Large data limit for a phase transition model with the $p$-Laplacian on point clouds

Citation for published version:
Cristoferi, R & Thorpe, M 2018, 'Large data limit for a phase transition model with the $p$-Laplacian on point clouds', European Journal of Applied Mathematics, pp. 1-47. https://doi.org/10.1017/S0956792518000645

Digital Object Identifier (DOI):
10.1017/S0956792518000645

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Peer reviewed version

Published in:
European Journal of Applied Mathematics

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Large Data Limit for a Phase Transition Model with the $p$-Laplacian on Point Clouds

Riccardo Cristoferi$^1$ and Matthew Thorpe$^2$

$^1$Department of Mathematics, Heriot-Watt University, Edinburgh, Scotland, UK EH14 4AS

$^2$Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, CB3 0WA, UK

July 2018

Abstract

The consistency of a nonlocal anisotropic Ginzburg-Landau type functional for data classification and clustering is studied. The Ginzburg-Landau objective functional combines a double well potential, that favours indicator valued functions, and the $p$-Laplacian, that enforces regularity. Under appropriate scaling between the two terms minimisers exhibit a phase transition on the order of $\varepsilon = \varepsilon_n$ where $n$ is the number of data points. We study the large data asymptotics, i.e. as $n \to \infty$, in the regime where $\varepsilon_n \to 0$. The mathematical tool used to address this question is $\Gamma$-convergence. It is proved that the discrete model converges to a weighted anisotropic perimeter.

1 Introduction

The analysis of big data is one of the most important challenges we currently face. A typical problem concerns partitioning the data based on some notion of similarity. When the method makes use of a (usually small) subset of the data for which there are labels then this is known as a classification problem. When the method only uses geometric features, i.e. there are are no a-priori known labels, then this is known as a clustering problem. We refer to both problems as labelling problems.

A popular method to represent the geometry of a given data set is to construct a graph embedded in an ambient space $\mathbb{R}^d$. Typically the labelling task is fulfilled via a minimization procedure. In the machine learning community, successfully implemented approaches include minimizing graph cuts and total variation (see, for instance, [4,10,13–16,39,47,48,50,51]).

Of capital importance for evaluating a labelling method is whether it is 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. Indeed, knowing whether a specific minimization strategy is an approximation of a limit (minimizing) object can help explain properties of the finite data algorithm. For a consistent methodology properties of the large data limit will be evident when a large, but finite, number of data points is being considered. 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.

This paper is part of an ongoing project aimed at justifying analytically the consistency of several models for soft labelling used by practitioners. Here we consider a generalization of the approach introduced by Bertozzi and Flenner in [7] (see also [17], for an introduction on this topic see [56]), where a Ginzburg-Landau (or Modica-Mortola, see [40,41]) type functional is used as the underlying energy to minimize in the context of the soft classification problem. The functional we consider is a discretisation of the non-local Ginzburg-Landau functional studied by Alberti and Bellettini [1,2] with the generalisation that we consider non-uniform densities and $\ell^p$ (rather than $\ell^2$) cost on finite differences. Our goal is to prove the consistency of the model.

There are multiple extensions to the approach we consider here; for instance our Ginzburg-Landau functional is based on the $p$-Laplacian, one can also consider the normalised $p$-Laplacian or the random walk Laplacian (see [42,47]). Further open problems concern the extention to multi-phase labelling (see [30]) and convergence of the associated gradient flows.

The paper is organized as follows: in the following subsection we define the discrete model, and in Subsection 1.2 we define the continuum limiting problem. The main results are given in Section 1.3 with the proofs presented in Sections 4 and 5. In Section 1.4 we give an overview on the related literature. In Sections 1.5 and 1.6 we include two examples with the purpose of demonstrating key properties of our functional; in particular how the choice of $p$ effects minimizers of our Ginzburg-Landau functional and an example to image segmentation. Section 2 contains some preliminary material we include for the convenience of the reader. Finally, Section 3 is devoted to the proofs of some technical results that are of interest in their own right, and are later used in the proofs in Section 4.

1.1 Finite Data Model

In the graph representation of a data set, vertexes are points $X_n := \{x_i\}_{i=1}^n \subset X$, where $X \subset \mathbb{R}^d$ is a connected, bounded and open set, with weighted edges $\{W_{ij}\}_{i,j=1}^n$, where each $W_{ij} \geq 0$ is meant to represent similarities between the vertexes $x_i$ and $x_j$, and in some sense encode the geometry. The larger $W_{ij}$ is the more similar the points $x_i$ and $x_j$ are and "the closer they are on the graph".

Let us consider the problem of partitioning a set of data into two classes. A partition of the set of points $X_n$ is a map $u : X_n \to \{-1, 1\}$, where $-1$ and $1$ represent the two classes. This is referred to as hard labelling, since $u$ can only assume a finite number of values. From the computational point of view it is preferable to work with functions whose values range in the whole interval $[-1, 1]$, i.e., labellings $u : X_n \to [-1, 1]$, thus allowing for a soft labelling. Labels that are close to $1$, or to $-1$, are supposed to be in the same class. The model used to obtain the binary classification should then force the labelling to be either $1$ or $-1$ when the number of data points is large.

In order to scale the weights on the edges of the graph we define $W_{ij}$ through a kernel $\eta : \mathbb{R}^d \to \mathbb{R}$. More precisely, we define the graph weights by $W_{ij} = \eta_\varepsilon(x_i - x_j) = \frac{1}{\varepsilon^d} \eta((x_i - x_j)/\varepsilon)$ where $\varepsilon$ controls the scale of interactions on the graph; in particular choosing $\varepsilon$ large implies the graph is dense, and choosing $\varepsilon$ small implies the graph is disconnected. Later assumptions, see Remark 1.4, imply that we scale $\varepsilon = \varepsilon_n$ such that the graph is eventually connected (with probability one).

We now introduce the discrete functional we are going to study.

Definition 1.1. For $p \geq 1$ and $n \in \mathbb{N}$ define the functional $G_n^{(p)} : L^1(X_n) \to [0, \infty)$ by

$$G_n^{(p)}(u) := \frac{1}{n^2 \varepsilon_n^2} \sum_{i,j=1}^n W_{ij} |u(x_i) - u(x_j)|^p + \frac{1}{\varepsilon_n n} \sum_{i=1}^n V(u(x_i)),$$
where
\[ W_{ij} := \eta_{\varepsilon_n}(x_i - x_j) := \frac{1}{\varepsilon_n} \eta \left( \frac{x_i - x_j}{\varepsilon_n} \right) \] (1)
and \( V \) is a double well potential (see Assumption (B4)).

The first term in \( G_n \) plays the role of penalising oscillations, intuitively one wants a labelling solution such that if \( x_i \) and \( x_j \) are close on the graph then the labels are also close. This term, when \( p = 2 \), can also be written as \( \frac{1}{\varepsilon_n} \langle u, Lu \rangle_{\mu_n} \) where \( L \) is the graph Laplacian. The second term penalises soft labellings. In particular, we assume that \( V(t) = 0 \) if and only if \( t \in \{ \pm 1 \} \) and \( V(t) > 0 \) for all \( t \neq \pm 1 \). Hence any soft labelling is given a penalty of \( \frac{1}{\varepsilon_n} \sum_{i=1}^{n} V(u(x_i)) \), as \( \varepsilon_n \to 0 \) this penalty blows up unless \( u \) takes the values \( \pm 1 \) almost everywhere.

The function \( \eta \) plays the role of a mollifier, and that explains the definition of \( \eta_{\varepsilon_n} \). Moreover, to justify the scaling \( \frac{1}{\varepsilon_n} \) we reason as follows: assume \( \eta \) has support contained in a ball, we get
\[ |u(x_i) - u(x_j)|^p \sim \varepsilon_n^p |\nabla u|^p. \]
So that, dividing by \( \varepsilon_n \) will give us the typical form of the singular perturbation used in the gradient theory of phase transitions (see [40]), namely
\[ \int X \varepsilon_n V(u) + \varepsilon_n^{p-1} |\nabla u|^p. \]

The consistency of the model is studied using \( \Gamma \)-convergence (see Section 2.4), a very important tool introduced by De Giorgi in the 70’s to understand the limiting behavior of a sequence of functionals (see [21]). This kind of variational convergence gives, almost immediately, convergence of minimizers.

1.2 Infinite Data Model

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

**Definition 1.2.** 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
\[ C(x, \nu) := \left\{ C \subset \nu^\perp : C \text{ is a } (d - 1)\text{-dimensional cube centred at } x \right\}. \]

For \( C \in C(x, \nu) \), we denote by \( v_1, \ldots, v_{d-1} \) 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 \( i = 1, \ldots, d - 1 \).

Finally, we consider the following space of functions:
\[ \mathcal{U}(C, \nu) := \left\{ u : \mathbb{R}^d \to [-1, 1] : u \text{ is } C \text{-periodic, } \lim_{y, \nu \to \infty} u(y) = 1, \text{ and } \lim_{y, \nu \to -\infty} u(y) = -1 \right\}. \]

We now define the limiting (continuum) model.

**Definition 1.3.** Let \( p \geq 1 \) and \( X \subset \mathbb{R}^d \) be open and bounded, and \( \rho \in L^\infty(X) \) a positive function. Define the functional \( G_\infty^{(p)} : L^1(X) \to [0, \infty] \) by
\[ G_\infty^{(p)}(u) := \left\{ \begin{array}{ll}
\int_{\partial^*\{u=1\}} \sigma^{(p)}(x, \nu_u(x)) \rho(x) \, d\mathcal{H}^{d-1}(x) & \text{if } u \in BV(X; \{ \pm 1 \}), \\
+\infty & \text{else,}
\end{array} \right. \]
where
\[ \sigma^{(p)}(x, \nu) := \inf \left\{ \frac{1}{H^{d-1}(C)} G^{(p)}(u, \rho(x), T_C) : C \in \mathcal{C}(x, \nu), u \in \mathcal{U}(C, \nu) \right\}, \]
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
\[ G^{(p)}(u, \lambda, A) := \lambda \int_A \int_{\mathbb{R}^d} \eta(h)|u(z+h)−u(z)|^p \, dh \, dz + \int_A V(u(z)) \, dz. \]
Here \( \partial^* \{ u = 1 \} \) denotes the reduced boundary of \( \{ u = 1 \} \) and \( \nu_x(x) \) is the measure theoretic exterior normal to the set \( \{ u = 1 \} \) at the point \( x \in \partial^* \{ u = 1 \} \) (see Definition 2.15).

Notice that the discrete functional \( G^{(p)}_{\infty} \) is nonlocal while the functional \( G^{(p)}_{\infty} \) is local. The minimization problem defining \( \sigma^{(p)} \) is called the cell problem and it is common in phase transitions problems (see related works in Section 1.4). Although not explicit, we have at least information on the form of the limiting functional: an anisotropic weighted perimeter. This shows that minimizers of \( G^{(p)}_{\infty} \) are sets \( E \subset X \) whose boundary \( \partial E \) (or, to be precise, the reduced boundary \( \partial^* E \)) will likely be in the region where \( \rho \) is small and orthogonal to directions \( \nu \) for which \( \sigma^{(p)}(x, \nu) \) is small.

Finally, we want to point out that one of the main issues we have to deal with is that, for each \( n \in \mathbb{N} \), the data set \( X_n \) is a discrete set, while in the limit the data is given by a probability measure \( \mu \) on the set \( X \), hence why we call \( G^{(p)}_{\infty} \) the continuum model. Thus, we need to compare functions (the labeling) defined on different sets. To do so we will implement the strategy introduced by García Trillos and Slepčev in [34], that consists in extending a function \( u : X_n \to \mathbb{R} \) to a function \( v : X \to \mathbb{R} \) in an optimal piecewise constant way. Optimal here is meant in the sense of optimal transportation. In particular, a sequence of functions \( \{ u_n \}_{n=1}^{\infty} \) with \( u_n \in L^1(X_n) \), is said to converge in the \( TL^1 \) topology to a function \( u \in L^1(X) \) if there exists a sequence \( \{ T_n \}_{n=1}^{\infty} \subset L^1(X; X_n) \) converging to the identity map in \( L^1(X) \) and with
\[ \mu(T_n^{-1}(B)) = \frac{1}{n} \# \{ x_i \in B : i = 1, 2, \ldots, n \} \]
for every Borel set \( B \subset X \), such that \( u_n \circ T_n \to u \) in \( L^1(X) \). We review the \( TL^1 \) topology in more detail in Section 2.2.

1.3 Main Results

This section is devoted to the precise statements of the main results of this paper.

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 \( \rho \) 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 \( \mu_n \) be the associated empirical measure (see Definition 2.1). Let \( \{ \epsilon_n \}_{n=1}^{\infty} \) be a positive sequence converging to zero and such that the following rate of convergence holds:

(A2) \( \frac{\text{dist}_{\infty}(\mu_n, \mu)}{\epsilon_n} \to 0 \), where \( \text{dist}_{\infty}(\mu_n, \mu) \) is the \( \infty \)-Wasserstein distance between the measures \( \mu_n \) and \( \mu \), see Definition 2.3.
Theorem 1.6. When \( x_i \stackrel{\text{i.i.d.}}{\sim} \mu \) then (with probability one), Assumption (A2) is implied by \( \varepsilon_n \gg \delta_n \), where \( \delta_n \) is defined in Theorem 2.10. Notice that for \( d \geq 3 \) this lower bound on \( \varepsilon_n \) ensures that the graph with vertices \( x_n \) and edges weighted by \( W_{ij} \) (see (1)) is eventually connected (see [44, 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 \( \mu_n \) converges to the uniform measure and the lower bound is given by \( \varepsilon_n \gg n^{-\frac{1}{d}} \).

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 \([-1, 1]\).

The assumptions on \( V \) imply that in the limit there are only two phases \( \pm 1 \). Assumption (B3) is used to establish compactness, in particular it is used to show that minimisers can be bounded in \( L^\infty \) by 1. The prototypical example of a function \( V : \mathbb{R}^d \to \mathbb{R} \) satisfying Assumptions (B1-4) is given by \( V(s) := (s^2 - 1)^2 \).

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

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

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

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

(C4) For all \( \delta > 0 \) there exists \( c_\delta, \alpha_\delta \) 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 1.5. Note that Assumptions (C3) and (C4) imply that \( \|\eta\|_{L^\infty} < \infty \) and, in particular, \( \int_{\mathbb{R}^d} \eta(x)|x| \, dx < \infty \). Indeed, given \( \delta > 0 \), it is possible to cover \( B(0, R_\eta) \) with a finite family \( \{\tilde{B}_\delta(x_1), \ldots, \tilde{B}_\delta(x_r)\} \) of sets of the form
\[
\tilde{B}_\delta(x_i) := \{\alpha_\delta z : |z - x_i| < \delta\}.
\]

Assumption (C2) is justified by the fact that \( \eta \) plays the role of an interaction potential. Finally, Assumption (C4) is a version of continuity of \( \eta \) 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 [54, Proposition 2.2].

The main result of the paper is the following theorem.

Theorem 1.6. Let \( p \geq 1 \) and assume Assumptions (A1-2), (B1-4) and (C1-4) are in force. Then, the following holds:

- (compactness) let \( u_n \in L^1(\mu_n) \) satisfy \( \sup_{n \in \mathbb{N}} \mathcal{G}_n^{(p)}(u_n) < \infty \), then \( u_n \) is relatively compact in \( TL^1 \) and each cluster point \( u \) has \( \mathcal{G}_\infty^{(p)}(u) < \infty \);

- (\( \Gamma \)-convergence) \( \Gamma \)-\( \lim_{n \to \infty} (TL^1) \mathcal{G}_n^{(p)} = \mathcal{G}_\infty^{(p)} \).
The proof of compactness is in Section 4.1 and the $\Gamma$-convergence is proved in Sections 4.2 and 4.3.

Since the proof of Theorem 1.6 is quite long, we briefly sketch here the main idea behind the $\Gamma$-convergence result. We approximately follow the method of [34] where the authors considered the continuum limit of total variation on point clouds. We will show the convergence of the discrete nonlocal functional $G^{(p)}_{\varepsilon_n}$ to the continuum local one $G^{(p)}_\infty$ via an intermediate nonlocal continuum functional $F^{(p)}_{\varepsilon_n}$ (defined in (4)). In particular, we will prove that:

(i) the functionals $F^{(p)}_{\varepsilon_n}$ $\Gamma$-converge in $L^1(X)$ to $G^{(p)}_\infty$, see Section 3 where we implement a strategy similar to the one of [1], where the authors considered the functional $F^{(p)}_{\varepsilon_n}$ with $\rho \equiv 1$ and $p = 2$,

(ii) it is possible to bound from below $G^{(p)}_{\varepsilon_n}$ with $F^{(p)}_{\varepsilon_n}$ (see (43)), where $\lim_{n \to \infty} \frac{\varepsilon_n}{\varepsilon} = 1$, from which the liminf inequality follows,

(iii) if $u \in BV(X; \{ \pm 1 \})$ we use the same recovery sequence $u_\varepsilon$ as in [1] to show $\limsup_{\varepsilon \to 0} F^{(p)}_{\varepsilon_n}(u_\varepsilon) \leq G^{(p)}_\infty$, after which we can get an upper bound of $G^{(p)}_{\varepsilon_n}(u_n)$ in terms of $F^{(p)}_{\varepsilon_n}(u_\varepsilon)$ where $u_n : X_n \to [0, \infty)$ is defined at each $x_i \in X_n$ as a suitable average of $u_{\varepsilon_n}$ around the point $x_i$ and $\lim_{n \to \infty} \frac{\varepsilon_n}{\varepsilon} = 1$. This will give us the limsup inequality.

Similarly, the compactness property follows by comparing $G^{(p)}_{\varepsilon_n}$ with the intermediary functional $F^{(p)}_{\varepsilon_n}$.

As an application of the Theorem 1.6, we consider the functional $G^{(p)}_{\varepsilon_n}$ with a data fidelity term.

**Definition 1.7.** Let $k_n : X_n \times \mathbb{R} \to \mathbb{R}$ and $k_\infty : X \times \mathbb{R} \to \mathbb{R}$. Define the functionals $K_n : TL^1(X_n) \to \mathbb{R}$ and $K_\infty : TL^1(X) \to \mathbb{R}$ by

$$K_n(u, \nu) = \begin{cases} \frac{1}{n} \sum_{i=1}^{n} k_n(x_i, u(x_i)) & \text{if } \nu = \mu_n, \\ +\infty & \text{else}, \end{cases}$$

and

$$K_\infty(u, \nu) = \begin{cases} \int_X k_\infty(x, u(x)) \rho(x) \, dx & \text{if } \nu = \mu, \\ +\infty & \text{else}, \end{cases}$$

respectively.

We make the following assumptions on $k_n, k_\infty$:

(D1) $k_n \geq 0, k_\infty \geq 0$.

(D2) There exist $\beta > 0$ and $q \geq 1$ such that $k_n(x, u) \leq \beta (1 + |u|^q)$, for all $n \in \mathbb{N}$ and almost all $x \in X_n$.

(D3) For almost every $x \in X$ the following holds: let $u_n \to u$ be a converging real valued sequence and $x_n \to x$, then

$$\lim_{n \to \infty} k_n(x_n, u_n) = k_\infty(x, u).$$

**Remark 1.8.** For example, we can use this form of $K_n, K_\infty$ to include a data fidelity term in a specific subset of $X$. Let $B \subset X$ be an open set with $\text{Vol}(B) > 0$ and $\text{Vol}(\partial B) = 0$. Let $\lambda_n \geq 0$ with $\lambda_n \to \lambda$ as $n \to \infty$. Let $\sup_{n \in \mathbb{N}} \|y_n\|_{L^\infty} < \infty$ and such that $y_n(x_i) \to y_\infty(x)$ for almost every $x \in X$ and any sequence $x_{i_n} \to x$. Define

$$k_n(x, u) := \begin{cases} \lambda_n |y_n(x) - u|^q & \text{in } B \cap X_n, \\ 0 & \text{on } X_n \setminus B, \end{cases}$$
Then \( k_n \) and \( k_\infty \) satisfy Assumptions (D1-3). Indeed, Assumption (D1) follows directly from the definition of the fidelity terms, while Assumption (D3) holds thanks to continuity and the fact that \( \text{Vol}(\partial B) = 0 \). Finally, in order to prove Assumption (D2) we simply notice that

\[
\begin{align*}
    k_n(x, u) &= \lambda_n |y_n(x) - u|^q \leq \sup_{n \in \mathbb{N}} \lambda_n 2^{q-1} \left( \|y_n\|_{L^\infty}^q + |u|^q \right) \
    &\leq \beta (1 + |u|^q) \text{ for some } \beta > 0.
\end{align*}
\]

for some \( \beta > 0 \).

We now consider the minimisation problem

\[
\text{minimise } G_n^{(p)}(u) + K_n(u) \text{ over } u \in L^1(X_n).
\]

**Corollary 1.9.** In addition to Assumptions (A1-3), (B1-2), (C1-4), (D1-3), assume that for the same density function \( \Gamma \) a similar \( 1.4 \) Related Works

We now consider the minimisation problem

\[
\text{minimise } G_n^{(p)}(u) + K_n(u) \text{ over } u \in L^1(X_n).
\]

**Corollary 1.9.** In addition to Assumptions (A1-3), (B1-2), (C1-4), (D1-3), assume that for the same \( q \geq 1 \) as in Assumption (D2) there exists \( \tau, R_V > 0 \) such that for all \( |s| \geq R_V \) that \( V(s) \geq \tau |s|^q \). Then any sequence of almost minimizers of \( G_n^{(p)} + K_n \) is compact in \( TL^1 \). And furthermore, any cluster point of almost minimizers is a minimizer of \( G_\infty^{(p)} + K_\infty \) in \( L^1(X) \).

We prove the corollary in Section 5.

Finally, we comment on the hypothesis \( \rho \geq c_1 > 0 \). If it is omitted, we can still get the following result:

**Corollary 1.10.** Let \( p \geq 1 \) and assume Assumptions (A2), (B1-3) and (C1-4) are in force and that \( \rho \in [0, c_2] \), for some \( c_2 < \infty \). Set \( X_+ := \{ x \in X : \rho(x) > 0 \} \) and define the functional \( \tilde{G}_\infty^{(p)} : L^1(X) \to [0, +\infty] \) as

\[
\tilde{G}_\infty^{(p)}(u) := \begin{cases}
    \int_{\partial^* \{ u = 1 \}} \sigma^{(p)}(x, \nu_u(x)) \rho(x) dH^{d-1}(x) & \text{if } u \in BV_{\text{loc}}(X_+; \{ \pm 1 \}), \\
    +\infty & \text{else},
\end{cases}
\]

where \( BV_{\text{loc}}(X_+; \{ \pm 1 \}) \) denotes the space of functions \( u \in L^1(X; \{ \pm 1 \}) \) such that \( u \in BV(K; \{ \pm 1 \}) \) for any compact set \( K \subset X_+ \). Then, the following holds:

- (compactness) for any compact set \( K \subset X_+ \) we have that any sequence \( \{ u_n \}_{n=1}^\infty \subset L^1(K; \mu_n) \) satisfying \( \sup_{n \in \mathbb{N}} G_n^{(p)}(u_n) < \infty \) is relatively compact in \( TL^1 \) and each cluster point \( u \) has \( \tilde{G}_\infty^{(p)}(u) < \infty \);

- (\( \Gamma \)-convergence) \( \Gamma \)-\( \lim_{n \to \infty} (TL^1) G_n^{(p)} = \tilde{G}_\infty^{(p)} \).

We omit a rigorous proof of the corollary and just explain the details. Indeed, for fixed a compact set \( K \subset X_+ \), the continuity of \( \rho \) implies that \( \min_K \rho \geq c_1 > 0 \). Thus the result of Corollary 1.10 follows by applying Theorem 1.6.

### 1.4 Related Works

The functional \( G_n^{(p)} \) in the case \( p = 1 \) has been considered by the second author and Theil in [54], where a similar \( \Gamma \)-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 [55] 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 $T L^p$ topology, as introduced by García Trillos and Slepčev [34], provides a notion of convergence upon which the $\Gamma$-convergence framework can be applied. This method has now been applied in many works, see, for instance, [20, 23, 31, 32, 34, 36, 48, 54]. Further studies on this topology can be found in [34, 35, 52, 53].

The literature on phase transitions problems is quite extensive. Here we just recall some of the main results, starting from the pioneering work [41] of Modica and Mortola and of Mortola [40] (see also Sternberg [49]), where the scalar isotropic case has been studied. The vectorial case has been considered by Kohn and Sternberg in [37], Fonseca and Tartar in [28] and Baldo [5]. A study of the anisotropic case has been carried out by Bouchitté [8] and Owen [43] in the scalar case, and by Barroso and Fonseca [6] and Fonseca and Popovici [27] in the vectorial case.

Nonlocal approximations of local functionals of the perimeter type go back to the work [1] of Alberti and Bellettini (see also [2]). Several variants and extensions have been considered since then (see, for instance, Savin and Valdinoci [46] and Esedoğlu and Otto [25]). In particular, nonlocal functionals have been used by Brezis, Bourgain and Mironescu in [9] to characterize Sobolev spaces (see also the work [45] of Ponce).

Approximations of (anisotropic) perimeter functionals via energies defined in the discrete setting have been carried out by Braides and Yip in [12] and by Chambolle, Giacomini and Lussardi in [18].

1.5 The Choice of $p$

By allowing for any $p \geq 1$ in the definition of our functionals $\mathcal{G}^{(p)}_n$ (rather than $p = 2$ usually considered) we allow for greater flexibility. In particular the choice of $p$ has an important consequence regarding the balance between minimising the length of the perimeter and avoiding regions of high density. As for Euclidean norms, as $p \to \infty$ the Dirichlet energy $E^{(p)}_n$, defined by
\[
E^{(p)}_n(u) = \frac{1}{\varepsilon_n} \frac{1}{n} \sum_{i,j=1}^{n} W_{ij} |u(x_i) - u(x_j)|^p + \frac{1}{\varepsilon_n} \frac{1}{n} \sum_{i=1}^{n} V^p(u(x_i)) \] converges as $p \to \infty$ to $E^{(\infty)}_n$, defined by
\[
E^{(\infty)}_n(u) = \max_{i,j=1,\ldots,n} W_{ij} |u(x_i) - u(x_j)|.
\]

In fact the convergence also holds in a variational sense [24, 38]. It is therefore of no surprise that the same holds in the Ginzburg-Landau setting (where the normalisation has been adjusted so that terms are of order 1 with respect to $p$).

**Proposition 1.11.** Fix $n \in \mathbb{N}$, and assume $V : \mathbb{R} \to [0, \infty)$ is continuous and weights $W_{ij} \geq 0$ for all $i, j = 1, \ldots, n$. Let
\[
\tilde{\mathcal{G}}^{(p)}_n(u) = \frac{1}{\varepsilon_n n^2} \sum_{i,j=1}^{n} W^{p} |u(x_i) - u(x_j)|^p + \frac{1}{\varepsilon_n n} \sum_{i=1}^{n} V^p(u(x_i)) \]
\[
\mathcal{G}^{(\infty)}_n(u) = \max_{i,j=1,\ldots,n} W_{ij} |u(x_i) - u(x_j)|, \quad \frac{1}{\varepsilon_n} \max_{i=1,\ldots,n} V(u(x_i)) \]

Then,
\[
\Gamma\lim_{p \to \infty} \tilde{\mathcal{G}}^{(p)}_n = \mathcal{G}^{(\infty)}_n.
\]
we assume $L^0(X_n) \ni u_p \to u \in L^0(X_n)$ then,

$$
\tilde{G}_n^{(p)}(u_p) \geq \left[ \frac{1}{\varepsilon_n n^2} W^p_{rt}|u_p(x_r) - u_p(x_t)|^p + \frac{1}{\varepsilon_n n^p} V^p(u_p(x_\ell)) \right]^{\frac{1}{p}}
$$

$$
\geq \max \left\{ \frac{1}{\varepsilon_n n^\frac{p}{2}} W^p_{rt}|u_p(x_r) - u_p(x_t)|, \frac{1}{\varepsilon_n n^p} V(u_p(x_\ell)) \right\}
$$

$$
\to \max \left\{ W^p_{rt}|u(x_r) - u(x_t)|, \frac{1}{\varepsilon_n} V(u(x_\ell)) \right\}
$$

for any $r, t, \ell \in \{1, \ldots, n\}$. Choosing $(r, t) = \arg\max_{i,j} W_{ij}|u(x_i) - u(x_j)|$ and $\ell = \arg\max_{i} V(u(x_i))$ implies

$$
\liminf_{p \to \infty} \tilde{G}_n^{(p)}(u_p) \geq G_n^{(\infty)}(u).
$$

For the recovery sequence we consider the sequence $u_p = u$ then,

$$
\tilde{G}_n^{(p)}(u_p) \leq \max \left\{ \frac{1}{\varepsilon_n} \max_{i,j=1,\ldots,n} W_{ij}|u(x_i) - u(x_j)|, \frac{1}{\varepsilon_n} \max_{i=1,\ldots,n} V(u(x_i)) \right\} \to G_n^{(\infty)}(u)
$$

as required. \(\square\)

There are several ways we could have renormalised $G_n^{(p)}$ to obtain a well defined limit as $p \to \infty$. We choose a normalisation $\tilde{G}_n^{(p)}$ that closely resembled $G_n^{(p)}$ and such that the limit $G_n^{(\infty)}$ will exhibit a phase transition as $n \to \infty$. In particular, the limit functional $G_n^{(\infty)}$ is chosen so that minimisers $u_n$ have the property that $u_n(x) \to \{\pm 1\}$ for almost every $x$, and $\max_{i,j} W_{ij} |u_n(x_i) - u_n(x_j)| = O(1)$.

The important property to note is that the $p = \infty$ limit depends only on the data through its support. In particular, if one considers the semi-supervised learning problem (similar to Remark 1.8 with $\lambda_n = \infty$):

$$
\text{minimise } \tilde{G}_n^{(p)}(u) \text{ subject to } u(x_i) = y_i \text{ for } x_i \in X'
$$

then the density of the data in $X \setminus X'$ is irrelevant for $p = \infty$.

To illustrate this we consider a simple example. We consider a density $\rho$ as in Figure 1 in particular the density is given by $\rho(x) \propto \phi(x_1)$ where $\phi$ is the density of a normal random variable, with mean at 0 and standard deviation 0.25, and $X$ is the bean shape. The regions of labelled data are chosen to be the balls $B((-0.3, 0), 0.08)$ and $B((0.3, 0), 0.08)$, with the first ball labelled $-1$ and the second ball labelled $+1$. The distribution is chosen so that the density is largest where the bean is narrowest. The optimal partitioning, given by minimizing $G_n^{(p)}$ is therefore a trade off between minimizing the length of the perimeter of the partitioning and avoiding high density regions. We plot the optimal partitioning for two choices of $p$, $p = 2$ and $p = 100$, using the same data set. We see that for $p = 2$ the partitioning avoids the region of high density at the cost of increasing the length of the boundary. On the other hand, when $p = 100$ the partitioning is much closer to the centre of the bean where the length of the partitioning is minimized.

### 1.6 Application to Image Segmentation

Let us consider segmentation in images. Let $x_i \in [0, 1]^2$ denote the location of pixels (in particular $X_n = \{x_i\}_{i=1}^n$ form a regular grid with distance $\frac{1}{\sqrt{n}}$ between neighbouring pixels). Let $y_i \in \mathbb{R}^3$ be the RGB
The density of the data generating distribution $\rho$.

(b) The minimiser of $G^{(2)}_n$.

(c) The minimiser of $G^{(100)}_n$.

Figure 1: How the choice of $p$ affects the trade-off between avoiding high density regions and minimising the length of the classification boundary. The experiment setup is described in Section 1.5 and we computed minimisers using a gradient flow (known as the Allen-Cahn equation).

values of a given image at the pixel located at $x_i$. We assume $I \subset \{1, \ldots, n\}$ indexes the a-priori labelled pixels and $f : \{x_i\}_{i \in I} \to \mathbb{R}$ are the given labels. We wish to label the remaining pixels $\{x_i\}_{i \notin I}$. Define

$$W_{i,j} = \frac{1}{\varepsilon^2} \Phi(|y_i - y_j|) \chi_{|x_i - x_j| \leq \varepsilon}$$

where $\Phi : \mathbb{R} \to \mathbb{R}$ is a decreasing function and $\varepsilon$ determines the number of neighbouring pixels we connect, e.g. $\varepsilon = \frac{1}{\sqrt{n}}$ connects the four neighbouring pixels, $\varepsilon = \frac{\sqrt{2}}{\sqrt{n}}$ connects the eight (including diagonals) neighbouring pixels, etc. Note that we are thinking of this problem in 2D where the kernel $\eta$ depends not just on the spatial difference but also on the values of pixels, i.e. $\eta(x_i - x_j) = \eta(x_i - x_j; y_i, y_j)$. This falls slightly outside of our theoretical framework as the function $\eta$ now depends on more than just the difference between $x_i$ and $x_j$.

The segmentation is defined by minimising

$$G^{(p)}_n(u) + K_n(u) = \lambda \frac{1}{2} \sum_{i \in I} |f(x_i) - u(x_j)|^2 + \frac{1}{\varepsilon n^2} \sum_{i,j \in \{1, \ldots, n\}} W_{ij} |u(x_i) - u(x_j)|^p + \frac{1}{\varepsilon n} \sum_{i=1}^n V(u(x_i)).$$

This setup is similar to Bertozzi and Flenner [7] and Calatroni, van Gennip, Schönlieb, Rowland and Flenner [17] however they embed the problem into a much higher dimensional space. In particular, in [7,17] the vertex corresponding to pixel $i$ is the vector consisting of the RGB values of the $M$ nearest neighbours, so that the graph is embedded in $\mathbb{R}^{3M}$ compared to the ambient dimension in our setup which is $\mathbb{R}^2$. We refer to methods in [7,17] as the segmentation in colour space method and our method as segmentation in pixel space.

It is not within the scope of this paper to comprehensively compare the differences in approach (in fact we note that the theoretical results in this paper apply to [7,17]). However, we point out that the method proposed here has the advantage that the parameter $\varepsilon$ directly controls the number of pixels connected which can make it easier for parameter selection. Choosing the parameter $\varepsilon$ in the colour space segmentation method is not as clear as one needs to compute the connectivity radius of the graph.

On the other hand, the pixel space segmentation method proposed here is spatially local which can make it difficult for similar colours that are spatially separated to be placed in the same class. In the colour space segmentation method similar colours are close on the graph regardless of the pixel locations.

In Figure 2 we plot an example of an image segmented into background and foreground. Given the image in Figure 2(a), where regions have been labelled either background (marked green), or foreground (marked
Figure 2: An image segmentation example with experimental setup as described in Section 1.6. The image is from the Microsoft image database and also appeared in [7].

(a) Input images with a-priori given labels marked. (b) The minimiser of $G_n^{(2)} + K_n$.

We construct a graph as described above with

$$\Phi(t) = \exp\left(-\frac{t^2}{\tau}\right), \quad \tau = 5 \times 10^{-4}$$

and choose $\varepsilon = \frac{1}{\sqrt{n}}$ and $p = 2$. Using a gradient flow to minimise $G_n^{(p)} + K_n$ we produce the output given in Figure 2(b). Since $\varepsilon$ determines the number of connections of each pixel it was easy to select parameters with a desirable output.

2 Background

2.1 Notation

In the following $\chi_E$ will denote the characteristic function of a set $E \subset \mathbb{R}^d$, while $\text{Vol}(E) = \mathcal{L}^d(E)$ will denote its $d$-dimensional Lebesgue measure and $\mathcal{H}^{d-1}(E)$ its $(d-1)$-Hausdorff measure. Moreover, with $B(x, r)$ we will denote the ball centered at $x \in \mathbb{R}^d$ with radius $r > 0$ and we set $S_{d-1} := \partial B(0,1)$. The identity map will be denoted by $\text{Id}$.

Given an open set $X \subset \mathbb{R}^d$, we define the space

$$\mathcal{P}(X) := \{ \text{Radon measures } \mu \text{ on } X \text{ with } \mu(X) = 1 \}.$$ 

Given a set of data points $\{x_i\}_{i=1}^n$ we define the empirical measure as follows.

**Definition 2.1.** For all $n \in \mathbb{N}$ let $X_n := \{x_i\}_{i=1}^n$ be a set of $n$ random variables. We define the empirical measure $\mu_n$ as

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i},$$

where $\delta_{x_i}$ denotes the Dirac delta centered at $x$.

We state our results in terms of a general sequence of empirical measures $\mu_n$ that converge weak* to some $\mu \in \mathcal{P}(X)$. An important special case is when $x_i$ are independent and identically distributed (which we abbreviate to iid) from $\mu$.

**Remark 2.2.** When $x_i \overset{\text{iid}}{\sim} \mu$ then $\mu_n$ converges, with probability one, to $\mu$ weakly* in the sense of measures, see for example [22, Theorem 11.4.1], (and we write $\mu_n \overset{w^*}{\rightharpoonup} \mu$), i.e.

$$\int_X \varphi \, d\mu_n \rightarrow \int_X \varphi \, d\mu$$
as \( n \to \infty \), for all \( \varphi \in C_c(X) \).

We write \( L^p(X, \mu; Y) \) for the space of \( L^p \) integrable, with respect to \( \mu \), functions from \( X \) to \( Y \). We will often suppress the \( Y \) dependence and just write \( L^p(X, \mu) \). Moreover, if \( \mu = \mathcal{L}^d \) then we will often write \( L^p(X) = L^p(X, \mu) \). If \( \mu = \mu_n \) is the empirical measure we also write \( L^p(X_n) = L^p(X, \mu_n) \).

### 2.2 Transportation theory

In this section we collect the 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 and \( X_n \subset X \) is a finite set of points. This is fundamental in stating our \( \Gamma \)-convergence result (Theorem 1.6). The \( TL^p \) space was introduced in [34] and consists of comparing \( w \) and a piecewise constant extension of the function \( u \) in \( L^p \). In particular, we take a map \( T : X \to X_n \) and we consider the function \( v : X \to \mathbb{R} \) defined as \( v := u \circ T \). In order that this defines a metric one needs to impose conditions on \( T \), the natural conditions are that \( T \) "matches the measure \( \mu \) with \( \mu_n \" and is optimal in the sense that matching moves as little mass as possible (see Theorem 2.10). This will be done by using the optimal transport distance that we recall now (see also [57] for background on optimal transport and [34, 53] for a further description of the \( TL^p \) space).

**Definition 2.3.** Let \( X \subset \mathbb{R}^d \) be an open set and let \( \mu, \lambda \in \mathcal{P}(X) \). We define the set of couplings \( \Gamma(\mu, \lambda) \) between \( \mu \) and \( \lambda \) as

\[
\Gamma(\mu, \lambda) := \{ \pi \in \mathcal{P}(X \times X) : \pi(A \times X) = \mu(A), \pi(X \times A) = \lambda(A), \text{ for all measurable } A \subset X \}.
\]

For \( p \in [1, +\infty] \), we define the \( p \)-Wasserstein distance between \( \mu \) and \( \lambda \) as follows:

- when \( 1 \leq p < \infty \),
  \[
  \text{dist}_p(\mu, \lambda) := \inf \left\{ \left( \int_{X \times X} |x - y|^p \, d\pi(x, y) \right)^{\frac{1}{p}} : \pi \in \Gamma(\mu, \lambda) \right\},
  \]

- when \( p = \infty \),
  \[
  \text{dist}_\infty(\mu, \lambda) := \inf \left\{ \text{esssup}_\pi \left\{ |x - y| : (x, y) \in X \times X \right\} : \pi \in \Gamma(\mu, \lambda) \right\},
  \]

where \( \text{esssup}_\pi \) denotes the essential supremum with respect to the measure \( \pi \).

**Remark 2.4.** The infimum problems in the above definition are known as the Kantorovich optimal transport problems and the distance is commonly called the \( \text{dist}_p \) Wasserstein distance or sometimes the earth movers distance. It is possible to see (see [57]) that the infimum is actually achieved. Moreover, the metric \( \text{dist}_p \) is equivalent to the weak* convergence of probability measures \( \mathcal{P}(X) \) (plus convergence of \( p \)th moments).

We now consider the case we are interested in: take \( \mu \in \mathcal{P}(X) \) with \( \mu = \rho \mathcal{L}^d \) (where \( \mathcal{L}^d \) is the \( d \)-dimensional Lebesgue measure on \( \mathbb{R}^d \)) and assume the density \( \rho \) is such that \( 0 < c_1 \leq \rho \leq c_2 < \infty \). In this case the Kantorovich optimal transport problem is equivalent to the Monge optimal transport problem (see [29]). In particular, for \( p \in [1, +\infty] \) it holds that

\[
\text{dist}_p(\mu, \lambda) = \min \left\{ \|\text{Id} - T\|_{L^p(X, \mu)} : T : X \to X \text{ Borel, } T_\# \mu = \lambda \right\},
\]

where

\[
\|\text{Id} - T\|_{L^p(X, \mu)}^p := \int_X |x - T(x)|^p \rho(x) \, dx.
\]
and we define the push forward measure $T\#\mu \in \mathcal{P}(X)$ as $T\#\mu(A) := \mu(T^{-1}(A))$ for all $A \subset X$. In the case $p = +\infty$ we get
\[
\text{dist}_{\infty}(\mu, \lambda) = \inf \left\{ \|\text{Id} - T\|_{L^{\infty}(X, \mu)} : T : X \to X \text{ Borel}, \ T\#\mu = \lambda \right\},
\]
where
\[
\|\text{Id} - T\|_{L^{\infty}(X, \mu)} := \text{ess sup}_{x} \rho(x)[x - T(x)].
\]
A map $T$ is called a transport map between $\mu$ and $\lambda$ if $T\#\mu = \lambda$.

Throughout the paper we will assume the empirical measures $\mu_n$ converges weakly* to $\mu$ (see Remark 2.2 for iid samples) so by Remark 2.4 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
\[
\lim_{n \to \infty} \|\text{Id} - T_n\|_{L^p(X, \mu)}^p = 0.
\]
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 2.5.** 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 $\mu_n$ converges weak* to $\mu$. We say that $u_n \rightharpoonup w$ in $TL^p(X)$, and we write $u_n \rightharpoonup TL^p w$, if there exists a sequence of stagnating transport maps $\{T_n\}_{n=1}^{\infty}$ between $\mu$ and $\mu_n$ such that
\[
\|u_n - w\|_{L^p(X, \mu)} \to 0,
\]
as $n \to \infty$, where $v_n := u_n \circ T_n$.

**Remark 2.6.** One can show that if (2) holds for one sequence of stagnating maps, then it holds for all sequences of stagnating maps [34, Proposition 3.12]. Moreover, since $\rho$ is bounded above and below it holds
\[
\|v_n - \text{Id}\|_{L^p(X, \mu)} \to 0 \quad \Leftrightarrow \quad \|v_n - \text{Id}\|_{L^p(X)} \to 0.
\]

We have introduced $TL^p$ convergence $u_n \rightharpoonup TL^p u$ by defining transport maps $T_n : X \to X_n$ which "optimally partition" the space $X$ after which we define a piecewise constant extension of $u_n$ to the whole of $X$. This constructionist approach is how we use $TL^p$ convergence in our proofs. However, this description hides the metric properties of $TL^p$. We briefly mention here the metric structure which characterises the convergence given in Definition 2.5. We define the $TL^p(X)$ space as the space of couplings $(u, \mu)$ 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
\[
d_{TL^p}((u, \mu), (v, \lambda)) := \min_{\pi \in \Gamma(\mu, \nu)} \left(\int_{X^2} |x - y|^p + |u(x) - v(y)|^p \, d\pi(x, y)\right)^{\frac{1}{p}}
\]
\[
= \inf_{T\#\mu = \lambda} \left(\int_{X} |x - T(x)|^p + |u(x) - v(T(x))|^p \, d\mu(x)\right)^{\frac{1}{p}},
\]
or for $p = +\infty$ by
\[
d_{TL^\infty}((u, \mu), (v, \lambda)) := \inf_{\pi \in \Gamma(\mu, \nu)} \left(\text{ess inf}_{\pi} \{ |x - y| + |u(x) - v(y)| : (x, y) \in X \times X\}\right)
\]
\[
= \min_{T\#\mu = \lambda} \left(\text{ess inf}_{\mu} \{ |x - T(x)| + |u(x) - v(T(x))| : x \in X\}\right).
\]
Proposition 2.7. 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 $\mu_n \rightharpoonup \mu$ and there exists a sequence of stagnating transport maps $\{T_n\}_{n=1}^{\infty}$ between $\mu$ and $\mu_n$ such that $\|u_n \circ T_n - u\|_{L^p(X, \mu)} \to 0$.

The proof is given in [34, Remark 3.4 and Proposition 3.12]. Note that Definition 2.5 characterises $TL^p$ convergence.

In order to be able to write the discrete functional we will need the following result.

Lemma 2.8. Let $\lambda \in \mathcal{P}(X)$ and let $T : X \to X$ be a Borel map. Then, for any $u \in L^1(X, \lambda)$ it holds

$$\int_X u \, dT#\lambda = \int_X u \circ T \, d\lambda.$$ 

Proof. Let $s : X \to \mathbb{R}$ be a simple function. Write $s = \sum_{i=1}^{k} a_i \chi_{U_i}$.

Then

$$\int_X s \, dT#\lambda = \sum_{i=1}^{k} a_i T#\lambda(U_i) = \sum_{i=1}^{k} a_i \lambda(T^{-1}(U_i)) = \int_X s \circ T \, d\lambda.$$ 

The result then follows directly from the definition of the integral. \qed

Remark 2.9. Applying the above result to the empirical measures $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ and $u \in L^1(X)$ we get

$$\frac{1}{n} \sum_{i=1}^{n} u(x_i) = \int_X v_n(x) \, d\mu(x),$$

where $v_n := u \circ T_n$ for any $T_n$ such that $(T_n)#\mu = \mu_n$.

In [33] the authors, García Trillos and Slepče, obtain the following rate of convergence for a sequence of stagnating maps. This is of crucial importance for applying the results of this paper to the iid setting.

Theorem 2.10. Let $X \subset \mathbb{R}^d$ be a bounded, connected and open set with Lipschitz boundary. Let $\mu \in \mathcal{P}(X)$ be of the form $\mu = \rho \mathcal{L}^d$ with $0 < c_1 \leq \rho \leq c_2 < \infty$. Let $\{x_i\}_{i=1}^{\infty}$ be a sequence of independent and identically distributed random variables distributed on $X$ according to the measure $\mu$, and let $\mu_n$ be the associated empirical measure. Then, there exists a constant $C > 0$ such that, with probability one, there exists a sequence $\{T_n\}_{n=1}^{\infty}$ of maps $T_n : X \to X$ with $(T_n)#\mu = \mu_n$ and

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

where

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

Remark 2.11. The proof for $d = 1$ is simpler and follows from the law of iterated logarithms. Notice that the connectedness of $X$ is essential in order to get the above result.

By the above theorem our main result, Theorem 1.6 holds with probability one when $x_i \iid \mu$ and the graph weights are scaled by $\varepsilon_n$ with $\varepsilon_n \gg \delta_n$. 

14
2.3 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 2.12.** Let $E \subset \mathbb{R}^d$ with $\text{Vol}(E) < \infty$ and let $X \subset \mathbb{R}^d$ be an open set. We say that $E$ has finite perimeter in $X$ if

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

**Remark 2.13.** 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 $X$ such that

$$\int_{\mathbb{R}^d} \varphi \, dD\chi_E = \int_E \text{div}\varphi \, dx$$

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

**Definition 2.14.** Let $X \subset \mathbb{R}^d$ be an open set and let $u \in L^1(X; \{\pm 1\})$ with $\|u\|_{L^1(X)} < \infty$. We say that $u$ is of bounded variation in $X$, and we write $u \in BV(X; \{\pm 1\})$, if $\{u = 1\} := \{x \in X : u(x) = 1\}$ has finite perimeter in $X$.

**Definition 2.15.** 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 2.16.** 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) $D\chi_E = \nu_E \mathcal{H}^{d-1} \setminus \partial^* E$,

(iii) the reduced boundary $\partial^* E$ is $\mathcal{H}^{d-1}$-rectifiable, i.e., there exists 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 2.17. Using the above result it is possible to prove that (see [26])

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

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

The construction of the recovery sequences in Section 3.4 and Section 4.3 will be done for a special class of functions, that we introduce now.

**Definition 2.18.** We say that a function \( u \in L^1(X; \{ \pm 1 \}) \) is polyhedral if \( u = \chi_E - \chi_{X \setminus E} \), where \( E \subset X \) is a set whose boundary is a Lipschitz manifold contained in the union of finitely many affine hyperplanes. In particular, \( u \in BV(X, \{ \pm 1 \}) \).

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 2.19.** Let \( u \in BV(X; \{ \pm 1 \}) \). Then there exists a sequence \( \{ u_n \}_{n=1}^{\infty} \subset BV(X; \{ \pm 1 \}) \) of polyhedral functions such that \( u_n \to u \) in \( L^1(X) \) and \( |Du_n|(X) \to |Du|(X) \). In particular \( Du_n \rightharpoonup^* Du \).

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

**Theorem 2.20.** 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 \( D \chi_{E_n} \rightharpoonup^* D \chi_E \) 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

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

Remark 2.21. A set of finite perimeter can have a very wild reduced boundary. Since the limiting energy density \( \sigma^{(p)} \) depends on the exterior theoretic normal \( \nu \), it would be very difficult to build an explicit recovery sequence for a set \( \{ u = 1 \} \) that is only assumed to have finite perimeter. For this reason, we consider the family of polyhedral sets, for which we provide an explicit construction, and then we use the density result Theorem 2.19 together with Theorem 2.20 to address the general case.

### 2.4 \( \Gamma \)-convergence

We recall the basic notions and properties of \( \Gamma \)-convergence (in metric spaces) we will use in the paper (for a reference, see [11, 19]).

**Definition 2.22.** Let \((A, d)\) be a metric space. We say that \( F_n : A \to [-\infty, +\infty] \) \( \Gamma \)-converges to \( F : A \to [-\infty, +\infty] \), and we write \( F_n \Gamma \rightharpoonup_d F \) or \( F = \Gamma \lim(d)_{n \to \infty} F_n \), 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

\[ \limsup_{n \to \infty} F_n(x_n) \leq F(x). \]
With a small abuse of notation we will write \( \Gamma\text{-}\lim(L^1) \) and \( \Gamma\text{-}\lim(TL^1) \) for \( \Gamma \)-convergence with respect to the \( L^1 \) metric \( d_{L^1} : L^1(X) \times L^1(X) \to [0, \infty) \) and the \( TL^1 \) metric \( d_{TL^1} : TL^1(X) \times TL^1(X) \to [0, \infty) \).

The notion of \( \Gamma \)-convergence has been designed in order for the following convergence of minimisers and minima result to hold (see for example [17][19]).

**Theorem 2.23.** Let \((A, d)\) be a metric space and let \( F_n \xrightarrow[\Gamma]{\text{d}} F \), 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 \( \varepsilon_n \)-minimizers for \( F_n \), that is

\[
F_n(x_n) \leq \max \left\{ \inf_{A} F_n + \frac{1}{\varepsilon_n}, \frac{1}{\varepsilon_n} \right\}.
\]

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

**Remark 2.24.** The condition defining an \( \varepsilon \)-minimizer takes into account the fact that the infimum of the functional can be \(-\infty\). In the case \( \inf_{n \in \mathbb{N}} \inf_{A} F_n > -\infty \), condition (3) reduces, for \( n \) sufficiently large, to \( F_n(x_n) \leq \inf_{A} F_n + \frac{1}{\varepsilon_n} \).

In the context of this paper we apply Theorem 2.23 in order to prove Corollary 1.9. In particular, we show that \( G_n^{(p)} + K_n \) \( \Gamma \)-converges to \( G_{\infty} + K_{\infty} \) and satisfies a compactness property. We note that in general

\[
\Gamma\text{-}\lim_{n \to \infty} (d) (F_n + G_n) \neq \Gamma\text{-}\lim_{n \to \infty} (d) F_n + \Gamma\text{-}\lim_{n \to \infty} (d) G_n.
\]

However, with a suitable strong notion of convergence of \( G_n \to G \) we can infer the additivity of \( \Gamma \)-limits.

**Proposition 2.25.** Let \((A, d)\) be a metric space and let \( F_n \xrightarrow[\Gamma]{\text{d}} F \). Assume \( G_n(u_n) \to G(u) \) and \( G(u) > -\infty \) for any sequence \( u_n \to u \) with \( \sup_{n \in \mathbb{N}} F_n(u_n) < +\infty \) and \( F(u) < +\infty \) then

\[
\Gamma\text{-}\lim_{n \to \infty} (d) (F_n + G_n) = \Gamma\text{-}\lim_{n \to \infty} (d) F_n + \Gamma\text{-}\lim_{n \to \infty} (d) G_n.
\]

The assumption in the above proposition is similar to the notion of continuous convergence, see [19, Definition 4.7 and Proposition 6.20]. In our context continuous convergence is not quite the right concept, indeed the fidelity terms \( K_n \) defined in Definition 1.7 do not continuously converge to \( K_{\infty} \). However we show, in Section 3 that \( K_n \) does satisfy the assumptions in Proposition 2.25.

### 3 Convergence of the Non-Local Continuum Model

We first introduce the intermediary functional \( F^{(p)}_{\varepsilon}(\cdot, A) : L^{1}(X) \to [0, \infty] \) that is a non-local continuum approximation of the discrete functional \( G^{(p)}_{\varepsilon} \).

**Definition 3.1.** Let \( p \geq 1, \varepsilon > 0, s_\varepsilon > 0, \) and let \( A \subset X \) be an open and bounded set. Define the functional \( F^{(p)}_{\varepsilon}(\cdot, A) : L^{1}(X) \to [0, \infty] \) by

\[
F^{(p)}_{\varepsilon}(u, A) = \frac{s_\varepsilon}{\varepsilon} \int_{A \times A} \eta_\varepsilon(x - z) |u(x) - u(z)|^p \rho(x) \rho(z) \, dx \, dz + \frac{1}{\varepsilon} \int_{A} V(u(x)) \rho(x) \, dx.
\]

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

This section is devoted to proving the following result.
Theorem 3.2. Let \( p \geq 1 \), and assume \( s_\varepsilon \to 1 \) as \( \varepsilon \to 0 \). Under conditions (A1), (B1-3) and (C1-3) the following holds:

- (Compactness) Let \( \varepsilon_n \to 0^+ \) and \( u_n \in L^1(X, \mu) \) satisfy \( \sup_{n \in \mathbb{N}} F^{(p)}_{\varepsilon_n}(u_n) < \infty \), then \( u_n \) is relatively compact in \( L^1(X, \mu) \) and each cluster point \( u \) has \( G^{(p)}_\infty(u) < \infty \);

- (\( \Gamma \)-convergence) \( \Gamma \)-\( \lim_{\varepsilon \to 0} (L^1) F^{(p)}_\varepsilon = G^{(p)}_\infty \) and furthermore, if \( u \in L^1(X, \mu) \) is a polyhedral function then for any \( \zeta_\varepsilon \to 0 \) the recovery sequence \( u_\varepsilon \) can be chosen to satisfy the following: each \( u_\varepsilon \) is Lipschitz continuous with \( \text{Lip}(u_\varepsilon) = \frac{1}{\varepsilon} \) and \( u_\varepsilon(x) = u(x) \) for all \( x \) satisfying \( \text{dist}(x, \partial^* \{ u = 1 \}) > \frac{\varepsilon}{\zeta_\varepsilon} \).

See Section 3.1 for compactness and Sections 3.3 and 3.4 for the \( \Gamma \)-convergence.

The result of Theorem 3.2 is a generalization of a result by Alberti and Bellettini (see [1]) that we are going to recall for the reader’s convenience. First, we introduce notation for the special case when \( p = 2 \) and \( \rho \equiv 1 \) on \( X \).

Definition 3.3. Let \( \varepsilon > 0 \), and define the functionals \( E_\varepsilon, E_0 : L^1(X) \to [0, \infty] \) by

\[
E_\varepsilon(u) := \frac{1}{\varepsilon} \int_X \int_X \eta_\varepsilon(x, z) |u(x) - u(z)|^2 \, dx \, dz + \frac{1}{\varepsilon} \int_X V(u(x)) \, dx
\]

\[
E_0(u) := \begin{cases} \int_{\partial^* \{ u = 1 \}} \hat{\sigma}(\nu_u(x)) \, d\mathcal{H}^{d-1}(x) & \text{if } u \in BV(X, \{ \pm 1 \}), \\ +\infty & \text{else}, \end{cases}
\]

respectively, where

\[
\hat{\sigma}(\nu) := \min \left\{ E(f; \nu) \mid f : \mathbb{R} \to \mathbb{R}, \lim_{t \to \infty} f(t) = 1, \lim_{t \to -\infty} f(t) = -1 \right\}
\]

\[
E(f; \nu) := \int_{-\infty}^{\infty} \int_{\mathbb{R}^d} \eta(h) |f(t + h_\nu) - f(t)|^2 \, dh \, dt + \int_{-\infty}^{\infty} V(f(t)) \, dt
\]

and \( h_\nu := h \cdot \nu \).

Remark 3.4. We make the following observations.

1. The fact that there exists a minimizer of \( \hat{\sigma}(\nu) \) follows from [2] Theorem 2.4.

2. The minimum in \( \hat{\sigma}(\nu) \) can be taken over non-decreasing functions with \( tf(t) \geq 0 \) for all \( t \in \mathbb{R} \).

3. If \( \eta \) is isotropic, i.e., \( \eta(h) = \eta(|h|) \), then \( E(f; \nu) \) is independent of \( \nu \) and therefore \( \sigma(\nu) \) is a constant and the functional \( E_0 \) is a multiple of the perimeter \( \mathcal{H}^{d-1}(\partial^* \{ u = 1 \}) \).

The following theorem summerizes two results, namely [1] Theorem 1.4 and [2] Theorem 3.3.

Theorem 3.5. Assume that \( X \subset \mathbb{R}^d \) is open, \( V \) satisfies conditions (B1-3), and \( \eta \) satisfies conditions (C1-3). Then the following holds:

- (Compactness) Any sequence \( \varepsilon_n \to 0^+ \) and \( \{ u_n \} \subset L^1(X) \) with \( \sup_{n \in \mathbb{N}} E_{\varepsilon_n}(u_n) < \infty \) is relatively compact in \( L^1(X) \), furthermore any cluster point \( u \) satisfies \( E_0(u) < \infty \);
\(\text{(\(\Gamma\)-convergence) } \lim_{\varepsilon \to 0}(L^1)\mathcal{E}_\varepsilon = \mathcal{E}_0.\)

In fact the above theorem is true if (C1) is relaxed to \(\eta \geq 0\) (i.e. positivity and continuity at the origin are not needed). The proof of Theorem [3.2] is based on the proof of [1, Theorem 1.4], where we have to deal with the fact that, in our case, we are considering a generic exponent \(p \geq 1\), and that we have a density \(\rho\). The compactness proof is in Section 3.1 and the \(\Gamma\)-convergence in Sections 3.3 and 3.4.

3.1 Compactness

The aim of this section is to show that any sequence \(\{u_n\}_{n=1}^{\infty} \subset L^1(X, \mu)\) with \(\sup_{n\in\mathbb{N}} \mathcal{F}^{(p)}_{\varepsilon_n}(u_n) < \infty\) is relatively compact in \(L^1(X, \mu)\) and that \(\mathcal{G}^{(p)}_{\infty}(u) < \infty\) for any cluster point \(u \in L^1(X, \mu)\). This will prove the first part of Theorem [3.2].

The strategy of the proof is to apply the Alberti and Bellettini compactness result in Theorem 3.5. When \(p = 2\) this follows from the upper and lower bounds on \(\rho\) that imply an ‘equivalence’ between \(\mathcal{E}_{\varepsilon_n}\) and \(\mathcal{F}^{(2)}_{\varepsilon_n}\). When \(p \neq 2\) we approximate \(v_n\) with a sequence \(v_n\) satisfying \(v_n(x) \in \{\pm 1\}\) then since \(|v_n(x) - v_n(z)|^p = 2^{2-p}|v_n(x) - v_n(z)|^p\) we can easily find an equivalence between \(\mathcal{E}_{\varepsilon_n}\) and \(\mathcal{F}^{(p)}_{\varepsilon_n}\). We start with the preliminary result that shows \(\mathcal{E}_0(u) < \infty \iff \mathcal{G}^{(p)}_{\infty}(u) < \infty\).

**Proposition 3.6.** Let \(X \subset \mathbb{R}^d\) be open and bounded, and let \(u \in L^1(X; \{\pm 1\})\). Under Assumptions (A1), (B1-2), (C1,3) we have

\[
\int_{\partial^*\{u=1\}} \hat{\sigma} (\nu_u(x)) \, d\mathcal{H}^{d-1}(x) < +\infty \iff \mathcal{H}^{d-1}(\partial^*\{u = 1\}) < +\infty
\]

\[
\implies \int_{\partial^*\{u=1\}} \sigma^{(p)}(x, \nu_u(x)) \rho(x) \, d\mathcal{H}^{d-1}(x) < +\infty.
\]

**Remark 3.7.** The above proposition implies that

\(\mathcal{E}_0(u) < \infty \iff u \in BV(X; \{\pm 1\}) \iff \mathcal{G}^{(p)}_{\infty}(u) < \infty\)

since by definition of \(\mathcal{G}^{(p)}_{\infty}\) if \(\mathcal{G}^{(p)}_{\infty}(u) < \infty\) then \(u \in BV(X; \{\pm 1\})\).

**Remark 3.8.** The missing implication,

\[
\int_{\partial^*\{u=1\}} \sigma^{(p)}(x, \nu_u(x)) \rho(x) \, d\mathcal{H}^{d-1}(x) < +\infty \implies \mathcal{H}^{d-1}(\partial^*\{u = 1\}) < +\infty,
\]

is also true. One can show that the minimisation problem in \(\sigma^{(p)}(x, \nu)\) (for any \(x \in X\) and \(\nu \in \mathbb{S}^{d-1}\)) can be reduced to a minimisation problem over functions \(f : \mathbb{R} \to \mathbb{R}\) similar to \(\hat{\sigma}(\nu)\) (but with an additional \(x\) dependence); see Lemma [3.9]. Since the missing implication is not needed then we do not include it.

**Proof of Proposition 3.6.** Step 1. We show

\[
\mathcal{H}^{d-1}(\partial^*\{u = 1\}) < +\infty \implies \int_{\partial^*\{u=1\}} \hat{\sigma} (\nu_u(x)) \, d\mathcal{H}^{d-1}(x) < +\infty.
\]

Choose \(\hat{f}(t) = +1\) if \(t \geq 0\) and \(\hat{f}(t) = -1\) if \(t < 0\). Then,

\[
E(\hat{f}; \nu) = \int_{-\infty}^{\infty} \int_{\mathbb{R}^d} \eta(h) |\hat{f}(t + h \nu) - \hat{f}(t)|^2 \, dh \, dt = \int_{\mathbb{R}^d} \eta(h) |h \nu| \, dh \leq R_\eta \int_{\mathbb{R}^d} \eta(h) =: C
\]

19
where \( C \in (0, \infty), h_\nu = h \cdot \nu \) and \( R_\eta \) is given by assumption Assumption (C3). Note that \( C \) is independent of \( \nu \). So,

\[
\int_{\partial^* \{u=1\}} \hat{\sigma}(\nu_u(x)) \, d\mathcal{H}^{d-1}(x) \leq \int_{\partial^* \{u=1\}} E(\hat{f}; \nu_u(x)) \, d\mathcal{H}^{d-1}(x) \leq C \mathcal{H}^{d-1}(\partial^* \{u=1\}).
\]

**Step 2.** We show

\[
\int_{\partial^* \{u=1\}} \hat{\sigma}(\nu_u(x)) \, d\mathcal{H}^{d-1}(x) < +\infty \iff \mathcal{H}^{d-1}(\partial^* \{u=1\}) < +\infty.
\]

Using Assumption (C1) it is possible to find \( a > 0 \) and \( c > 0 \) satisfying \( \eta(h) \geq c \) for all \( |h| \leq a \). Let \( \hat{f} : \mathbb{R} \to \mathbb{R} \) such that

\[
\hat{f} \text{ is non-decreasing, } \lim_{t \to \infty} \hat{f}(t) = -\lim_{t \to -\infty} \hat{f}(t) = 1, \quad \hat{f}(t) \geq 0 \text{ for } t \in \mathbb{R}.
\]

If \( \hat{f}(\frac{a}{2}) \leq \frac{1}{2} \) then \( \hat{f}(t) \in [0, \frac{1}{2}] \) for \( t \in [0, \frac{a}{2}] \) and

\[
E(\hat{f}; \nu) \geq \frac{a}{2} \inf_{t \in [0, \frac{1}{2}]} V(t) =: \tilde{a}_1 > 0.
\]

Otherwise, if \( \hat{f}(\frac{a}{2}) > \frac{1}{2} \) then for \( h \geq \frac{a}{2} \)

\[
\int_{\mathbb{R}} |\hat{f}(t + h) - \hat{f}(t)|^2 \, dt \geq \int_{\frac{a}{2} - h}^{0} |\hat{f}(t + h) - \hat{f}(t)|^2 \, dt \geq \frac{h - \frac{a}{2}}{4}.
\]

Similarly, if \( h \leq -\frac{a}{2} \) we have

\[
\int_{\mathbb{R}} |\hat{f}(t + h) - \hat{f}(t)|^2 \, dt \geq \frac{\frac{a}{2} - h}{4}.
\]

So, for all \( |h| \geq \frac{a}{2} \) it holds

\[
\int_{\mathbb{R}} |\hat{f}(t + h) - \hat{f}(t)|^2 \, dt \geq \frac{|h - \frac{a}{2}|}{4}.
\]

Therefore

\[
E(\hat{f}; \nu) \geq \int_{\mathbb{R}^d} \eta(h) \int_{\mathbb{R}} |\hat{f}(t + h\nu) - \hat{f}(t)|^2 \, dt \, dh
\]

\[
\geq \frac{1}{4} \int_{\{h \in \mathbb{R}^d : |h\nu| \geq \frac{a}{2}\}} \eta(h) \, |h\nu - \frac{a}{2}| \, dh
\]

\[
\geq \frac{ac}{16} \int_{\{h \in \mathbb{R}^d : |h\nu| \geq \frac{a}{2}\}} \chi_{B(0,a)} \, dh
\]

\[
=: \tilde{a}_2 > 0.
\]

Set \( \tilde{a} := \min\{\tilde{a}_1, \tilde{a}_2\} > 0 \). Using (6) and (7) we have, for any \( f : \mathbb{R} \to \mathbb{R} \) satisfying (5), that \( E(\hat{f}; \nu) \geq \tilde{a} \).

**Step 3.** We show

\[
\mathcal{H}^{d-1}(\partial^* \{u=1\}) < +\infty \iff \int_{\partial^* \{u=1\}} \sigma^{(p)}(x, \nu_u(x)) \rho(x) \, d\mathcal{H}^{d-1}(x) < +\infty.
\]

20
Following the same argument as in the first part of the proof we let \( \hat{f}(t) = 1 \) for \( t \geq 0 \) and \( \hat{f}(t) = -1 \) for \( t < 0 \). Fix \( \nu \in \mathbb{S}^{d-1} \) and let \( f_\nu(x) = \hat{f}(x \cdot \nu) \). For any \( \bar{x} \in X \) and \( C \in \mathcal{C}(\bar{x}, \nu) \) we clearly have \( f_\nu \in \mathcal{U}(C, \nu) \). So,

\[
\frac{1}{\mathcal{H}^{d-1}(C)} G^{(p)}(f_\nu, \rho(\bar{x}), T_C) = \rho(\bar{x}) \int_{-\infty}^{\infty} \int_{\mathbb{R}^d} \eta(h) |\hat{f}(t + h \nu) - \hat{f}(t)|^p \, dh \, dt \\
\leq 2^p c_2 \int_{\mathbb{R}^d} \eta(h) |h \nu| \, dh \\
\leq \tilde{c}
\]

where \( \tilde{c} \in (0, \infty) \) can be chosen to be independent of \( \nu \) and in the last step we used Assumption (C3). Then,

\[
\int_{\partial^* \{ u = 1 \}} \sigma^{(p)}(x, \nu_\nu(x)) \rho(x) \, d\mathcal{H}^{d-1}(x) \leq \int_{\partial^* \{ u = 1 \}} \frac{1}{\mathcal{H}^{d-1}(C)} G^{(p)}(f_\nu, \rho(\bar{x}), T_C) \rho(\bar{x}) \, d\mathcal{H}^{d-1}(\bar{x}) \\
\leq c_2 \tilde{c} \mathcal{H}^{d-1}(\partial^* \{ u = 1 \}).
\]

This concludes the proof. \( \square \)

By the above proposition it is enough to show that any sequence \( \{ u_n \}_{n=1}^{\infty} \subset L^1(X, \mu) \) satisfying \( \sup_{n \in \mathbb{N}} F_{\varepsilon_n}(u_n) < +\infty \) is relatively compact and any cluster point \( u \) satisfies \( E_0(u) < \infty \). We do this by a direct comparison with \( E_{\varepsilon_n} \).

**Proof of Theorem 3.2 (Compactness).** Assume \( \varepsilon_n \to 0^+ \), \( s_n := s_{\varepsilon_n} \to 1 \) and \( \sup_{n \in \mathbb{N}} F_{\varepsilon_n}^{(p)}(u_n) < +\infty \). Let

\[
v_n(x) := \text{sign}(u_n) := \begin{cases} +1 & \text{if } u_n(x) \geq 0 \\ -1 & \text{if } u_n(x) < 0. \end{cases}
\]

We claim that

\[
\| u_n - v_n \|_{L^1} \to 0, \quad (8)
\]

\[
\sup_{n \in \mathbb{N}} F_{\varepsilon_n}^{(p)}(v_n) < +\infty. \quad (9)
\]

**Step 1.** Let us first prove (8). Fix \( \delta > 0 \) and let

\[
K_n^{(\delta)} = \{ x \in X : |u_n(x)| \geq 1 + \delta \} \\
L_n^{(\delta)} = \{ x \in X : |u_n(x)| \leq 1 - \delta \}.
\]

Note that for \( x \in X \setminus (K_n^{(\delta)} \cup L_n^{(\delta)}) \) we have \( |v_n(x) - u_n(x)| \leq \delta \). Now,

\[
\int_X |u_n(x) - v_n(x)| \, dx \leq \delta \text{Vol}(X) + \int_{K_n^{(\delta)}} |u_n(x) - v_n(x)| \, dx + \int_{L_n^{(\delta)}} |u_n(x) - v_n(x)| \, dx \\
\leq \delta \text{Vol}(X) + \int_{K_n^{(\delta)}} |u_n(x)| \, dx + \text{Vol}(K_n^{(\delta)}) + 2\text{Vol}(L_n^{(\delta)}).
\]

Since \( V \) is continuous and zero only at \( \pm 1 \) then there exists \( \gamma_\delta > 0 \) such that \( V(t) \geq \gamma_\delta \) for all \( t \in (-\infty, -1 - \delta) \cup (-1 + \delta, 1 - \delta) \cup (1 + \delta, +\infty) \). Hence \( V(u_n(x)) \geq \gamma_\delta \) for all \( x \in K_n^{(\delta)} \cup L_n^{(\delta)} \). This implies,

\[
\text{Vol}(K_n^{(\delta)}) \leq \frac{1}{\gamma_\delta} \int_{K_n^{(\delta)}} V(u_n(x)) \, dx \leq \frac{\varepsilon_n}{c_1 \gamma_\delta} F_{\varepsilon_n}^{(p)}(u_n).
\]

21
By the same calculation \( \text{Vol}(L_n) \leq \frac{\varepsilon_n}{c_1\gamma_0} F_{\varepsilon_n}^{(p)}(u_n) \). Furthermore,

\[
\int_{K_n^{(p)}} \left| u_n(x) \right| \, dx = \int_{K_n^{(p)} \cap \{|u_n(x)| \leq R_V\}} \left| u_n(x) \right| \, dx + \int_{\{|u_n(x)| > R_V\}} \left| u_n(x) \right| \, dx
\]
\[
\leq R_V \text{Vol}(K_n^{(p)}) + \frac{1}{\tau} \int_{\{|u_n(x)| > R_V\}} V(u_n(x)) \, dx
\]
\[
\leq \left( \frac{R_V}{\gamma_0} + \frac{1}{\tau} \right) \frac{\varepsilon_n F_{\varepsilon_n}^{(p)}(u_n)}{c_1}.
\]

So, \( \lim_{n \to \infty} \|u_n - v_n\|_{L^1} \leq \delta \text{Vol}(X) \). Since this is true for all \( \delta > 0 \) then we have \( \lim_{n \to \infty} \|u_n - v_n\|_{L^1} = 0 \) which proves claim (8).

**Step 2.** In order to prove (9) we reason as follows. If \( |u_n(x)| \geq \frac{1}{2} \) then \( \text{sign}(u_n(x)) \neq \text{sign}(u_n(y)) \) implies \( |u_n(x) - u_n(y)| \geq \frac{1}{2} \). Now since,

\[
|v_n(x) - v_n(y)| = \begin{cases} 0 & \text{if sign}(u_n(x)) = \text{sign}(u_n(y)) \\ 2 & \text{if sign}(u_n(x)) \neq \text{sign}(u_n(y)) \end{cases}
\]

then \( |v_n(x) - v_n(y)| \leq 4|u_n(x) - u_n(y)| \) when \( |u_n(x)| \geq \frac{1}{2} \).

Let \( M_n = \{x \in X : |u_n(x)| \leq \frac{1}{2}\} \). We have,

\[
F_{\varepsilon_n}^{(p)}(v_n) = \frac{s_n}{\varepsilon_n} \int_{M_n \times X} \eta_{\varepsilon_n}(x - z) |v_n(x) - v_n(z)|^p \rho(x) \rho(z) \, dx \, dz + \frac{s_n}{\varepsilon_n} \int_{M_n \times X} \eta_{\varepsilon_n}(x - z) |v_n(x) - v_n(y)|^p \rho(x) \rho(y) \, dx \, dz
\]
\[
\leq \frac{2p c_2 s_n}{\varepsilon_n} \int_{M_n \times X} \eta_{\varepsilon_n}(x - z) \, dx \, dz + \frac{4p s_n}{\varepsilon_n} \int_{M_n \times X} \eta_{\varepsilon_n}(x - z) |u_n(x) - u_n(z)|^p \rho(x) \rho(z) \, dx \, dz
\]
\[
\leq \frac{2p c_2 s_n}{\varepsilon_n} \int_{\mathbb{R}^d} \eta(w) \, dw \text{Vol}(M_n) + 4p F_{\varepsilon_n}^{(p)}(u_n).
\]

Now,

\[
\frac{1}{\varepsilon_n} \text{Vol}(M_n) \leq \frac{1}{\gamma c_1} \int_{\{|u_n(x)| \leq \frac{1}{2}\}} V(u_n(x)) \, dx \leq \frac{1}{\gamma c_1} F_{\varepsilon_n}^{(p)}(u_n)
\]

where \( V(t) \geq \gamma > 0 \) for all \( t \in [-\frac{1}{2}, \frac{1}{2}] \). Hence \( \sup_{n \in \mathbb{N}} F_{\varepsilon_n}^{(p)}(v_n) < +\infty \).

**Step 3.** We conclude the proof by noticing that, since \( v_n \in L^1(X; \{\pm 1\}) \) we have \( F_{\varepsilon_n}^{(p)}(v_n) = 2^{p-2} s_n \varepsilon_{\varepsilon_n}(v_n) \). So \( v_n \) is relatively compact in \( L^1 \) by Theorem 3.5 and by Proposition 3.6 \( G_{\infty}^{(p)}(u) < +\infty \) for any cluster point \( u \) of \( \{v_n\}_{n=1}^{\infty} \). Using (8) the same holds true for the sequence \( \{u_n\}_{n \in \mathbb{N}} \).

### 3.2 Preliminary results

Here we prove some technical results needed in the proof of the \( \Gamma \)-convergence result stated in Theorem 3.2. We start by stating the result that the minimisation in the cell problem \( \sigma^{(p)}(x, \nu) \) can be restricted to functions \( u \) that only vary in the direction \( \nu \).
Lemma 3.9. Fix $\nu \in \mathbb{S}^{d-1}$ and let $C \subset \nu^\perp$. Let $p \geq 1$ and assume Assumptions (B1-3) and (C1-3) are in force. Then, for any $\lambda > 0$, the minimisation problem

$$\min \left\{ \frac{1}{\mathcal{H}^{d-1}(C)} G^{(p)}(u, \lambda, T_C) : u \in \mathcal{U}(C, \nu) \right\}$$

has a solution $u^\dagger$ that can be written $u^\dagger(x) = \bar{u}(x \cdot \nu)$ where $\bar{u} : \mathbb{R} \to [-1, 1]$ is increasing.

The proof of the lemma is a simple adaptation of [2] Theorem 3.3. Indeed, one can absorb $\lambda$ into the mollifier $\eta$ and then the only difference between Lemma 3.9 and [2] Theorem 3.3 is the exponent $p$. By modifying [2] Proposition 3.7 to treat this more general case we can conclude Lemma 3.9 holds. We omit the proof of Lemma 3.9.

Remark 3.10. By Lemma 3.9 we can write

$$\sigma^{(p)}(x, \nu) = \inf \left\{ \rho(x) \int_{\mathbb{R}} \int_{\mathbb{R}^d} \eta(h)|u(z + h \cdot \nu) - u(z)|^p dh dz + \int \mathcal{V}(u(z)) dz \right\}$$

where the infimum is taken over functions $u : \mathbb{R} \to [-1, 1]$ satisfying $\lim_{x \to \infty} u(x) = 1$, $\lim_{x \to -\infty} u(x) = -1$.

We continue by proving some continuity properties of the function $\sigma^{(p)}$.

Lemma 3.11. Under Assumptions (A1) and (C3) the followings hold:

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

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

Proof. Proof of (i). Fix $\bar{x} \in X$ and $\nu \in \mathbb{S}^{d-1}$ and let $\{x_n\}_{n=1}^{\infty} \subset X$ and $\{\nu_n\}_{n=1}^{\infty} \subset \mathbb{S}^{d-1}$ be such that $x_n \to \bar{x}$ and $\nu_n \to \nu$ as $n \to \infty$. Let $R_n$ be a rotation such that $R_n\nu = \nu_n$. Fix $t > 0$ and let $D \in \mathcal{C}(\bar{x}, \nu)$ and $w \in \mathcal{U}(D, \nu)$ be such that

$$\frac{1}{\mathcal{H}^{d-1}(D)} G^{(p)}(w, \rho(\bar{x}), T_D) = \sigma^{(p)}(\bar{x}, \nu) + t.$$

By Remark 3.10 we can assume that $w(x) = \bar{w}(x \cdot \nu)$ for some increasing function $\bar{w} : \mathbb{R} \to [-1, 1]$. Then,

$$\int_{T_D} |w(z + h) - w(z)|^p \, dz \leq 2^{p-1} \mathcal{H}^{d-1}(D) \int_{\mathbb{R}} \int_{\mathbb{R}^d} |\bar{w}(z_{\nu} + h_{\nu}) - \bar{w}(z_{\nu})| \, dz_{\nu} \, dh_{\nu}$$

where $h_{\nu} = h \cdot \nu$

$$= 2^{p-1} \mathcal{H}^{d-1}(D) \lim_{M \to \infty} \sum_{i=-M}^{M} \int_{0}^{h_{\nu}} |\bar{w}(y + (i+1)h_{\nu}) - \bar{w}(y + ih_{\nu})| \, dy$$

$$= 2^{p-1} \mathcal{H}^{d-1}(D) \lim_{M \to \infty} \int_{0}^{h_{\nu}} \bar{w}(y + (M+1)h_{\nu}) - \bar{w}(y - Mh_{\nu}) \, dy$$

$$= 2^{p} \mathcal{H}^{d-1}(D) h_{\nu}$$

with the last line following from the monotone convergence theorem. Hence,

$$\int_{T_D} |w(z + h) - w(z)|^p \, dz \leq 2^{p} \mathcal{H}^{d-1}(D) h \cdot \nu . \quad (10)$$

For $n \in \mathbb{N}$ define $C_n \in \mathcal{C}(x_n, \nu_n)$ and $u_n \in \mathcal{U}(C_n, \nu_n)$ by

$$C_n := R_n(D - \bar{x}) + x_n ,$$
We claim that 
\[ \delta \]
where in the last step we used (10). In order to show that the second term in the parenthesis vanishes, we use an argument similar to the one for proving that translations are continuous in 
\[ \mathbb{R} \].

Since \( r > L \), we get
\[ \tilde{n} \rightarrow \sigma \] for all \( n \) sufficiently large. Then
\[ \rho \] is a continuous function with support in \( B(0, 2\tilde{R}_{\eta}) \) such that \( \| \eta - \tilde{\eta}_{s} \|_{L_{1}(\mathbb{R}^{d})} < \varepsilon \). Then, for every \( r > 0 \) there exists \( \tilde{n} \in \mathbb{N} \) such that \( |\eta_{s}(R_{n}h) - \tilde{\eta}_{s}(h)| < r \) for all \( h \in \mathbb{R}^{d} \) and all \( n \geq \tilde{n} \). So that, for \( n \geq \tilde{n} \)

\[ \int_{\mathbb{R}^{d}} |\eta_{s}(R_{n}h) - \eta(h)| \; dh \leq \| \eta - \tilde{\eta}_{s} \|_{L_{1}(\mathbb{R}^{d})} + \int_{\mathbb{R}^{d}} |\tilde{\eta}_{s}(R_{n}h) - \tilde{\eta}_{s}(h)| \; dh \leq s + r \text{Vol}(B(0, 2\tilde{R}_{\eta})) . \]

Since \( r \) and \( s \) are arbitrary, we conclude that
\[ \int_{\mathbb{R}^{d}} |\eta_{s}(R_{n}h) - \eta(h)| \; dh \rightarrow 0 \] as \( n \rightarrow \infty \) and, in turn, that \( \delta_{n} \rightarrow 0 \) as \( n \rightarrow \infty \).

**Proof of (ii).** Fix \( \nu \in \mathbb{S}^{d-1} \), \( \bar{x} \in \mathbb{X} \) and let \( x_{n} \rightarrow \bar{x} \). By part (i) we have that \( \sigma^{(p)}(\bar{x}, \nu) \geq \limsup_{n \rightarrow \infty} \sigma^{(p)}(x_{n}, \nu) \).

We claim that \( \sigma^{(p)}(\bar{x}, \nu) \leq \liminf_{n \rightarrow \infty} \sigma^{(p)}(x_{n}, \nu) \). Without loss of generality, let us assume that

\[ \liminf_{n \rightarrow \infty} \sigma^{(p)}(x_{n}, \nu) = \lim_{n \rightarrow \infty} \sigma^{(p)}(x_{n}, \nu) < \infty . \]

For every \( n \in \mathbb{N} \) let \( C_{n} \in \mathcal{C}(x_{n}, \nu) \) and \( u_{n} \in \mathcal{U}(C_{n}, \nu) \) be such that

\[ \frac{1}{\mathcal{H}_{d-1}(C_{n})} \mathcal{G}^{(p)}(u_{n}, \rho(x_{n}), T_{C_{n}}) \leq \sigma^{(p)}(x_{n}, \nu) + \frac{1}{n} . \]
Set \( \lambda_n := \bar{x} - x_n \) and define
\[
\tilde{C}_n := C_n + \lambda_n, \quad \tilde{u}_n(x) := u_n(x - \lambda_n).
\]

So, \( \tilde{C}_n \in C(\bar{x}, \nu) \) and \( \tilde{u}_n \in U(\tilde{C}_n, \nu) \). Using (11), we get
\[
\sigma^{(p)}(\bar{x}, \nu) \leq \frac{1}{\mathcal{H}^{d-1}(\tilde{C}_n)} G^{(p)}(\tilde{u}_n, \rho(\bar{x}), T_{\tilde{C}_n})
\leq \frac{G^{(p)}(u_n, \rho(x_n), T_{C_n})}{\mathcal{H}^{d-1}(C_n)} + \frac{1}{\mathcal{H}^{d-1}(C_n)} \left| G^{(p)}(\tilde{u}_n, \rho(\bar{x}), T_{\tilde{C}_n}) - G^{(p)}(u_n, \rho(x_n), T_{C_n}) \right|
\leq \sigma^{(p)}(x_n, \nu) + \frac{1}{n} \left[ G^{(p)}(\tilde{u}_n, \rho(\bar{x}), T_{\tilde{C}_n}) - G^{(p)}(u_n, \rho(x_n), T_{C_n}) \right]
\leq \left[ \sigma^{(p)}(x_n, \nu) + \frac{1}{n} \right].
\]

To estimate the last term, we reason as follows. First of all, we notice that
\[
\int_{T_{\tilde{C}_n}} V(u_n(z)) \, dz = \int_{T_{\tilde{C}_n}} V(\tilde{u}_n(z)) \, dz .
\]

Fix \( \varepsilon > 0 \). Using the continuity of \( \rho \), there exists \( \bar{n} \in \mathbb{N} \), such that \( |\rho(x_n) - \rho(\bar{x})| < \varepsilon \) for all \( n \geq \bar{n} \). From (13) we get that
\[
\frac{1}{\mathcal{H}^{d-1}(C_n)} \left| G^{(p)}(\tilde{u}_n, \rho(\bar{x}), T_{\tilde{C}_n}) - G^{(p)}(u_n, \rho(x_n), T_{C_n}) \right|
\leq \frac{1}{\mathcal{H}^{d-1}(C_n)} |\rho(x_n) - \rho(\bar{x})| \int_{T_{\tilde{C}_n}} \int_{\mathbb{R}^d} \eta(h) |u_n(z + h) - u_n(z)|^p \, dh \, dz
\leq \frac{\varepsilon}{c_1} \left[ \sigma^{(p)}(x_n, \nu) + \frac{1}{n} \right].
\]

By the above and (12) we have
\[
\left( 1 - \frac{\varepsilon}{c_1} \right) \sigma^{(p)}(\bar{x}, \nu) \leq \left( 1 + \frac{\varepsilon}{c_1} \right) \sigma^{(p)}(x_n, \nu) + \frac{1}{n}
\]

Using the arbitrariness of \( \varepsilon \) we conclude that
\[
\sigma^{(p)}(\bar{x}, \nu) \leq \liminf_{n \to \infty} \sigma^{(p)}(x_n, \nu)
\]
as required. \( \Box \)

Remark 3.12. Notice that the above result did not require the existence of a solution for the infimum problem defining \( \sigma^{(p)} \).

We notice that the main feature of the \( \Gamma \)-convergence of \( \mathcal{F}^{(p)}_{\varepsilon_n} \) to \( \mathcal{G}^{(p)}_{\infty} \) is that we recover, in the limit, a local functional starting from a nonlocal functional. To be more precise, let \( A, B \subset X \) be disjoint sets. Then, it holds that
\[
\mathcal{F}^{(p)}_{\varepsilon_n}(u, A \cup B) = \mathcal{F}^{(p)}_{\varepsilon_n}(u, A) + \mathcal{F}^{(p)}_{\varepsilon_n}(u, B) + 2\tilde{\Lambda}_{\varepsilon_n}(u, A, B)
\]
where we define the nonlocal deficit
\[
\tilde{\Lambda}_{\varepsilon}(u, A, B) := \frac{s_{\varepsilon}}{\varepsilon} \int_A \int_B \eta_{\varepsilon}(x - z) |u(x) - u(z)|^p \rho(x) \rho(z) \, dx \, dz.
\]
On the other hand, for the limiting functional we have

\[ G^{(p)}_{\infty}(u, A \cup B) = G^{(p)}_{\infty}(u, A) + G^{(p)}_{\infty}(u, B), \]

where, for \( u \in BV(X; \{\pm 1\}) \), we set

\[ G^{(p)}_{\infty}(u, A) := \int_{\partial^* \{u = 1\} \cap A} \sigma^{(p)}(x, \nu_u(x)) \rho(x) \, d\mathcal{H}^{d-1}(x). \]

Identity (14) states that the functionals \( F^{(p)}_{\varepsilon_n} \) are nonlocal, while (16) is the locality property of the limiting functional \( G^{(p)}_{\infty} \). Thus, we expect the nonlocal deficit to disappear in the limit, i.e., if \( u_{\varepsilon_n} \to u \) in \( L^1(X) \), then

\[ \Lambda_{\varepsilon_n}(u_{\varepsilon_n}, A, B) \to 0, \quad \text{as } n \to \infty. \]

For technical reasons we also consider the nonlocal deficits without weighting by \( \rho \) or \( s_{\varepsilon} \):

\[ \Lambda_\varepsilon(u, A, B) := \frac{1}{\varepsilon} \int_A \int_B \eta_\varepsilon(x-z) |u(x) - u(z)|^p \, dx \, dz. \]

By continuity of \( \rho \) if \( A \) and \( B \) are sets in \( X \) that are close to \( \bar{x} \) then \( \Lambda_\varepsilon(u, A, B) \approx s_{\varepsilon} \rho^2(\bar{x}) \Lambda_\varepsilon(u, A, B) \).

In [1] the authors prove that the limit of the nonlocal deficit is determined by the behavior of \( u_{\varepsilon_n} \) close to the boundaries of \( A \) and \( B \), and, in turn, that (17) holds in certain cases of interest. Here we only state the main technical result of [1] in a version we need in the paper, addressing the interested reader to the paper by Alberti and Bellettini for the details.

**Proposition 3.13.** Assume Assumptions (A1) and (C1-3) hold. 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 \( \nu \in S^{d-1} \) the following holds: given \( C \in C(\bar{x}, \nu) \) consider the strip \( T_C \) and any cube \( Q \subset \mathbb{R}^d \) whose intersection with \( \nu^\perp \) is \( C \), 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 \( n \to \infty \).

**Remark 3.14.** The boundedness assumption on the sequence \( \{v_n\}_{n=1}^\infty \) allows one to obtain the proof of the above result directly from [1] Proposition 2.5 and Theorem 2.8. In particular, in [1] the authors prove the result for \( p = 1 \) and \( \rho \equiv 1 \), using the \( L^\infty \) bound on \( v_n \) one can easily bound the more general case considered here by the \( L^1 \) case. With similar computations it is also possible to obtain the same result without the \( L^\infty \) bound.

Finally, notice that when \( A, B \in \mathbb{R}^d \) are disjoint sets with \( d(A, B) > 0 \), using the fact that the function \( \eta \) has support in the ball \( B(0, R) \) (see Assumption (C3)), it is easy to prove that there exists \( \bar{n} \in \mathbb{N} \) such that for all \( n \geq \bar{n} \) it holds

\[ \Lambda_{\varepsilon_n}(v_n, A, B) = 0. \]

For technical reasons we need to introduce a scaled version of the functional \( G^{(p)} \).

**Definition 3.15.** For \( \varepsilon > 0, p \geq 1, u : \mathbb{R}^d \to \mathbb{R}, \lambda \in \mathbb{R}, \) and \( A \subset \mathbb{R}^d \), we define

\[ G^{(p)}_\varepsilon(u, \lambda, A) := \frac{\lambda}{\varepsilon} \int_A \int_{\mathbb{R}^d} \eta_\varepsilon(h) |u(z + h) - u(z)|^p \, dh \, dz + \frac{1}{\varepsilon} \int_A V(u(z)) \, dz. \]
Let \( r > 0 \) and \( x \in X \). For a set \( A \subset \mathbb{R}^d \), we define \( x + rA := \{ x + ry : y \in A \} \). Moreover, for a function \( u : \mathbb{R}^d \to \mathbb{R} \), we set
\[
R_{x,r} u(y) := u(x + ry).
\]
Using a change of variables, it is easy to see that the following scaling property holds true:
\[
G_{\varepsilon}^{(p)}(u, \lambda, x + rA) = r^{d-1}G_{\varepsilon/r}^{(p)}(R_{x,r} u, \lambda, A).
\]

3.3 The Liminf Inequality

This section is devoted to proving the following: let \( u_{\varepsilon_n} \to u \) in \( L^1 \), then
\[
G_{\varepsilon}^{(p)}(u) \leq \lim \inf_{n \to \infty} F_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}).
\]
We will follow the proof of [1, Theorem 1.4], with some modifications due to the presence of the density \( \rho \).

**Proof of Theorem 3.2 (Liminf).** Let \( \varepsilon_n \to 0^+ \) and \( u_{\varepsilon_n} \to u \) in \( L^1(X, \mu) \). Assume without loss of generality that
\[
\lim \inf_{n \to \infty} F_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}) = \lim_{n \to \infty} F_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}) < \infty.
\]

**Step 1.** By compactness (see Section 3.1) it holds
\[
u = \chi_A \text{ for some set } A \subset X \text{ of finite perimeter in } X.
\]
In order to prove (19) we use the strategy introduced by Fonseca and Müller in [26]. Write
\[
F_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}) = \int_X g_{\varepsilon_n}(x)dx
\]
and set \( d\lambda_{\varepsilon_n} := g_{\varepsilon_n} d\mathcal{L}^d \mathbf{1} X \), so that
\[
|\lambda_{\varepsilon_n}|(X) = F_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}).
\]
Using (20), (21), (22), up to a subsequence (not relabeled) it holds
\[
\lambda_{\varepsilon_n} \xrightarrow{\ast} \lambda \text{ for some finite Radon measure } \lambda \text{ on } X.
\]
Then
\[
|\lambda|(X) \leq \lim \inf_{n \to \infty} |\lambda_{\varepsilon_n}|(X).
\]
In view of (22) and (23), the liminf inequality (19) is implied by the following claim: for \( \mathcal{H}^{d-1}\)-a.e. \( \bar{x} \in \partial^* \{ u = 1 \} \) it holds
\[
\sigma^{(p)}(\bar{x}, \nu(\bar{x})) \rho(\bar{x}) \leq \frac{d\lambda}{d\theta}(\bar{x}),
\]
where \( \theta := \mathcal{H}^{d-1} \mathbf{1} \partial^* \{ u = 1 \} \). In order to prove the claim we reason as follows. For \( \mathcal{H}^{d-1}\)-a.e. \( \bar{x} \in \partial^* \{ u = 1 \} \) it is possible to find the density of \( \lambda \) with respect to \( \theta \) via (recall Remark 2.17)
\[
\frac{d\lambda}{d\theta}(\bar{x}) = \lim_{r \to 0} \frac{\lambda(\bar{x} + rQ)}{r^{d-1}}.
\]
where \( Q \) is a unit cube centered at the origin and having \( \nu(\bar{x}) \), the measure theoretic exterior normal to \( A \) at \( \bar{x} \), as one of its axes. Let \( \bar{x} \in \partial^* \{ u = 1 \} \). Theorem 2.16 implies that
\[
R_{\bar{x},r} u \to v_{\bar{x}}
\]
in \( L^1_{loc}(\mathbb{R}^N) \) as \( r \to 0 \), where
\[
v_{\bar{x}}(x) := \begin{cases} -1 & x \cdot \nu(\bar{x}) \geq 0, \\ 1 & x \cdot \nu(\bar{x}) < 0. \end{cases}
\]
Let \( \bar{x} \in \partial^* \{ u = 1 \} \) be a point for which \((24)\) and \((25)\) hold. Without loss of generality, we can assume that

\[
\frac{d\lambda}{d\theta}(\bar{x}) < \infty.
\]  

(26)

Since \( u_{\varepsilon_n} \to u \) in \( L^1 \) and \( \lambda_{\varepsilon_n} \rightharpoonup \lambda \), it is possible to find a (not relabeled) subsequence \( \{ \varepsilon_n \}_{n=1}^\infty \) and a sequence \( \{ r_n \}_{n=1}^\infty \) with \( r_n \to 0^+ \) and \( \varepsilon_n \to 0^+ \), such that

\[
\frac{d\lambda}{d\theta}(\bar{x}) = \lim_{n \to \infty} \frac{\lambda_{\varepsilon_n}(\bar{x} + r_n Q)}{r_n^{d-1}}
\]  

(27)

and

\[
R_{\bar{x},r_n} u_{\varepsilon_n} \to u_\bar{x}.
\]

Using the fact that \( X \) is open, we can assume that \( \bar{x} + r_n Q \subset X \) for all \( n \in \mathbb{N} \). Thus

\[
\frac{\lambda_{\varepsilon_n}(\bar{x} + r_n Q)}{r_n^{d-1}} \geq \frac{\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \bar{x} + r_n Q)}{r_n^{d-1}}.
\]  

(28)

**Step 2.** We claim that

\[
\delta_n := \frac{\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \bar{x} + r_n Q) - \rho(\bar{x}) \mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \rho(\bar{x}), \bar{x} + r_n Q)}{r_n^{d-1}} \to 0,
\]  

(29)

as \( n \to \infty \), where

\[
\mathcal{F}_{\varepsilon_n}^{(p)}(u, \xi, A) := \frac{\xi \theta_\varepsilon}{\varepsilon} \int_A \int_A \eta_\varepsilon(x-z)|u(x) - u(z)|^p \, dx \, dz + \frac{1}{\varepsilon} \int_A V(u(x)) \, dx,
\]

for \( \varepsilon > 0, A \subset X, u : A \to \mathbb{R} \) and \( \xi \in \mathbb{R} \). Indeed, fix \( t > 0 \). Thanks to Assumption (A1) the function \( \rho \) is continuous in \( X \). Then, it is possible to find \( \bar{\eta} \in \mathbb{N} \) such that for all \( n \geq \bar{\eta} \) and all \( y \in \bar{x} + r_n Q \) it holds

\[
|\rho(y) - \rho(\bar{x})| < t.
\]

Thus,

\[
\delta_n \leq \frac{t}{r_n^{d-1} \varepsilon_n} \int_{\bar{x} + r_n Q} V(u_{\varepsilon_n}(x)) \, dx + \frac{t \varepsilon_n \rho(\bar{x})}{r_n^{d-1} \varepsilon_n} \int_{\bar{x} + r_n Q} \int_{\bar{x} + r_n Q} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \, dx \, dz
\]

\[
+ \frac{ts_{\varepsilon_n}}{r_n^{d-1} \varepsilon_n} \int_{\bar{x} + r_n Q} \int_{\bar{x} + r_n Q} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(z) \, dx \, dz
\]

\[
\leq \frac{t}{r_n^{d-1} \varepsilon_n c_1} \int_{\bar{x} + r_n Q} V(u_{\varepsilon_n}(x)) \rho(x) \, dx
\]

\[
+ \frac{t s_{\varepsilon_n} (c_1 + c_2)}{r_n^{d-1} \varepsilon_n c_1^2} \int_{\bar{x} + r_n Q} \int_{\bar{x} + r_n Q} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(z) \rho(x) \, dx \, dz
\]

\[
\leq \frac{t (c_1 + c_2) \lambda_{\varepsilon_n}(\bar{x} + r_n Q)}{r_n^{d-1} c_1^2},
\]

where in the last step we used \((28)\). By \((26)\) and \((27)\) \( \lim_{n \to \infty} \delta_n \leq Ct \) for some constant \( C < \infty \). Since \( t > 0 \) is arbitrary, this proves the claim.
Step 3. Observe that for any \( \lambda \geq 0, \varepsilon > 0, r > 0 \) and \( v \in L^1 \) we have

\[
\min \left\{ 1, \frac{s_n}{s_{\varepsilon}} \right\} \hat{\mathcal{F}}^{(p)}(R_{\bar{x}, r} v, \lambda, Q) \leq \frac{1}{r^{d-1}} \hat{\mathcal{F}}^{(p)}(v, \lambda, \bar{x} + rQ) \leq \max \left\{ 1, \frac{s_n}{s_{\varepsilon}} \right\} \hat{\mathcal{F}}^{(p)}(R_{\bar{x}, r} v, \lambda, Q).
\]

Let \( C = Q \cap \nu(\bar{x})^{-1} \subseteq C(\bar{x}, \nu(\bar{x})) \). Define the function \( w_n : \mathbb{R}^d \rightarrow \mathbb{R} \) as the periodic extension of the function that is \( R_{\bar{x}, r_n} u_{\varepsilon_n} \) in \( Q \) and \( \nu_{\bar{x}} \) in \( T_C \setminus Q \). Set \( \varepsilon_n := \frac{\varepsilon_n}{r_n} \) and \( s_n' = \min \left\{ 1, \frac{s_n}{s_{\varepsilon_n}} \right\} \). Using (28) and (29) together with the scaling identity (18) we get

\[
\frac{\lambda \varepsilon_n(\bar{x} + r_nQ)}{r_n^{d-1}} \geq \rho(\bar{x}) \frac{\hat{\mathcal{F}}^{(p)}(u_{\varepsilon_n}, \rho(\bar{x}), \bar{x} + r_nQ)}{r_n^{d-1}} - \delta_n
\]

\[
\geq s_n' \rho(\bar{x}) \hat{\mathcal{F}}^{(p)}(R_{\bar{x}, r_n} u_{\varepsilon_n}, \rho(\bar{x}), Q) - \delta_n
\]

\[
= s_n' \rho(\bar{x}) \hat{\mathcal{F}}^{(p)}(w_n, \rho(\bar{x}), Q) - \delta_n
\]

\[
\geq s_n' \rho(\bar{x}) G^{(p)}(w_n, \rho(\bar{x}), T_C) - \delta_n
\]

\[
\geq s_n' \rho(\bar{x}) G^{(p)}(w_n, \rho(\bar{x}), T_C) - \delta_n
\]

\[
- s_n' \rho(\bar{x}) \left| \hat{\mathcal{F}}^{(p)}(w_n, \rho(\bar{x}), Q) - G^{(p)}(w_n, \rho(\bar{x}), T_C) \right|.
\]

(30)

We would like to say that

\[
\left| \frac{\hat{\mathcal{F}}^{(p)}(w_n, \rho(\bar{x}), Q) - G^{(p)}(w_n, \rho(\bar{x}), T_C)}{r_n} \right| \rightarrow 0
\]

as \( n \rightarrow \infty \). Unfortunately, this might not be true. In order to overcome this difficulty, take \( t \in (0, 1) \). Notice that we can bound

\[
\frac{r_n \rho(\bar{x})}{\varepsilon_n} \int_{tQ} \int_{tQ} \eta_n(y - z) |w_n(y) - w_n(z)|^p \, dy \, dz.
\]

(31)

Now,

\[
\frac{r_n \rho(\bar{x})}{\varepsilon_n} \int_{tQ} \int_{tQ} \eta_n(y - z) |w_n(y) - w_n(z)|^p \, dy \, dz \leq G^{(p)}(w_n, \rho(\bar{x}), T_C)
\]

(32)

Moreover, using Proposition 3.13 we get that for a.e. \( t \in (0, 1) \) it holds

\[
\Lambda^{(p)}(w_n, tQ, tT_C \setminus tQ) \rightarrow 0, \quad \Lambda^{(p)}(w_n, tT_C \setminus tT_C) \rightarrow 0
\]

(33)

as \( n \rightarrow \infty \). Finally, using Remark 3.14 and the fact that \( w_n \) is constant on \( T_C \setminus Q \) it is easy to see that

\[
\lim_{t \rightarrow 1} \lim_{n \rightarrow \infty} \Lambda^{(p)}(w_n, tT_C \setminus tQ, tT_C \setminus tQ) = 0.
\]

(34)

Hence, from (29), (30), (31), (32), (33) and (34) and recalling that \( s_n' \rightarrow 1 \) we get

\[
\lim_{n \rightarrow \infty} \frac{\lambda \varepsilon_n(\bar{x} + r_nQ)}{r_n^{d-1}} \geq \rho(\bar{x}) \sigma^{(p)}(\bar{x}, \nu(\bar{x}))
\]

as required.
3.4 The Limsup Inequality

This section is devoted at proving the following: let $u \in BV(X, \{\pm 1\})$, then it is possible to find $\{u_{\varepsilon_n}\}_{n=1}^\infty \subset L^1(X)$ with $u_{\varepsilon_n} \to u$ in $L^1(X, \mu)$ such that

$$\limsup_{n \to \infty} F^{(p)}(u_{\varepsilon_n}) \leq G^{(p)}(u) .$$

(35)

Without loss of generality, we can assume $G^{(p)}(u) < \infty$, namely $u \in BV(X; \{\pm 1\})$. The proof will follow the lines of the argument used to prove [1, Theorem 5.2].

Proof of Theorem 3.2 (limsup). We first prove the result for polyhedral functions then, via a diagonalisation argument, generalise to arbitrary functions in $BV(X; \{\pm 1\})$. We fix the sequence $\varepsilon_n \to 0^+$ now.

Step 1. Polyhedral functions. Assume $u \in BV(X, \{\pm 1\})$ is a polyhedral function (see Definition 2.18). Then we claim that there exists a sequence $\{u_{\varepsilon_n}\}_{n=1}^\infty$ with $|u_{\varepsilon_n}| \leq 1$, converging uniformly to $u$ on every compact set $K \subset X \setminus \partial^*\{u = 1\}$, and in particular $u_{\varepsilon_n} \to u$ in $L^1(X, \mu)$, such that (35) holds.

Let us denote by $E$ the polyhedral set $\{u = 1\}$ and by $E_1, \ldots, E_k$ its faces. It is possible to cover $\partial E \cap X$ with a finite family of sets $A_1, \ldots, \tilde{A}_k$, where each $A_i$ is an open set satisfying the following properties (see Figure 3):

(i) $\partial A_i$ can be written as the union of two Lipschitz graphs over the face $E_i$,

(ii) every point in the relative interior of $E_i$ belongs to $A_i$,

(iii) $H^{d-1}(\bar{A}_i \cap \bigcup_{j \neq i} E_j) = 0$,

(iv) $A_i \cap A_j = \emptyset$ if $i \neq j$.

Set

$$A_0 := \{u = 1\} \setminus \bigcup_{i=1}^k \bar{A}_i , \quad A_{k+1} := X \setminus \bigcup_{i=0}^k \bar{A}_i .$$

We then define $u_{\varepsilon_n}$ in each $A_i$ separately. Set $u_{\varepsilon_n}(x) := 1$ for $x \in A_0$ and $u_{\varepsilon_n}(x) := -1$ for $x \in A_{k+1}$. Now fix $i \in \{1, \ldots, k\}$ and $n \in \mathbb{N}$. In order to define $u_{\varepsilon_n}$ in $A_i$, we reason as follows. Denote by $\nu$ the normal of the hyperplane containing $E_i$. Without loss of generality, we can assume $\nu = e_d$ and $E_i \subset \{x_d = 0\}$. A point $x \in \mathbb{R}^d$ will be denoted as

$$x = (x', x_d), \quad x' \in \mathbb{R}^{d-1}, \quad x_d \in \mathbb{R} .$$

Fix $\xi > 0$. Using the continuity of $\sigma$ (see Lemma 3.11) and of $\rho$ in $X$ (see Assumption (A1)) it is possible to find a finite family of $d-1$ dimensional disjoint cubes $\{Q_j\}_{j=1}^{M_\xi}$, for some $M_\xi \in \mathbb{N}$, of side $r_\xi > 0$ lying in the hyperplane containing $E_i$, having $E_i \subseteq \bigcup_{j=1}^{M_\xi} Q_j$, $E_i \cap Q_j \neq \emptyset$, and satisfying the following properties: denoting by $\{x_j\}_{j=1}^{M_\xi}$ their centers (or, in the case the center of a cube $Q_j$ is not contained in $E_i$, a point of $Q_j \cap E_i$) we have

$$\left| \int_{E_i} \sigma^{(p)}(x, u) \rho(x) dH^{d-1}(x) - r_{\xi}^{d-1} \sum_{j=1}^{M_\xi} \sigma^{(p)}(x_j, \nu) \rho(x_j) \right| < \xi .$$

(36)

It is possible to find, for every $j = 1, \ldots, M_\xi$, $C_j \subset C(x_j, \nu)$ and $w_j \in \mathcal{U}(C_j, \nu)$ such that

$$\frac{1}{H^{d-1}(C_j)} G^{(p)}(w_j, \rho(x_j), T_{C_j}) < \sigma^{(p)}(x_j, \nu) + \frac{\xi}{\rho(x_j) M_\xi r_{\xi}^{d-1}} .$$

(37)
We can assume $|w_j| \leq 1$. For every $j = 1, \ldots, M_\xi$, let $L_j \in \mathbb{N}$ be such that
\[
\frac{1}{\mathcal{H}^{d-1}(C_j)} \left| G^{(p)}(w_j, \rho(x_j), T_{C_j}) - G^{(p)}(\tilde{w}_j, \rho(x_j), T_{C_j}) \right| < \frac{\xi}{\rho(x_j)M_\xi r_\xi^{d-1}},
\]
where
\[
\tilde{w}_j(x) := \begin{cases} 
  w_j(x) & \text{if } |x_d| < L_j, \\
  +1 & \text{if } x_d > L_j, \\
  -1 & \text{if } x_d < -L_j. 
\end{cases}
\]
By mollifying, $\tilde{w}_j$, i.e. $\tilde{w}_j' = J_\beta * \tilde{w}_j$, we have that $\tilde{w}_j'$ is Lipschitz continuous with $\text{Lip}(\tilde{w}_j') \leq \frac{1}{\beta}$ and $\tilde{w}_j'(x) \in \{\pm 1\}$ for $|x_d| > L_j + \beta$. We can choose $\beta = \beta_\xi$ such that
\[
\frac{1}{\mathcal{H}^{d-1}(C_j)} \left| G^{(p)}(w_j, \rho(x_j), T_{C_j}) - G^{(p)}(\tilde{w}_j', \rho(x_j), T_{C_j}) \right| < \frac{2\xi}{\rho(x_j)M_\xi r_\xi^{d-1}}. 
\tag{38}
\]

For every $j = 1, \ldots, M_\xi$ cover $Q_j$ with copies of $\varepsilon_n C_j$. Denote them by $\{Q_{j,s}^{(n)}\}_{s=1}^{k_j^{(n)}}$ and by $\{y_{j,s}^{(n)}\}_{s=1}^{k_j^{(n)}}$ their centers. Notice that it might be necessary to consider the intersection of some cubes with $Q_j$, and that
\[
k_j^{(n)} = \left\lceil \frac{1}{\varepsilon_n^{d-1}} r_\xi^{d-1} \left( \mathcal{H}^{d-1}(C_j) \right)^{-1} \right\rceil, 
\tag{39}
\]
where $\lceil x \rceil$ denotes the smallest natural number greater than or equal to $x \in \mathbb{R}$. Define the function $v_j^{(n)} : \mathbb{R}^d \to [-1, +1]$ as the periodic extension of
\[
v_j^{(n)}(x) := \sum_{s=1}^{k_j^{(n)}} \tilde{w}_j'(x - (x_j + y_{j,s}^{(n)}) / \varepsilon_n) \chi_{Q_{j,s}^{(n)}}(x').
\]
We are now in position to define the function $u_{\varepsilon_n}$ in $A_i$: for $x \in A_i$ define

$$u_{\varepsilon_n}(x) := \sum_{j=1}^{M\| \varepsilon_n \|} \chi_{Q_{j,n}^i}(x') v_j^{(n)}(x).$$

Note that $u_{\varepsilon_n}$ has Lipschitz constant $\text{Lip}(u_{\varepsilon_n}) \leq \frac{1}{2\| \varepsilon_n \|}$ and $u_{\varepsilon_n}(x) = u(x)$ for any $x$ with $\text{dist}(x, \partial^* \{u = 1\}) \geq \max_j (L_j + \beta_k \varepsilon_n)$. Hence, $u_{\varepsilon_n} \to u$ as $n \to \infty$ uniformly on compact sets $K \subset X \setminus \partial^* \{u = 1\}$. We now prove the validity of inequality (35). We claim that:

1. $\tilde{\Lambda}_{\varepsilon_n} (u_{\varepsilon_n}, A_i, A_j) \to 0$ as $n \to \infty$ for all $i \neq j$, where $\tilde{\Lambda}$ is defined by (13);
2. $\limsup_{n \to \infty} F_{\varepsilon_n}^{(p)} (u_{\varepsilon_n}, A_i) \leq G_{\infty}^{(p)} (u, A_i)$ for all $i = 0, \ldots, k + 1$.

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

$$\limsup_{n \to \infty} F_{\varepsilon_n}^{(p)} (u_{\varepsilon_n}) \leq \sum_{i=0}^{k+1} \limsup_{n \to \infty} F_{\varepsilon_n}^{(p)} (u_{\varepsilon_n}, A_i) + 2 \sum_{i<j}^{k+1} \limsup_{n \to \infty} \tilde{\Lambda}_{\varepsilon_n} (u_{\varepsilon_n}, A_i, A_j)$$

Note that $u_{\varepsilon_n}$ has Lipschitz constant $\text{Lip}(u_{\varepsilon_n}) \leq \frac{1}{2\| \varepsilon_n \|}$ and $u_{\varepsilon_n}(x) = u(x)$ for any $x$ with $\text{dist}(x, \partial^* \{u = 1\}) \geq \max_j (L_j + \beta_k \varepsilon_n)$. Hence, $u_{\varepsilon_n} \to u$ as $n \to \infty$ uniformly on compact sets $K \subset X \setminus \partial^* \{u = 1\}$. We now prove the validity of inequality (35). We claim that:

1. $\tilde{\Lambda}_{\varepsilon_n} (u_{\varepsilon_n}, A_i, A_j) \to 0$ as $n \to \infty$ for all $i \neq j$, where $\tilde{\Lambda}$ is defined by (13);
2. $\limsup_{n \to \infty} F_{\varepsilon_n}^{(p)} (u_{\varepsilon_n}, A_i) \leq G_{\infty}^{(p)} (u, A_i)$ for all $i = 0, \ldots, k + 1$.

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

$$\limsup_{n \to \infty} F_{\varepsilon_n}^{(p)} (u_{\varepsilon_n}) \leq \sum_{i=0}^{k+1} \limsup_{n \to \infty} F_{\varepsilon_n}^{(p)} (u_{\varepsilon_n}, A_i) + 2 \sum_{i<j}^{k+1} \limsup_{n \to \infty} \tilde{\Lambda}_{\varepsilon_n} (u_{\varepsilon_n}, A_i, A_j)$$

We start by proving claim (ii). It is easy to see that it holds true for $i = 0, k + 1$. Fix $i \in \{1, \ldots, k\}$. Noticing that

$$A_i \subset \bigcup_{j=1}^{M} \bigcup_{s=1}^{k_j} T_{Q_{j,s}^{(n)}},$$

we get

$$F_{\varepsilon_n}^{(p)} (u_{\varepsilon_n}, A_i) \leq F_{\varepsilon_n}^{(p)} \left( u_{\varepsilon_n}, \bigcup_{j=1}^{M} \bigcup_{s=1}^{k_j} T_{Q_{j,s}^{(n)}} \right)$$

Note that, $\varepsilon_n$ we are using the whole cube $Q_{j,s}^{(n)}$, not only its intersection with $Q_j$.

Here, for every $j$ and $s$ we are using the whole cube $Q_{j,s}^{(n)}$, not only its intersection with $Q_j$.

Note that,

$$A_{j,s}^{(n)} := \left\{ (x,z) \in \mathbb{R}^d \times T_{Q_{j,s}^{(n)}} : \eta_{\varepsilon_n} (x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \neq 0 \right\}$$

$$\subseteq B(x_j, \sqrt{2d\varepsilon_n}) \times \left( T_{Q_{j,s}^{(n)}} \cap \{|z| \leq \varepsilon_n(L_j + R_{\eta})\} \right)$$

32
for $\varepsilon_n$ sufficiently small (compared to $r_\xi$). Hence there exists $\delta_\xi > 0$ such that for all $(x, z) \in A_{j,s}^{(n)}$ we have $|\rho(x) - \rho(x_j)| \leq \delta_\xi$ and $|\rho(z) - \rho(x_j)| \leq \delta_\xi$ where $\delta_\xi \to 0$ as $\xi \to 0^+$. This implies

$$s_{\varepsilon_n}(x, z) \rho(x) \rho(z) \leq |s_{\varepsilon_n} - 1| |\rho(x)\rho(z)| + |\rho(x)\rho(z)|\delta_\xi c_2 \leq |s_{\varepsilon_n} - 1| |\rho(x)\rho(z)| + 2\delta_\xi c_2.$$ 

Hence,

$$\frac{s_{\varepsilon_n}}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} \int_{R^d} \eta_{\varepsilon_n}(x - z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(x)\rho(z) \, dx \, dz \leq \frac{|s_{\varepsilon_n} - 1| |\rho(x)\rho(z)| + 2\delta_\xi c_2 + \rho^2(x_j)}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} \int_{R^d} \eta_{\varepsilon_n}(x - z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \, dx \, dz.$$ 

Similarly,

$$\frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} V(u_{\varepsilon_n}(x))\rho(x) \, dx \leq (\rho(x_j) + \delta_\xi) \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} V(u_{\varepsilon_n}(x)) \, dx.$$ 

Let $\gamma_{\xi,n} = 1 + \frac{2\delta_\xi}{c_2} + |s_{\varepsilon_n} - 1|$ and notice that

$$\lim_{\xi \to 0} \lim_{n \to \infty} \gamma_{\xi,n} = 1$$

(taking a subtle dependence that requires $\varepsilon_n$ be small compared to $r_\xi$ which means it is necessary to first take the limits in this order). Then

$$\frac{s_{\varepsilon_n}}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} \int_{R^d} \eta_{\varepsilon_n}(x - z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(x)\rho(z) \, dx \, dz + \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} V(u_{\varepsilon_n}(x))\rho(x) \, dx$$

$$\leq \gamma_{\xi,n}\rho(x_j) \left( \frac{\rho(x_j)}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} \int_{R^d} \eta_{\varepsilon_n}(x - z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \, dx \, dz + \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}}^{(n)}} V(u_{\varepsilon_n}(x)) \, dx \right)$$

$$\leq \gamma_{\xi,n}\rho(x_j) G_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \rho(x_j), T_{Q_{j,s}}^{(n)}).$$

Hence, using \((36), (37), (38), (39)\) and \((40)\), we get

$$F_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, A_i) \leq \gamma_{\xi,n} \sum_{j=1}^{M_z} \sum_{s=1}^{k_j^{(n)}} \rho(x_j) G_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \rho(x_j), T_{Q_{j,s}}^{(n)})$$

$$= \gamma_{\xi,n} \sum_{j=1}^{M_z} \sum_{s=1}^{k_j^{(n)}} \rho(x_j) \varepsilon_n^{d-1} G_1^{(p)}(\tilde{w}^i_j, \rho(x_j), T_{C_j})$$

$$= \gamma_{\xi,n} \sum_{j=1}^{M_z} \rho(x_j) \varepsilon_n^{d-1} H^{d-1}(C_j) G_1^{(p)}(\tilde{w}^i_j, \rho(x_j), T_{C_j})$$

$$\leq \gamma_{\xi,n} \left( \varepsilon_n^{d-1} \sum_{j=1}^{M_z} \rho(x_j) \sigma^{(p)}(x_j, \nu) + 3\xi \right)$$
\[ \leq \gamma_{\xi,n} \left( \int_{E_i} \sigma^{(p)}(x, \nu) \rho(x) \, d\mathcal{H}^{d-1}(x) + 4\xi \right). \]

Taking \( \limsup \) as \( n \to \infty \) and \( \xi \to 0 \) implies (ii). Note that \( u_{\xi_n} \) is not the recovery sequence, rather each \( u_{\xi_n} \) depended on \( \xi \) (through \( L_j \) and \( \beta_{\xi} \)). Hence, if we make the \( \xi \) dependence explicit, we showed \( \limsup \sup_{n \to \infty} \mathcal{F}^{(p)}_\varepsilon(u_{\xi_n}, A_i) \leq \limsup_{n \to \infty} \gamma_{\xi,n} \mathcal{G}^{(p)}(u, A_i) \). By a diagonalisation argument we can find a sequence \( \xi_n \to 0 \) such that \( \limsup_{n \to \infty} \mathcal{F}^{(p)}_\varepsilon(u_{\xi_n}, A_i) \leq \mathcal{G}^{(p)}(u, A_i) \). Moreover if the sequence \( \xi_n \to 0 \) sufficiently slowly then we can infer \( \max L_j + \beta_{\xi_n} \geq \frac{1}{\varepsilon_n} \) and \( \beta_{\xi_n} \leq \varepsilon_n \) hence, \( \text{Lip}(u_{\xi_n}) \leq \frac{1}{\varepsilon_n} \) and \( u_{\xi_n}(x) = u(x) \) whenever \( \text{dist}(x, \partial^* \{u = 1\}) \geq \frac{\alpha_n}{\varepsilon_n} \) for any sequence \( \xi_n \to 0 \).

We are thus left to prove claim (i). Analogously to Remark \[3.14\], \( \Lambda_{\xi_n}(u_{\xi_n}, A_i, A_j) \to 0 \) as \( n \to \infty \) for all \( i, j \) for which \( d(A_i, A_j) > 0 \). Let us now consider indexes \( i \neq j \) for which \( d(A_i, A_j) = 0 \). Write

\[ \Lambda_{\xi_n}(u_{\xi_n}, A_i, A_j) = \frac{1}{\varepsilon_n} \int_{B(0,R_n)} \int_{A_{\varepsilon_n h}} \eta(h) |u_{\varepsilon_n}(z + \varepsilon_n h) - u_{\varepsilon_n}(z)|^p \rho(z + \varepsilon_n h) \rho(z) \, dz \, dh, \]

where \( A_{\varepsilon_n h} := \{ z \in A_i : z + \varepsilon_n h \in A_j \} \). Since \( A_i \) and \( A_j \) are Lipschitz graphs over sets lying on two different hyperplanes, it holds that \( \text{Vol}(A_{\varepsilon_n h}) = O(\varepsilon_n^d) \). Thus

\[ \Lambda_{\xi_n}(u_{\xi_n}, A_i, A_j) \to 0, \]

as \( n \to \infty \).

**Step 2. The general case.** Let \( u \in BV(X; \{\pm 1\}) \). Using Theorem \[2.19\], it is possible to find a sequence \( \{v_n\}_{n=1}^\infty \) of polyhedral function such that \( v_n \to u \) in \( L^1 \) (which, in turn, implies that \( Dv_n \xrightarrow{w^*} Du \) and \( |Dv_n|(X) \to |Du|(X) \)). Using Step 1 and a diagonalisation argument we get that there exists a sequence \( \{u_n\}_{n=1}^\infty \) with \( u_n \to u \) in \( L^1(X) \) such that

\[ \mathcal{F}^{(p)}_\varepsilon(u_n) \leq \mathcal{G}^{(p)}_{\infty}(v_n) + \frac{1}{n}. \]

Then, Theorem \[2.20\] together with Lemma \[3.11\] gives us that

\[ \limsup_{n \to \infty} \mathcal{G}^{(p)}_{\infty}(v_n) = \mathcal{G}^{(p)}_{\infty}(u) . \]

This concludes the proof. \( \square \)

4  **Convergence of the Graphical Model**

In this Section we prove Theorem \[1.6\]. In particular, in Section 4.1 we prove the compactness part of Theorem \[1.6\] and in Sections 4.2, 4.3 we prove the \( \Gamma \)-convergence result.

4.1 **Compactness**

**Proof of Theorem \[1.6\] (Compactness).** \( \textbf{Step 1.} \) We first show that there exist \( c_n, \alpha_n > 0 \) with \( \alpha_n, c_n \to 1 \) as \( n \to \infty \), such that

\[ \eta \left( \frac{T_n(x) - T_n(z)}{\varepsilon_n} \right) \geq c_n \eta \left( \frac{\alpha_n(x - z)}{\varepsilon_n} \right). \] (41)

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

\[ \eta(a) \geq c_n \eta(\alpha_n b). \] (42)
Since by Assumption (A2) we have that \( \delta_n \to 0 \) then \( \alpha_n, c_n \) 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

\[
|a - b| = \frac{|T_n(x) - T_n(z) + z - x|}{\varepsilon_n} \leq 2\|T_n - 1d\|_{L^\infty} = \delta_n
\]

and therefore, by (42), 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 3.** Since by Assumption (A2) we have that \( \delta_n \to 0 \) then \( \alpha_n, c_n \) 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

\[
|a - b| = \frac{|T_n(x) - T_n(z) + z - x|}{\varepsilon_n} \leq 2\|T_n - 1d\|_{L^\infty} = \delta_n
\]

and therefore, by (42), 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.** Let \( \nu_n := u_n \circ T_n \). Using Lemma 2.8 and (41) we have that

\[
\mathcal{G}_n^{(p)}(\nu_n) = \frac{1}{\varepsilon_n^{d+1}} \int_X \int_X \eta \left( \frac{T_n(x) - T_n(z)}{\varepsilon_n} \right) |v_n(x) - v_n(z)|^p \rho(x) \rho(z) \, dx \, dz
\]

\[
+ \frac{1}{\varepsilon_n} \int_X V(v_n(x)) \rho(x) \, dx
\]

\[
\geq \frac{c_n}{\varepsilon_n^{d+1}} \int_X \int_X \eta \left( \frac{\alpha_n(x - z)}{\varepsilon_n} \right) |v_n(x) - v_n(z)|^p \rho(x) \rho(z) \, dx \, dz
\]

\[
+ \frac{1}{\varepsilon_n} \int_X V(v_n(x)) \rho(x) \, dx
\]

\[
= \frac{c_n}{\alpha_n \varepsilon_n^{d+1} \varepsilon_n^{\prime \prime}} \int_X \int_X \eta \left( \frac{x - z}{\varepsilon_n} \right) |v_n(x) - v_n(z)|^p \rho(x) \rho(z) \, dx \, dz
\]

\[
+ \frac{1}{\varepsilon_n} \int_X V(v_n(x)) \rho(x) \, dx
\]

\[
= \frac{1}{\alpha_n} \mathcal{F}_n^{(p)}(\nu_n)
\]

(43)

where \( \varepsilon_n := \frac{\varepsilon_n}{\alpha_n} \) and \( \mathcal{F}_n^{(p)} \) is defined by (4) with \( s_n := \frac{\varepsilon_n}{\alpha_n} \) (in (4) \( s \) depended on \( \varepsilon \) not on \( n \), since we have fixed the sequence \( \varepsilon_n \) then clearly we could write \( n \) in terms of \( \varepsilon_n \), however this would make the notation cumbersome).

**Step 3.** Since \( \lim_{n \to \infty} \alpha_n = 1 \), we can infer that \( \mathcal{F}_n^{(p)}(\nu_n) \) is bounded and hence by Theorem 3.2 the sequence \( \{\nu_n\}_{n=1}^\infty \) is relatively compact in \( L^1 \). Therefore the sequence \( \{\mu_n\}_{n=1}^\infty \) is relatively compact in \( TL^1 \), with any limit \( u \) satisfying \( \mathcal{G}_\infty^{(p)}(u) < \infty \).

**4.2 The Liminf Inequality**

*Proof of Theorem 1.6 (Liminf).* For any \( u \in L^1(\mu) \) and any \( u_n \in L^1(\mu_n) \) with \( u_n \to u \) in \( TL^1 \) we claim that

\[
\liminf_{n \to \infty} \mathcal{G}_n^{(p)}(u_n) \geq \mathcal{G}_\infty^{(p)}(u).
\]

Indeed, taking the liminf on both sides of (43) and using Theorem 3.2 we have

\[
\liminf_{n \to \infty} \mathcal{G}_n^{(p)}(u_n) \geq \liminf_{n \to \infty} \frac{1}{\alpha_n} \mathcal{F}_n^{(p)}(\nu_n) \geq \mathcal{G}_\infty^{(p)}(u)
\]

since \( \lim_{n \to \infty} \frac{1}{\alpha_n} = 1 \).

**4.3 The Limsup Inequality**

*Proof of Theorem 1.6 (Limsup).* The aim of this section is to prove the following: given \( u \in L^1(X, \mu) \) it is possible to find a sequence \( \{u_n\}_{n=1}^\infty \subset L^1(X_n) \) with \( u_n \to u \) in \( TL^1(X) \) such that

\[
\limsup_{n \to \infty} \mathcal{G}_n^{(p)}(u_n) \leq \mathcal{G}_\infty^{(p)}(u).
\]
Without loss of generality we can assume $G^{(p)}_{\infty}(u) < +\infty$. In particular, $u \in BV(\mathcal{X}; \{\pm 1\})$.

We divide the proof in two cases: we first assume that $u$ is a polyhedral function and then we extend the argument to any function $u$ with $G^{(p)}_{\infty}(u) < +\infty$ via a diagonalisation argument.

Case 1. Assume that $u$ is a polyhedral function (see Definition 2.18). Let $w_n \in L^1(\mathcal{X}, \mu)$ be a recovery sequence for the $\Gamma$-convergence of $F^{(p)}_{\varepsilon_n}$ to $G^{(p)}_{\infty}$ where $\varepsilon_n' = \frac{n}{\alpha_n}$ and $\alpha_n$ is a sequence we fix shortly (see step 2) i.e., we assume $w_n \to u$ in $L^1(\mathcal{X}, \mu)$ and $\lim_{n \to \infty} F^{(p)}_{\varepsilon_n}(w_n) = G^{(p)}_{\infty}(u)$. Let $u_n(x_i) = n \int_{T_n^{-1}(x_i)} w_n(x) \, dx$ where $T_n$ is any sequence of transport maps satisfying the conclusions of Theorem 2.10

Step 1. We show that $u_n \to u$ in $TL^1$. Setting $v_n = u_n \circ T_n$, we are required to show $v_n \to u$ in $L^1$. Let $\zeta_n \to 0$ with $\zeta_n \gg \sqrt{\frac{\|T_n - \text{Id}\|_{L^\infty}}{\varepsilon_n}}$. By Theorem 3.2 we can assume that $w_n$ are Lipschitz continuous with $\text{Lip}(w_n) \leq \frac{1}{\zeta_n \varepsilon_n}$ and $w_n(x) = u(x)$ for all $x$ satisfying $\text{dist}(x, \partial^*\{u = 1\}) > \frac{\varepsilon_n}{\zeta_n}$. Hence, $v_n(x) = w_n(x) = u(x)$, for $n$ sufficiently large, for all $x$ such that $\text{dist}(x, \partial^*\{u = 1\}) > \frac{3\varepsilon_n}{\zeta_n}$. Therefore,

$$\|v_n - w_n\|_{L^1(\mathcal{X})} = \sum_{i=1}^{n} \left| \int_{T_n^{-1}(x_i)} |v_n(x) - w_n(x)| \, dx \right|$$

$$\leq n \sum_{i=1}^{n} \left| \int_{T_n^{-1}(x_i)} \int_{T_n^{-1}(x_i)} |w_n(y) - w_n(x)| \, dy \, dx \right|$$

$$\leq \frac{2\|T_n - \text{Id}\|_{L^\infty}}{\zeta_n \varepsilon_n} \# \left\{ i : \text{dist}(x_i, \partial^*\{u = 1\}) \leq \frac{3\varepsilon_n}{\zeta_n} \right\}.$$

Now,

$$\# \left\{ i : \text{dist}(x_i, \partial^*\{u = 1\}) \leq \frac{3\varepsilon_n}{\zeta_n} \right\} \leq \frac{3\varepsilon_n}{\zeta_n} = \eta \left( \left\{ x : \text{dist}(T_n(x), \partial^*\{u = 1\}) \leq \frac{3\varepsilon_n}{\zeta_n} \right\} \right)$$

$$\leq n c_2 \text{Vol} \left( \left\{ x : \text{dist}(x, \partial^*\{u = 1\}) \leq \frac{4\varepsilon_n}{\zeta_n} \right\} \right)$$

$$= O \left( \frac{n \varepsilon_n}{\zeta_n} \right).$$

Hence,

$$\|v_n - w_n\|_{L^1(\mathcal{X})} = O(\varepsilon_n). \quad (44)$$

Moreover,

$$\|v_n - u\|_{L^1(\mathcal{X})} \leq \|v_n - w_n\|_{L^1(\mathcal{X})} + \|w_n - u\|_{L^1(\mathcal{X})} \to 0,$$

so $u_n \to u$ in $TL^1$.

Step 2. We show that there exists $\hat{\alpha}_n, \hat{\varepsilon}_n \to 1$ such that

$$\eta \left( \left| T_n(x) - T_n(z) \right| \right) \leq \hat{\varepsilon}_n \eta \left( \left| \hat{\alpha}_n(x) - \hat{z}_n \right| \right). \quad (45)$$

To show the validity of (45) we use the following subclaim: for all $\hat{\delta} > 0$ sufficiently small there exists $\hat{\alpha}, \hat{\varepsilon}_\delta > 0$ such that $\hat{\alpha}_\delta \to 1, \hat{\varepsilon}_\delta \to 1$, as $\hat{\delta} \to 0$, and, for any $\hat{a}, \hat{b} \in \mathbb{R}^d$, it holds

$$|\hat{a} - \hat{b}| < \hat{\delta} \quad \Rightarrow \quad \hat{\varepsilon}_\delta \eta(\hat{\alpha}_\delta \hat{b}) \geq \eta(\hat{a}). \quad (46)$$

Then (45) can be obtained as follows: for any $n \in \mathbb{N}$ take

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

36
and let $\hat{\alpha}_n$, $\hat{c}_n$ be the numbers given by the subclaim for which (46) 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 infer (45).

To prove the subclaim, we let $\alpha_\delta, c_\delta$ 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 $\delta := \min \left\{ 1, \frac{\hat{\delta}}{\inf_{\gamma \in (0,1]} \alpha_\gamma} \right\}$, trivially $\delta \to 0$ as $\hat{\delta} \to 0$. We assume that

$$\frac{\delta}{\inf_{\gamma \in (0,1]} \alpha_\gamma} \leq 1.$$  

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,

$$|a - b| \leq \frac{\hat{\delta}}{\alpha_\delta} \leq \frac{\delta}{\inf_{\gamma \in (0,1]} \alpha_\gamma} = \delta$$

then

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

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

**Step 3.** Using Lemma 2.8 and (45) we get

$$G_n^{(p)}(u_n) = \frac{1}{\varepsilon_n} \int_X \int_X \eta_{\varepsilon_n}(T_n(x) - T_n(z))|v_n(x) - v_n(z)|^p \rho(x) \rho(z) \, dx \, dz + \frac{1}{\varepsilon_n} \int_X V(v_n(x)) \rho(x) \, dx$$

$$\leq \frac{\hat{c}_n}{\hat{\alpha}_n+1} \frac{1}{\varepsilon_n} \int_X \int_X \eta_{\varepsilon_n}(x - z)|v_n(x) - v_n(z)|^p \rho(x) \rho(z) \, dx \, dz + \frac{1}{\varepsilon_n} \int_X V(v_n(x)) \rho(x) \, dx$$

$$= \frac{\hat{c}_n}{\hat{\alpha}_n+1} \frac{1}{\varepsilon_n} \int_X \int_X \eta_{\varepsilon_n}(x - z)|w_n(x) - w_n(z)|^p \rho(x) \rho(z) \, dx \, dz + \frac{1}{\varepsilon_n} \int_X V(w_n(x)) \rho(x) \, dx$$

$$+ a_n + b_n$$

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

$$a_n := \frac{\hat{c}_n}{\hat{\alpha}_n+1} \frac{1}{\varepsilon_n} \int_X \int_X \eta_{\varepsilon_n}(x - z) (|v_n(x) - v_n(z)|^p - |w_n(x) - w_n(z)|^p) \rho(x) \rho(z) \, dx \, dz,$$

$$b_n := \frac{1}{\varepsilon_n} \int_X (V(v_n(x)) - V(w_n(x))) \rho(x) \, dx.$$  

We recall the followings inequalities: $\forall \delta > 0$ there exists $C_\delta > 0$ such that for any $a, b \in \mathbb{R}^d$ we have

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

(47)

Inequality (47) follows by noticing that, for every given $\delta > 0$, the function $x \mapsto \frac{|x|^{p-1}}{|x-1|^p}$ is bounded in $[0, \infty)$, and then setting $x = \frac{|a|}{|b|}$. Moreover, for all $p \geq 1$ and all $a, b \in \mathbb{R}$, it holds

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

(48)

Fix $\delta > 0$. Using (47) and (48) we infer

$$a_n \leq \frac{\hat{c}_n \delta}{\hat{\alpha}_n+1} \frac{1}{\varepsilon_n} \int_X \int_X \eta_{\varepsilon_n}(x - z)|w_n(x) - w_n(z)|^p \rho(x) \rho(z) \, dx \, dz + d_n$$

where

$$d_n := \frac{2^{p-1} \hat{c}_n C_\delta}{\hat{\alpha}_n+1} \frac{1}{\varepsilon_n} \int_X \int_X \eta_{\varepsilon_n}(x - z) (|v_n(x) - w_n(x)|^p + |v_n(z) - w_n(z)|^p) \rho(x) \rho(z) \, dx \, dz.$$
We show $d_n \to 0$. We have that

$$d_n \leq \frac{2^p \hat{c}_n C p^2}{C_n \alpha_n + 1} \int_X |v_n(x) - w_n(x)|^p \, dx \int_{\mathbb{R}^d} \eta(x) \, dx$$

$$\leq \frac{2^{p-1} \hat{c}_n C p^2}{C_n \alpha_n + 1} \|v_n - w_n\|_{L^1(X)} \int_{\mathbb{R}^d} \eta(x) \, dx$$

$$= O \left( \frac{C \|v_n - w_n\|_{L^1(X)}}{\varepsilon_n} \right),$$

where we used the fact that $\hat{c}_n \to 1$, $\hat{\alpha}_n \to 1$ as $n \to \infty$. From Equation (44) we get that

$$d_n \to 0,$$

(49)
as $n \to \infty$.

Let $L_V$ be the Lipschitz constant for $V$ (as given by Assumption (B4)). Since,

$$b_n \leq \frac{c_2}{\varepsilon_n} \int_X |V(v_n(x)) - V(w_n(x))| \, dx \leq \frac{c_2 L_V}{\varepsilon_n} \|v_n - w_n\|_{L^1(X)}$$

then $b_n \to 0$ by (44).

Combining the above estimates we have,

$$G^{(p)}(u_n) \leq \frac{\hat{c}_n(1 + \delta)}{\hat{\alpha}_n + 1} \int_X \int_X \eta_{\varepsilon_n}(x - z) |w_n(x) - w_n(z)|^p \rho(x) \rho(z) \, dx \, dz$$

$$+ \frac{1}{\varepsilon_n} \int_X V(w_n(x)) \rho(x) \, dx + b_n + d_n.$$  

(50)

**Step 4.** We now conclude as follows. Using (50) we get

$$G^{(p)}(u_n) \leq \frac{1 + \delta}{\hat{\alpha}_n} \mathcal{F}^{(p)}(w_n) + b_n + d_n$$

for $\mathcal{F}^{(p)}_{\varepsilon_n}$ defined as in (4) with $s_n := \frac{\hat{c}_n}{\hat{\alpha}_n}$. Taking the limsup on both sides and using (49), together with Theorem 3.2 we have

$$\limsup_{n \to \infty} G^{(p)}(u_n) \leq \limsup_{n \to \infty} (1 + \delta) \mathcal{F}^{(p)}(w_n) \leq (1 + \delta) G^{(p)}(u).$$

Taking $\delta \to 0$ completes the proof for case 1.

**Case 2.** To extend to arbitrary functions $u \in BV(X; \{\pm 1\})$ we apply the following diagonalisation argument. Using Theorem 2.19 together with Lemma 3.11 and Theorem 2.20 we can find a sequence of polyhedral functions $\{u^{(m)}\}_{m=1}^{\infty}$ such that

$$\|u - u^{(m)}\|_{L^1} \leq \frac{1}{m}, \quad G^{(p)}(u^{(m)}) \leq G^{(p)}(u) + \frac{1}{m}.$$ 

Using the result of Case 1, for each $m \in \mathbb{N}$ we have

$$\limsup_{n \to \infty} G^{(p)}(u^{(m)}) \leq G^{(p)}(u^{(m)}),$$

38
where \( u_n^{(m)} \) is the recovery sequence for \( u^{(m)} \) in the \( \Gamma \)-convergence \( G^{(p)}_\infty = \Gamma \lim_{n \to \infty} F^{(p)}_{ε_n} \). For each \( m \in \mathbb{N} \) let \( n_m \in \mathbb{N} \) be such that
\[
G^{(p)}_n(u_n^{(m)}) \leq G^{(p)}_\infty (u^{(m)}) + \frac{1}{m} \quad \text{and} \quad \| u_n^{(m)} \circ T_n - u^{(m)} \|_{L^1} \leq \frac{1}{m}
\]
for all \( n \geq n_m \). At the cost of increasing \( n_m \) we assume that \( n_{m+1} > n_m \) for all \( m \). Let \( u_n := u_n^{(m)} \) for \( n \in [n_m, n_{m+1}) \). Then,
\[
\limsup_{n \to \infty} G^{(p)}_n(u_n) = \limsup_{m \to \infty} \sup_{n \in [n_m, n_{m+1})} G^{(p)}_n(u_n^{(m)}) \leq \limsup_{m \to \infty} \left( G^{(p)}_\infty (u^{(m)}) + \frac{1}{m} \right) \leq G^{(p)}_\infty (u) .
\]

Similarly,
\[
\lim_{n \to \infty} \| u_n \circ T_n - u \|_{L^1} \leq \limsup_{m \to \infty} \sup_{n \in [n_m, n_{m+1})} \left( \| u_n^{(m)} \circ T_n - u^{(m)} \|_{L^1} + \frac{1}{m} \right) \leq \lim_{m \to \infty} \frac{2}{m} = 0
\]
therefore \( u_n \) converges to \( u \) in \( TL^1 \). Hence \( u_n \) is a recovery sequence for \( u \). This completes the proof. \( \square \)

5 Convergence of Minimizers with Data Fidelity

In this section we prove Corollary 1.9.

**Proof of Corollary 1.9.** In view of Theorem 2.23 and Proposition 2.25 it is enough to prove that, for any \( u \in L^1(X, \mu) \) and any \( \{u_n\}_n^{\infty} \) with \( u_n \in L^1(X_n) \) such that \( u_n \to u \) in \( TL^1(X) \) it holds that
\[
\lim_{n \to \infty} K_n(u_n) = K_\infty (u) .
\]
We can restrict ourselves to sequences \( u_n \to u \) satisfying
\[
\sup_{n \in \mathbb{N}} G^{(p)}_n(u_n) < +\infty . \tag{51}
\]
Let
\[
v_n(x) := \begin{cases} u_n(x) & \text{if } x \in X_n(u_n) , \\
1 & \text{otherwise} . \end{cases}
\]
where
\[
X_n(u_n) := \{ x \in X_n : |u_n(x)|^q \leq R_V \} ,
\]
and \( R_V \) is as in Assumption (B3).

**Step 1.** We claim that
\[
\lim_{n \to \infty} |K_n(u_n) - K_n(v_n)| \to 0 , \tag{52}
\]
as \( n \to \infty \). Indeed we have that
\[
\lim_{n \to \infty} |K_n(u_n) - K_n(v_n)| = \lim_{n \to \infty} \left| \frac{1}{n} \sum_{x_i \not \in X_n(u_n)} (k_n(x_i, u_n(x_i)) - k_n(x_i, 1)) \right|
\]
\[
= \beta \lim_{n \to \infty} \frac{1}{n} \sum_{x_i \not \in X_n(u_n)} (3 + |u_n(x_i)|^q) ,
\]
where in the last equality we used (D2). Now,

\[
\frac{1}{n} \sum_{x_i \notin X_n(u_n)} 1 \leq \frac{1}{nR_V} \sum_{x_i \notin X_n(u_n)} |u_n(x_i)|^q \leq \frac{1}{nR_V} \sum_{x_i \notin X_n(u_n)} V(u_n(x_i)) \leq \frac{e_n G_n^{(p)}(u_n)}{\tau R_V}.
\]

Using (51) we conclude (52).

**Step 2.** We claim that \(v_n \to u\) in \(TL^1(X)\). By a direct computation we get

\[
\|u_n - v_n\|_{L^1(\mu_n)} = \frac{1}{n} \sum_{x_i \notin X_n(u_n)} |u_n(x_i) - 1| \\
\leq \frac{1}{n} \sum_{x_i \notin X_n(u_n)} (1 + |u_n(x_i)|) \\
\leq \frac{1}{n} \sum_{x_i \notin X_n(u_n)} (1 + |u_n(x_i)|^q) \to 0,
\]

so \(v_n \to u\) in \(TL^1\).

**Step 3.** We show that \(K_n(v_n) \to K_\infty(u)\). Consider any subsequence of \(v_n\) which we do not relabel. From Step 2 we have that \(v_n \circ T_n \to u\) in \(L^1\). Thus there exists a further subsequence (not relabeled) such that \(v_n(T_n(x)) \to u(x)\) for almost every \(x \in X\). Using Assumption (D3) we get

\[
\lim_{n \to \infty} k_n(T_n(x), v_n(T_n(x))) = k_\infty(x, u(x)).
\]

Moreover it holds

\[
k_n(T_n(x), v_n(T_n(x))) \leq \beta (1 + |v_n(T_n(x))|^q) \leq \beta (1 + R_V) \tag{54}
\]

Using (53), (54) and applying the Lebesgue’s dominated convergence theorem we get

\[
\lim_{n \to \infty} K_n(v_n) = \lim_{n \to \infty} \int_X k_n(T_n(x), v_n(T_n(x))) \rho(x) \, dx \\
= \int_X k_\infty(x, u(x)) \rho(x) \, dx \\
= K_\infty(u) \tag{55}
\]

Since any subsequence of \(v_n\) has a further subsequence such that \(\lim_{n \to \infty} K_n(v_n) = K_\infty(u)\) then we can conclude the convergence is over the full sequence.

**Step 4.** Using (52) and (55) we conclude that

\[
\lim_{n \to \infty} K_n(u_n) \to K_\infty(u)
\]

as required \(\square\)

**Acknowledgements**

The authors thank the Center for Nonlinear Analysis at Carnegie Mellon University and the Cantab Capital Institute for the Mathematics of Information at the University of Cambridge for their support during the
preparation of the manuscript. In addition the authors would like to thank Dejan Slepčev and Nicolás García Trillos for discussions and references that improved the manuscript. The matlab implementation of the bean used in the example in Section 1.5 was written by Dejan Slepčev. The example in Section 1.6 was aided by Luca Calatroni. The research of R.C. was funded by National Science Foundation under Grant No. DMS-1411646. Part of the research of M.T. was funded by the National Science Foundation under Grant No. CCT-1421502. Both authors would like to thank the referees for valuable feedback that improved the paper.

References


