates (3.13) and (3.14)

$$\|\rho_k - \rho_{k+1}\|_{L^\infty((0,1)^d)} \leq \|\rho_k\|_{L^\infty((0,1)^d)}\left\|\frac{\rho_{k+1}}{\rho_k} - 1\right\|_{L^\infty((0,1)^d)} \leq 2\lambda\left(\alpha 2^k \frac{\ln n}{n}\right)^{1/2},$$

with probability at least $1 - n^{-\alpha/2}$. Hence from Lemma 3.2.4 and Remark 3.2.3, we deduce that for all $Q \in \mathscr{F}_k$

$$d_\infty(\mu_k|_Q, \mu_{k+1}\llcorner_Q) \leq C(\lambda, d)\,\mathrm{diam}(Q)\left(\alpha 2^k \frac{\ln n}{n}\right)^{1/2} \leq C(\lambda, d)\frac{1}{2^{k/d}}\left(\alpha 2^k \frac{\ln n}{n}\right)^{1/2}.$$

Using (3.15) and the previous inequalities, we conclude that except on a set with probability $O(n^{-\alpha/2})$, for every $k = 0, \ldots, k_n$

$$d_\infty(\mu_k, \mu_{k+1}) \leq C\frac{1}{2^{k/d}}\left(2^k \frac{\ln n}{n}\right)^{1/2},$$

for some constant C depending only on λ, α and d. From the triangle inequality and (3.16), we obtain

$$
\begin{aligned}
d_\infty(\nu,\nu_n) &\le \sum_{k=1}^{k_n} d_\infty(\mu_{k-1},\mu_k) + d_\infty(\mu_{k_n},\nu_n) \\
&\le C\left(\sum_{k=1}^{k_n} \frac{1}{2^{k/d}}\left(\alpha 2^k \frac{\ln n}{n}\right)^{1/2} + \frac{(\ln n)^{1/d}}{n^{1/d}}\right) \\
&\le C\left(\left(\frac{\ln n}{n}\right)^{1/2} \sum_{k=1}^{k_n} 2^{k(1/2-1/d)} + \frac{(\ln n)^{1/d}}{n^{1/d}}\right).
\end{aligned}
$$

Given that $d \ge 3$, the term $\sum_{k=1}^{k_n} 2^{k(1/2-1/d)}$ is $O(1)$ and thus the previous expression is $O(\frac{(\ln n)^{1/d}}{n^{1/d}})$. In summary, except on a set with probability $O(n^{-\alpha/2})$

$$
d_\infty(\nu,\nu_n) \le C\frac{(\ln n)^{1/d}}{n^{1/d}},
$$

where C is a constant that depends on α, λ and d only. □

3.2.2 The matching results for $(0,1)^2$

Now we prove Proposition 1.0.8 for $D = (0,1)^2$. We actually state and prove a stronger result which is in agreement with the result by Talagrand in [66]. The improvement with respect to the statement of Proposition 1.0.8, has to do with the speed of decay of the tail probability of the transportation distance. Proposition 1.0.8 is an immediate consequence of the following.

Proposition 3.2.5. *Suppose that $\rho : (0,1)^2 \to (0,\infty)$ is a density function satisfying*

$$
\frac{1}{\lambda} \le \rho \le \lambda \tag{3.17}
$$

for some $\lambda > 1$. Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ be i.i.d samples from ρ and denote by ν_n the empirical measure

$$
\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{x}_i}.
$$

Then, there is a constant $L > 0$ depending only on λ, such that except on a set with probability $L\exp(-(\ln n)^{3/2}/L)$, we have

$$
d_\infty(\nu,\nu_n) \le L\frac{(\ln n)^{3/4}}{n^{1/2}}.
$$

In order to match the empirical measure ν_n with the measure ν, we consider a partition of $(0,1)^2$ into n rectangles $Q_1,\dots,Q_n$, each of which has ν-measure equal to $1/n$. We then look for a bijection between the set of points $\mathbf{x}_1,\dots,\mathbf{x}_n$ and the set $\{Q_1,\dots,Q_n\}$, in such a way that every data point is matched to a nearby rectangle. Note however, that in order to guarantee that all points within a rectangle are close to the corresponding data point we should be able to control the diameter of all the Q_is. This is is important since we want to obtain estimates on $d_\infty(\nu,\nu_n)$. With a slight modification to the construction preceding Lemma 3.2.4 we obtain the following.

Lemma 3.2.6. *Let $\rho:(0,1)^2\to(0,\infty)$ be a density function satisfying* (3.17)*, and let ν be the measure $\mathrm{d}\nu=\rho\mathrm{d}x$. Then, for any $n\in\mathbb{N}$ there exists a collection of rectangles $\{Q_i : i=1,\dots,n\}$ that partitions $[0,1]^2$, such that the aspect ratio of all rectangles is less than $3\lambda^2$ and their volume according to ν is $1/n$. In particular, for every Q_i*

$$\operatorname{diam}(Q_i)\le\frac{C(\lambda)}{\sqrt{n}},\tag{3.18}$$

where $C(\lambda)$ is a constant only depending on λ.

The task now is to show that with high probability we can indeed find a matching between the points $\mathbf{x}_1,\dots,\mathbf{x}_n$ and the rectangles $Q_1,\dots,Q_n$, in such a way that every point is close to its matched rectangle. When $\rho\equiv 1$, the previous statement is directly related to the result of Leighton and Shor [49]. The proof of Leighton and Shor depends on discrepancy estimates over all regions R formed by squares from a suitable regular grid G' defined on D. By discrepancy we mean the difference between $\nu(R)$ and $\nu_n(R)$ for a given region R. Obtaining a uniform bound on the discrepancy over all regions R can be interpreted as obtaining probabilistic estimates on the supremum of a stochastic process indexed by the mentioned class of regions R. A conceptually clear and efficient proof of this matching result, based on obtaining upper bounds of stochastic processes, was presented by Talagrand [66, 67]. In order to prove Proposition 1.0.8 we follow the framework of Talagrand and start by stating a general result on obtaining bounds on the supremum of more general stochastic processes (Section 1 in [66]).

Let (Y,d) be an arbitrary metric space. For $n\in\mathbb{N}$ define,

$$e_n(Y,d)=\inf\sup_{y\in Y}d(y,Y_n),$$

where the infimum is taken over all subsets Y_n of Y with cardinality less than 2^{2^n}. Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of partitions of Y. This sequence of partitions is called *admissible* if it is increasing (in the sense that for every n, A_{n+1} is a refinement of A_n) and it is such that the cardinality of A_n is no bigger than 2^{2^n}. For a given $y\in Y$ and $\{A_n\}_{n\in\mathbb{N}}$ admissible, $A_n(y)$ represents the unique set in A_n containing y. For an $\alpha>0$, consider

$$\gamma_\alpha(Y,d)=\inf\sup_{y\in Y}\sum_{n\ge 0}2^{n/\alpha}\operatorname{diam}(A_n(y)),$$

where $\operatorname{diam}(A_n(y))$ represents the diameter of the set $A_n(y)$ using the distance function d and where the infimum is taken over all $\{A_n\}_{n\in\mathbb{N}}$ admissible sequences of partitions of Y. With these definitions we can now state Theorem 1.2.9 in [66].

Lemma 3.2.7. *Let Y be a set and let d_1,d_2 be two distance functions defined on Y. Let $\{Z_y\}_{y\in Y}$ be a stochastic process satisfying: for all $y,y'\in Y$ and all $u>0$*

$$\mathbb{P}\left(|Z_y - Z_{y'}| \geq u\right) \leq 2\exp\left(-\min\left\{\frac{u^2}{d_2(y,y')^2}, \frac{u}{d_1(y,y')}\right\}\right), \tag{3.19}$$

and also $\mathbb{E}[Z_y]=0$ for all $y\in Y$. Then, there is a constant $L>0$ large enough, such that for all $u_1,u_2>0$

$$\mathbb{P}(\sup_{y\in Y}|Z_y - Z_{y_0}| \geq L(\gamma_1(Y,d_1)+\gamma_2(Y,d_2)) + u_1D_1 + u_2D_2) \leq L\exp(-\min\left\{u_2^2,u_1\right\}), \tag{3.20}$$

where $D_1 = 2\sum_{n\geq 0} e_n(Y,d_1)$ and $D_2 = 2\sum_{n\geq 0} e_n(Y,d_2)$.

One of the consequences of the previous lemma is the following: in order to prove a tail estimate of the supremum of the stochastic process $\{Z_y\}_{y\in Y}$, like the one in (3.20), one needs to do two things. First, estimate the quantities $\gamma_1(Y,d_1)$, $\gamma_2(Y,d_2)$, D_1 and D_2. Note that these quantities depend only on the distances d_1,d_2 and hence are not a priori related to the process $\{Z_y\}_{y\in Y}$. Secondly, relate the stochastic process $\{Z_y\}_{y\in Y}$ with the distances d_1,d_2 by establishing condition (3.19).

We are now ready to prove Proposition 3.2.5. As mentioned earlier, this result is an adaptation of the proof by Talagrand of Leighton and Shor theorem. We sketch some of the main steps in the proof by Talagrand and give the details on how to generalize it to non-constant densities.

Proof of Proposition 3.2.5 . In what follows $L>0$ is a constant that may increase from line to line.

Discrepancy estimates. Let l_1 be the largest integer such that $2^{-l_1} \geq \frac{(\ln n)^{3/4}}{\sqrt{n}}$. Consider G to be the regular grid of mesh 2^{-l_1} given by

$$G = \left\{(x_1,x_2)\in[0,1]^2\,;\, 2^{l_1}x_1\in\mathbb{N} \text{ or } 2^{l_1}x_2\in\mathbb{N}\right\}. \tag{3.21}$$

A *vertex* of the grid G is a point (x_1,x_2) in $[0,1]^2$ such that $2^{l_1}x_1\in\mathbb{N}$ and $2^{l_1}x_2\in\mathbb{N}$. A *square* of the grid G is a square of side length equal to 2^{-l_1} and whose edges belong to G. The edges are included in the squares.

For a given vertex w of G and a given integer k, consider $\mathscr{C}(w,k)$ the set of simple closed curves that lie on G which contain the vertex w and have length $l(C)\leq 2^k$. Note that every closed simple curve C in $\mathbb{R}^2$ divides the space into two regions, one of which is bounded. We call this set the interior of the curve C and we denote it by C°. For $C,C'\in\mathscr{C}(w,k)$ set $d_1(C,C')=1$ if $C\neq C'$ and $d_1(C,C')=0$ if $C=C'$. Also set $d_2(C,C')=\sqrt{n}\|\chi_{C^\circ}-\chi_{C'^\circ}\|_{L^2(D)}$.

Claim 1: For a given vertex w of G and a given integer k with $k \leq l_1+2$, there exists $L>0$ large enough such that with probability at least $1-L\exp(-(\ln n)^{3/2}/L)$

$$\sup_{C\in\mathscr{C}(w,k)} \Big|\sum_{i\leq n}(\chi_{C^\circ}(\mathbf{x}_i)-\nu(C^\circ))\Big| \leq L2^k\sqrt{n}(\ln n)^{3/4}. \tag{3.22}$$

To prove the claim, the idea is to study the supremum of the stochastic process $\{Z_C\}_{C\in\mathscr{C}(w,k)}$ where

$$Z_C := \frac{1}{L}\sum_{i\leq n}(\chi_{C^\circ}(\mathbf{x}_i)-\nu(C^\circ)).$$

For fixed $C, C' \in \mathscr{C}(w,k)$ one can write the difference $Z_C - Z_{C'}$ as

$$Z_C - Z_{C'} = \sum_{i\leq n} Z_i,$$

where $Z_i = \frac{1}{L}(\chi_{C^\circ}(\mathbf{x}_i)-\chi_{C'^\circ}(\mathbf{x}_i)-\nu(C^\circ)+\nu(C'^\circ))$. The random variables $\{Z_i\}_{i\leq n}$ are independent and identically distributed with mean zero, they satisfy $|Z_i| \leq \frac{2}{L}$ and furthermore, their variance σ^2 is bounded by

$$\sigma^2 \leq \frac{1}{L^2}\mathbb{E}\left[|\chi_{C^\circ}(\mathbf{x}_i)-\chi_{C'^\circ}(\mathbf{x}_i)|^2\right] \leq \frac{\lambda}{L^2}\|\chi_{C^\circ}-\chi_{C'^\circ}\|^2_{L^2(D)}.$$

Using Bernstein's inequality and choosing $L>0$ to be large enough, we obtain

$$\mathbb{P}(|Z_C - Z_{C'}| \geq u) \leq 2\exp\left(-\frac{u^2}{n\|\chi_{C^\circ}-\chi_{C'^\circ}\|^2_{L^2(D)}+u}\right) = 2\exp\left(-\min\left\{\frac{u^2}{d_2(C,C')^2}, \frac{u}{d_1(C,C')}\right\}\right).$$

In the proof of proposition 3.4.3 in [66], the estimates $\gamma_1(\mathscr{C}(w,k),d_1) \leq L2^k\sqrt{n}$, $\gamma_2(\mathscr{C}(w,k),d_2) \leq L2^k\sqrt{n}(\ln n)^{3/4}$, $D_1 \leq 2(k+l_1+1)$ and $D_2 \leq L2^{k+1}\sqrt{n}$ are established. Setting $u_1 = (\ln n)^{3/2}$ and $u_2 = (\ln n)^{3/4}$ one can use Lemma 3.2.7 (with $Y = \mathscr{C}(w,k)$, d_1, d_2 as above and $y_0 = \{w\}$) to prove the claim.

Considering all possible vertices w of G and all possible integers k with $-l_1 \leq k \leq l_1+2$. It is a direct consequence of Claim 1 above that with probability at least $1-L\exp(-(\ln n)^{3/2}/L)$,

$$\sup_C \Big|\sum_{i\leq n}(\chi_{C^\circ}(\mathbf{x}_i)-\nu(C^\circ))\Big| \leq Ll(C)\sqrt{n}(\ln n)^{3/4}, \tag{3.23}$$

where the supremum is taken over all C closed, simple curves on G. See the proof of Theorem 3.4.2 in [66]. We denote by Ω_n the event for which (3.23) holds.

Enlarging Regions. Consider an integer l_2 with $l_2 < l_1$. We consider G' the grid defined as in (3.21) but with mesh size 2^{-l_2}. Note that in particular $G' \subseteq G$. Let R be a union of squares of the grid G'. One can define R' to be the region formed by taking the union of all the squares in G' with at least one side contained in R. With no change in the proof of Theorem 3.4.1 in [66], it follows from the discrepancy estimates obtained previously that in the event Ω_n one has

$$\nu(R') \geq \nu_n(R) \tag{3.24}$$

for all regions R formed with squares from G', provided that $2^{-l_2} \geq \frac{2^6 L}{\sqrt{n}}(\ln n)^{3/4}$.

What this is saying is that given the discrepancy estimates obtained previously, in the event Ω_n, for any region R formed by taking the union of squares in G', one can enlarge R a bit to obtain a region R' in such a way that the area of the enlarged region R' according to ν is greater than the area of the original region R according to ν_n. It is worth remarking that the restriction to the number 2^{-l_2} (the mesh size of G'), for this to be possible, coincides with the scaling for the transportation cost we are after.

Matching between rectangles and data points. We choose l_2 to be the largest integer satisfying $2^{-l_2} \geq \frac{2^6 L}{\sqrt{n}}(\ln n)^{3/4}$. Consider $\{Q_1, \ldots, Q_n\}$ the rectangles constructed from Lemma 3.2.6. For $i \in \{1, \ldots, n\}$ let $B_i = \left\{ j \leq n : \operatorname{dist}(\mathbf{x}_i, Q_j) \leq 2\sqrt{2} \cdot 2^{-l_2} \right\}$.

Claim 2: In the event Ω_n, there is a bijection $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ with $\pi(i) \in B_i$ for all i.

By Hall's marriage Theorem, to prove this claim it is enough to prove that for every $I \subseteq \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the cardinality of $\cup_{i \in I} B_i$ is greater than the cardinality of I. Fix $I \subseteq \{1, \ldots, n\}$ and denote by R_I the region formed with the squares of G' that contain at least one of the points $\mathbf{x}_i$ with $i \in I$. Now, take $J = \left\{ j \leq n : Q_j \cap (R_I)' \neq \emptyset \right\}$, then, $J \subseteq \cup_{i \in I} B_i$. From the properties of the boxes Q_i and from (3.24) it follows that $\# \cup_{i \in I} B_i \geq \# J = n\nu(\cup_{j \in J} Q_j) \geq n\nu((R_I)') \geq \# I$. This proves the claim.

Finally, we construct a transportation map T_n between ν and ν_n. Indeed, for x in Q_i, set $T_n(x) = X_{\pi^{-1}(i)}$. From the properties of the boxes Q_i, it is straightforward to check that $T_{n\sharp}\nu = \nu_n$ and that $\|T_n - Id\|_{L^\infty(D)} \leq L\frac{(\ln n)^{3/4}}{\sqrt{n}}$ due to the estimate on the diameter of the rectangles Q_i in (3.18). □

3.3 The matching results for general D

The goal of this section is to prove the optimal bounds on matching for all open, connected, bounded domains D with Lipschitz boundary. In order to achieve this, we first prove Proposition 3.0.17 for general domains D. It is useful to consider first a class of domains D which are *well partitioned*.

Definition 3.3.1. *Let $D \subseteq \mathbb{R}^d$. We say that D satisfies the (WP) property with k polytopes if D is an open, bounded and connected set and is such that there exists a finite family of closed convex polytopes $\{A_i\}_{i=1}^k$ covering D and satisfying: For all $i, j = 1, \ldots, k$*

1. $\operatorname{int}(A_i) \cap D \neq \emptyset$.

2. *If $i \neq j$ then* $\operatorname{int}(A_i) \cap \operatorname{int}(A_j) = \emptyset$.

3. *$A_i \cap \overline{D}$ is bi-Lipschitz homeomorphic to a closed cube.*

The class of domains satisfying the (WP) property is convenient for our purposes for two reasons. The first one because as we see below, in order to prove the matching results for sets with the (WP) property, we can use induction on the number of polytopes. The second reason, has to do with the fact that the class of sets which are well partitioned contains the class of open, bounded, connected domains with smooth boundary. This is the content of the next proposition.

Proposition 3.3.2. *Let $D \subseteq \mathbb{R}^d$ be an open, bounded and connected domain with smooth boundary. Then, D satisfies the (WP) property with k polytopes for some $k \in \mathbb{N}$.*

Proof. Consider D to be a bounded open set with smooth boundary. For $\varepsilon > 0$ we denote by $\partial_\varepsilon D$ the set of points $x \in \mathbb{R}^d$ with $d(x,\partial D) \leq \varepsilon$. The fact that ∂D is a smooth compact manifold implies that there exists $0 < \varepsilon_0 < 1$ such that for every $x \in \partial_{\varepsilon_0} D$ there is a unique point $P(x)$ on ∂D closest to x. Furthermore the function $P : x \in \partial_{2\varepsilon_0} D \mapsto P(x)$ is smooth.

For a given $z \in \partial D$ we let $\vec{n}_z$ be the unit outer normal vector to ∂D at the point z. The fact that ∂D is a smooth manifold in $\mathbb{R}^d$ also implies that the outer unit normal vector field changes smoothly over ∂D.

We consider the signed distance function to ∂D, $g : \partial_{2\varepsilon_0} D \longrightarrow \mathbb{R}$

$$g(y) := \begin{cases} \operatorname{dist}(y,\partial D), & \text{if } y \in D^c \\ -\operatorname{dist}(y,\partial D), & \text{if } y \in D. \end{cases} \tag{3.25}$$

This function is smooth and its gradient is given by

$$\nabla g(y) = \vec{n}_{P(y)}. \tag{3.26}$$

We remark that for every $y \in \partial_{\varepsilon_0} D$, $g(y) = |y - P(y)|$ if $y \notin D$ and $g(y) = -|y - P(y)|$ if $y \in D$.

For a fixed $0 < \varepsilon < \varepsilon_0$ consider the family of open balls $\left\{B(x,\varepsilon^2)\right\}_{x\in\partial D}$. This is an open cover of the set ∂D which is compact. Hence, there exists a finite subcover $\left\{B(x_1,\varepsilon^2),\ldots,B(x_N,\varepsilon^2)\right\}$ of ∂D. To fix some notation, we let $\vec{n}_i$ be the vector $\vec{n}_{x_i}$ and we let T_i be the tangent plane to ∂D at the point x_i. Let $V_1,\ldots V_N$ be the Voronoi cells induced by the points $x_1,\ldots,x_N$; that is, we let V_i be the set

$$V_i := \left\{ y \in \mathbb{R}^d \,:\, |x_i - y| \leq |x_j - y|,\ \forall j \neq i \right\}.$$

Note that for every $t \in [-\varepsilon,\varepsilon]$ we have $P(x_i + t\vec{n}_i) = x_i$. In particular,

$$|x_i + t\vec{n}_i - x_i| < |x_i + t\vec{n}_i - x_j|, \tag{3.27}$$

for every $j \neq i$. Consider $\tilde{x}_i$ to be the point $\tilde{x}_i := -\frac{\varepsilon}{2}\vec{n}_i + x_i$ and let $T_i^+ := \varepsilon\vec{n}_i + T_i$, $T_i^- := \varepsilon\vec{n}_i + T_i$ be the planes parallel to T_i passing though the points $\varepsilon\vec{n}_i + x_i$ and $-\varepsilon\vec{n}_i + x_i$ respectively. We denote by S_i the closed strip delimited by the planes T_i^+ and T_i^- and let $A_i := V_i \cap S_i$. See Figure 3.1.

We first want to show that the region A_i is contained in a circular cylinder whose axis is the line passing through the point x_i with direction $\vec{n}_i$ and whose radius is small compared to ε. To achieve this, for a point $y \in \mathbb{R}^d$ denote by y_i the projection of y along the line passing through x_i with direction $\vec{n}_i$.

Claim 1: For all $0 < \varepsilon < \frac{\varepsilon_0}{2}$ small enough, $y \in A_i$ implies that $|y - y_i| \leq 4\varepsilon^{3/2}$.

To prove the claim suppose for the sake of contradiction that there is $y \in A_i$ with $|y - y_i| \geq 4\varepsilon^{3/2}$. Since $y \in S_i$, in particular $|y_i - x_i| = \mathrm{dist}(y_i, \partial D) \leq \varepsilon$. Consider a point $\tilde{y}$ in the segment $[y, y_i]$ such that $4\varepsilon^{3/2} \geq |\tilde{y} - y_i| \geq 3\varepsilon^{3/2}$. Then $|\tilde{y} - x_i| \leq |\tilde{y} - y_i| + |y_i - x_i| < 4\varepsilon^{3/2} + \varepsilon < 2\varepsilon$ if ε is small enough. Thus $|\tilde{y} - P(\tilde{y})| < 2\varepsilon$. Note also that $y \in A_i$ and $y_i \in A_i$ (from (3.27)). Since the set A_i is convex, we conclude that $\tilde{y} \in A_i$. To get to a contradiction we want to show that $|\tilde{y} - x_k| < |\tilde{y} - x_i|$ for some k; this would imply that $\tilde{y} \notin V_i$ which indeed would be a contradiction given that $\tilde{y} \in A_i$.

Note that $P(\tilde{y}) \in B(x_k, \varepsilon^2)$ for some k. Thus

$$\begin{aligned} |\tilde{y} - x_k|^2 &\leq (|\tilde{y} - P(\tilde{y})| + |P(\tilde{y}) - x_k|)^2 \\ &= |\tilde{y} - P(\tilde{y})|^2 + 2|\tilde{y} - P(\tilde{y})| \cdot |P(\tilde{y}) - x_k| + |P(\tilde{y}) - x_k|^2 \\ &\leq |\tilde{y} - P(\tilde{y})|^2 + 4\varepsilon^3 + \varepsilon^4. \end{aligned} \tag{3.28}$$

Furthermore, note that

$$\begin{aligned} |\tilde{y} - x_i|^2 &= |y_i - x_i|^2 + |\tilde{y} - y_i|^2 \\ &= g(y_i)^2 + |\tilde{y} - y_i|^2 \\ &= g(\tilde{y})^2 + g(y_i)^2 - g(\tilde{y})^2 + |\tilde{y} - y_i|^2 \\ &\geq |\tilde{y} - P(\tilde{y})|^2 - |g(y_i)^2 - g(\tilde{y})^2| + |\tilde{y} - y_i|^2. \end{aligned} \tag{3.29}$$

Since g is smooth in $\overline{\partial_{\varepsilon_0} D}$, there exists M such that $M \geq \|D^2 g(x)\|$ for all $x \in \overline{\partial_{\varepsilon_0} D}$. By (3.26), the gradient of the signed distance function g at the point y_i is equal to $\vec{n}_i$. Since $\tilde{y} - y_i$ is orthogonal to $\vec{n}_i$, by Taylor expansion $|g(\tilde{y}) - g(y_i)| = |g(\tilde{y}) - g(y_i) - Dg(y_i) \cdot (\tilde{y} - y_i)| \leq M|\tilde{y} - y_i|^2$. Thus $|g(\tilde{y})^2 - g(y_i)^2| = |g(\tilde{y}) - g(y_i)| \cdot |g(\tilde{y}) + g(y_i)| \leq 3M\varepsilon|\tilde{y} - y_i|^2$. Using (3.29) we deduce that

$$|\tilde{y} - x_i|^2 \geq |\tilde{y} - P(\tilde{y})|^2 + (1 - 3M\varepsilon)|\tilde{y} - y_i|^2.$$

Therefore for small enough $\varepsilon > 0$

$$|\tilde{y} - x_i|^2 \geq |\tilde{y} - P(\tilde{y})|^2 + 5\varepsilon^3.$$

Combining the previous inequality with (3.28) we deduce that $|\tilde{y} - x_i| > |\tilde{y} - x_k|$. This proves the claim.

Consider the circular cylinder whose axis is the line passing through the point x_i with direction $\vec{n}_i$ and whose radius is $4\varepsilon^{3/2}$. We let C_i^+ be the portion of the cylinder contained in S_i.

By (3.27) we can find a circular cylinder of smaller radius, whose axis is the same as that of C_i^+, but such that the portion of it contained in S_i, denoted by C_i^-, satisfies:

$$C_i^- \subseteq A_i \subseteq C_i^+.$$

See Figure 3.1.

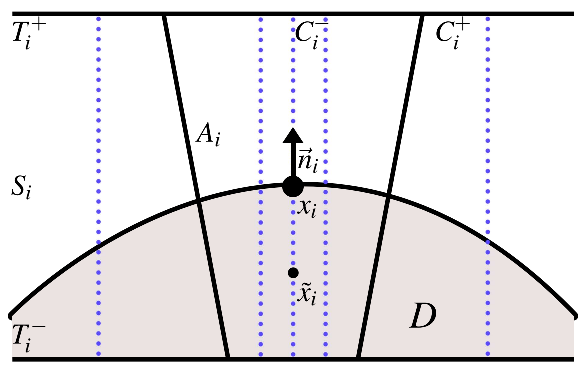


Figure 3.1: Close-up of a piece of the boundary of D

Claim 2. Let $0 < \varepsilon < \frac{\varepsilon_0}{2}$ be small enough. Then, there exists a map $\Phi_i : A_i \cap \overline{D} \to A_i$ which is a bi-Lipschitz homeomorphism. In particular, since A_i is a closed convex body with nonempty interior, we conclude that $A_i \cap \overline{D}$ is bi-Lipschitz homeomorphic to the unit cube.

To prove the claim fix $\varepsilon > 0$ so that in particular the conclusions from Claim 1 hold. From the bound on the second derivative of g and since the radius of C_i^+ is $4\varepsilon^{3/2}$, we deduce that there exists a universal constant $L > 0$ such that

$$|\vec{n}_z - \vec{n}_i| \leq L\varepsilon^{3/2}, \ \forall z \in \partial D \cap A_i, \tag{3.30}$$

due to the fact that $A_i \subseteq C_i^+$.

We now turn to constructing the bi-Lipschitz mapping between $\overline{D} \cap A_i$ and A_i. We do that by linear mappings along rays emanating from $\tilde{x}_i$. Consider $\mathscr{S}^{d-1}$ the set of all unit vectors in $\mathbb{R}^d$. For $\vec{n} \in \mathscr{S}^{d-1}$ define $s_{\vec{n}}$ and $t_{\vec{n}}$ by

$$s_{\vec{n}} := \sup\left\{s > 0 \, : \, \tilde{x}_i + s\vec{n} \in \overline{D} \cap A_i\right\},$$

$$t_{\vec{n}} := \sup\left\{t > 0 \, : \, \tilde{x}_i + t\vec{n} \in A_i\right\}.$$

Since $C_i^- \subseteq A_i \subseteq C_i^+$, we deduce that both functions $\vec{n} \in \mathscr{S}^{d-1} \mapsto s_{\vec{n}}$ and $\vec{n} \in \mathscr{S}^{d-1} \mapsto t_{\vec{n}}$ are bounded above and below by positive constants.

Now, note that for every $\vec{n} \in \mathscr{S}^{d-1}$, we have $s_{\vec{n}} \leq t_{\vec{n}}$. Moreover, by (3.30) and the fact that $A_i \subseteq C_i^+$, we deduce that if $s_{\vec{n}} < t_{\vec{n}}$ then

$$|\vec{n}_i - \vec{n}| \leq L\varepsilon^{3/2},$$

where L is a universal constant which is not necessarily the same as in (3.30). In particular, by choosing ε to be small enough we can assume that if $s_{\vec{n}} < t_{\vec{n}}$ then, the ray starting at $\tilde{x}_i$ with direction $\vec{n}$ only intersects $\partial D \cap A_i$ at one point. This fact, together with the smoothness of the outer normal vector field implies that the map $\vec{n} \in \mathscr{S}^{d-1} \mapsto s_{\vec{n}}$ is Lipschitz. On the other hand, since the set A_i is a convex set with piecewise smooth boundary (a convex polytope), we deduce that the function $\vec{n} \in \mathscr{S}^{d-1} \mapsto t_{\vec{n}}$ is Lipschitz as well.

Consider the map $\Phi_i : \overline{D} \cap A_i \to A_i$ defined as follows. Set $\Phi_i(\tilde{x}_i) = \tilde{x}_i$. For $x \in \overline{D} \cap A_i$, $x \neq \tilde{x}_i$ we can write $x = \tilde{x}_i + s\vec{n}$, for some $\vec{n} \in \mathscr{S}^{d-1}$ and for some $0 < s \leq s_{\vec{n}}$; we let $\Phi_i(x)$ be

$$\Phi_i(x) := \tilde{x}_i + \frac{s t_{\vec{n}}}{s_{\vec{n}}} \vec{n}.$$

Since both functions $\vec{n} \in \mathscr{S}^{d-1} \mapsto s_{\vec{n}}$ and $\vec{n} \in \mathscr{S}^{d-1} \mapsto t_{\vec{n}}$ are bounded above and below by positive constants and are Lipschitz, we deduce that the map Φ_i is a bi-Lipschitz homeomorphism between $\overline{D} \cap A_i$ and A_i. This proves the claim.

Claim 3. For any $\varepsilon < 1$ it holds that $\partial D \cap (V_i \setminus S_i) = \emptyset$. To prove this claim, assume for the sake of contradiction that there exists $x \in \partial D \cap (V_i \setminus S_i)$. Since $x \notin S_i$, it follows that $|x - x_i| \geq \varepsilon$. On the other hand, given that $x \in \partial D$, we know there exists k such that $x \in B(x_k, \varepsilon^2)$. Since $\varepsilon < 1$, we deduce that $|x - x_k| < |x - x_i|$ and thus $x \notin V_i$. This is a contradiction.

Now we have all the ingredients needed to prove Proposition 3.3.2. Indeed, take $\varepsilon > 0$ small enough so that all of the conclusions of all the previous claims hold. From Claim 3, we deduce that every V_i can be partitioned into three convex polytopes. One which intersects ∂D, namely $A_i = V_i \cap S_i$ and other two polytopes, one which is contained in $\mathrm{int}(D^c)$ and another one contained in D. We denote the later one by $\hat{A}_i$. We consider the family $\left\{A_1, \hat{A}_1, \ldots, A_N, \hat{A}_N\right\}$ of convex polytopes. This family covers D and is such that properties (1) and (2) from Definition 3.3.1 are satisfied. Moreover, given that $\hat{A}_i \subseteq D$ and given that $\hat{A}_i$ is convex, we deduce that $\hat{A}_i$ satisfies property (3) automatically, since all closed convex bodies with piecewise smooth boundary are bi-Lipschitz homeomorphic. Finally, Claim 2 implies that property (3) holds for each of the A_i. All together this implies that D satisfies the (WP) property. □

We now prove a lemma that prepares the ground for an inductive argument to be used in the proof of the matching results for domains with the (WP) property.

Lemma 3.3.3. *Suppose that D is a domain which satisfies hypothesis (WP) with k polytopes ($k > 1$). Let $\{A_i\}_{i=1}^k$ be associated polytopes. Then there exists j such that $D' := D \setminus A_j$ is connected.*

Proof. We say that $A_l \sim A_m$ if $\mathrm{relint}(\partial A_m) \cap \mathrm{relint}(\partial A_l) \cap D \neq \emptyset$, where $\mathrm{relint}(\partial A_i)$ is the union of the relative interiors of the facets of A_i ($(d-1)$-dimensional faces). This relation induces a graph $G = (V, E)$ where the set of nodes V is the set of polytopes A_i and where an edge between A_m and A_l ($m \neq l$) belongs to the graph if and only if $A_m \sim A_l$. We claim that G is a connected graph.

Indeed, consider $m \neq l$. We want to show that there exists a path in the graph G connecting A_m with A_l. For this purpose consider $x \in \operatorname{int}(A_m) \cap D$ and $y \in \operatorname{int}(A_l) \cap D$. Denote by C the union of all the ridges ($(d-2)$-dimensional faces) of all the polytopes A_i. Given that C is the union of finitely many $(d-2)$-dimensional objects in $\mathbb{R}^d$, we conclude that $D \setminus C$ is a connected open set and as such it is path connected. Since $x \in \operatorname{int}(A_m) \cap D$ and $y \in \operatorname{int}(A_l) \cap D$, in particular $x, y \in D \setminus C$ and so there exists a continuous function $\gamma : [0,1] \to D \setminus C$ such that $\gamma(0) = x$ and $\gamma(1) = y$. Let $A_{i_0}, A_{i_1}, \ldots, A_{i_N}$ be the polytopes visited by the path γ in order of appearance; this list satisfies $A_{i_s} \neq A_{i_{s+1}}$ for all s, $A_{i_0} = A_m$ and $A_{i_N} = A_l$. Now, note that for any given s, the path γ intersects $\partial A_{i_s} \cap \partial A_{i_{s+1}}$ at a point which belongs to the relative interior of a facet ($d-1$ dimensional face) of A_{i_s} and of $A_{i_{s+1}}$; this because γ lies in $D \setminus C$. From this fact we conclude that $A_{i_s} \sim A_{i_{s+1}}$ and hence there is a path in G connecting A_m and A_l. This proves that G is connected.

From the fact that G is connected, we deduce that it has a spanning tree G'. That is, there exists a subgraph G' of G which is a tree and includes all of the vertices of G. Let A_j be a leave of the spanning tree G'. It is now straightforward to show that A_j is the desired polytope from the statement. □

Remark 3.3.4. *Consider D and A_j as in the statement of Lemma 3.3.3. Then $D' := D \setminus A_j$ satisfies the property (WP) with $(k-1)$ polytopes and $D'' := D \cap A_j$ satisfies the property (WP) with one polytope.*

Let A_j be the polytope as in statement of Lemma 3.3.3. Note that there exists $i \neq j$ such that $\operatorname{relint}(\partial A_i) \cap \operatorname{relint}(\partial A_j) \cap D \neq \emptyset$; we denote this polytope by $\tilde{A}_j$. Let $\tilde{x} \in \operatorname{relint}(\partial \tilde{A}_j) \cap \operatorname{relint}(\partial A_j) \cap D$. Note that necessarily $F := \operatorname{relint}(\partial \tilde{A}_j) \cap \operatorname{relint}(\partial A_j)$ is contained in a hyperplane and hence we can consider e a unit vector which is orthogonal to F. Take $r > 0$ such that $B(\tilde{x}, r) \subseteq \operatorname{int}((\tilde{A}_j \cup A_j) \cap D)$. Let $z_1 := \tilde{x} + re$ and let $z_{-1} := \tilde{x} - re$. Without loss of generality we can assume that $z_1 \in \operatorname{int}(\tilde{A}_j)$. Denote by C_1 the set of points of the form $tz_1 + (1-t)y$ where $t \in [0,1]$ and $y \in B(\tilde{x}, r) \cap F$, similarly, denote by C_{-1} the set of points of the form $tz_{-1} + (1-t)y$ where $t \in [0,1]$ and where $y \in B(\tilde{x}, r) \cap F$. Let $z_{-1/2} := \tilde{x} - \frac{r}{2}e$ and consider the set $C_{-1/2}$ defined analogously to the way C_1 and C_{-1} are defined. We can think of C_1 and C_{-1} as *gates* connecting the sets $D' = D \setminus A_j$ and $D'' = D \cap A_j$. We illustrate the construction on Figures 3.2 and 3.3.

We claim that there is a function $\psi : D'' \cup C_1 \to D''$ which is a bi-Lipschitz homeomorphism. In fact, for a given point $y \in F \cap B(\tilde{x}, r)$ consider the line with direction e passing through the point y. This line intersects ∂C_1, at the points y and y_1, it intersects ∂C_{-1} at the points y and y_{-1} and finally it intersects $\partial C_{-1/2}$ at the points y and $y_{-1/2}$. We set $\psi(y_1) := y$, $\psi(y) := y_{-1/2}$ and $\psi(y_{-1}) := y_{-1}$. On the segments $[y_{-1}, y]$, $[y, y_1]$ we define ψ to be continuous and piecewise linear. In this way we define ψ for all points in $C_1 \cup C_{-1}$. Finally, set ψ to be the identity on $D'' \setminus C_{-1}$. It is straightforward to check that ψ constructed in this way is a bi-Lipschitz homeomorphism.

Now we are ready to prove Proposition 3.0.17 for general domains.

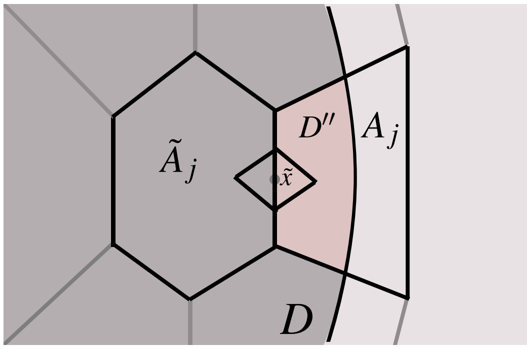


Figure 3.2: Polytope A_j with neighbor $\tilde{A}_j$.

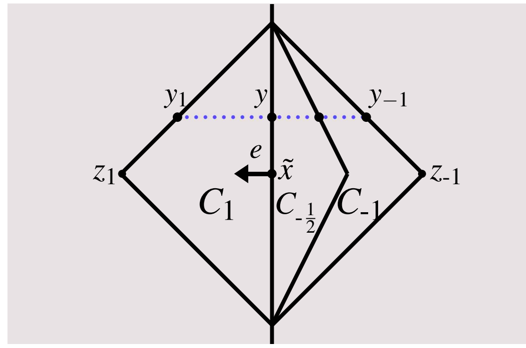


Figure 3.3: Gate, enlarged.

Proof of Proposition 3.0.17. Step 1: Instead of proving the result for domains as in the statement, we first prove the result for domains D satisfying the (WP) property. The proof is by induction on the number of polytopes k.

We remark that the constant $D(d,\lambda)$ may change (increase) from line to line in the proof.

Base case. Suppose $k = 1$. In this case there exists $\psi : \overline{D} \to [0,1]^d$ a bi-Lipschitz homeomorphism between $\overline{D}$ and the unit box. We use the map ψ to obtain measures $\tilde{\nu}_1, \tilde{\nu}_2$ on $(0,1)^d$ by setting

$$\tilde{\nu}_i := \psi_\sharp \nu_i \quad \text{for } i = 1,2.$$

Using the fact that ψ is bi-Lipschitz, we can use the change of variables formula to deduce that $\tilde{\nu}_1$ and $\tilde{\nu}_2$ are absolutely continuous with respect to the Lebesgue measure with densities

$$\tilde{\rho}_i(y) = \rho_i(\psi^{-1}(y))|\det(J\psi^{-1}(y))| \quad \text{for } i = 1,2.$$

Here, $J\psi^{-1}$ represents the Jacobian matrix of ψ^{-1}.

Using the fact that ψ is bi-Lipschitz, we deduce that

$$\frac{1}{\tilde{\lambda}} \leq \tilde{\rho}_1, \tilde{\rho}_2 \leq \tilde{\lambda}$$

where $\tilde{\lambda} = \max\{\mathrm{Lip}(\psi)^d, \mathrm{Lip}(\psi^{-1})^d\}$. By Proposition 3.0.17 applied to the unit cube,

$$d_\infty(\tilde{\nu}_1, \tilde{\nu}_2) \leq C(\tilde{\lambda}, d)\|\tilde{\rho}_1 - \tilde{\rho}_2\|_{L^\infty((0,1)^d)}.$$

Consequently,

$$d_\infty(\nu_1, \nu_2) \leq \mathrm{Lip}(\psi^{-1}) d_\infty(\tilde{\nu}_1, \tilde{\nu}_2) \leq C\|\tilde{\rho}_1 - \tilde{\rho}_2\|_{L^\infty((0,1)^d)} \leq C\|\rho_1 - \rho_2\|_{L^\infty(D)}.$$

for some constant C depending on λ and D only.

Inductive Step. Suppose that for any domain in $\mathbb{R}^d$ satisfying the (WP) property with $(k-1)$ polytopes the proposition is true. Let D be a domain satisfying the (WP) property with

k polytopes and let $\rho_1,\rho_2 : D \to (0,\infty)$ be functions as in the statement. By relabeling the functions if necessary, we can assume without loss of generality that $\int_{D'}\rho_1(x)\mathrm{d}x - \int_{D'}\rho_2(x)\mathrm{d}x \geq 0$, where D' is as in Remark 3.3.4. Since there is more mass in D' according to ν_1 than according to ν_2, we decide to transfer this excess of mass from the set D' to the set D''. To achieve this, we first move the excess of mass on D' to the gate C_1, so that we can subsequently move it to the set D''. In mathematical terms, we consider an intermediate distribution $\mathrm{d}\tilde{\nu}_1 = \tilde{\rho}_1\mathrm{d}x$ where

$$\tilde{\rho}_1(x) := \begin{cases} \rho_2(x) & \text{if } x \in D' \setminus C_1, \\ \beta\rho_1(x) & \text{if } x \in C_1, \\ \rho_1(x) & \text{if } x \in D'', \end{cases}$$

and where

$$\beta = \frac{\int_{D'}(\rho_1(x)-\rho_2(x))\mathrm{d}x + \int_{C_1}\rho_2(x)\mathrm{d}x}{\int_{C_1}\rho_1(x)\mathrm{d}x};$$

the idea is to compare ν_1 with $\tilde{\nu}_1$ and then compare $\tilde{\nu}_1$ with ν_2.

First, note that there is a $\lambda' > 1$ depending only on λ and D such that

$$\frac{1}{\lambda'} \leq \rho_1, \tilde{\rho}_1 \leq \lambda'.$$

Since by construction $\nu_1(D') = \tilde{\nu}_1(D')$, we use Remark 3.3.4 and the induction hypothesis to conclude that:

$$d_\infty(\nu_{1\llcorner D'}, \tilde{\nu}_{1\llcorner D'}) \leq C(\lambda',D')\|\rho_1 - \tilde{\rho}_1\|_{L^\infty(D')} = C(\lambda,D)\|\rho_1 - \tilde{\rho}_1\|_{L^\infty(D')},$$

where $\nu_{1\llcorner D'}$ denotes the measure ν_1 restricted to D' and $\tilde{\nu}_1|_{D'}$ the measure $\tilde{\nu}_1$ restricted to D'; notice that we can write $C(\lambda',D') = C(\lambda,D)$ because λ' depends on λ and D only. An immediate consequence of the previous estimate is that

$$d_\infty(\nu_1,\tilde{\nu}_1) \leq C(\lambda,D)\|\rho_1 - \tilde{\rho}_1\|_{L^\infty(D)}. \tag{3.31}$$

Given the definition of β, it is straightforward to show that

$$\|\rho_1 - \tilde{\rho}_1\|_{L^\infty(D)} \leq C(\lambda,D)\|\rho_1 - \rho_2\|_{L^\infty(D)}$$

for some constant $C(\lambda,D)$ only depending on D and λ. The previous inequality combined with (3.31) gives:

$$d_\infty(\nu_1,\tilde{\nu}_1) \leq C(\lambda,D)\|\rho_1 - \rho_2\|_{L^\infty(D)}.$$

Now we compare $\tilde{\nu}_1$ with ν_2. First of all note that $\tilde{\nu}_1(D_1'') = \nu_2(D_1'')$, where $D_1'' := D'' \cup C_1$. From the discussion proceeding Remark 3.3.4 we know that D_1'' is bi-Lipschitz homeomorphic to the set D'' which in turn is bi-Lipschitz homeomorphic to the unit box. Thus, D_1'' is bi-Lipschitz homeomorphic to the unit box and hence proceeding as in the base case, we conclude that

$$d_\infty(\tilde{\nu}_{1\llcorner D_1''}, \nu_{2\llcorner D_1''}) \leq C(\lambda,D)\|\tilde{\rho}_1 - \rho_2\|_{L^\infty(D_1'')}$$

and consequently

$$d_\infty(\tilde{\nu}_1,\nu_2) \leq C(\lambda,D)\|\tilde{\rho}_1-\rho_2\|_{L^\infty(D)}.$$

A straightforward computation shows that $\|\tilde{\rho}_1-\rho_2\|_{L^\infty(D)} \leq C(\lambda,D)\|\rho_1-\rho_2\|_{L^\infty(D)}$ and thus

$$d_\infty(\tilde{\nu}_1,\nu_2) \leq C(\lambda,D)\|\rho_1-\rho_2\|_{L^\infty(D)}.$$

Using the previous inequality, (3.31) and the triangle inequality we obtain the desired result.

Step 2: Now consider an open, connected bounded domain D with Lipschitz boundary. By Remark 5.3 in [12] there exists an open set $\tilde{D}$ with smooth boundary which is bi-Lipschitz homeomorphic to D. In particular $\tilde{D}$ is bounded and connected. By propositions 3.3.2 and Step 1, the result holds for $\tilde{D}$. Proceeding as in the base case in Step 1 and using the fact that D and $\tilde{D}$ are bi-Lipschitz homeomorphic we obtain the desired result. □

Now we are ready to prove Proposition 1.0.8.

Proof of Proposition 1.0.8 . Let us consider the function $\phi:\mathbb{N}\to(0,\infty)$, which is given by

$$\phi(n)=\begin{cases}\frac{(\ln n)^{1/d}}{n^{1/d}}, & \text{if } d\geq 3\\ \frac{(\ln n)^{3/4}}{n^{1/2}}, & \text{if } d=2.\end{cases} \tag{3.32}$$

Step 1. We first prove the result for domains D satisfying the (WP) property. The proof is by induction on k, the number of polytopes used in the definition of the property (WP). In what follows C may change from line to line, but always represents a constant that depends only on λ and D. Furthermore, since the probability that a sample point belongs to a boundary of one of the k polytopes is zero, we assume without the loss of generality that no sample point belongs to the boundary of any of the polytopes considered.

Base Case. Suppose that D is a domain satisfying property (WP) with one polytope. Then, $\overline{D}$ is bi-Lipschitz homeomorphic to the unit box. That is, there exists a bi-Lipschitz mapping $\psi:\overline{D}\to[0,1]^d$. Given a density $\rho:D\to(0,\infty)$ satisfying (1.10), we define measure $\tilde{\nu}$ on $(0,1)^d$ to be the push-forward of ν by ψ:

$$\tilde{\nu} := \psi_\sharp \nu.$$

Given the i.i.d. random points $\mathbf{x}_1,\ldots,\mathbf{x}_n$ on D distributed according to ν we note that

$$\tilde{X}_i=\psi(\mathbf{x}_i) \text{ for } i=1,\ldots,n$$

are i.i.d random points on $(0,1)^d$ distributed according to $\tilde{\nu}$.

As in the proof of Proposition 3.0.17 we use the fact that ψ is bi-Lipschitz to deduce that $\tilde{\nu}$ has a density $\tilde{\rho}$ satisfying

$$\frac{1}{\tilde{\lambda}}\leq\tilde{\rho}\leq\tilde{\lambda}$$

where $\tilde{\lambda} = \lambda \max\{\mathrm{Lip}(\psi)^d, \mathrm{Lip}(\psi^{-1})^d\}$. From Proposition 1.0.8 applied to the unit cube, we know that for $\alpha > 2$, except on a set with probability $O(n^{-\alpha/2})$,

$$d_\infty(\tilde{\nu},\tilde{\nu}_n) \le C\phi(n),$$

which implies

$$d_\infty(\nu,\nu_n) \le \mathrm{Lip}(\psi^{-1}) d_\infty(\tilde{\nu},\tilde{\nu}_n) \le C\phi(n).$$

where C only depends on λ, D and α.

Inductive Step. Suppose that the theorem is true for any domain in $\mathbb{R}^d$ satisfying the (WP) property with $k-1$ polytopes. Let D be a domain satisfying the (WP) property with k polytopes and let $\rho : D \to (0,\infty)$ be a density function satisfying (1.10). Consider $\tilde{\rho}_n : D \to D$ the density function given by

$$\tilde{\rho}_n(x) = \begin{cases} \frac{\nu_n(D')}{\nu(D')}\rho(x), & \text{if } x \in D' \\ \frac{\nu_n(D'')}{\nu(D'')}\rho(x), & \text{if } x \in D'', \end{cases} \tag{3.33}$$

where D' and D'' are as in Remark 3.3.4. Let $\tilde{\nu}_n$ be the measure $d\tilde{\nu}_n = \tilde{\rho}_n dx$ and note that $\nu_n(D') = \tilde{\nu}_n(D')$ and $\nu_n(D'') = \tilde{\nu}(D'')$. Also, notice that

$$\|\rho - \tilde{\rho}_n\|_{L^\infty(D)} \le C|\nu_n(D') - \nu(D')|, \tag{3.34}$$

for some constant C that depends only on λ and D.

To give some probabilistic estimates on $|\nu_n(D') - \nu(D')|$, we use Chernoff's inequality (3.9) to conclude that

$$\mathbb{P}\left(|\nu_n(D') - \nu(D')| > \sqrt{\frac{\alpha \ln n}{n}}\right) \le 2n^{-2\alpha}. \tag{3.35}$$

Denote by Ω_n the event in which $|\nu_n(D') - \nu(D)| \le \sqrt{\frac{\alpha \ln n}{n}}$. By (3.34) and Proposition 3.0.17 (from its proof, it holds for well partitioned domains), given Ω_n we have:

$$d_\infty(\nu,\tilde{\nu}_n) \le C\frac{(\ln n)^{1/2}}{n^{1/2}}. \tag{3.36}$$

We use the fact that $\nu_n(D') = \tilde{\nu}_n(D')$ and $\nu_n(D'') = \tilde{\nu}_n(D'')$ to estimate $d_\infty(\tilde{\nu}_n,\nu_n)$. Indeed, by the induction hypothesis, given the event Ω_n, with probability at least $1 - cn^{-\alpha/2}$

$$d_\infty(\tilde{\nu}_{n \llcorner D'}, \nu_{n \llcorner D'}) \le C\phi(n) \;\text{ and }\; d_\infty(\tilde{\nu}_{n \llcorner D''}, \nu_{n \llcorner D''}) \le C\phi(n).$$

In case the previous inequalities hold we conclude that

$$d_\infty(\tilde{\nu}_n,\nu_n) \le \max\left\{d_\infty(\tilde{\nu}_{n \llcorner D'}, \nu_{n \llcorner D'}), d_\infty(\tilde{\nu}_{n \llcorner D''}, \nu_{n \llcorner D''})\right\} \le C\phi(n).$$

Thus, given Ω_n, with probability at least $1-cn^{-\alpha/2}$

$$d_\infty(\tilde{\nu}_n, \nu_n) \leq C\phi(n).$$

From the previous discussion, (3.35) and (3.36) we conclude that with probability at least $1-cn^{-\alpha/2}$,

$$d_\infty(\nu, \nu_n) \leq C\phi(n) + C\frac{(\ln n)^{1/2}}{n^{1/2}} \leq C\phi(n).$$

Step 2. To prove the theorem for an arbitrary open, connected, bounded domain D with Lipschitz boundary it is enough to notice that by Remark 5.3 in [12] there exists an open set $\tilde{D}$ with smooth boundary which is bi-Lipschitz homeomorphic to D. In particular $\tilde{D}$ is bounded and connected. By Proposition 3.3.2 the result holds for $\tilde{D}$ by Step 1. Proceeding as in the base case in Step 1 and using the fact that D and $\tilde{D}$ are bi-Lipschitz homeomorphic we obtain the desired result. □

Chapter 4

Total variation in the continuum

In this chapter, the goal is to present the rigorous definition of weighted total variation and establish Proposition 1.0.7. We remark that when ρ is constant, the proof of the Γ-convergence part of Proposition 1.0.7 may be found in the Appendix of [4] in case D is a convex set, and in [58] for a general domain D satisfying the assumptions in the statement. In case ρ is not constant the results are obtained in a straightforward way by adapting the arguments presented in [58]. For the compactness statement of the proof new arguments were required, due to the presence of domain boundary and lack of L^∞-control. Part of the proof on compactness in [4] is used. As a corollary, we show that if one considers only functions uniformly bounded in $L^\infty(D)$, the compactness holds for open and bounded domains D regardless of the regularity of its boundary.

4.1 Weighted total variation

Let D be an open subset of $\mathbb{R}^d$ and let $\psi : D \to (0,\infty)$ be a continuous function. Consider the measure $\mathrm{d}\nu(x) = \psi(x)\mathrm{d}x$. As done in the previous chapters, we denote by $L^1(D,\nu)$ the L^1-space with respect to the measure ν and by $||\cdot||_{L^1(D,\nu)}$ its corresponding norm ; and we restrict the use $L^1(D)$ in the special case $\psi \equiv 1$ and in that case $||\cdot||_{L^1(D)}$ represents its corresponding norm. 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, [10], for $u \in L^1(D,\psi)$ define

$$TV(u;\psi;D) = \sup\left\{ \int_D u \,\mathrm{div}(\phi)\mathrm{d}x \,:\, (\forall x \in D)\ |\phi(x)| \leq \psi(x)\,,\ \phi \in C_c^\infty(D:\mathbb{R}^d) \right\} \tag{4.1}$$

the *weighted total variation of u in D with respect to the weight* ψ . When D is clear from the context, we write $TV(u;\psi)$ instead of $TV(u;\psi;D)$. 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 the dependence on ψ and write $BV(D)$ and $TV(u)$. Finally, for measurable subsets $E \subseteq D$, we define the weighted perimeter in D as the weighted total variation of the characteristic function of the set, i.e., $\mathrm{Per}(E;\psi) = TV(\mathbf{1}_E;\psi)$.

Throughout the paper we restrict our attention to the case where D is a bounded set and ψ is bounded from below and from above by positive constants. Occasionally we use D to be the $\mathbb{R}^d$ in which case we explicitly state that the functions we consider are defined on $\mathbb{R}^d$. Finally, in most of the remainder we consider $\psi = \rho^2$, where ρ is continuous and bounded below and above by positive constants, and occasionally we consider $\psi \equiv 1$ in which case as explained above we write $TV(\cdot)$ instead of $TV(\cdot;1)$.

Remark 4.1.1. *If 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 4.1.2. *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 [10].

Proposition 4.1.3. *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 $|Du|_\psi$-a.e. $x \in D$ and such that $\forall \phi \in C_c^\infty(D,\mathbb{R}^d)$*

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

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 to u. In case $\psi \equiv 1$, we denote $|Du|_\psi$ by $|Du|$ and we call it the *total variation measure* associated to 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

$$\mathrm{d}|Du|_\psi(x) = \psi(x)\mathrm{d}|Du|(x). \tag{4.2}$$

In particular,

$$TV(u;\psi) = \int_D \psi(x)\mathrm{d}|Du|(x). \tag{4.3}$$

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 $u_n \overset{L^1(D,\psi)}{\longrightarrow} u$ as $n\to\infty$, then

$$\liminf_{n\to\infty} TV(u_n;\psi) \geq TV(u;\psi). \tag{4.4}$$

We finish this section with the following approximation result.

Proposition 4.1.4. *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 $u_n \overset{L^1(D)}{\longrightarrow} u$ and $\int_D |\nabla u_n|\psi(x)\mathrm{d}x \to TV(u;\psi)$ as $n\to\infty$.*

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 [50] 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 $u_n \overset{L^1(D)}{\longrightarrow} u$ and with $\int_D |\nabla u - \nabla u_n|\psi(x)\mathrm{d}x \to 0$ as $n\to\infty$. Using a diagonal argument we conclude that in order to prove Proposition 4.1.4 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 $u_n \overset{L^1(D)}{\longrightarrow} u$ and with $\int_D |\nabla u_n|\psi(x)\mathrm{d}x \to TV(u;\psi)$ as $n\to\infty$.

Step 1: If ψ is Lipschitz this is precisely the content of Theorem 3.4 in [10].

Step 2 If ψ is not necessarily Lipschitz 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 $\psi_k \searrow \psi$. In fact, it is straightforward to verify that the functions

$$\psi_k(x) := \sup_{y\in D} \psi(y) - k|x-y|$$

satisfy the above conditions. 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 $u_{n,k} \overset{L^1(D)}{\longrightarrow} u$ and with $\int_D |\nabla u_{n,k}|\psi_k(x)\mathrm{d}x \to TV(u;\psi_k)$ as $n\to\infty$. By (4.3) and by the dominated convergence theorem we know that $TV(u;\psi_k) = \int_D \psi_k(x)\mathrm{d}|Du|(x) \to \int_D \psi(x)\mathrm{d}|Du|(x) = TV(u;\psi)$ 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, $u_{n,k_n} \overset{L^1(D)}{\longrightarrow} u$ and $\int_D |\nabla u_n|\psi_{k_n}(x)\mathrm{d}x \to TV(u;\psi)$ as $n\to\infty$. Taking $u_n := u_{n,k_n}$ and using the fact that $\psi \leq \psi_{k_n}$ we obtain:

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

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

4.2 Γ-convergence of non-local total variation $TV_\varepsilon(\cdot;\rho)$

In this section we prove the Γ-convergence of the nonlocal functionals $TV_\varepsilon(\cdot;\rho)$ to the weighted total variation $TV(\cdot;\rho^2)$. We recall that

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

where η satisfies conditions (K1)-(K3) in the introduction. We adopt the following notation: ε is a short-hand notation for ε_k where $\{\varepsilon_k\}_{k\in\mathbb{N}}$ is an arbitrary sequence of positive real numbers converging to zero as $k\to\infty$. Limits as $\varepsilon\to 0$ simply mean limits as $k\to\infty$ for every such sequence. The next lemma follows ideas present in [58, 18].

Lemma 4.2.1. *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*

$$\sup_{\varepsilon>0}\left\{||\nabla u_\varepsilon||_{L^\infty(\mathbb{R}^d)} + ||D^2u_\varepsilon||_{L^\infty(\mathbb{R}^d)}\right\} < \infty. \tag{4.5}$$

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

$$\lim_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho) = \sigma_\eta\int_D |\nabla u(x)|\rho^2(x)\mathrm{d}x. \tag{4.6}$$

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

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

First we show that

$$\lim_{\varepsilon\to 0}|TV_\varepsilon(u_\varepsilon;\rho) - H_\varepsilon(u_\varepsilon)| = 0. \tag{4.7}$$

For this purpose, note that by Taylor's theorem and by (4.5), for $x,y\in D$ $x\neq y$ and $\varepsilon>0$

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

where $||D^2 u_\varepsilon||_{L^\infty(\mathbb{R}^d)}$ 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

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

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 (4.7).

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

$$\tilde{H}_\varepsilon(w)=\frac{1}{\varepsilon}\int_D\int_{x+h\in D}\eta_\varepsilon(h)\,|\nabla w(x)\cdot h|\,\rho^2(x)\mathrm{d}h\mathrm{d}x. \tag{4.8}$$

We claim that

$$\lim_{\varepsilon\to 0}\left|H_\varepsilon(u_\varepsilon)-\tilde{H}_\varepsilon(u_\varepsilon)\right|=0. \tag{4.9}$$

Indeed, using the fact that ρ is Lipschitz,

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

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

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_D\int_{x+h\in\mathbb{R}^d\setminus D}\eta_\varepsilon(h)\,|\nabla u_\varepsilon(x)\cdot h|\,\rho^2(x)\mathrm{d}h\mathrm{d}x=0. \tag{4.10}$$

Note that,

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

Using (4.5) 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 (4.10).

Step 4: Using steps 1, 2, and 3 in order to obtain (4.6) it is enough to prove that

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_D\int_{\mathbb{R}^d}\eta_\varepsilon(h)|\nabla u_\varepsilon(x)\cdot h|\rho^2(x)\mathrm{d}h\mathrm{d}x=\sigma_\eta\int_D|\nabla u|\rho^2(x)\mathrm{d}x. \tag{4.11}$$

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

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

Taking ε to zero in the previous expression we obtain (4.11), and consequently (4.6). $\square$

Proof of Proposition 1.0.7. Let us first start showing the Liminf inequality.

Case 1: ρ is Lipschitz. Consider an arbitrary $u\in L^1(D,\rho)$ and suppose that $u_\varepsilon\overset{L^1(D,\rho)}{\longrightarrow}u$ as $\varepsilon\to 0$. Recall that given the assumptions on ρ this is equivalent to $u_\varepsilon\overset{L^1(D)}{\longrightarrow}u$ as $\varepsilon\to 0$. We want to show that $\liminf_{\varepsilon\to 0}TV_\varepsilon(u_\varepsilon;\rho)\geq\sigma_\eta TV(u;\rho^2)$. 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 4.2.1 (see [58,18]). The plan is to first regularize the functions u_ε to obtain a new sequence of functions $\left\{u_{\varepsilon,\delta}\right\}_{\varepsilon>0}$ ($\delta>0$ 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 4.2.1.

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 $\int_{\mathbb{R}^d}J(z)dz=1$. We set J_δ to be $J_\delta(z)=\frac{1}{\delta^d}J\left(\frac{z}{\delta}\right)$. Note that $\int_{\mathbb{R}^d}J_\delta(z)dz=1$ for every $\delta>0$.

Fix D' an open domain 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 $w\in L^1(D)$ we define the mollified function $w_\delta\in L^1(\mathbb{R}^d)$ by setting $w_\delta(x)=\int_{\mathbb{R}^d}J_\delta(x-z)w(z)\mathrm{d}z=\int_{\mathbb{R}^d}J_\delta(z)w(x-z)\mathrm{d}z$ where we have extended w to be zero outside of D. The functions w_δ are smooth, and satisfy $w_\delta\overset{L^1(D')}{\longrightarrow}w$ as $\delta\to 0$, see for example [50]. Furthermore

$$\nabla w_\delta(x)=\int_{\mathbb{R}^d}\nabla J_\delta(z)w(x-z)\mathrm{d}z=\frac{1}{\delta}\int_{\mathbb{R}^d}\frac{1}{\delta^d}\nabla J\left(\frac{z}{\delta}\right)w(x-z)\mathrm{d}z. \tag{4.12}$$

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

$$||\nabla w_\delta||_{L^\infty(\mathbb{R}^d)}\leq\frac{C}{\delta}||w||_{L^1(D)}\quad\text{and}\quad||D^2w_\delta||_{L^\infty(\mathbb{R}^d)}\leq\frac{C}{\delta^2}||w||_{L^1(D)}. \tag{4.13}$$

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 $w = u_\varepsilon$ in inequalities (4.13) 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 (4.12) to express $\nabla u_{\varepsilon,\delta}$ and ∇u_δ, it is straightforward to deduce that

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

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 4.2.1 (taking D to be D') to infer that

$$\begin{aligned}\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{D'}\int_{D'} \eta_\varepsilon(x-y)|u_{\varepsilon,\delta}(x) - u_{\varepsilon,\delta}(y)|\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y \\ = \sigma_\eta \int_{D'} |\nabla u_\delta(x)|\rho^2(x)\mathrm{d}x\mathrm{d}y.\end{aligned} \tag{4.14}$$

To measure the approximation error in the energy, we set

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

and estimate

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

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 (4.14) imply that

$$\liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho) \geq \liminf_{\varepsilon\to 0} a_{\varepsilon,\delta} + \sigma_\eta\int_{D'} |\nabla u_\delta(x)|(\rho(x))^2\mathrm{d}x. \tag{4.15}$$

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

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

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 (4.15),

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

Given that $u_\delta \xrightarrow{L^1(D')} u$ as $\delta\to 0$, we can use the lower semicontinuity of the weighted total variation, (4.4), to obtain

$$
\liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho) \geq \sigma_\eta \liminf_{\delta\to 0}\int_{D'}|\nabla u_\delta|(\rho(x))^2\mathrm{d}x \geq \sigma_\eta |Du|_{\rho^2}(D'). \tag{4.16}
$$

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

$$
\rho_k(x) := \inf_{y\in D}\rho(y)+k|x-y|. \tag{4.17}
$$

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 $\rho_k\leq\rho$ to conclude that

$$
\liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho) \geq \liminf_{\varepsilon\to 0} TV_\varepsilon(u_\varepsilon;\rho_k) \geq \sigma_\eta TV(u;\rho_k^2). \tag{4.18}
$$

Using (4.3) 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 (4.18) yields the desired result.

Now we turn our attention to the Limsup inequality.

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 $u \in L^1(D,\rho)$:

$$\limsup_{\varepsilon\to 0} TV_\varepsilon(u;\rho) \le \sigma_\eta TV(u;\rho^2). \tag{4.19}$$

It suffices to show (4.19) for functions $u \in BV(D)$ (if the right hand side of (4.19) 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 [6] to obtain an extension $\hat{u} \in BV(\mathbb{R}^d)$ of u to the entire space $\mathbb{R}^d$ with $|D\hat{u}|(\partial D) = 0$. In particular from (4.2) we obtain

$$|D\hat{u}|_{\rho^2}(\partial D) = 0. \tag{4.20}$$

We split the proof of (4.19) in two cases:

Step 1: Suppose that η has compact support, i.e. assume there is $\alpha > 0$ such that if $|h| \ge \alpha$ 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 [10] and our assumptions on ρ provide a sequence of functions $\{w_k\}_{k\in\mathbb{N}} \in C^\infty(D_\varepsilon) \cap BV(D_\varepsilon)$ such that as $k \to \infty$

$$w_k \xrightarrow{L^1(D_\varepsilon)} \hat{u} \quad \text{and} \quad \int_{D_\varepsilon} |\nabla w_k(x)|\rho^2(x)\mathrm{d}x \to |D\hat{u}|_{\rho^2}(D_\varepsilon). \tag{4.21}$$

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

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

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

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

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

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

Using (4.21) in particular we obtain that $v_k \xrightarrow{L^1(D)} u$ as $k\to\infty$. This together with continuity of $TV_\varepsilon(\cdot;\rho)$ with respect to L^1-convergence implies that $TV_\varepsilon(w_k;\rho)\to TV_\varepsilon(u;\rho)$ as $k\to\infty$. Therefore, from the previous chain of inequalities and from (4.21) we conclude that

$$TV_\varepsilon(u;\rho)\le\sigma_\eta|D\hat{u}|_{\rho^2}(D_\varepsilon)+\limsup_{k\to\infty}a_{\varepsilon,k}\le\sigma_\eta|D\hat{u}|_{\rho^2}(D_\varepsilon)+C\varepsilon|D\hat{u}|_{\rho^2}(D_\varepsilon). \tag{4.22}$$

Using (4.20), 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 (4.22) implies the desired estimate, (4.19).

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:

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

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 [50]. By Step 1 we conclude that:

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

Taking α to infinity and using condition (K3) on $\boldsymbol{\eta}$ (see introduction) implies (4.19).

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

$$\rho_k(x) := \sup_{y\in D}\rho(y) - k|x-y|. \tag{4.23}$$

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:

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

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

$$\limsup_{\varepsilon\to 0} TV_\varepsilon(u;\rho) \le \limsup_{\varepsilon\to 0} TV_\varepsilon(u;\rho_k) \le \sigma_\eta TV(u;\rho_k^2). \tag{4.24}$$

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 (4.24) provides the desired result. □

Remark 4.2.2. *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 $u\in L^1(D,\rho)$:*

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

Compactness

The only remaining point left to be proved from Proposition 1.0.7 is compactness. We first establish it for regular domains D and then extend it to more general ones.

Lemma 4.2.3. *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

$$\sup_{\varepsilon>0} TV_\varepsilon(v_\varepsilon;\rho) < \infty. \tag{4.25}$$

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, (4.25) 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 $d(x,\partial D)\le\delta$ there exists a unique closest point on ∂D. For all $x\in U := \{x\in\mathbb{R}^d : d(x,D)<\delta\}$ let Px be the closest point to x in $\overline{D}$. We define the local reflection mapping from U to $\overline{D}$ by $\hat{x} = 2Px - x$. Let ξ be a smooth cut-off function such that $\xi(s)=1$ if $s\le\delta/8$ and $\xi(s)=0$ if $s\ge\delta/4$. We define an auxiliary function $\hat{v}_\varepsilon$ 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:

$$\sup_{\varepsilon>0}\frac{1}{\varepsilon}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\eta_\varepsilon(x-y)|\tilde{v}_\varepsilon(x)-\tilde{v}_\varepsilon(y)|\mathrm{d}x\mathrm{d}y<\infty. \tag{4.26}$$

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 $y\in D$

$$|\hat{x}-y| < 2|x-y|. \tag{4.27}$$

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 $x,y\in W$

$$\frac{1}{4}|x-y| < |\hat{x}-\hat{y}| < 4|x-y|. \tag{4.28}$$

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

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

Squaring and straightforward algebra yield

$$|z-Px|^2 \ge 2\delta(z-Px)\cdot\frac{x-Px}{|x-Px|}. \tag{4.29}$$

For $x \in W$ and $y \in D$, using (7.20) we obtain

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

Therefore $|y-\hat{x}|^2 \leq 3|y-x|^2$, which establishes (4.27).

For distinct $x, y \in W$ using (7.20), with $z = Py$ and with $z = Px$, follows

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

Therefore

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

Since the roles on x, y and $\hat{x}, \hat{y}$ can be reversed it follows that $|x-y| \leq 4|\hat{x}-\hat{y}|$. These estimates establish (4.28).

We now return to proving (4.26). For ε small enough,

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

where the first inequality follows from (4.27) and the second follows from the fact that the change of variables $x \mapsto \hat{x}$ is bi-Lipschitz as shown in (4.28). Also,

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

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

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

where we used (4.28) and change of variables to establish the last inequality. Also,

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

The first inequality is obtained thanks to the fact that $|\xi(y)| \leq 1$ and (4.28), while the second inequality is obtained by a change of variables.

Using that

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

by combining the above inequalities we conclude that

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

Using the proof of Proposition 3.1 in [4] 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 4.2.4. *We remark that the difference between the compactness result we proved above and the one proved in Proposition 3.1 in [4] is the fact that we consider functions bounded in L^1, instead of bounded in L^∞ as was assumed in [4]. Nevertheless, after extending the functions to the entire $\mathbb{R}^d$ as above, one can directly apply the proof in [4] to obtain the desired compactness result.*

Now we are ready to establish the compactness from Proposition 1.0.7.

Proof of Proposition 1.0.7 (Compactness). Suppose $\{u_\varepsilon\}_{\varepsilon>0} \subseteq L^1(D)$ is as in the statement. As in Lemma 4.2.3, we can assume that $\rho \equiv 1$. By Remark 5.3 in [12], 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:

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

and

$$\begin{aligned} C \int_D \int_D \eta_\varepsilon(x-y) \left| u_\varepsilon(x) - u_\varepsilon(y) \right| \mathrm{d}x\mathrm{d}y &\geq \int_{\tilde{D}} \int_{\tilde{D}} \eta_\varepsilon \left(\Theta(x) - \Theta(y) \right) \left| v_\varepsilon(x) - v_\varepsilon(y) \right| \mathrm{d}x\mathrm{d}y \\ &\geq \int_{\tilde{D}} \int_{\tilde{D}} \hat{\eta}_\varepsilon(x-y) \left| v_\varepsilon(x) - v_\varepsilon(y) \right| \mathrm{d}x\mathrm{d}y. \end{aligned}$$

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 hypothesis of Lemma 4.2.3 (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 4.2.5. *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 4.2.3. 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 $\lim_{k\to\infty} \sup_{\varepsilon>0} \|u_\varepsilon\|_{L^1(D\setminus D_k,\rho)} = 0$. □

Chapter 5

Γ-convergence of graph total variation and its implications

In this chapter we establish Theorem 1.0.2. A few remarks help clarify the hypotheses and conclusions of Theorem 1.0.2. The scaling condition $\varepsilon_n \gg (\log n)^{p_d} n^{-1/d}$ (where $p_2 = 3/4$ and $p_d = 1/d$ for $d \geq 3$) comes directly from the existence of transportation maps from Proposition 1.0.9. This means that ε_n must decay more slowly than the maximal distance a point in D has to travel to match its corresponding data point in V_n. In other words, the similarity graph G_n must contain information on a larger scale than that on which the intrinsic randomness operates. Additionally, we must remark that the conclusion of the theorem still holds if the partitions $\{Y_n^*, Y_n^{*c}\}$ only approximate an optimal balanced cut, that is, if the energies of $\{Y_n^*, Y_n^{*c}\}$ satisfy

$$\lim_{n\to\infty} \left(\frac{\mathrm{Cut}(Y_n{}^*, Y_n^{*c})}{\mathrm{Bal}(Y_n^*, Y_n^{*c})} - \min_{Y \subsetneq V_n} \frac{\mathrm{Cut}(Y, Y^c)}{\mathrm{Bal}(Y, Y^c)} \right) = 0.$$

As seen from the proof of Theorem 1.0.2 in Section 5.2, this follows from Proposition 1.1.5.

It is also important to remark the optimality of scaling of ε_n for $d \geq 3$. In fact, if $d \geq 3$ then the rate presented in the statement of Theorem 1.0.2 is sharp in terms of scaling. Namely for $D = (0,1)^d$, ν being the Lebesgue measure on D and η compactly supported, it is known from graph theory (see [39, 41, 56]) that there exists a constant $c > 0$ such that if $\varepsilon_n < c\frac{(\log n)^{1/d}}{n^{1/d}}$ then the weighted graph associated to (V_n, W_n) is disconnected with high probability. The resulting optimal discrete cuts have zero energy, but may be very far from the optimal continuum cuts.

In case $d = 2$ on the other hand, the connectivity threshold for a random geometric graph is $\varepsilon_n = c\frac{\log(n)^{1/2}}{n^{1/2}}$, which is below the rate for which we can establish the consistency of balanced cuts. Thus, an interesting open problem is to determine if the consistency results are still valid when the parameter ε_n is taken below the rate $\frac{\log(n)^{3/4}}{n^1/2}$ we obtained the proof for, but above the connectivity rate. In particular we are interested in determining if connectivity is the determining factor in order to obtain consistency of balance graph cuts.

As stated in the Introduction, before we establish Theorem 1.0.2, we first prove Theorem 1.0.4, Corollary 1.0.5 and Theorem 1.0.6. We rely on the results established in the previous chapters.

5.1 Γ-convergence of GTV_{n,ε_n}

In this section we present the proof of Theorem 1.0.4.

Proof of Theorem 1.0.4. We use the sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from Proposition 1.0.9. Let $\omega \in \Omega$ be such that (1.22) and (1.23) hold in cases $d=2$ and $d\geq 3$ respectively. By Proposition 1.0.9 the complement in Ω of such ωs 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 from Proposition 1.0.9 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 (2.4) it follows that

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

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

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

Thanks to the assumptions on $\{\varepsilon_n\}_{n\in\mathbb{N}}$ ((1.22) and (1.23) in cases $d=2$ and $d\geq 3$ respectively), for large enough $n\in\mathbb{N}$:

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

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

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

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

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

Note that $\frac{\tilde{\varepsilon}_n}{\varepsilon_n} \to 1$ as $n \to \infty$ and that $u_n \xrightarrow{TL^1} u$ implies $\tilde{u}_n \xrightarrow{L^1(D,\rho)} u$ as $n \to \infty$. We deduce from Proposition 1.0.7 that $\liminf_{n\to\infty} TV_{\tilde{\varepsilon}_n}(\tilde{u}_n;\rho) \geq \sigma_\eta TV(u;\rho^2)$ and hence:

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

Limsup inequality: By Remark 1.1.2 and Proposition 4.1.4, 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 $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The fact that u is Lipschitz implies that $u_n \xrightarrow{TL^1} u$. Now, consider $\tilde{\varepsilon}_n := \varepsilon_n + \frac{2}{b}\|Id - T_n\|_{L^\infty(D)}$ and let $\tilde{u}_n = u_n \circ T_n$. Then note that for Lebesgue almost every $(x,y) \in D \times D$

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

Then for all n

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

Also

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

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

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

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

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

where the last equality follows from the last part of Proposition 1.0.7.

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_{n,ε_n}^k 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:

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

Limsup inequality: By Remark 1.1.2 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

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

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 $k \to \infty$ a.e. Denote by GTV_{n,ε_n}^k the graph TV corresponding to $\boldsymbol{\eta}_k$. If $u_n \xrightarrow{TL^1} u$ as $n \to \infty$, by Step 2 $\sigma_{\eta_k} TV(u;\rho^2) \leq \liminf_{n\to\infty} GTV_{n,\varepsilon_n}^k(u_n) \leq \liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n)$ for every $k \in \mathbb{N}$. The monotone convergence theorem implies that $\lim_{k\to\infty} \sigma_{\eta_k} = \sigma_\eta$ and so we conclude that $\sigma_\eta TV(u;\rho^2) \leq \liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n)$.

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

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 $t \leq \alpha$ and $\boldsymbol{\eta}_\alpha(t) = 0$ for $t > \alpha$. We denote by $GTV_{n,\varepsilon_n}^\alpha$ the graph total variation using $\boldsymbol{\eta}_\alpha$. Then

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

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 $|x-y| \leq |T_n(x) - T_n(y)| + 2\|Id - T_n\|_{L^\infty(D)}$ and $|T_n(x) - T_n(y)| \leq |x-y| + 2\|Id - T_n\|_{L^\infty(D)}$ we can use the fact that $\frac{\|Id-T_n\|_{L^\infty(D)}}{\varepsilon_n} \to 0$ as $n \to \infty$ to conclude that for large enough n, for almost every $(x,y) \in D \times D$ for which $|T_n(x) - T_n(y)| > \alpha\varepsilon_n$ it holds that $|x-y| \leq 2|T_n(x) - T_n(y)|$ and $|T_n(x) - T_n(y)| \leq 2|x-y|$. We conclude that for large enough n

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

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:

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

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

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

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

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

□

We now present the proof of Theorem 1.0.6 on compactness.

Proof of Theorem 1.0.6. 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 4.2.3 and the compactness part of Proposition 1.0.7 without loss of generality we can assume that $\boldsymbol{\eta}$ is of the form $\boldsymbol{\eta}(t) = a$ if $t < b$ and $\boldsymbol{\eta}(t) = 0$ for $t \geq b$, for some a and b positive constants.

Consider the sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from Proposition 1.0.9. Since $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfies (1.19), estimates (1.22) and (1.23) imply that for Lebesgue a.e. $z,y \in D$

with $|T_n(z)-T_n(y)|>b\varepsilon_n$ it holds that $|z-y|>b\varepsilon_n-2\|Id-T_n\|_{L^\infty(D)}$. For large enough n, we set $\tilde{\varepsilon}_n:=\varepsilon_n-\frac{2\|Id-T_n\|_{L^\infty(D)}}{b}>0$. We conclude that for large n and Lebesgue a.e. $z,y\in D$:

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

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

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

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 $\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\to1$ 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 1.0.7 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 1.0.5 on the Γ-convergence of graph perimeter.

Proof of Corollary 1.0.5. Note that if $\{Y_n\}_{n\in\mathbb{N}}$ is such that $Y_n\subseteq\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ and $\mathbf{1}_{Y_n}\xrightarrow{TL^1}\mathbf{1}_A$ as $n\to\infty$ for some $A\subseteq D$, then the liminf inequality follows automatically from the liminf inequality in Theorem 1.0.4. The limsup inequality is not immediate, since we cannot use the density of Lipschitz functions as we did in the proof of Theorem 1.0.4 given that we restrict our attention to characteristic functions.

We follow the proof of Proposition 3.5 in [28] 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 1.0.4, we know there exists a sequence $\{u_n\}_{n\in\mathbb{N}}$ (with $u_n\in L^1(D,\nu_n)$) such that $u_n\xrightarrow{TL^1}\mathbf{1}_A$ and $\limsup_{n\to\infty}GTV_{n,\varepsilon_n}(u_n)\leq\sigma_\eta TV(\mathbf{1}_A;\rho^2)$. It is straightforward to verify that the functionals GTV_{n,ε_n} satisfy the coarea formula:

$$GTV_{n,\varepsilon_n}(u_n)=\int_{-\infty}^{\infty}GTV_{n,\varepsilon_n}(\mathbf{1}_{\{u_n>s\}})\mathrm{d}s.$$

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

$$\int_{\delta}^{1-\delta} GTV_{n,\varepsilon_n}(\mathbf{1}_{\{u_n > s\}})\mathrm{d}s \le GTV_{n,\varepsilon_n}(u_n).$$

For every n there is $s_n \in (\delta, 1-\delta)$ such that $GTV_{n,\varepsilon_n}(\mathbf{1}_{\{u_n > s_n\}}) \le \frac{1}{1-2\delta} GTV_{n,\varepsilon_n}(u_n)$. Define $A_n^{\delta} := \{u_n > s_n\}$. It is straightforward to show that $\mathbf{1}_{A_n^{\delta}} \xrightarrow{TL^1} \mathbf{1}_A$ as $n \to \infty$ and that $\limsup_{n\to\infty} \mathrm{Per}_{n,\varepsilon_n}(A_n^{\delta}) \le \frac{1}{1-2\delta} \sigma_\eta \,\mathrm{Per}(A;\rho^2)$. Taking $\delta \to 0$ and using a diagonal argument provides sets $\{A_n\}_{n\in\mathbb{N}}$ such that $\mathbf{1}_{A_n} \xrightarrow{TL^1} \mathbf{1}_A$ as $n \to \infty$ and $\limsup_{n\to\infty} \mathrm{Per}_{n,\varepsilon_n}(A_n) \le \sigma_\eta \,\mathrm{Per}(A;\rho^2)$. □

Remark 5.1.1. *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 1.0.4. In this case, instead of approximating by Lipschitz functions, one would approximate $\mathbf{1}_A$ 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 smooth boundary. As in the proof of Theorem 1.0.4, the key is to show that for step kernels ($\boldsymbol{\eta}(t) = b$ if $0 \le t < a$ and zero otherwise)*

$$\lim_{n\to\infty} GTV_{n,\varepsilon_n}(\mathbf{1}_G) = TV(\mathbf{1}_G;\rho^2).$$

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 such that for all n for which $||Id - T_n||_{L^\infty(D)} \le \delta$,*

$$\int_D |\mathbf{1}_G(x) - \mathbf{1}_G(T_n(x))|\mathrm{d}x \le 4\,\mathrm{Per}(E)\,||Id - T_n||_{L^\infty(D)}.$$

This estimate follows from the fact that if $\mathbf{1}_G(x) \ne \mathbf{1}_G(T_n(x))$ then $d(x,\partial E) \le |x - T_n(x)|$ and the fact that, for δ small enough, $|\{x \in \mathbb{R}^d : d(x,\partial E) < \delta\}| \le 4\,\mathrm{Per}(E)\delta$, which follows from Weyl's formula [74] 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 [6]) and using Remark 1.1.2 we obtain the limsup inequality for the characteristic function of any measurable set.

5.1.1 Extension to different sets of points

Consider the setting of Theorem 1.0.4. The only information about the points $\mathbf{x}_i$ that the proof requires is the upper bound on the ∞-transportation distance between ν and the empirical measure ν_n. Theorem 1.0.9 provides such bounds when $\mathbf{x}_i$ 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 ε_n in Theorem 1.0.4 for which the Γ-convergence holds.

That is, if $\mathbf{x}_1, \dots, \mathbf{x}_n, \dots$ are the given points, let ν_n still be $\frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{x}_i}$. If one can find transportation maps T_n from ν to ν_n such that

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

for some nonnegative function $f : \mathbb{N} \to (0,\infty)$ then Theorem 1.0.4 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 $V = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ in D, $\sup_{y\in D} \mathrm{dist}(y, V) \geq cn^{-1/d}$ and thus $n^{1/d}\|Id - T_n\|_\infty \geq c$.

One special case is when $D = (0,1)^d$, ν is the Lebesgue measure and $\mathbf{x}_1, \dots, \mathbf{x}_n, \dots$ 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 [28] and [21] 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 2.2.13.

5.2 Consistency of Cheeger and ratio graph cuts

We note that convergence in TL^1 was only defined for functions, and thus it is important to clarify what is meant by TL^1-convergence for partitions. In fact, when defining a notion of convergence for sequences of partitions $\{Y_1^n, \dots, Y_R^n\}$, we need to address the inherent ambiguity that arises from the fact that both $\{Y_1^n, \dots, Y_R^n\}$ and $\{Y_{P(1)}^n, \dots, Y_{P(R)}^n\}$ refer to the same partition for any permutation P of $\{1, \dots, R\}$. Having the previous observation in mind, the convergence of partitions is defined in a natural way.

Definition 5.2.1. *The sequence $\{Y_1^n, \dots, Y_R^n\}_{n\in\mathbb{N}}$, where $\{Y_1^n, \dots, Y_R^n\}$ is a partition of V_n, converges in the TL^1-sense to the partition $\{A_1, \dots, A_R\}$ of D, if there exists a sequence of permutations $\{P_n\}_{n\in\mathbb{N}}$ of the set $\{1, \dots, R\}$, such that for every $r \in \{1, \dots, R\}$,*

$$\mathbf{1}_{Y^n_{P_n(r)}} \xrightarrow{TL^1} \mathbf{1}_{A_r}, \quad \textit{as } n \to \infty.$$

We start the proof of Theorem 1.0.2 by showing that (1.5) actually has a minimizer. The first step is to reformulate (1.5) in a way that allows us to handle the balance term. We extend the balance term to arbitrary functions $u \in L^1(D,\nu)$:

$$B_{\mathrm{R}}(u) = \int_D |u(x) - \mathrm{mean}_\rho(u)|\rho(x)\,\mathrm{d}x \quad \text{and} \quad B_{\mathrm{C}}(u) = \min_{c\in\mathbb{R}} \int_D |u(x) - c|\rho(x)\,\mathrm{d}x, \tag{5.7}$$

where $\mathrm{mean}_\rho(u)$ denotes the mean/expectation of $u(x)$ with respect to the measure $\mathrm{d}\nu = \rho\mathrm{d}x$. From here on, we use B to represent either B_R or B_C depending on the context. We have the relations (see (1.3))

$$B_\mathrm{R}(\mathbf{1}_A) = \mathrm{Bal}_\mathrm{R}(A,A^c), \;\; B_\mathrm{C}(\mathbf{1}_A) = \mathrm{Bal}_\mathrm{C}(A,A^c), \tag{5.8}$$

for every measurable subset A of D. We also consider *normalized indicator functions* $\tilde{\mathbf{1}}_A$ given by

$$\tilde{\mathbf{1}}_A := \frac{\mathbf{1}_A}{B(\mathbf{1}_A)}, \quad A \subseteq D,$$

and consider the set

$$\mathrm{Ind}(D) := \left\{u \in L^1(\nu) : u = \tilde{\mathbf{1}}_A \text{ for some measurable set } A \subseteq D \text{ with } B(\mathbf{1}_A) \neq 0\right\}. \tag{5.9}$$

Then for $u = \tilde{\mathbf{1}}_A \in \mathrm{Ind}(D)$

$$TV(u;\rho^2) = TV(\tilde{\mathbf{1}}_A;\rho^2) = TV\left(\frac{\mathbf{1}_A}{B(\mathbf{1}_A)};\rho^2\right) = \frac{TV(\mathbf{1}_A;\rho^2)}{B(\mathbf{1}_A)} = \frac{2\,\mathrm{Per}(A;\rho^2)}{\mathrm{Bal}(A,A^c)}. \tag{5.10}$$

Thus, we deduce that problem (1.5) is equivalent to :

$$\text{Minimize} \qquad E(u) := \begin{cases} TV(u;\rho^2) & \text{if } u \in \mathrm{Ind}(D) \\ +\infty & \text{otherwise.} \end{cases} \tag{5.11}$$

Before we show that the above problem actually has a minimizer we need the following lemma.

Lemma 5.2.2. *(i) The balance functions B are continuous in $L^1(D,\nu)$.*

(ii) The set $\mathrm{Ind}(D)$ *is closed in $L^1(D,\nu)$.*

Proof. Let us start by proving (i). We first consider the balance term $B_\mathrm{C}(u)$ that corresponds to the Cheeger Cut. Suppose that $u_k \to u$ in $L^1(D,\nu)$, and let c_k, c_∞ denote medians of u_k and u respectively. By definition, c_k and c satisfy

$$c_k \in \operatorname*{argmin}_{c\in\mathbb{R}} \int_D |u_k(x) - c|\; \rho(x)\,\mathrm{d}x, \qquad c_\infty \in \operatorname*{argmin}_{c\in\mathbb{R}} \int_D |u(x) - c|\; \rho(x)\,\mathrm{d}x.$$

This implies that

$$\int_D |u_k(x) - c_k|\rho(x)\,\mathrm{d}x \leq \int_D |u_k(x) - c|\rho(x)\,\mathrm{d}x$$

for any $c \in \mathbb{R}$, so that in particular we have

$$\begin{aligned} &\int_D |u_k - c_k|\rho(x)\,\mathrm{d}x - \int_D |u - c_\infty|\;\rho(x)\,\mathrm{d}x \\ &\leq \int_D |u_k - c_\infty|\rho(x)\mathrm{d}x - \int_D |u - c_\infty|\rho(x)\mathrm{d}x \leq \int_D |u_k - u|\rho(x)\mathrm{d}x = \|u_k - u\|_{L^1(D,\nu)}. \end{aligned}$$

Exchanging the role of u_k and u in this argument implies that the inequality

$$\int_D |u - c_\infty| \mathrm{d}\rho - \int_D |u_k - c_k| \rho(x) \mathrm{d}x \leq \int_D |u - u_k| \rho(x) \mathrm{d}x \leq \|u_k - u\|_{L^1(D,\nu)}$$

also holds. Combining these inequalities shows that $|B(u_k) - B(u)| \leq \|u_k - u\|_{L_1(D,\nu)} \to 0$ as desired. Now consider the balance term $B_{\mathrm{R}}(u)$ that corresponds to the ratio Cut. For the ratio cut, the inequality $||a| - |b|| \leq |a - b|$ immediately implies

$$\begin{aligned}
&\left| \int_D |u_k - \mathrm{mean}_\rho(u_k)| \rho(x) \mathrm{d}x - \int_D |u - \mathrm{mean}_\rho(u)| \rho(x) \mathrm{d}x \right| \\
&\leq \int_D |u_k - u| \rho(x) \mathrm{d}x + \int_D |\,\mathrm{mean}_\rho(u_k) - \mathrm{mean}_\rho(u)| \rho(x) \mathrm{d}x \\
&\leq \int_D |u_k - u| \rho(x) \mathrm{d}x + |\,\mathrm{mean}_\rho(u_k) - \mathrm{mean}_\rho(u)|.
\end{aligned}$$

Since $u_k \to u$ in $L^1(D,\nu)$ we have that $\mathrm{mean}_\rho(u_k) \to \mathrm{mean}_\rho(u)$ and therefore $|B(u_k) - B(u)| \leq \|u_k - u\|_{L_1(D,\nu)} + |\,\mathrm{mean}_\rho(u_k) - \mathrm{mean}_\rho(u)| \to 0$ as desired.

In order to prove (ii) suppose that $\{u_k\}_{n\in\mathbb{N}}$ is a sequence in $\mathrm{Ind}(D)$ converging in $L^1(D,\nu)$ to some $u \in L^1(D,\nu)$, we need to show that $u \in \mathrm{Ind}(D)$. By (i) we know that $B(u_k) \to B(u)$ as $k \to \infty$. Since $u_k \in \mathrm{Ind}(D)$, in particular $B(u_k) = 1$. Thus, $B(u) = 1$. On the other hand, $u_k \in \mathrm{Ind}(D)$ implies that u_k has the form $u_k = \alpha_k \mathbf{1}_{A_k}$. Since this is true for every k, in particular we must have that u has the form $u = \alpha \mathbf{1}_A$for some real number α and some measurable subset A of D. Finally, the fact that B is 1-homogeneous implies that $1 = B(u) = \alpha B(\mathbf{1}_A)$. In particular $B(\mathbf{1}_A) \neq 0$ and $\alpha = \frac{1}{B(\mathbf{1}_A)}$. Thus $u = \tilde{\mathbf{1}}_A$ with $B(\mathbf{1}_A) \neq 0$ and hence $u \in \mathrm{Ind}(D)$. □

Lemma 5.2.3. *Let D and ν be as stated at the beginning of this section. There exists a measurable set $A \subseteq D$ with $0 < \nu(A) < 1$ such that $\tilde{\mathbf{1}}_A$ minimizes* (5.11)*.*

Proof. The statement follows by the direct method of the calculus of variations. Since the functional is bounded from below it suffices to show that it is lower semicontinuous with respect to the $L^1(D,\nu)$ norm and that a minimizing sequence is precompact in $L^1(D,\nu)$. To show lower semi-continuity it is enough to consider a sequence $u_n = \mathbf{1}_{A_n} \in \mathrm{Ind}(D)$ converging in $L^1(D,\nu)$ to $u \in L^1(D,\nu)$. From Lemma 5.2.2 it follows that $u \in$ Ind(D) and hence $u = \tilde{\mathbf{1}}_A$ for some A with $B(A) > 0$. Therefore $\mathbf{1}_{A_n} \to \mathbf{1}_A$ as $n \to \infty$ in $L^1(D,\nu)$. The lower semi-continuity then follows from the lower semi-continuity of the total variation (4.4), the continuity of B and the fact that since $B(\mathbf{1}_A) > 0$, $1/B(\mathbf{1}_{A_n}) \to 1/B(\mathbf{1}_A)$ as $n \to \infty$. The precompactness of any minimizing sequence of (5.11) follows directly from Theorem 5.1 in [10], which completes the proof. □

The next step in the proof of Theorem 1.0.2 is to reformulate problem 1.2 in a similar way to the way we reformulated 1.5. For $u_n \in L^1(D,\nu_n)$, we define

$$B^n_{\mathrm{R}}(u_n) := \int_D |u_n(x) - \mathrm{mean}_n(u_n)| \mathrm{d}\nu_n(x) \quad \text{and} \quad B^n_{\mathrm{C}}(u_n) := \min_{c\in\mathbb{R}} \int_D |u_n(x) - c| \mathrm{d}\nu_n(x). \tag{5.12}$$

Here $\text{mean}_n(u_n) = \int_D u_n(x)\mathrm{d}\nu_n(x)$. A straightforward computation shows that for $Y_n \subseteq V_n$

$$B^n_{\mathrm{R}}(\mathbf{1}_{Y_n}) = \mathrm{Bal}_{\mathrm{R}}(Y_n, Y_n^c), \;\; B^n_{\mathrm{C}}(\mathbf{1}_{Y_n}) = \mathrm{Bal}_{\mathrm{C}}(Y_n, Y_n^c). \tag{5.13}$$

From here on we write B_n to represent either B^n_{R} or B^n_{C} depending on the context. Given $Y_n \subseteq V_n$ with $B_n(\mathbf{1}_{Y_n}) \neq 0$, the *normalized indicator function* $\tilde{\mathbf{1}}_{Y_n}(x)$ is defined by

$$\tilde{\mathbf{1}}_{Y_n}(x) = \mathbf{1}_{Y_n}(x)/B^n_{\mathrm{C}}(\mathbf{1}_{Y_n}) \quad \text{or} \quad \tilde{\mathbf{1}}_{Y_n}(x) = \mathbf{1}_{Y_n}(x)/B^n_{\mathrm{R}}(\mathbf{1}_{Y_n}).$$

Note that $B_n(\tilde{\mathbf{1}}_{Y_n}) = 1$. We also restrict the minimization of $E_n(u)$ to the set

$$\mathrm{Ind}_n(D) := \{u_n \in L^1(D, \nu_n) : u_n = \tilde{\mathbf{1}}_{Y_n} \text{ for some } Y_n \subseteq V_n \text{ with } B_n(\mathbf{1}_{Y_n}) \neq 0\}. \tag{5.14}$$

Now, suppose that $u_n \in \mathrm{Ind}_n(D)$, i.e. that $u_n = \tilde{\mathbf{1}}_{Y_n}$ for some set Y_n with $B_n(\mathbf{1}_{Y_n}) > 0$. Using (5.8) together with the fact that GTV_{n,ε_n} is one-homogeneous implies, as in (5.10)

$$GTV_{n,\varepsilon_n}(u_n) = \frac{2}{n^2 \varepsilon_n^{d+1}} \frac{\mathrm{Cut}(Y_n, Y_n^c)}{\mathrm{Bal}(Y_n, Y_n^c)}. \tag{5.15}$$

Thus, the minimization problem

$$\text{Minimize} \quad E_n(u_n) := \begin{cases} GTV_{n,\varepsilon_n}(u_n) & \text{if } u_n \in \mathrm{Ind}_n(D) \\ +\infty & \text{otherwise.} \end{cases} \tag{5.16}$$

is equivalent to the balanced graph-cut problem (1.2) on the graph $G_n = (V_n, W_n)$ constructed from the first n data points.

Now, note that

$$u_n^*(x) := \tilde{\mathbf{1}}_{Y_n^*}(x), \;\; u_n^{**}(x) := \tilde{\mathbf{1}}_{Y_n^{*c}}(x) \qquad \text{minimize} \quad E_n(u_n) \;\; \text{over all} \;\; u_n \in L^1(D, \nu_n), \tag{5.17}$$

and that

$$u^*(x) := \tilde{\mathbf{1}}_{A^*}(x), \;\; u^{**}(x) := \tilde{\mathbf{1}}_{A^{*c}}(x) \quad \text{minimize} \quad E(u) \;\; \text{over all} \;\; u \in L^1(D, \nu). \tag{5.18}$$

We show that the approximating functionals E_n Γ-converge to $\sigma_\eta E$ in the TL^1-sense. In Lemma 5.2.6 we establish that u_n^* and u_n^{**} exhibit the required compactness. Thus, they must converge toward the normalized indicator functions u^* and u^{**} up to relabeling (see Proposition 1.1.5). If $\{A^*, A^{*c}\}$ is the unique minimizer of 1.5, the convergence of the whole sequence follows. The convergence of the partition $\{Y_n^*, Y_n^{*c}\}$ toward the partition $\{A^*, A^{*c}\}$ in the sense of Definition 5.2.1 is a direct consequence. The convergence (1.16) follows from (1.25) in Proposition 1.1.5.

Γ-convergence

Proposition 5.2.4. (Γ-Convergence) *Let the domain D, measure ν, kernel η, sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$, sample points $\{\boldsymbol{x}_i\}_{i\in N}$, and graph G_n satisfy the assumptions of Theorem 1.0.2. Let E_n be as defined in* (5.16) *and E as in* (5.11). *Then*

$$E_n \xrightarrow{\Gamma} \sigma_n E \quad \text{with respect to } TL^1 \text{ metric as } n \to \infty.$$

We leverage Theorem 1.0.4 to prove this claim. We first need a preliminary lemma which allows us to handle the presence of the additional balance terms in (5.16) and (5.11).

Lemma 5.2.5. *(i) If $\{u_n\}_{n\in\mathbb{N}}$ is a sequence with $u_n \in L^1(D,\nu_n)$ and $u_n \xrightarrow{TL^1} u$ for some $u \in L^1(D,\nu)$, then $B_n(u_n) \to B(u)$.*

(ii) If $u_n = \tilde{\mathbf{1}}_{Y_n}$, where $Y_n \subset V_n$, converges to $u = \tilde{\mathbf{1}}_A$ in the TL^1-sense, then $\mathbf{1}_{Y_n}$ converges to $\mathbf{1}_A$ in the TL^1-sense.

Proof. To prove (i), suppose that $u_n \in L^1(D,\nu_n)$ and that $u_n \xrightarrow{TL^1} u$. Let us consider $\{T_n\}_{n\in\mathbb{N}}$ a stagnating sequence of transportation maps between ν and $\{\nu_n\}_{n\in\mathbb{N}}$. Then, we have $u_n \circ T_n \xrightarrow{L^1(D,\nu)} u$ and thus by (i), we have that $B(u_n \circ T_n) \to B(u)$. To conclude the proof we notice that $B(u_n \circ T_n) = B_n(u_n)$ for every n. In fact, by the change of variables (2.4) we have that for every $c \in \mathbb{R}$

$$\int_D |u_n(x) - c| \mathrm{d}\nu_n(x) = \int_D |u_n \circ T_n(x) - c| \mathrm{d}\nu(x). \tag{5.19}$$

In particular we have $B_C^n(u_n) = B_C(u_n \circ T_n)$. Applying the change of variables (2.4), we obtain $\mathrm{mean}_n(u_n) = \mathrm{mean}_\rho(u_n \circ T_n)$ and combining with (5.19) we deduce that $B_R^n(u_n) = B_R(u_n \circ T_n)$. The proof of (ii) is straightforward. ☐

Now we turn to the proof or Proposition 5.2.4.

Proof of Proposition 5.2.4. Liminf Inequality. For arbitrary $u \in L^1(D,\nu)$ and arbitrary sequence $\{u_n\}_{n\in\mathbb{N}}$ with $u_n \in L^1(D,\nu_n)$ and with $u_n \xrightarrow{TL^1} u$, we need to show that

$$\liminf_{n\to\infty} E_n(u_n) \geq \sigma_\eta E(u).$$

First assume that $u \in \mathrm{Ind}(D)$. In particular $E(u) = TV(u;\rho^2)$. Now, note that working along a subsequence we can assume that the liminf is actually a limit and that this limit is finite (otherwise the inequality would be trivially satisfied). This implies that for all n large enough we have $E_n(u_n) < +\infty$, which in particular implies that $E_n(u_n) = GTV_{n,\varepsilon_n}(u_n)$. Theorem 1.0.4 then implies that

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

Now let as assume that $u \notin \mathrm{Ind}(D)$. Let us consider a stagnating sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ between $\{\nu_n\}_{n\in\mathbb{N}}$ and ν. Since $u_n \xrightarrow{TL^1} u$ then $u_n \circ T_n \xrightarrow{L^1(D,\nu)} u$. By Lemma 5.2.5 , the set $\mathrm{Ind}(D)$ is a closed subset of $L^1(D,\nu)$. We conclude that $u_n \circ T_n \notin \mathrm{Ind}(D)$ for all large enough n. From the proof of Lemma 5.2.5 we know that $B_n(u_n) = B(u_n \circ T_n)$ and from this fact, it is straightforward to show that $u_n \circ T_n \notin \mathrm{Ind}(D)$ if and only if $u_n \notin \mathrm{Ind}_n(D)$. Hence, $u_n \notin \mathrm{Ind}_n(D)$ for all large enough n and in particular $\liminf_{n\in\mathbb{N}} E_n(u_n) = +\infty$ which implies that the desired inequality holds in this case.

Limsup Inequality. We now consider $u \in L^1(D,\nu)$. We want to show that there exists a sequence $\{u_n\}_{n\in\mathbb{N}}$ with $u_n \in L^1(D,\nu_n)$ such that $u_n \xrightarrow{TL^1} u$ and

$$\limsup_{n\to\infty} E_n(u_n) \leq \sigma_\eta E(u).$$

Let us start by assuming that $u \notin \mathrm{Ind}(D)$. In this case $E(u) = +\infty$. From Theorem 1.0.4 we know there exists at least one sequence $\{u_n\}_{n\in\mathbb{N}}$ with $u_n \in L^1(D,\nu_n)$ such that $u_n \xrightarrow{TL^1} u$. Since $E(u) = +\infty$, the inequality is trivially satisfied in this case.

On the other hand, if $u \in \mathrm{Ind}(D)$, we know that $u = \tilde{\mathbf{1}}_A$ for some measurable subset A of D with $B(\mathbf{1}_A) \neq 0$. By Theorem 1.0.5, there exists a sequence $\{Y_n\}_{n\in\mathbb{N}}$ with $Y_n \subseteq V_n$, satisfying $\mathbf{1}_{Y_n} \xrightarrow{TL^1} \mathbf{1}_A$ and

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(\mathbf{1}_{Y_n}) \leq \sigma_\eta TV(\mathbf{1}_A;\rho^2). \tag{5.20}$$

Since $\mathbf{1}_{Y_n} \xrightarrow{TL^1} \mathbf{1}_A$ Lemma 5.2.5 implies that

$$B_n(\mathbf{1}_{Y_n}) \to B(\mathbf{1}_A). \tag{5.21}$$

In particular $B_n(\mathbf{1}_{Y_n}) \neq 0$ for all n large enough, and thus we can consider the function $u_n := \tilde{\mathbf{1}}_{Y_n} \in \mathrm{Ind}_n(D)$. From (5.21) it follows that $u_n \xrightarrow{TL^1} u$ and together with (5.20) it follows that

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) = \limsup_{n\to\infty} \frac{1}{B_n(Y_n)} GTV_{n,\varepsilon_n}(\mathbf{1}_{Y_n}) \leq \frac{1}{B(\mathbf{1}_A)} \sigma_\eta TV(\mathbf{1}_A;\rho^2) = \sigma_\eta TV(u;\rho^2).$$

Since, $u_n \in \mathrm{Ind}_n(D)$ for all n large enough, in particular we have $GTV_{n,\varepsilon_n}(u_n) = E_n(u_n)$ and also since $u \in \mathrm{Ind}(D)$, we have $E(u) = TV(u;\rho^2)$. These facts together with the previous chain of inequalities imply the result. □

Compactness

Lemma 5.2.6 (Compactness)**.** *Any subsequence of $\{u_n^*\}_{n\geq 1}$ or $\{u_n^{**}\}_{n\geq 1}$ of minimizers of E_n (defined in* (5.17) *and* (5.18)*) has a further subsequence that converges in the TL^1-sense.*

Proof. Let u_n^*, u_n^{**} denote minimizing sequences. Thanks to Theorem 1.0.6, to show that any subsequence of u_n^* has a convergent subsequence, it suffices to show that both

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n^*) < +\infty \tag{5.22}$$

$$\limsup_{n\to\infty} \|u_n^*\|_{L^1(\nu_n)} < +\infty, \tag{5.23}$$

hold. From the Γ-convergence established in Proposition 5.2.4 and from the proof of Proposition 1.1.5 it follows that (5.22) is satisfied for both minimizing sequences. Recall that $u_n^* = \mathbf{1}_{Y_n^*}/B_n(\mathbf{1}_{Y_n^*})$ and that $u_n^{**} = \mathbf{1}_{Y_n^{*c}}/B_n(\mathbf{1}_{Y_n^{*c}})$, where Y_n^* denotes an optimal balanced cut.

To show (5.23), consider first the balance term that corresponds to the Cheeger Cut. Define a sequence v_n as follows. Set $v_n := u_n^*$ if $|Y_n^*| \le |Y_n^{*c}|$ and $v_n = u_n^{**}$ otherwise. It then follows that

$$\|v_n\|_{L^1(\nu_n)} = \frac{\min\{|Y_n^*|, |Y_n^{*c}|\}}{\min\{|Y_n^*|, |Y_n^{*c}|\}} = 1.$$

Also, note that $GTV_{n,\varepsilon_n}(v_n) = GTV_{n,\varepsilon_n}(u_n^*)$. Thus (5.22) and (5.23) hold for v_n, so that any subsequence of v_n has a convergent subsequence in the TL^1-sense. Let $v_{n_k} \xrightarrow{TL^1} v$ denote a convergent subsequence. Now observe that by construction v_{n_k} minimizes E_{n_k} for every k. Thus, it follows from Proposition 5.2.4 and general properties of Γ-convergence (see Proposition 1.1.5), that v minimizes E and in particular v is a normalized characteristic function, that is, $v = \mathbf{1}_A/B(\mathbf{1}_A)$ for some $A \subseteq D$ with $B(\mathbf{1}_A) \neq 0$. Since $B_{n_k}(\mathbf{1}_{Y_{n_k}^*}) = B_{n_k}(\mathbf{1}_{Y_{n_k}^{*c}})$, $v_{n_k} \xrightarrow{TL^1} v$ implies that

$$\frac{1}{B_{n_k}(Y_{n_k}^*)} \to \frac{1}{B(A)}.$$

Therefore, for large enough k we have

$$\|u_{n_k}^*\|_{L^1(D,\nu_{n_k})} \le \frac{1}{B_{n_k}(Y_{n_k}^*)} \le \frac{2}{B(A)}$$

and

$$\|u_{n_k}^{**}\|_{L^1(D,\nu_{n_k})} \le \frac{1}{B_{n_k}(Y_{n_k}^{*c})} = \frac{1}{B_{n_k}(Y_{n_k}^*)} \le \frac{2}{B(A)}.$$

We conclude that $\|u_{n_k}^*\|_{L^1(D,\nu_{n_k})}$ and $\|u_{n_k}^{**}\|_{L^1(D,\nu_{n_k})}$ remain bounded, so that both minimizing subsequences satisfy (5.23) and (5.22) simultaneously. This yields compactness in the Cheeger Cut case.

Now consider the balance term $B(u) = B_{\mathrm{R}}(u)$ that corresponds to the Ratio Cut. Define a sequence $v_n := u_n^* - \mathrm{mean}_n(u_n^*)$, and note that $GTV_{n,\varepsilon_n}(v_n) = GTV_{n,\varepsilon_n}(u_n^*)$ since the total variation is invariant with respect to translation. It then follows that

$$\|v_n\|_{L^1(D,\nu)} = \int_D |u_n^*(x) - \mathrm{mean}_\rho(u_n^*)|\rho(x)\,\mathrm{d}x = B(u_n^*) = 1.$$

Thus the sequence $\{v_n\}_{n\in\mathbb{N}}$ is precompact in TL^1. Let $v_{n_k} \xrightarrow{TL^1} v$ denote a convergent subsequence. Using a stagnating sequence of transportation maps $\{T_{n_k}\}_{k\in\mathbb{N}}$ between ν and the sequence of measures $\{\nu_{n_k}\}_{k\in\mathbb{N}}$, we have that $v_{n_k}\circ T_{n_k} \xrightarrow{L^1(D,\nu)} v$. By passing to a further subsequence if necessary, we may assume that $v_{n_k}\circ T_{n_k}(x) \to v(x)$ for ν-almost every x in D.

For any such x, we have that either $T_{n_k}(x) \in Y^*_{n_k}$ or $T_{n_k}(x) \in Y^{*c}_{n_k}$ so that either

$$v_{n_k}\circ T_{n_k}(x) = \frac{1}{2|Y^*_{n_k}|} \quad \text{or} \quad v_{n_k}\circ T_{n_k}(x) = -\frac{1}{2|Y^{*c}_{n_k}|}.$$

Now, by continuity of the balance term, we have

$$B(v) = \lim_{k\to\infty} B_{n_k}(v_{n_k}) = 1,$$

and also

$$\mathrm{mean}_\rho(v) = \lim_{k\to\infty} \mathrm{mean}_{n_k}(v_{n_k}) = 0.$$

In particular the measure of the region in which v is positive is strictly greater than zero, and likewise the measure of the region in which v is negative is strictly greater than zero. It follows that both $|Y^*_{n_k}|$ and $|Y^{*c}_{n_k}|$ remain bounded away from zero for all k sufficiently large. As a consequence, the fact that

$$\|u^*_{n_k}\|_{L^1(D,\nu_{n_k})} = \frac{1}{2|Y^{*c}_{n_k}|}, \qquad \|u^{**}_{n_k}\|_{L^1(D,\nu_{n_k})} = \frac{1}{2|Y^*_{n_k}|},$$

implies that both (5.22) and (5.23) hold along a subsequence, yielding the desired compactness. □

Conclusion of the proof of Theorem 1.0.2

Proof of Theorem 1.0.2. From Proposition 1.1.5, we know that any limit point of $\{u^*_n\}_{n\in\mathbb{N}}$ (in the TL^1-sense) must equal u^* or u^{**}. As a consequence, for any subsequence $u^*_{n_k}$ that converges to u^* we have that $\mathbf{1}_{Y^*_{n_k}} \xrightarrow{TL^1} \mathbf{1}_{A^*}$ by Lemma 5.2.5, while $\mathbf{1}_{Y^*_{n_k}} \xrightarrow{TL^1} \mathbf{1}_{A^{*c}}$ if the subsequence converges to u^{**} instead. Moreover, in the first case we would also have $\mathbf{1}_{Y^{*c}_{n_k}} \xrightarrow{TL^1} \mathbf{1}_{A^{*c}}$ and in the second case $\mathbf{1}_{Y^c_{n_k}} \xrightarrow{TL^1} \mathbf{1}_{A^*}$. Thus in either case we have

$$\left\{Y^*_{n_k}, Y^{*c}_{n_k}\right\} \xrightarrow{TL^1} \left\{A^*, A^{*c}\right\}.$$

Thus, for any subsequence of $\{Y^*_n, Y^{*c}_n\}_{n\in\mathbb{N}}$ it is possible to obtain a further subsequence converging to $\{A^*, A^{*c}\}$, and thus the full sequence converges to $\{A^*, A^{*c}\}$. □

Chapter 6

Consistency of multiway balanced cuts

In this chapter we establish Theorem 1.0.10. Overall, its proof follows similar arguments to the ones in the proof of Theorem 1.0.2, where the notion of Γ-convergence plays the leading role. Just as what we did in the two-class case, we reformulate both the balanced graph-cut problem (1.4) and the analogous balanced domain-cut problem (1.9) as equivalent minimizations defined over spaces of functions and not just spaces of partitions or sets.

We let $B_n(u_n) := \mathrm{mean}_n(u_n)$ for $u_n \in L^1(D, \nu_n)$ and $B(u) := \mathrm{mean}_\rho(u)$ for $u \in L^1(D, \nu)$, to be the corresponding balance terms. Given this balance terms, we let $\mathrm{Ind}_n(D)$ and $\mathrm{Ind}(D)$ be defined as in (5.14) and (5.9) respectively. We can then let the sets $\mathscr{M}_n(D)$ and $\mathscr{M}(D)$ to consist of those collections $\mathscr{U} = (u_1, \ldots, u_R)$ comprised of exactly R disjoint, normalized indicator functions that cover D. The sets $\mathscr{M}_n(D)$ and $\mathscr{M}(D)$ are the multi-class analogues of $\mathrm{Ind}_n(D)$ and $\mathrm{Ind}(D)$ respectively. Specifically, we let

$$\mathscr{M}_n(D) = \left\{ (u_1^n, \ldots, u_R^n) \, : u_r^n \in \mathrm{Ind}_n(D), \; \int_D u_r^n(x) u_s^n(x) \, \mathrm{d}\nu_n(x) = 0 \; \text{ if } \; r \neq s, \; \sum_{r=1}^R u_r^n > 0 \right\}, \tag{6.1}$$

$$\mathscr{M}(D) = \left\{ (u_1, \ldots, u_R) \, : u_r \in \mathrm{Ind}(D), \; \int_D u_r(x) u_s(x) \, \mathrm{d}\nu(x) = 0 \; \text{ if } \; r \neq s, \; \sum_{r=1}^R u_r > 0 \right\}. \tag{6.2}$$

Note for example that if $\mathscr{U} = (u_1, \ldots, u_R) \in \mathscr{M}(D)$, then the functions u_r are normalized indicator functions, $u_r = \mathbf{1}_{A_r}/|A_r|$ for $1 \leq r \leq R$, and the orthogonality constraints imply that $\{A_1, \ldots, A_R\}$ is a collection of pairwise disjoint sets (up to Lebesgue-null sets). Additionally, the condition that $\sum_{r=1}^R u_r > 0$ holds almost everywhere implies that the sets $\{A_1, \ldots, A_R\}$ cover D up to Lebesgue-null sets.

With these definitions at hand, we may follow the same argument in the two-class case to conclude that that the minimization

$$\text{Minimize} \quad E_n(\mathscr{U}_n) := \begin{cases} \sum_{r=1}^{R} GTV_{n,\varepsilon_n}(u_r^n) & \text{if } \mathscr{U}_n \in \mathscr{M}_n(D) \\ +\infty & \text{otherwise} \end{cases} \tag{6.3}$$

is equivalent to the balanced graph-cut problem (1.4), while the minimization

$$\text{Minimize} \quad E(\mathscr{U}) := \begin{cases} \sum_{r=1}^{R} TV(u_r;\rho^2) & \text{if } \mathscr{U} \in \mathscr{M}(D) \\ +\infty & \text{otherwise} \end{cases} \tag{6.4}$$

is equivalent to the balance domain-cut problem (1.9).

At this stage, the proof of Theorem 1.0.10 is completed by following the same steps as in the two-class case. In particular we want to show that E_n defined in (6.3) Γ-converges in the TL^1-sense to $\sigma_\eta E$, where E is defined in 6.4. That is, we want to prove the following.

Proposition 6.0.7. (Γ-Convergence) *Let the domain D, measure ν, kernel η, sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$, sample points $\{\boldsymbol{x}_i\}_{i\in N}$, and graph G_n satisfy the assumptions of Theorem 1.0.2. Consider the functional E_n as in* (6.3) *and the functional E as in* (6.4)*. Then*

$$E_n \xrightarrow{\Gamma} \sigma_n E \quad \text{with respect to } (TL^1)^R \text{ metric as } n \to \infty.$$

That is

1. *For any $\mathscr{U} \in (L^1(D,\nu))^R$ and any sequence $\mathscr{U}_n \in (L^1(D,\nu_n))^R$ that converges to $\mathscr{U}$ in the TL^1 sense,*
$$E(\mathscr{U}) \le \liminf_{n\to\infty} E_n(\mathscr{U}_n). \tag{6.5}$$

2. *For any $\mathscr{U} \in (L^1(D,\nu))^R$ there exists at least one sequence $\mathscr{U}_n$ that both converges to $\mathscr{U}$ in the TL^1-sense and also satisfies*
$$\limsup_{n\to\infty} E_n(\mathscr{U}_n) \le E(\mathscr{U}). \tag{6.6}$$

In the above, $(TL^1)^R := TL^1 \times \cdots \times TL^1$ (R times). $(L^1(D,\nu))^R$ is defined analogously.

Remark 6.0.8. *We remark that all the types of convergence for vector-valued functions are to be understood as component-wise convergence in the corresponding topology. This helps us clarify the way the TL^1 -convergence is considered in Proposition 6.0.7.*

Moreover, we establish the corresponding compactness.

Proposition 6.0.9 (Compactness)**.** *Any subsequence of $\{\mathscr{U}_n^*\}_{n\ge 1}$ of minimizers to* (6.3) *has a further subsequence that converges in the TL^1-sense.*

In order to establish Proposition 6.0.7 however, there is an extra difficulty in relation to the equivalent result in the two class case which has to do with the following. In Remark 5.1.1 we use an approximation of arbitrary sets with sets that have smooth boundary and whose perimeter approximates that of the original set. This is done in order to establish the Γ-convergence of the graph perimeter. Such approximation is a classical result in geometric measure theory (see [6, 50]). To establish Proposition 6.0.7, it would be desirable to have a similar approximation procedure. The problem is that when one considers a partition of a domain into three or more sets, triple junctions appear, implying that it is not possible to approximate an arbitrary partition with partitions that consist of sets with smooth boundary. Moreover, there is a priori no obvious way to find a "smoother" partition, which approximates the original partition, and recovers the perimeter of the sets that form the original partition. For this reason the first step in order to establish Proposition 6.0.7, and ultimately Theorem 1.0.10, is to provide such approximation step. In Section 6.2 we establish Proposition 6.0.7 and Proposition 6.0.9.

6.1 Density of partitions consisting of piecewise smooth sets

Definition 6.1.1. *We say that an open and bounded set $A \subseteq \mathbb{R}^d$ has* piecewise (PW) smooth *boundary if its boundary is a subset of the union of finitely many $d-1$-dimensional manifolds embedded in $\mathbb{R}^d$. Finally, we say that $\{A_1,\dots,A_R\}$ is a partition of D induced by piecewise smooth sets, if $A_r = Q_r \cap D$ where for all r, Q_r is a subset of $\mathbb{R}^d$ with piecewise smooth boundary such that*

$$|D\mathbf{1}_{Q_r}|_{\rho^2}(\partial D) = 0.$$

In the above and throughout this section, we assume that $\rho : D \to \mathbb{R}$ has been extended to a lower semicontinuous function $\rho : \mathbb{R}^d \to \mathbb{R}$ which is bounded above and below by the same constants bounding ρ. This can be achieved for example by setting $\rho \equiv \frac{1}{\lambda}$ on $\mathbb{R}^d \setminus D$. In this section we write the dependence of the weighted total variation in terms of the set D or $\mathbb{R}^d$ as in (4.1).

We show that for any $\mathscr{U} = (\tilde{\mathbf{1}}_{A_1},\dots,\tilde{\mathbf{1}}_{A_R})$ where each of the sets A_r has finite perimeter, there exists a sequence $\left\{\mathscr{U}_m = (\tilde{\mathbf{1}}_{A_1^m},\dots,\tilde{\mathbf{1}}_{A_R^m})\right\}_{m\in\mathbb{N}}$, where each of the $\mathscr{U}_m$ is induced by piecewise smooth sets, and such that for every $r \in \{1,\dots,R\}$

$$\mathbf{1}_{A_r^m} \xrightarrow{L^1(D,\nu)} \mathbf{1}_{A_r},$$

and

$$\lim_{m\to\infty} TV(\mathbf{1}_{A_r^m};\rho^2;D) = TV(\mathbf{1}_{A_r};\rho^2;D).$$

Note that by establishing the existence of such approximating sequence, it immediately follows that $\mathscr{U}_m \to \mathscr{U}$ in $(L^1(D,\nu))^R$ and that $\lim_{m\to\infty} E(\mathscr{U}_m) = E(\mathscr{U})$ (by continuity of the balance terms). The approximation can be obtained from the Appendix in [11]. Here we present our own construction which relies on a simple observation (Lemma 6.1.2 below) and the usual mollification-truncation argument (see Section 13.6 in [50]).

Lemma 6.1.2. *Let $\{A_1,\dots,A_R\}$ denote a collection of open and bounded sets with smooth boundary in $\mathbb{R}^d$ that satisfy*

$$\mathscr{H}^{d-1}(\partial A_r \cap \partial A_s) = 0\,,\, \forall r \neq s, \tag{6.7}$$

where $\mathscr{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. Let D denote an open and bounded set. Then there exists a permutation $\pi : \{1,\dots,R\} \to \{1,\dots,R\}$ such that

$$TV(\mathbf{1}_{A_{\pi(r)}\setminus \bigcup_{s=r+1}^{R} A_{\pi(s)}};\rho^2;D) \leq TV(\mathbf{1}_{A_{\pi(r)}};\rho^2;D),\, \forall r \in \{1,\dots,R\}\,.$$

Proof. The proof is by induction on R. *Base case:* Note that if $R=1$ there is nothing to prove. *Inductive Step:* Suppose that the result holds when considering any $R-1$ sets as described in the statement. Let $A_1,\dots,A_R$ be a collection of sets as in the statement. By the induction hypothesis it is enough to show that we can find $r \in \{1,\dots,R\}$ such that

$$TV(\mathbf{1}_{A_r \setminus \bigcup_{s\neq r} A_s};\rho^2;D) \leq TV(\mathbf{1}_{A_r};\rho^2;D). \tag{6.8}$$

To simplify notation, denote by Γ_r the set ∂A_r and define a_{rs} as the quantity

$$a_{rs} := \int_{\Gamma_r \cap (A_s \setminus \bigcup_{k\neq r, k\neq s} A_k)\cap D} \rho^2(x)\, \mathrm{d}\mathscr{H}^{d-1}(x).$$

Hypothesis (6.7) and the smoothness of the sets A_r imply that the equality

$$TV(\mathbf{1}_{A_r \setminus \bigcup_{s\neq r} A_s};\rho^2;D) = \int_{\Gamma_r \cap (\bigcup_{k\neq r} A_k)^c \cap D} \rho^2(x)\, \mathrm{d}\mathscr{H}^{d-1}(x) + \sum_{s\colon s\neq r} a_{sr} \tag{6.9}$$

holds for every $r \in \{1,\dots,R\}$, as does the inequality

$$TV(\mathbf{1}_{A_r};\rho^2;D) \geq \int_{\Gamma_r \cap (\bigcup_{k\neq r} A_k)^c \cap D} \rho^2(x)\, \mathrm{d}\mathscr{H}^{d-1} + \sum_{s\colon s\neq r} a_{rs}. \tag{6.10}$$

If $TV(\mathbf{1}_{A_r \setminus \bigcup_{s\neq r} A_s};\rho^2;D) > TV(\mathbf{1}_{A_r};\rho^2;D)$ for every r then (6.10) and (6.9) would imply that

$$\sum_{s\colon s\neq r} a_{sr} > \sum_{s\colon s\neq r} a_{rs},\ \forall r,$$

which after summing over r would imply

$$\sum_{r=1}^{R}\sum_{s\colon s\neq r} a_{sr} > \sum_{r=1}^{R}\sum_{s\colon s\neq r} a_{rs} = \sum_{r=1}^{R}\sum_{s\colon s\neq r} a_{sr}.$$

This would be a contradiction. Hence there exists at least one r for which (6.8) holds. $\square$

Lemma 6.1.3. *Let D denote an open, bounded domain in $\mathbb{R}^d$ with Lipschitz boundary and let $(Q_1,\dots,Q_R)$ denote a collection of R bounded and mutually disjoint subsets of $\mathbb{R}^d$ that satisfy*

$$\text{(i)}\ TV(\mathbf{1}_{Q_r};\rho^2;\mathbb{R}^d) < +\infty \quad , \quad \text{(ii)}\ |D\mathbf{1}_{Q_r}|_{\rho^2}(\partial D) = 0 \quad \textit{and} \quad \text{(iii)}\ D \subseteq \cup_{r=1}^R Q_r.$$

Then there exists a sequence of mutually disjoint sets $\{A_1^m,\dots,A_R^m\}$ with piecewise smooth boundaries which cover D and satisfy

$$\mathbf{1}_{A_r^m} \xrightarrow{L^1(\mathbb{R}^d)} \mathbf{1}_{Q_r} \quad \textit{and} \quad \lim_{m\to\infty} TV(\mathbf{1}_{A_r^m};\rho^2;D) = TV(\mathbf{1}_{Q_r};\rho^2;D) \tag{6.11}$$

for all $1 \le r \le R$.

Proof. Let $\{\gamma_k\}_{k\in\mathbb{N}}$ denote some sequence of positive reals converging to zero and

$$J_k(x) := \frac{1}{\gamma_k^d} J\left(\frac{|x|}{\gamma_k}\right), \quad J \ge 0, \quad J \in C_c^\infty([0,1]), \quad \int_{\mathbb{R}^d} J(x)\mathrm{d}x = 1,$$

a corresponding sequence of positive, radially symmetric mollifiers. Let $D_k := \{x \in \mathbb{R}^d : \mathrm{dist}(x,D) < \gamma_k\}$ denote the open γ_k-neighborhood of the domain D. For each $k \in \mathbb{N}$ and each Q_r in the collection let

$$u_r^k := J_k * \mathbf{1}_{Q_r}$$

denote a smoothed version of the characteristic function.

For any test function $\Phi \in C_c^1(D:\mathbb{R}^d)$ with $|\Phi(x)| \le \rho^2(x)$, we have

$$\int_D u_r^k \,\mathrm{div}(\Phi(x))\,\mathrm{d}x = -\int_{D_k} \mathbf{1}_{Q_r}\,\mathrm{div}(J_k * \Phi(y))\,\mathrm{d}y \le |D\mathbf{1}_{Q_r}|_{\rho^2}(D_k).$$

The equality follows from the symmetry of J_k and the fact that $J_k * \Phi$ has support within D_k while the inequality follows from the fact that $|J_k * \Phi| \le \rho^2$ so it produces an admissible test function in the definition of the total variation. As a consequence,

$$\limsup_{k\to\infty} TV(u_r^k;D) \le \limsup_{k\to\infty} |D\mathbf{1}_{Q_r}|_{\rho^2}(D_k) = |D\mathbf{1}_{Q_r}|_{\rho^2}(\overline{D}) = |D\mathbf{1}_{Q_r}|_{\rho^2}(D)$$

due to the second assumption in the statement of the lemma. The fact that $u_r^k \xrightarrow{L^1(\mathbb{R}^d)} \mathbf{1}_{Q_r}$ combines with the lower-semicontinuity of the total variation to imply

$$TV(\mathbf{1}_{Q_r};\rho^2;D) \le \liminf_{k\to\infty} TV(u_r^k;\rho^2;D) \le \limsup_{k\to\infty} TV(u_r^k;\rho^2;D) \le TV(\mathbf{1}_{Q_r};\rho^2;D).$$

In other words, these sequences satisfy

$$u_r^k \xrightarrow{L^1(\mathbb{R}^d)} \mathbf{1}_{Q_r}, \qquad TV(u_r^k;\rho^2;D) \to TV(\mathbf{1}_{Q_r};\rho^2;D), \qquad 0 \le u_r^k(x) \le 1 \ \ \forall x \in \mathbb{R}^d.$$

The $(u_1^k, \ldots, u_R^k)$ also satisfy one additional property that will prove useful: there exists a constant $\alpha > 0$ so that

$$\Sigma^k(x) := \sum_{r=1}^{R} u_r^k(x) \geq \alpha > 0 \quad \text{for all} \quad x \in D.$$

To see this, note that the fact that D is an open and bounded set with Lipschitz boundary implies that there exists a cone $C \subseteq \mathbb{R}^d$ with non-empty interior, a family of rotations $\{R_x\}_{x \in D}$ and $\zeta > 0$ such that for every $x \in D$,

$$x + R_x(C \cap B(0, \zeta)) \subseteq D.$$

The fact that J is radially symmetric then implies that for every $x \in D$,

$$\int_D J_k(x-y)dy \geq \int_{x+R_x(C \cap B(0,\zeta))} J_k(x-y)dy = \\ \int_{C \cap B(0,\zeta)} J_k(y)dy = \int_{C \cap B(0, \frac{\zeta}{\gamma_k})} J(y)dy \geq \alpha > 0$$

for some positive constant α. The summation $\Sigma^k(x)$ of all u_r^k therefore satisfies the pointwise estimate

$$\Sigma^k(x) := \sum_{r=1}^{R} u_r^k(x) = \int_{\mathbb{R}^d} J_k(x-y) \sum_{r=1}^{R} \mathbf{1}_{Q_r}(y) \, \mathrm{dy} \geq \int_D J_k(x-y)dy \geq \alpha$$

for all $x \in D$ as claimed.

Step 1: Now, for each u_r^k and each $t \in (0,1)$ consider the superlevel set

$$Q_r^k(t) := \left\{ u_r^k > t \right\}.$$

The first claim is that, for any fixed t in $(0,1)$, the characteristic function $\mathbf{1}_{Q_r^k(t)}$ converges in $L^1(\mathbb{R}^d)$ to the characteristic function of the original set. To see this, note that

$$Q_r^k(t) \setminus Q_r \subset \left\{ |u_r^k - \mathbf{1}_{Q_r}| \geq t \right\}.$$

By Chebyshev's/Markov's inequality, if $\mathscr{L}_d$ denotes Lebesgue measure in $\mathbb{R}^d$ then

$$\mathscr{L}_d(Q_r^k(t) \setminus Q_r) \leq \mathscr{L}_d\left(\left\{ |u_r^k - \mathbf{1}_{Q_r}| > t \right\} \right) \leq \frac{1}{t} \|u_r^k - \mathbf{1}_{Q_r}\|_{L^1} \to 0.$$

In a similar fashion,

$$\mathscr{L}_d(Q_r \setminus Q_r^k(t)) \leq \mathscr{L}_d\left(\left\{ |u_r^k - \mathbf{1}_{Q_r}| \geq (1-t) \right\} \right) \leq \frac{1}{1-t} \|u_r^k - \mathbf{1}_{Q_r}\|_{L^1} \to 0.$$

As a consequence, it follows that

$$\int_{\mathbb{R}^d} |\mathbf{1}_{Q_r^k(t)} - \mathbf{1}_{Q_r}| \,\mathrm{d}x = \mathscr{L}_d(Q_r^k(t) \setminus Q_r) + \mathscr{L}_d(Q_r \setminus Q_r^k(t)) \to 0$$

as claimed.

Step 2: The next claim is that there exists a set $\mathscr{T} \subset (0,1)$ of full Lebesgue measure with the following property: if $t \in \mathscr{T}$ then $Q_r^k(t)$ has a smooth boundary for all k and all sets $(Q_1^k(t), \ldots, Q_R^k(t))$ in the collection. To see this, note Sard's lemma (see for example [50]) implies that for any fixed $k \in \mathbb{N}$ the set $Q_r^k(t)$ has smooth boundary up to an exceptional set $\mathscr{T}_{k,r} \subset (0,1)$ of Lebesgue measure zero. Now define the set $\mathscr{T}$ as

$$\mathscr{T} = (0,1) \setminus \bigcup_{k=1}^{\infty} \bigcup_{r=1}^{R} \mathscr{T}_{k,r}.$$

Note that $\mathscr{T}$ has full measure since a countable union of Lebesgue-null sets has measure zero. If $t \in \mathscr{T}$ then it does not lie in any of the exceptional sets, meaning that for each k and each r the set $Q_r^k(t)$ has a smooth boundary.

Step 3: We use a diagonal argument to construct an approximating sequence of partitions that are not necessarily disjoint, but satisfy the hypotheses of Lemma (6.7).

For the set Q_1, Step 1 and lower semi-continuity of the total variation imply that for all $t \in (0,1)$

$$TV(\mathbf{1}_{Q_1};\rho^2;D) \leq \liminf_{k\to\infty} TV(\mathbf{1}_{Q_1^k(t)};\rho^2;D).$$

On the other hand, Fatou's lemma combined with the co-area formula imply

$$\int_0^1 \liminf_{k\to\infty} TV(\mathbf{1}_{Q_1^k(t)};\rho^2;D)\,\mathrm{d}t \leq \lim_{k\to\infty} \int_0^1 TV(\mathbf{1}_{Q_1^k(t)};\rho^2;D)\,\mathrm{d}t = \lim_{k\to\infty} TV(u_1^k;\rho^2;D) = TV(\mathbf{1}_{Q_1};\rho^2;D).$$

In other words,

$$TV(\mathbf{1}_{Q_1};\rho^2;D) \leq \liminf_{k\to\infty} TV(\mathbf{1}_{Q_1^k(t)};\rho^2;D)$$

and

$$\int_0^1 \liminf_{k\to\infty} TV(\mathbf{1}_{Q_1^k(t)};\rho^2;D)\,\mathrm{d}t = TV(\mathbf{1}_{Q_1};\rho^2;D),$$

which imply

$$\liminf_{k\to\infty} TV(\mathbf{1}_{Q_1^k(t)};\rho^2;D) = TV(\mathbf{1}_{Q_1};\rho^2;D)$$

almost everywhere. In particular, there exists a $t_1 \in \mathscr{T}$ with $0 < t_1 < \alpha/R$ and a subsequence $\{k_m\}_{m\in\mathbb{N}}$ with the property that

$$\partial Q_1^{k_m}(t_1) \text{ is smooth } \forall m, \quad \lim_{m\to\infty} TV(\mathbf{1}_{Q_1^{k_m}(t_1)};\rho^2;D) = TV(\mathbf{1}_{Q_1};\rho^2;D), \quad \mathbf{1}_{Q_1^{k_m}(t_1)} \xrightarrow{L^1(\nu)} \mathbf{1}_{Q_1}. \tag{6.12}$$

We now pass to the set Q_2. As $\partial Q_1^{k_m}(t_1)$ is smooth and bounded for all m, it has zero Lebesgue measure for all m in particular. As $u_2^{k_m}$ is smooth, Lemma 2.95 in [6] implies that

$$\mathscr{H}^{d-1}\left(\partial Q_1^{k_m}(t_1)\cap\partial Q_2^{k_m}(t)\right)=0$$

for almost every $t\in(0,1)$. Let $\mathscr{T}_{2,m}$ denote the m^{th} exceptional set for which this property does not hold. Define the set

$$\mathscr{T}_2:=\mathscr{T}\setminus\bigcup_{m=1}^{\infty}\mathscr{T}_{2,m},$$

which has full Lebesgue measure. By definition, if $t\in\mathscr{T}_2$ then $\partial Q_2^{k_m}(t)$ is smooth for all m and

$$\mathscr{H}^{d-1}\left(\partial Q_1^{k_m}(t_1)\cap\partial Q_2^{k_m}(t)\right)=0$$

for all m as well. Along the subsequence $\{k_m\}$, the lower semi-continuity property still holds,

$$TV(\mathbf{1}_{Q_2};\rho^2;D)\leq\liminf_{k\to\infty}TV(\mathbf{1}_{Q_2^{n_k}(t)};\rho^2;D),$$

as does the argument based on Fatou's lemma and the co-area formula. In particular, there exists a further subsequence $\{k_{m_l}\}_{l\in\mathbb{N}}$ and a $t_2\in\mathscr{T}_2$ with $0<t_2<\alpha/R$ so that (6.12) holds along this subsequence. The analogous properties hold for the sets $\{Q_2^{k_{m_l}}(t_2)\}$ as well. Moreover, the relation

$$\mathscr{H}^{d-1}\left(\partial Q_1^{k_{m_l}}(t_1)\cap\partial Q_2^{k_{m_l}}(t_2)\right)=0$$

also holds along this subsequence. By extracting $(R-2)$ more subsequences in this way, we obtain a subsequence taht we denote simply by k_m of the original sequence together with a sequence of sets $Q_r^{k_m}(t_r)$ with $0<t_r<\alpha/R$ that satisfy

$$\begin{aligned}&\partial Q_r^{k_m}(t_r)\text{ is smooth }\forall m,\quad\lim_{m\to\infty}TV(\mathbf{1}_{Q_r^{k_m}(t_r)};\rho^2;D)=TV(\mathbf{1}_{Q_r};\rho^2;D),\quad\mathbf{1}_{Q_r^{k_m}(t_r)}\xrightarrow{L^1(\nu)}\mathbf{1}_{Q_r},\\&\mathscr{H}^{d-1}\left(\partial Q_r^{k_m}(t_r)\cap\partial Q_s^{k_m}(t_s)\right)=0\end{aligned}\tag{6.13}$$

for all m and all $r\neq s$.

Step 4: We now use the sets constructed in the previous step and lemma 6.1.2 to complete the proof. Let $Q_r^m:=Q_r^{k_m}(t_r)$. We claim that the sets $(Q_1^m,\dots,Q_R^m)$ cover D as well. To see this, suppose there exists

$$x\in D\setminus\left(\bigcup_{r=1}^{R}Q_r^m\right).$$

This would imply that $u_r^{k_m}(x)\leq t_r$ for all r by definition. In turn,

$$\Sigma^{k_m}(x)\leq\sum_{r=1}^{R}t_r<\alpha,$$

which contradicts the estimate on Σ^{k_m} obtained earlier. Due to (6.13) and Lemma 6.1.2, for each $m \in \mathbb{N}$ there exists a permutation $\pi_m : \{1,\ldots,R\} \to \{1,\ldots,R\}$ with the property that

$$TV\left(\mathbf{1}_{A_r^m};\rho^2;D\right) \le TV(\mathbf{1}_{Q_r^m};\rho^2;D)$$

for all $1 \le r \le R$, where A_r^m denotes the set

$$A_r^m := Q_r^m \setminus \bigcup_{s=\pi_m^{-1}(r)+1}^{R} Q_{\pi_m(s)}^m.$$

Each A_r^m has a piecewise smooth boundary for all m due to the fact that each Q_r^m has a smooth boundary. The disjointness of $(Q_1,\ldots,Q_R)$ combines with the L^1-convergence of $\mathbf{1}_{Q_r^m}$ to $\mathbf{1}_{Q_r}$ to show that

$$\mathbf{1}_{A_r^m} \overset{L^1(\mathbb{R}^d)}{\longrightarrow} \mathbf{1}_{Q_r}$$

as well. This combines with lower semi-continuity of the total variation to imply

$$\begin{aligned} TV(\mathbf{1}_{Q_r};\rho^2;D) &\le \liminf_{m\to\infty}; TV\left(\mathbf{1}_{A_r^m};\rho^2;D\right) \\ &\le \limsup_{m\to\infty} TV\left(\mathbf{1}_{A_r^m};\rho^2;D\right) \le \limsup_{m\to\infty} TV(\mathbf{1}_{Q_r^m};\rho^2;D) = TV(\mathbf{1}_{Q_r};\rho^2;D). \end{aligned}$$

Finally, noting that

$$D \subset \bigcup_{r=1}^{R} Q_r^m = \bigcup_{r=1}^{R} A_r^m$$

and that the A_r^m are pairwise disjoint yields the claim. □

To complete the construction we intended at the beginning at this section, we we need to verify the hypotheses $(\mathrm{i}-\mathrm{ii})$ of the previous lemma. This is the content of our final lemma.

Lemma 6.1.4. *Let D be an open bounded domain with Lipschitz boundary and let $\{A_1,\ldots,A_R\}$ denote a disjoint collection of sets that satisfy*

$$A_r \subseteq D \quad \textit{and} \quad TV(\mathbf{1}_{A_r};\rho^2;D) < \infty.$$

Then, there exists a disjoint collection of bounded sets $(Q_1,\ldots,Q_R)$ that satisfy $Q_r \cap D = A_r$ together with the properties

$$\text{(i)}\ \ TV(\mathbf{1}_{Q_r};\rho^2;\mathbb{R}^d) < +\infty \quad \textit{and} \quad \text{(ii)}\ \ |D\mathbf{1}_{Q_r}|_{\rho^2}(\partial D) = 0.$$

Proof. The proof follows from Remark 3.43 in [6] (which with minimal modifications applies to total variation with weight ρ^2). □

6.2 Γ-convergence

Let us first establish Proposition 6.0.7. We start with a Lemma which is the multi-class analogue of Lemmas 5.2.2 and 5.2.5 combined.

Lemma 6.2.1. (i) *If* $\mathcal{U}_k \to \mathcal{U}$ *in* $(L^1(D,\nu))^R$ *then* $B(u_r^k) \to B(u_r)$ *for all* $1 \le r \le R$. (ii) *The set* $\mathcal{M}(D)$ *is closed in* $L^1(D,\nu)$. (iii) *If* $\{\mathcal{U}_n\}$ *is a sequence with* $\mathcal{U}_n \in (L^1(\nu_n))^R$ *and* $\mathcal{U}_n \xrightarrow{TL^1} \mathcal{U}$ *for some* $\mathcal{U} \in (L^1(D,\nu))^R$*, then* $B_n(u_r^n) \to B(u_r)$ *for all* $1 \le r \le R$. (iv) *If* $u_n = \tilde{\mathbf{1}}_{Y_n}$*, where* $Y_n \subset V_n$*, converges to* $u = \tilde{\mathbf{1}}_A$ *in the* TL^1*-sense, then* $\mathbf{1}_{Y_n}$ *converges to* $\mathbf{1}_A$ *in the* TL^1*-sense.*

Proof. Statements (i), (iii) and (iv) follow directly from the proof of Proposition 5.2.5. In order to prove the second statement, suppose that a sequence $\{\mathcal{U}_k\}_{k\in\mathbb{N}}$ in $\mathcal{M}(D)$ converges to some $\mathcal{U}$ in $(L^1(D,\nu))^R$. We need to show that $\mathcal{U} \in \mathcal{M}(D)$. First of all note that for every $1 \le r \le R$, $u_r^k \xrightarrow{L^1(D,\nu)} u_r$. Since $u_r^k \in \mathrm{Ind}(D)$ for every $k \in \mathbb{N}$, and since $\mathrm{Ind}(D)$ is a closed subset of $L^1(D,\nu)$ (by Proposition 5.2.5), we deduce that $u_r \in \mathrm{Ind}(D)$ for every r.

The orthogonality condition follows from Fatou's lemma. In fact, working along a subsequence we can without the loss of generality assume that for every r, $u_r^k \to u_r$ for almost every x in D. Hence, for $r \neq s$ we have

$$0 \le \int_D u_r(x)u_s(x)d\nu(x) = \int_D \liminf_{k\to\infty}(u_r^k(x)u_s^k(x))\mathrm{d}\nu(x) \le \liminf_{k\to\infty}\int_D u_r^k(x)u_s^k(x)d\nu(x) = 0.$$

Now let us write $u_r^k = \mathbf{1}_{A_r^k}/B(\mathbf{1}_{A_r^k})$ and $u_r = \mathbf{1}_{A_r^k}/B(\mathbf{1}_{A_r})$. As in the proof of Proposition (5.2.5) we must have $B(\mathbf{1}_{A_r^k}) \to B(\mathbf{1}_{A_r})$ as $k \to \infty$. Thus, for almost every $x \in D$

$$\sum_{r=1}^R u_r(x) = \lim_{k\to\infty}\sum_{r=1}^R u_r^k(x) \ge \lim_{k\to\infty}\min_{r=1,\dots,R}\frac{1}{B(\mathbf{1}_{A_r^k})} = \min_{r=1,\dots,R}\frac{1}{B(\mathbf{1}_{A_r})} > 0.$$

□

Proof of Proposition 6.0.7. Liminf inequality. The proof of (6.5) follows the approach used in the two-class case. Let $\mathcal{U}_n \xrightarrow{TL^1} \mathcal{U}$ denote an arbitrary convergent sequence. As $\mathcal{M}(D)$ is closed, if $\mathcal{U} \notin \mathcal{M}(D)$ then as in the two-class case, it is easy to see that $\mathcal{U}_n \notin \mathcal{M}_n(D)$ for all n sufficiently large. The inequality (6.5) is then trivial in this case, as both sides of it are equal to infinity. Conversely, if $\mathcal{U} \in \mathcal{M}(D)$ then we may assume that $\mathcal{U}_n \in \mathcal{M}_n(D)$ for all n, since only those terms with $\mathcal{U}_n \in \mathcal{M}_n(D)$ can contribute non-trivially to the limit inferior. In this case we easily have

$$\begin{aligned}\liminf_{n\to\infty} E_n(\mathcal{U}_n) = \liminf_{n\to\infty}\sum_{r=1}^R GTV_{n,\varepsilon_n}(u_r^n) &\ge \sum_{r=1}^R \liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_r^n)\\ &\ge \sigma_\eta \sum_{r=1}^R TV(u_r;\rho^2) = \sigma_\eta E(\mathcal{U}).\end{aligned}$$

The last inequality follows from Theorem 1.0.4. This establishes the first statement in Proposition 6.0.7.

Limsup inequality. We now turn to the proof of (6.6), Borrowing terminology from the Γ-convergence literature, we say that $\mathscr{U} \in (L^1(D,\nu))^R$ has a *recovery sequence* when there exists a sequence $\mathscr{U}_n \in (L^1(\nu_n))^R$ such that (6.6) holds. To show that each $\mathscr{U} \in (L^1(\nu))^R$ has a recovery sequence, we first remark that we can assume that $E(\mathscr{U}) < \infty$ and that due to Remark 1.1.2 and the results from Section 6.1, we can assume that $\mathscr{U}$ is of the form $(u_1,\dots,u_R)$, where $u_r = \tilde{\mathbf{1}}_{A_r}$ and where the partition $\{A_1,\dots,A_R\}$ is induced by piecewise smooth sets $(A_r = D \cap Q_r)$.

Let $c_0 := \max\{B(\mathbf{1}_{A_1}),\dots,B(\mathbf{1}_{A_R})\}$ denote the size of the largest set in the collection. The fact that $E(\mathscr{U}) < \infty$ then implies

$$TV(\mathbf{1}_{A_r};\rho^2) \le c_0\, TV(u_r;\rho^2) \le c_0 \sum_{r=1}^{R} TV(u_r;\rho^2) < \infty,$$

so that all sets $\{A_1,\dots,A_R\}$ in the collection defining $\mathscr{U}$ have finite perimeter. Additionally because $\mathscr{U} \in \mathscr{M}(D)$ implies that any two sets A_r, A_s with $r \neq s$ have empty intersection up to a Lebesgue-null set, we may freely assume without the loss of generality that the sets $\{A_1,\dots,A_R\}$ are mutually disjoint.

Let $Y_r^n = A_r \cap V_n$ denote the restriction of A_r to the first n data points. We consider the transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from Proposition 1.0.9. We let A_n^r be the set for which $\mathbf{1}_{A_r^n} = \mathbf{1}_{Y_n^r} \circ T_n$.

We first notice that the fact that Q_r has a piecewise smooth boundary in $\mathbb{R}^d$ and the fact that $||Id - T_n||_{L^\infty(D)} \to 0$, imply that

$$||\mathbf{1}_{A_r^n} - \mathbf{1}_{A_r}||_{L^1(D,\nu)} \le C_0(Q_r)\, ||Id - T_n||_{L^\infty(D)}, \tag{6.14}$$

where $C_0(Q_r)$ denotes some constant that depends on the set Q_r. This inequality follows from the formulas for the volume of tubular neighborhoods (see [74]). In particular, note that by the change of variables (2.4) we have, $|Y_r^n| = |A_r^n| \to |A_r|$ as $n \to \infty$, so that in particular we can assume that $|Y_r^n| \neq 0$. We define $u_r^n := \mathbf{1}_{Y_r^n}/|Y_r^n|$ as the corresponding normalized indicator function. We claim that $\mathscr{U}_n := (u_1^n,\dots,u_R^n)$ furnishes the desired recovery sequence.

To see that $\mathscr{U}_n \in \mathscr{M}_n(D)$ we first note that each $u_r^n \in \mathrm{Ind}_n(D)$ by construction. On the other hand, the fact that $\{A_1,\dots,A_R\}$ forms a partition of D implies that $\{Y_1^n,\dots,Y_R^n\}$ defines a partition of V_n. As a consequence,

$$E_n(\mathscr{U}_n) = \sum_{r=1}^{R} GTV_{n,\varepsilon_n}(u_r^n)$$

by definition of the E_n functionals.

Using (6.14), we can proceed as in Remark 5.1 5.1.1. In particular, we can assume that η has the form $\eta(z) = a$ for $z < b$ and $\eta(z) = 0$ otherwise. We set $\tilde{\varepsilon}_n := \varepsilon_n + \frac{2}{b}||Id - T_n||_{L^\infty(D)}$.

Recall that by assumption $||Id - T_n||_\infty \ll \varepsilon_n$, and thus $\tilde{\varepsilon}_n$ is a small perturbation of ε_n. As in the proof of Theorem 1.0.4, we have

$$\frac{\varepsilon_n^{d+1}}{\tilde{\varepsilon}_n^{d+1}} GTV_{n,\varepsilon_n}(\mathbf{1}_{Y_r^n}) \leq TV_{\tilde{\varepsilon}_n}(\mathbf{1}_{A_r^n};\rho).$$

A straightforward computation shows that there exists a constant K_0 such that

$$|TV_{\tilde{\varepsilon}_n}(\mathbf{1}_{A_r^n};\rho) - TV_{\tilde{\varepsilon}_n}(\mathbf{1}_{A_r};\rho)| \leq \frac{K_0}{\tilde{\varepsilon}_n}\|\mathbf{1}_{A_r^n} - \mathbf{1}_{A_r}\|_{L^1(D,\nu)} \leq K_0 C_0(Q_r)\frac{||Id - T_n||_{L^\infty(D)}}{\tilde{\varepsilon}_n}.$$

Since $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$, the previous inequalities imply that

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(\mathbf{1}_{Y_r^n}) \leq \limsup_{n\to\infty} TV_{\tilde{\varepsilon}_n}(\mathbf{1}_{A_r^n};\rho) = \limsup_{n\to\infty} TV_{\tilde{\varepsilon}_n}(\mathbf{1}_{A_r};\rho).$$

Finally, from (1.21) we deduce that

$$\limsup_{n\to\infty}\ TV_{\tilde{\varepsilon}_n}(\mathbf{1}_{A_r^n};\rho) \leq \sigma_\eta TV(\mathbf{1}_{A_r};\rho^2),$$

and thus we conclude that $\limsup_{n\to\infty}\ GTV_{n,\varepsilon_n}(\mathbf{1}_{A^r}) \leq \sigma_\eta TV(\mathbf{1}_{A_r};\rho^2)$. As a consequence we have

$$\limsup_{n\to\infty}\ GTV_{n,\varepsilon_n}(u_r^n) = \limsup_{n\to\infty}\ \frac{GTV_{n,\varepsilon_n}(\mathbf{1}_{Y_r^n})}{B_n(\mathbf{1}_{Y_r^n})} \leq \sigma_\eta \frac{TV(\mathbf{1}_{A_r};\rho^2)}{B(\mathbf{1}_{A_r})}$$

for each r, by continuity of the balance term. From the previous computations we conclude that $E_n(\mathscr{U}_n) \to E(\mathscr{U})$, and from 6.14, we deduce that $\mathscr{U}_n \to \mathscr{U}$ in the TL^1-sense, so that $\mathscr{U}_n$ does furnish the desired recovery sequence. □

Having Proposition 6.0.7, Proposition 6.0.9 can be obtained by similar arguments to the ones we used in the two-class case. With Proposition 6.0.7 and Proposition 6.0.9, the arguments presented in Section 5.2 can be adapted in a straightforward way to complete the proof of Theorem 1.0.10.

Chapter 7

Pointwise convergence of graph perimeter

The purpose of this chapter is to present some results on the pointwise convergence of the graph perimeter towards continuous perimeter. For simplicity we consider $D = (0,1)^d$ and $\rho \equiv 1$. Moreover, the graph G_n based on $\mathbf{x}_1, \dots, \mathbf{x}_n$ uniformly distributed on $(0,1)^d$ is constructed using the kernel $\boldsymbol{\eta}$ which is given by $\boldsymbol{\eta}(t) = 1$ if $0 \le t \le 1$ and $\boldsymbol{\eta}(t) = 0$ if $t > 1$. We let σ_d be the surface tension associated to this kernel, which is found to be

$$\sigma_d := \frac{2 s_{d-2}}{(d+1)(d-1)} \tag{7.1}$$

where s_{d-2} is the area of the $d-2$-dimensional unit sphere (the boundary of the unit ball in $\mathbb{R}^{d-1}$). In figure 7.1 below we illustrate how random geometric graphs are used to estimate the perimeter with respect to D of a fixed set Q. Throughout this section, we use $\mathrm{Per}_{n,\varepsilon_n}(Q)$ to represent $\mathrm{Per}_{n,\varepsilon_n}(Q \cap V_n)$.

We show the following.

Theorem 7.0.2. *Let $p \ge 1$ and and let $Q \subseteq D$ be a set with finite perimeter. Assume $\varepsilon_n \to 0$ as $n \to \infty$. Then*

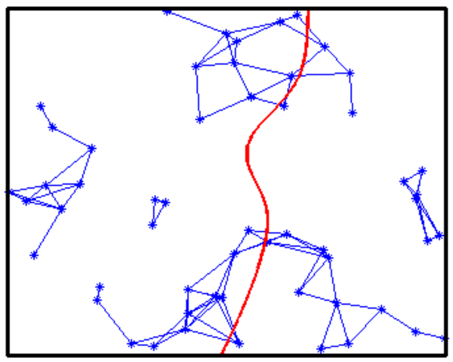
$$\mathbb{E}(|\mathrm{Per}_{n,\varepsilon_n}(Q) - \mathbb{E}(\mathrm{Per}_{n,\varepsilon_n}(Q))|^p) \le C\,(f(n))^p \tag{7.2}$$

where

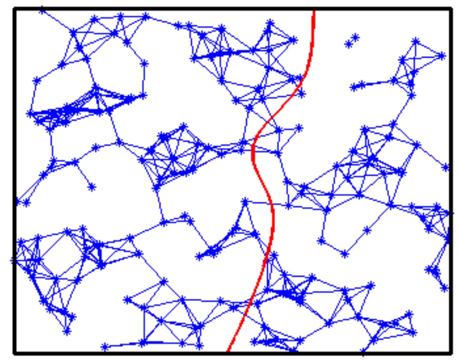
$$f(n) := \begin{cases} \frac{1}{\sqrt{n\varepsilon_n}} & \text{if } \frac{1}{n^{1/d}} \le \varepsilon_n \\ \frac{1}{n\varepsilon_n^{(d+1)/2}} & \text{if } \frac{1}{n^{2/(d+1)}} \le \varepsilon_n \le \frac{1}{n^{1/d}}. \end{cases} \tag{7.3}$$

and where $C = C(p,Q)$ is a constant that depends only on p and the perimeter $\mathrm{Per}(Q)$ of Q. In particular, if $n^{-\frac{2}{(d+1)}} \ll \varepsilon_n \ll 1$, then

$$\mathrm{Per}_{n,\varepsilon_n}(Q) \to \sigma_d \,\mathrm{Per}(Q), \text{ almost surely as } n \to \infty.$$



Graph cut with $n = 60$ and $\varepsilon = 0.15$.



Graph cut with $n = 200$ and $\varepsilon = 0.1$.

Figure 7.1: The graph perimeter is the appropriately rescaled number of edges between Q and Q^c. The red line represents the boundary of Q in D

Note that $\mathbb{E}(\mathrm{Per}_{n,\varepsilon_n}(Q)) = \mathrm{Per}_{\varepsilon_n}(Q) = TV_{\varepsilon_n}(\mathbf{1}_Q)$. Thus, the last part of the previous theorem follows from (1.21), the moment estimates (7.2), Markov's inequality, and the Borel-Cantelli lemma which imply that

$$\mathrm{Per}_{n,\varepsilon_n}(Q) - \mathbb{E}\left(\mathrm{Per}_{n,\varepsilon_n}(Q)\right) \to 0 \quad \text{a.s.}$$

We note that this a.s. convergence holds for rather sparse graphs. Namely the typical degree of a node is $\alpha_d n \varepsilon^d$, where α_d is the volume of the unit ball in d dimensions. When $n^{-\frac{2}{(d+1)}} \ll \varepsilon_n \ll n^{-\frac{1}{d}}$ the a.s. convergence holds, while the average degree of a vertex converges to zero. The convergence is still possible because the expected number of edges crossing $\partial_D Q = \partial Q \cap D$ goes to infinity. The relevant point is that the pointwise convergence holds for very sparse graphs, where consistency of optimal balanced cuts on the graphs towards an optimal domain cut is impossible.

After finding moment estimates for $\mathrm{Per}_{n,\varepsilon_n}(Q)$, we concentrate on finding explicit estimates for $|\mathrm{Per}_{\varepsilon_n}(Q) - \sigma_d \,\mathrm{Per}(Q)|$, that is we find bias estimates. To obtain these estimates we assume that Q is a set with smooth relative boundary. It proves straightforward to check that $|\mathrm{Per}_{\varepsilon_n}(Q) - \sigma_d \,\mathrm{Per}(Q)| = O(\varepsilon_n)$ for general subsets $Q \subseteq [0,1]^d$ with smooth relative boundary. We show in Section 7.2 that the error is actually quadratic in ε_n

$$|\mathbb{E}(\mathrm{Per}_{n,\varepsilon_n}(Q)) - \sigma_d \,\mathrm{Per}(Q)| = |\mathrm{Per}_{\varepsilon_n}(Q) - \sigma_d \,\mathrm{Per}(Q)| = O(\varepsilon_n^2) \tag{7.4}$$

under the extra condition that $Q \subset\subset D$.

Lemma 7.0.3. *Let Q be a set with smooth boundary, such that* $\mathrm{dist}(Q, \partial D) > 0$. *Then*

$$\mathrm{Per}_\varepsilon(Q) = \sigma_d \,\mathrm{Per}(Q) + O(\varepsilon^2). \tag{7.5}$$

Combining the bias and variance estimates allows us to obtain the rates of convergence for the error $|\mathrm{Per}_{n,\varepsilon_n}(Q) - \sigma_d \mathrm{Per}(Q)|$. In particular we estimate the "standard deviation"

$$\mathrm{std}(n) := \mathbb{E}\left((\mathrm{Per}_{n,\varepsilon_n}(Q) - \sigma_d \mathrm{Per}(Q))^2\right)^{1/2},$$

which we may quantify precisely by using the variance-bias decomposition

$$\mathrm{std}^2(n) = \mathrm{Var}(\mathrm{Per}_{n,\varepsilon_n}(Q)) + (\mathbb{E}(\mathrm{Per}_{n,\varepsilon_n}(Q)) - \sigma_d \mathrm{Per}(Q))^2.$$

Using the special case $p = 2$ of Theorem 7.0.2 to estimate for the variance and using Lemma 7.0.3 to estimate the bias we obtain the following.

Theorem 7.0.4. *Let $Q \subset D$ be an open set with smooth boundary such that $Q \subset\subset (0,1)^d$. Assume that $n^{-\frac{2}{d+1}} \leq \varepsilon_n \ll 1$. Consider $f(n)$ defined via* (7.3)*. Then, the error of approximating $\sigma_d \mathrm{Per}(Q)$ by $\mathrm{Per}_{n,\varepsilon_n}(Q)$ satisfies*

$$\mathrm{std}(n) = O(f(n) + \varepsilon_n^2).$$

In general, if $Q \subseteq D$ is an open set with smooth boundary such that $\partial Q \setminus [0,1]^d \neq \emptyset$, then

$$\mathrm{std}(n) = O(f(n) + \varepsilon_n).$$

7.1 Moment estimates

We establish Theorem 7.0.2. In order to understand the asymptotic behavior of the graph perimeter $\mathrm{Per}_{n,\varepsilon_n}(Q)$ for fixed Q, we first define a symmetric kernel $\phi_\varepsilon : D \times D \to (0,\infty)$ by

$$\phi_\varepsilon(x,y) = \frac{\mathbf{1}_{\{|x-y| \leq \varepsilon\}}}{\varepsilon^{d+1}} |\mathbf{1}_Q(x) - \mathbf{1}_Q(y)|.$$

Using the kernel ϕ_{ε_n}, we can then write $\mathrm{Per}_{n,\varepsilon_n}(Q)$ as

$$\mathrm{Per}_{n,\varepsilon_n}(Q) = \frac{2}{n(n-1)\varepsilon_n^{d+1}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \phi_{\varepsilon_n}(\mathbf{x}_i, \mathbf{x}_j), \tag{7.6}$$

which is a *U-statistic* in the terminology of Hoeffding [47].

Let us first note that Hoeffding's decomposition theorem for U-statistics of order two (see [48]) implies that $\mathrm{Per}_{n,\varepsilon_n}(Q)$ can be written as:

$$\mathrm{Per}_{n,\varepsilon_n}(Q) - \mathrm{Per}_{\varepsilon_n}(Q) = 2U_{n,1} + U_{n,2}, \tag{7.7}$$

where $U_{n,1}$ is a U-statistic of order one (just a sum of centered independent random variables) and $U_{n,2}$ is a U-statistic of order two which is *canonical* or completely degenerate (see [48]). In order to define the variables $U_{n,1}$ and $U_{n,2}$, let us introduce the functions

$$\begin{aligned} \bar{\phi}_{\varepsilon}(x) &:= \int_D \phi_{\varepsilon}(x,z)\mathrm{d}z, && x \in D, \\ g_{n,1}(x) &:= \bar{\phi}_{\varepsilon_n}(x) - \mathrm{Per}_{\varepsilon_n}(Q), && x \in D, \\ g_{n,2}(x,y) &:= \phi_{\varepsilon_n}(x,y) - \bar{\phi}_{\varepsilon_n}(x) - \bar{\phi}_{\varepsilon_n}(y) + \mathrm{Per}_{\varepsilon_n}(Q), && x,y \in D. \end{aligned} \tag{7.8}$$

With the previous definitions, we can now define

$$\begin{aligned} U_{n,1} &= \frac{1}{n}\sum_{i=1}^{n} g_{n,1}(\mathbf{x}_i), \\ U_{n,2} &= \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} g_{n,2}(\mathbf{x}_i,\mathbf{x}_j). \end{aligned} \tag{7.9}$$

We remark that $\int_D g_{n,1}(z)\mathrm{d}z = 0$ and that $\int_D g_{n,2}(x,z)\mathrm{d}z$ for all $x \in D$. Because of this, $U_{n,1}$ and $U_{n,2}$ are said to be canonical statistics of order one and two respectively (see [48]). Now, Bernstein's inequality [15] implies that

$$\mathbb{E}(|U_{n,1}|^p) \leq \frac{C_p}{n^p} \max\left(A_{n,1}^p, B_{n,1}^p\right), \tag{7.10}$$

where

$$A_{n,1} := ||g_{n,1}||_{L^\infty(D)}, \quad B_{n,1} := \sqrt{n}||g_{n,1}||_{L^2(D)}. \tag{7.11}$$

and C_p is a universal constant. See also [38] for a slight generalization of the previous result.

On the other hand some of the moment estimates in [38] for canonical U-statistics of order two can be used to prove that

$$\mathbb{E}\left(|U_{n,2}|^p\right) \leq \frac{C_p}{n^{2p}} \max\left(A_{n,2}^p, B_{n,2}^p, C_{n,2}^p\right), \tag{7.12}$$

where

$$A_{n,2} := ||g_{n,2}||_{L^\infty(D)}, \quad B_{n,2} := n||g_{n,2}||_{L^2(D)}, \quad (C_{n,2})^2 := n|| \int_D g_{n,2}^2(\cdot,y)\mathrm{d}y||_{L^\infty(D)}. \tag{7.13}$$

and C_p is a universal constant. From the decomposition (7.7) it follows that for $p \geq 1$

$$\mathbb{E}\left(|\mathrm{Per}_{n,\varepsilon_n}(Q) - \mathrm{Per}_{\varepsilon_n}(Q)|^p\right) \leq C_p(\mathbb{E}\left(|U_{n,1}|^p\right) + \mathbb{E}\left(|U_{n,2}|^p\right)).$$

Thus in order to obtain the moment estimates for $\mathrm{Per}_{n,\varepsilon_n}(Q)$ in Theorem 7.0.2, we focus on finding estimates for the quantities in (7.11) and (7.13).

We first compute the moments of $U_{n,1}$ and so we start computing the quantities $A_{n,1}$ and $B_{n,1}$ from (7.11). Denote by T_ε the ε-*tube* around ∂Q, that is, consider the set

$$T_\varepsilon := \left\{ x \in \mathbb{R}^d \,:\, \mathrm{dist}(x, \partial Q) \le \varepsilon \right\}. \tag{7.14}$$

We also consider the *half tubes* T_ε^- and T_ε^+,

$$T_\varepsilon^- := \{ x \in Q \,:\, \mathrm{dist}(x, \partial Q) \le \varepsilon \}, \quad T_\varepsilon^+ := \{ x \in Q^c \,:\, \mathrm{dist}(x, \partial Q) \le \varepsilon \}. \tag{7.15}$$

With these definitions it is straightforward to check that

$$\bar{\phi}_{\varepsilon_n}(x) = \begin{cases} |B_d(x, \varepsilon_n) \cap Q| / \varepsilon_n^{d+1} & \text{if } x \in T_{\varepsilon_n}^+ , \\ |B_d(x, \varepsilon_n) \cap Q^c| / \varepsilon_n^{d+1} & \text{if } x \in T_{\varepsilon_n}^- , \\ 0 & \text{if } x \notin T_{\varepsilon_n} . \end{cases} \tag{7.16}$$

Since $|B_d(x, \varepsilon_n) \cap Q|$ and $|B_d(x, \varepsilon_n) \cap Q^c|$ are bounded by $\alpha_d \varepsilon_n^d$, where α_d is the volume of the d-dimensional unit ball, we deduce that

$$A_{n,1} = O\left(\frac{1}{\varepsilon_n} \right).$$

In order to compute the quantity $B_{n,1}$ we use the following lemma.

Lemma 7.1.1. *Let $p \ge 1$ and let $Q \subseteq D$, be a set with finite perimeter. Then, for all $\varepsilon > 0$ we have*

$$\int_D \bar{\phi}_\varepsilon^p(x) \mathrm{d}x \le \frac{\alpha_d^{p-1} \sigma_d}{\varepsilon^{p-1}} \mathrm{Per}(Q).$$

Proof. The proof follows the same argument used to establish the limsup inequality in Proposition 1.0.7 or Theorem 6.2 in [4]. We assume that $\mathrm{dist}(Q, \partial D) > 0$. We remark that a slight modification of the argument we present below proves the result in the general case and hence we omit the details.

First we prove that for any function $u : \mathbb{R}^d \to [0,1]$ with $u \in W^{1,1}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$ and for all $\varepsilon > 0$ we have

$$\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \frac{\mathbf{1}_{|x-y| \le \varepsilon}}{\varepsilon^{d+1}} |u(y) - u(x)| \mathrm{d}y \right)^p \mathrm{d}x \le \frac{\alpha_d^{p-1} \sigma_d}{\varepsilon^{p-1}} \int_{\mathbb{R}^d} |\nabla u(x)| \mathrm{d}x, \tag{7.17}$$

Inequality (7.17) follows from

$$
\begin{aligned}
&\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\frac{\mathbf{1}_{|x-y|\leq\varepsilon}}{\varepsilon^{d+1}}|u(y)-u(x)|\mathrm{d}y\right)^p\mathrm{d}x=\frac{1}{\varepsilon^p}\int_{\mathbb{R}^d}\left(\int_{B_d(0,1)}|u(x+\varepsilon h)-u(x)|\mathrm{d}h\right)^p\mathrm{d}x\\
&\leq\frac{\alpha_d^{p-1}}{\varepsilon^p}\int_{\mathbb{R}^d}\int_{B_d(0,1)}|u(x+\varepsilon h)-u(x)|^p\mathrm{d}h\,\mathrm{d}x\leq\frac{\alpha_d^{p-1}}{\varepsilon^p}\int_{\mathbb{R}^d}\int_{B_d(0,1)}|u(x+\varepsilon h)-u(x)|\mathrm{d}h\,\mathrm{d}x\\
&=\frac{\alpha_d^{p-1}}{\varepsilon^{p-1}}\int_{\mathbb{R}^d}\int_{B_d(0,1)}\left|\int_0^1\nabla u(x+t\varepsilon h)\cdot h\mathrm{d}t\right|\mathrm{d}h\,\mathrm{d}x\leq\frac{\alpha_d^{p-1}}{\varepsilon^{p-1}}\int_{\mathbb{R}^d}\int_{B_d(0,1)}\int_0^1|\nabla u(x+t\varepsilon h)\cdot h|\mathrm{d}t\,\mathrm{d}h\,\mathrm{d}x\\
&=\frac{\alpha_d^{p-1}}{\varepsilon^{p-1}}\int_0^1\int_{B_d(0,1)}\int_{\mathbb{R}^d}|\nabla u(x)\cdot h|\mathrm{d}x\,\mathrm{d}h\,\mathrm{d}t=\frac{\alpha_d^{p-1}}{\varepsilon^{p-1}}\int_0^1\int_{\mathbb{R}^d}|\nabla u(x)|\int_{B_d(0,1)}\left|\frac{\nabla u(x)}{|\nabla u(x)|}\cdot h\right|\mathrm{d}h\,\mathrm{d}x\,\mathrm{d}t\\
&=\frac{\alpha_d^{p-1}\sigma_d}{\varepsilon^{p-1}}\int_{\mathbb{R}^d}|\nabla u(x)|\mathrm{d}x,
\end{aligned}
$$

where in the first equation we used the change of variables $h=\frac{x-y}{\varepsilon}$, in the first inequality we used Jensen's inequality and in the second inequality the fact that u takes values in $[0,1]$.

Now, for any set $Q\subseteq D$ as in the statement, we can find a sequence of functions $\{u_k\}_{k\in\mathbb{N}}$ with $u_k:\mathbb{R}^d\to[0,1]$, $u_k\in W^{1,1}(\mathbb{R}^d)\cap C^\infty(\mathbb{R}^d)$ and such that

$$
u_k\overset{L^1(\mathbb{R}^d)}{\longrightarrow}\mathbf{1}_Q,\quad\lim_{k\to\infty}\int_{\mathbb{R}^d}|\nabla u_k(x)|\mathrm{d}x=\mathrm{Per}(Q).\tag{7.18}
$$

Such sequence can be obtained for example with the aid of standard mollifiers (see Theorem 13.9 in [50] for example). It follows from (7.17) and from (7.18) that

$$
\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\frac{\mathbf{1}_{|x-y|\leq\varepsilon}}{\varepsilon^{d+1}}|\mathbf{1}_Q(y)-\mathbf{1}_Q(x)|\mathrm{d}y\right)^p\mathrm{d}x\leq\frac{\alpha_d^{p-1}\sigma_d}{\varepsilon^{p-1}}\mathrm{Per}(Q).
$$

Finally, notice that

$$
\int_D\bar{\phi}_\varepsilon^p(x)\mathrm{d}x=\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\frac{\mathbf{1}_{|x-y|\leq\varepsilon}}{\varepsilon^{d+1}}|\mathbf{1}_Q(y)-\mathbf{1}_Q(x)|\mathrm{d}y\right)^p\mathrm{d}x\leq\frac{\alpha_d^{p-1}\sigma_d}{\varepsilon^{p-1}}\mathrm{Per}(Q).
$$

□

Using the previous lemma with $p=2$ we deduce that $\int_D\bar{\phi}_{\varepsilon_n}^2(x)\mathrm{d}x=O\left(\frac{1}{\varepsilon_n}\right)$, and since

$$
\int_D g_{n,1}^2(x)\mathrm{d}x=\int_D\bar{\phi}_{\varepsilon_n}^2(x)\mathrm{d}x-\left(\mathrm{Per}_{\varepsilon_n}(Q)\right)^2,
$$

we conclude that

$$
B_{n,1}=O\left(\sqrt{\frac{n}{\varepsilon_n}}\right).
$$

From the previous computations, we deduce that

$$\mathbb{E}(|U_{n,1}|^p) \leq C \max\left(\frac{1}{n^p \varepsilon_n^p}, \frac{1}{n^{p/2}\varepsilon_n^{p/2}}\right),$$

where C may depend on the set Q through its perimeter and p. If $\frac{1}{n^{2/(d+1)}} \leq \varepsilon_n$, so that in particular $\frac{1}{n\varepsilon_n}$ is $o(1)$, then

$$\mathbb{E}(|U_{n,1}|^p) \leq \frac{C}{n^{p/2}\varepsilon_n^{p/2}}. \tag{7.19}$$

Now we turn to the task of obtaining moment estimates for $U_{n,2}$. We estimate the quantities $A_{n,2}$, $B_{n,2}$ and $C_{n,2}$ from (7.13). Let us start by estimating $A_{n,2}$. Note that for any $(x,y) \in D \times D$, $\bar{\phi}_{\varepsilon_n}(x)$ and $\bar{\phi}_{\varepsilon_n}(y)$ are of order $\frac{1}{\varepsilon_n}$ and that $\mathrm{Per}_{\varepsilon_n}(Q)$ is of order one. Thus, it is clear from the definition of $g_{n,2}$ in (7.8) that

$$A_{n,2} = O\left(\frac{1}{\varepsilon_n^{d+1}}\right).$$

On the other hand, a direct computation allows us to deduce that for every $x \in D$,

$$\begin{aligned}
\int_D g_{n,2}^2(x,y)\mathrm{d}y = &\int_D \phi_{\varepsilon_n}^2(x,y)\mathrm{d}y - \bar{\phi}_{\varepsilon_n}^2(x) + 2\theta_n \overline{\phi}_n(x) \\
&- 2\int_D \phi_{\varepsilon_n}(x,y)\bar{\phi}_{\varepsilon_n}(y)\mathrm{d}y + \int_D \bar{\phi}_{\varepsilon_n}^2(y)\mathrm{d}y - \theta_n^2 \\
= &\frac{1}{\varepsilon_n^{d+1}}\bar{\phi}_{\varepsilon_n}(x) - \bar{\phi}_{\varepsilon_n}^2(x) + 2\theta_n \overline{\phi}_n(x) \\
&- 2\int_D \phi_{\varepsilon_n}(x,y)\bar{\phi}_{\varepsilon_n}(y)\mathrm{d}y + \int_D \bar{\phi}_{\varepsilon_n}^2(y)\mathrm{d}y - \theta_n^2,
\end{aligned} \tag{7.20}$$

where we are using $\theta_n := \mathrm{Per}_{\varepsilon_n}(Q)$. From this, it follows that

$$C_{n,2} = O\left(\sqrt{\frac{n}{\varepsilon_n^{d+2}}}\right).$$

Finally, upon integration of (7.20) and direct computations, we obtain

$$||g_{n,2}||_{L^2(D)}^2 = \frac{\theta_n}{\varepsilon_n^{d+1}} - 2\int_D \bar{\phi}_{\varepsilon_n}^2(y)\mathrm{d}y + \theta_n^2,$$

which implies that

$$B_{n,2} = O\left(\frac{n}{\varepsilon_n^{(d+1)/2}}\right).$$

Thus, from (7.12) we deduce that

$$\mathbb{E}(|U_{n,2}|^p) \leq K_p \max\left(\frac{1}{n^{2p}\varepsilon_n^{p(d+1)}}, \frac{1}{n^p \varepsilon_n^{p(d+1)/2}}, \frac{1}{n^{3p/2}\varepsilon_n^{p(d+2)/2}}\right),$$

where K_p depends on the set Q through its perimeter. Hence, if $\frac{1}{n^{2/(d+1)}} \leq \varepsilon_n$, we have

$$\mathbb{E}(|U_{n,2}|^p) \leq \frac{K_p}{n^p \varepsilon_n^{p(d+1)/2}}. \tag{7.21}$$

Combining (7.19) and (7.21) and using the canonical decomposition (7.7), we obtain (7.2).

7.1.1 Sharpness of the rate for pointwise convergence

A very simple argument shows that the rates for ε_n that guarantee the almost sure convergence of the graph perimeter to the actual perimeter in Theorem 7.0.2 are optimal in terms of scaling.

In fact, suppose $n^2\varepsilon_n^{d+1} = o(1)$ and let e_n denote the random variable that counts the number of edges that cross the interface between Q and its complement. In other words, we define

$$e_n := \varepsilon_n^{d+1} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \phi_{\varepsilon_n}(\mathbf{x}_i, \mathbf{x}_j).$$

As a consequence, if Q has finite perimeter then we have

$$\mathrm{Per}_{\varepsilon_n,n}(Q) = \frac{2}{n(n-1)\varepsilon_n^{d+1}} e_n, \qquad \mathbb{E}(e_n) = \frac{n(n-1)\varepsilon_n^{d+1}}{2} \mathrm{Per}_{\varepsilon_n}(Q). \tag{7.22}$$

Note that e_n takes integer values in the range $\{0, 1, \ldots, N\}$ for $N = n(n-1)/2$, so that

$$\mathbb{E}(e_n) = \sum_{k=1}^{N} k p_k^n \qquad p_k^n := \mathbb{P}(e_n = k).$$

The fact that $p_0^n + \cdots + p_N^n = 1$ implies

$$\mathbb{E}(e_n) = \sum_{k=1}^{N} k p_k^n \geq \sum_{k=1}^{N} p_k^n = (1 - p_0^n).$$

In particular, from (7.22) and (1.21) we deduce that if $n^2\varepsilon_n^{d+1} \to 0$ and Q has finite perimeter then

$$(1 - p_0^n) \leq \mathbb{E}(e_n) = o(1).$$

On the other hand, note that for any given $\gamma > 0$ it is true that $\mathrm{Per}_{n,\varepsilon_n}(Q) > \gamma$ implies that $e_n \neq 0$. In turn

$$\mathbb{P}\left(\mathrm{Per}_{n,\varepsilon_n}(Q) > \gamma\right) \leq \mathbb{P}\left(e_n \neq 0\right) = 1 - p_0^n = o(1).$$

We conclude that if $n^2\varepsilon_n^{d+1} \to 0$ then $\mathrm{Per}_{\varepsilon_n,n}(Q)$ converges in probability to zero. Therefore, if Q has a non-zero, finite perimeter then $\mathrm{Per}_{\varepsilon_n,n}(Q)$ does not converge to $\sigma_d \, \mathrm{Per}(Q)$ in probability (nor almost surely, either).

7.2 Bias estimates

The bias estimates are obtained by a series of computations whose starting point is writing $\mathrm{Per}_\varepsilon(Q)$ in terms of an iterated integral, the outer one taken over the manifold ∂Q and the inner one taken along the normal line to ∂Q at an arbitrary point $x \in \partial Q$. Such computations show that the first order term of $\mathrm{Per}_{\varepsilon_n}(Q)$ on ε_n vanishes.

Given that we assume $Q \subset\subset (0,1)^d$ and Q has smooth boundary, we conclude that its perimeter can be written

$$\mathrm{Per}(Q) = \int_{\partial Q} \mathrm{d}\mathscr{H}^{d-1} = \mathscr{H}^{d-1}(\partial Q).$$

Additionally, for all $\varepsilon \leq \delta := \mathrm{dist}(Q, \partial D)$ we have that

$$\mathrm{Per}_\varepsilon(Q) = \frac{2}{\varepsilon^{d+1}} \int_Q |B_d(x,\varepsilon) \cap Q^c| \,\mathrm{d}x,$$

where $B_d(x,r)$ denotes the ball of radius r in $\mathbb{R}^d$ centered at x and Q^c denotes the complement of Q in all of space. Moreover, since ∂Q is a compact smooth manifold, we can assume without the loss of generality (by taking ε small enough) that for every $x \in T_\varepsilon$ there is a unique point $P(x)$ in ∂Q closest to x. Furthermore, we can assume that the map P is smooth. We may further write

$$\mathrm{Per}_\varepsilon(Q) = \frac{2}{\varepsilon^{d+1}} \int_{T_\varepsilon^-} |B_d(x,\varepsilon) \cap Q^c| \,\mathrm{d}x,$$

where T_ε^- is defined in (7.15). This reformulation makes it natural to write the previous integral as an iterated integral; the outer integral is taken over the manifold ∂Q and the inner integral is taken along the normal line to ∂Q at an arbitrary point x along the boundary.

To make this idea precise, we first let $\mathbf{N}(x)$ denote the outer unit normal to ∂Q at $x \in \partial Q$ and then consider the transformation $(x,t) \in \partial Q \times (0,1) \mapsto x - t\varepsilon\mathbf{N}(x)$ for all ε sufficiently small. The Jacobian of this transformation equals $\varepsilon \det(I + t\varepsilon\mathbf{S}_x)$, where $\mathbf{S}_x$ denotes the shape operator (or second fundamental form) of ∂Q at x, see [40] for instance. For all ε sufficiently small, we may therefore conclude that

$$\frac{1}{\varepsilon} \int_{T_\varepsilon^-} |B_d(x,\varepsilon) \cap Q^c| \,\mathrm{d}x = \int_{\partial Q} \left(\int_0^1 |B_d(x - t\varepsilon\mathbf{N}(x), \varepsilon) \cap Q^c| \det(I + t\varepsilon\mathbf{S}_x) \,\mathrm{d}t \right) \mathrm{d}\mathscr{H}^{d-1}(x).$$

As a consequence, we also have that

$$\mathrm{Per}_\varepsilon(Q) = \frac{2}{\varepsilon^d} \int_{\partial Q} \left(\int_0^1 |B_d(x - t\varepsilon\mathbf{N}(x), \varepsilon) \cap Q^c| \det(I + t\varepsilon\mathbf{S}_x) \,\mathrm{d}t \right) \mathrm{d}\mathscr{H}^{d-1}(x). \tag{7.23}$$

With the expression (7.23) in hand, we may now proceed to establish (7.5) by expanding $\mathrm{Per}_\varepsilon(Q)$ in terms of ε and appealing to some elementary computations that show that the first order term in ε vanishes.

For a fixed $x \in \partial Q$, we first wish to understand the behavior of the function

$$g_x(\varepsilon) := \frac{1}{\varepsilon^d}\left(\int_0^1 |B_d(x - t\varepsilon\mathbf{N}(x), \varepsilon) \cap Q^c| \det(I + t\varepsilon\mathbf{S}_x)\, \mathrm{d}t\right)$$

for ε in a neighborhood of zero. Without loss of generality, we may assume that $x = 0$, that $\mathbf{N}(x) = e_d$ and that around x the boundary ∂Q coincides with the graph $\hat{x} = (x_1, \ldots, x_{d-1}) \mapsto (\hat{x}, f(\hat{x})) \in \mathbb{R}^d$ of a smooth function $f(\hat{x})$ that satisfies both $f(0) = 0$ and $\nabla f(0) = 0$ simultaneously. By symmetry of the shape operator $\mathbf{S}_x$, there exists an orthonormal basis for $\mathbb{R}^{d-1}$ (where we identify $\mathbb{R}^{d-1}$ with the hyperplane $\{(\hat{x}, x_d) : x_d = 0\}$) consisting of eigenvectors of the shape operator. We let $v_1, \ldots, v_{d-1}$ denote the eigenvectors of $\mathbf{S}_x$ and $\kappa_1, \ldots, \kappa_{d-1}$ the corresponding eigenvalues (also known as principal curvatures). In particular, whenever $\|\hat{x}\| \leq \varepsilon$ we have that

$$f(\hat{x}) = \frac{1}{2}\sum_{i=1}^{d-1} \kappa_i \langle \hat{x}, v_i\rangle^2 + O(\varepsilon^3), \tag{7.24}$$

where curvatures $\kappa_i = \kappa_i(x)$ and the $O(\varepsilon^3)$ error term can be uniformly bounded.

With these reductions in place, we first define $u(\hat{y}) := \sqrt{\varepsilon^2 - \|\hat{y}\|^2}$ and then let

$$h(\hat{y}, t; \varepsilon) := \begin{cases} 2u(\hat{y}) & \text{if} \quad f(\hat{y}) + \varepsilon t < -u(\hat{y}), \\ u(\hat{y}) - \varepsilon t - f(\hat{y}) & \text{if} \quad -u(\hat{y}) \leq f(\hat{y}) + \varepsilon t \leq u(\hat{y}), \\ 0 & \text{otherwise.} \end{cases}$$

A direct calculation then shows that

$$|B_d(x - t\varepsilon\mathbf{N}(x), \varepsilon) \cap Q^c| = \int_{B_{d-1}(0,\varepsilon)} h(\hat{y}, t; \varepsilon)\, \mathrm{d}\hat{y}, \tag{7.25}$$

and an application of (7.24) shows that $h(\hat{y}, t; \varepsilon) = 2\sqrt{\varepsilon^2 - \|\hat{y}\|^2}$ only if

$$\|\hat{y}\|^2 = \varepsilon^2 - O(\varepsilon^4) \qquad \text{and} \qquad u(\hat{y}) = O(\varepsilon^2).$$

It therefore follows that

$$\int_{B_{d-1}(0,\varepsilon) \cap \{f(\hat{y}) + \varepsilon t < -u(\hat{y})\}} h(\hat{y}, t; \varepsilon)\, \mathrm{d}\hat{y} \leq O(\varepsilon^2) \int_{B_{d-1}(0,\varepsilon) \cap \{\|\hat{y}\| \geq \sqrt{\varepsilon^2 - O(\varepsilon^4)}\}} \mathrm{d}\hat{y} = O(\varepsilon^{d+3}).$$

We then let A_t^ε denote the set $A_t^\varepsilon := \{\hat{y} \in B_{d-1}(0, \varepsilon) : -u(\hat{y}) \leq f(\hat{y}) + \varepsilon t \leq u(\hat{y})\}$ and use the previous estimate in (7.25) to uncover

$$|B_d(x - t\varepsilon\mathbf{N}(x), \varepsilon) \cap Q^c| = \int_{B_{d-1}(0,\varepsilon) \cap A_t^\varepsilon} u(\hat{y}) - \varepsilon t - f(\hat{y})\, \mathrm{d}\hat{y} + O(\varepsilon^{d+3}). \tag{7.26}$$

We may then note that

$$\det(I + \varepsilon t\mathbf{S}_x) = (1 + t\varepsilon\kappa_1) \ldots (1 + t\varepsilon\kappa_{d-1}) = 1 + t\varepsilon H_x + O(\varepsilon^2),$$

where $H_x := \sum_{i=1}^{d-1} \kappa_i$ represents the mean curvature. Using this fact in (7.26) then yields

$$g_x(\varepsilon) = \frac{1}{\varepsilon^d} \int_0^1 \left(\int_{B_{d-1}(0,\varepsilon)\cap A_t^\varepsilon} u(\hat{y}) - \varepsilon t - f(\hat{y})\, \mathrm{d}\hat{y} \right) (1 + t\varepsilon H_x)\, \mathrm{d}t + O(\varepsilon^2).$$

Now let $f^\varepsilon(z) := \frac{1}{\varepsilon} f(\varepsilon z)$ and define the corresponding subset C_t^ε of $(0,1) \times B_{d-1}(0,1)$ as

$$C_t^\varepsilon := \left\{ (t,z) \in B_{d-1}(0,1) : -\sqrt{1-\|z\|^2} \le f^\varepsilon(z) + t \le \sqrt{1-\|z\|^2} \right\},$$

then make the change of variables $\hat{y} = \varepsilon z$ to see that

$$g_x(\varepsilon) = \int_{C_t^\varepsilon} \left(\sqrt{1-\|z\|^2} - t - f^\varepsilon(z) \right) (1 + t\varepsilon H_x)\, \mathrm{d}z\mathrm{d}t + O(\varepsilon^2).$$

Recalling (7.24) shows that

$$f^\varepsilon(z) = \frac{\varepsilon}{2} \sum_{i=1}^{d-1} \kappa_i \langle z, v_i \rangle^2 + O(\varepsilon^2), \tag{7.27}$$

which then allows us to obtain an expansion of $g_x(\varepsilon)$ in terms of ε according to the relation

$$\begin{aligned} g_x(\varepsilon) &= \int_{C_t^\varepsilon} \left(\sqrt{1-\|z\|^2} - t \right) \mathrm{d}t\mathrm{d}z \\ &\quad + \varepsilon \int_{C_t^\varepsilon} \left(tH_x(\sqrt{1-\|z\|^2} - t) - \frac{1}{2} \sum_{i=1}^{d-1} \kappa_i \langle z, v_i \rangle^2 \right) \mathrm{d}t\mathrm{d}z + O(\varepsilon^2). \end{aligned} \tag{7.28}$$

The bias estimate (7.5) then directly follows after computing each of these terms individually.

We begin by considering the first term in the expansion, i.e.

$$\mathrm{I} := \int_{C_t^\varepsilon} \left(\sqrt{1-\|z\|^2} - t \right) \mathrm{d}t\mathrm{d}z.$$

Given $\varepsilon > 0$ and $z \in B_{d-1}(0,1)$ define $c(z) := \max\{-\sqrt{1-\|z\|^2} - f^\varepsilon(z), 0\}$ and $C(z) := \min\{\sqrt{1-\|z\|^2} - f^\varepsilon(z), 1\}$, so that we may easily write

$$\mathrm{I} = \int_{B_{d-1}(0,1)} (C(z) - c(z)) \left(\sqrt{1-\|z\|^2} - \frac{C(z)+c(z)}{2} \right) \mathrm{d}z.$$

As the set where $c(z) \neq 0$ has measure at most $O(\varepsilon^2)$, we easily conclude that

$$\mathrm{I} = \int_{B_{d-1}(0,1)} C(z) \left(\sqrt{1-\|z\|^2} - \frac{C(z)}{2} \right) \mathrm{d}z + O(\varepsilon^2).$$

If $C(z)=1$ then $\sqrt{1-\|z\|^2}-\frac{C(z)}{2}=\frac{1}{2}(1-\|z\|^2)+O(\varepsilon^2)$ as well. In any case, it follows that

$$\mathrm{I}=\frac{1}{2}\int_{B_{d-1}(0,1)}(1-\|z\|^2)\,\mathrm{d}z+O(\varepsilon^2)=\frac{\sigma_d}{2}+O(\varepsilon^2). \tag{7.29}$$

We now proceed to compute the second term in the expansion

$$\mathrm{II}:=H_x\int_{C_t^\varepsilon}\left(t\sqrt{1-\|z\|^2}-t^2\right)\,\mathrm{d}t\mathrm{d}z=H_x\int_{B_{d-1}(0,1)}C^2(z)\left(\frac{\sqrt{1-\|z\|}}{2}-\frac{C(z)}{3}\right)\,\mathrm{d}z+O(\varepsilon^2)$$

and the third term in the expansion

$$\mathrm{III}:=\frac{1}{2}\sum_{i=1}^{d-1}\kappa_i\int_{C_t^\varepsilon}\langle z,v_i\rangle^2\,\mathrm{d}t\mathrm{d}z=\frac{1}{2}\sum_{i=1}^{d-1}\kappa_i\int_{B_{d-1}(0,1)}\langle z,v_i\rangle^2C(z)\,\mathrm{d}z+O(\varepsilon^2)$$

in a similar fashion. We always have $C(z)=\sqrt{1-\|z\|^2}+O(\varepsilon)$, so that

$$\mathrm{II}=\frac{H_x}{6}\int_{B_{d-1}(0,1)}(1-\|z\|^2)^{3/2}\,\mathrm{d}z+O(\varepsilon)=\frac{H_x\mathrm{vol}(\mathscr{S}^{d-2})}{6}\int_0^1(1-r^2)^{3/2}r^{d-2}\,\mathrm{d}r+O(\varepsilon). \tag{7.30}$$

The third term follows similarly by appealing to spherical coordinates, in that we have

$$\begin{aligned}\mathrm{III}&=\frac{1}{2}\sum_{i=1}^{d-1}\kappa_i\int_{B_{d-1}(0,1)}\sqrt{1-\|z\|^2}\langle z,v_i\rangle^2\,\mathrm{d}z+O(\varepsilon)\\&=\frac{H_x\mathrm{vol}(\mathscr{S}^{d-2})}{2(d-1)}\int_0^1\sqrt{1-r^2}r^d\,\mathrm{d}r+O(\varepsilon)=\mathrm{II}+O(\varepsilon)\end{aligned}$$

thanks to an integration by parts in the final term. We therefore have that $\mathrm{I}=\sigma_d/2+O(\varepsilon^2)$ and $\mathrm{II}-\mathrm{III}=O(\varepsilon)$, so that $g_x(\varepsilon)=\sigma_d/2+O(\varepsilon^2)$ and

$$\mathrm{Per}_\varepsilon(Q)=2\int_{\partial Q}g_x(\varepsilon)\,\mathrm{d}\mathscr{H}^{d-1}=\sigma_d\,\mathrm{Per}(Q)+O(\varepsilon^2)$$

as desired.

We may also show that when Q is a fixed ball, say $Q=B_d(x_c,\frac{1}{3})$ for $x_c\in\mathbb{R}^d$ the center point of $[0,1]^d$, that the absolute value of the difference between $\mathrm{Per}_\varepsilon(Q)$ and $\sigma_d\,\mathrm{Per}(Q)$ remains bounded from below by $c\varepsilon^2$ for $c>0$ some positive constant. The proof proceeds similarly to the proof of the bias estimate above. In particular, this shows that the bound in Lemma 7.0.3 is optimal in terms of scaling for general sets with smooth boundary.

Bibliography

[1] Martial Agueh. Finsler structure in the p-Wasserstein space and gradient flows. *C. R. Math. Acad. Sci. Paris*, 350(1-2):35–40, 2012.

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

[3] G. Alberti. Variational models for phase transitions, an approach via Γ-convergence. In *Calculus of variations and partial differential equations (Pisa, 1996)*, pages 95–114. Springer, Berlin, 2000.

[4] Giovanni Alberti and Giovanni Bellettini. A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies. *European J. Appl. Math.*, 9(3):261–284, 1998.

[5] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. Birkhäuser Basel, 2008.

[6] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[7] Ery Arias-Castro and Bruno Pelletier. On the convergence of maximum variance unfolding. *The Journal of Machine Learning Research*, 14(1):1747–1770, 2013.

[8] Ery Arias-Castro, Bruno Pelletier, and Pierre Pudlo. The normalized graph cut and Cheeger constant: from discrete to continuous. *Adv. in Appl. Probab.*, 44(4):907–937, 2012.

[9] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009.

[10] Annalisa Baldi. Weighted BV functions. *Houston J. Math.*, 27(3):683–705, 2001.

[11] Sisto Baldo. Minimal interface criterion for phase transitions in mixtures of Cahn-Hilliard fluids. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 7(2):67–90, 1990.

[12] John M Ball and Arghir Zarnescu. Partial regularity and smooth topology-preserving approximations of rough domains. *arXiv preprint arXiv:1312.5156*, 2013.

[13] Mikhail Belkin and Partha Niyogi. Convergence of Laplacian eigenmaps. *Advances in Neural Information Processing Systems (NIPS)*, 19:129, 2007.

[14] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. System Sci.*, 74(8):1289–1308, 2008.

[15] Sergei Bernstein. On a modification of Chebyshevs inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math*, 1(4):38–49, 1924.

[16] Emmanuel Boissard. Simple bounds for convergence of empirical and occupation measures in 1-Wasserstein distance. *Electron. J. Probab.*, 16:no. 83, 2296–2333, 2011.

[17] François Bolley, Arnaud Guillin, and Cédric Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields*, 137(3-4):541–593, 2007.

[18] Jean Bourgain, Haim Brezis, and Petru Mironescu. Another look at sobolev spaces. In *in Optimal Control and Partial Differential Equations*, pages 439–455, 2001.

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

[20] Andrea Braides and Anneliese Defranceschi. *Homogenization of multiple integrals*, volume 12 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1998.

[21] Andrea Braides and Nung Kwan Yip. A quantitative description of mesh dependence for the discretization of singularly perturbed nonconvex problems. *SIAM J. Numer. Anal.*, 50(4):1883–1898, 2012.

[22] X. Bresson and T. Laurent. Asymmetric Cheeger cut and application to multi-class unsupervised clustering. CAM report 12-27, UCLA, 2012.

[23] X. Bresson, T. Laurent, D. Uminsky, and J. von Brecht. Convergence and energy landscape for Cheeger cut clustering. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1394–1402, 2012.

[24] Xavier Bresson, Thomas Laurent, David Uminsky, and James von Brecht. Multiclass total variation clustering. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1421–1429. 2013.

[25] Xavier Bresson, Thomas Laurent, David Uminsky, and James H von Brecht. An adaptive total variation algorithm for computing the balanced cut of a graph. *arXiv preprint arXiv:1302.2717*, 2013.

[26] Xavier Bresson, Xue-Cheng Tai, Tony F Chan, and Arthur Szlam. Multi-class transductive learning based on l1 relaxations of cheeger cut and mumford-shah-potts model. *UCLA CAM Report*, pages 12–03, 2012.

[27] Charles Castaing, Paul Raynaud de Fitte, and Michel Valadier. *Young measures on topological spaces*, volume 571 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2004.

[28] Antonin Chambolle, Alessandro Giacomini, and Luca Lussardi. Continuous limits of discrete perimeters. *M2AN Math. Model. Numer. Anal.*, 44(2):207–230, 2010.

[29] Thierry Champion, Luigi De Pascale, and Petri Juutinen. The ∞-Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.*, 40(1):1–20, 2008.

[30] J. Cheeger. A Lower Bound for the Smallest Eigenvalue of the Laplacian. *Problems in Analysis*, pages 195–199, 1970.

[31] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952.

[32] F. R. K. Chung. *Spectral Graph Theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

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

[34] V. Dobrić and J. E. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.*, 8(1):97–118, 1995.

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

[36] Nicolás García Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *to appear in Canad. J. Math.*, 2015.

[37] Evarist Giné and Vladimir Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 238–259. Inst. Math. Statist., Beachwood, OH, 2006.

[38] Evarist Giné, Rafał Latała, and Joel Zinn. Exponential and moment inequalities for U-statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 13–38. Birkhäuser Boston, Boston, MA, 2000.

[39] Ashish Goel, Sanatan Rai, and Bhaskar Krishnamachari. Sharp thresholds for monotone properties in random geometric graphs. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 580–586 (electronic), New York, 2004. ACM.

[40] A. Gray. *Tubes*. Progress in Mathematics. Birkhäuser Basel, 2004.

[41] Piyush Gupta and P. R. Kumar. Critical power for asymptotic connectivity in wireless networks. In *Stochastic analysis, control, optimization and applications*, Systems Control Found. Appl., pages 547–566. Birkhäuser Boston, Boston, MA, 1999.

[42] L. Hagen and A. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Trans. Computer-Aided Design*, 11:1074 –1085, 1992.

[43] J. Hartigan. Consistency of single linkage for high density clusters. *J. Amer. Statist. Assoc.*, 76:388–394., 1981.

[44] J.A. Hartigan. *Clustering algorithms*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1975.

[45] M. Hein and S. Setzer. Beyond Spectral Clustering - Tight Relaxations of Balanced Graph Cuts. In *Advances in Neural Information Processing Systems (NIPS)*, 2011.

[46] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. From graphs to manifolds–weak and strong pointwise consistency of graph Laplacians. In *Learning theory*, pages 470–485. Springer, 2005.

[47] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[48] V. S. Koroljuk and Yu. V. Borovskich. *Theory of U-statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.

[49] T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica*, 9(2):161–187, 1989.

[50] Giovanni Leoni. *A first course in Sobolev spaces*, volume 105 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2009.

[51] Markus Maier, Ulrike von Luxburg, and Matthias Hein. How the result of graph clustering methods depends on the construction of the graph. *ESAIM: Probability and Statistics*, 17:370–418, 1 2013.

[52] Luciano Modica. The gradient theory of phase transitions and the minimal interface criterion. *Arch. Rational Mech. Anal.*, 98(2):123–142, 1987.

[53] Luciano Modica and Stefano Mortola. Un esempio di Γ-convergenza. *Boll. Un. Mat. Ital. B (5)*, 14(1):285–299, 1977.

[54] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[55] Pablo Pedregal. *Parametrized measures and variational principles*. Progress in Nonlinear Differential Equations and their Applications, 30. Birkhäuser Verlag, Basel, 1997.

[56] Mathew Penrose. A strong law for the longest edge of the minimal spanning tree. *Ann. Probab.*, 27(1):246–260, 1999.

[57] David Pollard. Strong consistency of *k*-means clustering. *The Annals of Statistics*, 9(1):135–140, 1981.

[58] Augusto C. Ponce. A new approach to Sobolev spaces and connections to Γ-convergence. *Calc. Var. Partial Differential Equations*, 19(3):229–255, 2004.

[59] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[60] P. W. Shor and J. E. Yukich. Minimax grid matching and empirical measures. *Ann. Probab.*, 19(3):1338–1348, 1991.

[61] A. Singer. From graph to manifold Laplacian: the convergence rate. *Appl. Comput. Harmon. Anal.*, 21(1):128–134, 2006.

[62] Arthur Szlam and Xavier Bresson. A total variation-based graph clustering algorithm for cheeger ratio cuts. *UCLA CAM Report*, pages 1–12, 2009.

[63] Arthur Szlam and Xavier Bresson. Total variation and cheeger cuts. In Johannes Fnkranz and Thorsten Joachims, editors, *ICML*, pages 1039–1046. Omnipress, 2010.

[64] M. Talagrand. The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.*, 22(2):919–959, 1994.

[65] M. Talagrand and J. E. Yukich. The integrability of the square exponential transportation cost. *Ann. Appl. Probab.*, 3(4):1100–1111, 1993.

[66] Michel Talagrand. *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005. Upper and lower bounds of stochastic processes.

[67] Michel Talagrand. *Upper and lower bounds of stochastic processes*, volume 60 of *Modern Surveys in Mathematics*. Springer-Verlag, Berlin Heidelberg, 2014.

[68] Daniel Ting, Ling Huang, and Michael I Jordan. An analysis of the convergence of graph Laplacians. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.

[69] Yves van Gennip and Andrea L. Bertozzi. Γ-convergence of graph Ginzburg-Landau functionals. *Adv. Differential Equations*, 17(11-12):1115–1180, 2012.

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

[71] U. von Luxburg, M. Belkin, and Bousquet O. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

[72] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[73] Yen-Chuen Wei and Chung-Kuan Cheng. Towards efficient hierarchical designs by ratio cut partitioning. In *Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International Conference on*, pages 298–301. IEEE, 1989.

[74] Hermann Weyl. On the Volume of Tubes. *Amer. J. Math.*, 61(2):461–472, 1939.

[75] Rui Xu and D. Wunsch, II. Survey of clustering algorithms. *Trans. Neur. Netw.*, 16(3):645–678, May 2005.