Rend. Sem. Mat. Univ. Pol. Torino
Vol. xx, x (xxxx), 1 – ??

R. Cristoferi*

CLUSTERING OF BIG DATA: CONSISTENCY OF A NONLOCAL GINZBURG-LANDAU TYPE MODEL

Abstract. After introducing the main ideas and reviewing some of the literature on the subject, we consider a discrete non-local variational model for clustering in the context of soft-classification semi-supervised learning. The functional is inspired by a similar model studied by Alberti and Bellettini (see [1]) in the context of phase transition for ferromagnetic materials. A parameter ε_n controls both the non-local term, as well as the size of the phase transition layer. We identify the Γ-limit of the variational functional as $\varepsilon_n \to 0$. In the machine learning community, this is known as the study of the *consistency* of the model. The limiting functional is given by a fidelity term plus weighted anisotropic perimeter.

1. Introduction

This paper is based on a seminar given by the author at the *Analysis and applications: contribution from young researchers* workshop that took place on 8,9 April 2019 in Politecnico di Torino. The work presented here is part of an ongoing project in collaboration with Matthew Thorpe (see [24]).

In the modern era, people are producing a tremendous amount of data. Every second there are 5 new Facebook profiles created, over 400 thousand tweets are sent on Tweeter every minute, and more that 300 million of new pictures appear every day on Facebook (see [42]). A first operation one has to do in order to make some use of these collection of data, is to partition them in classes according to some notion of similarity. Of course, this notion of similarity depends on the particular nature of the data, and on the tasks that will be subsequently have to be performed on them. In practice, what we have to do is to assign a label to each element of the data set, where each label represent a different class. There are three scenarios: when no information is known on the labeling of the data set we talk about *unsupervised* learning, and of *semi-supervised* learning when some labels are already know. Here we focus on the latter. The amount of known labels we have to propagate to the whole data set affects the methods that need to be implemented to reach the goal. The methods we are going to consider seem to perform better, at least from a numerical point of view, than other methods used in machine learning.

Among several methods proposed by many authors over the years, the ones falling into the category of *variational methods* have been very successful. The general idea is that the partition of the set of data in classes should satisfy some opti-

*Department of Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom, r.cristoferi@hw.ac.uk

mality condition. Thus, behind every variational method there is a functional that has to be minimized. This functional will be defined on the set of possible labeling of the data set, and the partition of the data set will be chosen among the labeling that minimize the functional (in general, no uniqueness of the minimizer is guaranteed). Examples of such a variational methods for labeling that have been successfully implemented include minimizing graph cuts and total variation (see, for instance, [6, 11, 15, 17, 16, 18, 43, 51, 54, 55, 53]).

Usually, each model posses some parameters that practitioners can vary in order to obtain better results for each particular case where the method is applied (we note that there are very few theoretical results connecting the role of the parameters to the features of the partition one obtains by using that model). Some of the parameters depend on the number of elements of the data set. Of capital importance for evaluating a labeling method is whether the method *consistent* or not; namely it is desirable that the minimization procedure approaches some limit minimization method when the number of elements of the data set goes to infinity. Due to the extremely large amount of data usually considered in applications, this limiting operation is not just a merely mathematical study. Moreover, knowing whether a specific minimization strategy is an approximation of a limit (minimizing) object can help explain properties of the finite data method. In particular, this can also be used to justify, a posteriori, the use of a certain procedure in order to obtain some desired features of the classification. Furthermore, understanding the large data limits can open up new algorithms.

1.1. The structure of the data set

Assume that each element of the data set can be encoded into a point of a vector space $\mathbb{R}^d$. For instance, a point can represent a pixel of an image, or the whole image. The construction of the points in the data set (sometimes called *point cloud*) is something we are not interested in, but we assume we are given. A data set would then be a collection of points $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$. We refer to the example by Jeff Calder described in Figure 1 as a justification for the need of the additional structure we will endow the data set with.

In order to take into consideration the geometry of the data set, and to being able to decide the relevance each element of the data set has with respect to the others, we construct a graph over the data set X_n. The idea is the following: since we are using a variational method, the underlining functional we have to minimize should penalize a lot the case where elements of the data set we want to be in the same class, are given different labels. Thus, the weight on each edge of the graph should represent the degree of similarity of the two elements of the data set connected by the vertex. Again, we are not interested in the specific way such a graph is constructed, but we briefly describe the general method used to define the weights. Let x_i, and x_j be two points on the data set. The weight W^{ij} between them will depend on the points themselves, on features we are interested in, and on how much we want to penalize the difference in these features. From the mathematical point of view, we need a *feature extraction* function

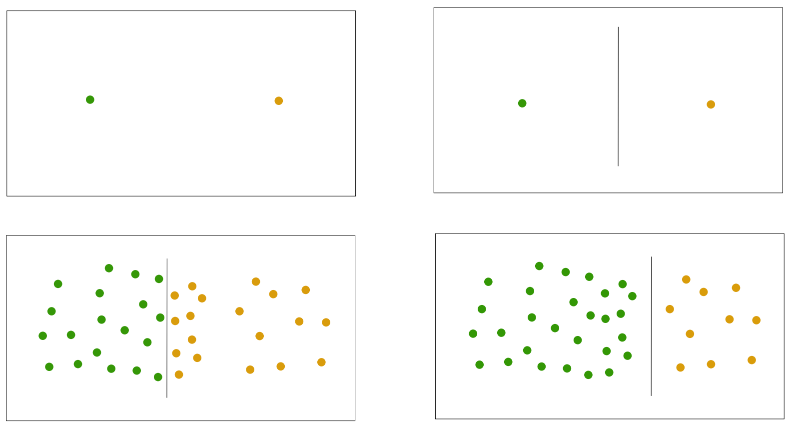

Figure 1: Assume we are only given labels on two points (top-left figure). Without any knowledge on the data set, the best we can do is assign to each point the same label to the one of the two pre-labeled points it is close to (top-right figure). Of course, for certain geometry of the data set, this procedure might give a partition of the data set (bottom-left figure) that is different from the expected one (bottom-right figure).

$\pi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, and a *feature penalization* function $\varphi : \mathbb{R} \to [0, +\infty)$. In this paper we assume the feature extraction function π to be of the form

$$\pi(x_i, x_j) = \pi(x_i - x_j),$$

and such that $\operatorname{supp}\pi \subset B_R(0)$, for some $R > 0$. Note that the feature extraction function is anisotropic. The weight W^{ij} between x_i and x_j is then defined as

$$W^{ij} := \varphi(\pi(x_i - x_j)) =: \eta(x_i - x_j).$$

Note that $W^{ij} = 0$ means that we do not care about whether x_i and x_j have different labels or not.

1.2. The discrete model

The discrete functional we consider will be introduced one term at a time, in order to justify all the choices we make.

A first observation is the following. Since we are going to consider what happens when the number of points n of the data set goes to infinity, we expect the region around each point in the data set to be more crowd as we increase n. This implies that the number of edges each point in data set is connected to blows up with n. There are two reasons we do not want this to happen: first of all, the problem would be unfeasible to treat from a numerical point of view (since sparse graphs are usually preferred).

Figure 2: As the points of the data set get closer and closer to each other we want to look at the local geometry of the data set.

Moreover, we want to localize the geometry in order for the partition to care about the local structure of the data set (see the example in Figure 2). For these reasons, we rescale the weights W^{ij} as

$$W_n^{ij} := \eta_{\varepsilon_n}(x_i - x_j) := \frac{1}{\varepsilon_n^d}\eta\left(\frac{x_i - x_j}{\varepsilon_n}\right),$$

where $\varepsilon_n \to 0$ as $n \to \infty$. Note that now $W_n^{ij} = 0$ is $|x_i - x_j| > \varepsilon_n R$. The reader should also note that, from an analytical point of view, η_n will play the role of a family of mollifier. Hence, the first term of the functional would be

$$\frac{1}{\varepsilon_n n^2}\sum_{i,j=1}^{n} W_n^{ij}|u(x_i) - u(x_j)|^p,$$

where $p \geq 1$. This term plays the role of penalising oscillations: intuitively one wants a labeling solution such that if x_i and x_j are close on the graph then the labels are also close. Moreover, the higher the weight W_n^{ij} is, the more we penalise a labeling that put x_i and x_j in different classes. When $p = 2$, can also be written as $\frac{1}{\varepsilon_n n}\langle u, Lu\rangle_{\mu_n}$ where L is the graph Laplacian [52]. Finally, the rescaling $\frac{1}{\varepsilon_n}$ has been chosen in order to get a meaningful limiting functional.

Hereafter, for simplicity, we will consider only two classes, that we will denote by ± 1. All the stated results hold true, mutatis mutandi, also in the case more classes are considered. From the numerical point of view, the *hard* classification problem is difficult to treat. For this reason, it is convenient to relax the constrain of the labeling to take value only in the classes ± 1, and allow for a *soft* classification, *i.e.*, for labeling $u : X_n \to \mathbb{R}$. Of course, in the limiting model we would like to recover the hard labeling.

Thus, we introduce a penalization for the labeling not to take value in the classes, as follows:

$$\frac{1}{\varepsilon_n n}\sum_{i=1}^{n} V(u(x_i))\,,$$

where $V:\mathbb{R}\to[0,+\infty)$ is such that $V^{-1}(0)=\{\pm 1\}$.

Finally, instead of asking a labeling u to match the given labels $\overline{u}$, we penalise it for not matching them via the term

$$\lambda\frac{1}{n}\sum_{i=1}^{n}|u(x_i)-\overline{u}(x_i)|^q \mathbb{1}_B(x_i)\,,$$

where $\lambda>0$, $q\geq 1$, and $B\subset\mathbb{R}^d$ is the set where we assume to have given the labels.

Thus, the discrete functional $\mathcal{G}_n : L^1(X_n)\to[0,+\infty]$ reads as

$$\mathcal{G}_n(u):=\frac{1}{\varepsilon_n n^2}\sum_{i,j=1}^{n} W_n^{ij}|u(x_i)-u(x_j)|^p+\frac{1}{\varepsilon_n n}\sum_{i=1}^{n} V(u(x_i))+\lambda\frac{1}{n}\sum_{i=1}^{n}|u(x_i)-\overline{u}(x_i)|^q \mathbb{1}_B(x_i)$$

Here $L^1(X_n)$ is with respect to the empirical measure. Note that, since X_n is finite, every labeling $u:X_n\to\mathbb{R}$ is in the domain of the functional $\mathcal{G}_n$.

We note that the parameters of the discrete model are: the exponent p, the constant λ, and the couple classes and potential V. A change in each of these parameters will reflect in a change of the limiting functional, as well as in a change on the (properties of the) minimal partitions.

Before introducing the continuum model, we want to draw the attention with the similarities between the functional $\mathcal{G}_n$ and models for phase transitions problems. A well studied mathematical model for describing equilibrium configurations of a fluid under isothermal conditions confined in a container $\Omega\subset\mathbb{R}^N$ and having two stable phases (or a mixture of two immiscible and non-interacting fluids with two stable phases) is the following

$$E_\varepsilon(u):=\int_\Omega\left[\frac{1}{\varepsilon}V(u)+\varepsilon|\nabla u|^2\right]\mathrm{d}x\,. \tag{1}$$

Here $u\in H^1(\Omega)$ is the phase variable, and $\varepsilon>0$ is a small parameter related to the thickness of the transition layer between the different phases. The zeros of the potential V corresponds to the stable phases of the liquid. This model has been firstly derived by Van der Waals in his pioneering work [59], and then rediscovered by Cahn and Hilliard in [19]. It was conjectured by Gurtin (see [40]) that for $0<\varepsilon\ll 1$ the minimizer u_ε of the energy E_ε will approximate a piecewise constant function, u, taking values in the zero set of the potential V, and minimizing the surface area of the interface separating the two phases. This conjecture has been proved by Modica and Mortola in [45] (see also [44]) by using the tools of Γ-convergence. Since then, the mathematical literature on the energy E_ε and its variants has blossomed immensely. Here we just recall some of the main classical papers on the subject. The vectorial case has been considered by Kohn and Sternberg in [41], Fonseca and Tartar in [33] and Baldo [7]. A study of the anisotropic case has been carried out by Bouchitté [9] and Owen [46] in the scalar case, and by Barroso and Fonseca [8] and Fonseca and Popovici [32] in the vectorial case. A general case has been considered by Ambrosio in [2], while solid-solid phase transitions, when higher derivatives are considered in the energy have been the focus of the works [21], and [22]. Several variants and extensions have been investigated by, for instance, Savin and Valdinoci [50] and Esedoḡlu and Otto [30]. In particular, nonlocal functionals have been used by Brezis, Bourgain and Mironescu in [10] to characterize Sobolev spaces (see also the work [48] of Ponce). Approximations of (anisotropic) perimeter functionals via energies defined in the discrete setting have been carried out by Braides and Yip in [13] and by Chambolle, Giacomini and Lussardi in [20]. More recently, the interaction between phase transitions and other physical phenomena like homogenization has been the focus of the work of some authors. See, for instance [4, 5, 14, 23, 27, 28].

A *discrete non-local* version of the energy E_ε has been studied, in the context of Ising spin systems on lattices, by Alberti and Bellettini in [1] and [1]. The non-local energy they considered reads as

$$\widetilde{E}_\varepsilon(u) := \frac{1}{\varepsilon^{d+1}} \int_\Omega \int_\Omega J\left(\frac{y-x}{\varepsilon}\right)(u(y)-u(x))^2 \, \mathrm{d}x\,\mathrm{d}y + \frac{1}{\varepsilon}\int_\Omega V(u(x))\,\mathrm{d}x. \tag{2}$$

Here J is a non-negative kernel (ferromagnetic Kac potential) satisfying similar assumptions as η. In their works, they were able to identify the Γ-limit of the family of functionals $\widetilde{E})_\varepsilon$, by introducing fundamental tools needed to study this class of problems. Indeed, different techniques have to be used in order to study the asymptotic behaviours of the functionals (1) and (2).

The functional studied by Alberti and Bellettini is related to the functional we are studying. Indeed, by using the results recalled in Section 2.1, it is possible to write teh functional $\mathcal{G}_n$ as (apart from the fidelity term)

$$\begin{aligned} \mathcal{G}_n(u) = \frac{1}{\varepsilon}\int_{A\times A} \eta_\varepsilon(T_n(x)-T_n(z))|u(T_n(x))-u(T_n(z))|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\ + \frac{1}{\varepsilon}\int_A V(u(T_n(x)))\rho(x)\,\mathrm{d}x, \end{aligned}$$

for some maps $T_n : \Omega \to X_n$. This writing brings to light the connection between our model for labeling and the non-local model for phase transitions (2) studied by Alberti and Bellettini. Indeed, as explained above, the rationale behind the choice of the energy $\mathcal{G}_n$ is to view labeling of a data set as a problem of partitioning the data set into classes. Therefore, apart from the passage from the discrete non-local model to the continuum non-local model, from this last one we will be able to suitably adapt the techniques introduced in [1].

1.3. The continuum model

In order to define the limiting functional, we first need to introduce some notation.

DEFINITION 1. *Let $\nu \in \mathbb{R}^d$. Define $\nu^\perp := \{z \in \mathbb{R}^d \,:\, z\cdot\nu = 0\}$. Moreover, for $x \in \mathbb{R}^d$, set*

$$\mathcal{C}(x,\nu) := \Big\{ C \subset \nu^\perp \,:\, C \textit{ is a } (d-1)\textit{-dimensional cube centred at } x \Big\}\,.$$

For $C \in \mathcal{C}(x,\nu)$, we denote by $v_1,\dots,v_{d-1}$ its principal directions (where each v_i is a unit vector normal to the i^{th} face of C), and we say that a function $u:\mathbb{R}^d \to \mathbb{R}$ is C-periodic *if $u(y+rv_i) = u(y)$ for all $y \in \mathbb{R}^d$, all $r \in \mathbb{N}$ and all $i = 1,\dots,d-1$.*

Finally, we consider the following space of functions:

$$\mathcal{U}(C,\nu) := \Big\{ u:\mathbb{R}^d \to [-1,1] \,:\, u \textit{ is } C\textit{-periodic}, \lim_{y\cdot\nu\to\infty} u(y) = 1, \textit{ and } \lim_{y\cdot\nu\to-\infty} u(y) = -1 \Big\}\,.$$

DEFINITION 2. *Let $p \geq 1$ and $X \subset \mathbb{R}^d$ be open and bounded. Define the functional $\mathcal{G}_\infty : L^1(X) \to [0,\infty]$ by*

$$\mathcal{G}_\infty(u) := \begin{cases} \displaystyle\int_{\partial^*\{u=1\}} \sigma(x,\nu_u(x))\rho(x)\,\mathrm{d}\mathcal{H}^{d-1}(x) + \lambda\int_B |u(x)-\overline{u}(x)|^q \mathrm{d}x \\ \qquad\qquad\qquad \textit{if } u \in BV(X;\{\pm1\})\,, \\ +\infty \qquad\qquad\qquad\qquad\qquad\qquad \textit{else}\,, \end{cases}$$

where

$$\sigma(x,\nu) := \inf\left\{ \frac{1}{\mathcal{H}^{d-1}(C)} G(u,\rho(x),T_C) \,:\, C \in \mathcal{C}(x,\nu)\,, u \in \mathcal{U}(C,\nu) \right\},$$

and, for $C \in \mathcal{C}(x,\nu)$, we set $T_C := \{z + t\nu \,:\, z \in C, t \in \mathbb{R}\}$. Finally, for $\lambda \in \mathbb{R}$ and $A \subset \mathbb{R}^d$ define

$$G(u,\lambda,A) := \lambda \int_A \int_{\mathbb{R}^d} \eta(h)|u(z+h)-u(z)|^p\,\mathrm{d}h\,\mathrm{d}z + \int_A V(u(z))\,\mathrm{d}z\,.$$

Here $\partial^*\{u=1\}$ denotes the reduced boundary of $\{u=1\}$ and $\nu_u(x)$ is the measure theoritic exterior normal to the set $\{u=1\}$ at the point $x \in \partial^*\{u=1\}$ (see Definition 6).

REMARK 1. It is possible to see that the the function $(x,\nu) \mapsto \sigma(x,\nu)$ is upper semi-continuous on $X \times \mathbb{S}^{d-1}$, while, for every $\nu \in \mathbb{S}^{d-1}$, the function $x \mapsto \sigma(x,\nu)$ is continuous on X.

Notice that the discrete functional $\mathcal{G}_n$ is nonlocal while the functional $\mathcal{G}_\infty$ is local. The reason for this localization is because we rescaled the weights W_n^{ij}. The minimization problem defining σ is called the *cell problem* and it is common in phase transitions problems (see related works in Section **??**). Although not explicit, we have at least information on the form of the limiting functional: an anisotropic weighted perimeter plus an L^q fidelity term.

1.4. Main results

In this section we state the main result proved in [24]. We refer to Section 2 for the definitions. Let $X \subset \mathbb{R}^d$ be a bounded, connected and open set with Lipschitz boundary. Fix $\mu \in \mathcal{P}(X)$ and assume the following.

(A1) $\mu \ll \mathcal{L}^d$, has a continuous density $\rho : X \to [c_1, c_2]$ for some $0 < c_1 \leq c_2 < \infty$.

We extend ρ to a function defined in the whole space $\mathbb{R}^d$ by setting $\rho(x) := 0$ for $x \in \mathbb{R}^d \setminus X$. For all $n \in \mathbb{N}$, consider a point cloud $X_n = \{x_i\}_{i=1}^n \subset X$ and let μ_n be the associated empirical measure. Let $\{\varepsilon_n\}_{n=1}^\infty$ be a positive sequence converging to zero and such that the following rate of convergence holds:

(A2) $\dfrac{\mathrm{dist}_\infty(\mu_n, \mu)}{\varepsilon_n} \to 0\,.$

The double well potential $V : \mathbb{R} \to \mathbb{R}$ satisfies the following.

(B1) V is continuous.

(B2) $V^{-1}(0) = \{\pm 1\}$ and $V \geq 0$.

(B3) There exists $\tau > 0, R_V > 1$ such that for all $|s| \geq R_V$ that $V(s) \geq \tau |s|$.

(B4) V is Lipschitz continuous on $[-1, 1]$.

Recall that the graph weights are defined by $W_n^{ij} = \eta_{\varepsilon_n}(x_i - x_j)$. We assume that $\eta : \mathbb{R}^d \to [0, \infty)$ is a measurable functions satisfying the following.

(C1) $\eta \geq 0$, $\eta(0) > 0$ and η is continuous at $x = 0$.

(C2) η is an even function, i.e. $\eta(-x) = \eta(x)$.

(C3) η has support in $B(0, R_\eta)$, for some $R_\eta > 0$.

(C4) For all $\delta > 0$ there exists c_δ, α_δ such that if $|x - z| \leq \delta$ then $\eta(x) \geq c_\delta \eta(\alpha_\delta z)$, furthermore $c_\delta \to 1$, $\alpha_\delta \to 1$ as $\delta \to 0$.

REMARK 2. We collect here the comments on the several hypothesis listed above. The assumption $0 < c_1 \leq \rho \leq c_2 < +\infty$ is usually satisfied in practical applications: usually concentration phenomena happen on manifold, and thus the whole problem can be stated directly on that manifold. Moreover, if $\rho = 0$ is some regions, it is possible to neglect the set $\{\rho \leq \delta\}$, for $\delta \ll 1$. We note that these two conditions are just technical, and do not affect any the limiting functional.

When $x_i \overset{\text{iid}}{\sim} \mu$ then (with probability one), hypothesis (A2) is implied by $\varepsilon_n \gg \delta_n$, where δ_n is defined in Theorem 2. Notice that for $d \geq 3$ this lower bound on ε_n ensures that the graph with vertices x_n and edges weighted by W^{ij} is eventually connected (see [47, Theorem 13.2]). The lower bound can potentially be improved when x_i are not independent. For example if $\{x_i\}_{i=1}^n$ form a regular graph then μ_n converges to the uniform measure and the lower bound is given by $\varepsilon_n \gg n^{-\frac{1}{d}}$.

Assumption (B3) is used to establish compactness, in particular it is used to show that minimisers can be bounded in L^∞ by 1. The prototypical example of a function $V : \mathbb{R}^d \to \mathbb{R}$ satisfying (B1-4) is given by $V(s) := (s^2-1)^2$.

Note that (C3) and (C4) imply that $\|\eta\|_{L^\infty} < \infty$ and, in particular, $\int_{\mathbb{R}^d} \eta(x)|x|\,\mathrm{d}x < \infty$. Indeed, given $\delta > 0$, it is possible to cover $B(0,R_\eta)$ with a finite family $\widetilde{B}_\delta(x_1),\ldots,\widetilde{B}_\delta(x_r)$ of sets of the form $\widetilde{B}_\delta(x_i) := \{\,\alpha_\delta z : |z-x_i| < \delta\,\}$. Hypothesis (C2) is justified by the fact that η plays the role of an interaction potential. Finally, hypothesis (C4) is a version of continuity of η we need in order to perform our technical computations. We note that (C4) is general enough to include $\eta(x) = \chi_A$ where $A \subset \mathbb{R}^d$ is open, bounded, convex and $0 \in A$, see [58, Proposition 2.2].

The main result of the paper is the following.

THEOREM 1. *Let $p,q \geq 1$ and assume (A1-2), (B1-4) and (C1-4) are in force. Let $B \subset X$ be an open set with $|B| > 0$, and $|\partial B| = 0$. Then, every sequence $\{u_n\}_{n=1}^\infty$, with $u_n \in L^1(\mu_n)$, such that*

$$\sup_{n\in\mathbb{N}} \mathcal{G}_n(u_n) < \infty$$

is relatively compact in TL^1, and each cluster point $u \in L^1(X,\mu)$ has $\mathcal{G}_\infty(u) < \infty$.

Moreover, the sequence of functionals $\{\mathcal{G}_n\}_n$ Γ-converges to the functional $\mathcal{G}_\infty$ in the TL^1 topology.

The above result ensure the consistency of the discrete model we are considering. Using well-known properties of Γ-convergence, it is possible to prove that cluster points of minimizers of the discrete functionals $\mathcal{G}_n$ are minimizers of the limiting functional $\mathcal{G}_\infty$.

The functional $\mathcal{G}_n$ in the case $p = 1$ has been considered by Thorpe and Theil in [58], where a similar Γ-convergence result has been proved. The difference is that, in the case $p = 1$, the limit energy density function $\sigma^{(1)}$ can be given explicitly, via an integral. In [60] van Gennip and Bertozzi studied the Ginzburg-Landau functional on 4-regular graphs for $d = 2$ and $p = 2$ proving limits for $\varepsilon \to 0$ and $n \to \infty$ (both simultaneously and independently).

The TL^p topology, as introduced by García Trillos and Slepčev [37], provides a notion of convergence upon which the Γ-convergence framework can be applied. This method has now been applied in many works, see, for instance, [37, 58, 38, 29, 26, 39, 34, 35, 53]. Further studies on this topology can be found in [37, 38, 56, 57].

1.5. Future directions

The result presented in this paper is part of an ongoing project in collaboration with Matthew Thorpe. We are currently studying the effect of non-locality and phase transition separately, namely when the rescaling of the weights is done with a parameter δ_n that is different from the parameter ε_n used for the phase transition (rescaling of the functional). Several regimes depending on $\lim_{n\to+\infty}\frac{\varepsilon_n}{\delta_n}$ are considered. Moreover, we want to understand whether the Γ-convergence of the energies can be used to infer convergence of the solutions of the corresponding gradient flows.

Finally, we also aim at studying the consistency of variational models for the classification of big data, where different operators on graphs, like the normalized Laplacian, are used to define the discrete functional.

2. Background

We collect here the main definitions and results that will be needed in the paper.

2.1. Transportation theory

In this section we collect the fundamental material needed in order to explain how to compare functions defined in different spaces, namely a function $w\in L^1(X,\mu)$ and a function $u\in L^1(X_n,\mu_n)$, where $X\subset\mathbb{R}^d$ is an open set, $X_n=\{x_i\}_{i=1}^n\subset X$ is a finite set of points, and $\mu_n:=\frac{1}{n}\sum_{i=1}^n\delta_{x_i}$ denotes the empirical measures relative to X_n. This is fundamental in stating our Γ-convergence result (Theorem 1). The TL^p space was introduced in [37] (see also [57]). For a general introduction on optimal transportation, see [49, 61, 62]. Here the definitions we introduce tailored the special case we are interested in.

DEFINITION 3. *Let $X\subset\mathbb{R}^d$ be an open set, $\mu=\rho\mathcal{L}^d$ be a probability measure on X, and assume the density ρ is such that $0<c_1\leq\rho\leq c_2<\infty$. For $p\in[1,+\infty)$, and a positive Radon measure λ on X, we define the p-Wasserstein distance between μ and λ as*

$$\mathrm{dist}_p(\mu,\lambda):=\inf\left\{\,\|\mathrm{Id}-T\|_{L^p(X,\mu)}\,:\,T:X\to X\ \textit{Borel},\ T_{\#}\mu=\lambda\,\right\},$$

where

$$\|\mathrm{Id}-T\|^p_{L^p(X,\mu)}:=\int_X|x-T(x)|^p\rho(x)\,\mathrm{d}x$$

and we define the push forward *measure $T_{\#}\mu$ as $T_{\#}\mu(A):=\mu\left(T^{-1}(A)\right)$ for all $A\subset X$. In the case $p=+\infty$ we get*

$$\mathrm{dist}_\infty(\mu,\lambda)=\inf\left\{\,\|\mathrm{Id}-T\|_{L^\infty(X,\mu)}\,:\,T:X\to X\ \textit{Borel},\ T_{\#}\mu=\lambda\,\right\},$$

where

$$\|\mathrm{Id}-T\|_{L^\infty(X,\mu)}:=\operatorname*{ess\,sup}_X\rho(x)[x-T(x)].$$

A map T is called a transport map *between μ and λ if $T_{\#}\mu=\lambda$.*

REMARK 3. It is possible to see (see [61]) that the infimum is actually achieved. Moreover, the metric dist_p is equivalent to the weak* convergence of probability measures plus convergence of p^{th} moments.

Throughout the paper we will assume the empirical measures μ_n converges weakly* to μ, so by Remark 3 there exists a sequence of Borel maps $\{T_n\}_{n=1}^{\infty}$ with $T_n : X \to X_n$ and $(T_n)_\# \mu = \mu_n$ such that

$$\lim_{n\to\infty} \|\mathrm{Id} - T_n\|^p_{L^p(X,\mu)} = 0\,.$$

Such a sequence of functions $\{T_n\}_{n=1}^{\infty}$ will be called *stagnating*. We are now in position to define the notion of convergence for sequences $u_n \in L^p(X_n)$ to a continuum limit $u \in L^p(X,\mu)$.

DEFINITION 4. *Let $u_n \in L^p(X_n)$, $w \in L^p(X,\mu)$ where $X_n = \{x_i\}_{i=1}^n$ and assume that the empirical measure μ_n converges weak* to μ. We say that $u_n \to w$ in $TL^p(X)$, and we write $u_n \xrightarrow{TL^p} w$, if there exists a sequence of stagnating transport maps $\{T_n\}_{n=1}^{\infty}$ between μ and μ_n such that*

$$\|v_n - w\|_{L^p(X,\mu)} \to 0\,, \tag{3}$$

as $n \to \infty$, where $v_n := u_n \circ T_n$, and T_n is a sequence of stagnating maps.

The above definition of the TL^p convergence hides the underling metric structure. Since in the proof of the Γ-convergence result we want to exploit the metric structure of the TL^p spaces, we briefly describe them here. We define the $TL^p(X)$ space as the space of couplings (u,μ) where $\mu \in \mathcal{P}(X)$ has finite p^{th} moment and $u \in L^p(\mu)$. We define the distance $d_{TL^p} : TL^p(X) \times TL^p(X) \to [0,+\infty)$ for $p \in [1,+\infty)$ by

$$d_{TL^p}((u,\mu),(v,\lambda)) := \inf_{T_\#\mu=\lambda} \left(\int_X |x - T(x)|^p + |u(x) - v(T(x))|^p \,\mathrm{d}\mu(x) \right)^{\frac{1}{p}},$$

or for $p = +\infty$ by

$$d_{TL^\infty}((u,\mu),(v,\lambda)) := \inf_{T_\#\mu=\lambda} \left(\operatorname*{ess\,inf}_{\mu} \{ |x - T(x)| + |u(x) - v(T(x))| \,:\, x \in X \} \right).$$

We now state the relation between the TL^p-distance and the TL^p-convergence. For a proof we refer to [37, Remark 3.4 and Proposition 3.12].

PROPOSITION 1. *The distance d_{TL^p} is a metric and furthermore,*

$$d_{TL^p}((u_n,\mu_n),(u,\mu)) \to 0$$

if and only if $\mu_n \overset{w^*}{\rightharpoonup} \mu$ and there exists a sequence of stagnating transport maps $\{T_n\}_{n=1}^{\infty}$ between μ and μ_n such that $\|u_n \circ T_n - u\|_{L^p(X,\mu)} \to 0$.

Finally, we state a rate of convergence for a sequence of stagnating maps proved by García Trillos and Slepčev in [36] we will make use in the proof of the main result.

THEOREM 2. *Let $X \subset \mathbb{R}^d$ be a bounded, connected and open set with Lipschitz boundary. Let $\mu := \rho\mathcal{L}^d$ be a probability measure on X, with $0 < c_1 \leq \rho \leq c_2 < \infty$, and let $\{x_i\}_{i=1}^{\infty}$. Assume $\mu_n \overset{w^*}{\rightharpoonup} \mu$, where μ_n is the empirical measure associated to $\{x_i\}_{i=1}^{\infty}$. Then, there exist a constant $C > 0$, and a sequence $\{T_n\}_{n=1}^{\infty}$ of maps $T_n : X \to X$ with $(T_n)_{\#}\mu = \mu_n$ and*

$$\limsup_{n\to\infty} \frac{\|T_n - \mathrm{Id}\|_{L^\infty(X)}}{\delta_n} \leq C\,,$$

where

$$\delta_n := \begin{cases} \sqrt{\frac{\log\log n}{n}} & \text{if } d = 1\,, \\ \frac{(\log n)^{\frac{3}{4}}}{\sqrt{n}} & \text{if } d = 2\,, \\ \left(\frac{\log n}{n}\right)^{\frac{1}{d}} & \text{if } d \geq 3\,. \end{cases}$$

2.2. Sets of finite perimeter

In this section we recall the definition and basic facts about sets of finite perimeter. We refer the reader to [3] for more details.

DEFINITION 5. *Let $E \subset \mathbb{R}^d$ with $|E| < \infty$ and let $X \subset \mathbb{R}^d$ be an open set. We say that E has* finite perimeter *in X if*

$$|D\chi_E|(X) := \sup\left\{ \int_E \mathrm{div}\varphi\,\mathrm{d}x \,:\, \varphi \in C^1_c(X;\mathbb{R}^d)\,,\, \|\varphi\|_{L^\infty} \leq 1 \right\} < \infty\,.$$

REMARK 4. If $E \subset \mathbb{R}^d$ is a set of finite perimeter in X it is possible to define a finite vector valued Radon measure $D\chi_E$ on A such that

$$\int_{\mathbb{R}^d} \varphi\,\mathrm{d}D\chi_E = \int_E \mathrm{div}\varphi\,\mathrm{d}x$$

for all $\varphi \in C^1_c(X;\mathbb{R}^d)$.

DEFINITION 6. *Let $E \subset \mathbb{R}^d$ be a set of finite perimeter in the open set $X \subset \mathbb{R}^d$. We define $\partial^* E$, the* reduced boundary *of E, as the set of points $x \in \mathbb{R}^d$ for which the limit*

$$\nu_E(x) := -\lim_{r\to 0} \frac{D\chi_E(x + rQ)}{|D\chi_E|(x + rQ)}$$

exists and is such that $|\nu_E(x)| = 1$. *Here* Q *denotes the unit cube of* $\mathbb{R}^d$ *centered at the origin with sides parallel to the coordinate axes. The vector* $\nu_E(x)$ *is called the* measure theoretic exterior normal *to* E *at* x.

We now recall the structure theorem for sets of finite perimeter due to De Giorgi, see [3, Theorem 3.59] for a proof of the following theorem.

THEOREM 3. *Let* $E \subset \mathbb{R}^d$ *be a set with finite perimeter in the open set* $X \subset \mathbb{R}^d$. *Then*

(i) for all $x \in \partial^* E$ *the set* $E_r := \frac{E-x}{r}$ *converges locally in* $L^1(\mathbb{R}^d)$ *as* $r \to 0$ *to the halfspace orthogonal to* $\nu_E(x)$ *and not containing* $\nu_E(x)$,

(ii) $D\chi_E = \nu_E \,\mathcal{H}^{d-1} \llcorner \partial^* E$,

(iii) the reduced boundary $\partial^* E$ *is* $\mathcal{H}^{d-1}$*-rectifiable,* i.e.*, there exist Lipschitz functions* $f_i : \mathbb{R}^{d-1} \to \mathbb{R}^d$ *such that*

$$\partial^* E = \bigcup_{i=1}^{\infty} f_i(K_i)\,,$$

where each $K_i \subset \mathbb{R}^{d-1}$ *is a compact set.*

REMARK 5. Using the above result it is possible to prove that (see [31])

$$\nu_E(x) = -\lim_{r \to 0} \frac{D\chi_E(x + rQ)}{r^{d-1}}$$

for all $x \in \partial^* E$, where Q is a unit cube centred at 0 with sides parallel to the co-ordinate axis.

Using the result [3, Theorem 3.42] and the fact that it is possible to approximate every smooth surface with polyhedral sets, it is possible to obtain the following density result.

THEOREM 4. *Let* $E \subset \Omega$ *be a set of finite perimeter. Then there exists a sequence* $\{E_n\}_{n=1}^{\infty}$ *of sets of finite perimeter in* Ω, *such that* ∂E_n *is a Lipschitz manifold contained in the union of finitely many affine hyperplanes,* $\chi_{E_n} \to \chi_E$, *and* $|D\chi_{E_n}|(\Omega) \to |D\chi_E|(\Omega)$.

Finally, we recall a result due to Reshetnvyak in the form we will need in this paper (for a proof of the general case see, for instance, [3, Theorem 2.38]).

THEOREM 5. *Let* $\{E_n\}_{n=1}^{\infty}$ *be a sequence of sets of finite perimeter in the open set* $X \subset \mathbb{R}^d$ *such that* $D\chi_{E_n} \overset{w^*}{\rightharpoonup} D\chi_E$ *and* $|D\chi_{E_n}|(X) \to |D\chi_E|(X)$, *where* E *is a set of finite perimeter in* X. *Let* $f : X \times \mathbb{S}^{d-1} \to [0,\infty)$ *be an upper semi-continuous function.*

Then

$$\limsup_{n\to\infty}\int_{\partial^* E_n\cap X} f\left(x,\nu_{E_n}(x)\right)\,\mathrm{d}\mathcal{H}^{d-1}(x)\leq\int_{\partial^* E\cap X} f\left(x,\nu_E(x)\right)\,\mathrm{d}\mathcal{H}^{d-1}(x)\,.$$

2.3. Γ-convergence

We recall the basic notions and properties of Γ-convergence (in metric spaces) we will use in the paper (for a reference, see [12, 25]).

DEFINITION 7. *Let (A,d) be a metric space. We say that $F_n : A\to[-\infty,+\infty]$ Γ-converges to $F : A\to[-\infty,+\infty]$, and we write $F_n\stackrel{\Gamma\text{-(d)}}{\longrightarrow}F$ or $F=\Gamma\text{-}\lim(d)_{n\to\infty}F_n$, if the following hold true:*

(i) for every $x\in A$ and every $x_n\to x$ we have $F(x)\leq\liminf_{n\to\infty}F_n(x_n)$;

(ii) for every $x\in A$ there exists $\{x_n\}_{n=1}^\infty\subset A$ (the so called recovery sequence*) with $x_n\to x$ such that $\limsup_{n\to\infty}F_n(x_n)\leq F(x)$.*

The notion of Γ-convergence has been designed in order for the following convergence of minimisers and minima result to hold (see for example [12, 25]).

THEOREM 6. *Let (A,d) be a metric space and let $F_n\stackrel{\Gamma-(\mathrm{d})}{\longrightarrow}F$, where F_n and F are as in the above definition. Let $\{\varepsilon_n\}_{n=1}^\infty$ with $\varepsilon_n\to 0^+$ as $n\to\infty$ and let $x_n\in A$ be* ε_n-minimizers *for F_n, that is*

$$F_n(x_n)\leq\max\left\{\inf_A F_n+\frac{1}{\varepsilon_n}\,,\,-\frac{1}{\varepsilon_n}\right\}. \tag{4}$$

Then every cluster point of $\{x_n\}_{n=1}^\infty$ is a minimizer of F.

3. The non-local continuum functional

Note that the fidelity term

$$\lambda\frac{1}{n}\sum_{i=1}^n|u(x_i)-\overline{u}(x_i)|^q\mathbb{1}_B(x_i)$$

does not pose particular problems for Γ-convergence. Therefore, from now on, we will consider the functionals $\mathcal{G}_n$ and $\mathcal{G}_\infty$ with $\lambda=0$.

We introduce here a non-local continuum functional that will be used as an intermediate step in the proof of compactness and Γ-convergence.

DEFINITION 8. *Let $p\geq 1$, $\varepsilon>0$, $s_\varepsilon>0$, and let $A\subset X$ be an open and bounded set. Define the functional $\mathcal{F}_\varepsilon(\cdot,A) : L^1(X)\to[0,\infty]$ by*

$$\mathcal{F}_\varepsilon(u,A)=\frac{s_\varepsilon}{\varepsilon}\int_{A\times A}\eta_\varepsilon(x-z)|u(x)-u(z)|^p\rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z+\frac{1}{\varepsilon}\int_A V(u(x))\rho(x)\,\mathrm{d}x\,. \tag{5}$$

When $A = X$, we will simply write $\mathcal{F}_\varepsilon(u)$.

Assume $s_\varepsilon \to 1$ as $\varepsilon \to 0$. The aim of this section is to prove the following result.

THEOREM 7. *Let $p \geq 1$, $\{\varepsilon_n\}_{n\in\mathbb{N}}$ with $\varepsilon_n \to 0$ as $n \to \infty$. Under conditions (A1), (B1-3) and (C1-3) the following hold:*

(i) Any $u_n \in L^1(X,\mu)$ satisfying

$$\sup_{n\in\mathbb{N}} \mathcal{F}_{\varepsilon_n}(u_n) < \infty$$

is relatively compact in $L^1(X,\mu)$, and each cluster point $u \in L^1(X,\mu)$ has $\mathcal{G}_\infty(u) < \infty$;

(ii) The functional $\mathcal{F}_{\varepsilon_n}$ Γ-converges to $\mathcal{G}_\infty$ in the L^1 topology.

The proof of Theorem 7 is a careful adaptation of the techniques used in [1] by Alberti and Bellettini, where the functional $\mathcal{F}_\varepsilon$ in the case $s_\varepsilon \equiv 1$, $p = 2$ and $\rho \equiv 1$ is studied. From the technical point of view, the continuity of the density ρ together with the bounds $0 < c_1 \leq \rho \leq c_2 < +\infty$ allows to locally treat it as a constant. A bit more care has to be taken in the generalization from the exponent $p = 2$ to a general exponent $p > 1$.

A fundamental technical tool provided in [1] is the control of the nonlocality

$$\Lambda_\varepsilon(u,A,B) := \frac{s_\varepsilon}{\varepsilon}\int_A\int_B \eta_\varepsilon(x-z)|u(x)-u(z)|^p\rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z\,.$$

in terms of a suitable notion of traces at the boundary of the sets.

PROPOSITION 2. *Let $v_n \to v$ in $L^1(X)$ with $|v_n| \leq 1$. Then, for all $\bar{x} \in \mathbb{R}^d$ and for all $\nu \in \mathbb{S}^{d-1}$ the following holds: given $C \in \mathcal{C}(\bar{x},\nu)$ consider the strip T_C and any cube $Q \subset \mathbb{R}^d$ whose intersection with $\nu^\perp$ is C. Then, for a.e. $t > 0$:*

(i) $\Lambda_{\varepsilon_n}(v_n, tT_C, \mathbb{R}^d \setminus tT_C) \to 0$ as $n \to \infty$,

(ii) $\Lambda_{\varepsilon_n}(v_n, tQ, tT_C \setminus tQ) \to 0$ as $n \to \infty$.

Another useful result we will make use of is the following continuity property of σ proved in [24].

LEMMA 1. *Under assumptions (A1) and (C3) the followings hold:*

(i) the function $(x,\nu) \mapsto \sigma(x,\nu)$ is upper semi-continuous on $X \times \mathbb{S}^{d-1}$,

(ii) for every $\nu \in \mathbb{S}^{d-1}$, the function $x \mapsto \sigma(x,\nu)$ is continuous on X.

Proof of Theorem 7. Step 1: proof of (i). Let $v_n(x) := \mathrm{sign}(u_n)$. It is possible to see that

$$\|u_n - v_n\|_{L^1} \to 0, \qquad \sup_{n\in\mathbb{N}} \mathcal{F}_{\varepsilon_n}(v_n) < +\infty. \tag{6}$$

Using (6), we conclude by noting that, since $v_n \in L^1(X;\{\pm 1\})$ we have

$$\mathcal{F}^{(p)}_{\varepsilon_n}(v_n) = 2^{p-2} s_n \mathcal{F}^{(2)}_{\varepsilon_n}(v_n)\,,$$

where we stressed the dependence on p of the functionals $\mathcal{F}^{(p)}_{\varepsilon_n}$. The result of Alberti and Bellettini then gives the desired compactness. Moreover, it is not difficult to see that $\mathcal{G}^{(2)}_\infty(u) < +\infty$ implies $\mathcal{G}^{(p)}_\infty(u) < +\infty$. Therefore, $\mathcal{G}_\infty(u) < +\infty$ for any cluser point $u \in L^1(X,\mu)$ of $\{u_n\}_{n=1}^\infty$.

Step 2: proof of (ii) - liminf inequality. The proof of the liminf inequality is based on the blow-up method (see [31]). In particular, it suffices to prove the following: let $x \in \Omega$, $\nu \in \mathbb{S}^{d-1}$, $Q \subset \mathbb{R}^d$ a unitary cube centered at the origin with two faces orthogonal to ν. Take a sequence $\{u_n\}_{n\in\mathbb{N}} \subset L^1(Q)$ with $u_n \to u$ in L^1, where $u(x) := \mathrm{sign}(x\cdot\nu)$, for $x \in Q$. Then we claim that

$$\lim_{n\to\infty} \frac{\mathcal{F}_{\varepsilon_n}(u_n, x + r_n Q)}{r_n^{d-1}} \geq \sigma(x,\nu)\,, \tag{7}$$

where $r_n \to 0$. In order to avoid useless technicalities, we will now assume $\rho \equiv 1$, and just describe the main ideas of the proof.

Let $C := Q \cap \nu(x)^\perp \in \mathcal{C}(x,\nu(x))$, and let $t \in (0,1)$. For $x \in Q$, and $r > 0$, we set $R_{x,r}u(y) := u(x+ry)$. Define the function $w_n : \mathbb{R}^d \to \mathbb{R}$ as the periodic extension of the function that is $R_{x,r_n} u_{\varepsilon_n}$ in Q and v_x in $T_C \setminus Q$. Set $\varepsilon'_n := \frac{\varepsilon_n}{r_n}$ and $s'_n = \min\left\{1, \frac{s_{\varepsilon_n}}{s_{\varepsilon'_n}}\right\}$. We get

$$\begin{aligned}
\frac{\lambda_{\varepsilon_n}(x + r_n t Q)}{r_n^{d-1}} &\geq \frac{\mathcal{F}_{\varepsilon_n}(u_{\varepsilon_n}, x + r_n t Q)}{r_n^{d-1}} \\
&\geq s'_n \mathcal{F}_{\varepsilon'_n}(R_{x,r_n} u_{\varepsilon_n}, tQ) \\
&= s'_n \mathcal{F}_{\varepsilon'_n}(w_n, tQ) \\
&\geq s'_n G_{\varepsilon'_n}(w_n, 1, T_C) \\
&\qquad - s'_n \left| \mathcal{F}_{\varepsilon'_n}(w_n, tQ) - G_{\varepsilon'_n}(w_n, 1, tT_C) \right| \\
&= s'_n \left(\frac{\varepsilon_n}{r_n}\right)^{d-1} G\left(R_{0,\varepsilon'_n} w_n, 1, \frac{r_n}{\varepsilon_n} t T_C\right) \\
&\qquad\qquad - s'_n \left| \mathcal{F}_{\varepsilon'_n}(w_n, tQ) - G_{\varepsilon'_n}(w_n, 1, tT_C) \right| \\
&\geq s'_n t^{d-1} \sigma(x,\nu) - s'_n \left| \mathcal{F}_{\varepsilon'_n}(w_n, tQ) - G_{\varepsilon'_n}(w_n, 1, tT_C) \right|\,.
\end{aligned} \tag{8}$$

Using Lemma 2, we get that, for a.e. $t \in (0,1)$,

$$\lim_{n\to\infty} \left| \mathcal{F}_{\varepsilon'_n}(w_n, tQ) - G_{\varepsilon'_n}(w_n, tT_C) \right| = 0\,.$$

Therefore, from (8), together with $s_\varepsilon \to 1$ as $\varepsilon \to 0$, we finally get (7).

Step 3: proof of (ii): limsup inequality. Using the upper semi-continuity of σ (see Lemma 1) together with Theorem 5, it is possible to construct a recovery sequence only for the class of polyhedral sets. We then conclude since this class is dense in the family of sets of finite perimeter (see Theorem 4).

Consider a polyhedral set E, and let Σ one of its faces, with normal ν. We will provide the construction of the recovery sequence only for the face Σ. We recall that there are no regularity requirements for the recovery sequence, other than being in L^1. Fix $\delta > 0$. By continuity of σ in the first variable (see Lemma 1) along with the continuity of ρ, it is then possible to partition Σ in a finite number of (parts of) cubes $U_1, \dots, U_{N_\delta}$ each of side r_δ, such that

$$\left| \int_{U_i} \sigma(x,\nu)\rho(x)\,\mathrm{d}\mathcal{H}^{d-1}(x) - r_\delta^{d-1} \sum_{j=1}^{N_\delta} \sigma(x_i,\nu)\rho(x_i) \right| < \delta . \tag{9}$$

where $x_i \in U_i$. Fix $i \in \{1, \dots, N_\delta\}$, and let $w_i \in \mathcal{U}(U_i, \nu)$ be such that

$$\frac{1}{\mathcal{H}^{d-1}(C_i)} G(w_i, \rho(x_i), T_{C_i}) < \sigma(x_i,\nu) + \frac{\delta}{\rho(x_i) N_\delta r_\delta^{d-1}} . \tag{10}$$

Define the map $v^i_{\varepsilon_n}$ as the periodic extension to U_i of the rescaled map $w_i\left(\frac{x}{\varepsilon_n}\right)$, and let

$$u_{\varepsilon_n}(x) := \sum_{i=1}^{N_\delta} v^i_{\varepsilon_n}(x) \mathbb{1}_{U_i}(x) .$$

We claim that:

(i) For all $i \neq j$
$$\lim_{n\to\infty} \Lambda_{\varepsilon_n}(u_{\varepsilon_n}, U_i, U_j) \to 0;$$

(ii) For all $i = 1, \dots, N_\delta$
$$\limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}(u_{\varepsilon_n}, U_i) \leq \mathcal{G}_\infty(u, U_i).$$

If the above claims hold true, then we can conclude as follows: we have

$$\begin{aligned}
\limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}(u_{\varepsilon_n}) &\leq \sum_{i=1}^{N_\delta} \limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}(u_{\varepsilon_n}, U_i) + 2 \sum_{i<j=0}^{N_\delta} \limsup_{n\to\infty} \Lambda_{\varepsilon_n}(u_{\varepsilon_n}, U_i, U_j) \\
&\leq \sum_{i=1}^{N_\delta} \mathcal{G}_\infty(u, U_i) \\
&= \mathcal{G}_\infty(u) .
\end{aligned}$$

Claim (i) follows from Proposition 2, since $v^i_{\varepsilon_n} \to v$ as $n \to \infty$, where $v(x) := \mathrm{sign}(x \cdot \nu)$. On the other hand, the continuity of ρ together with (9) and (10), ensures the validity of (ii). $\square$

4. Proof of Theorem 1

We first prove a technical result that will be needed in what follows.

LEMMA 2. *Let $u_n \to u$ in $TL^1(X)$. We claim that there exist $\alpha_n, c_n, \hat{\alpha}_n, \hat{c}_n > 0$ with $\alpha_n, c_n, \hat{\alpha}_n, \hat{c}_n \to 1$ such that*

$$c_n \eta\left(\frac{\alpha_n(x-z)}{\varepsilon_n}\right) \leq \eta\left(\frac{T_n(x)-T_n(z)}{\varepsilon_n}\right) \leq \hat{c}_n \eta\left(\frac{\hat{\alpha}_n(x-z)}{\varepsilon_n}\right).$$

Proof. Step 1: First inequality. Let $\delta_n := \frac{2\|T_n - \mathrm{Id}\|_{L^\infty}}{\varepsilon_n}$. By Assumption (C4) we can find α_n, c_n such that, for all $a, b \in \mathbb{R}^d$ with $|a-b| \leq \delta_n$, we have

$$\eta(a) \geq c_n \eta(\alpha_n b). \tag{11}$$

Since by assumption (A2) we have that $\delta_n \to 0$ then α_n, c_n can be chosen such that $\alpha_n \to 1, c_n \to 1$. Now if we let $a := \frac{T_n(x)-T_n(z)}{\varepsilon_n}$ and $b := \frac{x-z}{\varepsilon_n}$ we have

$$|a-b| = \frac{|T_n(x)-T_n(z)+z-x|}{\varepsilon_n} \leq \frac{2\|T_n - \mathrm{Id}\|_{L^\infty}}{\varepsilon_n} = \delta_n$$

and therefore, by (11), we get $\eta\left(\frac{T_n(x)-T_n(z)}{\varepsilon_n}\right) \geq c_n \eta\left(\frac{\alpha_n(x-z)}{\varepsilon_n}\right)$ as required.

Step 2: Second inequality. We use the following subclaim: for all $\hat{\delta} > 0$ sufficiently small there exists $\hat{\alpha}_{\hat{\delta}}, \hat{c}_{\hat{\delta}} > 0$ such that $\hat{\alpha}_{\hat{\delta}} \to 1, \hat{c}_{\hat{\delta}} \to 1$, as $\hat{\delta} \to 0$, and, for any $\hat{a}, \hat{b} \in \mathbb{R}^d$, it holds

$$|\hat{a} - \hat{b}| < \hat{\delta} \quad \to \quad \hat{c}_{\hat{\delta}} \eta(\hat{\alpha}_{\hat{\delta}} \hat{b}) \geq \eta(\hat{a}). \tag{12}$$

Then the desired inequality can be obtained as follows: for any $n \in \mathbb{N}$ take

$$\hat{a} := \frac{T_n(x)-T_n(z)}{\varepsilon_n}, \qquad \hat{b} := \frac{x-z}{\varepsilon_n}, \qquad \hat{\delta} := \frac{2\|T_n - \mathrm{Id}\|_{L^\infty}}{\varepsilon_n},$$

and let $\hat{\alpha}_n, \hat{c}_n$ be the numbers given by the subclaim for which (12) holds. Note that $\hat{\alpha}_n$ is chosen independently from w_n (since w_n depends on $\hat{\alpha}_n$ there is therefore no circular argument). Then, since $|\hat{a} - \hat{b}| \leq \delta_n$, we conclude.

To prove the subclaim, we let α_δ, c_δ be as in Assumption (C4) and let $\hat{\delta} > 0$. Without loss of generality we assume that $\inf_{\gamma \in (0,1]} \alpha_\gamma \in (0, \infty)$. We choose $\delta := \min\left(b1, \frac{\hat{\delta}}{\inf_{s \in (0,1]} \alpha_s}\right) b$, trivially $\delta \to 0$ as $\hat{\delta} \to 0$. We assume that $\frac{\hat{\delta}}{\inf_{\gamma \in (0,1]} \alpha_\gamma} \leq 1$. Let $\hat{a}, \hat{b} \in \mathbb{R}^d$ with $|\hat{a} - \hat{b}| < \hat{\delta}$, and define $a := \frac{\hat{a}}{\alpha_\delta}$ and $b := \frac{\hat{b}}{\alpha_\delta}$. Since, $|a-b| \leq \frac{\hat{\delta}}{\alpha_\delta} \leq \frac{\hat{\delta}}{\inf_{\gamma \in (0,1]} \alpha_\gamma} = \delta$ then

$$\eta(b) \geq c_\delta \eta(\alpha_\delta a) \quad \to \quad \frac{1}{c_\delta} \eta\left(\frac{\hat{b}}{\alpha_\delta}\right) \geq \eta(\hat{a}).$$

Let $\hat{c}_{\hat{\delta}} := \frac{1}{c_\delta}$, $\hat{\alpha}_{\hat{\delta}} := 1/\alpha_\delta$ then $\hat{\delta} \to 0$ implies $\delta \to 0$ which in turn implies $\alpha_\delta, c_\delta \to 1$ and therefore $\hat{\alpha}_{\hat{\delta}}, \hat{c}_{\hat{\delta}} \to 1$. This proves the claim. $\square$

We are now in position to prove Theorem 1.

Step 1: Compactness. Let $\{u_n\}_{n=1}^\infty$ with $u_n \in L^1(X_n)$ be such that

$$\sup_{n\in\mathbb{N}} \mathcal{G}_n(u_n) < +\infty.$$

Let $\{T_n\}_{n=1}^\infty$ be the corresponding transport maps given by Theorem 2. Set $v_n := u_n \circ T_n$. Then, using Lemma 2, we get

$$\begin{aligned}
\mathcal{G}_n(u_n) &= \frac{1}{\varepsilon_n^{d+1}} \int_X \int_X \eta\left(\frac{T_n(x)-T_n(z)}{\varepsilon_n}\right) |v_n(x)-v_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\quad + \frac{1}{\varepsilon_n}\int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&\geq \frac{c_n}{\varepsilon_n^{d+1}} \int_X \int_X \eta\left(\frac{\alpha_n(x-z)}{\varepsilon_n}\right) |v_n(x)-v_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\quad + \frac{1}{\varepsilon_n}\int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&= \frac{c_n}{\alpha_n^{d+1}(\varepsilon_n')^{d+1}} \int_X \int_X \eta\left(\frac{x-z}{\varepsilon_n'}\right) |v_n(x)-v_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\quad + \frac{1}{\varepsilon_n}\int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&= \frac{1}{\alpha_n}\mathcal{F}_{\varepsilon_n'}(v_n) \qquad (13)
\end{aligned}$$

where $\varepsilon_n' := \frac{\varepsilon_n}{\alpha_n}$. Therefore, the compactness follows from Theorem 7.

Step 2: Liminf inequality. Let $u \in L^1(\mu)$ and take $u_n \in L^1(\mu_n)$ with $u_n \to u$ in TL^1. From Lemma 2, and using Theorem 7 we have

$$\liminf_{n\to\infty} \mathcal{G}_n(u_n) \geq \liminf_{n\to\infty} \frac{\varepsilon_n'}{\varepsilon_n}\mathcal{F}_{\varepsilon_n'}(v_n) \geq \mathcal{G}_\infty(u)$$

since $\lim_{n\to\infty}\frac{\varepsilon_n'}{\varepsilon_n} = 1$.

Step 3: Limsup inequality Using the density of polyhedral sets in the family of sets of finite perimeter (Theorem 4), together with Theorem 5, it suffices to provide a recovery sequence in the case $\partial\{u=1\}$ is a polyhedral set.

Let $\{w_n\}_{n=1}^\infty \subset L^1(X)$ be the recovery sequence provided by Theorem 7 such that

$$w_n \to u\,, \;\text{ in } L^1(X)\,, \qquad \lim_{n\to\infty}\mathcal{F}_{\varepsilon_n}(w_n) = \mathcal{G}_\infty(u)\,.$$

For each $n \in \mathbb{N}$, set

$$u_n(x_i) = n \int_{T_n^{-1}(x_i)} w_n(x)\,\mathrm{d}x\,,$$

where $X_n = \{x_1, \dots, x_n\}$.

Let $\zeta_n \to 0$ with $\zeta_n \gg \sqrt{\frac{\|T_n - \mathrm{Id}\|_{L^\infty}}{\varepsilon_n}}$. A careful analysis of the way the recovery sequence in Theorem 7 has been constructed, allows to choose $\{w_n\}_{n=1}^\infty$ in such a way that each w_n is Lipschitz continuous with $\mathrm{Lip}(w_n) = \frac{1}{\zeta_{\varepsilon_n}\varepsilon_n}$ and $u_{\varepsilon_n}(x) = u(x)$ for all x satisfying

$$\mathrm{dist}(x, \partial^*\{u = 1\}) > \frac{\varepsilon_n}{\zeta_{\varepsilon_n}}\,.$$

Then, using the fact that $w_n \to u$ in $L^1(X)$, it is possible to show that

$$u_n \to u\,, \qquad \text{in } TL^1\,.$$

We now show that

$$\limsup_{n\to\infty} \mathcal{G}_n(u_n) \le \mathcal{G}_\infty(u)\,.$$

For, write

$$\begin{aligned}
\mathcal{G}_n(u_n) &= \frac{1}{\varepsilon_n}\int_X\int_X \eta_{\varepsilon_n}(T_n(x)-T_n(z))|v_n(x)-v_n(z)|^p\rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z + \frac{1}{\varepsilon_n}\int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&\le \frac{\hat{c}_n}{\hat{\alpha}_n^{d+1}\varepsilon_n'}\int_X\int_X \eta_{\varepsilon_n'}(x-z)|v_n(x)-v_n(z)|^p\rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z + \frac{1}{\varepsilon_n}\int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&= \frac{\hat{c}_n}{\hat{\alpha}_n^{d+1}\varepsilon_n'}\int_X\int_X \eta_{\varepsilon_n'}(x-z)|w_n(x)-w_n(z)|^p\rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z + \frac{1}{\varepsilon_n}\int_X V(w_n(x))\rho(x)\,\mathrm{d}x \\
&\qquad + a_n + b_n
\end{aligned}$$

where we recall $\varepsilon_n' := \frac{\varepsilon_n}{\hat{\alpha}_n}$ and

$$\begin{aligned}
a_n &:= \frac{\hat{c}_n}{\hat{\alpha}_n^{d+1}\varepsilon_n'}\int_X\int_X \eta_{\varepsilon_n'}(x-z)\left(|v_n(x)-v_n(z)|^p - |w_n(x)-w_n(z)|^p\right)\rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z\,, \\
b_n &:= \frac{1}{\varepsilon_n}\int_X \left(V(v_n(x)) - V(w_n(x))\right)\rho(x)\,\mathrm{d}x\,.
\end{aligned}$$

Using the fact that $\|w_n - v_n\|_{L^1(X)} \to 0$ as $n \to \infty$, together with the fact that for all $\delta > 0$ there exits $C_\delta > 0$ such that for any $a, b \in \mathbb{R}^d$ we have

$$|a|^p \le (1+\delta)|b|^p + C_\delta|a-b|^p\,,$$

and

$$|a+b|^p \le 2^{p-1}\left(|a|^p + |b|^p\right)\,,$$

it is possible to show that $a_n, b_n \to 0$ as $n \to \infty$. Therefore, we get

$$\limsup_{n\to\infty} \mathcal{G}_n(u_n) \le (1+\delta)\,\mathcal{G}_\infty(u)\,.$$

We conclude since $\delta > 0$ is arbitrary.

Acknowledgements

The author would like to thank the hospitality of Politecnico of Torino, and the Progetto di Ricerca INdAM Giovani Ricercatori “Optimal Shapes in Boundary Value Problems" for supporting the organisation of the workshop *Analysis and applications: contribution from young researchers*. The research was supported by the grant EP/R013527/2 “Designer Microstructure via Optimal Transport Theory" of David Bourne.

References

[1] G. ALBERTI AND G. BELLETTINI, *A nonlocal anisotropic model for phase transitions: Asymptotic behaviour of rescaled energies*, European Journal of Applied Mathematics, 9 (1998).

[2] L. AMBROSIO, *Metric space valued functions of bounded variation*, Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, 17 (1990), pp. 439–478.

[3] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 2000.

[4] N. ANSINI, A. BRAIDES, AND V. CHIADÒ PIAT, *Interactions between homogenization and phase-transition processes*, Tr. Mat. Inst. Steklova, 236 (2002), pp. 386–398.

[5] N. ANSINI, A. BRAIDES, AND V. CHIADÒ PIAT, *Gradient theory of phase transitions in composite media*, Proc. Roy. Soc. Edinburgh Sect. A, 133 (2003), pp. 265–296.

[6] S. ARORA, S. RAO, AND U. VAZIRANI, *Expander flows, geometric embeddings and graph partitioning*, Journal of the ACM, 56 (2009), pp. Art. 5, 37.

[7] S. BALDO, *Minimal interface criterion for phase transitions in mixtures of Cahn-Hilliard fluids*, Annales de l'Institut Henri Poincaré. Analyse Non Linéaire, 7 (1990), pp. 67–90.

[8] A. C. BARROSO AND I. FONSECA, *Anisotropic singular perturbations the vectorial case*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 124 (1994), pp. 527–571.

[9] G. BOUCHITTE, *Singular perturbations of variational problems arising from a two-phase transition model*, Applied Mathematics and Optimization, 21 (1990), pp. 289–314.

[10] J. BOURGAIN, H. BREZIS, AND P. MIRONESCU, *Another look at Sobolev spaces*, in Optimal Control and Partial Differential Equations, IOS, Amsterdam, 2001, pp. 439–455.

[11] Y. Boykov, O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2001), pp. 1222–1239.

[12] A. Braides, *Γ-Convergence for Beginners*, Oxford University Press, 2002.

[13] A. Braides and N. K. Yip, *A quantitative description of mesh dependence for the discretization of singularly perturbed nonconvex problems*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 1883–1898.

[14] A. Braides and C. I. Zeppieri, *Multiscale analysis of a prototypical model for the interaction between microstructure and surface energy*, Interfaces Free Bound., 11 (2009), pp. 61–118.

[15] X. Bresson and T. Laurent, *Asymmetric Cheeger cut and application to multi-class unsuprevised clustering*, CAM report, (2012).

[16] X. Bresson, T. Laurent, D. Uminsky, and J. H. von Brecht, *Convergence and energy landscape for cheeger cut clustering*, in Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), 2012, pp. 1385–1393.

[17] ——, *Multiclass total variation clustering*, in Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS), 2013, pp. 1421–1429.

[18] X. Bresson, X.-C. Tai, T. F. Chan, and A. Szlam, *Multi-class transductive learning based on l^1 relaxations of cheeger cut and mumford-shah-potts model*, Journal of Mathematical Imaging and Vision, 49 (2014), pp. 191–201.

[19] J. W. Cahn and J. E. Hilliard, *Free energy of a nonuniform system. i. interfacial free energy*, J. Chem. Ph.m, 28 (1958), pp. 258–267.

[20] A. Chambolle, A. Giacomini, and L. Lussardi, *Continuous limits of discrete perimeters*, M2AN Mathematical Modelling and Numerical Analysis, 44 (2010), pp. 207–230.

[21] S. Conti, I. Fonseca, and G. Leoni, *A Γ-convergence result for the two-gradient theory of phase transitions*, Comm. Pure Appl. Math., 55 (2002), pp. 857–936.

[22] S. Conti and B. Schweizer, *A sharp-interface limit for a two-well problem in geometrically linear elasticity*, Archive for Rational Mechanics and Analysis, 179 (2006), pp. 413–452.

[23] R. Cristoferi, I. Fonseca, A. Hagerty, and C. Popovici, *A homogenization result in the gradient theory of phase transitions*, Interf. Free Boundaries, 21 (2019), pp. 367–408.

[24] R. CRISTOFERI AND M. THORPE, *Large data limit for a phase transition model with the p-laplacian on point clouds*. To appear on European J. Appl. Math.

[25] G. DAL MASO, *An Introduction to* Γ*-Convergence*, Springer, 1993.

[26] E. DAVIS AND S. SETHURAMAN, *Consistency of modularity clustering on random geometric graphs*, arXiv preprint arXiv:1604.03993, (2016).

[27] N. DIRR, M. LUCIA, AND M. NOVAGA, Γ*-convergence of the Allen-Cahn energy with an oscillating forcing term*, Interfaces Free Bound., 8 (2006), pp. 47–78.

[28] N. DIRR, M. LUCIA, AND M. NOVAGA, *Gradient theory of phase transitions with a rapidly oscillating forcing term*, Asymptot. Anal., 60 (2008), pp. 29–59.

[29] M. M. DUNLOP, D. SLEPČEV, A. M. STUART, AND M. THORPE, *Large data and zero noise limits of graph-based semi-supervised learning algorithms*, In Preperation, (2018).

[30] S. ESEDOḠLU AND F. OTTO, *Threshold dynamics for networks with arbitrary surface tensions*, Communications on Pure and Applied Mathematics, 68 (2015), pp. 808–864.

[31] I. FONSECA AND S. MÜLLER, *Relaxation of quasiconvex functionals in* $\mathrm{BV}(\Omega, \mathbf{R}^p)$ *for integrands* $f(x,u,\nabla u)$, Archive for Rational Mechanics and Analysis, 123 (1993), pp. 1–49.

[32] I. FONSECA AND C. POPOVICI, *Coupled singular perturbations for phase transitions*, Asymptotic Analysis, 44 (2005), pp. 299–325.

[33] I. FONSECA AND L. TARTAR, *The gradient theory of phase transitions for systems with two potential wells*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 111 (1989), pp. 89–102.

[34] N. GARCÍA-TRILLOS, *Variational limits of k-nn graph based functionals on data clouds*, arXiv preprint arXiv:1607.00696, (2016).

[35] N. GARCÍA TRILLOS AND R. MURRAY, *A new analytical approach to consistency and overfitting in regularized empirical risk minimization*, European Journal of Applied Mathematics, 28 (2017), pp. 886–921.

[36] N. GARCÍA TRILLOS AND D. SLEPČEV, *On the rate of convergence of empirical measures in* ∞*-transportation distance*, Canadian Journal of Mathematics, 67 (2015), pp. 1358–1383.

[37] ——, *Continuum limit of total variation on point clouds*, Archive for Rational Mechanics and Analysis, 220 (2016), pp. 193–241.

[38] ——, *A variational approach to the consistency of spectral clustering*, Applied and Computational Harmonic Analysis, (2016).

[39] N. García Trillos, D. Slepčev, J. von Brecht, T. Laurent, and X. Bresson, *Consistency of cheeger and ratio graph cuts*, The Journal of Machine Learning Research, 17 (2016), pp. 6268–6313.

[40] M. E. Gurtin, *Some results and conjectures in the gradient theory of phase transitions*, in Metastability and incompletely posed problems, Springer, 1987, pp. 135–146.

[41] R. Kohn and P. Sternberg, *Local minimizers and singular perturbations*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 111 (1989), pp. 69–84.

[42] B. Marr, *How much data do we create every day? the mind-blowing stats everyone should read*. https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read.

[43] E. Merkurjev, T. Kostić, and A. L. Bertozzi, *An MBO scheme on graphs for classification and image processing*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1903–1930.

[44] L. Modica, *The gradient theory of phase transitions and the minimal interface criterion*, Archive for Rational Mechanics and Analysis, 98 (1987).

[45] L. Modica and S. Mortola, *Un esempio di Γ-convergenza*, Bollettino dell'Unione Matematica Italiana B, 14 (1977), pp. 285–299.

[46] N. C. Owen and P. Sternberg, *Nonconvex variational problems with anisotropic perturbations*, Nonlinear Analysis, 16 (1991), pp. 705–719.

[47] M. Penrose, *Random geometric graphs*, vol. 5 of Oxford Studies in Probability, Oxford University Press, Oxford, 2003.

[48] A. C. Ponce, *A new approach to Sobolev spaces and connections to Γ-convergence*, Calculus of Variations and Partial Differential Equations, 19 (2004), pp. 229–255.

[49] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Birkhäuser, 2015.

[50] O. Savin and E. Valdinoci, *Γ-convergence for nonlocal phase transitions*, Annales de l'Institut Henri Poincaré. Analyse Non Linéaire, 29 (2012), pp. 479–500.

[51] J. Shi and J. Malik, *Normalized cuts and image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (2000), pp. 888–905.

[52] D. Slepčev and M. Thorpe, *Analysis of p-laplacian regularization in semi-supervised learning*. To appear on SIAM Journal on Mathematical Analysis.

[53] ———, *Analysis of p-Laplacian regularization in semi-supervised learning*, arXiv preprint arXiv:1707.06213, (2017).

[54] A. SZLAM AND X. BRESSON, *A total variation-based graph clustering algorithm for cheeger ratio cuts*, UCLA CAM Report, (2009), pp. 1–12.

[55] ———, *Total variation and cheeger cuts*, in Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML), 2010, pp. 1039–1046.

[56] M. THORPE, S. PARK, S. KOLOURI, G. K. ROHDE, AND D. SLEPČEV, *A transportation L^p distance for signal analysis*, to appear in the Journal of Mathematical Imaging and Vision, arXiv preprint arXiv:1609.08669, (2017).

[57] M. THORPE AND D. SLEPČEV, *Transportation L^p distances: Properties and extensions*, In preparation, (2017).

[58] M. THORPE AND F. THEIL, *Asymptotic analysis of the Ginzburg-Landau functional on point clouds*, to appear in the Proceedings of the Royal Society of Edinburgh Section A: Mathematics, arXiv preprint arXiv:1604.04930, (2017).

[59] J. D. VAN DER WAALS, *The thermodynamics theory of capillarity under the hypothesis of a continuous variation of density*, Verhaendel Kronik. Akad. Weten. Amsterdam, 1 (1893), pp. 386–398.

[60] Y. VAN GENNIP AND A. L. BERTOZZI, *Γ-convergence of graph Ginzburg-Landau functionals*, Advances in Differential Equations, 17 (2012), pp. 1115–1180.

[61] C. VILLANI, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[62] C. VILLANI, *Optimal Transport: Old and New*, Springer, 2009.