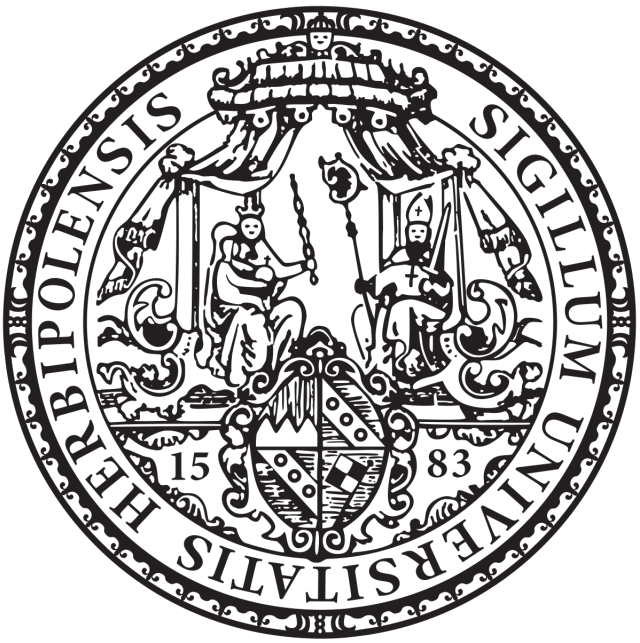
Variational Approach to Graph Clustering

Nick Burk

30. Dezember 2025



Julius-Maximilians Universität Würzburg
Institut für Mathematik
Lehrstuhl für Mathematik des Maschinellen Lernens

Masterarbeit zur Erlangung des akademischen Grades „Master of Science“

Vorgelegt an der Julius-Maximilians Universität Würzburg im Wintersemester 2025

Betreut durch
Prof. Dr. Leon Bungert
Dr. Eloi Martinet

Contents

1 Introduction

In data science and machine learning one of the most fundamental task is the extraction of meaningful structure from given high-dimensional data. In theses point clouds the search for meaningful clusters is challenging yet very important. A class of approaches that has emerged as potent and reliable are graph based methods, where a graph is created out of the data cloud by connecting nearby points. As graphs naturally encode relational information and geometric properties inherent in the data, leveraging them allows for high quality clustering. On the other hand, both the number of fields where Graphs occur as canonical data structure and their importance have increased significantly in recent years. For example in social sciences, bioinformatics or recommendation systems their ability to model complex relationships for large quantities of data are quite valuable. In these domains, Graph partitioning helps to reduce computational complexity and to extract usable insights.

In general, graph clustering approaches rely on an objective function which measures the quality of the partition by penalizing cuts between clusters. A solution obtained from optimizing said objective leads to clusters that are well-separated. In order to ensure that clusters are of meaningful size and robust to outliers, "balance" terms are introduced. These give rise to functionals such as the Cheeger cut and related version that focus on edge measures.

In this thesis, we consider data clouds, $X_n = \{x_1, \dots, x_n\}$, obtained as independent and identically distributed (i.i.d.) samples from a measure ν with density ρ on a bounded domain D. The measure ν represents the ground truth from which X_n is sampled. As the sample size n grows, clustering methods should exhibit consistency, meaning the clustering of the data X_n should converge to a specific partition of the underlying domain D. A notion of convergence suitable for showing the convergence of minimizers of the Graph objective functionals towards a minimizer of the limit is Γ-convergence. It was introduced by De Giorgi in the 70's, allows for studying the limits of families of problems and can be considered as one of the standard tools for proving variational convergence.

Our approach to formulate a consistent problem is as follows. Given the i.i.d sample points X_n, we build a graph using a kernel η that relies on some threshold ε. This gives us a weighted graph where the further the distance between points, the smaller their shared weight. for distances larger than the connectivity threshold ε the weight is zero. To calculate the partition, rather than relying on a combinatorical heuristics, we use a variational approach. We achieve this by defining operators on graphs that are consistent with their continuum counterparts. This also allows us to reformulate

the graph functional so that it can be adapted to a continuum setting in a natural way. Eventually, we also introduce an appropriate formulation to use soft labeling functions instead of indicators. Specifically, we formulate a functional inspired by the Ginzburg-Landau or Allen-Cahn approach. The minimizers of this are phase field functions with a small boundary between sets. To prove Γ-convergence we use a approach similar to Modica in [17]. Furhtermore, we use optimal transport theory to bridge the gap between discrete and continuous functions. And we provide a generalization to the multi-class balanced Cheer-cut.

This approach allows us to prove that the continuum solution can be recovered as the number of points tends to infinity. Central to our analysis is the concept of Γ-convergence and its properties. To emphasis the practicability of our theoretical findings, we provide computational examples for various distributions, demonstrating the robustness and flexibility of our approach. Our approach contrasts with existing algorithms like DBSCAN, which rely on subjective definitions of density or proximity. Furthermore, we improve on geometric methods like k-Nearest-Neighbors and k-Means, that provide little theoretical rigour and are susceptible to outliers or non-spherical clusters.

This thesis builds on the work of García Trillos et al. [14] in which the convergence of graph Cheeger cuts in the form of graph total variation and its minimizers was first demonstrated. It also incorporates the tools introduced by García Trillos and Slepčev [13], which were developed for the study of consistency of minimization problems on random point clouds in a random discrete setting. In particular, the proof of Γ-convergence of total variation on graphs was provided. For a limited class of graph structure settings, work related to the Γ-convergence of grap functionals to continuous functionals involving perimeter includes that of van Gennip at all [18].

The idea of using functionals with smooth minimizers that allow for small boarders of transition originates from Cahn and Hillard [8] while Cristoferi and Thropes [9] establish an analogous result on graphs. They provide a general approach involving the p-Laplacian by considering a discretisation of the non-local Ginzburg–Landau functional studied by Alberti and Bellettini [1] which considers non-local functionals that converge to total variation. The α-Cheeger-cut is inspired by the work of Bogosel et al. [4] who demonstrate the convergence of a Ginzburg–Landau-style functional divided by a balance term, which is in turn weighted by an additional parameter. A notion of convergence suitable for showing the convergence of minimizers of approximating objective functionals converge towards a minimizer of the limit functional is the notion of Γ-convergence, introduced by De Giorgi in the 70's and representing a standard notion of variational convergence. See the books by Braides [6] and Dal Maso [10] for a detailed exposition of the properties of Γ-convergence.

This thesis is organised as follows: Chapter 2 begins with a general introduction to graphs, providing the necessary background information and definitions. This foundational material enables us to formulate graph cut problems, which we then link to perimetric problems in the continuum setting.

Chapter 3 focuses on continuum perimeter problems and how they are formulated. We introduce functions of bounded total variation and discuss the use of weighted total variation, which generalises the canonical Euclidean case to other measures, such as Gaussian measures. This framework enables us to work with any measure that has a bounded, absolutely continuous density with respect to the Lebesgue measure. The chapter also addresses the extension of these problems to multiple sets or clusters, resulting in multi-phase field problems that require only an additional orthogonality constraint on the sets.

Chapter 4 addresses the challenge posed by the discrete nature of graphs when comparing solutions or proving convergence. To overcome this issue, we use the theory of optimal transport, based on a method developed by García Trillos that establishes an upper bound on the optimal transport distance between graph points and their continuum counterparts.

Chapter 5 shifts the focus from solving perimeter problems for a fixed volume — a restrictive assumption in many applications — to incorporating volume into the optimization process. We introduce the multi-class balanced cut, also known as the α-Cheeger-Cut, which aims to maximize the volume while maintaining a minimal perimeter. By incorporating a parameter α into the objective function, we can adjust the balance between volume and perimeter, thereby gaining control over the clustering process. The balanced cut is defined for both continuum domains and graphs. We demonstrate that our graph definitions Γ-converges to the continuum ones, building on and slightly generalizing an existing result from García Trillos.

Chapter 6 focuses on providing a computationally favourable approach by introducing a functional that can be minimized by continuously differentiable functions. To this end, we introduce a small border region of width $\varepsilon > 0$ and approximate the characteristic function of a set using a smooth phase-field function that takes values in $[0, 1]$ The theory of Γ-convergence ensuresensures that, under certain assumptions, this phase-field function converges to an indicator function of a suitable set. A significant part of the proof involves the convergence of a non-local functional to the weighted perimeter, with the recovery sequence constructed in a manner similar to that in the original Modica-Mortola proof. We also extend these results to a multi-phase setting, enabling graphs to be clustered into an arbitrary number of clusters on a solid theoretical foundation.

In Chapter 7 we present a variety of numerical examples to illustrate the theoretical results. These include examples of different densities in a graph setting, the effect of the balance parameter on the solution and comparisons with continuum solutions obtained using finite difference methods. We also consider general clustering problems, such as the 'two moons' dataset with outliers.

Finally, in Chapter 8, we propose further questions and present possible ideas on how to generalise our results on existence and asymptotic behaviour.

2 Graphs

2.1 Introduction

A graph is a mathematical structure consisting of a set of objects called nodes or vertices that are related in some sense. This relationship is expressed by edges, which connect two nodes. When solving problems computationally, we encounter two types of graph: those where the entire graph is given canonically, and those where we construct one from a given set of vertices using a metric.

Graphs occur naturally in the context of networks such as the power grid, internet servers, and social interactions. By mapping these out, the graph's structure — namely its nodes and edges — encodes a pre-existing external object. In this context, we cannot simply add or remove parts of the graph, as this would alter the object in question.

Conversely, when given a set of vertices, we can operate much more freely. The points are usually assumed to be elements of a metric space, ideally $\mathbb{R}^d$. Sometimes it may be necessary to translate the characteristics of the points in a suitable way. For example, the pixels of an image could be considered as elements of $\mathbb{R}^5$, with two dimensions in space and three in colour. However, there are many ways to construct a variety of graphs using points in a metric space. For instance, we could connect any two points whose distance is smaller than a given $\varepsilon > 0$. Clearly, for any non-trivial set, different values of e can produce wildly different graphs. As such, constructing the graph itself can pose challenges, though this is not an issue we investigate in detail. However, we require this ε to be scaled correctly, as we need a sufficiently connected graph for our proof to hold and our numerical experiments to work. We will see later how this can be done for points sampled from a suitable probability distribution in $\mathbb{R}^d$. In this context, the graph serves as a tool for encoding the geometric relations shared by any two points. This additional information is a powerful tool that allows for robust convergence.

In our case, we work with the modelling assumption that the vertices are sampled from a bounded continuous density. The underlying function for a set of observations is also referred to as the 'ground truth' in data science and machine learning. In supervised machine learning, the main task is to uncover this ground truth. Although our objective differs slightly, it has implications for supervised or semi-supervised learning with graphs. Our focus is on recovering the continuum solution: if we sample a sufficiently large number of points, the results should closely resemble the expected solution in the continuum space from which we sampled.

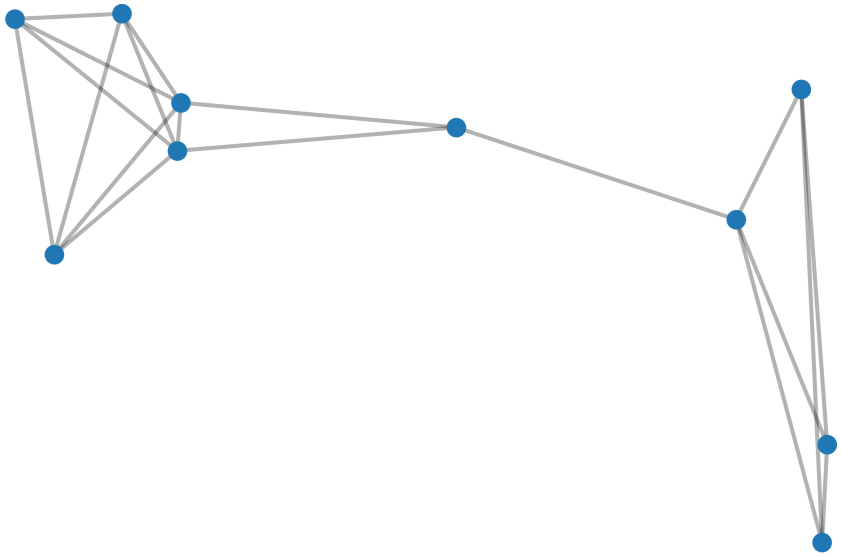


Figure 2.1: Example of a simple graph with two potential clusters.

The next logical question, then, is what kind of problems we are interested in, and what sort of continuum solution we expect.

Put simply, we need to know how to cluster the vertices of a given graph. A cluster describes a subset of data that shares an exclusive characteristic. The aim is to partition the original set into several clusters such that the objects within each cluster are more similar to one another than to objects in any other cluster. Most of the time, this characteristic is geometrical. A solution consists of areas of high density separated by low-density space, as seen in the Figure 2.1. In the continuum setting, this objective translates to finding subsets of a given domain such that their boundaries have small values in the given measure.

In this chapter, we provide a definition of graphs as we will use them, and we give some background on their construction from a vertex set. We then define the necessary operators on graphs, which are similar to those in the continuum case. These operators are used to introduce clusters and graph cuts, which are closely related.

2.2 Graphs

Definition 2.1 (Weighted Graphs)**.** A weighted graph is a tuple $G = (V, w)$, where V is a countable set of vertices and $w : V \times V \to [0, \infty)$ is a weight function. G is called undirected if $w(x, y) = w(y, x)$ for all $x, y \in V$ and directed otherwise. The set of edges is defined as $E = \{(x, y) \in V \times V : w(x, y) > 0\}$. Furthermore, we will

oftentimes write $w_{xy} := w(x, y)$ and adhere to the convention that $w(x, x) = 0$.

There are countless weight functions to choose from. Some of the most popular ones are of the the form $w : V \times V \to \{0, 1\}$. These lead to what we call an unweighted graph. In this case defining a graph as the tuple $G = (V, E)$, with the set of edges $E = \{(x, y) \in V \times V : w_{xy} = 1\}$ is equivalent. In the literature, the term graph often refers to an unweighted graph. The definition via a set of edges instead of a weight function is also common. As we will work with distribution functions, kernels and functionals on continuous sets in later proofs, it is more convenient to use the weight function directly. From now on, we will only consider weighted, undirected graphs as defined above. Nevertheless, let us briefly consider three common approaches to creating graphs before introducing some weight functions.

First, assume that our vertex set is simply $V = \{0, ..., n\}$, with no additional geometric information. One approach is to assign neighbours to each node according to a probabilistic generating model. The best-known example of this method is probably the Erdős–Rényi graph, for which the generating model is binomial $B(n, p)$. Here, edges are formed independently in pairs with probability p.

Another important generation model is the Watts–Strogatz model. Starting from a regular lattice, edges are rewired with a certain probability, resulting in networks known as scale-free, which reflect the power-law degree distributions observed in many real-world systems.

For points in a metric space, a good approach is to connect each point to its k nearest neighbours to create a k-NN graph. This creates a graph with a very regular structure and no isolated points. However, it may obscure the original data distribution, particularly if it was not very regular. An alternative approach is to connect any two vertices that lie within a fixed distance ε of each other and weight them accordingly. These ε-ball graphs preserve the underlying distribution, but choosing the right ε can be challenging. When using a metric to define weights in this way, there is often a trade-off between capturing global and local properties. If the threshold ε is chosen too small, some nodes may become disconnected from the rest. Conversely, if it is chosen too large, the graph will become overly dense and connect vertices that share no significance. Furthermore, computing the weights for a large number of node pairs can be computationally expensive and numerically unstable. Therefore, it is often advisable to work with a sparser graph.

An important property of a weight function based on a metric is how it decays relative to distance. For example, given the points $\{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, the weight function $w_{ij} = \max(0, 1 - \|x_i - x_j\|)$ decays linearly, whereas $w_{ij} = \exp\left(-\frac{\|x_i - x_j\|}{c}\right)$ does so exponentially. The choice depends on priorities, ranging from preserving local structure to ensuring computational tractability.

For our numerical examples, we will use ε-ball graphs with a Gaussian similarity kernel. Let $|\cdot|$ denote the Euclidean norm and $c > 0$ be a constant. Then, the weight function

$w : \mathbb{R}^d \to [0,1]$ is given by

$$w_{ij} = \begin{cases} \exp\left(-\frac{c|x_i - x_j|}{\varepsilon}\right) & \text{if } |x_i - x_j| < \varepsilon, \\ 0 & \text{else.} \end{cases}$$

When discussing the vertices, it is useful to refer to them by their indices in the set $\{1, \dots, n\}$. We can then use these indices to define the so called weight matrix W, where $W_{ij} = w(i,j)$, for $i, j \in \{1, \dots, n\}$. This $n \times n$ matrix has full rank if and only if all vertices have at least one edge. A graph is fully connected if all vertices are connected to each other, meaning the weight matrix has no entries equal to zero apart from on the diagonal. The weight matrix is especially helpful in implementations.

As mentioned above, it is important that the graph is properly connected in order to represent the underlying domain and avoid single disconnected vertices. We now formalise the notion of connectedness.

Definition 2.2 (Connectedness)**.** A Graph $G = (V, w)$ is called connected if for every $x, y \in V$ there exist a $k \in \mathbb{N}$ and points $x_1, \dots, x_k \in V$, such that $x = x_1$ as well as $y = x_k$ and for $i \in \{1, \dots, k-1\}$ it holds $w(x_i, x_{i+1}) > 0$.

We now prove that random geometric ε-ball graphs are connected with high probability. The above definition and the proposition below are taken from [7].

Proposition 2.1 (Connectedness of random geometric graphs)**.** *Let $\Omega = [0,1]^d$ be a hypercube and $\mu \in \mathcal{P}(\Omega)$ a probability measure which has the density ρ with respect to the d-dimensional Lebesgue measure. Assume that there exists a constant $c_\rho > 0$ such that $c_\rho \leq \rho$ almost everywhere in Ω. Let $V = \{x_1, \dots, x_n\} \subset \Omega$ be i.i.d. points sampled from μ and weight function w such that $w(x_i, x_j) > 0$ if $|x_i - x_j| < \varepsilon$. Then there exist constants $C_1, C_2 > 0$ depending only on d and c_ρ such that the graph $G = (V, w)$ is connected with probability at least $1 - C_1 n \exp(-C_2 n \varepsilon^d)$.*

Proof. The proof idea is to cover Ω with boxes, such that the maximal distance between two points in neighboring boxes is at most ε. Then the connectedness of the graph is equivalent to there being no empty boxes. So, we cover Ω with $M = \left\lceil 2^d d^{\frac{d}{2}} \varepsilon^{-d} \right\rceil$ disjoint boxes $\{B_i\}_{i=1,\dots,M}$ of side length $h \leq \frac{\varepsilon}{2\sqrt{d}}$. We denote the event of G being disconnected by $N := \{\exists B_i : V \cap B_i = \emptyset\}$, i.e. there is an empty box. Using a union bound and the fact that the graph points are i.i.d., we get

$$\begin{aligned} \mathbb{P}(N) &\leq \mathbb{P}\left(\bigcup_{i=1}^{M} \{B_i \cap V = \emptyset\}\right) \\ &\leq \sum_{i=1}^{M} \mathbb{P}\left(B_i \cap V = \emptyset\right) \\ &= \sum_{i=1}^{M} \mathbb{P}\left(x_1 \notin B_i\right)^n. \end{aligned}$$

From the lower bound on ρ it follows

$$\mathbb{P}(x_1 \notin B_i) = 1 - \int_{B_i} \rho \, \mathrm{d}x \leq 1 - |B_i| c_\rho = 1 - C_2 \varepsilon^d,$$

where C_2 depends on d and c_ρ. Now, if $n \geq \varepsilon^{-d}$, we can use the elementary inequality $1 - t \leq \exp(-t)$ for all $t \in \mathbb{R}$ to get

$$\mathbb{P}(N) \leq M(1 - C_2 \varepsilon^d)^n \leq C_1 \varepsilon^{-d} \exp(-C_2 n \varepsilon^d) \leq C_1 n \exp(-C_2 n \varepsilon^d).$$

On the other hand if $n \leq \varepsilon^{-d}$, choosing $C_1 \geq \exp(C_2)$ it holds the estimate

$$\mathbb{P}(N) \leq 1 \leq C_1 n \exp(-C_2) \leq C_1 n \exp(-C_2 n \varepsilon^d).$$

□

Connectedness provides a meaningful scale for ε and is closely related to the transportation distance Proposition 4.4. If the graph is not connected, we will not obtain the correct continuum partition. While we may obtain an optimal cut on the graph, the convergence fails, hindering us from making meaningful statements.

2.3 Operators, Cuts and Clusters

Let V be a subset of $\mathbb{R}^d$. If V is a set of sampled points from some given density, we will write X_n to indicate this. In our definition, the edges or the weight function only encode the binary relationship between vertices. Due to of their implicit representation via a weight function, the edges themselves are not objects of interest. Any function is solely defined on the vertices of the graph; for example, the characteristic function of a set is defined in this way.

Since the number of vertices is finite, any function $u : G \to \mathbb{R}$ has an equivalent representation as a vector in $\mathbb{R}^n$. This implies that we can treat any graph function as a vector in $\mathbb{R}^n$ and that the space of graph functions is finite-dimensional.

Definition 2.3 (Space of Graph functions)**.** Let $G = (V, w)$ be a graph. We define the finite-dimensional space of functions on the graph as $l^2(V) := \{u : V \to \mathbb{R}\}$ and equip it with the scalar product

$$\langle u, v \rangle = \sum_{x \in V} u(x) v(x).$$

In the above definition, the scalar product is the discrete version of the one implying the L^2-norm in the continuum setting. This set of functions constitutes a Hilbert space. Therefore, it is quite straightforward to treat functions of graphs, as well as to define the following operators. However, before we do that, let us introduce another important quantity: the degree of a vertex, defined as

$$\deg(x) = \sum_{y \in V} w_{xy}. \tag{2.1}$$

Definition 2.4 (Operators on Graphs)**.** Given a graph $G = (V, w)$ we define the gradient of a function $u : G \to \mathbb{R}$ at the point $x \in V$ by

$$\nabla u(y) := \sqrt{w_{xy}}(u(y) - u(x)).$$

Furthermore we define the graph Laplace operator and the degree operator $L, D : l^2 \to \mathbb{R}$ via

$$Lu(x) := \sum_{y \in V} w_{xy}(u(y) - u(x)),$$
$$Du(x) := \deg(x)u(x).$$

As with the weight functions previously, it is equivalent to express these operators as matrices. While the degree matrix is a quite simply $D_{ij} = \delta_{ij} \deg(i)$ for any $i, j \in V$, the Laplace matrix is given by

$$L = W - D.$$

To see this, let us rewrite $u \in \mathbb{R}^N$ with $u_i = u(i)$, a notation that we will use frequently. Using the same natural formulation, we can define a vector $Lu \in \mathbb{R}^N$, where $(Lu)_i = Lu(i)$. By definition of the Laplace operator and the degree of a node, it follows that

$$(Lu)_i = \sum_{j \in V} w_{ij}u_j - \deg(i)u_i.$$

As this is simply the product of u and the transposed vector $w_j \in \mathbb{R}^N$, where the i-th entry equals $\deg(i)$, the matrix representation follows. As in a continuous setting, the Laplace operator Lu is of course equivalent to $\langle \nabla, \nabla u \rangle$. Alternatively, it can be derived from the divergence operator, which is the adjoint of the gradient operator. If G is undirected, then D and L are self-adjoint operators, meaning that for any $u, v \in V$ it holds that $\langle Du, v \rangle = \langle u, Dv \rangle$ and $\langle Lu, v \rangle = \langle u, Lv \rangle$.

A subset $A \subset G$ of a graph is quite simply defined by its vertices, with the edges being implicitly defined via the weight function. The complement of A is defined as $A^c = V \backslash A$. An important next question for a given set A is how to define its perimeter in V. One approach is to start at the boundary $\partial A := \{x \in A | \exists y \in V \backslash A \text{ with } w(x, y) > 0\}$. A vertex lies on the boundary of A if it shares an edge with a vertex outside A. To measure this boundary, we simply need to keep track of all the edges, which gives us the definition of the perimeter of a set.

Definition 2.5 (Graph Perimeter and Volume)**.** Given a graph $G = (V, w)$ and a subset $A \subset G$, the perimeter of A is defined as

$$\mathrm{Per}(A) := \sum_{x \in A} \sum_{y \in A^c} w_{xy}.$$

The volume of any set on the graph is defined by the proportion of vertices within $|A| = \frac{\#\{x \in A\}}{n}$ for all $A \subset V$.

On a graph, the perimeter is also known as a 'cut', referring to the number of edges that would need to be cut for this set to become disconnected from the rest of the graph. A cut is commonly called expensive if it takes a large value, and cheap otherwise — just as it is harder to cut a tough material.

Now that we have a framework for sets, we can discuss clusters and our objectives. Although the concept of a cluster as a set of closely related points is straightforward, its rigorous definition is not. So, let us set out the requirements for a good cluster. For any given point, we want to obtain a set of densely connected peers that are clearly distinct from any other cluster. In a graph, this means a high number of relatively heavy edges. Also, each point should belong to only one cluster. Using the previous definition, these requirements can be formulated as a minimization problem with constraints. For a non-empty set A such a problem is of the form

$$\min_{A \subset V} \mathrm{Per}(A).$$

In the case of two clusters, the aim is to find a function $u : V \to \{0, 1\}$ that defines the set $A = \{x \in V : u(x) = 1\}$ and minimizes $\mathrm{Per}(A)$ such that $|A| > 0$.

Now, all the vertices belong to exactly one class, and there is no area of lower density to separate the two. Otherwise, a set and its complement would exist with fewer edges in between. However, note that this solution is not unique. Without additional constraints, such a solution can be trivial; a single isolated vertex might constitute the second set. The only way to counteract such trivial solutions is to introduce additional constraints, mainly concerning the volume of both sets. To this end, we can prescribe a fixed volume constraint or introduce a volume term into the minimization objective. In our setting, the first option relates to the classical isoperimetric problem, and the second to the Cheeger–Cuts, both of which we will discuss later. In our setting, the first option relates to the classical isoperimetric problem and the second to the Cheeger–Cut problem. We will discuss both of these later.

Although the motivation and solutions on graphs are all about clusters, it should be clear why we will now focus on the more general concept of perimeter problems. The perimeter is a quantity that describes many properties, such as shapes or clusters. In the next chapter, we will introduce the concept of perimeter to general continuous L^1 functions.

3 Perimeter Problems

3.1 The Isoperimetric Problem

Given a domain $D \subset \mathbb{R}^d$, a perimeter problem is one that deals with finding a set $A \subset D$ with a minimal perimeter given certain constraints. We will look at the definition of perimeter for arbitrary sets shortly; for now, let us simply assume that we can measure it and that the perimeter is always positive. Of course, the minimum could be zero, so constraints are necessary to avoid these trivial solutions. In the simplest case, the only constraint is that the volume $\mathrm{Vol}(A) = m$, where m is a constant. This is called the isoperimetric problem

$$\begin{aligned} &\min_{A \subset D} \mathrm{Per}(A) \\ &\text{s.t.} \mathrm{Vol}(A) = m. \end{aligned}$$

It has a long history that reaches back at least to classical antiquity. Even then, it was known that the shape that solves this problem is a ball. Due to its simple setup, the isoperimetric problem is ideal for introducing the tools required to solve more advanced perimeter problems. However, since we also want to solve perimeter problems computationally, there are two issues that need to be addressed first. Firstly, when looking for a subset of D, it would be natural to search the space of all possible subsets. However, solving this computationally would make it a combinatorial problem, which is notoriously difficult to solve. Instead, we would prefer to work with smooth functions that are suitable for established optimisation approaches. Secondly, we need a universal yet flexible definition of perimeter that works with characteristic functions, but which can also be relaxed to differentiable approximations consistently. For this purpose, we introduce the concept of total variation.

Definition 3.1 (Total Variation)**.** The Total Variation of a function $f \in L^1(D)$ is defined via

$$TV(f) := \sup \left\{ \int_D f \mathrm{div}\phi \, \mathrm{d}x \mid \phi \in C_c^1(D, \mathbb{R}^d), |\phi| \leq 1 \right\}.$$

Here $C_c^1(D, \mathbb{R}^d)$ denotes the subspace of differentiable functions with compact support and $|\cdot|$ is the euclidean norm. We call the set of functions where $TV(f) < \infty$, the space of functions of Bounded Variation and denote it with $BV(D)$.

If we replace f with the characteristic function $\chi_A : D \to \{0, 1\}$ of the set $A \subset D$, the only subsets where χ_A "varies" are those forming the boundary of A, as its value changes. This provides a means of measuring the perimeter. The set A has finite perimeter in D if $\chi_A \in BV(D)$. For smooth sets, the total variation corresponds to the standard definition of the perimeter $\int_{\partial A} \mathrm{d}x$.

Functions of bounded variation are intuitive yet also very useful, particularly for our use case. In the next section, we will therefore spend some time introducing and studying them and their properties.

3.2 Functions of Bounded Variation

This section provides a very brief overview of functions of bounded variations (BV), focusing on the aspects that will be required later. It is based on [11, Chapter 5]. The same book also provides a comprehensive overview of measures, in particular differentiation of Radon measures, and Sobolev function which are closely related to BV functions. Function of bounded variation on $\mathbb{R}^d$ are functions whose weak first partial derivatives are Radon measures. This is essentially the weakest form in which a function can be differentiable in the measure theoretic sense. As these first derivatives are merely measures, it is not at all obvious that any of the usual rules of calculus apply. We start by introducing the Structure Theorem 3.1, the main building block for this assertion as Radon measures as we will see shortly. Next, we make things more concrete by linking the total variation to strong derivatives. The main properties that concern us are the lower semi-continuity of BV functions Proposition 3.2 and the approximation of BV functions by smooth functions Proposition 3.3.

Throughout this section let $U \subset \mathbb{R}^d$ be an open set. We recall the definition of the weak partial derivative, which is a generalisation of the concept of derivatives for functions that are not necessarily differentiable, but are only assumed to be integrable.

Definition 3.2 (Weak Partial Derivative)**.** Given a function $f \in L^1(U)$, we call $g_i \in L^1(U)$ the weak partial derivative of f with respect to $x_i, i = 1, \ldots, d$ if

$$\int_U f \partial_i \phi \, \mathrm{d}x = -\int_U g_i \phi \, \mathrm{d}x \tag{3.1}$$

for all $\phi \in C_c^1(U)$.

Therefore, any function that satisfies the partial integration formula qualifies as a weak derivative. If two such functions are equal Lebesgue almost everywhere and f is differentiable in the conventional sense, then its weak derivative is identical to the strong derivative. According to the structure theorem, for a function of bounded variation, the weak derivative is given by a Radon measure μ with respect to some μ-measurable density σ.

Theorem 3.1 (Structure Theorem)**.** . *Let $f \in BV(U)$. Then there exists a Radon measure μ and a function $\sigma : U \to \mathbb{R}^n$, such that $|\sigma| = 1$ μ-a.e. and for all $\phi \in C_c^1(U, \mathbb{R}^d)$, we have*

$$\int_U f \mathrm{div}\phi \, \mathrm{d}x = -\int_U \phi \cdot \sigma \, \mathrm{d}\mu.$$

We do not show the result here, but we note some points of interest from the proof [11, Theorem 5.1]. In said proof the Riesz Representation Theorem is applied to the linear functional $L(\phi) = -\int_U f \mathrm{div}\phi \, \mathrm{d}x$. It should be noted that the total variation is in fact the supremum of linear functionals and is finite on compact sets.

To see how this all connects assume that $f \in W^{1,1}(U)$, i.e. a Sobolev function in $L^1(U)$ with first weak derivative $(g_1, \ldots, g_d)$. This derivative is uniquely defined $\mathcal{L}^n$-a.e. and we denote it with $Df := (g_1, \ldots, g_d) \in \mathbb{R}^d$. Recall the definition of the divergence operator as$\mathrm{div}\phi := \sum_{i=1}^d \partial_i \phi$. Then for each $\phi \in C_c^1(U)$ with $|\phi| \leq 1$ it follows from the structure theorem

$$\int_U f \mathrm{div}\phi \, \mathrm{d}x = -\int_U Df \cdot \phi \, \mathrm{d}x \leq \int_U |Df| \mathrm{d}x. \tag{3.2}$$

The inequality follows from $|\phi| \leq 1$ and the the fact that, the left-hand side is maximal if $\mathrm{sign}(\phi) = -\mathrm{sign}(Df)$. Since $f \in W^{1,1}(U)$, it holds $\int_U |Df| \mathrm{d}x < \infty$, implying $W^{1,1}(U) \subset BV(U)$**.** Furthermore, Proposition 3.1 provides a measure $\|Df\| = \mathcal{L}^d \llcorner |Df|$, i.e. with density $|Df|$ with respect to the Lebesgue measure $\mathcal{L}^d$. We call this the variation measure of f, since it constitutes an upper bound on the functional $L(\phi)$ and is therefore equal to the total variation of f. Note that, in the case $Df \neq 0$ the σ from the structure theorem, it is given by $\sigma = \frac{Df}{|Df|}$, or $\sigma = 0$ otherwise.

A similar calculation holds for a smooth, open set $E \subset \mathbb{R}^d$ with finite Hausdorff measure $\mathcal{H}^{n-1}(\partial E \cap K) < \infty, K \subset U$ compact. Then for ϕ as above and ν the outward unit normal along ∂E

$$\int_E \mathrm{div}\phi \, \mathrm{d}x = \int_{U \cap \partial E} \phi \cdot \nu \, \mathrm{d}\mathcal{H}^{n-1} \leq \mathcal{H}^{n-1}(\partial E \cap U).$$

Hence E has finite perimeter in V and we can define the perimeter measure $\|\partial E\| = \mathcal{H}^{n-1} \llcorner \nu$. But note, $\chi_E \notin W^{1,1}(U)$, so not every function of bounded variation is a Sobolev function.

The following proposition as found in [11, Theorem 1.30] is a version of the Radon-Nikodym-Theorem. Here $D_\nu \mu$ denotes the density of μ w.r.t. ν.

Proposition 3.1 (Differential of Radon measures)**.** *Let ν, μ be Radon measures on $\mathbb{R}^d$,with $\mu << \nu$. Then*

$$\mu(A) = \int_A D_\nu \mu \, d\nu$$

for all ν-measurable sets $A \subseteq \mathbb{R}^d$.

This proposition not only proves that ν has a density with respect to μ, but also that this density can be computed via the derivative w.r.t. ν. Compare this result with (3.2) to obtain the promised variation measure. Furthermore, we can now see that functions of bounded variation are those with first weak derivative equal to Radon measure. In our case $\nu = \mathcal{L}^d$ and for the weak first derivative we obtain the measure of variation of f or a perimeter measure if $f = \chi_E$. We will denote this measure with $\|Df\|$ and its density by $|Df|$. If $f \in C^1(U)$, the weak derivative is equal to the strong derivative, leading to the much more explicit equivalence

$$TV(f) = \int_U |\nabla f(x)| \, \mathrm{d}x.$$

So far we looked at the Lebesgue measure only, but of course the results above are not limited to that. For this reason we generalize the notion of total variation to bounded probability measures supported on a set.

Definition 3.3 (Weighted Total Variation)**.** Let $U \subset \mathbb{R}^d$ be open and bounded with Lipschitz boundary. Assume $\rho : U \to (0, \infty)$ to be a continuous density to the probability measure ν supported on U. Furthermore, ρ is bounded on the U by constants $\Lambda \geq \lambda > 0$ with $\lambda \leq \rho \leq \Lambda$. Given a function $u \in L^1(U, \nu)$, we define the weighted total variation of u w.r.t. ρ^2 by

$$TV(u, \nu) = \sup \left\{ \int_U u \text{ div } \phi \, \mathrm{d}x \mid \phi \in C_c^1(U, \mathbb{R}^n), |\phi(x)| \leq \rho^2(x), \forall x \in U \right\},$$

where in the $C_c^1(U, \mathbb{R}^d)$ denotes the set of C^1-functions from U to $\mathbb{R}^d$, whose support is compactly contained in U.

Since $\rho(x)^2$ is lower semi-continuous and bounded below and above by positive constants it belongs to the class of weights considered in [3] where the weighted total variation is studied. There exists a weighted pendant to the structure theorem, see [3, Theorem 3.3]. Consequently, the above arguments hold and if u is sufficiently regular, e.g. C^1, we still have

$$TV(u, \nu) = \int_U |\nabla u| \rho^2 \mathrm{d}x.$$

Moreover, if $u = 1_A$ is the characteristic function of a set $A \subset \mathbb{R}^d$ with C^1-boundary the weighted total variation is equal to the perimeter of A. Specifically, it can be written as

$$TV(1_A, \nu) = \int_{\partial A \cap U} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1}.$$

In case ν is the uniform distribution, ρ is a constant and the functional $TV(\cdot, \nu)$ reduces to a multiple of the classical total variation. For a set $A \subset \mathbb{R}^d$ $TV(1_A, \nu)$ describes the multiplied surface area of ∂A contained in U. Since ρ is bounded above and below by positive constants, a function $u \in L^1(U)$ has finite weighted total variation if and only if it has finite classical total variation. As before, the distributional derivative is

therefore a Radon measure. We denote the space of functions with bounded weighted total variation by $BV_\nu(U)$.

Having introduced the concept of BV functions, we can now consider the two very important properties that will be needed in the later chapters. All of these results are provided as found in [11], with some adaption for the weight $\rho^2(x)$ from [3]. The first property, deals with the weak convergence of variation measures implied by a sequence of functions.

Proposition 3.2 (Lower semi-continuity of variation measures)**.** *Let $f_k \in BV_\nu(U)$ and $f_k \to f$ in $L^1(U,\nu)$. Then*

$$\|Df\|_\nu(U) \leq \liminf_{k\to\infty} \|Df_k\|_\nu(U).$$

Proof. Let $\phi_k \in C_c^1(U)$ and $|\phi_k| \leq \rho^2(x)$ as in the definition of the weighted total variation. Since $TV(f_k) < \infty$ it follows that $\mathrm{div}\phi_k$ is bounded by some constant $M > 0$ almost everywhere. Since $f_k \in BV_\nu(U)$ it holds

$$\begin{aligned}\left|\int_U f_k \mathrm{div}\phi \, \mathrm{d}\nu\right| &\leq \int_U |f_k||\mathrm{div}\phi| \, \mathrm{d}\nu \\ &\leq M\|f_k\|_{L^1(\nu)}.\end{aligned}$$

Therefore the functional $f \mapsto \int_U f\mathrm{div}\phi \, \mathrm{d}\nu$ is continuous in $L^1(\nu)$. Since $\|Df\|_\nu(U)$ is defined as a supremum over a family of continuous functions, lower semi-continuity follows. □

Proposition 3.3 (Local approximation by smooth functions)**.** *Assume $f \in BV_\nu(U)$. Then there exists a sequence of functions $\{f_k\}_{k\in\mathbb{N}} \subset BV_\nu(U) \cap C^\infty(U)$ such that*

$$f_k \overset{L^1(\nu)}{\longrightarrow} f \quad \text{and} \quad \|Df_k\|_\nu(U) \to \|Df\|_\nu(U),$$

as $k \to \infty$.

This proof is rather long and technical; we therefore refer the reader to [11, Theorem 5.3]. A generalisation for the weighted total variation works analogously and can be found in [3, Theorem 3.4]. Both proofs rely on an approximation using smooth mollifier functions (for an introduction see [11, Section 4.2.1]). We find this result is particularly helpful later on, as we work in a setting where f is the characteristic function of a set, which we need to approximate for the so-called recovery sequence. There the boundary of the set might be very unsmooth, which makes it difficult to work with. The above proposition makes our lives much easier by providing a pleasant sequence of smooth functions that we can work with.

3.3 General Perimeter Problems

The isoperimetric problem, which was introduced in the first section, serves as an excellent starting point for general perimeter problems. Generally speaking, the two aspects that distinguish different problems are the measures and the constraints.

In perimeter problems, the constraints can vary widely. They often include a requirement for the set's volume, either explicitly or implicitly. Adhering to particular topological features or boundary assumptions is also common. These are often used to account for physical limitations or additional information, such as labelled sets, that influence the problem. An introduction to these goes beyond the scope of this work, so we will restrict ourselves to the simple volume constraint. Although, in the next chapter, we will explore an alternative approach to addressing the volume.

For a general isoperimetric problem, we require both a volume and a perimeter measure. Given a suitable measure ν, we have already introduced the a corresponding perimeter measure with the weighted total variation in Definition 3.3. If ν is a probability measure supported on D and $\rho : D \to (0, \infty)$ its continuous density with respect to the Lebesgue measure, then

$$\mathrm{vol}(A) = \nu(A) = \int_A \rho \mathrm{d}x \quad \text{and} \quad \mathrm{Per}(A) = TV(1_A, \nu) = \int_A \rho^2 \mathrm{d}\mathcal{H}^{d-1}.$$

Here $A \subset D$ and $\mathcal{H}^{d-1}$ the $d-1$-dimensional Hausdorff measure. To prove the existence of a solution of the isoperimetric problem, we need two lemmas. The first of these will also come in quite handy later on.

Lemma 3.1. *Let $A \subset D$ where $D \subset \mathbb{R}^d$. Then the set of characteristic functions 1_A is closed in $L^1(U, \nu)$. Also, the subset of characteristic functions of sets with volume $\nu(A) = m \in (0, 1]$, is closed in $L^1(U, \nu)$.*

Proof. Let $\{1_{A_k}\}_{k\in\mathbb{N}}$ be sequence of characteristic functions, w.o.l.g $\nu(A_k) > 0$. Let $f \in L^1(U, \nu)$ and $1_{A_k} \xrightarrow{L^1} f$. We need to show f is a characteristic function. It holds

$$\begin{aligned} 0 &= \lim_{k\to\infty} \|1_{A_k} - f\|_{L^1(u,\nu)} \\ &= \lim_{k\to\infty} \int_U |1_{A_k} - f| \, \mathrm{d}\nu(x) \\ &= \lim_{k\to\infty} \int_{A_k} |1 - f| \, \mathrm{d}\nu(x) + \int_{A_k^c} |f| \, \mathrm{d}\nu(x). \end{aligned}$$

We deduce that there exits some set $A \subset D$, such that $\nu(A_k \backslash A) \overset{k\to\infty}{\longrightarrow} 0$, where $f(x) = 1 \, \forall x \in A$ and $\int_{A^c} f \, \mathrm{d}\nu(x) = 0$.

Given $m \in \mathbb{R}$, if it holds $\int_U 1_{A_k} \mathrm{d}\nu(x) = m$ for all k, then by $\nu(A_k \backslash A) \overset{k\to\infty}{\longrightarrow} 0$ we conclude $\nu(A) = m$. □

Remark. If $f_k = a_k 1_{A_k}$ and the convergences $a_k \to a$ and $f_k \to f$ hold, then the above lemma implies that f is of the from $a1_A$.

Lemma 3.2 (Compactness of BV functions)**.** *Let $D \subset \mathbb{R}^d$ be open and bounded with Lipschitz boundary ∂D. Assume $f_k \in BV_\nu(D)$ is a sequence satisfying $\sup_k \|f_k\|_{BV_\nu(D)} < \infty$. Then there exists a subsequence f_{k_j} and function $f \in BV(D)$, such that for $j \to \infty$*

$$f_{k_j} \to f \text{ in } L^1(\nu).$$

Proof. For $k = 1, 2, \ldots$, choose $g_k \in C^\infty(D, \nu)$ such that

$$\int_U |f_k - g_k| \, d\nu < \frac{1}{k} \text{ and } \int_U \|Dg_k\|_\nu \, d\nu < \infty.$$

According to Proposition 3.3 such functions exist, since we can locally approximate f_k in L^1 by smooth functions of bounded variation. [3, Remark 7] states that $g_k \in W^{1,1}(D, \nu)$. By construction it holds $\|f_k - g_k\|_{L^1} < \frac{1}{k}$ and also for the subsequence $f_{k_j} \to f$ in $L^1(D)$. Furthermore, $\|Df\|(D) \leq \liminf_{j\to\infty} \|Df_{k_j}\|(D) < \infty$, so $f \in BV(D)$. □

Example 3.1 (Existence isoperimetric solution)**.** Let $D \subset \mathbb{R}^n$ be an open and bounded set. There exists a solution to the isoperimetric problem

$$\min_{A \subset D : \nu(A) = m} TV(1_A, \nu)$$

for $\Omega \subset D$ and $0 < m < \mathcal{L}(D)$.

Our set of admissible functions is $A = \{f \in BV(U, \{0, 1\}) : \int_D f \, dx = m\}$. Since $\mathcal{L}^n(D) < \infty$, there exist a covering with countable balls of finite radius, meaning we can find N such that $\mathcal{L}^n(\bigcup_{k \leq N} B_k) \geq m$. This implies there exists a set Ω with $\mathcal{L}^n(\Omega) = m$ and finite perimeter $\mathrm{Per}(\Omega) \leq \mathrm{Per}\left(\bigcup_{k \leq N} B_k\right) \leq \sum_{k \leq N} \mathrm{Per}(B_k) < \infty$. As such A is not empty.

For any minimizing sequence $f_k \in A$ there exists a converging subsequence $f_{k_j} \to f^*$in $L^1(D)$ by Proposition 3.2. According to Lemma 3.1 it follows $f \in A$. Finally, the lower semi-continuity of variation measures Proposition 3.2 guarantees

$$\|Df^*\|(D) \leq \liminf_{j\to\infty} \|Df_{k_j}\|(D),$$

meaning f^* is the characteristic function with minimal total variation and there for the solution of the isoperimetric problem.

Perimeter problems are not limited to single sets. For example, consider partitioning the domain $D \subset \mathbb{R}^d$ into $R \in \mathbb{N}$ sets, which is a straightforward instance of a multi-set perimeter problem. We are therefore looking for a collection of sets $A_1, \ldots, A_R$ that

are pairwise disjoint and satisfy $\bigcup_{r=1}^{R} A_r = D$. As in many cases, there are several equivalent ways to formulate these two constrains. The latter, for example, could be expressed almost everywhere as $\sum_{r=1}^{R} 1_{A_r} > 0$ or via volume $\mathrm{vol}(D) = \sum_{r=1}^{R} \mathrm{vol}(A_r)$. The disjointness of the sets is equivalent to an orthogonality constraints on their characteristic functions. We can either require that $\sum_{r=1}^{R} 1_{A_r} \leq 1$ a.e. or use a pairwise approach $\int_D u_r u_s \mathrm{d}\nu = 0$, for $r, s \in \{1, \ldots, R\}$ and $r \neq s$. Depending on the context or techniques used, one approach may be preferable to the others, but finding suitable relaxations of the constraints is almost always important, especially when related to applications.

Apart from the given constraints, the optimization of any single set is independent of the others. We formulate a general isoperimetric problem for R sets and volumes $m \in \mathbb{R}$ over a tuple of functions $U = (u_1, \ldots, u_R)$ as

$$\begin{aligned} \min_{U \in L^1(D,\nu)^R} & \sum_{r=1}^{R} TV(u_r, \nu) \\ \text{s.t.} & \int_D u_r u_s \mathrm{d}\nu = 0 \text{ and } \int_D u_r \mathrm{d}\nu = m. \end{aligned} \tag{3.3}$$

4 Discrete to Continuum

4.1 Graph Total Variation

In the second chapter, we discussed graph clusters and their relationship to perimeter problems. We also emphasised the importance of consistency between graph solutions and the continuum solution involving the sampled density. We then saw how the perimeter can be formulated in a continuum setting using total variation and its implied variation measure. Currently, these perimeter concepts are unrelated, so it is not possible to draw any comparisons. To address this, we introduce Graph Total Variation: a rescaled graph cut involving a special class of weight functions. This allows us to study the graph cut within a measure-theoretic framework, enabling direct comparisons using the theory of optimal transport. Furthermore, graph problems will become easier to view as a relaxation of the continuum case, and we will be able to prove the convergence of minimizers using the concept of Γ-convergence.

Let $D \subset \mathbb{R}^d$ be an open bounded subset and $\rho : D \to (0, \infty)$ a continuous density to the probability measure ν, supported on D. We draw samples $X_n := \{x_i\}_{i=1,\dots,n}$ from ρ for a vertex set $V = X_n$. For the weight function we choose a kernel $\eta : \mathbb{R}^d \to \mathbb{R}$ together with a parameter $\varepsilon > 0$.

From this combinations the graph weights are given as

$$w_{ij} = \frac{1}{\varepsilon^d}\eta\Big(\frac{x_i - x_j}{\varepsilon}\Big).$$

We will also use the notation $\eta_\varepsilon(x_i - x_j) := \frac{1}{\varepsilon^d}\eta\Big(\frac{x_i - x_j}{\varepsilon}\Big)$. One example for such a kernel is the Gaussian similarity function. In this example the kernel is symmetric thanks to the norm used in the denominator, creating an undirected graph. The ε has the same role as before, namely for larger values of ε the graph becomes more densely connected while it might be disconnected for values that are too small. There are many assumptions one can impose on the kernels used to create the graph, especially if only given a set of vertexes. If the graph is given on the other hand, that is of course a different matter. Common assumptions imposed on kernels include positivity and continuity at zero. While these assumptions are not necessary for defining graphs, they are often quite natural.

The quantity we will actually work with is called 'Graph Total Variation' for reasons that will become apparent later.

Definition 4.1 (Graph Total Variation)**.** Given a set of points $X_n := \{x_1, ..., x_n\} \subset D$. Let $\eta : \mathbb{R}^d \to [0, \infty)$ be a similarity kernel that compactly supported on D, non-increasing, continuous at 0 and satisfies $\eta(0) > 0$. Let G be the graph whose edge weights are given by $w_{ij} = \frac{1}{\varepsilon^d}\eta\Big(\frac{x_i - x_j}{\varepsilon}\Big)$ for $1 \leq i, j \leq n$. Then we denote the Graph Total Variation by

$$GTV_{n,\varepsilon_n}(u) := \frac{1}{\varepsilon_n n^2} \sum_{i,j=1}^{n} w_{ij}|u(x_i) - u(x_j)|.$$

Recall our definition $|Y| = \frac{\#Y}{n}$ for $Y \subset V$. We can extend this to a measure on D with the Dirac measure $\delta_x(y) = 1$ iff $x = y$ and zero otherwise. We obtain the discrete measure $\nu_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}(A)$ for $A \subset D$, which allows us to reformulate using integrals instead of sums via

$$GTV_{n,\varepsilon_n}(u_n) = \frac{1}{\varepsilon_n} \int_{D\times D} \eta_\varepsilon(x - y)|u(x) - u(y)| \, \mathrm{d}\nu_n(x)\mathrm{d}\nu_n(y).$$

This should also make it clear that the total variation of a graph depends on the underlying measure. Note that we will work with $GTV_{n,\varepsilon_n} : TL^1(D) \to [0, \infty]$ (see (4.5)) defined by

$$GTV_{n,\varepsilon_n}(\mu, u_n) := \begin{cases} GTV_{n,\varepsilon_n}(u_n) & \text{if } \mu = \nu_n \\ +\infty & \text{otherwise.} \end{cases}$$

Since we only consider GTV_{n,ε_n} for $\mu = \nu_n$, we will neglect this argument in the following and always write $GTV_{n,\varepsilon_n}(u_n)$.

4.2 Γ-Convergence

Problem relaxations are a very common approach in applications. The problem is modified so that it can be solved more easily, and the solution serves as a good starting point for a subsequent, more closely related problem. The idea is to work with a sequence of relaxations, where successive tightening of the problem leads to convergence of the minimisers to the original solution. The graph cut problem can be interpreted as a relaxed version of the continuum case, in which we limit ourselves to a finite number of points. A well-known tool for proving the convergence of relaxed problems is called Γ-convergence.

Later, we will work with graphs based on a data sample from a probability distribution. The functional depends on the realisation of a random variable, making it a random functional. In this section, let (X, d_X) be a metric space and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $F_n : X \times \Omega \to [0, \infty]$ be a sequence of random functionals. We simply write$F_n(x)$ if the element $\omega \in \Omega$ has been fixed. Here ω describes a realization of the some distribution that takes values in (X, d_X).

Definition 4.2 (Γ-Convergence)**.** We say the sequence of random functionals $\{F_n\}_{n\in\mathbb{N}}$ Γ-converges with respect to metric d_X to the deterministic functional $F : X \to [0,\infty]$ as $n \to \infty$ if for $\mathbb{P}$-almost every ω, the following conditions hold:

(i) **Liminf inequality**: For every $x \in X$ and every sequence $\{x_n\}_{n\in\mathbb{N}}$ converging to x,
$$\liminf_{n\to\infty} F_n(x_n) \geq F(x),$$

(ii) **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 call F the Γ-limit of the sequence F_n and write $F_n \xrightarrow{\Gamma} F$.

In practice, the liminf property is often relatively easy to verify. The limsup property on the other hand is much harder to prove, since we must construct a converging sequence for every single point $x \in X$. This is the so called the recovery sequence. Sometimes it is sufficient to use the constant sequence$x_n := x$, but not in our case though. However, there is another property that we can make use of. Rather than proving the limsup inequality directly for all $x \in X$, we show it on a dense subset $X' \subset X$, where it is somewhat easier to prove. We can then deduce that the inequality holds for the whole set X.

Proposition 4.1 (Γ-Convergence on dense subsets)**.** *Let X be a metric space and functionals $F_n : \Omega \times X \to [0,\infty]$, $F : X \to [0,\infty]$. Let $X' \subset X$ be a dense subset, such that for all $x \in X'$ there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$, $x_n \to x$ where*
$$\limsup_{n\to\infty} F_n(x_n) \leq F(x).$$
If F is continuous in X, then the inequality holds for all $x \in X$.

Proof. Let F be continuous in X and let the inequality hold for all $x' \in X'$. We take an arbitrary $x \in X$. Since X' is a dense subset there exists a sequence $x_n \subset X'$ with $x_n \to x$ and $F(x_n) \to F(x)$ by assumption. For n large enough there exists $\tilde{x} \in X'$ such that $d_X(\tilde{x} - x) \leq \delta$ and $|F(\tilde{x}) - F(x)| \leq \varepsilon$ by continuity. Now, without loss of generality we assume $F(\tilde{x}) - F(x) \geq 0$, otherwise we look at $F(x) - F(\tilde{x})$. As $\tilde{x} \in X'$, by assumption there exists a sequence $\tilde{x}_n$ with
$$\limsup_{n\to\infty} F_n(\tilde{x}_n) \leq F(\tilde{x}) \leq F(x) + \varepsilon.$$
This is equivalent to the limsup inequality by [6, Theorem 1.17 (iii), approximate limsup inequality (ii)"] □

This property is not related to the randomness of the functionals in any way. Let us look at some other properties of the Γ-limit that will come in useful later.

Lemma 4.1 (Simple Properties of the Γ-Limit)**.** *Let F_n be sequences of non-negative functionals that Γ-converge to F and let $G : X \to \mathbb{R}$ be continuous, such that $G(x) \neq 0, \forall x \in X$. Then the following properties hold for the limit*

(i) $\Gamma\text{-}\lim_{n\to\infty} (F_n + G) = F + G.$

(ii) $\Gamma\text{-}\lim_{n\to\infty} \frac{F_n}{G} = \frac{F}{G}.$

Proof. Let $F_n \xrightarrow{\Gamma} F$ on $X \subset \mathbb{R}^d$ and let $G : X \to \mathbb{R}$ be continuous. We have to show the liminf and limsup property for the combined limit, using the Γ-limit of F_n and continuity of G. Let $x_n \to x$ be an arbitrary converging sequence and let z_n be recovery sequence of F_n to any $z \in X$.

(i) It holds:

$$\begin{aligned}\liminf_{n\to\infty}(F_n + G)(x_n) =& \liminf_{n\to\infty} F_n(x_n) + \lim_{n\to\infty} G(x_n)\\ =& \liminf_{n\to\infty} F_n(x_n) + G(x)\\ \geq& F(x) + G(x).\end{aligned}$$

We can use z_n for the recovery sequence, since

$$\begin{aligned}\limsup_{n\to\infty}(F_n + G)(z_n) =& \limsup_{n\to\infty} F_n(z_n) + \limsup_{n\to\infty} G(z_n)\\ \geq& F(z) + \lim_{n\to\infty} G(z_n)\\ =& F(z) + G(z)\end{aligned}$$

(ii) follows completely analogously. □

Definition 4.3 (Compactness Property)**.** We say that the sequence of non-negative random functionals $\{F_n\}_{n\in\mathbb{N}}$ satisfies the compactness property if for $\mathbb{P}$-almost every ω , the following statement holds: any sequence $\{x_n\}_{n\in\mathbb{N}}$ bounded in X and for which

$$\limsup_{n\to\infty} F_n(x_n) < \infty,$$

is relatively compact in X.

The main reason why Γ-convergence is such a popular techniques is its very strong property of converging minimizers if they satisfy the compactness property. The following proposition assumes that we have a sequence of relaxed functionals that we successively tighten for $n \to \infty$. Any such functional F_n has a corresponding solution x_n. The proposition states that if the difference between this solution and the infimum of the functional converge to zero, then any clustering point of the sequence x_n is in fact a minimizer of the Γ-limit.

Proposition 4.2 (Convergence of minimizers)**.** *Let X be a metric space and $F_n : \Omega \times X \to [0, \infty]$ random functionals, satisfying the compactness property and Γ-converging to $F : X \to [0, \infty)$. Suppose that for $\mathbb{P}$ almost every ω there is a bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ such that*

$$\lim_{n\to\infty} \left(F_n(x_n) - \inf_{x\in X} F_n(x) \right) = 0. \tag{4.1}$$

Then with $\mathbb{P}$ probability one any such sequence is relatively compact and each of it's clustering points is a minimizer of F. It holds

$$\lim_{n\to\infty} \inf_{x\in X} F_n(x) = \min_{x\in X} F(x). \tag{4.2}$$

Proof. This proof can be found in [14, Proposition 17]. Consider fixed $\omega \in \Omega$ such that the assumptions of the statement hold. Let $\{x_n\}_{n\in\mathbb{N}}$ be sequence as described above and let $\tilde{x} \in X$ arbitrary. By the limsup inequality there exists a sequence $\{\tilde{x}_n\}_{n\in\mathbb{N}}$ with $\tilde{x}_n \to \tilde{x}$ and $\limsup_{n\to\infty} F_n(\tilde{x}_n) \leq F(\tilde{x})$. By the convergence (4.1) it holds

$$\limsup_{n\to\infty} F_n(x_n) = \limsup_{n\to\infty} \inf_{x\in X} F_n(x) \leq \limsup_{n\to\infty} F_n(\tilde{x}_n) \leq F(\tilde{x}). \tag{4.3}$$

Since $\tilde{x}$ was arbitrary it holds for any $x \in X$, including the infimum of F, that

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

Furthermore, this means $\limsup_{n\to\infty} F_n(x_n) \leq \infty$. As the sequence $\{x_n\}_{n\in\mathbb{N}}$was assumed to be bounded,we can conclude that it is relatively compact.

Now we show the minimizing properties. Let x^*be any accumulation point of $\{x_n\}_{n\in\mathbb{N}}$, of which at least one exists because of compactness. Without loss of generality, assume $x_n \to x^*$ otherwise we consider a subsequence. By the liminf inequality and definition of the infimum we deduce

$$\inf_{x\in X} F(x) \leq F(x^*) \leq \liminf_{n\to\infty} F_n(x_n). \tag{4.4}$$

With the above observations it follows that for arbitrary$\tilde{x}$

$$F(x^*) \leq \liminf_{n\to\infty} F_n(x_n) \leq \limsup_{n\to\infty} F_n(x_n) \leq F(\tilde{x})$$

and thus x^* is a minimizer of F. In particular we have that $\inf_{x\in X} F(x) = \min_{x\in X} F(x)$. We conclude the proof by establishing (4.2) through

$$\begin{aligned}
\lim_{n\to\infty} \inf_{x\in X} F_n(x) &\leq \limsup_{n\to\infty} \inf_{x\in X} F_n(x) \leq \inf_{x\in X} F(x), \\
\lim_{n\to\infty} \inf_{x\in X} F_n(x) &\geq \liminf_{n\to\infty} \inf_{x\in X} F_n(x) \geq \inf_{x\in X} F(x),
\end{aligned}$$

by (4.3) and (4.4). □

4.3 Optimal Transport Maps

To address the obvious question of how the minimum cut solution behaves on graphs compared to the continuum partition of the domain, we need additional tools. Specifically, we need a way to compare functionals evaluated in different spaces.

More precisely, given a point cloud $X_n = \{x_1, \dots, x_n\} \subset D$ we need to compare its empirical measure ν_n with the measure ν from which's density we sampled. In this section, we introduce a concept of convergence to compare the discrete and continuum partitions by examining their associated characteristic functions. To achieve this, we need a method of comparing L^1 functions with respect to different measures.

We denote by $\mathcal{B}(D)$ the Borel σ-algebra on D and by $\mathcal{P}(D)$ the set of Borel probability measures on D. The set of objects of our interest is

$$TL^p(D) := \{(\mu, f) : \mu \in \mathcal{P}(D),\ f \in L^p(\mu)\}.\acute{} \tag{4.5}$$

The elements of $TL^p(D)$ are tuples of p-integrable functions with "their" respective measures. Therefore, for two distinct measures $\mu, \theta \in \mathcal{P}(D)$, (μ, f) and (θ, f) are separate elements. In particular, for $Y_n \subset X_n$ and $A \subset D$ note that $(\nu_n, 1_{Y_n})$ and $(\nu, 1_A)$ both belong to $TL^p(D)$. For comparing two measures $\mu, \theta \in \mathcal{P}(D)$ we focus on the case where one of the measures, say μ, is absolutely continuous with respect to the Lebesgue measure. Given a Borel-measurable map $T : D \to D$ and $\mu \in \mathcal{P}(D)$, we define $T_{\#}\mu \in \mathcal{P}(D)$ the push-forward measure of μ by T as

$$T_{\#}\mu(A) := \mu(T^{-1}(A))$$

for $A \in \mathcal{B}(D)$. T is called a transportation map if $T_{\#}\mu = \theta$. Then the following change of variables formula holds for $f \in L^1(\theta)$

$$\int_D f \,\mathrm{d}\theta(x) = \int_D T(f(x)) \,\mathrm{d}\mu(x).$$

Note that the notion of transport maps can be generalized to couplings with transportation plans, see e.g. [13, Part 3]. In the case of absolute continuity with respect to the Lebesgue measure, however, transport maps enable us to reduce the challenging problem of comparing measures to the relatively straightforward comparison of maps. More precisely, we compare the transport map to the identity map under one of the measures. Now, for convergence in the space $TL^p(D)$, we need a concept for a sequence of transport maps that becomes closer to the identity. Let $\mu, \mu_n \in \mathcal{P}(D)$, for $n \in \mathbb{N}$, and a sequence of transport maps $\{T_n\}_{n\in\mathbb{N}}$ such that $\mu_n = T_{n\#}\mu$. We call this sequence stagnating if

$$\left(\int_D |T_n - x|^p \,\mathrm{d}\mu(x)\right)^{\frac{1}{p}} \underset{n\to\infty}{\longrightarrow} 0.$$

So an empirical measure ν_n converges to the measure ν if the mass of ν needs to be moved very little to match the distribution of ν_n as $n \to \infty$. The following proposition

formalizes this interpretation of the convergence in TL^p, note the denotation of the weak convergence of measures with $\xrightarrow{w}$. The complete result with proof can be found in [13, Proposition 3.12].

Proposition 4.3 (TL^p-Convergence)**.** *Consider $\mu \in \mathcal{P}(D)$ an absolutely continuous measure with respect to the Lebesgue measure. Let $(\mu, f) \in TL^1(D)$ and $\{(\mu_n, f_n)\}_{n\in\mathbb{N}}$as sequences in $TL^p(D)$. The following statements are equivalent*

(i) $(\mu_n, f_n) \xrightarrow{TL^1} (\mu, f)$ as $n \to \infty$.

(ii) $\mu_n \xrightarrow{w} \mu$ and there exists a stagnating sequence of transportation maps $T_{n\#}\mu = \mu_n$such that
$$\left(\int_D |f(x) - f_n(T_n(x))|^p \,\mathrm{d}\mu(x)\right)^{\frac{1}{p}} \underset{n\to\infty}{\longrightarrow} 0.$$

(iii) $\mu_n \xrightarrow{w} \mu$ and for any stagnating sequence of transportation maps $T_{n\#}\mu = \mu_n$it holds
$$\left(\int_D |f(x) - f_n(T_n(x))|^p \,\mathrm{d}\mu(x)\right)^{\frac{1}{p}} \underset{n\to\infty}{\longrightarrow} 0.$$

So for convergence of a sequence of tuples (μ_n, f_n) to converge to (μ, f) in the TL^p sense, it suffices to find one sequence of stagnating transport maps and show $L^p(\mu)$-convergence of $f_n \circ T_n$ to f. In particular, there exist several sequences of stagnating transport maps, we are free to choose either of them. Also note, if $\mu_n = \mu$ for all n, convergence in TL^p is equivalent to convergence in $L^p(\mu)$. In case of a sequence of probability measures $\mu_n \xrightarrow{w} \mu$, we will simplify the notation to $f_n \xrightarrow{TL^p} f$ if it is clear what the corresponding measures are.

Next, we need precise information on the rate of convergence and as such on the transportation distance $T_n(x) - x$.

Proposition 4.4 (Bound on Transport Distance)**.** *Let $D \subset \mathbb{R}^d$ be an open, connected and set with Lipschitz boundary. Let ν be a probability measure on D with density ρ, such that $\lambda \leq \rho(x) \leq \Lambda$ for positive constants λ, Λ. Let $x_1, \ldots, x_n$ be a sequence of independent random points sampled from D distributed according to ρ. Define $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Then, there exists a constant $C > 0$ (depending on D, ρ), such that with probability one a sequence of transportation maps $\{T_n\}_{n\in\mathbb{N}}$ from ν to ν_n exists and*

$$\limsup_{n\to\infty} \frac{n^{1/d}}{(\log n)^{1/p_d}} \|Id - T_n\|_{L^\infty(\nu)} \leq C. \tag{4.6}$$

Here, the power $p_d = 1/d$ if $d \geq 3$ and $p_d = 3/4$ if $p = 2$.

The result is a combination of [12, Theorem 1.1] and [12, Theorem 1.2]. The optimality of the upper bound is discussed in the same paper. The scaling condition comes directly

from the existence of transportation maps for this Proposition. It 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 X_n. In other words, the graph must contain information on a larger scale than that on which the intrinsic randomness operates. As such, the result is based on the the same principle as the connectedness result Proposition 2.1, where we utilize the fact that with large probability none of the points x_i are contained in a ball of radius approximately $\left(\frac{n}{\log n}\right)^{1/d}$. This estimate assumes the points x_i to be i.i.d. distributed according to ν, a common and quite reasonable randomness assumption when modelling randomly obtained data. If however, the points are more regularly distributed or given deterministically, much tighter bounds on the transportation distance could be obtained. This would also translate into better bounds for all the Γ-convergence results presented in the following two chapters.

An additional problem arises when considering the concept of TL^p convergence for sets. Namely, the ambiguity of partitions under permutations. Therefore, the following definition defines the convergence for partitions in a natural way.

Definition 4.4. Let $Y_n = \{Y_n^1, \dots, Y_n^R\}$ be a partition of X_n. The sequence $\{Y_n\}_{n\in\mathbb{N}}$ converges in the TL^1-sense to the partition $A_n = \{A_n^1, \dots, A_n^R\}$ of D, if there exists a sequence of permutations $\{P_n\}_{n\in\mathbb{N}}$ such that

$$(\nu_n, 1_{Y_n^{P_n(r)}}) \xrightarrow{TL^1} (\nu, 1_{A_n^r}),$$

for every $r \in \{1, \dots, R\}$ as $n \to \infty$.

Another implication of this definition is that the convergence is componentwise for multiple sets. We therefore define on the space of R-tuples of $L^p(\mu)$ functions, namely

$$TL^p(D)^R := \{(\mu, U) : \mu \in \mathcal{P}(D), U = (u_1, \dots, u_R), u_i \in L^p(\mu) \text{ for } i = 1, \dots, R\}. \tag{4.7}$$

Furthermore, we denote the space of R-tuples of $L^p(\mu)$ functions $L^p(\mu)^R$. Any function applied to a function $U \in L^p(\mu)^R$ is defined componentwise; for example, consider the balance term in the next chapter.

5 Cheeger-Cuts

5.1 Introduction

The isoperimetric problems we have considered so far have had hard volume constraints on the objective function. This kind of fixed constraint requires advanced knowledge of the application, and if the optimal volume is unknown beforehand, it becomes a hyperparameter of the optimization problem. In contrast, the Cheeger approach incorporates a volume term, known as the balance term, into the optimization objective. More precisely, for a single set $A \subset D$, these optimization problems are of the form

$$\min_{A \subset D} \frac{\mathrm{Per}(A)}{\mathrm{Vol}(A)}.$$

In general, reducing the perimeter or increasing the volume of A will diminish the objective value. Of course, there is a trade-off when trying to minimize the perimeter while maximizing the volume at the same time. Therefore, part of the question is how much of an increase in perimeter are we willing to accept for an increase in volume. This design choice will be controlled via a new parameter, α.

The content of this chapter is primarily based on [14] with the inspiration for introducing α taken from [4]. While [14] aim for a partition of the domain D, this constraint must be dropped for α to have any impact. Since we permit "empty" regions in a controlled manner, the results in this chapter are slightly more generalised than those in the aforementioned.

Throughout this chapter let $D \subset \mathbb{R}^d$ be open, bounded and connected with Lipschitz boundary. Furthermore, we make to following assumptions to be referenced later.

Assumptions 5.1. Let $\rho : D \to (0, \infty)$ be a continuous density to the probability measure ν supported on D. We assume ρ is bounded on D by constants $\Lambda \geq \lambda > 0$ with $\lambda \leq \rho \leq \Lambda$. The set $X_n = \{x_1, \ldots, x_n\}$ are sampled points from density ρ. We consider an isotropic similarity kernel η with radial profile $\boldsymbol{\eta} : [0, \infty) \to [0, \infty)$ and define it as $\eta(x) := \boldsymbol{\eta}(|x|)$. We assume

(i) $\boldsymbol{\eta}(0) > 0$ and $\boldsymbol{\eta}$ continuous at 0.

(ii) $\boldsymbol{\eta}$ is non-increasing.

(iii) $\sigma_\eta := \int_{\mathbb{R}^d} \eta(h)|h_1| \mathrm{d}h < \infty$.

The quantity σ_η is also referred to as the surface tension. Here, $h_1$denotes the first coordinate of the vector h but could be replaced with $\langle h, e\rangle$, for e some unit vector in $\mathbb{R}^d$. The class of admissible kernels is broad and includes both Gaussian kernels and discontinuous kernels of the form $\boldsymbol{\eta} = 1$ for $x \leq 1$ and $\boldsymbol{\eta} = 0$ if $x > 1$.

The perimeter is measured by the weighted total variation $TV(u, \nu)$ and a suitable quantity for what we denoted with $\mathrm{Vol}(A)$ will be introduced in the next section. For the resulting functional, we show a Γ-convergence and compactness result that allow us to infer the consistency of graph cuts as the number of vertices approaches ∞.

5.2 The Balance Term

Definition 5.1 (Continuum Balance Term)**.** Given $u \in L^1(\nu)$ we define continuum balance term with respect to measure ν and parameter $\alpha \in (0, \infty)$ as

$$B^\alpha(u) = \left(\int u \,\mathrm{d}\nu(x)\right)^\alpha = \left(\int u(x)\rho(x)\,\mathrm{d}x\right)^\alpha. \tag{5.1}$$

The parameter α allows us to control the relation of volume and perimeter. Since ν is a probability measure, if u is an indicator function, $\alpha > 1$ diminishes the value of B^α while $\alpha < 1$ increases it. In any case it holds $0 \leq B^\alpha(1_A) \leq 1$ for some set $A \subset D$. We mostly use the balancing term for characteristic function and we may simplify the notation to $B^\alpha(A) = B^\alpha(1_A)$. In the instance of a Cheeger cut for single sets, we will use the so called Cheeger Balance Term $\min(B^\alpha(1_A), B^\alpha(1_{A^c}))$. Otherwise we may run into problems because of the ambiguity of partitions under permutations. For now this not relevant although it will be in the next chapter.

To define the corresponding discrete balancing term, recall that we defined the volume of a set on a graph as $|Y| = \frac{\#Y}{n}$ for $Y \subset X_n$. We can extend this to a measure on D with the dirac measure $\delta_x(y) = 1$ iff $x = y$ and zero otherwise. Given X_n we obtain the discreet measure $\nu_n(A) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}(A)$ for $A \subset D$.

Definition 5.2 (Graph Balance Term)**.** Let $u_n \in L^1(\nu_n)$. Where we define for $A_n \subset X_n$ the discrete measure $\nu_n(A_n) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}(A_n)$. Then

$$B_n^\alpha(u_n) = \left(\int_D u_n \,\mathrm{d}\nu_n(x)\right)^\alpha = \left(\frac{1}{n}\sum_{i=1}^n u_n(x_i)\right)^\alpha. \tag{5.2}$$

Alternatively, one could also define a graph balance term using the degree of a vertex. Broadly speaking, the additional sum in the degree definition (2.1) leads to a scaling t $\frac{\deg(A_n)}{\deg(X_n)} \sim C_\eta \int_A \rho^2 \,\mathrm{d}x$, see [14, Section 1.3] for some details.

In order to utilize existing results for the total variation, we work with indicator functions scaled by their balance term $\tilde{1}_A := \frac{1_A}{B^\alpha(A)}$. Instead of reformulating the total variation we adapt the sets of functions to consider. We define

$$Ind(D) = \{u \in L^1(\nu) : \ u = \tilde{1}_Y \text{for some measureable set } Y \subset D,\, B^\alpha(Y) \neq 0\}. \tag{5.3}$$

For $u \in Ind(D)$ the continuity of the balance term, Lemma 5.1, and the 1-homogeneity of the total variation allow us to formulate the Cheeger cut objective as

$$TV(u,\nu) = TV(\tilde{1}_A,\nu) = TV\Big(\frac{1_A}{B^\alpha(A)},\nu\Big) = \frac{TV(1_A,\nu)}{B^\alpha(A)}.$$

Using the empirical measure mentioned above we define the discrete pendant as

$$Ind_n(D) = \{u_n \in L^1(\nu_n) :\ u_n = \tilde{1}_{Y_n} \text{for some } Y_n \subset X_n,\ B_n^\alpha(Y_n) \neq 0\}. \tag{5.4}$$

Analogously it holds for $u_n \in Ind_n(D)$

$$GTV_{n,\varepsilon_n}(u_n,\nu_n) = \frac{GTV_{n,\varepsilon_n}(1_{A_n},\nu_n)}{B_n^\alpha(A_n)}.$$

We saw in (3.3) that, for the perimeter problems involving multiple sets, we need an additional orthogonality constraint. Functions that do not satisfy said constrain are of no interest to the problem at hand. With that in mind we define the sets of functions for the multi class Cheeger cut problems as

$$\mathcal{M}(D) = \{(u^1,\dots,u^R) :\ u^r \in Ind(D),\ \int_D u^r(x)u^s(x)\,\mathrm{d}\nu(x) = 0 \text{ if } r \neq s\}, \tag{5.5}$$

$$\mathcal{M}_n(D) = \{(u_n^1,\dots,u_n^R) :\ u_n^r \in Ind_n(D),\ \int_D u_n^r(x)u_n^s(x)\,\mathrm{d}\nu_n(x) = 0 \text{ if } r \neq s\}. \tag{5.6}$$

The Lemma below contains two results which are absolutely essential to our proof. Firstly, the continuity of the balancing term allows us to utilize Lemma 4.1, meaning it is sufficient to show the Γ-convergence of the cut term, greatly reducing the complexity of the proof. Secondly, the sets we minimize over are closed. As such, any sequence has a converging subsequence to an accumulation point, a fact we rely upon for the compactness property. The following Lemma and its proof are based on [14, Lemma 7 and Lemma 26].

Lemma 5.1. *For the continuum balance term defined above, the following holds:*

(i) B^α *is continuous in* $L^1(\nu)$*, as is* $\min(B^\alpha, 1-B^\alpha)$.

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

(iii) The set $\mathcal{M}(D)$ *is closed in* $L^1(\nu)^R$.

Proof. **(i)** Let $u \in L^1(\nu)$ and a sequence $\{u_n\}_{n\in\mathbb{N}} \subset L^1(\nu)$ converging to u in $L^1(\nu)$ a.e. From the L^1convergence $\int_D |u_n - u|\,\mathrm{d}\nu(x) \to 0$ directly follows pointwise convergence along a subsequence to the same limit and so $u_n \to u$. Then, by continuity of the map $x \mapsto x^\alpha$ and the dominated convergence theorem, it holds

$$B^\alpha(u_n) = \Big(\int_D u_n\,\mathrm{d}\nu(x)\Big)^\alpha \to \Big(\int_D u\,\mathrm{d}\nu(x)\Big)^\alpha = B^\alpha(u).$$

Since we proofed B^α to be continuous, the Lipschitz continuity of the minimum function together with the linearity of the transformation $1-B^\alpha$ directly implies the second part of the statement.

(ii) Let $\{u_n\}_{n\in\mathbb{N}}$be a sequence in $Ind(D)$. We need to show that for any $u \in L^1(\nu)$ with $u_n \overset{L^1(\nu)}{\to} u$ it follows $u \in Ind(D)$. First we note that the set of indicator functions is closed in $L^1(\nu)$ as of Lemma 3.1. Since u_n are of the form $a_n 1_{A_n}$ by assumption and $B^\alpha(u_n) \to B^\alpha(u)$, said Lemma implies $u = a1_A$ for some $a \in \mathbb{R}$ and $\nu(A) > 0$. By construction B^α is α-homogeneous, meaning $B^\alpha(a1_A) = (a\int_D 1_A \,\mathrm{d}\nu(x))^\alpha = a^\alpha B^\alpha(1_A)$. Furthermore, the special form of u_n gives us

$$\begin{aligned} B^\alpha(u_n) &= \left(\int_D u_n \,\mathrm{d}\nu(x)\right)^\alpha = \left(\int_D \frac{1_{A_n}}{B^\alpha(A_n)} \,\mathrm{d}\nu(x)\right)^\alpha \\ &= \frac{(\int_D 1_{A_n}\rho \,\mathrm{d}x)^\alpha}{(B^\alpha(A_n))^\alpha} = \frac{B^\alpha(A_n)}{(B^\alpha(A_n))^\alpha}. \end{aligned}$$

The continuity of B^α in $L^1(\nu)$ implies a convergence we can rewrite as

$$\begin{aligned} |B^\alpha(u_n) - B^\alpha(u)| &= |\frac{B^\alpha(A_n)}{(B^\alpha(A_n))^\alpha} - B^\alpha(a1_A)| \\ &= |\frac{B^\alpha(A_n) - B^\alpha(a1_A)\,(B^\alpha(A_n))^\alpha}{(B^\alpha(A_n))^\alpha}| \underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

By definition of $Ind(D)$ it holds $0 < \nu(A_n) \leq \nu(D)$ and so it follows

$$\nu(A_n)^\alpha - a^\alpha\,\nu(A)^\alpha\,(\nu(A_n)^\alpha)^\alpha \underset{n\to\infty}{\longrightarrow} 0.$$

This in turn implies

$$1 - a\,\nu(A)\,\frac{\nu(A_n)^\alpha}{\nu(A_n)} \underset{n\to\infty}{\longrightarrow} 0.$$

By our previous argumentation, it needs to hold $\nu(A_n) \to \nu(A)$. Hence, we deduce

$$a\,\nu(A_n)^\alpha \underset{n\to\infty}{\longrightarrow} 1$$

and $a = \frac{1}{\nu(A)^\alpha} = \frac{1}{B^\alpha(A)}$.Thus, $u = \frac{1_A}{B^\alpha(A)}$ with $B^\alpha(A) \neq 0$ finishing our argument to show $u \in Ind(D)$.

(iii) For some sequence $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ in $\mathcal{M}(D)$ converging to some $\mathcal{U} \in L^1(\nu)^R$ we need to show $\mathcal{U} \in \mathcal{M}(D)$. First of all, note that by definition of $TL^p(D)^R$ in (4.7), it holds a componentwise convergence $u_n^r \overset{L^1(\nu)}{\to} u^r$ for $1 \leq r \leq R$. By (ii) it is therefore enough to show the orthogonality. Again, w.o.l.g. we may assume $u_n^r \to u^r$ for almost every $x \in D$. For $r \neq s$, we have

$$\begin{aligned} 0 &\leq \int_D u^r(x)u^s(x) \,\mathrm{d}\nu(x) \\ &= \int_D \liminf_{n\to\infty} u_n^r(x)u_n^s(x) \,\mathrm{d}\nu(x) \\ &\leq \liminf_{n\to\infty} \int_D u_n^r(x)u_n^s(x) \,\mathrm{d}\nu(x) = 0. \end{aligned}$$

The first inequality follows from $u^r, u^s \in Ind(D)$, while the second one is an application of Fatou's Lemma. □

5.3 Γ-Convergence of α-Cheeger-Cuts

In this section, we will prove Γ-convergence of the graph functionals, one of our main building blocks in showing the consistency of graph cuts. We define the continuum multi class Cheeger cut functional $E : TL^1(D)^R \to [0, \infty]$, where

$$E(\mu, \mathcal{U}) := \begin{cases} \sum_{r=1}^{R} TV(u^r, \nu) & \text{if } \mu = \nu \text{ and } \mathcal{U} \in \mathcal{M}(D) \\ +\infty & \text{else.} \end{cases} \tag{5.7}$$

Recall, that the elements of $\mathcal{M}(D)$ are scaled indicator functions, so we obtain the desired Cheeger objective of the form $TV(u^r, \nu) = \frac{TV(1_{A^r}, \nu)}{B^\alpha(A^r)}$. Likewise, the graph multi class Cheeger cut functional is $E_n : TL^1(D)^R \to [0, \infty]$ with

$$E_n(\mu, \mathcal{U}_n) := \begin{cases} \sum_{r=1}^{R} GTV_{n,\varepsilon_n}(u_n^r) & \text{if } \mu = \nu_n \text{ and } \mathcal{U}_n \in \mathcal{M}_n(D) \\ +\infty & \text{else.} \end{cases} \tag{5.8}$$

For reasons explained in the last chapter, both functionals are defined in the same space: $TL^1(D)^R$. Of course, only a fraction of these possible functions, with their corresponding measures, are actually of use to us. Specifically, those functions that satisfy the constraints of $\mathcal{M}_n(D)$ and $\mathcal{M}(D)$ respectively, as mentioned in the section on the balancing term already. Furthermore only the measure ν and its empirical realization ν_n are of interest, so we neglect the argument μ.

Theorem 5.1 (Γ-Convergence Graph Total Variation)**.** If the assumptions 5.1 hold, it follows

$$E_n \xrightarrow{\Gamma} \sigma_n E$$

with respect to $TL^1(D)^R$ metric for $n \to \infty$.

The proof of this theorem is split into several parts. First, we state two preliminary results we rely upon. Then, we prove the liminf and limsup properties of Γ-convergence, where we construct the recovery sequence only on a subset. Therefore, we need to prove this subset lies dense in $L^1(\nu)^R$, which we do in the next section. That is necessary, because an arbitrary partition of the domain D into more than two sets can not be approximated by smooth partitions, as multiple junctions appear when more than two sets in the partition meet.

We start with a preliminary Lemma to handle the presence of the balance term.

Lemma 5.2. *(i) If $\{U_n\} \subset L^1(\nu_n)^R$ is sequence with $U_n \xrightarrow{TL^1} U$ for some $U \in L^1(\nu)^R$, then $B_n^\alpha(u_n^r) \to B^\alpha(u^r)$ for all $1 \le r \le R$.*

(ii) Let $Y_n \subset X_n$. If $\tilde{1}_{Y_n} \overset{TL^1}{\to} \tilde{1}_Y$, then it also holds $1_{Y_n} \overset{TL^1}{\to} 1_Y$.

Proof. **(i)** Let $\{U_n\} \subset L^1(\nu_n)^R$a sequence with $U_n \overset{TL^1}{\to} U$ for some $U \in L^1(\nu)^R$. Then, with probability one there exists a stagnating sequence of transport maps T_n between ν_n and ν by Proposition 4.4. As such it holds $\mathcal{U}_n \circ T_n \overset{L^1(\nu)}{\to} \mathcal{U}$ and by Lemma 5.1 $B^\alpha(\mathcal{U}_n \circ T_n) \to B^\alpha(\mathcal{U})$. Furthermore, note the equivalence

$$B^\alpha(\mathcal{U}_n \circ T_n) = \int_D \mathcal{U}_n \circ T_n \mathrm{d}\nu(x) = \int_{D^R} \mathcal{U}_n \mathrm{d}\nu_n(x) = B_n^\alpha(\mathcal{U}_n).$$

By our definition of $TL^1(D)^R$ (4.7) the convergence $\mathcal{U}_n \overset{TL^1(\nu)}{\longrightarrow} \mathcal{U}$ occurs elementwise as $u_n^r \overset{TL^1}{\to} u^r$ for $1 \leq r \leq R$ providing the result.

(ii) Let $Y_n \subset X_n$ and $Y \subset D$, with $\tilde{1}_{Y_n} \overset{TL^1}{\to} \tilde{1}_Y$. We assume $B^\alpha(Y) > 0$, otherwise $\nu(Y) = \nu(Y_n) = 0$ and the claim follows trivially. From (i) and the fact that ν is a probability measure for n sufficiently large it follows directly

$$0 = \lim_{n\to\infty} \int_D |\frac{1_{Y_n} \circ T_n}{B_n^\alpha(Y_n)} - \frac{1_Y}{B^\alpha(Y)}| \,\mathrm{d}\nu(x) \geq \lim_{n\to\infty} \int_D |1_{Y_n} \circ T_n - 1_Y| \,\mathrm{d}\nu(x) \geq 0.$$

□

Before we start with Γ-convergence proof, we state one more result. Our proof heavily relies on the following intermediary non-local convergence result.

Theorem 5.2 (Γ-convergence of non-local TV)**.** *Consider D, ν and ρ as stated in the assumptions* 5.1*. For $u \in L^1(D, \nu)$and $\varepsilon > 0$ define the non-local total variation as*

$$TV_\varepsilon(u) := \frac{1}{\varepsilon^{d+1}} \int_{D\times D} \eta\Big(\frac{x-y}{\varepsilon}\Big)|u(x) - u(y)| d\nu(x)\nu d(y).$$

Then TV_ε Γ-converges with respect to the $L^1(D, \nu)$-metric to $\sigma_\eta TV(\cdot, \nu)$. Moreover, the functionals TV_ε satisfy the compactness property Definition 4.3 with respect to the $L^1(D, \nu)$-metric.

For a proof thereof, we refer the reader to [13, Theorem 4.1]. The definition of Γ-convergence for a family of functionals is given in terms of sequences indexed by real numbers. For the non-local we adopt a notation, where ε is a short-hand for ε_n, an element of an arbitrary sequence of positive real numbers $\varepsilon_n \overset{n\to\infty}{\longrightarrow} 0$. Limits as $\varepsilon \to 0$ simply mean limits as $n \to \infty$ for every such sequence.

Now we are ready to commence the proof of our own result, following the reasoning and structure of [14, Proposition 24].

Proof of Theorem 5.1. **Liminf inequality.** For an arbitrary sequence $\{\mathcal{U}_n\} \subset L^1(\nu_n)^R$ with $\mathcal{U}_n \overset{TL^1}{\to} \mathcal{U}$ for some $\mathcal{U} \in L^1(\nu)^R$, we need to show

$$\liminf_{n\to\infty} E_n(\mathcal{U}_n) \geq \sigma_\eta E(\mathcal{U}).$$

We know from Lemma 5.1 that $\mathcal{M}(D)$ is closed in $L^1(\nu)^R$. This means, that if $\mathcal{U} \notin \mathcal{M}(D)$, it also holds $\mathcal{U}_n \circ T_n \notin \mathcal{M}(D)$, since $\mathcal{U}_n \circ T_n \overset{L^1(\nu)}{\to} \mathcal{U}$. Furthermore, we know $B^\alpha(\mathcal{U}_n \circ T_n) = B_n^\alpha(\mathcal{U}_n)$. These two facts together imply that $\mathcal{U}_n \circ T_n \notin \mathcal{M}(D)$ if and only if $\mathcal{U}_n \notin \mathcal{M}_n(D)$. In any other case $E_n(\mathcal{U}_n)$ and $E(\mathcal{U})$ are both equal to $+\infty$. So we can assume $\mathcal{U} \in \mathcal{M}(D)$, implying that for n large enough it also holds $\mathcal{U}_n \in \mathcal{M}_n(D)$.

We will start with a simple constant kernel η and successively generalize the result to kernels satisfying the assumptions. Furthermore, to reduce notation complexity we use a single $u \in L^1(\nu)$ in the proof of the intermediary results.

Step 1: Assume η has the form

$$\eta(z) = \begin{cases} a & |z| < b \\ 0 & \text{otherwise,} \end{cases}$$

for some $a, b \in \mathbb{R}$. We will need to compare the transported discrete realizations of ν_n with the continuous measure ν. For this simple kernel, the comparison comes down to whether the argument $z = \frac{T_n(x)-T_n(y)}{\varepsilon_n}$ is larger than b. We want to leverage the convergence of the non-local total variation by showing that $GTV_{n,\varepsilon_n}(u_n)$ is larger or equal to $TV_{\tilde{\varepsilon}_n}(u_n)$ for an appropriate choice of $\tilde{\varepsilon}_n$. Such a choice is $\tilde{\varepsilon}_n := \varepsilon_n - \frac{2}{b}\|\mathrm{Id} - T_n\|_\infty$. To recognize this, recall that by assumption on the connectedness of the graph it holds $\|\mathrm{Id} - T_n\|_\infty \ll \varepsilon_n$, i.e. $\tilde{\varepsilon}_n$ is actually a small perturbation. Let us look at points $x, y \in D$ where $|T_n(x) - T_n(y)| > b\varepsilon_n$. Then, it holds

$$\begin{aligned} |T_n(x) - T_n(y)| &= |T_n(x) - \mathrm{Id}(x)| + |\mathrm{Id}(x) - \mathrm{Id}(y)| + |\mathrm{Id}(y) - T_n(y)| \\ &\leq 2\|\mathrm{Id} - T_n\|_\infty + |x - y|. \end{aligned}$$

So, for any such points x, y we can lower bound their distance by

$$|x - y| > b\varepsilon_n - 2\|\mathrm{Id} - T_n\|_\infty = b\tilde{\varepsilon}_n.$$

As the form of kernel implies $\eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big) = 0$ for $|x - y| > b\tilde{\varepsilon}_n$, it follows

$$\eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big) \leq \eta\Big(\frac{T_n(x) - T_n(y)}{\varepsilon_n}\Big). \tag{5.9}$$

Let $\tilde{u}_n := u_n \circ T_n$. From the inequality(5.9) we can deduce

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

For $n \to \infty$ it holds $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$ by construction. Furthermore, note that $u_n \xrightarrow{TL^1} u$ implies $\tilde{u}_n \xrightarrow{L^1(\nu)} u$. So we can apply the Theorem 5.2 to obtain $\liminf_{n\to\infty} TV_{\tilde{\varepsilon}_n}(\tilde{u}_n) \geq \sigma_\eta TV(u,\nu)$. Hence we conclude

$$\begin{aligned}\liminf_{n\to\infty} E_n(\mathcal{U}_n) &= \liminf_{n\to\infty} \sum_{r=1}^{R} GTV(u_n^r) \\ &\geq \liminf_{n\to\infty} \sum_{r=1}^{R} TV(u^r,\nu) = \liminf_{n\to\infty} \sigma_n E(\mathcal{U}).\end{aligned}$$

Step 2: Assume η to be a piecewise constant function with compact support. In this case $\eta = \sum_{k=1}^{l} \eta_k$ for some l and functions η_k as above. Denote by GTV^k_{n,ε_n} the corresponding graph total variation with η_k. Let $u_n \xrightarrow{TL^1} u$, then it holds by Step 1

$$\begin{aligned}\liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) &= \liminf_{n\to\infty} \sum_{k=1}^{l} GTV^k_{n,\varepsilon_n}(u_n) \geq \sum_{k=1}^{l} \liminf_{n\to\infty} GTV^k_{n,\varepsilon_n}(u_n) \\ &\geq \sum_{k=1}^{l} \sigma_{\eta_k} TV(u,\nu) = \sigma_\eta TV(u,\nu).\end{aligned}$$

Step 3: Assume η is compactly supported and satisfies all our assumptions. Then there exists an increasing sequence of piecewise constant functions η_k with $\eta_k \nearrow \eta$ a.e. as $k \to \infty$. Once again, it holds

$$\liminf_{n\to\infty} GTV_{n,\varepsilon_n}(u_n) \leq \liminf_{n\to\infty} GTV^k_{n,\varepsilon_n}(u_n) \leq \sigma_\eta TV(u,\nu),$$

as $u_n \xrightarrow{TL^1} u$ for every $k \in \mathbb{N}$. Note that monotone convergence theorem gives us $\lim_{k\to\infty} \sigma_{\eta_k} = \sigma_\eta$.

Step 4: The proof for general η follows analogously to Step 3. We can approximate the general kernel from below using a sequence of compactly supported kernels.

Limsup inequality. For an arbitrary $\mathcal{U} \in L^1(\nu)^R$, we need to supply a sequence $\{\mathcal{U}_n\} \subset L^1(\nu_n)^R$ with $\mathcal{U}_n \xrightarrow{TL^1} \mathcal{U}$ such that

$$\limsup_{n\to\infty} E_n(\mathcal{U}_n) \leq \sigma_\eta E(\mathcal{U}).$$

The main idea is to work with piecewise smooth boundaries by considering sets B^r that extend to the outside of D, to get $A^r = B^r \cap D$. From there, we will construct a recovery sequence by restriction to the first n points with transport maps. By Proposition 4.1, it suffices to then show the limsup inequality on a dense subset. Since we want to work the whole $\mathbb{R}^d$, it proves useful to consider an extension $\rho(x) = \lambda$ for $x \in \mathbb{R}^d \backslash D$. This extension is a lower semi-continuous function and has the same lower and upper bounds as the original ρ. Furthermore, just like in the liminf proof, it is

enough to consider the case $\mathcal{U} \in L^1(\nu_n)^R$ where $E(\mathcal{U}) < \infty$, as the other case is trivial and any sequence will do. But this also implies

$$TV(1_{A^r}, \nu) \leq B^\alpha(A^r) \cdot TV(u^r) \leq max\{B^\alpha(A^1), \dots B^\alpha(A^R)\} \cdot \sum_{r=1}^{R} TV(u^r) < \infty$$

for $r = 1, \dots, R$, i.e. the collection of sets $\{A^1, \dots, A^R\}$ defining $\mathcal{U}$ with $u^r = \tilde{1}_{A^r}$, have finite perimeter. Moreover, $\mathcal{U} \in \mathcal{M}(D)$ guarantees that $\mathcal{L}(A^r \cap A^s) = 0$ for any two sets A^r, A^s with $r \neq s$, so we may assume without the loss of generality that the sets of $\mathcal{U}$ are mutually disjoint.

We first, in Theorem 5.3, construct a recovery sequence for $\mathcal{U}$ with defining sets $\{A^1, \dots, A^R\}$ of the form $A^r = B_r \cap D$, where B_r has piecewise smooth boundary and satisfies $|D1_{B_r}|_{\rho^2}(\partial D) = 0$. Let $Y_n^r = A^r \cap X_n$denote the restriction of A^r to the first n data points. By Proposition 4.4, there exists a sequence of transport maps $\{T_n\}_{n\in\mathbb{N}}$ such that $1_{A_n^r} := 1_{Y_n^r} \circ T_n \to 1_{A^r}$ for $n \to \infty$. Note that, by the change of variables

$$\int_D 1_{A_n^r}(x)\ \mathrm{d}\nu(x) = \int_D 1_{Y_n^r}\ \mathrm{d}\nu_n(x),$$

we have $|Y_n^r| = |A_n^r| \to |A^r|$ as $n \to \infty$. As argued in Lemma 5.2, such a convergence implies $B_n^\alpha(Y_n^r) \to B^\alpha(A^r)$. In particular, we can assume $|Y_n^r| > 0$ and we can define $u_n^r = \frac{1_{Y_n^r}}{B^\alpha(Y_n^r)}$ as the corresponding normalized indicator function. We claim that $\mathcal{U}_n = (u_n^1, \dots, u_n^R)$ furnishes the desired recovery sequence.

To see that $\mathcal{U}_n \in \mathcal{M}_n(D)$ first note that each $u_n^r \in Ind_n(D)$ by construction, while $\{A^1, \dots, A^R\}$ being disjoint implies $\{Y_n^1, \dots, Y_n^R\}$ are as well. Consequently,

$$E_n(\mathcal{U}_n) = \sum_{r=1}^{R} GTV_{n,\varepsilon_n}(u_n^r)$$

by definition.

Step 1: As in the proof of the liminf inequality, we first assume η is of the form $\eta(|z|) = a$ if $|z| < b$ and zero otherwise. Again, we define a small perturbation $\tilde{\varepsilon}_n := \varepsilon_n + \frac{2}{b}\|\mathrm{Id} - T_n\|_\infty$and use non-local total variation $TV_{\tilde{\varepsilon}_n}$ of $u \in L^1(\nu)$. Analogously to the calculation for (5.9), $|x-y| > b\tilde{\varepsilon}_n = b\varepsilon_n + 2\|\mathrm{Id} - T_n\|_\infty$ implies $|T_n(x) - T_n(y)| > b\varepsilon_n$. Thus it follows

$$\eta\Big(\frac{T_n(x) - T_n(y)}{\varepsilon_n}\Big) \leq \eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big),$$

which in turn implies that

$$\frac{\varepsilon_n^{d+1}}{\tilde{\varepsilon}_n^{d+1}} GTV_{n,\varepsilon_n}(1_{Y_n^r}) \leq TV_{\tilde{\varepsilon}_n}(1_{A_n^r}).$$

Next we find a bound for $TV_{\tilde{\varepsilon}_n}$ by estimating the error of our projection A_n^r. It holds

$$\begin{aligned}
|TV_{\tilde{\varepsilon}_n}(1_{A_n^r}) - TV_{\tilde{\varepsilon}_n}(1_{A^r})| =& |\frac{1}{\tilde{\varepsilon}_n}\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)|1_{A_n^r}(x)-1_{A_n^r}(y)|\,\mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
&- \frac{1}{\tilde{\varepsilon}_n}\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)|1_{A^r}(x)-1_{A^r}(y)|\,\mathrm{d}\nu(x)\mathrm{d}\nu(y)| \\
\leq& \frac{1}{\tilde{\varepsilon}_n}\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)|1_{A_n^r}(x)-1_{A_n^r}(y)-1_{A^r}(x)+1_{A^r}(y)|\,\mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
\leq& \frac{2}{\tilde{\varepsilon}_n}\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)|1_{A_n^r}(x)-1_{A^r}(x)|\,\mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
\leq& \frac{2}{\tilde{\varepsilon}_n}\int_{D\times D}\eta_{\tilde{\varepsilon}_n}(x-y)\nu(x)\mathrm{d}\nu(y)\;\|1_{A_n^r}-1_{A^r}\|_{L^1(\nu)} \\
\leq& \frac{K_0}{\tilde{\varepsilon}_n}\|1_{A_n^r}-1_{A^r}\|_{L^1(\nu)}, \qquad (5.10)
\end{aligned}$$

for some constant K_0 that depends on η and D. From the result on transport maps, Proposition 4.4, we know that, for n large enough, the difference $1_{A_n^r} - 1_{A^r}$ is trivial far enough inside of A^r but might not be on a tubular neighbourhood depending on n via $\|\mathrm{Id} - T_n\|_\infty$. Weyl's volume formula for tubes [19] gives a constant depending on surface and volume of A^r, and therefore B^r. We obtain

$$\|1_{A_n^r} - 1_{A^r}\|_{L^1(\nu)} \leq C_0(B^r)\|\mathrm{Id} - T_n\|_\infty. \qquad (5.11)$$

Since $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$ the previous inequalities (5.10) and (5.11) imply

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(1_{Y_n^r}) \leq \limsup_{n\to\infty} TV_{\tilde{\varepsilon}_n}(1_{A_n^r}) = \limsup_{n\to\infty} TV_{\tilde{\varepsilon}_n}(1_{A^r}).$$

We deduce from Theorem 5.2 that

$$\limsup_{n\to\infty} TV_{\tilde{\varepsilon}_n}(1_{A_n^r}, \nu) \leq \sigma_\eta TV(1_{A^r}, \nu),$$

allowing us to finally conclude $\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(1_{Y_n^r}) \leq \sigma_\eta TV(1_{A^r}, \nu)$. Consequently, it holds

$$\limsup_{n\to\infty} GTV_{n,\varepsilon_n}(u_n^r) = \limsup_{n\to\infty} \frac{GTV_{n,\varepsilon_n}(1_{Y_n^r})}{B^\alpha(Y_n^r)} \leq \sigma_\eta \frac{TV(1_{A^r}, \nu)}{B^\alpha(A^r)} \qquad (5.12)$$

by continuity of the balance term. As such, we are able to determine the convergence $E_n(\mathcal{U}_n) \to E(\mathcal{U})$ while (5.11) implies $\mathcal{U}_n \xrightarrow{TL^1} \mathcal{U}$, i.e. $\mathcal{U}_n$ does indeed furnish the desired recovery sequence.

Step 2: The generalization to a piecewise constant function η with compact support works exactly as it does in the proof of the liminf equality.

Step 3: Analogously to the liminf proof, for η compactly supported, there exists decreasing sequence of piecewise constant functions η_k with $\eta_k \searrow \eta$ a.e. as $k \to \infty$.

Step 4: For general η, let $a > 0$ and define a compactly supported

$$\eta_a(z) := \begin{cases} \eta(z) & \text{for} z \leq a \\ 0 & \text{else.} \end{cases}$$

Since we use transport maps there, we need to bound the error when leaving the support of η_a, such that it goes to zero for $a \to \infty$. We denote with GTV^a_{n,ε_n} the graph total variation using η_a. With this approximation the functionals $GTV^a_{n,\varepsilon_n}$and GTV_{n,ε_n} only differ on the set $\Omega_a = \{(x,y) \in D : |T_n(x) - T_n(y)| > a\varepsilon_n\}$, so we obtain

$$GTV_{n,\varepsilon_n} = GTV^a_{n,\varepsilon_n} + \frac{1}{\varepsilon_n^{d+1}} \int_{\Omega_a} \eta\Big(\frac{T_n(x) - T_n(y)}{\varepsilon_n}\Big) |1_{A_n^r}(x) - 1_{A_n^r}(y)| \, \mathrm{d}\nu(x)\mathrm{d}\nu(y). \tag{5.13}$$

From the earlier calculation, we know $|T_n(x) - T_n(y)| \leq 2\|\mathrm{Id} - T_n\|_\infty + |x - y|$. As $\frac{\|\mathrm{Id}-T_n\|_\infty}{\varepsilon_n} \overset{n\to\infty}{\longrightarrow} 0$ we conclude that, for large enough n, for almost every $(x,y) \in D \times D$, for which $|T_n(x) - T_n(y)| > a\varepsilon_n$, it holds $\frac{1}{2}|T_n(x) - T_n(y)| \leq |x - y| \leq 2|T_n(x) - T_n(y)|$. With the change of variables $x = x$ and $h = \frac{x-y}{2\varepsilon_n}$, we deduce that

$$\begin{aligned} \frac{1}{\varepsilon^{d+1}} \int_{\Omega_a} \eta\Big(\frac{|T_n(x) - T_n(y)|}{\varepsilon}\Big) |1_{A_n^r}(x) - 1_{A_n^r}(y)| \, \mathrm{d}\nu(x)\mathrm{d}\nu(y) \\ \leq \frac{1}{\varepsilon^{d+1}} \int_{|x-y| > \frac{a\varepsilon_n}{2}} \eta\Big(\frac{|x-y|}{2\varepsilon_n}\Big) |1_{A_n^r}(x) - 1_{A_n^r}(y)| \, \mathrm{d}\nu(x)\mathrm{d}\nu(y) \\ \leq \frac{\Lambda^2}{\varepsilon^{d+1}} \int_{|h| > \frac{a}{4}} \eta(|h|) \, \mathrm{d}x\mathrm{d}y, \end{aligned}$$

which is obviously bounded by assumption on the surface tension of η. So, sending a to infinity, we can conclude that for all $r \in \{1, \dots, R\}$

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

By Theorem 5.3 and Proposition 4.1 the limsup inequality follows for any collection of sets with finite perimeter. The continuity of the balance term and Lemma 4.1 then allows us to deduce the result for all $\mathcal{U} \in \mathcal{M}(D)$. □

5.4 Construction of Piecewise Smooth Sets

The following theorem allows us to apply the result of Γ-convergence on dense subsets, Proposition 4.1, to the proof of Theorem 5.1. In it we show that the subset of sets with piecewise smooth boundaries is dense in the set of indicator functions with finite perimeter satisfying the orthogonality constraint. We say a set A is induced by a piecewise smooth set B, if it holds $A = B \cap D$, where $B \subset \mathbb{R}^d$ has piecewise smooth boundary and satisfies $|D1_{B_r}|_{\rho^2}(\partial D) = 0$.

Theorem 5.3 (Density of Piecewise Smooth Sets)**.** *Let* $\mathcal{U} = (\tilde{1}_{A_1}, \dots, \tilde{1}_{A_R}) \in L^1(D,\nu)^R$ *where each of the sets* A_r *has finite perimeter. Then there exists a sequence* $\{\mathcal{U}_m = (\tilde{1}_{A_1^m}, \dots, \tilde{1}_{A_R^m})\}_{m\in\mathbb{N}}$*, where each* $\mathcal{U}_m$ *is induced by piecewise smooth sets, and it holds*

$$1_{A_m^r} \overset{L^1(\nu)}{\to} 1_{A^r},$$

as well as

$$\lim_{n\to\infty} TV(1_{A_m^r}, \nu) = TV(1_{A^r}, \nu)$$

for every $r \in \{1, \dots, R\}$*.*

We provide the construction of the approximating sequence $\{\mathcal{U}_m\}_{m\in\mathbb{N}}$ through the sequence of three lemmas presented below. The entirety of this proof is in essence identical to [14, Proposition 24, Density].

Lemma 5.3. *Let* D *denote an open and bounded set. Let* $\{A_1, \dots, A_R\}$ *denote a collection of open, bounded sets with smooth boundary in* $\mathbb{R}^d$ *that satisfy*

$$\mathcal{H}^{d-1}(\partial A_r \cap \partial A_s)) = 0 \ \forall r \neq s. \tag{5.14}$$

Then there exists a permutation $\pi : \{1, \dots, R\} \to \{1, \dots, R\}$*, such that for all* $r \in \{1, \dots, R\}$

$$TV(1_{A_{\pi(r)} \setminus \bigcup_{s=r+1}^R A_{\pi(s)}}, \nu) \leq TV(1_{A_{\pi(r)}}, \nu).$$

Proof. The proof is done by induction on $R \in \mathbb{N}$, the number of sets. In the base case of $R = 1$ the claim holds true trivially. For the Inductive Step: let $\{A_1, \dots, A_R\}$ be as described in the assumptions of the lemma and suppose the result holds for any $R-1$ sets already. Then, we are only missing case of removing all other sets, so it is enough to show there exists a $r \in \{1, \dots, R\}$ such that

$$TV(1_{A_r \setminus \bigcup_{s\neq r} A_s}, \nu) \leq TV(1_{A_r}, \nu). \tag{5.15}$$

The idea of the proof is to look separately at the different parts of the boundary of this set A_r. To simplify notation we introduce

$$a_{rs} := \partial A_r \cap \Big(A_s \setminus \bigcup_{k\neq r, k\neq s} A_k\Big) \cap D.$$

This describes the part of the boundary A_r that is inside the the set A_s but none of the others. The total variation the perimeter of A_s in D without any other of the other sets is given by

$$TV(1_{A_r \setminus \bigcup_{s\neq r} A_s}, \nu) = \int_{\partial\left(A_r \setminus \bigcup_{s\neq r} A_s\right) \cap D} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1}. \tag{5.16}$$

For single, fixed open sets A_r, A_s it holds

$$\partial\Big(A_r \backslash A_s\Big) = (\partial A_r \cap A_s^c) \cup (\partial A_s \cap A_r) \cup (\partial A_r \cap \partial A_s)\,.$$

The last expression has measure zero by assumption. Thus replacing the right hand side (5.16) becomes

$$TV(1_{A_r \backslash \bigcup_{s=1}^R A_s}, \nu) = \int_{\partial A_r \cap \Big(\bigcup_{r\neq s} A_s\Big)^c \cap D} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1} + \sum_{s\neq r} \int_{a_{sr}} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1}.$$

Furthermore, regarding the whole ∂A_r it holds by the same reasoning

$$TV(1_{A_r}, \nu) = \int_{\partial A_r \cap \Big(\bigcup_{r\neq s} A_s\Big)^c \cap D} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1} + \sum_{s\neq r} \int_{a_{rs}} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1}.$$

Now let us assume (5.15) does not hold, i.e. for every $r \in \{1, \dots, R\}$ it is $TV(1_{A_r \backslash \bigcup_{s=1}^R A_s}, \nu) > TV(1_{A_r}, \nu)$. This would imply

$$\sum_{s\neq r} \int_{a_{sr}} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1} > \sum_{s\neq r} \int_{a_{rs}} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1}.$$

However summing over all r would then suggest

$$\sum_{r=1}^R \sum_{s\neq r} \int_{a_{sr}} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1} > \sum_{r=1}^R \sum_{s\neq r} \int_{a_{rs}} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1} = \sum_{r=1}^R \sum_{s\neq r} \int_{a_{sr}} \rho^2 \, \mathrm{d}\mathcal{H}^{d-1}.$$

This is a clear contradiction, hence (5.15) needs to hold for at least one $r \in \{1, \dots, R\}$. □

The following lemma asserts, that we can approximate the outer sets B with smooth sets. Since we extend our set $A \subset D$ to $B \subset \mathbb{R}^d$ such that $B \cap D = A$, we need to introduce some notation for the total variation. We denote the weighted total variation with respect to the extension $\rho(x) = \lambda$ for $x \in \mathbb{R}^d \backslash D$ by $TV(u, \mathbb{R}^d)$.

Lemma 5.4. *Let D denote an open, bounded domain in $\mathbb{R}^d$ with Lipschitz boundary and let $(B^1, \dots, B^R)$ denote a collection of R bounded and mutually disjoint subsets of $\mathbb{R}^d$ that satisfy*

$$TV(1_{B^r}, \mathbb{R}^d) < \infty \text{ and } |D1_{B^r}|_{\rho^2}(\partial D) = 0.$$

Then there exists a sequence of mutually disjoint sets $\{A_k^1, \dots, A_k^R\}$ with piecewise smooth boundaries which satisfy

$$1_{A_k^r} \overset{L^1(\nu)}{\longrightarrow} 1_{B^r} \text{ and } \lim_{k\to\infty} TV(1_{A_k^r}, \nu) = TV(1_{B^r}, \nu)$$

for all $1 \leq r \leq R$.

Proof. Recall we used the extension $\rho(x) = \lambda$ for $x \in \mathbb{R}^d \backslash D$ in the proof of Theorem 5.1. By Proposition 3.3 there exists a sequence of functions $u_k^r \in C^\infty(\mathbb{R}^d, [0,1])$ which satisfy

$$u_k^r \overset{L^1(\nu)}{\to} 1_{B^r} \text{ and } TV(u_k^r, \nu) \to TV(1_{B^r}, \nu)$$

for $k \to \infty$ and $r = 1, \dots, R$.

Denote the superlevel sets of $u_k^r$with $B_k^r(t) := \{x : u_k^r(x) > t\}$ for $k \in \mathbb{N}$, $t \in (0,1)$, $r = 1, \dots, R$. Sard's Lemma [15, Corollary 14.50] states almost all level sets of smooth functions are smooth manifolds, hence $\partial B_k^r(t)$ is smooth for all k. Moreover, it holds up to a subsequence, $1_{B_k^r(t)} \overset{L^1(\nu)}{\to} 1_{B^r}$ as $k \to \infty$ for a.e. $t \in (0,1)$. The general coarea formula [16, Proposition 4.2] tells us $TV(u_k^r, \nu) \geq TV(1_{B_k^r(t)}, \nu) \, \forall$ and therefore

$$\lim_{k\to\infty} TV(1_{B_k^r(t)}, \nu) \leq \lim_{k\to\infty} TV(u_k^r, \nu) = TV(1_{B^r}, \nu).$$

On the other hand by lower semi-continuity of the weighted total variation Proposition 3.2 it holds

$$\lim_{k\to\infty} TV(1_{B_k^r(t)}, \nu) \geq \liminf_{k\to\infty} TV(1_{B_k^r(t)}, \nu) \geq TV(1_{B^r}, \nu).$$

So follows for almost every $t \in (0,1)$ the equality

$$\lim_{k\to\infty} TV(1_{B_k^r(t)}, \nu) = TV(1_{B^r}, \nu).$$

So far we have a sequence of smooth sets with converging total variation, but these are not necessarily disjoint. For this reason we would like to apply Lemma 5.3. $\partial B_k^r(t)$ has $\mathcal{H}^d$-measure zero in $\mathbb{R}^d$, since it is a $(d-1)$-manifold for any $t \in \mathbb{R}$. Furthermore, since $\partial B_k^r(t)$ is smooth, we can apply [2, Lemma 2.95], and it follows that for any $r \neq s$ we can find $t_s, t_r \in \mathbb{R}$ such that

$$\mathcal{H}^{d-1}\left(\partial B_k^r(t_r) \cap \partial B_k^s(t_s)\right) = 0.$$

These sets satisfy the assumptions of Lemma 5.3, we denote them by $B_k^r := B_k^r(t_r)$. Hence said lemma provides disjoint sets $(A_k^1, \dots A_k^R)$ by

$$A^r := B_k^r \backslash \bigcup_{s=\pi_k^{-1}(r)+1}^{R} B_k^{\pi_k(s)},$$

where the permutation π_k guarantees

$$TV(1_{A_k^r}, \nu) \leq TV(1_{B_k^r}, \nu)$$

for every $r = 1, \dots, R$. Each A_k^r has a piecewise smooth boundary due to the fact that B_k^r have smooth boundaries. Moreover, the disjointness of $(B^1, \dots, B^R)$ combined

with the $L^1(\nu)$-convergence of $1_{B_k^r} \to 1_{B^r}$ shows $1_{A_k^r} \overset{L^1(\nu)}{\longrightarrow} 1_{B_k^r}$. At the same time the lower semi-continuity of the weighted total variation implies

$$\lim_{k\to\infty} TV(1_{A_k^r}, \nu) \geq \liminf_{k\to\infty} TV(1_{A_k^r}, \nu) \geq TV(1_{B^r}, \nu)$$

allowing us to conclude the result. □

To complete the construction, we need to verify the existence of sets $(B^1, \dots, B^R)$. These have satisfy the hypotheses of the previous lemma. This is the content of our final lemma for the proof of Theorem 5.3.

Lemma 5.5. *Let D be an open domain with Lipschitz boundary and let $(A^1, \dots, A^R)$ denote a collection of disjoint sets that satisfy*

$$A^r \subset D \quad \textit{and } TV(1_{A^r}, \nu) < \infty.$$

Then, there exists a disjoint collection of bounded sets $(B^1, \dots, B^R)$ that satisfy

$$B^r \cap D = A^r,\ TV(1_{B^r}, \mathbb{R}^d) < \infty,\ |D1_{B^r}|_{\rho^2}(\partial D) = 0.$$

Proof. Since D is a bounded set with Lipschitz boundary, it is a so called extension domain (see [2, Definition 3.20]) according to [2, Proposition 3.21]. For 1_A there then exists $v \in BV(\mathbb{R}^d)$, an extension with compact support such that $|Dv|(\partial D) = 0$ and $0 \leq v \leq 1$ by [2, Remark 3.43]. Since the density ρ is bounded, it follows $|Dv|_{\rho^2}(\partial D) = 0$, as well as $TV(v, \mathbb{R}^d) < \infty$. The coarea formula allows us to choose $t \in (0, 1)$ such that the superlevel set $B = \{v > t\}$ still conforms to $|D1_B|_{\rho^2}(\partial D) = 0$. Furthermore, due to the original disjointness of the A^r, choosing t close to 1 ensures that the superlevel sets do not overlap and sets B^r remain disjoint. □

Proof of Theorem 5.3. For any collection of disjoint sets with finite boundary $(A^1, \dots, A^R)$ where $A^r \subset D$ for $r \in \{1, \dots, R\}$, by Lemma 5.5, there exists a disjoint collection of bounded sets $(B^1, \dots, B^R)$, $B^r \subset \mathbb{R}^d$, that satisfy

$$B^r \cap D = A^r,\ TV(1_{B^r}, \mathbb{R}^d) < \infty,\ |D1_{B^r}|_{\rho^2}(\partial D) = 0.$$

As such, these sets $(B^1, \dots, B^R)$ satisfy the assumptions of Lemma 5.4. Hence exists for any of the B^r a sequence of mutually disjoint sets A_m^r with

$$1_{A_m^r} \overset{L^1(\nu)}{\to} 1_{A^r} \text{ and } \lim_{n\to\infty} TV(1_{A_m^r}, \nu) = TV(1_{A^r}, \nu)$$

for every $r \in \{1, \dots, R\}$. Moreover, by Lemma 5.2 (i) A_m^r immediately induce the sequence $\mathcal{U}_m$. □

5.5 Compactness

In this section, we aim to show that E_n satisfies the compactness property Definition 4.3. For the proof we rely on an existing result for non-local total variation once again.

Proposition 5.1 (Compactness Non-Local TV)**.** *Let D be a bounded, open, and connected set with Lipschitz boundary in $\mathbb{R}^d$. Suppose the sequence $\{u_\varepsilon\}_{\varepsilon<0} \subset L^1(D,\nu)$ satisfies*

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

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

The proposition is taken straight from [13, Proposition 4.6] where a proof can be found as well.

Theorem 5.4 (Compactness GTV)**.** *It holds, with probability one, that any sequence $\{\mathcal{U}_n\}$ with $\mathcal{U}_n \in L^1(\nu_n)^R$ with*

$$\limsup_{n\to\infty} E_n(\mathcal{U}_n) < \infty$$

is precompact in $TL^1(D)^R$.

Proof. Let $\{\mathcal{U}_n\}_{n\in\mathbb{N}} \in L^1(\nu_n)^R$ be a sequence with $\mathcal{U}_n = (u_n^1,\dots,u_n^R)$ such that $\limsup_{n\to\infty} E_n(\mathcal{U}_n) < \infty$. For sufficiently large n it needs to hold $\mathcal{U}_n \in \mathcal{M}_n(D)$ and therefore $u_n^r \in Ind_n(D)$ for $r \in \{1,\dots,R\}$. We want to leverage Proposition 5.1 so we need a sequence that satisfies its assumptions. Recall, u_n^r is of the form $\frac{1_{A_n^r}}{B_n^\alpha(A_n^r)}$ and we defined $|A_n| = \nu_n(A_n)$. Furthermore, $|A| \le |D| = 1$ for any subset $A \subset D$. In favour of readability let us drop the index r for now.

Similarly to the proof of the Γ-convergence, we can assume η is of the form $\eta(|z|) = a$ if $|z| < b$ and $\eta(|z|) = 0$ otherwise, without loss of generality. Furthermore, as in the reasoning for (5.9) we define a small perturbation $\tilde{\varepsilon}_n := \varepsilon_n - \frac{2}{b}\|\mathrm{Id} - T_n\|_\infty$ and deduce

$$\eta\Big(\frac{|x-y|}{\tilde{\varepsilon}_n}\Big) \le \eta\Big(\frac{|T_n(x)-T_n(y)|}{\varepsilon_n}\Big).$$

Let $\tilde{u}_n := u_n \circ T_n$. As before it follows

$$GTV_{n,\varepsilon_n}(u_n) \ge \frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \eta\Big(\frac{|x-y|}{\tilde{\varepsilon}_n}\Big)|\tilde{u}_n(x) - \tilde{u}_n(y)|\,\mathrm{d}\nu(x)\mathrm{d}\nu(y).$$

Since $\frac{\varepsilon_n^{d+1}}{\tilde{\varepsilon}_n^{d+1}} \to 1$ for $n\to\infty$, we conclude $TV_{\tilde{\varepsilon}_n}(\tilde{u_n},\nu) < \infty$.

Define the sequence $v_n := (u_n)^{\frac{1}{\alpha}}$, for it holds by definition of the balance term

$$GTV_{n,\varepsilon_n}(v_n) = \frac{GTV_{n,\varepsilon_n}\left((1_{A_n})^{\frac{1}{\alpha}}\right)}{\left(|A_n|^\alpha\right)^{\frac{1}{\alpha}}} = \frac{GTV_{n,\varepsilon_n}(1_{A_n})}{|A_n|}.$$

The $L^1(\nu)$ norm of v_n is given by

$$\|v_n\|_{L^1(\nu)} = \int_D \frac{(1_{A_n})^{\frac{1}{\alpha}}}{\left(B^\alpha(A_n)\right)^{\frac{1}{\alpha}}}\,\mathrm{d}\nu(x) = \int_D \frac{1_{A_n}}{|A_n|}\,\mathrm{d}\nu(x) = \frac{|A_n|}{|A_n|} = 1.$$

This also implies $B_n^\alpha(v_n) = 1$. If $\alpha \geq 1$ we use the fact $|A_n|^\alpha \leq |A_n|$ to deduce

$$GTV_{n,\varepsilon_n}(v_n) = \frac{GTV_{n,\varepsilon_n}(1_{A_n})}{|A_n|} \leq \frac{GTV_{n,\varepsilon_n}(1_{A_n})}{|A_n|^\alpha} = GTV_{n,\varepsilon_n}(u_n) < \infty.$$

On the other hand, for $0 < \alpha \leq 1$, as long as $|A_n| > 0$, we observe the relation

$$\begin{aligned} GTV_{n,\varepsilon_n}(v_n) &= \frac{GTV_{n,\varepsilon_n}(1_{A_n})}{|A_n|} \\ &= \frac{|A_n|^\alpha}{|A_n|}\frac{GTV_{n,\varepsilon_n}(1_{A_n})}{|A_n|^\alpha} \\ &= |A_n|^{\alpha-1} GTV_{n,\varepsilon_n}(u_n). \end{aligned}$$

Of course, since $u_n^r \in Ind_n(D)$ for $r \in \{1,\dots,R\}$ we can rely on $|A_n| \neq 0$. Therefore, whenever $GTV_{n,\varepsilon_n}(u_n) < \infty$ it holds $GTV_{n,\varepsilon_n}(v_n) < \infty$ as well as $TV_{\tilde{\varepsilon}_n}(T_n \circ v_n, \nu) < \infty$, for $\tilde{\varepsilon}_n$ defined as above.

By Proposition 5.1, $\{v_n\}_{n\in\mathbb{N}}$ is a precompact sequence in $TL^1(\nu)$ and there exists a converging subsequence $v_{n_k} \xrightarrow{TL^1} v$ for some $v \in L^1(\nu)$. Lemma 5.2 gives us $B^\alpha(v) = 1$. Moreover, the Γ-convergence of E_n implies $\sigma_\eta E(v) \leq \liminf_{k\to\infty} E_{n_k}(v_{n_k}) < \infty$, i.e. $v = \frac{1_A}{B^\alpha(A)}$ for some nonempty set $A \subset D$. The $L^1(\nu)$ convergence implies pointwise convergence in yet another subsequence $v_{n_{k_l}}$. For sufficiently large l the volume $|A_{n_{k_l}}|$ remains bounded away from zero, and therefore $\|u_{n_{k_l}}\|_{L^1(\nu)} < \infty$.

As any component u_n^r of $\mathcal{U}_n$ has a converging subsequence, the definition of $TL^1(D)^R$ allows us to conclude the precompactness of $\mathcal{U}_n \in L^1(\nu_n)^R$. □

5.6 Consistency of α-Cheeger-Cuts

Proposition 5.2 (Existence of solution)**.** *Under the assumptions* 5.1 *stated above, there exists a measurable set* $A^* \subset D$ *with* $0 < \nu(A^*) < 1$ *such that* $1_{A^*}^{\sim}$ *minimizes* $E(\nu, A)$ *in (5.7).*

Proof. Since the functional $E(\nu, A)$ is bounded from below it suffices to show that it is lower semi-continuous with respect to the $L^1(\nu)$ norm and that a minimizing sequence is precompact therein.

To show lower semi-continuity consider a sequence $u_n \in Ind(D)$ such that $u_n \overset{L^1(\nu)}{\longrightarrow} u$ for some $u \in L^1(\nu)$. From Lemma 5.1 it follows that $u \in Ind(D)$. Hence $u = \tilde{1}_A$ for some $A \subset D$ and $B^\alpha(1_A) > 0$, yielding $1_{A_n} \underset{n\to\infty}{\longrightarrow} 1_A$. The lower semi-continuity of the total variation Proposition 3.2 together with the continuity of B^α then implies $E(\nu, A)$ to be lower semi-continuous as well.

The pre-compactness of any minimizing sequence follows directly from [3, Theorem 5.1], which completes the proof. □

The theorems we have proven so far in this chapter, allow us to show the consistency of graph α-Cheeger Cuts with respect to the weighted total variation. This consistency is built on top of the three results of Γ-convergence, compactness and the ensuing convergence of minimizers.

Theorem 5.5 (Consistency of Cuts)**.** *Let the above assumptions* 5.1 *hold and let ε_n be a positive sequence of numbers satisfying the connectedness from Proposition 2.1. Assume $\{x_i\}_{i\in\mathbb{N}}$ is an i.i.d. sequence of points in D drawn from density ρ with $X_n = \{x_1, \dots, x_n\}$ the first n points. Denote by $G_n = (X_n, W_n)$ the graph whose weights are given by $w_{ij} := \eta\left(\frac{|x_i - x_j|}{\varepsilon_n}\right)$, $1 \le i, j \le n$.*

Let $Y_n = (Y_n^1, \dots, Y_n^R)$ denote an optimal graph balance cut and $\tilde{1}_{Y_n}$ the corresponding minimizer of E_n. If $\tilde{1}_A = (\tilde{1}_{A^1}, \dots, \tilde{1}_{A^R})$ is a minimizer of E in (5.7) , i.e. the optimal balance cut of D, then with probability one there exists a subsequence $\tilde{1}_{Y_{n_k}} = (\tilde{1}_{Y^1_{n_k}}, \dots, \tilde{1}_{Y^R_{n_k}})$ with

$$(\tilde{1}_{Y^1_{n_k}}, \dots, \tilde{1}_{Y^R_{n_k}}) \overset{TL^1}{\longrightarrow} (\tilde{1}_{A^1}, \dots, \tilde{1}_{A^R})$$

for $k \to \infty$. Furthermore, it holds

$$\lim_{n\to\infty} E_n(\nu_n, \tilde{1}_{Y_n}) = \sigma_\eta E(\nu, \tilde{1}_A). \tag{5.17}$$

Proof. Let $\tilde{1}_{Y_n}$ be a minimizer of E_n and $\tilde{1}_A$ a minimizer of E. In Theorem 5.1 we showed that the approximating functionals E_n do Γ-converge to $\sigma_\eta E$ in the TL^1-sense. With Theorem 5.4 we verified that $\tilde{1}_{Y_n}$ satisfies the required compactness to apply Proposition 4.2. Thus, $\tilde{1}_{Y_n}$ must converge toward some normalized indicator function $\tilde{1}_A$ up to relabeling. The convergence in sense of Definition 4.4 is a direct consequence. The convergence of (5.17) follows from (4.2) in Proposition 4.2. □

In general, minimizers of (5.7) are not unique. However, if the solution to the continuum Cheeger cut is unique, it is the only accumulation point of $(Y_n^1, \dots, Y_n^R)$ so the result below follows directly.

Corollary 5.1. *If* $(\tilde{1}_{A^1}, \ldots, \tilde{1}_{A^R})$ *is the unique solution of (5.7), with probability one it converges* $(Y_n^1, \ldots, Y_n^R) \xrightarrow{TL^1} (\tilde{1}_{A^1}, \ldots, \tilde{1}_{A^R})$ *for* $n \to \infty$.

6 Graph Ginzburg-Landau

6.1 Graph Ginzburg-Landau Functional

The total variation in the graph, as well as the continuum setting, requires us to work with indicator functions if we want to find sets of optimal perimeter. This makes numerical applications challenging. Ideally, we would use differentiable functions that smoothly approximate these sets. One of the most influential approaches in this context was first developed by Cahn and Hillard [8]. Their original aim was to describe the free energy of non-homogeneous mixtures of incompressible fluids. The model introduces an energy functional that depends on the local concentration at any given point, as well as a first-order term that models the change in concentration in the immediate neighbourhood. This is achieved by performing a Taylor expansion and making assumptions about isotropy, while also neglecting higher-order terms. The idea is to obtain smooth characteristic functions. To this end, the evaluated point of the Taylor expansion is replaced by a function that takes its minima in $\{0, 1\}$ to encourage separation. The minimiser of the resulting functional captures the gradual transition between phases rather than imposing a rigid separation. It mostly takes the values zero and one, but a thin layer exists where the transition occurs. This type of function is known as a phase field, while the function itself, or variations thereof, is known by several names, such as Ginzburg–Landau, Cahn–Hilliard or Allen–Cahn.

Traditionally, for $u : D \to \mathbb{R}$ the Ginzburg-Landau functional is given by

$$GL_\varepsilon(u) = \int_D \varepsilon \mid \nabla u \mid^2 + \frac{1}{\varepsilon} u^2(1-u)^2 \mathrm{d}\nu. \tag{6.1}$$

We will also write $W(u) = u^2(1-u)^2$, the so called double well potential, named after its shape. The double well is of course minimal for $u \in \{0, 1\}$ while the gradient of u change in its value. TThe functional requires additional constraints to avoid trivial solutions. For example, for a volume constraint, we obtain a set A with a boundary that is approximately ε thick as a minimizer. There are other ways to approximate the first order term using appropriate kernels, e.g. [1].

IIn this chapter, we work with the following setup. Note the difference in the final assumption. This adjustment is necessary in order to recover the correctly scaled limit, given that we now have squared differences in the functional.

Assumptions 6.1. Let $\rho : D \to (0, \infty)$ be a continuous density to the probability measure ν supported on D. We assume ρ is bounded on D by constants $\Lambda \geq \lambda > 0$

with $\lambda \leq \rho \leq \Lambda$. The set $X_n = \{x_1, \ldots, x_n\}$ are sampled points from density ρ. We consider an isotropic similarity kernel η with radial profile $\boldsymbol{\eta} : [0, \infty) \to [0, \infty)$ and define it as $\eta(x) := \boldsymbol{\eta}(|x|)$. We assume

(i) $\boldsymbol{\eta}(0) > 0$ and $\boldsymbol{\eta}$ continuous at 0.

(ii) $\boldsymbol{\eta}$ is non-increasing.

(iii) $\sigma_\eta^2 := \int_{\mathbb{R}^d} \eta(h)|h_1|^2 \mathrm{d}h < \infty$.

In general we consider $u \in L^1(D)$. We define the Graph Ginzburg-Landau functional $G_n : L^1(X_n) \to [0, \infty)$ as

$$G_n(u) := \frac{1}{\varepsilon_n^{d+1} n^2} \sum_{i,j=1}^{n} w_{ij}|u(x_i) - u(x_j)|^2 + \frac{1}{\varepsilon_n n^2} \sum_{i=1}^{n} W(u(x_i)) \deg(x_i),$$

where $w_{ij} := \eta(\frac{x_i - x_j}{\varepsilon_n})$ and $W(u(x_i)) = u(x_i)^2(1 - u(x_i))^2$. We can rewrite the graph functional, just as we did with the graph total variation, using empirical measure $\nu_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}(A)$ and (6.3) which yields

$$G_n(u) := \frac{1}{\varepsilon_n} \int_{D\times D} \eta_\varepsilon(x-y)|u(x)-u(y)|^2 \mathrm{d}\nu_n(x)\mathrm{d}\nu_n(y) + \frac{1}{\varepsilon_n} \int_{D\times D} \eta_\varepsilon(|x-y|)W(u(x))\ \mathrm{d}\nu_n(x)\ \mathrm{d}\nu_n(y)$$

The kernel function η_ε plays the role of a mollifier and is used to to approximate the derivative in the continuum case. The scaling with respect to ε_n differs from the total variation and can be justified as follows. If η were to be supported in a ball, the difference in u would scale as

$$|u(x_i) - u(x_j)|^p \sim \varepsilon^p |\nabla u|^p. \tag{6.2}$$

So dividing by ε_n give us the form used in the gradient theory of phase transitions [17].

Our aim is to use the Ginzburg-Landau type functional on graphs to approximate continuum Cheeger cuts. For this reason, the balancing term is defined as before via the integral over u. In the case of multiple phases we define

$$B_n^\alpha(u_n) = \left(\int_D u_n \,\mathrm{d}\nu_n(x)\right)^\alpha = \left(\frac{1}{n}\sum_{i=1}^{n} u_n(x_i)\right)^\alpha,$$

while we use

$$B_n^\alpha(u_n) = \min\left(\left(\int_D u_n \,\mathrm{d}\nu_n(x)\right)^\alpha, 1 - \left(\int_D u_n \,\mathrm{d}\nu_n(x)\right)^\alpha\right)$$

when solving for a single set. If u is a characteristic function of set Y , integrating u is equal to $|Y| = \frac{\#Y}{n}$, the ratio vertices from X_n that are in Y . Together with the graph Ginzburg-Landau functional this lets us define our model $\mathcal{G}_n : TL^1(D) \to [0, \infty]$

$$\mathcal{G}_n(\mu, u) := \begin{cases} \frac{G_n(u)}{B_n^\alpha(u)}. & \text{if } \mu = \nu \text{ and } B^\alpha(u) > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Once again, the consistency of the model is studied using Γ-convergence. For proving said convergence we need a suitable intermediary, non-local functional defined as

$$F_\varepsilon(u) := \frac{1}{\varepsilon}\int_{D\times D} \eta_\varepsilon(x-y)\,|u(x)-u(y)|^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y) + \frac{1}{\varepsilon}\int_D W(u(x))\rho(x)\,\mathrm{d}\nu(x).$$

This non-local model can also be seen as the expected values of the graph functional. Note the double well potential's additional $\rho(x)$ which is needed to recover the total variation weighted with $\rho(x)^2$. The consistency to the graph functional can be reasoned as follows. Recall $w_{xy} = \eta_\varepsilon(|x-y|)$ and $\deg(x) = \sum_{y\in V} w_{xy}$, so it follows

$$\begin{aligned}\frac{1}{n^2}\sum_{x\in V} W(u(x))\deg(x) &= \frac{1}{n^2}\sum_{x,y\in V} W(u(x))w_{xy}\\ &= \frac{1}{n^2}\sum_{x,y\in V} W(u(x))\eta_\varepsilon(|x-y|)\\ &= \int_{D\times D} \eta_\varepsilon(|x-y|)W(u(x))\,\mathrm{d}\nu_n(x)\,\mathrm{d}\nu_n(y). \end{aligned} \tag{6.3}$$

Furthermore, for $n\to\infty$ it holds

$$\begin{aligned}\int_{D\times D} \eta_\varepsilon(|x-y|)W(u(x))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y) &= \int_D W(u(x))\int_D \eta_\varepsilon(|x-y|)\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(x)\\ &= \int_D W(u(x))\int_D \eta_\varepsilon(|x-y|)\rho(y)\,\mathrm{d}y\,\mathrm{d}\nu(x).\end{aligned}$$

The kernel η_ε localizes the integral to a neighborhood of x, and the continuity ensures that the local average converges to the pointwise value $\rho(x)$. This means that $\lim_{\varepsilon\to 0}\int_\Omega \eta_\varepsilon(|x-y|)\rho(x)\,\mathrm{d}y = \rho(x)$. Therefore, we can conclude

$$\int_{D\times D} \eta_\varepsilon(|x-y|)W(u(x))\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y) \underset{\varepsilon\to 0}{\longrightarrow} \int_D W(u(x))\rho(x)\,\mathrm{d}\nu(x). \tag{6.4}$$

Let us define the continuum balancing term for a single phase u as

$$B^\alpha(u) = \min\left(\left(\int u\,\mathrm{d}\nu(x)\right)^\alpha, 1-\left(\int u\,\mathrm{d}\nu(x)\right)^\alpha\right)$$

and similarly for multiple phases

$$B^\alpha(u) = \left(\int u\,\mathrm{d}\nu(x)\right)^\alpha.$$

So the continuum local model is given through $\mathcal{G}: TL^1(D)\to[0,\infty]$

$$\mathcal{G}(\mu,u) = \begin{cases}\frac{TV(u,\nu)}{B^\alpha(u)} & \text{if } u\in BV_\nu(D,\{0,1\}), \mu=\nu \text{ and } B^\alpha(u)>0\\ +\infty & \text{otherwise.}\end{cases}$$

We use the weighted total variation once more to measure the perimeter of a set where the volume is not zero.

In the following two sections we will work towards a result of the form

$$\mathcal{G}_n \xrightarrow{\Gamma} \sigma_\eta^2 c_W \mathcal{G}$$

where

$$\sigma_\eta^2 = \int_{\mathbb{R}^d} \eta(h)|h_1|^2 \mathrm{d}h,$$
$$c_W = \int_0^1 \sqrt{W(s)}\ \mathrm{d}s.$$

Furthermore, we show that $\mathcal{G}_n$ satisfies the compactness property and we generalize the the convergence result to multiple phasefield functions.

6.2 Non-local to Local Convergence

In this section we prove the Γ-convergence of the nonlocal functionals F_ε to the weighted total variation with weight ρ^2. The methodology follows [13, Theorem 4.1], while the technique for bounding the liminf and limsup inequalities is based on [17].

Theorem 6.1 (Non-local to Local TV Convergence)**.** *Under the assumptions 6.1 stated above, it holds*

$$F_\varepsilon(u_\varepsilon) \xrightarrow{\Gamma} \begin{cases} \sigma_\eta^2 c_W TV(u,\nu) & \text{if } u \in BV_\nu(D,\{0,1\}) \\ +\infty & \text{otherwise,} \end{cases}$$

with respect to the $L^1(\nu)$-metric.

For the convergence proof of the non-local total variation smooth, mollified functions, so the total variation can be explicitly stated using its gradient. This makes handling the approximation via the kernel η much easier.

Proof of Theorem 6.1. **Liminf inequality.** Let $u \in L^1(\nu)$ and $u_\varepsilon \xrightarrow{L^1(\nu)} u$ for $\varepsilon \to 0$. We need to show that

$$\liminf_{\varepsilon\to 0} F_\varepsilon(u_\varepsilon) \geq \sigma_\eta^2 c_W TV(u,\nu).$$

Without the loss of generality, we may assume $u \in BV_\nu(D,\{0,1\})$ and $F_\varepsilon(u_\varepsilon)$ is bounded. Otherwise, the double well potential is not zero and the corresponding term blows up for $\varepsilon \to 0$. Therefore, the convergence $u_\varepsilon \xrightarrow{L^1(\nu)} u$ implies $\|u_\varepsilon\|_{L^1(\nu)} < \infty$ for any $\varepsilon > 0$. Also recall that, by the assumptions on ρ, the convergence $u_\varepsilon \xrightarrow{L^1(D,\nu)} u$ is equivalent to $u_\varepsilon \xrightarrow{L^1(D)} u$.

First **consider ρ Lipschitz**. The idea is smoothing the functions u_ε with mollifiers to apply some algebraic tricks, as smoothing with mollifiers does not increase the energy in the limit, while it gains the necessary regularity. To achieve a C^2 equivalent of u_ε consider $J : \mathbb{R}^d \to [0, \infty)$, a smooth radially symmetric function, supported in the closed unit ball $\overline{B(0,1)}$, with $\int_{\mathbb{R}^d} J(z)\,\mathrm{d}z = 1$. We set $J_\delta(z) := \frac{1}{\delta^d} J(\frac{z}{\delta})$.

Fix D', an open domain compactly contained in D. Then, there exists $\delta' > 0$ such that the cover $D'' = \bigcup_{x \in D'} B(x, \delta')$ is contained in D. For any $0 < \delta < \delta'$ and given $v \in L^1(\nu)$, we define the mollified function $v_\delta \in L^1(\mathbb{R}^d, \nu)$ as

$$v_\delta = \int_{\mathbb{R}^d} J_\delta(x - z) v(z)\,\mathrm{d}z = \int_{\mathbb{R}^d} J_\delta(z) v(x - z)\,\mathrm{d}z,$$

where we have extended v to be zero outside of D. The functions v_δ are smooth and satisfy $v_\delta \to v$ (see [11, Theorem 4.1]). Furthermore, by the assumptions on J combined with properties of convolution, we obtain the relation

$$\nabla v_\delta(x) = \int_{\mathbb{R}^d} \nabla J_\delta(x - z) v(z)\,\mathrm{d}z = \int_{\mathbb{R}^d} \nabla J_\delta(z) v(x - z)\,\mathrm{d}z.$$

Young's convolution inequality implies therefore

$$|\nabla v_\delta| \leq \int_{\mathbb{R}^d} |\nabla J_\delta(z)||v(x - z)|\,\mathrm{d}z.$$

From this expression follows that there exists a constant $C > 0$, only depending on the mollifier J, such that

$$\|\nabla v_\delta\|_{L^\infty(\mathbb{R}^d)} \leq \frac{C}{\delta} \int_{\mathbb{R}^d} |v(x - z)|\,\mathrm{d}z = \frac{C}{\delta} \|v\|_{L^1(D)}. \tag{6.5}$$

Analogously, taking the second derivative reveals the bound

$$\|D^2 v_\delta\|_{L^\infty(\mathbb{R}^d)} \leq \frac{C}{\delta^2} \|v\|_{L^1(D)}. \tag{6.6}$$

We can now set $u_{\varepsilon,\delta} := (u_\varepsilon)_\delta$. Comparing the gradients ∇u_δ and $\nabla u_{\varepsilon,\delta}$ of the mollified functions, we deduce

$$\int_{D'} |\nabla u_{\varepsilon,\delta}(x) - \nabla u_\delta(x)|\,\mathrm{d}x \leq \frac{C}{\delta} \int_D |u_\varepsilon(x) - u(x)|\,\mathrm{d}x \overset{\varepsilon \to 0}{\longrightarrow} 0.$$

In the following we will replace u_ε by $u_{\varepsilon,\delta}$ in an appropriate way and argue that the error term to the total variation vanishes. From there, we use the technique from [17] to lower bound the convergence in ε. Afterwards, we still need to argue the convergence to the final result as $\delta \to 0$.

We define the following term to approximate the error in the energy caused by the mollified functions

$$a_{\varepsilon,\delta} = \frac{1}{\varepsilon} \int_{D'' \times D''} \int_{\mathbb{R}^d} J_\delta(z) \eta_\varepsilon(x - y) |u_\varepsilon(x) - u_\varepsilon(y)|^2 (\rho(x)\rho(y) - \rho(x+z)\rho(y+z))\,\mathrm{d}z\mathrm{d}x\mathrm{d}y.$$

Now we can proceed to estimate, recalling $D' \subset D'' \subset D$,

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_{D\times D} \eta_\varepsilon(x-y)|u_\varepsilon(x)-u_\varepsilon(y)|^2\rho(x)\rho(y)\,\mathrm{d}x\mathrm{d}y \\
&\qquad \geq \frac{1}{\varepsilon}\int_{D''\times D''} \eta_\varepsilon(x-y)\,|u_\varepsilon(x)-u_\varepsilon(y)|^2\rho(x)\rho(y)\,\mathrm{d}x\mathrm{d}y \\
&\qquad = \frac{1}{\varepsilon}\int_{D''\times D''}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(x-y)\,|u_\varepsilon(x)-u_\varepsilon(y)|^2\rho(x)\rho(y)\,\mathrm{d}z\mathrm{d}x\mathrm{d}y \\
&\qquad = \frac{1}{\varepsilon}\int_{D''\times D''}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(x-y)\,|u_\varepsilon(x)-u_\varepsilon(y)|^2\rho(x+z)\rho(y+z)\mathrm{d}z\mathrm{d}x\mathrm{d}y \\
&\qquad\quad + a_{\varepsilon,\delta} \\
&\qquad \geq \frac{1}{\varepsilon}\int_{D'\times D'}\int_{\mathbb{R}^d} J_\delta(z)\eta_\varepsilon(\hat{x}-\hat{y})\,|u_\varepsilon(\hat{x}-z)-u_\varepsilon(\hat{y}-z)|^2\rho(\hat{x})\rho(\hat{y})\mathrm{d}z\mathrm{d}\hat{x}\mathrm{d}\hat{y} \\
&\qquad\quad + a_{\varepsilon,\delta} \\
&\qquad \geq \frac{1}{\varepsilon}\int_{D'\times D'} \eta_\varepsilon(\hat{x}-\hat{y})\,|\int_{\mathbb{R}^d} J_\delta(z)\big(u_\varepsilon(\hat{x}-z)-u_\varepsilon(\hat{y}-z)\big)|^2\mathrm{d}z\rho(\hat{x})\rho(\hat{y})\mathrm{d}\hat{x}\mathrm{d}\hat{y} \\
&\qquad\quad + a_{\varepsilon,\delta} \\
&\qquad = \frac{1}{\varepsilon}\int_{D'\times D'} \eta_\varepsilon(\hat{x}-\hat{y})\,|u_{\varepsilon,\delta}(\hat{x})-u_{\varepsilon,\delta}(\hat{y})|^2\rho(\hat{x})\rho(\hat{y})\mathrm{d}\hat{x}\mathrm{d}\hat{y} \;+ a_{\varepsilon,\delta},
\end{aligned}$$

where the second inequality is obtained using the change of variables $\hat{x} = x+z, \hat{y} = y+z$ together with the choice of $\delta < \delta'$. Jensen's inequality justifies the third one inequality. Next we introduce an error term between kernel and gradient

$$b_{\varepsilon,\delta} := \frac{1}{\varepsilon}\int_{D'\times D'} \eta_\varepsilon(x-y)\,|u_{\varepsilon,\delta}(x)-u_{\varepsilon,\delta}(y)|^2\mathrm{d}\nu(x)\mathrm{d}\nu(y) - \sigma_\eta\int_D \varepsilon|\nabla u_{\varepsilon,\delta}|^2\rho(x)^2\mathrm{d}x.$$

This allows us to write our current progress in the form

$$F_\varepsilon(u_\varepsilon) \geq \int_{D'} \varepsilon\sigma_\eta^2|\nabla u_{\varepsilon,\delta}|^2\rho(x)^2 + \frac{1}{\varepsilon}W(u_{\varepsilon,\delta})\rho(x)^2\mathrm{d}x + a_{\varepsilon,\delta} + b_{\varepsilon,\delta} \tag{6.7}$$

There are three terms left to estimate; $a_{\varepsilon,\delta} \to 0, b_{\varepsilon,\delta} \to 0$ and a lower bound on the integral term in form of the total variation. In (6.5) we showed that $|\nabla u_{\varepsilon,\delta}|$ is bounded. This means $\sigma_\eta\int_D \varepsilon|\nabla u_{\varepsilon,\delta}|^2\rho(x)^2\mathrm{d}x \underset{\varepsilon\to 0}{\longrightarrow} 0$. Furthermore, we look at the change of variables in η, where $\hat{h} := \frac{h}{\varepsilon}$ for $h \in \mathbb{R}^d$. The Jacobian of his coordinate transform $h \mapsto \hat{h}$ is given by ε^d. So we obtain the equivalence

$$\eta_\varepsilon(h)|h|^2 = \frac{1}{\varepsilon^d}\eta\Big(\frac{h}{\varepsilon}\Big)|h|^2 = \frac{\varepsilon^{d+2}}{\varepsilon^d}\eta(\hat{h})|\hat{h}|^2 = \varepsilon^2\eta(\hat{h})|\hat{h}|^2. \tag{6.8}$$

It follows, that $\frac{1}{\varepsilon}\int_{D'\times D'} \eta_\varepsilon(x-y)\,|u_{\varepsilon,\delta}(x)-u_{\varepsilon,\delta}(y)|^2\mathrm{d}\nu(x)\mathrm{d}\nu(y)$ also goes to zero as $\varepsilon \to 0$. Hence, we deduce $b_{\varepsilon,\delta} \to 0$ as $\varepsilon \to 0$. Using the Lipschitz continuity of ρ and

its upper bound Λ, we obtain

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

Since we assumed $F_\varepsilon(u_\varepsilon)$ is bounded and $u \in BV_\nu(D,\{0,1\})$, it follows $\liminf_{\delta\to 0}\liminf_{\varepsilon\to 0} a_{\varepsilon,\delta} = 0$.

In order to lower bound F_ε by the total variation, let us introduce the function $\phi(t) = \int_{1/2}^t \sqrt{W(s)}\mathrm{d}s$. Then with the simple algebraic inequality $a^2+b^2 \geq 2ab$, we are able to obtain

$$\begin{aligned}
\liminf_{\varepsilon\to 0} \int_{D'} \varepsilon|\nabla u_{\varepsilon,\delta}|^2\rho(x)^2 + \frac{1}{\varepsilon}W(u_{\varepsilon,\delta})\rho(x)^2\mathrm{d}x \geq & \liminf_{\varepsilon\to 0} \int_{D'} 2|\nabla u_{\varepsilon,\delta}|\sqrt{W(u_{\varepsilon,\delta})}\rho(x)^2\mathrm{d}x \\
= & \liminf_{\varepsilon\to 0} \int_{D'} |\nabla(\phi\circ u_{\varepsilon,\delta})|\rho(x)^2\mathrm{d}x \\
\geq & 2\int_{D'} |\nabla(\phi\circ u_\delta)|\rho(x)^2\mathrm{d}x.
\end{aligned}$$

Here the last inequality corresponds to the lower semi-continuity of the weighted total variation. Given that $u_\delta \xrightarrow{L^1(\nu)} u$ as $\delta\to 0$,

$$\begin{aligned}
\liminf_{\delta\to 0} \int_{D'} |\nabla(\phi\circ u_\delta)|\rho(x)^2\mathrm{d}x \geq & \int_{D'} |D(\phi\circ u)|\rho(x)^2\mathrm{d}x. \\
= & (\phi(1)-\phi(0))|Du|_{\rho^2}(D'). \qquad (6.9)
\end{aligned}$$

Finally, from (6.7) it follows

$$\liminf_{\varepsilon\to 0} F_\varepsilon(u_\varepsilon) \geq \sigma_\eta^2 c_W |Du|_{\rho^2}(D'),$$

where $c_W = 2(\phi(1)-\phi(0))$. As D' was an arbitrary open set compactly contained in D, we can take $D' \nearrow D$ to obtain the result.

Consider ρ is continuous but not necessarily Lipschitz. The idea is to approximate ρ from below by a family of Lipschitz functions $\{\rho_k\}_{k\in\mathbb{N}}$. We define $\rho_k : D\to\mathbb{R}$ by $\rho_k(x) := \inf_{y\in D}\rho(y) + k|x-y|$. These are Lipschitz and have the same bounds as ρ. Let $u \in L^1(\nu)$ and $u_\varepsilon \xrightarrow{L^1(\nu)} u$ for $\varepsilon\to 0$. Using the Lipschitz continuity and the fact $\rho_k \leq \rho$, we can conclude

$$\liminf_{\varepsilon\to 0} F_\varepsilon(u_\varepsilon,\rho) \geq \liminf_{\varepsilon\to 0} F_\varepsilon(u_\varepsilon,\rho_k) \geq TV(u,\nu_k).$$

By monotone convergence it follows $\lim_{k\to\infty} TV(u,\nu_k) = TV(u,\nu)$.

Limsup inequality. Let $u \in L^1(\nu)$ then we need to find a sequence $\{u_\varepsilon\}_{\varepsilon>0}$ such that for $\varepsilon \to 0$ it holds $u_\varepsilon \xrightarrow{L^1(\nu)} u$ and

$$\limsup_{\varepsilon\to 0} F_\varepsilon(u_\varepsilon) \leq \sigma_\eta^2 c_W TV(u,\nu).$$

It suffices to show the result for the non-trivial case of $u \in BV_\nu(D,\{0,1\})$. Also, we can obtain an extension to $\hat{u} \in BV_\nu(\mathbb{R}^d)$, such that $|D\hat{u}|_{\rho^2}\partial D = 0$ from [2, Proposition 3.21]. This of course implies there exists a non empty set $A \subset \mathbb{R}^d$ of finite perimeter with $u = 1_A$.

First **consider ρ Lipschitz**. *Step 1:* Assume η to have compact support on a set of radius a. The idea is to construct a one dimensional optimal profile $\chi_\varepsilon : \mathbb{R} \to \mathbb{R}$ that interpolates between zero and one while minimizing the one-dimensional equivalent to $F_\varepsilon(\chi_\varepsilon)$. See [17, Proof of Proposition 2] for more background on the construction.

We will approximate the set A from the outside. So is convenient to define the following distance function to the boundary of A

$$d(x) = \begin{cases} \operatorname{dist}(x,\partial A) & \text{if } x \in \mathbb{R}^d \backslash A \\ 0 & \text{otherwise.} \end{cases}$$

Note $d(x)$ is lipschitz continuous and the Jacobian of the transformation $|\nabla d(x)| = 1$ for all $x \notin A$. Furthermore, we denote the perimeter of the hull with distance smaller or equal ε by $\gamma_\varepsilon = \sup_{|t|\leq\varepsilon} \mathcal{H}^{d-1}(\{x \in D : d(x) = t\})$. Furthermore, if the set is weighted by say $\rho(x)$, we denote the weighted Hausdorff measure $\mathcal{H}_\rho^{d-1}(\{d(x) = t\}) := \int_{\{x\in D: d(x)=t\}} \rho(x)\, \mathrm{d}\mathcal{H}^{d-1}(x)$. We rely heavily on the fact

$$\gamma_\varepsilon \to \mathcal{H}^{d-1}(\partial A \cap D) \tag{6.10}$$

as $\varepsilon \to 0$, which holds by [17, Lemma 4].

For every $t \in \mathbb{R}$ we define

$$\phi_\varepsilon(t) = \int_0^t \frac{\varepsilon}{\sqrt{W(s)+\varepsilon}} \mathrm{d}s.$$

This allows us to define the recovery sequence for all $x \in \mathbb{R}^d$ as

$$v_\varepsilon(x) = \begin{cases} \phi_\varepsilon^{-1}(d(x)) & \text{if } 0 < d(x) \leq \phi_\varepsilon(1) \\ 1_A(x) & \text{otherwise.} \end{cases}$$

Note, for $x \in \mathbb{R}^d$ such that $d(x) \leq \phi_\varepsilon(1)$ the derivative of v_ε is given as

$$\nabla v_\varepsilon(x) = \frac{\nabla d(x)}{\phi_\varepsilon'(\phi_\varepsilon^{-1})} = \frac{\sqrt{W(\phi_\varepsilon^{-1}(d(x))) + \varepsilon}}{\varepsilon} \nabla d(x) \tag{6.11}$$

by the inverse function rule, since $\frac{\mathrm{d}}{\mathrm{d}t}\phi_\varepsilon(t) = \frac{\varepsilon}{\sqrt{W(t)+\varepsilon}}$. This shows, that v_ε is Lipschitz with constant of order $\frac{1}{\varepsilon}$, a property we require later in the discrete to local liminf equality.

First of all, we show that $v_\varepsilon \to u$ in $L^1(\nu)$. The difference $|v_\varepsilon(x) - u(x)|$ is non-trivial only on the set $S_\varepsilon = \{x \in \mathbb{R}^d : d(x, \partial A) \leq \phi_\varepsilon(1), x \notin A\}$ by construction of v_ε. Using the coarea formula for Lipschitz functions on d and the definition of γ_ε as supremum, we obtain

$$\begin{aligned}\int_D |v_\varepsilon(x) - u(x)|\mathrm{d}\nu &= \int_{S_\varepsilon} |v_\varepsilon(x)|\mathrm{d}\nu &(6.12)\\ &= \int_{S_\varepsilon} |\phi_\varepsilon^{-1}(d(x))|\rho(x)\ \mathrm{d}x &(6.13)\\ &= \int_0^{\phi_\varepsilon(1)} |\phi_\varepsilon^{-1}(t)||\nabla d|\mathcal{H}^{d-1}_\rho(\{d(x) = t\})\ \mathrm{d}t\\ &\leq 2\phi_\varepsilon(1)\gamma_{\phi_\varepsilon(1)}\Lambda \leq 4\varepsilon^{\frac{1}{2}}\gamma_{2\varepsilon^{1/2}}\Lambda.\end{aligned}$$

The expression on the right goes to zero for $\varepsilon \to 0$ as $\gamma_\varepsilon \to \mathcal{H}^{d-1}(\partial A \cap D)$.

Let $D_\delta := \{x \in \mathbb{R}^d : \mathrm{dist}(x, D) < a\delta\}$, the set containing D where $\eta\left(\frac{x-y}{\delta}\right) > 0$ for $y \in D$. It holds $D_\delta = D \cup B(y, a\delta)$ for $y \in D$. We introduce the parameter $\delta \geq \varepsilon$ to uncouple the convergence of v_ε from the hull D_δ for technical reasons. For every $\varepsilon > 0$ it holds

$$\begin{aligned}&\frac{1}{\varepsilon}\int_D\int_D \eta_\varepsilon(x-y)|v_\varepsilon(x) - v_\varepsilon(y)|^2\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y\\ &\qquad= \frac{1}{\varepsilon}\int_D\int_{D\cap B(y,a\varepsilon)} \eta_\varepsilon(x-y)|v_\varepsilon(x) - v_\varepsilon(y)|^2\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y\\ &\qquad= \frac{1}{\varepsilon}\int_D\int_{B(y,a\varepsilon)} \eta_\varepsilon(x-y)\Big|\int_y^x \nabla v_\varepsilon(t)\cdot(x-y)\mathrm{d}t\Big|^2\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y\\ &\qquad\leq \frac{1}{\varepsilon}\int_D\int_{B(y,a\varepsilon)}\int_y^x \eta_\varepsilon(x-y)|\nabla v_\varepsilon(t)\cdot(x-y)|^2\rho(x)\rho(y)\mathrm{d}t\mathrm{d}x\mathrm{d}y\\ &\qquad= \frac{1}{\varepsilon}\int_D\int_{|\tilde{h}|<a\varepsilon}\int_y^{y+\tilde{h}} \eta_\varepsilon(h)|\nabla v_\varepsilon(t)\cdot\tilde{h}|^2\rho(y+\tilde{h})\rho(y)\mathrm{d}t\mathrm{d}\tilde{h}\mathrm{d}y.\end{aligned}$$

The first equality uses η's support of radius a on D and the second simply by the fundamental theorem of calculus. The inequality is an application of the triangle inequality. Furthermore, we used $\tilde{h} = x - y$ and the equality $x = y + \tilde{h}$. Next we utilize a change of variables $z = y + t(x - y)$ and a transformation $h = \frac{\tilde{h}}{\varepsilon}$ as in (6.8). The change of variables allows us to rewrite $y = z - t\tilde{h}$ and $x = y + \tilde{h} = z + (1-t)\tilde{h}$. Let

$$a_{\delta,\varepsilon} = \varepsilon\int_{D_\delta}\int_{|h|<a}\int_0^1 \eta(h)|\nabla v_\varepsilon(z)\cdot h|^2\big(\rho(z - t\delta h)\rho(z + (1-t)\delta h) - \rho(z)^2\big)\mathrm{d}t\mathrm{d}h\mathrm{d}z.$$

Recall $\delta \geq \varepsilon$, to argue

$$\begin{aligned}
\frac{1}{\varepsilon}\int_D \int_{|\tilde{h}|<a\varepsilon} \int_y^{y+\tilde{h}} \eta_\varepsilon(\tilde{h})|\nabla v_\varepsilon(t)\cdot\tilde{h}|^2\rho(y+\tilde{h})\rho(y)\mathrm{d}t\mathrm{d}\tilde{h}\mathrm{d}y \\
=&\frac{1}{\varepsilon}\int_D \int_{|\tilde{h}|<a\varepsilon} \int_0^1 \eta_\varepsilon(\tilde{h})|\nabla v_\varepsilon(y+t\tilde{h})\cdot\tilde{h}|^2\rho(y+\tilde{h})\rho(y)\mathrm{d}t\mathrm{d}\tilde{h}\mathrm{d}y \\
\leq&\varepsilon\int_{D_\varepsilon}\int_{|h|<a}\int_0^1 \eta(h)|\nabla v_\varepsilon(z)\cdot h|^2\rho(z-t\varepsilon h)\rho(z+(1-t)\varepsilon h)\mathrm{d}t\mathrm{d}h\mathrm{d}z \\
\leq&\varepsilon\int_{D_\delta}\int_{|h|<a}\int_0^1 \eta(h)|\nabla v_\varepsilon(z)\cdot h|^2\rho(z-t\delta h)\rho(z+(1-t)\delta h)\mathrm{d}t\mathrm{d}h\mathrm{d}z \\
=&a_{\delta,\varepsilon}+\varepsilon\int_{D_\delta}\int_{|h|<a}\eta(h)|\nabla v_\varepsilon(z)\cdot h|^2\rho(z)^2\mathrm{d}h\mathrm{d}z \\
=&a_{\delta,\varepsilon}+\varepsilon\int_{D_\delta}\int_{|h|<a}\eta(h)\left|\frac{\nabla v_\varepsilon(z)}{|\nabla v_\varepsilon(z)|}\cdot h\right|^2|\nabla v_\varepsilon(z)|^2\rho(z)^2\mathrm{d}h\mathrm{d}z \\
=&a_{\delta,\varepsilon}+\varepsilon\int_{D_\delta}\int_{|h|<a}\eta(h)|h|^2\mathrm{d}h|\nabla v_\varepsilon(z)|^2\rho(z)^2\mathrm{d}z \\
=&a_{\delta,\varepsilon}+\sigma_\eta^2\int_{D_\delta}\varepsilon|\nabla v_\varepsilon(z)|^2\rho(z)^2\mathrm{d}z,
\end{aligned}$$

The third last equality requires isometry of the kernel η. So far we have shown

$$F_\varepsilon(v_\varepsilon)\leq\sigma_\eta^2\int_{D_\delta}\varepsilon|\nabla v_\varepsilon(x)|^2\rho(x)^2\mathrm{d}x+\frac{1}{\varepsilon}\int_{D_\delta}W(v_\varepsilon(x))\rho(x)^2\mathrm{d}x+a_{\delta,\varepsilon}. \tag{6.14}$$

Since ρ is Lipschitz and bound, the the error introduced above can be bound by

$$\begin{aligned}
a_{\delta,\varepsilon}\leq&\delta(\Lambda-\lambda)\mathrm{Lip}(\rho)\int_{D_\delta}\int_{|h|<a}\varepsilon\,\eta(h)|\nabla v_k(z)\cdot h|^2\rho(z)^2\mathrm{d}h\mathrm{d}z \\
\leq&\delta(\Lambda-\lambda)\mathrm{Lip}(\rho)\int_{D_\delta}\sigma_\eta^2\varepsilon|\nabla v_\varepsilon(z)|^2\rho(z)^2\mathrm{d}z \\
\leq&\delta(\Lambda-\lambda)\mathrm{Lip}(\rho)\sigma_\eta^2\int_{D_\delta}\varepsilon|\nabla v_\varepsilon(x)|^2\rho(x)^2+\frac{1}{\varepsilon}W(v_\varepsilon(x))\rho(x)\mathrm{d}x.
\end{aligned} \tag{6.15}$$

This term converges to zero as $\delta\to 0$, a fact that will become apparent shortly.

We define $\gamma_{\phi_\varepsilon(1),\rho}=\sup_{|t|\leq\phi_\varepsilon(1)}\int_{\{x\in\mathbb{R}^d:d(x)=t\}}\rho(x)^2\,\mathrm{d}\mathcal{H}^{d-1}(x)$, the weighted perimeter of the level set with distance less or equal $\phi_\varepsilon(1)$. As in (6.10) it holds $\gamma_{\phi_\varepsilon(1),\rho}\overset{\varepsilon\to 0}{\longrightarrow}|D\hat{u}|_{\rho^2}(D_\delta)$, i.e. the perimeter of A in D_δ. Recall the set $S_\varepsilon=\{x\in\mathbb{R}^d:d(x,\partial A)\leq$

$\phi_\varepsilon(1), x \notin A\}$, outside of which holds $v = 1_A$. So we can upper bound

$$\begin{aligned}
&\int_{D_\delta} \varepsilon|\nabla v_\varepsilon(x)|^2\rho(x)^2 + \frac{1}{\varepsilon}W(v_\varepsilon(x))\rho(x)^2 \mathrm{d}x \\
&\qquad = \int_{S_\varepsilon} \varepsilon|\nabla v_\varepsilon(x)|^2\rho(x)^2 + \frac{W(v_\varepsilon(x))}{\varepsilon}\rho(x)^2 \mathrm{d}x \\
&\qquad = \int_0^{\phi_\varepsilon(1)} \Big(\varepsilon|\phi_\varepsilon' \circ \phi_\varepsilon^{-1}(t)|^2 + \frac{W(\phi_\varepsilon^{-1}(t))}{\varepsilon}\Big)\mathcal{H}^{d-1}_{\rho^2}(\{d(x) = t\})\mathrm{d}t \\
&\qquad \leq \gamma_{\phi_\varepsilon(1),\rho} \int_0^{\phi_\varepsilon(1)} \varepsilon|\phi_\varepsilon' \circ \phi_\varepsilon^{-1}(t)|^2 + \frac{W(\phi_\varepsilon^{-1}(t))}{\varepsilon}\mathrm{d}t \\
&\qquad \leq \gamma_{\phi_\varepsilon(1),\rho} \int_0^{\phi_\varepsilon(1)} \varepsilon\frac{W(\phi_\varepsilon^{-1}(t)) + \varepsilon}{\varepsilon^2} + \frac{W(\phi_\varepsilon^{-1}(t))}{\varepsilon}\mathrm{d}t \\
&\qquad \leq \frac{2\gamma_{\phi_\varepsilon(1),\rho}}{\varepsilon} \int_0^{\phi_\varepsilon(1)} W(\phi_\varepsilon^{-1}(t)) + \varepsilon \ \mathrm{d}t \\
&\qquad = \frac{2\gamma_{\phi_\varepsilon(1),\rho}}{\varepsilon} \int_0^1 (W(s) + \varepsilon)\frac{\varepsilon}{\sqrt{W(s) + \varepsilon}} \ \mathrm{d}s \\
&\qquad = 2\gamma_{\phi_\varepsilon(1),\rho} \int_0^1 \sqrt{W(s) + \varepsilon} \ \mathrm{d}s.
\end{aligned}$$

In the second to last step, we applied the change of variable $s = \phi_\varepsilon(t)$. From (6.15) we deduce $\lim_{\delta\to 0}\lim_{\varepsilon\to 0} a_{\delta,\varepsilon} = 0$. For the term in (6.14), it follows

$$\lim_{\varepsilon\to 0} F_\varepsilon(v_\varepsilon) \leq \sigma_\eta^2 c_w |D\hat{u}|_{\rho^2}(D_\delta). \tag{6.16}$$

Furthermore, since $\hat{u}$ is an extension of u such that it holds $|D\hat{u}|_{\rho^2}\partial D = 0$, it holds

$$\lim_{\delta\to 0} |D\hat{u}|_{\rho^2}(D_\delta) = |D\hat{u}|_{\rho^2}(\bar{D}) = |D\hat{u}|_{\rho^2}(D) = TV(u, \nu). \tag{6.17}$$

Finally, combining equations (6.16) and (6.17) we obtain

$$\limsup_{\varepsilon\to 0} F_\varepsilon(v_\varepsilon) \leq \sigma_\eta^2 c_W TV(u, \nu). \tag{6.18}$$

Step 2: For general η see steps 2 to 4 in the proof of Theorem 5.1.

Assume ρ continuous but not necessarily Lipschitz. The idea is to approximate ρ from above by a family of Lipschitz functions $\{\rho_k\}_{k\in\mathbb{N}}$. We define $\rho_k : D \to \mathbb{R}$ by

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

These are Lipschitz and have the same bounds as ρ. The limsup inequality in for a Lipschitz density (6.18) and the fact $\rho \leq \rho_k$ imply

$$\limsup_{\varepsilon\to 0} F_\varepsilon(u, \rho) \leq \limsup_{\varepsilon\to 0} F_\varepsilon(u, \rho_k) \leq TV(u, \nu_k).$$

By the dominated convergence theorem we obtain the desired result

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

□

One additional remark on the behaviour of the double well potential as ε approaches zero. Continuity of W and Fatou's Lemma, allow us to conclude from the bound on F_ε, that

$$\int_D W(u)\rho(x)\ \mathrm{d}\nu \leq \liminf_{\varepsilon\to 0} \int_D W(u_\varepsilon)\rho(x)\ \mathrm{d}\nu \leq \liminf_{\varepsilon\to 0} \varepsilon F_\varepsilon(u_\varepsilon) = 0. \tag{6.19}$$

6.3 Graph Convergence

Theorem 6.2 (Γ-Convergence)**.** *Under the assumption 6.1 stated above, it holds*

$$\mathcal{G}_n \xrightarrow{\Gamma} \sigma_\eta^2 c_W \mathcal{G}$$

with respect to the $TL^1(D)$ metric.

Proof. The main challenge lies in showing the convergence for a kernel η of the form

$$\eta(t) := \begin{cases} a & \text{for } t < b \\ 0 & \text{else} \end{cases}$$

where after the generalization is similar to previous proofs. we will show the convergence for $G_n(u_n)$ to $TV(u,\nu)$, then it follows for $\mathcal{G}_n$ by continuity of B^α. Let $\{T_n\}_{n\in\mathbb{N}}$ be stagnating sequence of transportation maps as given by Proposition 4.4.

Liminf inequality: *Case 1:* Assume $u \in BV(D,\{0,1\})$ and $B^\alpha(u) > 0$. Let $u_n \overset{TL^1}{\to} u$ as $n \to \infty$. Exactly as in the proof of Theorem 5.1 for the perturbation $\tilde{\varepsilon}_n := \varepsilon_n - \frac{2}{b}\|\mathrm{Id} - T_n\|_\infty$ it holds

$$\eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big) \leq \eta\Big(\frac{T_n(x)-T_n(y)}{\varepsilon_n}\Big).$$

We define $\tilde{u}_n := u_n \circ T_n$ and

$$\begin{aligned}
G_n(u_n) =& \frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \eta\left(\frac{x-y}{\varepsilon_n}\right) |u_n(x) - u_n(y)|^2 \mathrm{d}\nu_n(x)\mathrm{d}\nu_n(y) \\
&+ \frac{1}{\varepsilon_n} \int_{D\times D} \eta_\varepsilon(|x-y|) W(u_n)\ \mathrm{d}\nu_n(x)\ \mathrm{d}\nu_n(y) \\
\geq& \frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \eta\left(\frac{T_n(x)-T_n(y)}{\varepsilon_n}\right) |\tilde{u}_n(x) - \tilde{u}_n(y)|^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
&+ \frac{1}{\varepsilon_n} \int_D W(\tilde{u}_n)\rho(x)\mathrm{d}\nu(x) + b_{\varepsilon_n} \\
\geq& \frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big) |\tilde{u}_n(x) - \tilde{u}_n(y)|^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
&+ \frac{1}{\varepsilon_n} \int_D W(\tilde{u}_n)\rho(x)\mathrm{d}\nu(x) + b_{\varepsilon_n} \\
=& \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{d+1} F_{\tilde{\varepsilon}}(\tilde{u}_n) + b_{\varepsilon_n},
\end{aligned}$$

where

$$b_{\varepsilon_n} = \frac{1}{\varepsilon_n} \int_{D\times D} \eta_\varepsilon(|x-y|) W(\tilde{u}_n)\ \mathrm{d}\nu_n(x)\ \mathrm{d}\nu_n(y) - \frac{1}{\varepsilon_n} \int_D W(\tilde{u}_n)\rho(x)\ \mathrm{d}\nu(x).$$

As we noted in (6.4) it holds $b_{\varepsilon_n} \to 0$ as $n \to \infty$. Furthermore, It holds $\frac{\tilde{\varepsilon}_n}{\varepsilon_n} \to 1$ and $u_n \overset{TL^1}{\to} u$ implies $\tilde{u}_n \overset{L^1(D,\nu)}{\longrightarrow} u$ as $n \to \infty$. By Theorem 6.1 it therefore follows

$$\liminf_{n\to\infty} G_n(u_n) \geq \liminf_{n\to\infty} F_{\tilde{\varepsilon}}(\tilde{u}_n) \geq \sigma_\eta^2 c_W TV(u).$$

Since B^α is continuous in $L^1(\nu)$ and $B^\alpha(u_n) \to B^\alpha(u)$ by Lemma 5.1 we can conclude

$$\liminf_{n\to\infty} \mathcal{G}_n(u_n) = \liminf_{n\to\infty} \frac{G_n(u_n)}{B_n^\alpha(u_n)} \geq \sigma_\eta^2 c_W \frac{TV(u)}{B^\alpha(u)} = \mathcal{G}(u).$$

Case 2: We have used several times already the fact that the set of all characteristic function is closed in $L^1(\nu)$, so the inequality holds for $u \notin BV(D, \{0,1\})$. Furthermore, in the case $B^\alpha(u) = 0$ it follows with the continuity $\lim_{n\to\infty} B^\alpha(u_n) = 0$, see Lemma 5.1, and the inequality holds.

Limsup inequality: Let $u \in BV(D, \{0,1\})$ and $v_n \in L^1(D,\nu)$ be a recovery sequence for the Γ-convergence of $F_{\tilde{\varepsilon}_n}(v_n)$ to $TV(u,\nu)$, where $\tilde{\varepsilon}_n := \varepsilon_n + \frac{2}{b}\|\mathrm{Id} - T_n\|_\infty$. We define u_n to be restriction of v_n to the first n datapoints, with T_n a stagnating sequence of transport maps. We know from (6.11) v_n is Lipschitz. We set $\tilde{u}_n := u_n \circ T_n$ and note $|\tilde{u}_n(x) - v_n(x)| \leq \mathrm{Lip}(v_n)|x - T_n(x)|$. Once for which it holds once again

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

We want to show $G_n(u_n) \leq F_{\tilde{\varepsilon}}(v_n)$ and leverage Theorem 6.1 in a similar fashion to the liminf proof.

$$\begin{aligned}
G_n(u_n) =& \frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \eta\Big(\frac{x-y}{\varepsilon_n}\Big)|u_n(x)-u_n(y)|^2 \mathrm{d}\nu_n(x)\mathrm{d}\nu_n(y) \\
&+ \frac{1}{\varepsilon_n}\int_{D\times D} \eta_\varepsilon(|x-y|)W(u_n)\ \mathrm{d}\nu_n(x)\ \mathrm{d}\nu_n(y) \\
=& \frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \eta\Big(\frac{T_n(x)-T_n(y)}{\varepsilon_n}\Big)|\tilde{u}_n(x)-\tilde{u}_n(y)|^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
&+ \frac{1}{\varepsilon_n}\int_{D} W(\tilde{u}_n)\rho(x)\mathrm{d}\nu(x) + b_{\varepsilon_n} \\
\leq& \frac{1}{\varepsilon_n^{d+1}} \int_{D\times D} \eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big)|\tilde{u}_n(x)-\tilde{u}_n(y)|^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
&+ \frac{1}{\varepsilon_n}\int_{D} W(\tilde{u}_n)\rho(x)\mathrm{d}\nu(x) + b_{\varepsilon_n} \\
=& \frac{1}{\tilde{\varepsilon}_n^{d+1}} \int_{D\times D} \eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big)|v_n(x)-v_n(y)|^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
&+ \frac{1}{\tilde{\varepsilon}_n}\int_{D} W(v_n)\rho(x)\mathrm{d}\nu(x) + a_{\varepsilon_n} + b_{\varepsilon_n},
\end{aligned}$$

where

$$\begin{aligned}
a_{\varepsilon_n} &= \frac{c_n}{\varepsilon_n}\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)\left(|\tilde{u}_n(x)-\tilde{u}_n(y)|^2 - |v_n(x)-v_n(y)|^2\right)\mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
b_{\varepsilon_n} &= \frac{1}{\varepsilon_n}\int_{D\times D} \eta_\varepsilon(|x-y|)W(\tilde{u}_n)\ \mathrm{d}\nu_n(x)\ \mathrm{d}\nu_n(y) - \frac{1}{\varepsilon_n}\int_D W(\tilde{u}_n)\rho(x)\ \mathrm{d}\nu(x).
\end{aligned}$$

Here c_n is a sequence to account for the difference $\varepsilon_n, \tilde{\varepsilon}_n$ such that $c_n \to 1$. The term $b_{\varepsilon_n} \to 0$ as $n \to \infty$ by (6.4). From Proposition 4.4 it follows $b_{\varepsilon_n} \to 0$ for $n \to \infty$. For the bound on a_{ε_n} we need two additional results. Given $a, b \in \mathbb{R}$ the generalized triangle inequality for $p \geq 1$ is given by

$$|a+b|^p \leq 2^{p-1}(|a|^p + |b|^p). \tag{6.20}$$

Furthermore, for $\tilde{a}, \tilde{b} \in \mathbb{R}$ by Lemma 6.1 it holds

$$|\tilde{a}|^p - |\tilde{b}|^p \leq \delta|\tilde{b}|^p + C_\delta|\tilde{a}-\tilde{b}|^p.$$

First we utilize Lemma 6.1 for fixed $\delta > 0$ and with the values $\tilde{a} = \tilde{u}_n(x) - \tilde{u}_n(y)$ and $\tilde{b} = v_n(x) - v_n(y)$ to estimate

$$\begin{aligned}
a_{\varepsilon_n} \leq C_\delta \frac{c_n}{\varepsilon_n}\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|\tilde{u}_n(x)-\tilde{u}_n(y)-v_n(x)+v_n(y)|^2\mathrm{d}\nu(x)\mathrm{d}\nu(y) \\
+ \delta\frac{c_n}{\varepsilon_n}\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|v_n(x)-v_n(y)|^2\mathrm{d}\nu(x)\mathrm{d}\nu(y).
\end{aligned}$$

By choice of v_n the second term is bound as $n \to \infty$, so it vanishes when taking $\delta \to 0$. The first term we denote with d_{ε_n} to treat it further. Next, set $a = \tilde{u}_n(x) - v_n(x)$ and $b = v_n(y) - \tilde{u}_n(y)$ in (6.20), this yields

$$d_{\varepsilon_n} \leq 4C_\delta \frac{c_n}{\varepsilon_n} \int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)|\tilde{u}_n(x) - v_n(x)|^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y).$$

Let $C_\eta = c_n \int_{\mathbb{R}^d} \eta_{\tilde{\varepsilon}_n}(h)\mathrm{d}h$, we can estimate

$$\begin{aligned} d_{\varepsilon_n} \leq & 4C_\delta \frac{c_n}{\varepsilon_n} \int_D \int_{\mathbb{R}^d} \eta_{\tilde{\varepsilon}_n}(h)\rho(h)\mathrm{d}h|\tilde{u}_n(x) - v_n(x)|^2\rho(x)\mathrm{d}x \\ \leq & 4C_\delta \Lambda^2 \frac{C_\eta}{\varepsilon_n} \int_D |\tilde{u}_n(x) - v_n(x)|^2\rho(x)\mathrm{d}x \\ \leq & 4C_\delta \Lambda^2 \frac{C_\eta}{\varepsilon_n} \mathrm{Lip}(v_n) \int_D |x - T_n(x)|^2\rho(x)\mathrm{d}x. \end{aligned}$$

We know $\mathrm{Lip}(v_n)$ scales as $\frac{1}{\varepsilon_n}$ while the integral term in the last equation scales as $\left(\frac{\log n}{n}\right)^{2/d}$. By the scaling assumptions on ε_n we conclude that the last term goes to zero for $n \to \infty$, and it follows $a_{\varepsilon_n} \to 0$. With the convergence of the non-local functional we can deduce

$$\limsup_{n\to\infty} G_n(u_n) \leq \limsup_{n\to\infty} F_{\tilde{\varepsilon}_n}(v_n) \leq \sigma_\eta^2 c_W TV(u,\nu).$$

Finally, with continuity of $B^\alpha(u)$ in $L^1(\nu)$ and $B^\alpha(u_n) \to B^\alpha(u)$ by Lemma 5.2, we conclude

$$\limsup_{n\to\infty} \mathcal{G}_n(u_n) \leq \sigma_\eta^2 c_W \mathcal{G}(u).$$

□

Lemma 6.1. *Let $a, b \in \mathbb{R}$ and $p \in \mathbb{N}$. For all $\delta > 0$ there exists a constant C_δ such that the following inequality holds*

$$|a|^p \leq (1+\delta)|b|^p + C_\delta |a-b|^p.$$

Proof. The inequality obviously holds for any $\delta > 0$ if $a = b$ or $a = 0$. For the case $a \neq b$ and $a \neq 0$ we may assume $b = 1$ without loss of generality. As such, it is sufficient to show

$$\frac{|a|^p - 1 - \delta}{|a-1|^p} \leq C_\delta.$$

Using the negative generalized triangle inequality (6.20) it holds

$$\begin{aligned} \frac{|a|^p - 1 - \delta}{|a-1|^p} &\leq \frac{|a|^p - 1 - \delta}{2^{p-1}|a|^p - 1} \\ &\leq \frac{|a|^p - 1 - \delta}{|a|^p - 1} \\ &= 1 - \frac{\delta}{|a|^p - 1}. \end{aligned}$$

The denominator of the last term is larger than -1. Therefore, we can choose $C_\delta = 1 + \delta$. □

6.4 Multi-phase convergence

We saw in (3.3) that for the perimeter problems involving multiple sets, we need an additional orthogonality constrain. Functions that do not satisfy said constrain are of no interest to the problem at hand. With that in mind we define the sets of functions for the multi phasefield cheeger cut problems

$$\mathcal{M}(D) = \{(u^1, \dots, u^R) : \ u^r \in BV(D, \{0,1\}), B^\alpha(u^r) > 0, \int_D u^r(x)u^s(x)\,\mathrm{d}\nu(x) = 0 \text{ if } r \neq s\}$$

$$\mathcal{M}_n(D) = \{(u_n^1, \dots, u_n^R) : \ u_n^r \in L^1(\nu_n), B_n^\alpha(u_n^r) > 0, \int_D u_n^r(x)u_n^s(x)\,\mathrm{d}\nu_n(x) = 0 \text{ if } r \neq s\}$$

where $B_n^\alpha(u) = (|Y_n|)^\alpha, B^\alpha(u) = (\int u\,\mathrm{d}\nu(x))^\alpha$. The Graph cut problem is then given by

$$E_n(\mu, \mathcal{U}_n) := \begin{cases} \sum_{r=1}^R \frac{G_n(u_n^r)}{B^\alpha(u_n^r)} & \text{if } \mu = \nu_n \text{ and } \mathcal{U}_n \in \mathcal{M}_n(D) \\ +\infty & \text{otherwise,} \end{cases}$$

while the corresponding continuum problem is

$$E(\mu, \mathcal{U}) := \begin{cases} \sum_{r=1}^R \frac{TV(u^r,\nu)}{B^\alpha(u^r)} & \text{if } \mu{=}\nu \text{ and } \mathcal{U} \in \mathcal{M}(D) \\ +\infty & \text{otherwise.} \end{cases}.$$

Theorem 6.3 (Multi Phase Γ-Convergence)**.** *If the assumptions 6.1 hold, it follows*

$$E_n \xrightarrow{\Gamma} \sigma_\eta^2 c_W E$$

with respect to $TL^1(D)^R$ metric forn $\to \infty$.

Proof. **Liminf inequality:** For a arbitrary sequence $\{\mathcal{U}_n\} \subset L^1(\nu_n)^R$ with $\mathcal{U}_n \overset{TL^1}{\to} \mathcal{U}$ for some $\mathcal{U} \in L^1(\nu)^R$, we need to show

$$\liminf_{n\to\infty} E_n(\mathcal{U}_n) \geq \sigma_\eta E(\mathcal{U}).$$

We showed in Lemma 3.1 that the set of indicator functions closed in $L^1(\nu)$. For any $u_n \xrightarrow{L^1(\nu)} u$ with $B^\alpha(u_n) > 0$ the continuity of B^α Lemma 5.1 allows to infer$B^\alpha(u) > 0$. Furthermore, for any $u_n^r, u_n^s \in \mathcal{M}(D)$ orthogonality of $u^r, u^s \in L^1(\nu)$ follows just like in Lemma 5.1 (iii) from Fatou's lemma. Therefore, $\mathcal{M}(D)$ is closed in $L^1(\nu)^R$ and

analogously to Theorem 5.1 we may assume $\mathcal{U} \in \mathcal{M}(D)$ and $\mathcal{U}_n \in \mathcal{M}_n(D)$. Then it directly follows from the result for a single set Theorem 6.2 that

$$\begin{aligned}\liminf_{n\to\infty} E_n(\mathcal{U}_n) =& \liminf_{n\to\infty} \sum_{r=1}^{R} \frac{G_n(u_n)}{B_n^\alpha(u_n^r)} \geq \sum_{r=1}^{R} \liminf_{n\to\infty} \frac{G_n(u_n)}{B_n^\alpha(u_n^r)} \\ \geq& \sigma_\eta^2 c_W \sum_{r=1}^{R} \frac{TV(u^r)}{B^\alpha(u^r)} = E_n(\mathcal{U}).\end{aligned}$$

Limsup inequality: Consider $\mathcal{U}$ with defining sets $\{A^1, \dots, A^R\}$, as we argued in Theorem 5.1 we may assume these sets to have finite perimeter and be mutually disjoint. We first construct a recovery sequence for $\mathcal{U}$ with defining sets $\{A^1, \dots, A^R\}$ of the form $A^r = B_r \cap D$, where B_r has piecewise smooth boundary and satisfies $|D1_{B_r}|_{\rho^2}(\partial D) = 0$. Let $Y_k^r = A^r \cap X_n$denote the restriction of A^r to the first n data points. By Proposition 4.4 there exists a sequence of transport maps $\{T_n\}_{n\in\mathbb{N}}$ such that $1_{A_n^r} := 1_{Y_k^r} \circ T_n \to 1_{A^r}$ for $n \to \infty$. Here A_n^r denotes in some sense the set after a sample has been transported back to A^r. Note that by the change of variables

$$\int_D 1_{A_n^r}(x)\,\mathrm{d}\nu_n(x) = \int_D 1_{A_n^r}(T_n(x))\,\mathrm{d}\nu(x)$$

we have, $|Y_k^r| = |A_n^r| \to |A^r|$ as $n \to \infty$. Again, we may assume η is of the form $\eta(|z|) = a$ if $|z| < b$ and zero otherwise. We define a small perturbation $\tilde{\varepsilon}_n := \varepsilon_n + \frac{2}{b}\|\mathrm{Id} - T_n\|_\infty$so it follows $\eta\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon_n}\right) \leq \eta\left(\frac{|x-y|}{\tilde{\varepsilon}_n}\right)$. This in turn implies

$$\frac{\varepsilon_n^{d+1}}{\tilde{\varepsilon}_n^{d+1}} G_n(1_{Y_n^r}) \leq F_{\tilde{\varepsilon}_n}(1_{A_n^r}).$$

Since the difference between indicator functions is either one or zero, we are able to bound the error by the same technique as in (5.12), namely

$$\begin{aligned}|F_{\tilde{\varepsilon}_n}(1_{A_n^r}) - F_{\tilde{\varepsilon}_n}(1_{A^r})| =& \Big|\frac{1}{\tilde{\varepsilon}_n}\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)\left(|1_{A_n^r}(x) - 1_{A_n^r}(y)|^2\right. \\ & \left. -|1_{A^r}(x) - 1_{A^r}(y)|^2\right)\mathrm{d}\nu(x)\mathrm{d}\nu(y)\Big| \\ =& \Big|\frac{1}{\tilde{\varepsilon}_n}\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)\left(|1_{A_n^r}(x) - 1_{A_n^r}(y)|\right. \\ & \left. -|1_{A^r}(x) - 1_{A^r}(y)|\right)\mathrm{d}\nu(x)\mathrm{d}\nu(y)\Big| \\ \leq& \frac{1}{\tilde{\varepsilon}_n}\int_{D\times D} \eta_{\tilde{\varepsilon}_n}(x-y)\mathrm{d}\nu(x)\mathrm{d}\nu(y)\ \|1_{A_n^r} - 1_{A^r}\|_{L^1(\nu)} \\ \leq& \frac{K_0}{\tilde{\varepsilon}_n}\|1_{A_n^r} - 1_{A^r}\|_{L^1(\nu)}\end{aligned}$$

for some constant $K_0 \leq \sigma_\eta^2$. Note that $\int_D W(1_A)\mathrm{d}\nu = 0$ for any set $A \subset \mathbb{R}^d$. From the result on transport maps Proposition 4.4 we know that for n large enough, the difference $1_{A_n^r} - 1_{A^r}$ is trivial far enough inside of A^r. The error on the tubular neighbourhood of the boundary depends on n via $\|\mathrm{Id} - T_n\|_\infty$. Weyl's volume formula for tubes [19] gives a constant depending on surface and volume of A^r and therefore B_r. We obtain

$$\|1_{A_n^r} - 1_{A^r}\|_{L^1(\nu)} \leq C_0(B_r)\|\mathrm{Id} - T_n\|_\infty.$$

Since $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$ the previous inequalities imply

$$\begin{aligned}\limsup_{n\to\infty} G_n(1_{Y^r}) &\leq \limsup_{n\to\infty} F_{\tilde{\varepsilon}_n}(1_{A_n^r}) = \limsup_{n\to\infty} F_{\tilde{\varepsilon}_n}(1_{A^r}) \\ &\leq \sigma_\eta^2 c_W TV(1_{A^r}).\end{aligned}$$

Recall we have shown in Theorem 5.3, that subsets with piecewise smooth boundaries are dense in the set of indicator functions with finite perimeter that satisfy the diagonality constrain. Together with the continuity of B^α and Lemma 5.2 the result follows. □

6.5 Compactness and Minimizers

Theorem 6.4 (Compactness $\mathcal{G}_n$)**.** *It holds with probability one, that any sequence* $\{u_n\}$ *with* $u_n \in L^1(\nu_n)$ *with*

$$\limsup_{n\to\infty} \mathcal{G}_n(u_n) < \infty \tag{6.21}$$

is precompact in $TL^1(D)$.

To show this Theorem, we have to rely on the following result from [9, Theorem 3.2 (i)].

Proposition 6.1. *Let* $\varepsilon_n \to 0$ *and* $u_n \in L^1(\nu)$ *satisfy* $\sup_{n\in\mathbb{N}} \mathcal{F}_{\varepsilon_n}^{(p)}(u_n) < \infty$*, where*

$$\mathcal{F}_\varepsilon^{(p)}(u) = \frac{1}{\varepsilon}\int_{D\times D} \eta_\varepsilon(x-z)|u(x)-u(z)|^p \rho(x)\rho(z)\,dx\,dz + \frac{1}{\varepsilon}\int_D W(u(x))\rho(x)\,dx.$$

Then, u_n *is relatively compact in* $L^1(\nu)$.

Proof. Similar to the compactness proof of the graph total variation, we want to leverage Proposition 6.1 for $p = 2$. We may assume $\sup_{n\in\mathbb{N}} \mathcal{G}_n(u_n) < \infty$, since the asymptotically result needs to hold for n sufficiently large by (6.21), and consequently it also holds $B_n^\alpha(u_n) > 0$.

Once again, we can assume η is of the form $\eta(|z|) = a$ if $|z| < b$ and $\eta(|z|) = 0$ otherwise, without loss of generality. Furthermore, as in the reasoning for (5.9) we define a small perturbation $\tilde{\varepsilon}_n := \varepsilon_n - \frac{2}{b}\|\mathrm{Id} - T_n\|_\infty$ and deduce

$$\eta\Big(\frac{x-y}{\tilde{\varepsilon}_n}\Big) \leq \eta\Big(\frac{T_n(x) - T_n(y)}{\varepsilon_n}\Big).$$

Let $\tilde{u}_n := u_n \circ T_n$ then as before, since $\frac{\varepsilon_n^{d+1}}{\tilde{\varepsilon}_n^{d+1}} \to 1$ for $n \to \infty$, it follows

$$G_n(u_n) \geq F_{\tilde{\varepsilon}_n}(\tilde{u}_n). \tag{6.22}$$

Furthermore, we can deduce $\mathcal{F}_\varepsilon^{(2)}(u) < \infty$ and the desired result follows immediately from Theorem 6.1 and the continuity of the balance term. □

The theorems we have proven in this chapter, allow for a consistency result identical to Theorem 5.5. This consistency is build on top the three results of Γ-convergence, compactness and the ensuing convergence of minimizers.

7 Numerical Examples

7.1 Problem Formulation and Gradients

In this section, we present a problem formulation that is suitable for numerical applications, and we derive all the quantities that are required in order to implement a solver for the graph Ginzburg–Landau functional. Our implementation and all examples are available on a GitHub-Repository.

We set our domain $D = [0,1]^2$and assume that the data points $\{x_1, \dots, x_n\} \subset D$ are sampled from a probability distribution ν. The function $u : D \to \mathbb{R}$ is represented as a vector $u \in \mathbb{R}^n$ with Laplacian matrix $L \in \mathbb{R}^{n\times n}$ and degree vector $d \in \mathbb{R}^n$. This enables us to rewrite the Graph Ginzburg-Landau functional as

$$\begin{aligned} G_n(u) =& \frac{1}{n^2\varepsilon^{d+1}} \sum_{i,j=1}^{n} w_{ij}(u_j^r - u_i^r)^2 + \frac{\beta}{n^2\varepsilon} \sum_{i=1}^{n} \deg(i)((u_i^r)^2 - 2(u_i^r)^3 + (u_i^r)^4) \\ =& \frac{2}{n^2\varepsilon^{d+1}} \langle -Lu, u\rangle + \frac{\beta}{n^2\varepsilon} \left\langle d(u)^2, (1-u)^2 \right\rangle. \end{aligned}$$

This reformulation is much more concise and allows for vectorization. We introduced the weight $\beta > 0$ for the double well potential to enable greater control in applications. The balancing term for multi-phase problems is given by

$$B^\alpha(u) = \left(\frac{1}{n}\sum_{i=1}^{n} |u_i|\right)^\alpha.$$

For a single phase, we use $B^\alpha(u) = \min\left(\frac{1}{n}\sum_{i=1}^{n}|u_i|, 1 - \frac{1}{n}\sum_{i=1}^{n}|u_i|\right)^\alpha$ instead. Note that, for u not an indicator function, the balancing term does not provide an accurate volume estimate. Various approaches have been developed to address this issue. For example, one could use a smooth indicator function. Alternatively, one could follow the approach proposed in [4], using an exponent u^q with $q = {}^{2d}/_{d-1}$ that supposedly improves the volume estimate. However, in our experiments, both approaches made the problem harder to solve without providing any value to the optimization.

The Cheeger-Cut is still defined as

$$\mathcal{G}_n(u) = \frac{G_n(u)}{B^\alpha(u)}.$$

When using a gradient-based approach, it is sometimes preferable to replace constraints with penalties. In our setting, these are the constraints of volume and orthogonality given by

$$\frac{\zeta}{\varepsilon}\left(m - \frac{1}{n}\sum_{i=1}^{n}|u_i|\right)^2 \quad \text{and} \quad \frac{\gamma}{\varepsilon n}\sum_{i=1}^{n}\sum_{r=1}^{R}\sum_{s=1,s\neq r}^{R}\left(u_i^r u_i^s\right)^2,$$

where $m \in [0,1]$ is the target volume for an isoperimetric problem. Although the order of summation of the orthogonality constraint can be interchanged, putting the summation over N outside allows the other operations to be vectorized.

Now, let us calculate the gradients. The gradient of G_n at vertex i is

$$\begin{aligned}
\nabla G_n(u)_i =& -\frac{2\varepsilon}{n^2}\sum_{j\in V} w_{ij}(u_j - u_i) + \frac{2\varepsilon}{n^2}\sum_{j\in V} w_{ji}(u_i - u_j) \\
&+ \frac{2}{n^2\varepsilon}\deg(i)(u_i - 3u_i^2 + 2u_i^3) \\
=& \frac{2\varepsilon}{n^2}\sum_{j\in V} 2w_{ij}u_i - \frac{2\varepsilon}{n^2}\sum_{j\in V} w_{ij}u_j + \frac{2}{n^2\varepsilon}\deg(i)(u_i - 3u_i^2 + 2u_i^3) \\
=& \frac{4\varepsilon}{n^2}Du_i - 4\varepsilon W u_i + \frac{2}{n^2\varepsilon}\deg(i)(u_i - 3u_i^2 + 2u_i^3) + \\
=& -\frac{4\varepsilon}{n^2}Lu_i + \frac{2}{n^2\varepsilon}\deg(i)(u_i - 3u_i^2 + 2u_i^3).
\end{aligned}$$

Note that we have used the symmetry of the weight function in the first equality. This allows us to obtain a vector for the Ginzburg–Landau gradient

$$\nabla G_n(u) = -\frac{4}{n^2\varepsilon^{d+1}}Lu + \frac{2\beta}{\varepsilon n^2}d(u - 3u^2 + 2u^3).$$

The derivative of the balance term for $u_i > 0$ is

$$\frac{\mathrm{d}}{\mathrm{d}u_i}B^\alpha(u) = \frac{\mathrm{d}}{\mathrm{d}u_i}\left(\frac{1}{n}\sum_i u_i\right)^\alpha = \frac{\alpha}{n}B^{\alpha-1}(u).$$

Using the quotient rule, the complete gradient of the cheeger cut functional is given as

$$\nabla\mathcal{G}_n(u)_i = \frac{\frac{\mathrm{d}}{\mathrm{d}u_i}G_n(u)B^\alpha(u) - \frac{\alpha}{n}B^{\alpha-1}(u)GL(u)}{B^{2\alpha}(u)}.$$

Note that the following equalities hold for exponents: $\frac{|A|^\alpha}{|A|^{2\alpha}} = |A|^{-\alpha}$ and $\frac{|A|^{\alpha-1}}{|A|^{2\alpha}} = |A|^{-(\alpha+1)}$. We can write the gradient more compactly and computational efficiently to be

$$\nabla\mathcal{G}_n(u) = \left[\nabla G_n(u) - \frac{\alpha}{n}\frac{G_n(u)}{B^1(u)}\right]\frac{1}{B^\alpha(u)}.$$

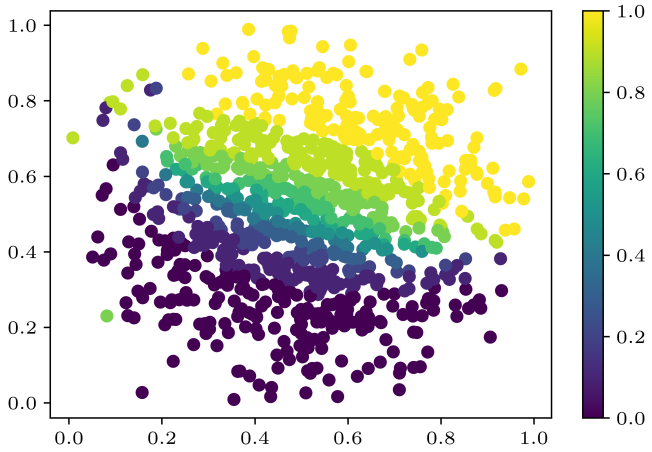


Figure 7.1: Solution of truncated Gaussian Isoperimetric Problem in two dimensions.

The gradients of constrains as penalties with respect to u_i^r are

$$\frac{2\zeta}{\varepsilon}\left(-m+\frac{1}{n}\sum_{i=1}^{n}|u_i^r|\right) \quad \text{and} \quad \frac{2\gamma}{\varepsilon n}\sum_{i=1}^{n}\sum_{r=1}^{R}\sum_{s=1,s\neq r}^{R} u_i^r(u_i^s)^2.$$

7.2 Flavours of Graphs Cuts

A key advantage of graph-based methods is their ability to handle various weighted measures without requiring explicit knowledge of the density function, a feature, which is particularly valuable in machine learning. One example of this is the Gaussian isoperimetric problem whichs solutions often manifest as half-spaces [5]. Although the Gaussian probability function is not compactly supported on $D = [0,1]^2$ the graph-based approaches necessitate that all points lie within the kernel's support. As any isolated outliers would be disconnected from the rest of the graph and it can be removed without altering the graph or the minimizer. Therefore, we resample any points falling outside of D. The solution, shown in Figure 7.1, shows a phase-field function with a distinct, relatively thick separating region whose width depends on the parameter ε.

To explore the impact of the parameter α in the balancing term, let us examine the behaviour of the solutions for different values of α. The points $\{x_1, \ldots, x_n\}$ are sampled uniformly at random within the domain

$$D = D_l \cup D_s,$$

where $D_l = [0,1]\times([0,0.3]\cup(0.4,0.6]\cup(0.7,1])$ and $D_s = [0.4,0.6]\times((0.3,0.4]\cup(0.6,0.7])$. When $\alpha = 1$, the standard Cheeger-cut is recovered. Figure 7.2 on the left, shows that,

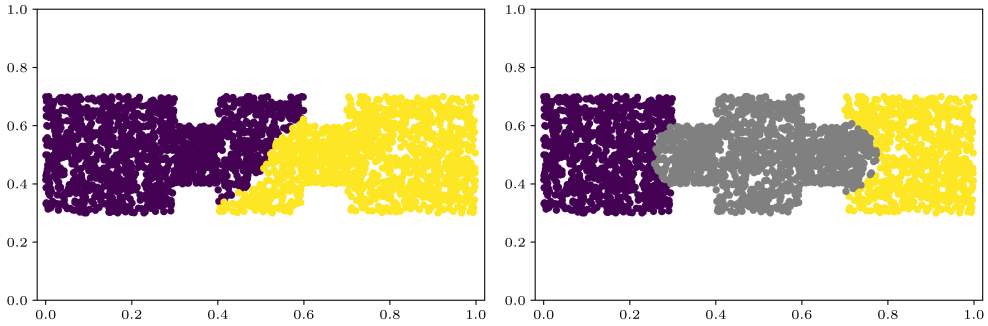


Figure 7.2: Solution to the two set Cheeger cut problem (with threshold 0.5). Left: $\alpha = 1$. Right: $\alpha = 0.5$.

at this value of α the solution exhibits minimal perimeter between tthe two phases at the centre of the domain, and the two volumes are very similar. Note that we use two phase-field functions instead of one to allow for empty regions. For the initialization of u we sample uniform random values in $[0, 1]$. To increase comparability and concentrate the sets at the sides of the domain, we initialized points close to the boundary with values uniformly at random in $[0.7, 1]$. This approach ensures more consistent solutions, as purely random initialization might lead to unpredictable regions of concentration. The same could, of course, also be observed with a random initialization, but it is not as replicable. Figure 7.2 on the right, does clearly shows a very different picture for $\alpha = 0.5$ as cut a core occurs in a much thinner region of the domain. This behaviour is consistent with importance of the perimeter increasing dramatically while the volume of the set plays a secondary role.

On the other hand for $\alpha > 1$ it is much harder to see the difference in behaviour because, for most partition tasks, the domain will already be covered. Consider an example from section 7.5: two clusters sampled from a Gaussian distribution with small variance, supplemented by random noise. The clusters are initialized with uniform random values, while the outliers are assigned the neutral values 0.5. See section 7.5 for details. As can be seen in Figure 7.8, the left image with $\alpha = 1$ shows a shorter cut, while the phase field covers all the available space in the right image with $\alpha = 5$. It is also worth noting, that the problem for $\alpha = 5$ was initialized with the optimal cut for $\alpha = 1$. For large values of α and random initialization, the solvers seem to overemphasize the importance of the balancing term, resulting in progress being locked into a local minimum quite soon. Conversely, starting with $\alpha < 1$ work quite well; however, using a solution to the problem with $\alpha = 1$ as initial guess, tends not to give up any volume, contrary to the example above. This behaviour is somewhat consistent. We would expect a solution to the Cheeger functional to always go for the maximum available volume, unless a much shorter perimeter solution is available. However, the value of α enables us to exert some control over this process.

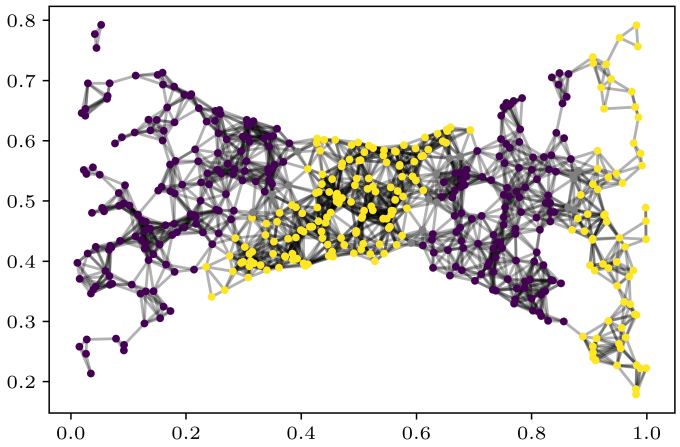


Figure 7.3: Optimal cut of Graph with 500 vertices and kernel support below connectivity threshold.

7.3 Parameter Choices

It is quite important that the ε used to create the graph satisfies the connectedness assumption. Even when relying on the connectedness result, the constant in Proposition 2.1 has to be chosen appropriately. In Figure 7.3 shows $n = 500$ data points sampled uniformly within the hourglass-shaped domain. The graph is constructed using $\varepsilon = c\frac{\log(n)^{3/4}}{n^{1/2}}$ and $c = 0.31$, so $\varepsilon \approx 0.05$. This results in a rather patchy network, where some vertices are connected by only one edge, or are even completely disconnected. The displayed solution is optimal in the sense that it provides an actual minimum cut given the initialization. However, it is undesirable due to a lack of consistency with the continuum partition of the domain. Compare this first result with the second plot in Figure 7.4. Here, the optimal graph cut is much closer to the expected value, even with relatively few points. This is because a more densely connected graph was constructed using the same set of points, albeit for $c = 0.8$, i.e. $\varepsilon \approx 0.14$. The plot on the left of Figure 7.4 shows that the domain is approximated much more closely.

One more thing that we want to mention is the importance of choosing an appropriate ε_n for the Ginzburg-Landau functional. If ε_n is chosen too small, the behaviour becomes similarly patchy, as seen in Figure 7.3, because the impact of the Laplacian term is too localized. A value two to three times grater than the graph-ε, sometimes even four times, tends to provide solid solutions. Conversely, an ε_n that is too large results in very broad borders that are often sub-optimal. Thresholding also becomes much less reliable, since the boarders of the phases may overlap considerably.

As we mentioned earlier, different values of α have an impact on the graph cut solution, with large values hindering numerical reachability. It is also worth noting that using large values of β, the weight of the double well potential or large values of γ for the separation constraint, can stiffen the problem and make it very difficult to solve.

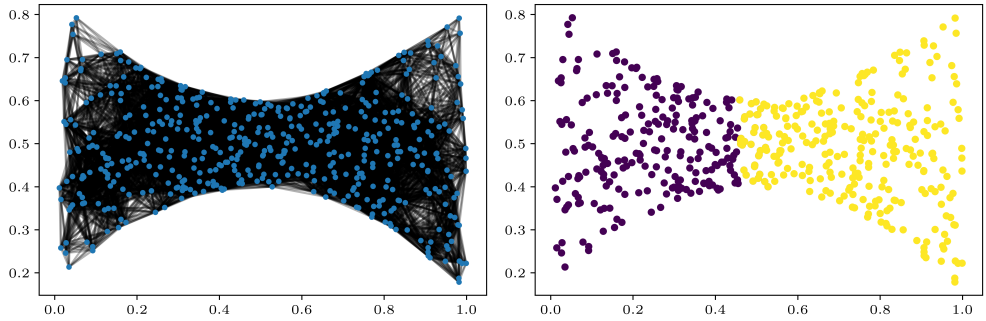


Figure 7.4: Left: a densely connected Graph that represents the domain well. Right: Optimal grap-cut of the domain.

Conversely, using very small values allows the Laplacian to dominate, which of course favours a constant solution.

Lastly, it is important to note that the choice of the initial guess is very important. The more points there are in the problem and the harder it is to find an optimal cut, the more the solution depends on the initial guess, as our problem is highly non-linear. In our experience, good solutions can be achieved with a completely random starting point. However, this method is far less consistent and convergence is not guaranteed. Often, providing just a small additional hint, such as concentrating the initial values of one phase in a specific area, can greatly improve the convergence and efficiency. In Figure 7.6 we present two example, where phase u^r was initialized uniformly at random with values in $[0, 1]$ and then the values within a small circle around a suitable m_r were updated to lie within $[0.7, 1]$. We would like to point out that, while random initial guesses are not a particularly reasonable assumption for most problems, providing this type of additional input can be considered a very weak form of semi-supervised learning. We did not provide a rigorous theoretical justification for this approach.

7.4 Approximation of the Continuum Limit

As the consistency with the continuum Cheeger cut is the most important property that we demonstrated, the following Figure 7.5 illustrates a sequence of discrete partitions, computed from the graph-based Cheeger cut problem, that converge to the optimal continuum Cheeger cut. As the number of points increases, we can see the graph-based Cheeger cut compared with the optimum solution at 0.5. For every experiment, the last n_{old} samples are retained and $m = n_{new} - n_{old}$ samples are drawn uniformly in $D = [0.35, 0.65] \times [0, 1]$. The solution of the last problem is used as initialization for the existing points, while the new ones are initialized randomly.

In the example shown here, we use a threshold of 0.5 for plotting. The shape is quite characteristic; generally, we observe a relatively straight cut that tends to be slightly

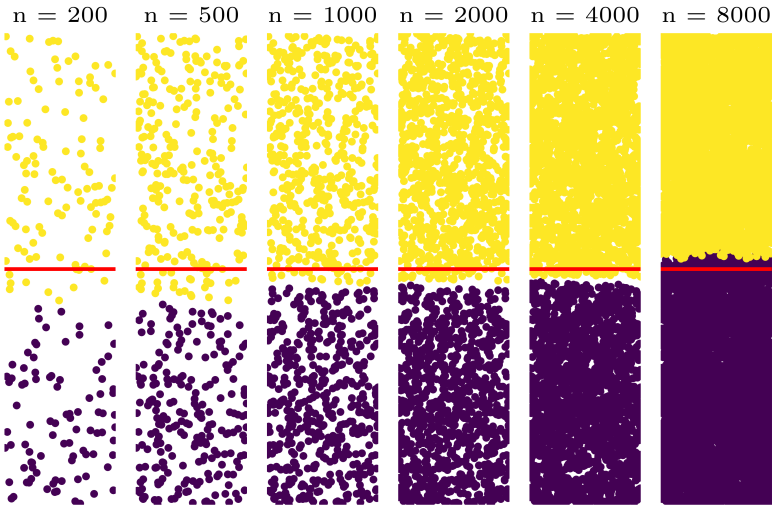


Figure 7.5: Convergence behaviour of the Graph Cheeger cut.

off-centre. The closer the solution moves to the continuum optimum, the more points we use. However, we did not observe the type of behaviour described in [14, Figure 2], where the shape of the graph cut varied significantly.

The paper [4] inspired the approach described here. In the context of optimal packing, they used the α-cheeger cut as initial guess for another algorithm. Therefore, we would like to compare the solutions we obtained on graphs with the continuum Cheeger cuts. Hence, we implemented a continuum solver using a finite difference approach. A notable difference between the two implementations is that the finite difference approach uses zero boundary conditions, whereas the graph has no boundary conditions.

To formulate the classical functional $GL_\varepsilon(u)$ (6.1) on a two-dimensional grid for a function $u : D_1 \times D_2 \to \mathbb{R}$, we us forward finite differences with grid size h_1 in x_1 and h_2 in x_2. Then, the gradient of u is given as

$$(\nabla u)_{ij} = \begin{pmatrix} \partial_j^+ u_{ij} \\ \partial_j^+ u_{ij} \end{pmatrix} = \begin{pmatrix} \frac{u_{ij+1}-u_{ij}}{h_1} \\ \frac{u_{i+1j}-u_{ij}}{h_2} \end{pmatrix}.$$

Therefore, $\mid \nabla u_{ij} \mid^2 = \left(\frac{u_{ij+1}-u_{ij}}{h_1}\right)^2 + \left(\frac{u_{i+1j}-u_{ij}}{h_2}\right)^2$ and we obtain a new functional on $N = n_i * n_j$ grid points, with $n_j = \frac{|D_1|}{h_1}, n_i = \frac{|D_2|}{h_2}$ as

$$GL(u) = \frac{\varepsilon}{N} \sum_{k=1}^{N} \left(\left(\frac{u_{k+1} - u_k}{h_1}\right)^2 + \left(\frac{u_{k+n_j} - u_k}{h_2}\right)^2 \right) + \frac{\beta}{\varepsilon N} \sum_{k=1}^{N} u_k^2 - 2u_k^3 + u_k^4.$$

The grid based balancing term in the multi phase case is simply given as $B^\alpha(u) =$

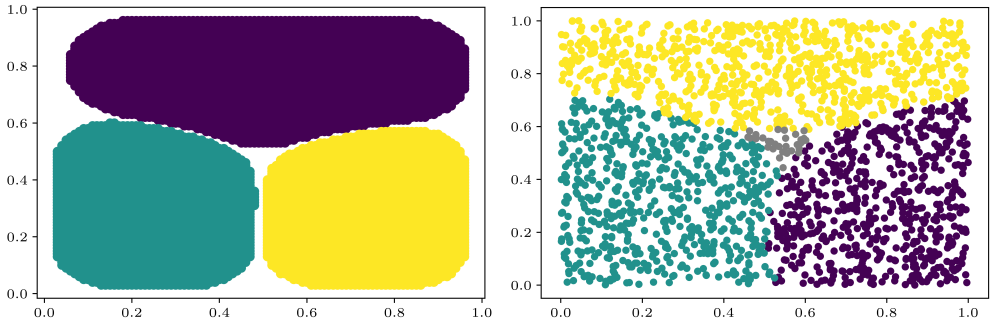


Figure 7.6: Comparison of a Graph partition to a continuum one.

$\left(\frac{1}{N}\sum_{k=1}^{N}|u_k|\right)^{\alpha}$. As before the Cheeger cut functional is

$$G(u) = \frac{GL(u)}{B^{\alpha}(u)}.$$

Let us calculate the gradient of $GL(u)$ with respect to discretization u_{ij}. Note u_{ij} also occurs in the role of $u_{ij+1}$and u_{i+1j} in the next summand, so we obtain

$$\begin{aligned}(\nabla GL(u))_{ij} =& -2\varepsilon\left(\left(\frac{u_{ij+1}-u_{ij}}{h_1^2}\right)+\left(\frac{u_{i+1j}-u_{ij}}{h_2^2}\right)\right)\\ &+2\varepsilon\left(\left(\frac{u_{ij}-u_{ij-1}}{h_1^2}\right)+\left(\frac{u_{ij}-u_{i-1j}}{h_2^2}\right)\right)\\ &+\frac{2}{\varepsilon}(u_{ij}-3u_{ij}^2+2u_{ij}^3)\\ =&2\varepsilon\left(\left(\frac{-u_{ij+1}+2u_{ij}-u_{ij-1}}{h_1^2}\right)+\left(\frac{-u_{i+1j}+2u_{ij}-u_{i-1j}}{h_2^2}\right)\right)\\ &+\frac{2}{\varepsilon}(u_{ij}-3u_{ij}^2+2u_{ij}^3),\end{aligned}$$

for $2 \leq i \leq n_i - 1$ and $2 \leq j \leq n_j - 1$. For the edges of grid set $u_{edge\pm1} = 0$. Note that the discrete version of the Laplace operator $\Delta u = \mathrm{div}\nabla u$ is exactly first expression.

In Figure 7.6 presents the continuum partition of the domain $D = [0,1]^2$ for three phases with $\alpha = 2$. For plotting the phases we used a threshold of 0.5. The white regions do not belong to any of the phases. Even though it may not look like it, all three phases have almost identical volume. On the left hand side of Figure 7.6 we illustrate the graph Cheeger cut for three phases with $n = 2000$ data points in an identical setting to the continuum example.

7.5 Hard Clustering Tasks

In this section, we will examine some common yet challenging clustering tasks involving outliers.

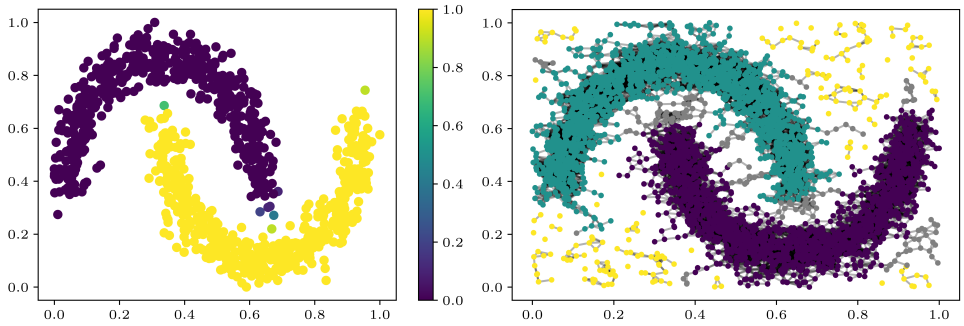


Figure 7.7: Example of a clustered Two Moons Dataset. Left: Cheeger cut solution to the classic problem. Right: With outliers and additional phase.

We begin with the Two Moons problem, see Figure 7.7, a classic benchmark often used to demonstrate the ability to perform non-linear classification. The domain is $D = [0, 1]^2$. For this dataset, the graph e has to be quite small so as not to bridge the gap between the tip of one moon and the middle of the other.

Clustering with outliers is one of the hardest tasks. Ideally, we would like all outliers to remain unclassified at the end of the optimization process. The idea is to leave some empty regions of low density by selecting the correct value of the parameter a that we introduced. As mentioned previously, initialization is crucial for certain problems, particularly the more challenging ones. Our examples include a significant proportion of outliers. As these are usually expected to be noise, distributed independently from the data points, they tend to be in much less dense regions — often building little islands by themselves. Using random initial values will lead to an equally random classification of the outliers. We introduce a simple heuristic for initialization that is similar to the technique used in DBScan. Note that a local quantity which makes an assertion about the density of a vertex's neighbourhood is its degree. We need to calculate this for the Laplacian regardless.

To illustrate this, imagine there are R classes that we want to cluster in a noisy sample. Given that outliers tend to lie in sparse regions and therefore have a smaller degree, we can define the set of potential outliers as follows

$$O = \{i \in V : \deg(i) \leq \deg_{min} + \frac{(\deg_{mean} - \deg_{min})}{3}\}.$$

The choice of lower bound is just an estimate based on our experience. We set $u^r(i) = 0.5 \ \forall i \in O$ and $r = 1, \ldots, R$, which is the maximum of the double well. The idea is that nodes close enough to a cluster will be pulled in by their neighbours, enabling them to be classified correctly, since the maximum is an unstable null point of the gradient. While this does provide a good solution, it is not robust, as can be seen in Figure 7.8. Crucially, for larger values of a, the behaviour is as expected: less emphasis on perimeter and therefore maximization of volume.

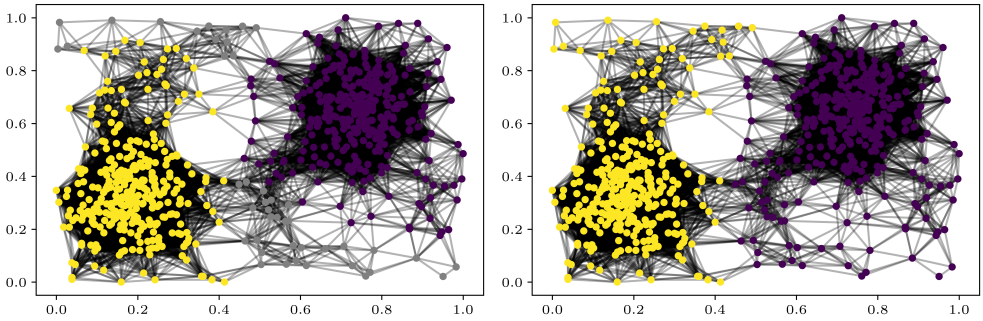


Figure 7.8: Solutions for two clusters with uniform noise. Left: $\alpha = 1$, right: $\alpha = 5$.

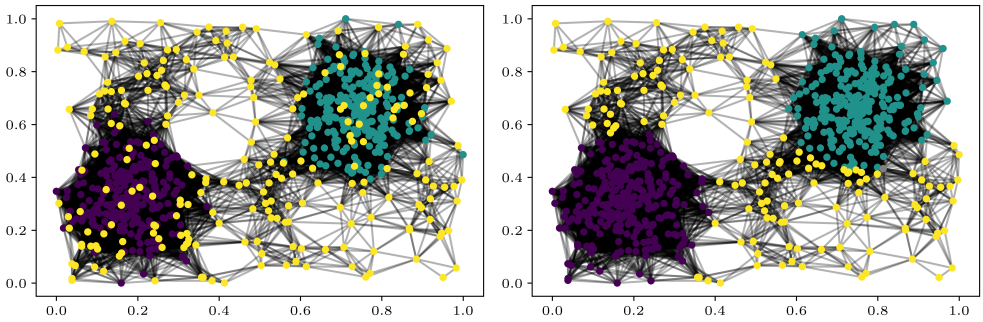


Figure 7.9: Solutions to two clusters with uniform noise using three phase fields. Left: Labeled distribution. Right: Solution for $\alpha = 5$.

To stabilize the procedure, we introduce an additional phase field for outliers only. We initialize it as $u^{R+1}(i) = 1$ if $i \in O$ and zero otherwise. Direct comparison with the labelled set of points in Figure 7.9 shows very convincing performance. Importantly, the larger values of α enforce the same set size for all phases, which is favourable in this case. However, it can be seen in the right cluster that, if not balanced correctly, the outlier phase can encroach on the regular ones. In general, using an additional phase tends to be more stable and can significantly improve performance when a large proportion of outliers is expected.

Lastly, be apply the heuristic introduced above to the two moons problem with outliers. In Figure 7.7 on the right hand sight we demonstrate how this very hard clustering problem can be solved with our approach. The problem consists of 4000 data points of which 20% are random uniform samples, i.e. noise. In the present of the noise it is advisable to use even smaller values for ε-graph as not to bridge the gap mentioned before. We solved this solution for three phases with the initialization discussed above. A value of $\alpha = 3$ was used to enforce a larger portion of outlier classification.

8 Conclusion and Discussion

In this thesis, we established the convergence of the α-Cheeger cut, thereby providing a rigorous basis for its use in graph-based applications. We demonstrated Γ-convergence to the continuum Total Variation and the compactness property of the graph total variation with the α-balance term. This enabled us to also conclude the convergence of minimizers. Furthermore, we introduced the graph Ginzburg-Landau functional and defined its corresponding non-local and local continuum versions. Using a classical approach similar to that employed by Modica in his original work, we proved the Γ-convergence for two phases to the continuum total variation, weighted by appropriate constant factors. This included both the convergence of the non-local to the local functional and the convergence of the graph to the local functional.

On the one hand, this approach offered a clear methodology, and our convergence goal was concrete. However, it required the introduction of a degree term to recover the squared density in the limit. We only considered isotropic kernels; for certain applications, it could be beneficial to generalise this approach. In this case, an approach similar to that used by Alberti and Bellatini or Christoferi and Thorpe might be more helpful.

As far as we are aware, the generalization of the Graph Ginzburg-Landau functional to several phase fields, is a novelty. We also included a compactness result, though this relied on the work of Christoferi and Thorpe. Ideally, we would have developed an independent result consistent with the methodology of Chapter 5. Furthermore, we did not investigate rates of convergence at all, but learning bounds might be of interest in this context.

With regard to the numerical applications, we provided a comprehensive set of examples and ensured that our methods were fully traceable. This included detailed observations on general behaviour, parameter selection and the influence of different settings. We explored the impact of varying the balancing term parameter α and thereby validated theoretical expectations regarding the interaction between volume and perimeter. However, we did not investigate the limiting behaviour for $\alpha \to \infty$ or towards a lower bound, neither in a theoretical or in a experimental setting. We presented comparisons between graph partitions and continuum results. Firstly, we visualized the behaviour of graph cuts as the number of points increased towards the expected continuum results. Secondly, we demonstrated that graph partitions are indeed quite similar to the continuum, even with a comparatively small number of points, as in our experiments.

We were able to introduce a reliable method to cluster data clouds while filtering out low density outliers. Although this approach requires minimal additional knowledge, its theoretical foundation was not provided. As the underlying theory has not yet been explored, but this could be an interesting subject for future research.

While our numerical results were generally robust and consistent, there were some challenges for single vertices of low degree, which were not always classified properly. Investigating normalized cuts and KNN graphs might help to might to address this issue. While it is theoretically feasible and probably not too difficult to implement, we did not provide any numerical results for tasks that are very common in the context of graph applications. One natural domain of application would be image segmentation, where the pixels of an image are embedded into $\mathbb{R}^5$, two dimensions for position and three for colour.

Acknowledgments

I would like to thank Prof. Dr. Leon Bungert for supervising this thesis, especially his keen advice and invaluable feedback were greatly appreciated. I am very grateful to my co-supervisor Dr. Eloi Martinet for his insightful feedback, helpful suggestions and overall patience that enabled the completion of this work. I would like to thank Patrick Schurk for some light proofreading and feedback, but more than anything for the countless hours working side by side in the libraries of Würzburg. His encouragement and distraction on the long working days were invaluable. Furthermore, I want to thank all my other friends and family whose love, patience, and encouragement helped keep me going during challenging times.

Versicherung zur selbständigen Leistungserbringung

Julius Maximilians
UNIVERSITÄT
WÜRZBURG

Versicherung zur selbständigen Leistungserbringung

Titel der Masterarbeit:
Variational Approach to Graph Clustering

Thema bereitgestellt von (Titel, Vorname, Nachname, Lehrstuhl):
Prof. Dr. Leon Bunger, Lehrstuhl für Mathematik des Maschinellen Lernens

Eingereicht durch (Vorname, Nachname, Matrikel):
Nick Burk, 3023232

Ich versichere, dass ich die vorstehende schriftliche Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die benutzte Literatur sowie sonstige Hilfsquellen sind vollständig angegeben. Wörtlich oder dem Sinne nach dem Schrifttum oder dem Internet entnommene Stellen sind unter Angabe der Quelle kenntlich gemacht.

Weitere Personen waren an der geistigen Leistung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines Ghostwriters oder einer Ghostwriting-Agentur in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar Geld oder geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Arbeit stehen.

☐ Mit dem Prüfungsleiter bzw. der Prüfungsleiterin wurde abgestimmt, dass für die Erstellung der vorgelegten schriftlichen Arbeit Chatbots (insbesondere ChatGPT) bzw. allgemein solche Programme, die anstelle meiner Person die Aufgabenstellung der Prüfung bzw. Teile derselben bearbeiten könnten, entsprechend den Vorgaben der Prüfungsleiterin bzw. des Prüfungsleiters eingesetzt wurden. Die mittels Chatbots erstellten Passagen sind als solche gekennzeichnet.

Der Durchführung einer elektronischen Plagiatsprüfung stimme ich hiermit zu. Die eingereichte elektronische Fassung der Arbeit ist vollständig. Mir ist bewusst, dass nachträgliche Ergänzungen ausgeschlossen sind.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht. Ich bin mir bewusst, dass eine unwahre Erklärung zur Versicherung der selbstständigen Leistungserbringung rechtliche Folgen haben kann.

Köngen, den 29.12.2025 N. Burk
Ort, Datum, Unterschrift

Ref. 2.3 / Stand: 24.11.2023

Bibliography

[1] Giovanni Alberti and Giovanni Bellettini. A non-local anisotropic model for phase transitions: Asymptotic behaviour of rescaled energies. *European Journal of Applied Mathematics*, 9(3):261–284, June 1998.

[2] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of Bounded Variation and Free Discontinuity*. Oxford Mathematical Monographs. Oxford University Press, 1st edition, 2000.

[3] Annalisa Baldi. Weighted BV Functions. *Huston Journal of Mathematics*, 27(3):683–705, January 2001.

[4] Beniamin Bogosel, Dorin Bucur, and Ilaria Fragalà. Phase Field Approach to Optimal Packing Problems and Related Cheeger Clusters. *Applied Mathematics & Optimization*, 81(1):63–87, February 2020.

[5] Christer Borell. The Brunn-Minkowski inequality in Gauss space. *Inventiones Mathematicae*, 30(2):207–216, June 1975.

[6] Andrea Braides. Γ-*Convergence for Beginners*. Oxford Lecture Series in Mathematics and Its Applications. Oxford University Press, 1st edition, 2002.

[7] Leon Bungert. PDEs on Graphs. 2024. Lecture notes, Winter semester 2024/25.

[8] John W Cahn and John E Hilliard. Free Energy of a Nonuniform System. I. Interfacial Free Energy. 2016.

[9] R. Cristoferi and M. Thorpe. Large data limit for a phase transition model with the p -Laplacian on point clouds. *European Journal of Applied Mathematics*, 31(2):185–231, April 2020.

[10] Gianni Dal Maso. *An Introduction to* Γ-*Convergence*. Birkhäuser Boston, 1993.

[11] Lawrence Craig Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions, Revised Edition*. Chapman and Hall/CRC, revised edition, April 2015.

[12] Nicolás García Trillos and Dejan Slepčev. On the Rate of Convergence of Empirical Measures in ∞-transportation Distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, December 2015.

[13] Nicolás García Trillos and Dejan Slepčev. Continuum limit of total variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220(1):193–241, April 2016.

[14] Nicolás García Trillos, Dejan Slepčev, James van Brecht, Thomas Laurant, and Xavier Bresson. Consistency of Cheeger and Ratio Graph Cuts. *The Journal of Machine Learning Research*, 17(1):1–46, October 2016.

[15] Giovanni Leoni. *A First Course in Sobolev Spaces.* Number 181 in Graduate Studies in Mathematics. American Mathematical Society, 2nd edition, 2017.

[16] Michele Miranda. Functions of bounded variation on "good" metric spaces. *Journal de Mathématiques Pures et Appliquées*, 82(8):975–1004, August 2003.

[17] Luciano Modica. The gradient theory of phase transitions and the minimal interface criterion. *Archive for Rational Mechanics and Analysis*, 98(2):123–142, June 1987.

[18] Yves Van Gennip, Nestor Guillen, Braxton Osting, and Andrea L. Bertozzi. Mean Curvature, Threshold Dynamics, and Phase Field Theory on Finite Graphs. *Milan Journal of Mathematics*, 82(1):3–65, June 2014.

[19] Hermann Weyl. On the volume of tubes. *American Journal of Mathematics*, 61(2):461–472, 1939.