Euro. Jnl of Applied Mathematics (2020), *vol.* 31, *pp.* 185–231 © Cambridge University Press 2018.
doi:10.1017/S0956792518000645

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

R. CRISTOFERI[1] and M. THORPE[2]

[1]*Department of Mathematics, Heriot-Watt University, Edinburgh, Scotland, EH14 4AS, UK*
e-mail: r.cristoferi@hw.ac.uk
[2]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, CB3 0WA, UK*
e-mail: m.thorpe@maths.cam.ac.uk

(*Received 24 February 2018*; *revised 16 July 2018*; *accepted 1 October 2018*; *first published online 14 November 2018*)

The consistency of a non-local 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 Γ-convergence. It is proved that the discrete model converges to a weighted anisotropic perimeter.

Key words: Ginzburg–Landau functional, PDEs on graphs, non-local variational problems, phase transitions, Γ-convergence

2010 Mathematics Subject Classification: 49J55, 49J45, 62G20

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 minimisation procedure. In the machine learning community, successfully implemented approaches include minimising 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 minimisation procedure approaches some limit minimisation method when the number of elements of the data set goes to infinity. Indeed, knowing whether a specific minimisation strategy is an approximation of a limit (minimising) object can help explain properties of the finite data algorithm. For a consistent methodology properties of the

† 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.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

large data limit will be evident when a large, but finite, number of data points are 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 generalisation 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 underlining energy to minimise 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 ℓ^p (rather than ℓ^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 extension to multi-phase labelling (see [30]) and convergence of the associated gradient flows.

The paper is organised as follows: in the following subsection we define the discrete model, and in Section 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 minimisers 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 $W_{ij} \geqslant 0$ 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 ε controls the scale of interactions on the graph; in particular, choosing ε large implies the graph is dense, and choosing ε 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.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

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

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

where

$$W_{ij} := \eta_{\varepsilon_n}(x_i - x_j) := \frac{1}{\varepsilon_n^d} \eta\left(\frac{x_i - x_j}{\varepsilon_n}\right), \tag{1.1}$$

and V is a double well potential (see Assumption (B4)).

The first term in $\mathcal{G}_n$ 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 $\frac{1}{\varepsilon_n n} \langle u, Lu \rangle_{\mu_n}$, 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 $\frac{1}{\varepsilon_n n} \sum_{i=1}^{n} V(u(x_i))$, as $\varepsilon_n \to 0$; this penalty blows up unless u takes the values ± 1 almost everywhere.

The function η plays the role of a mollifier, and that explains the definition of η_{ε_n}. Moreover, to justify the scaling $\frac{1}{\varepsilon_n}$ we reason as follows: assume η 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 ε_n will give us the typical form of the singular perturbation used in the gradient theory of phase transitions (see [40]), namely

$$\int_X \frac{1}{\varepsilon_n} V(u) + \varepsilon_n^{p-1} |\nabla u|^p.$$

The consistency of the model is studied using Γ-convergence (see Section 2.4), a very important tool introduced by De Giorgi in the 1970s to understand the limiting behaviour of a sequence of functionals (see [21]). This kind of variational convergence gives, almost immediately, convergence of minimisers.

1.2 Infinite data model

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

Definition 1.2 Let $v \in \mathbb{R}^d$. Define $v^\perp := \{z \in \mathbb{R}^d \,:\, z \cdot v = 0\}$. Moreover, for $x \in \mathbb{R}^d$, set

$$\mathcal{C}(x, v) := \left\{C \subset v^\perp \,:\, C \text{ is a } (d-1)\text{-dimensional cube centred at } x\right\}.$$

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

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

We now define the limiting (continuum) model.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Definition 1.3 Let $p \geqslant 1$ and $X \subset \mathbb{R}^d$ be open and bounded, and $\rho \in L^\infty(X)$ a positive function. Define the functional $\mathcal{G}_\infty^{(p)} : L^1(X) \to [0, \infty]$ by

$$\mathcal{G}_\infty^{(p)}(u) := \begin{cases} \displaystyle\int_{\partial^*\{u=1\}} \sigma^{(p)}(x, \nu_u(x))\rho(x)\, \mathrm{d}\mathcal{H}^{d-1}(x) & \text{if } u \in BV(X; \{\pm 1\}), \\ +\infty & \text{else,} \end{cases}$$

where

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

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

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

Here $\partial^*\{u = 1\}$ denotes the reduced boundary of $\{u = 1\}$, and $\nu_u(x)$ is the measure theoretic exterior normal to the set $\{u = 1\}$ at the point $x \in \partial^*\{u = 1\}$ (see Definition 2.13).

Notice that the discrete functional $\mathcal{G}_n^{(p)}$ is non-local while the functional $\mathcal{G}_\infty^{(p)}$ is local. The minimisation 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 in the form of the limiting functional: an anisotropic weighted perimeter. This shows that minimisers of $\mathcal{G}_\infty^{(p)}$ are sets $E \subset X$ whose boundary ∂E (or, to be precise, the reduced boundary $\partial^* E$) will likely be in the region where ρ is small and orthogonal to directions ν 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 are given by a probability measure μ on the set X, hence why we call $\mathcal{G}_\infty^{(p)}$ the continuum model. Thus, we will need to compare functions (the labelling) defined on different sets. To do so we will implement the strategy introduced by García Trillos and Slepčev in [34], which 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 $0 < c_1 \leqslant c_2 < \infty$.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

We extend ρ to a function defined in the whole space $\mathbb{R}^d$ by setting $\rho(x) := 0$ for $x \in \mathbb{R}^d \backslash X$. For all $n \in \mathbb{N}$, consider a point cloud $X_n = \{x_i\}_{i=1}^n \subset X$ and let μ_n be the associated empirical measure (see Definition 2.1). Let $\{\varepsilon_n\}_{n=1}^\infty$ be a positive sequence converging to zero and such that the following rate of convergence holds:

(A2) $\dfrac{\text{dist}_\infty(\mu_n, \mu)}{\varepsilon_n} \to 0$, where $\text{dist}_\infty(\mu_n, \mu)$ is the ∞-Wasserstein distance between the measures μ_n and μ; see Definition 2.3.

Remark 1.4 *When $x_i \overset{\text{iid}}{\sim} \mu$ then (with probability one), Assumption (A2) is implied by $\varepsilon_n \gg \delta_n$, where δ_n is defined in Theorem 2.8. Notice that for $d \geqslant 3$ this lower bound on ε_n ensures that the graph with vertices x_n and edges weighted by W_{ij} (see (1.1)) 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 μ_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 $V \geqslant 0$.
(B3) There exists $\tau > 0, R_V > 1$ such that for all $|s| \geqslant R_V$ that $V(s) \geqslant \tau |s|$.
(B4) V is Lipschitz continuous on $[-1, 1]$.

The assumptions on V imply that in the limit there are only two phases ± 1. Assumption (B3) is used to establish compactness, in particular it is used to show that minimisers can be bounded in L^∞ by 1. The prototypical example of a function $V : \mathbb{R}^d \to \mathbb{R}$ satisfying Assumptions (B1)–(B4) 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) $\eta \geqslant 0$, $\eta(0) > 0$ and η is continuous at $x = 0$.
(C2) η is an even function, i.e. $\eta(-x) = \eta(x)$.
(C3) η has support in $B(0, R_\eta)$, for some $R_\eta > 0$.
(C4) For all $\delta > 0$, there exists c_δ, α_δ such that if $|x - z| \leqslant \delta$, then $\eta(x) \geqslant c_\delta \eta(\alpha_\delta z)$; 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, $\int_{\mathbb{R}^d} \eta(x)|x| \, \mathrm{d}x < \infty$. Indeed, given $\delta > 0$, it is possible to cover $B(0, R_\eta)$ with a finite family $\widetilde{B}_\delta(x_1), \ldots, \widetilde{B}_\delta(x_r)$ of sets of the form*

$$\widetilde{B}_\delta(x_i) := \{\alpha_\delta z \, : \, |z - x_i| < \delta\}.$$

Assumption (C2) is justified by the fact that η plays the role of an interaction potential. Finally, Assumption (C4) is a version of continuity of η we need in order to perform our technical computations. We note that (C4) is general enough to include $\eta(x) = \chi_A$ where $A \subset \mathbb{R}^d$ is open, bounded, convex and $0 \in A$; see [54, Proposition 2.2].

The main result of the paper is the following theorem.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Theorem 1.6 *Let $p \geqslant 1$ and assume Assumptions (A1), (A2), (B1)–(B4) and (C1)–(C4) 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 $\mathcal{G}_\infty^{(p)}(u) < \infty$;*
- *(Γ-convergence) $\Gamma\text{-}\lim_{n\to\infty}(TL^1)\mathcal{G}_n^{(p)} = \mathcal{G}_\infty^{(p)}$.*

The proof of compactness is in Section 4.1 and the Γ-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 Γ-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 non-local functional $\mathcal{G}_n^{(p)}$ to the continuum local one $\mathcal{G}_\infty^{(p)}$ via an intermediate non-local continuum functional $\mathcal{F}_{\varepsilon_n}^{(p)}$ (defined in (3.1)). In particular, we will prove that:

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

(ii) it is possible to bound from below $\mathcal{G}_n^{(p)}$ with $\mathcal{F}_{\varepsilon_n'}^{(p)}$ (see (4.3)), where $\lim_{n\to\infty} \frac{\varepsilon_n'}{\varepsilon_n} = 1$, from which the liminf inequality follows; and

(iii) if $u \in BV(X;\{\pm 1\})$ we use the same recovery sequence u_ε as in [1] to show $\limsup_{\varepsilon\to 0} \mathcal{F}_\varepsilon^{(p)}(u_\varepsilon) \leqslant \mathcal{G}_\infty^{(p)}$, after which we can get an upper bound of $\mathcal{G}_n^{(p)}(u_n)$ in terms of $\mathcal{F}_{\varepsilon_n'}^{(p)}(u_{\varepsilon_n'})$, 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 $\lim_{n\to\infty} \frac{\varepsilon_n'}{\varepsilon_n} = 1$. This will give us the limsup inequality.

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

As an application of Theorem 1.6, we consider the functional $\mathcal{G}_n^{(p)}$ 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 $\mathcal{K}_n : TL^1(X_n) \to \mathbb{R}$ and $\mathcal{K}_\infty : TL^1(X) \to \mathbb{R}$ by

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

and

$$\mathcal{K}_\infty(u,\nu) = \begin{cases} \int_X k_\infty(x,u(x))\rho(x)\,\mathrm{d}x & \text{if } \nu = \mu, \\ +\infty & \text{else} \end{cases}$$

respectively.

We make the following assumptions on k_n, k_∞:

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

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

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

(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 $\mathcal{K}_n, \mathcal{K}_\infty$ to include a data fidelity term in a specific subset of X. Let $B \subset X$ be an open set with $\mathrm{Vol}(B) > 0$ and $\mathrm{Vol}(\partial B) = 0$. Let $\lambda_n \geqslant 0$ with $\lambda_n \to \lambda$ as $n \to \infty$. Let $y_n \in L^1(X_n)$ and $y_\infty \in L^1(X)$ with $\sup_{n\in\mathbb{N}} \|y_n\|_{L^\infty} < \infty$ and such that $y_n(x_{i_n}) \to y_\infty(x)$ for almost every $x \in X$ and any sequence $x_{i_n} \to x$. Define*

$$k_n(x, u) := \begin{cases} \lambda_n |y_n(x) - u|^q & \text{in } B \cap X_n, \\ 0 & \text{on } X_n \backslash B, \end{cases}$$

$$k_\infty(x, u) := \begin{cases} \lambda |y_\infty(x) - u|^q & \text{in } B, \\ 0 & \text{on } X \backslash B. \end{cases}$$

Then k_n and k_∞ satisfy Assumptions (D1)–(D3). Indeed, Assumption (D1) follows directly from the definition of the fidelity terms, while Assumption (D3) holds thanks to continuity and the fact that $\mathrm{Vol}(\partial B) = 0$. Finally, in order to prove Assumption (D2) we simply notice that

$$k_n(x, u) = \lambda_n |y_n(x) - u|^q \leqslant \sup_{n\in\mathbb{N}} \lambda_n 2^{q-1} \left(\|y_n\|_{L^\infty}^q + |u|^q \right) \leqslant \beta(1 + |u|^q),$$

for some $\beta > 0$.

We now consider the minimisation problem:

$$\text{minimise} \quad \mathcal{G}_n^{(p)}(u) + \mathcal{K}_n(u) \quad \text{over } u \in L^1(X_n).$$

Corollary 1 *In addition to Assumptions (A1)–(A3), (B1)–(B2), (C1)–(C4) and (D1)–(D3), assume that for the same $q \geqslant 1$ as in Assumption (D2) there exists $\tau, R_V > 0$ such that for all $|s| \geqslant R_V$ that $V(s) \geqslant \tau |s|^q$. Then any sequence of almost minimisers of $\mathcal{G}_n^{(p)} + \mathcal{K}_n$ is compact in TL^1. And furthermore, any cluster point of almost minimisers is a minimiser of $\mathcal{G}_\infty^{(p)} + \mathcal{K}_\infty$ in $L^1(X)$.*

We prove the corollary in Section 5.

Finally, we comment on the hypothesis $\rho \geqslant c_1 > 0$. If it is omitted, we can still get the following result.

Corollary 2 *Let $p \geqslant 1$ and assume Assumptions (A2), (B1)–(A3) and (C1)–(C4) 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 $\widetilde{\mathcal{G}}_\infty^{(p)} : L^1(X) \to [0, +\infty]$ as*

$$\widetilde{\mathcal{G}}_\infty^{(p)}(u) := \begin{cases} \displaystyle\int_{\partial^* \{u=1\}} \sigma^{(p)}(x, \nu_u(x)) \rho(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) & \text{if } u \in BV_{\mathrm{loc}}(X_+; \{\pm 1\}), \\ +\infty & \text{else}, \end{cases}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

where $BV_{\mathrm{loc}}(X_+;\{\pm 1\})$ *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* $K \subset X_0$ *we have that any sequence* $\{u_n\}_{n=1}^{\infty} \subset L^1(K;\mu_n)$ *satisfying* $\sup_{n\in\mathbb{N}} \mathcal{G}_n^{(p)}(u_n) < \infty$ *is relatively compact in* TL^1 *and each cluster point* u *has* $\widetilde{\mathcal{G}}_\infty^{(p)}(u) < \infty$*; and*
- *(*Γ*-convergence)* $\Gamma\text{-}\lim_{n\to\infty}(TL^1)\mathcal{G}_n^{(p)} = \widetilde{\mathcal{G}}_\infty^{(p)}$.

We omit a rigorous proof of the corollary and just explain the details. Indeed, for a fixed compact set $K \subset X_+$, the continuity of ρ implies that $\min_K \rho \geqslant c_1 > 0$. Thus the result of Corollary 2 follows by applying Theorem 1.6.

1.4 Related works

The functional $\mathcal{G}_n^{(p)}$ in the case $p = 1$ has been considered by the second author and Theil in [54], where a similar Γ-convergence result has been proved. The difference is that, in the case $p = 1$, the limit energy density function $\sigma^{(1)}$ can be given explicitly, via an integral. In [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 TL^p topology, as introduced by García Trillos and Slepčev [34], provides a notion of convergence upon which the Γ-convergence framework can be applied. This method has now been applied in many works, see, for instance, [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.

Non-local 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, non-local functionals have been used by Brezis, Bourgain and Mironescu in [9] to characterise 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 $p \geqslant 1$ in the definition of our functionals $\mathcal{G}_n^{(p)}$ (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 $\mathcal{E}_n^{(p)}$, defined by

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

$$\mathcal{E}_n^{(p)}(u) = \left[\sum_{i,j=1}^{n} W_{ij}^p |u(x_i) - u(x_j)|^p \right]^{\frac{1}{p}},$$

converges as $p \to \infty$ to $\mathcal{E}_n^{(\infty)}$, defined by

$$\mathcal{E}_n^{(\infty)}(u) = \max_{i,j=1,\dots,n} W_{ij} |u(x_i) - u(x_j)|.$$

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 *Fix $n \in \mathbb{N}$, and assume $V : \mathbb{R} \to [0, \infty)$ is continuous and weights $W_{ij} \geqslant 0$ for all $i, j = 1, \dots, n$. Let*

$$\tilde{\mathcal{G}}_n^{(p)}(u) = \left[\frac{1}{\varepsilon_n n^2} \sum_{i,j=1}^{n} W_{ij}^p |u(x_i) - u(x_j)|^p + \frac{1}{\varepsilon_n^p n} \sum_{i=1}^{n} V^p(u(x_i)) \right]^{\frac{1}{p}},$$

$$\mathcal{G}_n^{(\infty)}(u) = \max \left\{ \max_{i,j=1,\dots,n} W_{ij} |u(x_i) - u(x_j)|, \frac{1}{\varepsilon_n} \max_{i=1,\dots,n} V(u(x_i)) \right\}.$$

Then,

$$\Gamma\text{-}\lim_{p\to\infty} \tilde{\mathcal{G}}_n^{(p)} = \mathcal{G}_n^{(\infty)}.$$

Proof The proof is easy since the functionals are on discrete domains. In particular, for the liminf inequality we assume $L^0(X_n) \ni u_p \to u \in L^0(X_n)$ then,

$$\begin{aligned}
\tilde{G}_n^{(p)}(u_p) &\geqslant \left[\frac{1}{\varepsilon_n n^2} W_{rt}^p |u_p(x_r) - u_p(x_t)|^p + \frac{1}{\varepsilon_n^p n} V^p(u_p(x_\ell)) \right]^{\frac{1}{p}} \\
&\geqslant \max \left\{ \frac{1}{\varepsilon_n^{\frac{1}{p}} n^{\frac{2}{p}}} W_{rt} |u_p(x_r) - u_p(x_t)|, \frac{1}{\varepsilon_n n^{\frac{1}{p}}} V(u_p(x_\ell)) \right\} \\
&\to \max \left\{ W_{rt} |u(x_r) - u(x_t)|, \frac{1}{\varepsilon_n} V(u(x_\ell)) \right\},
\end{aligned}$$

for any $r, t, \ell \in \{1, \dots, n\}$. Choosing $(r, t) = \operatorname{argmax}_{i,j} W_{ij} |u(x_i) - u(x_j)|$ and $\ell = \operatorname{argmax}_i V(u(x_i))$ implies

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

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

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

as required. □

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

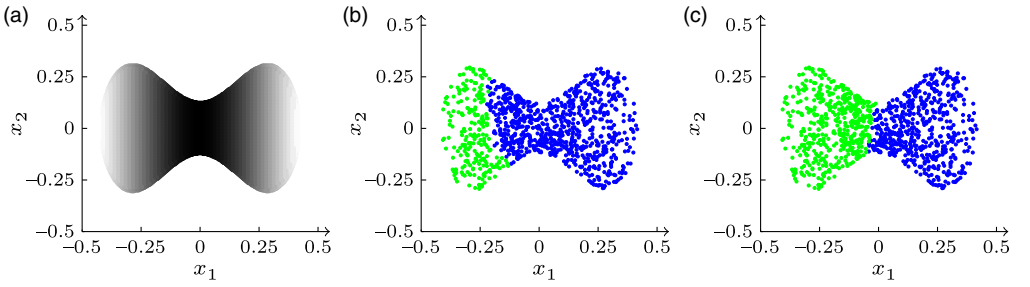


FIGURE 1. How the choice of p affects the trade-off between avoiding high-density regions and minimising the length of the classification boundary. The experiment setup is described in Section 1.5 and we computed minimisers using a gradient flow (known as the Allen–Cahn equation). (a) The density of the data generating distribution ρ. (b) The minimiser of $\mathcal{G}_n^{(2)}$. (c) The minimiser of $\mathcal{G}_n^{(100)}$.

There are several ways we could have renormalised $\mathcal{G}_n^{(p)}$ to obtain a well-defined limit as $p \to \infty$. We choose a normalisation $\tilde{\mathcal{G}}_n^{(p)}$ that closely resembled $\mathcal{G}_n^{(p)}$ and such that the limit $\mathcal{G}_n^{(\infty)}$ will exhibit a phase transition as $n \to \infty$. In particular, the limit functional $\mathcal{G}_n^{(\infty)}$ 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 $\lambda_n = \infty$):

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

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

To illustrate this we consider a simple example. We consider a density ρ as in Figure 1; in particular, the density is given by $\rho(x) \propto \phi(x_1)\lfloor_X$, where ϕ 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 minimising $\mathcal{G}_n^{(p)}$, is therefore a trade-off between minimising 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 minimised.

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 values of a given image at the pixel located at x_i. We assume $\mathcal{I} \subset \{1, \dots, n\}$ indexes the

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

a priori labelled pixels and $f : \{x_i\}_{i\in\mathcal{I}} \to \mathbb{R}$ are the given labels. We wish to label the remaining pixels $\{x_i\}_{i\notin\mathcal{I}}$. Define

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

where $\Phi : \mathbb{R} \to \mathbb{R}$ is a decreasing function and ε 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 η 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 η now depends on more than just the difference between x_i and x_j.

The segmentation is defined by minimising

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

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 ε directly controls the number of pixels connected which can make it easier for parameter selection. Choosing the parameter ε 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 red). We construct a graph as described above with

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

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

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

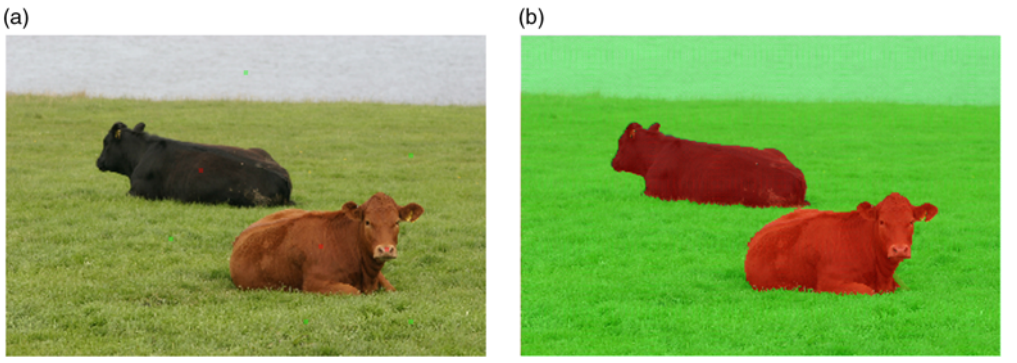


FIGURE 2. An image segmentation example with experimental setup as described in Section 1.6. The image is from the Microsoft image database and also appeared in [7]. (a) Input images with a priori given labels marked. (b) The minimiser of $\mathcal{G}_n^{(2)} + \mathcal{K}_n$.

2 Background

2.1 Notation

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

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

where δ_x denotes the Dirac delta centred at x.

We state our results in terms of a general sequence of empirical measures μ_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 μ.

Remark 2.2 *When $x_i \overset{\text{iid}}{\sim} \mu$ then μ_n converges, with probability one, to μ 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.*

$$\int_X \varphi \, \mathrm{d}\mu_n \to \int_X \varphi \, \mathrm{d}\mu$$

as $n \to \infty$, for all $\varphi \in C_c(X)$.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

We write $L^p(X, \mu; Y)$ for the space of L^p integrable, with respect to μ, 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 Γ-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 μ with μ_n' and is *optimal* in the sense that matching moves as little mass as possible (see Theorem 2.8). 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 μ and λ 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 μ and λ as follows:

- when $1 \leqslant p < \infty$,

$$\operatorname{dist}_p(\mu, \lambda) := \inf \left\{ \left(\int_{X \times X} |x - y|^p \, \mathrm{d}\pi(x, y) \right)^{\frac{1}{p}} \,:\, \pi \in \Gamma(\mu, \lambda) \right\},$$

- when $p = \infty$,

$$\operatorname{dist}_\infty(\mu, \lambda) := \inf \left\{ \operatorname{esssup}_\pi \{\, |x - y| \,:\, (x, y) \in X \times X \,\} \,:\, \pi \in \Gamma(\mu, \lambda) \right\},$$

  where $\operatorname{esssup}_\pi$ denotes the essential supremum with respect to the measure π.

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 pth Wasserstein distance or sometimes the earth movers distance. It is possible to see (see [57]) that the infimum is actually achieved. Moreover, the metric* dist_p *is equivalent to the weak* convergence of probability measures* $\mathcal{P}(X)$ *(plus convergence of pth 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 ρ such that

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

$0 < c_1 \leqslant \rho \leqslant c_2 < \infty$. 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

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

where

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

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

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

where

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

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

Throughout the paper we will assume the empirical measures μ_n converges weakly* to μ (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

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

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

Definition 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 μ_n converges weak* to μ. We say that $u_n \to w$ in $TL^p(X)$, and we write $u_n \xrightarrow{TL^p} w$ if there exists a sequence of stagnating transport maps $\{T_n\}_{n=1}^\infty$ between μ and μ_n such that

$$\|v_n - w\|_{L^p(X,\mu)} \to 0, \tag{2.1}$$

as $n \to \infty$, where $v_n := u_n \circ T_n$.

Remark 2.6 *One can show that if* (2.1) *holds for one sequence of stagnating maps, then it holds for all sequences of stagnating maps [34, Proposition 3.12]. Moreover, since ρ is bounded above and below it holds*

$$\|v_n - \mathrm{Id}\|_{L^p(X,\mu)} \to 0 \quad \Leftrightarrow \quad \|v_n - \mathrm{Id}\|_{L^p(X)} \to 0.$$

We have introduced TL^p convergence $u_n \xrightarrow{TL^p} u$ 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 $TL^p(X)$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

space as the space of couplings (u, μ), where $\mu \in \mathcal{P}(X)$ has finite pth 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

$$\begin{aligned} d_{TL^p}((u,\mu),(v,\lambda)) &:= \min_{\pi \in \Gamma(\mu,\nu)} \left(\int_{X^2} |x-y|^p + |u(x)-v(y)|^p \, \mathrm{d}\pi(x,y) \right)^{\frac{1}{p}} \\ &= \inf_{T_\# \mu = \lambda} \left(\int_X |x - T(x)|^p + |u(x) - v(T(x))|^p \, \mathrm{d}\mu(x) \right)^{\frac{1}{p}}, \end{aligned}$$

or for $p = +\infty$ by

$$\begin{aligned} d_{TL^\infty}((u,\mu),(v,\lambda)) &:= \inf_{\pi \in \Gamma(\mu,\nu)} \left(\operatorname{ess\,inf}_\pi \{ |x-y| + |u(x)-v(y)| \, : \, (x,y) \in X \times X \} \right) \\ &= \inf_{T_\# \mu = \lambda} \left(\operatorname{ess\,inf}_\mu \{ |x - T(x)| + |u(x) - v(T(x))| \, : \, x \in X \} \right). \end{aligned}$$

Proposition 2 *The distance d_{TL^p} is a metric and furthermore, $d_{TL^p}((u_n, \mu_n), (u, \mu)) \to 0$ if and only if $\mu_n \overset{w^*}{\rightharpoonup} \mu$ and there exists a sequence of stagnating transport maps $\{T_n\}_{n=1}^\infty$ between μ and μ_n such that $\|u_n \circ T_n - u\|_{L^p(X,\mu)} \to 0$.*

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 1 *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*

$$\int_X u \, \mathrm{d}T_\# \lambda = \int_X u \circ T \, \mathrm{d}\lambda.$$

Proof Let $s : X \to \mathbb{R}$ be a simple function. Write

$$s = \sum_{i=1}^k a_i \chi_{U_i}.$$

Then

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

The result then follows directly from the definition of the integral. □

Remark 2.7 *Applying the above result to the empirical measures $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $u \in L^1(X_n)$ we get*

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

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

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

In [33], the authors, García Trillos and Slepčev, 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 2.8 *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 $0 < c_1 \leqslant \rho \leqslant c_2 < \infty$. Let $\{x_i\}_{i=1}^{\infty}$ be a sequence of independent and identically distributed random variables distributed on X according to the measure μ, and let μ_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*

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

where

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

Remark 2.9 *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 $x_i \overset{\text{iid}}{\sim} \mu$ and the graph weights are scaled by ε_n with $\varepsilon_n \gg \delta_n$.

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

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

Remark 2.11 *If $E \subset \mathbb{R}^d$ is a set of finite perimeter in X, it is possible to define a finite vector-valued Radon measure $D\chi_E$ on A such that*

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

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

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Definition 2.12 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 2.13 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 $\mathbb{R}^d$ centred 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 2.14 *Let $E \subset \mathbb{R}^d$ be a set with finite perimeter in the open set $X \subset \mathbb{R}^d$. Then*

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

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

(iii) the reduced boundary $\partial^* E$ is $\mathcal{H}^{d-1}$-rectifiable, i.e. there 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 2.15 *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 Sections 3.4 and 4.3 will be done for a special class of functions that we introduce now.

Definition 2.16 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.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Theorem 2.17 *For every $u \in BV(X;\{\pm 1\})$, there exists a sequence $\{u_n\}_{n=1}^{\infty}$ of polyhedral functions such that $u_n \in BV(X;\{\pm 1\})$, $u_n \to u$ in $L^1(X)$ and $|Du_n|(X) \to |Du|(X)$. In particular $Du_n \overset{w^*}{\rightharpoonup} Du$.*

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

Theorem 2.18 *Let $\{E_n\}_{n=1}^{\infty}$ be a sequence of sets of finite perimeter in the open set $X \subset \mathbb{R}^d$ such that $D\chi_{E_n} \overset{w^*}{\rightharpoonup} D\chi_E$ and $|D\chi_{E_n}|(X) \to |D\chi_E|(X)$, where E is a set of finite perimeter in X. Let $f : X \times \mathbb{S}^{d-1} \to [0,\infty)$ be an upper semi-continuous function. Then*

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

Remark 2.19 *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 ν_u, 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 2.17 together with Theorem 2.18 to address the general case.*

2.4 Γ-convergence

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

Definition 2.20 Let (A, d) be a metric space. We say that $F_n : A \to [-\infty, +\infty]$ Γ-converges to $F : A \to [-\infty, +\infty]$, and we write $F_n \overset{\Gamma-(d)}{\longrightarrow} F$ or $F = \Gamma\text{-}\lim(d)_{n\to\infty} F_n$ if the following hold true:

(i) for every $x \in A$ and every $x_n \to x$, we have

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

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

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

With a small abuse of notation we will write $\Gamma\text{-}\lim(L^1)$ and $\Gamma\text{-}\lim(TL^1)$ when the underlying topology is given by 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 Γ-convergence has been designed in order for the following convergence of minimisers and minima result to hold (see e.g. [11, 19]).

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Theorem 2.21 *Let (A, d) be a metric space and let $F_n \xrightarrow{\Gamma-(d)} F$, 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 ε_n-*minimisers *for F_n, i.e.*

$$F_n(x_n) \leqslant \max\left\{ \inf_A F_n + \frac{1}{\varepsilon_n}, -\frac{1}{\varepsilon_n} \right\}. \tag{2.2}$$

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

Remark 2.22 *The condition defining an ε-minimiser 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 (2.2) reduces, for n sufficiently large, to $F_n(x_n) \leqslant \inf_A F_n + \frac{1}{\varepsilon_n}$.*

In the context of this paper we apply Theorem 2.21 in order to prove Corollary 1. In particular, we show that $\mathcal{G}_n^{(p)} + \mathcal{K}_n$ Γ-converges to $\mathcal{G}_\infty + \mathcal{K}_\infty$ 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 Γ-limits.

Proposition 3 *Let (A, d) be a metric space and let $F_n \xrightarrow{\Gamma-(d)} F$. 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 $\mathcal{K}_n$ defined in Definition 1.7 do not continuously converge to $\mathcal{K}_\infty$. However, we show, in Section 5, that $\mathcal{K}_n$ does satisfy the assumptions in Proposition 3.

3 Convergence of the non-local continuum model

We first introduce the intermediary functional $\mathcal{F}_\varepsilon^{(p)}$ that is a non-local continuum approximation of the discrete functional $\mathcal{G}_n^{(p)}$.

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

$$\mathcal{F}_\varepsilon^{(p)}(u, A) = \frac{s_\varepsilon}{\varepsilon} \int_{A\times A} \eta_\varepsilon(x - z)|u(x) - u(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z + \frac{1}{\varepsilon}\int_A V(u(x))\rho(x)\,\mathrm{d}x. \tag{3.1}$$

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

This section is devoted to proving the following result.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Theorem 3.2 *Let $p \geqslant 1$, and assume $s_\varepsilon \to 1$ as $\varepsilon \to 0$. Under conditions (A1), (B1)–(B3) and (C1)–(C3), the following holds:*

- *(Compactness) Let $\varepsilon_n \to 0^+$ and $u_n \in L^1(X,\mu)$ satisfy $\sup_{n\in\mathbb{N}} \mathcal{F}^{(p)}_{\varepsilon_n}(u_n) < \infty$, then u_n is relatively compact in $L^1(X,\mu)$ and each cluster point u has $\mathcal{G}^{(p)}_\infty(u) < \infty$.*
- *(Γ-convergence) $\Gamma\text{-}\lim_{\varepsilon\to 0}(L^1)\mathcal{F}^{(p)}_\varepsilon = \mathcal{G}^{(p)}_\infty$ and furthermore, if $u \in L^1(X,\mu)$ is a polyhedral function, then for any $\zeta_\varepsilon \to 0$ the recovery sequence u_ε can be chosen to satisfy the following: each u_ε is Lipschitz continuous with $\mathrm{Lip}(u_\varepsilon) = \frac{1}{\zeta_\varepsilon \varepsilon}$ and $u_\varepsilon(x) = u(x)$ for all x satisfying $\mathrm{dist}(x, \partial^*\{u = 1\}) > \frac{\varepsilon}{\zeta_\varepsilon}$.*

See Section 3.1 for compactness and Sections 3.3 and 3.4 for the Γ-convergence.

The result of Theorem 3.2 is a generalisation 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 $\mathcal{E}_\varepsilon, \mathcal{E}_0 : L^1(X) \to [0,\infty]$ by

$$\mathcal{E}_\varepsilon(u) := \frac{1}{\varepsilon}\int_X\int_X \eta_\varepsilon(x-z)|u(x)-u(z)|^2\,\mathrm{d}x\,\mathrm{d}z + \frac{1}{\varepsilon}\int_X V(u(x))\,\mathrm{d}x,$$

$$\mathcal{E}_0(u) := \begin{cases} \displaystyle\int_{\partial^*\{u=1\}} \hat{\sigma}(\nu_u(x))\,\mathrm{d}\mathcal{H}^{d-1}(x) & \text{if } u \in BV(X,\{\pm 1\}), \\ +\infty & \text{else,} \end{cases}$$

respectively, where

$$\hat{\sigma}(\nu) := \min\left\{ E(f;\nu) \,|\, f : \mathbb{R}\to\mathbb{R},\ \lim_{t\to\infty} f(t) = 1,\ \lim_{t\to-\infty} f(t) = -1 \right\},$$
$$E(f;\nu) := \int_{-\infty}^{\infty}\int_{\mathbb{R}^d} \eta(h)|f(t+h_\nu) - f(t)|^2\,\mathrm{d}h\,\mathrm{d}t + \int_{-\infty}^{\infty} V(f(t))\,\mathrm{d}t,$$

and $h_\nu := h\cdot\nu$.

Remark 3.4 *We make the following observations:*

(1) The fact that there exists a minimiser 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 $tf(t) \geqslant 0$ for all $t \in \mathbb{R}$.
(3) If η is isotropic, i.e. $\eta(h) = \eta(|h|)$, then $E(f;\nu)$ is independent of ν and therefore $\sigma(\nu)$ is a constant and the functional $\mathcal{E}_0$ is a multiple of the perimeter $\mathcal{H}^{d-1}(\partial^*\{u=1\})$.

The following theorem summarises 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)–(B3), and η satisfies conditions (C1)–(B3). Then the following holds:*

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

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

In fact the above theorem is true if (C1) is relaxed to $\eta \geqslant 0$ (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 $p \geqslant 1$, and that we have a density ρ. The compactness proof is in Section 3.1 and the Γ-convergence in Sections 3.3 and 3.4.

3.1 Compactness

The aim of this section is to show that any sequence $\{u_n\}_{n=1}^\infty \subset L^1(X,\mu)$ satisfying $\sup_{n\in\mathbb{N}} \mathcal{F}_{\varepsilon_n}^{(p)}(u_n) < \infty$ is relatively compact in $L^1(X,\mu)$ and that $\mathcal{G}_\infty^{(p)}(u) < \infty$ 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 ρ that imply an 'equivalence' between $\mathcal{E}_{\varepsilon_n}$ and $\mathcal{F}_{\varepsilon_n}^{(2)}$. When $p \neq 2$ we approximate u_n with a sequence v_n satisfying $v_n(x) \in \{\pm 1\}$, then since $|v_n(x) - v_n(z)|^2 = 2^{2-p}|v_n(x) - v_n(z)|^p$ we can easily find an equivalence between $\mathcal{E}_{\varepsilon_n}$ and $\mathcal{F}_{\varepsilon_n}^{(p)}$. We start with the preliminary result that shows $\mathcal{E}_0(u) < \infty \Leftrightarrow \mathcal{G}_\infty^{(p)}(u) < \infty$.

Proposition 4 *Let* $X \subset \mathbb{R}^d$ *be open and bounded, and let* $u \in L^1(X;\{\pm 1\})$. *Under Assumptions (A1), (B1)–(B2), (C1) and (C3), we have*

$$\int_{\partial^*\{u=1\}} \hat{\sigma}(\nu_u(x))\,\mathrm{d}\mathcal{H}^{d-1}(x) < +\infty \quad \Leftrightarrow \quad \mathcal{H}^{d-1}(\partial^*\{u=1\}) < +\infty$$
$$\Rightarrow \quad \int_{\partial^*\{u=1\}} \sigma^{(p)}(x,\nu_u(x))\rho(x)\,\mathrm{d}\mathcal{H}^{d-1}(x) < +\infty.$$

Remark 3.6 *The above proposition implies that*

$$\mathcal{E}_0(u) < +\infty \quad \Leftrightarrow \quad u \in BV(X;\{\pm 1\}) \quad \Leftrightarrow \quad \mathcal{G}_\infty^{(p)}(u) < +\infty,$$

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

Remark 3.7 *The missing implication,*

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

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 2. Since the missing implication is not needed we do not include it.*

Proof of Proposition 4 *Step 1.* We show

$$\mathcal{H}^{d-1}(\partial^*\{u=1\}) < +\infty \quad \Longrightarrow \quad \int_{\partial^*\{u=1\}} \hat{\sigma}(\nu_u(x))\,\mathrm{d}\mathcal{H}^{d-1}(x) < +\infty.$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

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

$$E(\hat{f}; \nu) = \int_{-\infty}^{\infty} \int_{\mathbb{R}^d} \eta(h) |\hat{f}(t + h_\nu) - \hat{f}(t)|^2 \, \mathrm{d}h \, \mathrm{d}t = \int_{\mathbb{R}^d} \eta(h) |h_\nu| \, \mathrm{d}h \leqslant R_\eta \int_{\mathbb{R}^d} \eta(h) =: C,$$

where $C \in (0, \infty)$, $h_\nu = h \cdot \nu$ and R_η is given by Assumption (C3). Note that C is independent of ν. So,

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

Step2. We show

$$\int_{\partial^*\{u=1\}} \hat{\sigma}(\nu_u(x)) \, \mathrm{d}\mathcal{H}^{d-1}(x) < +\infty \quad \Longrightarrow \quad \mathcal{H}^{d-1}(\partial^*\{u = 1\}) < +\infty.$$

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

$$\hat{f} \text{ is non-decreasing}, \quad \lim_{t\to\infty} \hat{f}(t) = -\lim_{t\to-\infty} \hat{f}(t) = 1, \quad \hat{f}(t)t \geqslant 0 \text{ for } t \in \mathbb{R}. \tag{3.2}$$

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

$$E(\hat{f}; \nu) \geqslant \frac{a}{2} \inf_{t\in[0,\frac{1}{2}]} V(t) =: \tilde{a}_1 > 0. \tag{3.3}$$

Otherwise, if $\hat{f}\left(\frac{a}{2}\right) > \frac{1}{2}$, then for $h \geqslant \frac{a}{2}$

$$\int_{\mathbb{R}} |\hat{f}(t + h) - \hat{f}(t)|^2 \, \mathrm{d}t \geqslant \int_{\frac{a}{2}-h}^{0} |\hat{f}(t + h) - \hat{f}(t)|^2 \, \mathrm{d}t \geqslant \frac{h - \frac{a}{2}}{4}.$$

Similarly, if $h \leqslant -\frac{a}{2}$ we have

$$\int_{\mathbb{R}} |\hat{f}(t + h) - \hat{f}(t)|^2 \, \mathrm{d}t \geqslant \frac{\frac{a}{2} - h}{4}.$$

So, for all $|h| \geqslant \frac{a}{2}$, it holds

$$\int_{\mathbb{R}} |\hat{f}(t + h) - \hat{f}(t)|^2 \, \mathrm{d}t \geqslant \frac{|h - \frac{a}{2}|}{4}.$$

Therefore,

$$\begin{aligned} E(\hat{f}; \nu) &\geqslant \int_{\mathbb{R}^d} \eta(h) \int_{\mathbb{R}} |\hat{f}(t + h_\nu) - \hat{f}(t)|^2 \, \mathrm{d}t \, \mathrm{d}h \\ &\geqslant \frac{1}{4} \int_{\{h\in\mathbb{R}^d \,:\, |h_\nu| \geqslant \frac{3a}{4}\}} \eta(h) \left| h_\nu - \frac{a}{2} \right| \, \mathrm{d}h \\ &\geqslant \frac{ac}{16} \int_{\{h\in\mathbb{R}^d \,:\, |h_\nu| \geqslant \frac{3a}{4}\}} \chi_{B(0,a)} \, \mathrm{d}h \\ &=: \tilde{a}_2 > 0. \end{aligned} \tag{3.4}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Set $\tilde{a} := \min\{\tilde{a}_1, \tilde{a}_2\} > 0$. Using (3.3) and (3.4) we have, for any $f : \mathbb{R} \to \mathbb{R}$ satisfying (3.2), that $E(\hat{f}; \nu) \geqslant \tilde{a}$. We get

$$\tilde{a}\mathcal{H}^{d-1}(\partial^*\{u = 1\}) \leqslant \int_{\partial^*\{u=1\}} \hat{\sigma}(\nu_u(x))\, \mathrm{d}\mathcal{H}^{d-1}(x) = \mathcal{E}_0(u) < \infty.$$

Step 3. We show

$$\mathcal{H}^{d-1}(\partial^*\{u = 1\}) < +\infty \quad \Longrightarrow \quad \int_{\partial^*\{u=1\}} \sigma^{(p)}(x, \nu_u(x))\rho(x)\, \mathrm{d}\mathcal{H}^{d-1}(x) < +\infty.$$

Following the same argument as in the first part of the proof, we let $\hat{f}(t) = 1$ for $t \geqslant 0$ 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,

$$\begin{aligned}\frac{1}{\mathcal{H}^{d-1}(C)} G^{(p)}(f_\nu, \rho(\bar{x}), T_C) &= \rho(\bar{x}) \int_{-\infty}^{\infty} \int_{\mathbb{R}^d} \eta(h)|\hat{f}(t + h_\nu) - \hat{f}(t)|^p \, \mathrm{d}h \, \mathrm{d}t \\ &\leqslant 2^p c_2 \int_{\mathbb{R}^d} \eta(h)|h_\nu| \, \mathrm{d}h \\ &\leqslant \tilde{c},\end{aligned}$$

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

$$\begin{aligned}&\int_{\partial^*\{u=1\}} \sigma^{(p)}(x, \nu_u(x))\rho(x)\, \mathrm{d}\mathcal{H}^{d-1}(x) \\ &\qquad \leqslant \int_{\partial^*\{u=1\}} \frac{1}{\mathcal{H}^{d-1}(C)} G^{(p)}(f_\nu, \rho(\bar{x}), T_C)\rho(\bar{x})\, \mathrm{d}\mathcal{H}^{d-1}(\bar{x}) \\ &\qquad \leqslant c_2\tilde{c}\mathcal{H}^{d-1}(\partial^*\{u = 1\}).\end{aligned}$$

This concludes the proof. □

By the above proposition it is enough to show that any sequence $\{u_n\}_{n=1}^{\infty} \subset L^1(X, \mu)$ satisfying $\sup_{n\in\mathbb{N}} \mathcal{F}_{\varepsilon_n}(u_n) < +\infty$ is relatively compact and any cluster point u satisfies $\mathcal{E}_0(u) < \infty$. We do this by a direct comparison with $\mathcal{E}_{\varepsilon_n}$.

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

$$v_n(x) := \operatorname{sign}(u_n) := \begin{cases} +1 \text{ if } u_n(x) \geqslant 0 \\ -1 \text{ if } u_n(x) < 0. \end{cases}$$

We claim that

$$\|u_n - v_n\|_{L^1} \to 0, \tag{3.5}$$

$$\sup_{n\in\mathbb{N}} \mathcal{F}^{(p)}_{\varepsilon_n}(v_n) < +\infty. \tag{3.6}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Step 1. Let us first prove (3.5). Fix $\delta > 0$ and let

$$K_n^{(\delta)} = \{x \in X \,:\, |u_n(x)| \geqslant 1 + \delta\},$$
$$L_n^{(\delta)} = \{x \in X \,:\, |u_n(x)| \leqslant 1 - \delta\}.$$

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

$$\begin{aligned}\int_X |u_n(x) - v_n(x)|\,\mathrm{d}x &\leqslant \delta \mathrm{Vol}(X) + \int_{K_n^{(\delta)}} |u_n(x) - v_n(x)|\,\mathrm{d}x + \int_{L_n^{(\delta)}} |u_n(x) - v_n(x)|\,\mathrm{d}x \\ &\leqslant \delta \mathrm{Vol}(X) + \int_{K_n^{(\delta)}} |u_n(x)|\,\mathrm{d}x + \mathrm{Vol}(K_n^{(\delta)}) + 2\mathrm{Vol}(L_n^{(\delta)}).\end{aligned}$$

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

$$\mathrm{Vol}(K_n^{(\delta)}) \leqslant \frac{1}{\gamma_\delta} \int_{K_n^{(\delta)}} V(u_n(x))\,\mathrm{d}x \leqslant \frac{\varepsilon_n}{c_1 \gamma_\delta} \mathcal{F}_{\varepsilon_n}^{(p)}(u_n).$$

By the same calculation $\mathrm{Vol}(L_n^{(\delta)}) \leqslant \frac{\varepsilon_n}{c_1\gamma_\delta} \mathcal{F}_{\varepsilon_n}^{(p)}(u_n)$. Furthermore,

$$\begin{aligned}\int_{K_n^{(\delta)}} |u_n(x)|\,\mathrm{d}x &= \int_{K_n^{(\delta)} \cap \{|u_n(x)| \leqslant R_V\}} |u_n(x)|\,\mathrm{d}x + \int_{\{|u_n(x)| > R_V\}} |u_n(x)|\,\mathrm{d}x \\ &\leqslant R_V \mathrm{Vol}(K_n^{(\delta)}) + \frac{1}{\tau} \int_{\{|u_n(x)| > R_V\}} V(u_n(x))\,\mathrm{d}x \\ &\leqslant \left(\frac{R_V}{\gamma_\delta} + \frac{1}{\tau}\right) \frac{\varepsilon_n \mathcal{F}_{\varepsilon_n}^{(p)}(u_n)}{c_1}.\end{aligned}$$

So, $\lim_{n\to\infty} \|u_n - v_n\|_{L^1} \leqslant \delta \mathrm{Vol}(X)$. Since this is true for all $\delta > 0$, we have $\lim_{n\to\infty} \|u_n - v_n\|_{L^1} = 0$ which proves claim (3.5).

Step 2. In order to prove (3.6) we reason as follows. If $|u_n(x)| \geqslant \frac{1}{2}$ then $\mathrm{sign}(u_n(x)) \neq \mathrm{sign}(u_n(y))$ implies $|u_n(x) - u_n(y)| \geqslant \frac{1}{2}$. Now since,

$$|v_n(x) - v_n(y)| = \begin{cases} 0 \text{ if } \mathrm{sign}(u_n(x)) = \mathrm{sign}(u_n(y)), \\ 2 \text{ if } \mathrm{sign}(u_n(x)) \neq \mathrm{sign}(u_n(y)), \end{cases}$$

then $|v_n(x) - v_n(y)| \leqslant 4|u_n(x) - u_n(y)|$ when $|u_n(x)| \geqslant \frac{1}{2}$.

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

$$\begin{aligned}\mathcal{F}_{\varepsilon_n}^{(p)}(v_n) = &\frac{s_n}{\varepsilon_n} \int_{M_n \times X} \eta_{\varepsilon_n}(x-z)|v_n(x) - v_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\ &+ \frac{s_n}{\varepsilon_n} \int_{M_n^c \times X} \eta_{\varepsilon_n}(x-z)|v_n(x) - v_n(y)|^p \rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}z\end{aligned}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

$$\leqslant \frac{2^p c_2^2 s_n}{\varepsilon_n} \int_{M_n \times X} \eta_{\varepsilon_n}(x-z)\,\mathrm{d}x\,\mathrm{d}z + \frac{4^p s_n}{\varepsilon_n} \int_{M_n^c \times X} \eta_{\varepsilon_n}(x-z)|u_n(x)-u_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \leqslant \frac{2^p c_2^2 s_n}{\varepsilon_n} \int_{\mathbb{R}^d} \eta(w)\,\mathrm{d}w \mathrm{Vol}(M_n) + 4^p \mathcal{F}_{\varepsilon_n}^{(p)}(u_n).$$

Now,

$$\frac{1}{\varepsilon_n}\mathrm{Vol}(M_n) \leqslant \frac{1}{\gamma \varepsilon_n} \int_{\{|u_n(x)| \leqslant \frac{1}{2}\}} V(u_n(x))\,\mathrm{d}x \leqslant \frac{1}{\gamma c_1} \mathcal{F}_{\varepsilon_n}^{(p)}(u_n),$$

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

Step 3. We conclude the proof by noticing that, since $v_n \in L^1(X;\{\pm 1\})$ we have $\mathcal{F}_{\varepsilon_n}^{(p)}(v_n) = 2^{p-2} s_n \mathcal{E}_{\varepsilon_n}(v_n)$. So v_n is relatively compact in L^1 by Theorem 3.5 and by Proposition 4 $\mathcal{G}_\infty^{(p)}(u) < +\infty$ for any cluster point u of $\{v_n\}_{n=1}^\infty$. Using (3.5) 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 Γ-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 ν.

Lemma 2 *Fix $\nu \in \mathbb{S}^{d-1}$ and let $C \subset \nu^\perp$. Let $p \geqslant 1$ and assume Assumptions (B1)–(B3) and (C1)–(C3) 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 λ into the mollifier η and then the only difference between Lemma 2 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 2 holds. We omit the proof of Lemma 2.

Remark 3.8 *By Lemma 2 we can write*

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

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 $\sigma^{(p)}$.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Lemma 3 *Under Assumptions (A1) and (C3), the followings hold:*

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

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

By Remark 3.8 we can assume that $w(x) = \bar{w}(x \cdot \nu)$ for some increasing function $\bar{w} : \mathbb{R} \to [-1, 1]$. Then,

$$\begin{aligned}
&\int_{T_D} |w(z+h) - w(z)|^p \, \mathrm{d}z \\
&\quad \leqslant 2^{p-1} \mathcal{H}^{d-1}(D) \int_{\mathbb{R}} |\bar{w}(z_\nu + h_\nu) - \bar{w}(z_\nu)| \, \mathrm{d}z_\nu \quad \text{where } h_\nu = h \cdot \nu \\
&\quad = 2^{p-1} \mathcal{H}^{d-1}(D) \lim_{M \to \infty} \sum_{i=-M}^{M} \int_0^{h_\nu} |\bar{w}(y + (i+1)h_\nu) - \bar{w}(y + ih_\nu)| \, \mathrm{d}y \\
&\quad = 2^{p-1} \mathcal{H}^{d-1}(D) \lim_{M \to \infty} \int_0^{h_\nu} \bar{w}(y + (M+1)h_\nu) - \bar{w}(y - Mh_\nu) \, \mathrm{d}y \\
&\quad = 2^p \mathcal{H}^{d-1}(D) h_\nu,
\end{aligned}$$

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

$$\int_{T_D} |w(z+h) - w(z)|^p \, \mathrm{d}z \leqslant 2^p \mathcal{H}^{d-1}(D) h \cdot \nu. \tag{3.7}$$

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

$$C_n := R_n(D - \bar{x}) + x_n,$$

$$u_n(x) := w(R_n^{-1}(x - x_n) + \bar{x}),$$

respectively. Then

$$\begin{aligned}
\sigma^{(p)}(x_n, \nu_n) &\leqslant \frac{1}{\mathcal{H}^{d-1}(C_n)} G^{(p)}(u_n, \rho(x_n), T_{C_n}) \\
&\leqslant \frac{1}{\mathcal{H}^{d-1}(D)} G^{(p)}(w, \rho(\bar{x}), T_D) + \delta_n \\
&\leqslant \sigma^{(p)}(\bar{x}, \nu) + t + \delta_n,
\end{aligned}$$

where

$$\delta_n := \frac{1}{\mathcal{H}^{d-1}(D)} \left| G^{(p)}(u_n, \rho(x_n), T_{C_n}) - G^{(p)}(w, \rho(\bar{x}), T_D) \right|.$$

We claim that $\delta_n \to 0$ as $n \to \infty$. Since $t > 0$ is arbitrary, this will prove the upper semi-continuity.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Fix $\varepsilon > 0$. Using the continuity of ρ, for n sufficiently large we have that $|\rho(x_n) - \rho(\bar{x})| < \varepsilon$. Thus, we get

$$\begin{aligned}
\delta_n &= \frac{1}{\mathcal{H}^{d-1}(D)} \left| \rho(x_n) \int_{T_D} \int_{\mathbb{R}^d} \eta(R_n h)\, |w(z+h) - w(z)|^p \, \mathrm{d}h \, \mathrm{d}z \right. \\
&\quad \left. - \rho(\bar{x}) \int_{T_D} \int_{\mathbb{R}^d} \eta(h)\, |w(z+h) - w(z)|^p \, \mathrm{d}h \, \mathrm{d}z \right| \\
&\leqslant \frac{|\rho(x_n) - \rho(\bar{x})|}{\mathcal{H}^{d-1}(D)} \int_{\mathbb{R}^d} \eta(h) \int_{T_D} |w(z+h) - w(z)|^p \, \mathrm{d}z \, \mathrm{d}h \\
&\quad + \frac{\rho(x_n)}{\mathcal{H}^{d-1}(D)} \int_{\mathbb{R}^d} |\eta(R_n h) - \eta(h)| \int_{T_D} |w(z+h) - w(z)|^p \, \mathrm{d}z \, \mathrm{d}h \\
&\leqslant 2^p \varepsilon R_\eta \|\eta\|_{L^1} + 2^p c_2 R_\eta \int_{\mathbb{R}^d} |\eta(R_n h) - \eta(h)| \, \mathrm{d}h,
\end{aligned}$$

where in the last step we used (3.7). 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 L^p. For every $s > 0$, let $\widetilde{\eta}_s : \mathbb{R} \to [0, \infty)$ be a continuous function with support in $B(0, 2R_\eta)$ such that $\|\eta - \widetilde{\eta}_s\|_{L^1(\mathbb{R}^d)} < s$. Then, for every $r > 0$ there exists $\bar{n} \in \mathbb{N}$ such that $|\widetilde{\eta}_s(R_n h) - \widetilde{\eta}_s(h)| < r$ for all $h \in \mathbb{R}^d$ and all $n \geqslant \bar{n}$. So that, for $n \geqslant \bar{n}$

$$\int_{\mathbb{R}^d} |\eta(R_n h) - \eta(h)| \, \mathrm{d}h \leqslant \|\eta - \widetilde{\eta}_s\|_{L^1(\mathbb{R}^d)} + \int_{\mathbb{R}^d} |\widetilde{\eta}_s(R_n h) - \widetilde{\eta}_s(h)| \, \mathrm{d}h \leqslant s + r \mathrm{Vol}(B(0, 2R_\eta)).$$

Since r and s are arbitrary, we conclude that $\int_{\mathbb{R}^d} |\eta(R_n h) - \eta(h)| \, \mathrm{d}h \to 0$ as $n \to \infty$ and, in turn, that $\delta_n \to 0$ as $n \to \infty$.

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

We claim that $\sigma^{(p)}(\bar{x}, \nu) \leqslant \liminf_{n\to\infty} \sigma^{(p)}(x_n, \nu)$. Without loss of generality, let us assume that

$$\liminf_{n\to\infty} \sigma^{(p)}(x_n, \nu) = \lim_{n\to\infty} \sigma^{(p)}(x_n, \nu) < \infty.$$

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

$$\frac{1}{\mathcal{H}^{d-1}(C_n)} G^{(p)}(u_n, \rho(x_n), T_{C_n}) \leqslant \sigma^{(p)}(x_n, \nu) + \frac{1}{n}. \tag{3.8}$$

Set $\lambda_n := \bar{x} - x_n$ and define

$$\widetilde{C}_n := C_n + \lambda_n, \qquad \widetilde{u}_n(x) := u_n(x - \lambda_n).$$

So, $\widetilde{C}_n \in \mathcal{C}(\bar{x}, \nu)$ and $\widetilde{u}_n \in \mathcal{U}(\widetilde{C}_n, \nu)$. Using (3.8), we get

$$\begin{aligned}
\sigma^{(p)}(\bar{x}, \nu) &\leqslant \frac{1}{\mathcal{H}^{d-1}(\widetilde{C}_n)} G^{(p)}(\widetilde{u}_n, \rho(\bar{x}), T_{\widetilde{C}_n}) \\
&\leqslant \frac{G^{(p)}(u_n, \rho(x_n), T_{C_n})}{\mathcal{H}^{d-1}(C_n)} \\
&\quad + \frac{1}{\mathcal{H}^{d-1}(C_n)} \left| G^{(p)}(\widetilde{u}_n, \rho(\bar{x}), T_{\widetilde{C}_n}) - G^{(p)}(u_n, \rho(x_n), T_{C_n}) \right|
\end{aligned}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

$$\leqslant \sigma^{(p)}(x_n, \nu) + \frac{1}{n} + \frac{1}{\mathcal{H}^{d-1}(C_n)} \left| G^{(p)}(\widetilde{u}_n, \rho(\bar{x}), T_{\widetilde{C}_n}) - G^{(p)}(u_n, \rho(x_n), T_{C_n}) \right|. \tag{3.9}$$

To estimate the last term, we reason as follows. First of all, we notice that

$$\int_{T_{C_n}} V(u_n(z))\,\mathrm{d}z = \int_{T_{\widetilde{C}_n}} V(\widetilde{u}_n(z))\,\mathrm{d}z. \tag{3.10}$$

Fix $\varepsilon > 0$. Using the continuity of ρ, there exists $\bar{n} \in \mathbb{N}$, such that $|\rho(x_n) - \rho(\bar{x})| < \varepsilon$ for all $n \geqslant \bar{n}$. From (3.10) we get that

$$\begin{aligned}
&\frac{1}{\mathcal{H}^{d-1}(C_n)} \left| G^{(p)}(\widetilde{u}_n, \rho(\bar{x}), T_{\widetilde{C}_n}) - G^{(p)}(u_n, \rho(x_n), T_{C_n}) \right| \\
&\quad \leqslant \frac{1}{\mathcal{H}^{d-1}(C_n)} |\rho(x_n) - \rho(\bar{x})| \int_{T_{C_n}} \int_{\mathbb{R}^d} \eta(h)\, |u_n(z+h) - u_n(z)|^p \,\mathrm{d}h\,\mathrm{d}z \\
&\quad \leqslant \frac{\varepsilon}{c_1} \left(\sigma^{(p)}(x_n, \nu) + \frac{1}{n} \right).
\end{aligned}$$

By the above and (3.9) we have

$$\left(1 - \frac{\varepsilon}{c_1 n}\right) \sigma^{(p)}(\bar{x}, \nu) \leqslant \left(1 + \frac{\varepsilon}{c_1}\right) \sigma^{(p)}(x_n, \nu) + \frac{1}{n}.$$

Using the arbitrariness of ε we conclude that

$$\sigma^{(p)}(\bar{x}, \nu) \leqslant \liminf_{n\to\infty} \sigma^{(p)}(x_n, \nu)$$

as required. □

Remark 3.9 *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 Γ-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 non-local functional. To be more precise, let $A, B \subset X$ be disjoint sets. Then, it holds that

$$\mathcal{F}^{(p)}_{\varepsilon_n}(u, A \cup B) = \mathcal{F}^{(p)}_{\varepsilon_n}(u, A) + \mathcal{F}^{(p)}_{\varepsilon_n}(u, B) + 2\widetilde{\Lambda}_{\varepsilon_n}(u, A, B), \tag{3.11}$$

where we define the non-local deficit

$$\widetilde{\Lambda}_{\varepsilon}(u, A, B) := \frac{s_\varepsilon}{\varepsilon} \int_A \int_B \eta_\varepsilon(x - z)|u(x) - u(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z. \tag{3.12}$$

On the other hand, for the limiting functional we have

$$\mathcal{G}^{(p)}_{\infty}(u, A \cup B) = \mathcal{G}^{(p)}_{\infty}(u, A) + \mathcal{G}^{(p)}_{\infty}(u, B), \tag{3.13}$$

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

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

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Identity (3.11) states that the functionals $\mathcal{F}_{\varepsilon_n}^{(p)}$ are non-local, while (3.13) is the locality property of the limiting functional $\mathcal{G}_\infty^{(p)}$. Thus, we expect the non-local deficit to disappear in the limit, i.e. that if $u_{\varepsilon_n} \to u$ in $L^1(X)$, then

$$\widetilde{\Lambda}_{\varepsilon_n}(u_{\varepsilon_n}, A, B) \to 0, \tag{3.14}$$

as $n \to \infty$. For technical reasons we also consider the non-local deficits without weighting by ρ or s_ε:

$$\Lambda_\varepsilon(u, A, B) := \frac{1}{\varepsilon} \int_A \int_B \eta_\varepsilon(x - z)|u(x) - u(z)|^p \, \mathrm{d}x \, \mathrm{d}z.$$

By continuity of ρ if A and B are sets in X that are close to $\bar{x}$ then $\widetilde{\Lambda}_\varepsilon(u, A, B) \approx s_\varepsilon \rho^2(\bar{x}) \Lambda_\varepsilon(u, A, B)$. In [1], the authors prove that the limit of the non-local deficit is determined by the behaviour of u_{ε_n} close to the boundaries of A and B and, in turn, that (3.14) 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 5 *Assume Assumptions (A1) and (C1)–(C3) hold. Let $v_n \to v$ in $L^1(X)$ with $|v_n| \leqslant 1$. Then, for all $\bar{x} \in \mathbb{R}^d$ and for all $\nu \in \mathbb{S}^{d-1}$, the following holds: given $C \in \mathcal{C}(\bar{x}, \nu)$ consider the strip T_C and any cube $Q \subset \mathbb{R}^d$ whose intersection with $\nu^\perp$ is C, for a.e. $t > 0$:*

(i) $\Lambda_{\varepsilon_n}(v_n, tT_C, \mathbb{R}^d \backslash tT_C) \to 0$ *as* $n \to \infty$,
(ii) $\Lambda_{\varepsilon_n}(v_n, tQ, tT_C \backslash tQ) \to 0$ *as* $n \to \infty$.

Remark 3.10 *The boundness 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^∞ 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^∞ bound.*

Finally, notice that when $A, B \in \mathbb{R}^d$ are disjoint sets with $\mathrm{d}(A, B) > 0$, using the fact that the function η has support in the ball $B(0, R_\eta)$ (see Assumption (C3)), it is easy to prove that there exists $\bar{n} \in \mathbb{N}$ such that for all $n \geqslant \bar{n}$ 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 3.11 For $\varepsilon > 0$, $p \geqslant 1$, $u : \mathbb{R}^d \to \mathbb{R}$, $\lambda \in \mathbb{R}$, and $A \subset \mathbb{R}^d$, we define

$$G_\varepsilon^{(p)}(u, \lambda, A) := \frac{\lambda}{\varepsilon} \int_A \int_{\mathbb{R}^d} \eta_\varepsilon(h)|u(z + h) - u(z)|^p \, \mathrm{d}h \, \mathrm{d}z + \frac{1}{\varepsilon} \int_A V(u(z)) \, \mathrm{d}z.$$

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

$$R_{x,r}u(y) := u(x + ry).$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Using a change of variables, it is easy to see that the following scaling property holds true:

$$G_\varepsilon^{(p)}(u,\lambda,x+rA)=r^{d-1}G_{\varepsilon/r}^{(p)}(R_{x,r}u,\lambda,A). \tag{3.15}$$

3.3 The liminf inequality

This section is devoted to proving the following: let $u_{\varepsilon_n}\to u$ in L^1, then

$$\mathcal{G}_\infty^{(p)}(u)\leqslant \liminf_{n\to\infty}\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}). \tag{3.16}$$

We will follow the proof of [1, Theorem 1.4], with some modifications due to the presence of the density ρ.

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

$$\liminf_{n\to\infty}\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n})=\lim_{n\to\infty}\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n})<\infty. \tag{3.17}$$

Step 1. By compactness (see Section 3.1) it holds $u=\chi_A$ for some set $A\subset X$ of finite perimeter in X. In order to prove (3.16) we use the strategy introduced by Fonseca and Müller in [26]. Write

$$\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n})=\int_X g_{\varepsilon_n}(x)\mathrm{d}x \tag{3.18}$$

and set $\mathrm{d}\lambda_{\varepsilon_n}:=g_{\varepsilon_n}\mathrm{d}\mathcal{L}^d\llcorner X$, so that

$$|\lambda_{\varepsilon_n}|(X)=\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}). \tag{3.19}$$

Using (3.17)–(3.19), up to a subsequence (not relabelled) it holds $\lambda_{\varepsilon_n}\overset{*}{\rightharpoonup}\lambda$ for some finite Radon measure λ on X. Then

$$|\lambda|(X)\leqslant\liminf_{n\to\infty}|\lambda_{\varepsilon_n}|(X). \tag{3.20}$$

In view of (3.19) and (3.20), the liminf inequality (3.16) is implied by the following claim: for $\mathcal{H}^{d-1}$-a.e. $\bar{x}\in\partial^*\{u=1\}$ it holds

$$\sigma^{(p)}(\bar{x},\nu(\bar{x}))\rho(\bar{x})\leqslant\frac{\mathrm{d}\lambda}{\mathrm{d}\theta}(\bar{x}),$$

where $\theta:=\mathcal{H}^{d-1}\llcorner\partial^*\{u=1\}$. 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 λ with respect to θ via (recall Remark 2.15)

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\theta}(\bar{x})=\lim_{r\to 0}\frac{\lambda(\bar{x}+rQ)}{r^{d-1}}, \tag{3.21}$$

where Q is a unit cube centred 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 2.14 implies that

$$R_{\bar{x},r}u\to v_{\bar{x}} \tag{3.22}$$

in $L^1_{\mathrm{loc}}(\mathbb{R}^N)$ as $r\to 0$, where

$$v_{\bar{x}}(x):=\begin{cases}-1 & x\cdot\nu(\bar{x})\geqslant 0,\\ 1 & x\cdot\nu(\bar{x})<0.\end{cases}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Let $\bar{x} \in \partial^*\{u = 1\}$ be a point for which (3.21) and (3.22) hold. Without loss of generality, we can assume that

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\theta}(\bar{x}) < \infty. \tag{3.23}$$

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

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\theta}(\bar{x}) = \lim_{n\to\infty} \frac{\lambda_{\varepsilon_n}(\bar{x} + r_n Q)}{r_n^{d-1}} \tag{3.24}$$

and

$$R_{\bar{x},r_n} u_{\varepsilon_n} \to v_{\bar{x}}.$$

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

$$\frac{\lambda_{\varepsilon_n}(\bar{x} + r_n Q)}{r_n^{d-1}} \geqslant \frac{\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \bar{x} + r_n Q)}{r_n^{d-1}}. \tag{3.25}$$

Step 2. We claim that

$$\delta_n := \frac{|\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \bar{x} + r_n Q) - \rho(\bar{x})\widetilde{\mathcal{F}}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \rho(\bar{x}), \bar{x} + r_n Q)|}{r_n^{d-1}} \to 0, \tag{3.26}$$

as $n \to \infty$, where

$$\widetilde{\mathcal{F}}_{\varepsilon}^{(p)}(u, \xi, A) := \frac{\xi s_\varepsilon}{\varepsilon} \int_A \int_A \eta_\varepsilon(x - z)|u(x) - u(z)|^p \,\mathrm{d}x\,\mathrm{d}z + \frac{1}{\varepsilon}\int_A V(u(x))\,\mathrm{d}x,$$

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 ρ is continuous in X. Then, it is possible to find $\bar{n} \in \mathbb{N}$ such that for all $n \geqslant \bar{n}$ and all $y \in \bar{x} + r_n Q$ it holds

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

Thus,

$$\begin{aligned}
\delta_n &\leqslant \frac{t}{r_n^{d-1}\varepsilon_n} \int_{\bar{x}+r_nQ} V(u_{\varepsilon_n}(x))\,\mathrm{d}x \\
&\quad + \frac{ts_{\varepsilon_n}\rho(\bar{x})}{r_n^{d-1}\varepsilon_n} \int_{\bar{x}+r_nQ}\int_{\bar{x}+r_nQ} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \,\mathrm{d}x\,\mathrm{d}z \\
&\quad + \frac{ts_{\varepsilon_n}}{r_n^{d-1}\varepsilon_n} \int_{\bar{x}+r_nQ}\int_{\bar{x}+r_nQ} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\leqslant \frac{t}{r_n^{d-1}\varepsilon_n c_1} \int_{\bar{x}+r_nQ} V(u_{\varepsilon_n}(x))\rho(x)\,\mathrm{d}x \\
&\quad + \frac{ts_{\varepsilon_n}(c_1 + c_2)}{r_n^{d-1}\varepsilon_n c_1^2} \int_{\bar{x}+r_nQ}\int_{\bar{x}+r_nQ} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(z)\rho(x)\,\mathrm{d}x\,\mathrm{d}z \\
&\leqslant \frac{t(c_1 + c_2)}{c_1^2} \frac{\lambda_{\varepsilon_n}(\bar{x} + r_n Q)}{r_n^{d-1}},
\end{aligned}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

where in the last step we used (3.25). By (3.23) and (3.24) $\lim_{n\to\infty}\delta_n \leqslant Ct$ for some constant $C<\infty$. Since $t>0$ is arbitrary, this proves the claim.

Step 3. Observe that for any $\lambda \geqslant 0$, $\varepsilon>0$, $r>0$ and $v\in L^1$ we have

$$\min\left\{1, \frac{s_\varepsilon}{s_{\frac{\varepsilon}{r}}}\right\}\tilde{\mathcal{F}}^{(p)}_{\frac{\varepsilon}{r}}(R_{\bar{x},r}v,\lambda,Q) \leqslant \frac{1}{r^{d-1}}\tilde{\mathcal{F}}^{(p)}_{\varepsilon}(v,\lambda,\bar{x}+rQ) \leqslant \max\left\{1, \frac{s_\varepsilon}{s_{\frac{\varepsilon}{r}}}\right\}\tilde{\mathcal{F}}^{(p)}_{\frac{\varepsilon}{r}}(R_{\bar{x},r}v,\lambda,Q).$$

Let $C = Q\cap \nu(\bar{x})^\perp \in \mathcal{C}(\bar{x},\nu(\bar{x}))$. Define the function $w_n:\mathbb{R}^d\to\mathbb{R}$ as the periodic extension of the function that is $R_{\bar{x},r_n}u_{\varepsilon_n}$ in Q and $v_{\bar{x}}$ in $T_C\backslash Q$. Set $\varepsilon_n' := \frac{\varepsilon_n}{r_n}$ and $s_n' = \min\left\{1, \frac{s_{\varepsilon_n}}{s_{\varepsilon_n'}}\right\}$. Using (3.25) and (3.26) together with the scaling identity (3.15) we get

$$\begin{aligned}
\frac{\lambda_{\varepsilon_n}(\bar{x}+r_nQ)}{r_n^{d-1}} &\geqslant \rho(\bar{x})\frac{\widetilde{\mathcal{F}}^{(p)}_{\varepsilon_n}(u_{\varepsilon_n},\rho(\bar{x}),\bar{x}+r_nQ)}{r_n^{d-1}} - \delta_n\\
&\geqslant s_n'\rho(\bar{x})\widetilde{\mathcal{F}}^{(p)}_{\varepsilon_n'}(R_{\bar{x},r_n}u_{\varepsilon_n},\rho(\bar{x}),Q) - \delta_n\\
&= s_n'\rho(\bar{x})\widetilde{\mathcal{F}}^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),Q) - \delta_n\\
&\geqslant s_n'\rho(\bar{x})G^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),T_C) - \delta_n\\
&\quad - s_n'\rho(\bar{x})\left|\widetilde{\mathcal{F}}^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),Q) - G^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),T_C)\right|\\
&= s_n'\rho(\bar{x})\left(\frac{\varepsilon_n}{r_n}\right)^{d-1}G^{(p)}\left(R_{0,\varepsilon_n'}w_n,\rho(\bar{x}),\frac{r_n}{\varepsilon_n}T_C\right) - \delta_n\\
&\quad - s_n'\rho(\bar{x})\left|\widetilde{\mathcal{F}}^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),Q) - G^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),T_C)\right|. \qquad (3.27)
\end{aligned}$$

We would like to say that

$$\left|\widetilde{\mathcal{F}}^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),Q) - G^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),T_C)\right| \to 0$$

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

$$\begin{aligned}
\left|\widetilde{\mathcal{F}}^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),tQ) - G^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),tT_C)\right| &\leqslant 2\rho(\bar{x})\Lambda_{\varepsilon_n'}(w_n,tQ,tT_C\backslash tQ)\\
&\quad + \rho(\bar{x})\Lambda_{\varepsilon_n'}(w_n,tT_C,\mathbb{R}^d\backslash tT_C) + \rho(\bar{x})\Lambda_{\varepsilon_n'}(w_n,tT_C\backslash tQ,tT_C\backslash tQ)\\
&\quad + \frac{r_n\rho(\bar{x})}{\varepsilon_n}|1-s_{\varepsilon_n'}|\int_{tQ}\int_{tQ}\eta_{\frac{\varepsilon_n}{r_n}}(y-z)|w_n(y)-w_n(z)|^p\,\mathrm{d}y\,\mathrm{d}z. \qquad (3.28)
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{r_n\rho(\bar{x})}{\varepsilon_n}\int_{tQ}\int_{tQ}\eta_{\varepsilon_n'}(y-z)|w_n(y)-w_n(z)|^p\,\mathrm{d}y\,\mathrm{d}z &\leqslant G^{(p)}_{\varepsilon_n'}(w_n,\rho(\bar{x}),T_C)\\
&= \left(\frac{\varepsilon_n}{r_n}\right)^{d-1}G^{(p)}\left(R_{0,\varepsilon_n'}w_n,\rho(\bar{x}),\frac{r_n}{\varepsilon_n}T_C\right). \qquad (3.29)
\end{aligned}$$

Moreover, using Proposition 5 we get that for a.e. $t\in(0,1)$ it holds

$$\Lambda_{\varepsilon_n'}(w_n,tQ,tT_C\backslash tQ)\to 0, \qquad \Lambda_{\varepsilon_n'}(w_n,tT_C,\mathbb{R}^d\backslash tT_C)\to 0 \qquad (3.30)$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

as $n \to \infty$. Finally, using Remark 3.10 and the fact that w_n is constant on $T_C \backslash Q$ it is easy to see that

$$\lim_{t\to 1} \lim_{n\to\infty} \Lambda_{\varepsilon_n'}(w_n, tT_C\backslash tQ, tT_C\backslash tQ) = 0. \tag{3.31}$$

Hence, from (3.26)–(3.31) and recalling that $s_n' \to 1$ we get

$$\lim_{n\to\infty} \frac{\lambda_{\varepsilon_n}(\bar{x} + r_n Q)}{r_n^{d-1}} \geqslant \rho(\bar{x})\sigma^{(p)}(\bar{x}, \nu(\bar{x}))$$

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

$$\limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}) \leqslant \mathcal{G}_{\infty}^{(p)}(u). \tag{3.32}$$

Without loss of generality, we can assume $\mathcal{G}_{\infty}^{(p)}(u) < \infty$, 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 2.16). Then we claim that there exists a sequence $\{u_{\varepsilon_n}\}_{n=1}^{\infty}$ with $|u_{\varepsilon_n}| \leqslant 1$, converging uniformly to u on every compact set $K \subset X \backslash \partial^*\{u = 1\}$, and in particular $u_{\varepsilon_n} \to u$ in $L^1(X, \mu)$, such that (3.32) holds.

Let us denote by E the polyhedral set $\{u = 1\}$ and by $E_1, \dots, E_k$ its faces. It is possible to cover $\partial E \cap X$ with a finite family of sets $\bar{A}_1, \dots, \bar{A}_k$, where each A_i is an open set satisfying the following properties (see Figure 3):

(i) ∂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) $\mathcal{H}^{d-1}\left(\bar{A}_i \cap \bigcup_{j\neq i} E_j\right) = 0$ and
(iv) $A_i \cap A_j = \emptyset$ if $i \neq j$.

Set

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

We then define u_{ε_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 $i \in \{1, \dots, k\}$ and $n \in \mathbb{N}$. In order to define u_{ε_n} in A_i, we reason as follows. Denote by ν 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

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

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

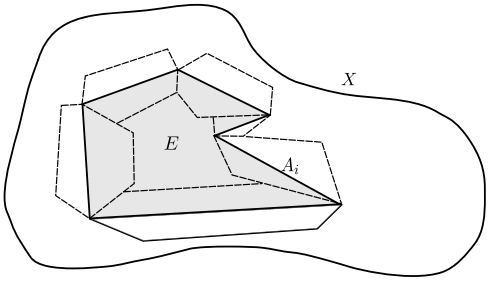


FIGURE 3. The polyhedral set E (shaded) and the sets A_i (dotted lines).

Fix $\xi > 0$. Using the continuity of σ (see Lemma 3) and of ρ 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 $E_i \subseteq \cup_{j=1}^{M_\xi} Q_j$, $E_i \cap Q_j \neq \emptyset$ and satisfying the following properties: denoting by $\{x_j\}_{j=1}^{M_\xi}$ their centres (or, in the case the centre of a cube Q_j is not contained in E_i, a point of $Q_j \cap E_i$) we have

$$\left| \int_{E_i} \sigma^{(p)}(x,\nu)\rho(x)\,\mathrm{d}\mathcal{H}^{d-1}(x) - r_\xi^{d-1} \sum_{j=1}^{M_\xi} \sigma^{(p)}(x_j,\nu)\rho(x_j) \right| < \xi. \tag{3.33}$$

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

$$\frac{1}{\mathcal{H}^{d-1}(C_j)} G^{(p)}(w_j, \rho(x_j), T_{C_j}) < \sigma^{(p)}(x_j,\nu) + \frac{\xi}{\rho(x_j) M_\xi r_\xi^{d-1}}. \tag{3.34}$$

We can assume $|w_j| \leqslant 1$. For every $j = 1, \dots, M_\xi$, let $L_j \in \mathbb{N}$ be such that

$$\frac{1}{\mathcal{H}^{d-1}(C_j)} \left| G^{(p)}(w_j, \rho(x_j), T_{C_j}) - G^{(p)}(\widetilde{w}_j, \rho(x_j), T_{C_j}) \right| < \frac{\xi}{\rho(x_j) M_\xi r_\xi^{d-1}},$$

where

$$\widetilde{w}_j(x) := \begin{cases} w_j(x) & \text{if } |x_d| < L_j, \\ +1 & \text{if } x_d > L_j, \\ -1 & \text{if } x_d < -L_j. \end{cases}$$

By mollifying, $\widetilde{w}_j$, i.e. $\widetilde{w}_j' = J_\beta * \widetilde{w}_j$, we have that $\widetilde{w}_j'$ is Lipschitz continuous with $\mathrm{Lip}(\widetilde{w}_j') \leqslant \frac{1}{\beta}$ and $\widetilde{w}_j'(x) \in \{\pm 1\}$ for $|x_d| > L_j + \beta$. We can choose $\beta = \beta_\xi$ such that

$$\frac{1}{\mathcal{H}^{d-1}(C_j)} \left| G^{(p)}(w_j, \rho(x_j), T_{C_j}) - G^{(p)}(\widetilde{w}_j', \rho(x_j), T_{C_j}) \right| < \frac{2\xi}{\rho(x_j) M_\xi r_\xi^{d-1}}. \tag{3.35}$$

For every $j = 1, \dots, M_\xi$ 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 centres. Notice that it might be necessary to consider the intersection of some cubes with Q_j, and that

$$k_j^{(n)} = \left\lceil \frac{1}{\varepsilon_n^{d-1}} r_\xi^{d-1} \left(\mathcal{H}^{d-1}(C_j)\right)^{-1} \right\rceil, \tag{3.36}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

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

$$v_j^{(n)}(x) := \sum_{s=1}^{k_j^{(n)}} \widetilde{w}_j' \left(\frac{x - (x_j + y_j^{(n)})}{\varepsilon_n} \right) \chi_{Q_{j,s}^{(n)}}(x').$$

We are now in position to define the function u_{ε_n} in A_i: for $x \in A_i$ define

$$u_{\varepsilon_n}(x) := \sum_{j=1}^{M_\xi} \chi_{Q_j^{(n)}}(x') v_j^{(n)}(x).$$

Note that u_{ε_n} has Lipschitz constant $\mathrm{Lip}(u_{\varepsilon_n}) \leqslant \frac{1}{\beta_\xi \varepsilon_n}$ and $u_{\varepsilon_n}(x) = u(x)$ for any x with $\mathrm{dist}(x, \partial^*\{u = 1\}) \geqslant \max_j (L_j + \beta_\xi)\varepsilon_n$. 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 (3.32). We claim that:

(i) $\widetilde{\Lambda}_{\varepsilon_n}(u_{\varepsilon_n}, A_i, A_j) \to 0$ as $n \to \infty$ for all $i \neq j$, where $\widetilde{\Lambda}$ is defined by (3.12), and
(ii) $\limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, A_i) \leqslant \mathcal{G}_\infty^{(p)}(u, A_i)$ for all $i = 0, \ldots, k+1$.

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

$$\begin{aligned}
\limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}) &\leqslant \sum_{i=0}^{k+1} \limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, A_i) + 2 \sum_{i<j=0}^{k+1} \limsup_{n\to\infty} \widetilde{\Lambda}_{\varepsilon_n}(u_{\varepsilon_n}, A_i, A_j) \\
&\leqslant \sum_{i=0}^{k+1} \mathcal{G}_\infty^{(p)}(u, A_i) \\
&= \mathcal{G}_\infty^{(p)}(u).
\end{aligned}$$

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

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

we get

$$\begin{aligned}
\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, A_i) &\leqslant \mathcal{F}_{\varepsilon_n}^{(p)} \left(u_{\varepsilon_n}, \bigcup_{j=1}^{M_\xi} \bigcup_{s=1}^{k_j^{(n)}} T_{Q_{j,s}^{(n)}} \right) \\
&\leqslant \sum_{j=1}^{M_\xi} \sum_{s=1}^{k_j^{(n)}} \left[\frac{s_{\varepsilon_n}}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} \int_{\mathbb{R}^d} \eta_{\varepsilon_n}(x - z) |u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(x) \rho(z) \, \mathrm{d}x \, \mathrm{d}z \right. \\
&\qquad \left. + \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} V(u_{\varepsilon_n}(x)) \rho(x) \, \mathrm{d}x \right]. \qquad (3.37)
\end{aligned}$$

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

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Note that,

$$\mathcal{A}_{j,s}^{(n)} := \left\{(x,z) \in \mathbb{R}^d \times T_{Q_{j,s}^{(n)}} \,:\, \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \neq 0\right\} \subseteq B(x_j, \sqrt{2d}r_\xi) \times \left(T_{Q_{j,s}^{(n)}} \cap \left\{|z_d| \leqslant \varepsilon_n(L_j + R_\eta)\right\}\right)$$

for ε_n sufficiently small (compared to r_ξ). Hence there exists $\delta_\xi > 0$ such that for all $(x,z) \in \mathcal{A}_{j,s}^{(n)}$ we have $|\rho(x) - \rho(x_j)| \leqslant \delta_\xi$ and $|\rho(z) - \rho(x_j)| \leqslant \delta_\xi$, where $\delta_\xi \to 0$ as $\xi \to 0^+$. This implies

$$\begin{aligned} s_{\varepsilon_n}\rho(x)\rho(z) &\leqslant |s_{\varepsilon_n} - 1|\rho(x)\rho(z) + \rho(x)\rho(z) \\ &\leqslant |s_{\varepsilon_n} - 1|c_2^2 + \rho(x_j)\rho(z) + \delta_\xi c_2 \\ &\leqslant |s_{\varepsilon_n} - 1|c_2^2 + \rho^2(x_j) + 2\delta_\xi c_2. \end{aligned}$$

Hence,

$$\begin{aligned} &\frac{s_{\varepsilon_n}}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} \int_{\mathbb{R}^d} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\ &\quad \leqslant \left(|s_{\varepsilon_n} - 1|c_2^2 + 2\delta_\xi c_2 + \rho^2(x_j)\right) \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} \int_{\mathbb{R}^d} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \,\mathrm{d}x\,\mathrm{d}z. \end{aligned}$$

Similarly,

$$\frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} V(u_{\varepsilon_n}(x))\rho(x)\,\mathrm{d}x \leqslant \left(\rho(x_j) + \delta_\xi\right) \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} V(u_{\varepsilon_n}(x))\,\mathrm{d}x.$$

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

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

(there is a subtle dependence that requires ε_n be small compared to r_ξ which means it is necessary to first take the limits in this order). Then

$$\begin{aligned} &\frac{s_{\varepsilon_n}}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} \int_{\mathbb{R}^d} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z + \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} V(u_{\varepsilon_n}(x))\rho(x)\,\mathrm{d}x \\ &\quad \leqslant \gamma_{\xi,n}\rho(x_j) \left(\frac{\rho(x_j)}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} \int_{\mathbb{R}^d} \eta_{\varepsilon_n}(x-z)|u_{\varepsilon_n}(x) - u_{\varepsilon_n}(z)|^p \,\mathrm{d}x\,\mathrm{d}z \right. \\ &\qquad \left. + \frac{1}{\varepsilon_n} \int_{T_{Q_{j,s}^{(n)}}} V(u_{\varepsilon_n}(x))\,\mathrm{d}x \right) \\ &\quad \leqslant \gamma_{\xi,n}\rho(x_j) G_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \rho(x_j), T_{Q_{j,s}^{(n)}}). \end{aligned}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Hence, using (3.33)–(3.37), we get

$$\begin{aligned}
\mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, A_i) &\leqslant \gamma_{\xi,n} \sum_{j=1}^{M_\xi} \sum_{s=1}^{k_j^{(n)}} \rho(x_j) G_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}, \rho(x_j), T_{Q_{j,s}^{(n)}}) \\
&= \gamma_{\xi,n} \sum_{j=1}^{M_\xi} \sum_{s=1}^{k_j^{(n)}} \rho(x_j) \varepsilon_n^{d-1} G_1^{(p)}(\widetilde{w}_j', \rho(x_j), T_{C_j}) \\
&= \gamma_{\xi,n} \sum_{j=1}^{M_\xi} \frac{r_\xi^{d-1} \rho(x_j)}{\mathcal{H}^{d-1}(C_j)} G_1^{(p)}(\widetilde{w}_j', \rho(x_j), T_{C_j}) \\
&\leqslant \gamma_{\xi,n} \left(r_\xi^{d-1} \sum_{j=1}^{M_\xi} \rho(x_j) \sigma^{(p)}(x_j, \nu) + 3\xi \right) \\
&\leqslant \gamma_{\xi,n} \left(\int_{E_i} \sigma^{(p)}(x, \nu) \rho(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) + 4\xi \right).
\end{aligned}$$

Taking limsup as $n \to \infty$ and $\xi \to 0$ implies (ii). Note that u_{ε_n} is not the recovery sequence, rather each u_{ε_n} depended on ξ (through L_j and β_ξ). Hence, if we make the ξ dependence explicit, we showed $\limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}^{(\xi)}, A_i) \leqslant \limsup_{n\to\infty} \gamma_{\xi,n} \mathcal{G}_\infty^{(p)}(u, A_i)$. Applying a diagonalisation argument we infer the existence of a sequence $\xi_n \to 0$ satisfying $\limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}^{(p)}(u_{\varepsilon_n}^{(\xi_n)}, A_i) \leqslant \mathcal{G}_\infty^{(p)}(u, A_i)$. Moreover if the sequence $\xi_n \to 0$ sufficiently slowly, then we can infer $\max L_j + \beta_{\xi_n} \geqslant \frac{1}{\zeta_{\varepsilon_n}}$ and $\beta_{\xi_n} \leqslant \zeta_{\varepsilon_n}$; hence, $\mathrm{Lip}(u_{\varepsilon_n}) \leqslant \frac{1}{\zeta_{\varepsilon_n} \varepsilon_n}$ and $u_{\varepsilon_n}(x) = u(x)$ whenever $\mathrm{dist}(x, \partial^*\{u = 1\}) \geqslant \frac{\varepsilon_n}{\zeta_{\varepsilon_n}}$ for any sequence $\zeta_{\varepsilon_n} \to 0$.

We are thus left to prove claim (i). Analogously to Remark 3.10, $\widetilde{\Lambda}_{\varepsilon_n}(u_{\varepsilon_n}, A_i, A_j) \to 0$ as $n \to \infty$ for all i, j for which $\mathrm{d}(A_i, A_j) > 0$. Let us now consider indexes $i \neq j$ for which $\mathrm{d}(A_i, A_j) = 0$. Write

$$\widetilde{\Lambda}_{\varepsilon_n}(u_{\varepsilon_n}, A_i, A_j) = \frac{1}{\varepsilon_n} \int_{B(0,R_\eta)} \int_{A_{\varepsilon_n h}} \eta(h) |u_{\varepsilon_n}(z + \varepsilon_n h) - u_{\varepsilon_n}(z)|^p \rho(z + \varepsilon_n h) \rho(z) \, \mathrm{d}z \, \mathrm{d}h,$$

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 $\mathrm{Vol}(A_{\varepsilon_n h}) = O(\varepsilon_n^d)$. Thus

$$\widetilde{\Lambda}_{\varepsilon_n}(u_{\varepsilon_n}, A_i, A_j) \to 0,$$

as $n \to \infty$.

Step 2. The general case. Let $u \in BV(X; \{\pm 1\})$. Using Theorem 2.17 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 $Dv_n \overset{w^*}{\rightharpoonup} Du$) 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

$$\mathcal{F}_{\varepsilon_n}^{(p)}(u_n) \leqslant \mathcal{G}_\infty^{(p)}(v_n) + \frac{1}{n}.$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Then, Theorem 2.18 together with Lemma 3 gives us that

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

This concludes the proof. □

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 and 4.3 we prove the Γ-convergence result.

4.1 Compactness

Proof of Theorem 1.6 (Compactness). *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

$$\eta\left(\frac{T_n(x) - T_n(z)}{\varepsilon_n}\right) \geqslant c_n \eta\left(\frac{\alpha_n(x-z)}{\varepsilon_n}\right). \tag{4.1}$$

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

$$\eta(a) \geqslant c_n \eta(\alpha_n b). \tag{4.2}$$

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

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

and therefore, by (4.2), we get $\eta\left(\frac{T_n(x)-T_n(z)}{\varepsilon_n}\right) \geqslant c_n \eta\left(\frac{\alpha_n(x-z)}{\varepsilon_n}\right)$ as required.

Step 2. Let $v_n := u_n \circ T_n$. Using Lemma 1 and (4.1) we have that

$$\begin{aligned}
\mathcal{G}_n^{(p)}(u_n) &= \frac{1}{\varepsilon_n^{d+1}} \int_X \int_X \eta\left(\frac{T_n(x) - T_n(z)}{\varepsilon_n}\right) |v_n(x) - v_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\quad + \frac{1}{\varepsilon_n} \int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&\geqslant \frac{c_n}{\varepsilon_n^{d+1}} \int_X \int_X \eta\left(\frac{\alpha_n(x-z)}{\varepsilon_n}\right) |v_n(x) - v_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\quad + \frac{1}{\varepsilon_n} \int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&= \frac{c_n}{\alpha_n^{d+1}(\varepsilon_n')^{d+1}} \int_X \int_X \eta\left(\frac{x-z}{\varepsilon_n'}\right) |v_n(x) - v_n(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\quad + \frac{1}{\varepsilon_n} \int_X V(v_n(x))\rho(x)\,\mathrm{d}x \\
&= \frac{1}{\alpha_n} \mathcal{F}_{\varepsilon_n'}^{(p)}(v_n),
\end{aligned} \tag{4.3}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

where $\varepsilon_n' := \frac{\varepsilon_n}{\alpha_n}$ and $\mathcal{F}_\varepsilon^{(p)}$ is defined by (3.1) with $s_n := \frac{c_n}{\alpha_n^d}$ (in (3.1) s depended on ε not on n, since we have fixed the sequence ε_n then clearly we could write n in terms of ε_n'; however, this would make the notation cumbersome).

Step 3. Since $\lim_{n\to\infty} \alpha_n = 1$, we can infer that $\mathcal{F}_{\varepsilon_n'}^{(p)}(v_n)$ is bounded and hence by Theorem 3.2 the sequence $\{v_n\}_{n=1}^\infty$ is relatively compact in L^1. Therefore, the sequence $\{u_n\}_{n=1}^\infty$ 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

$$\liminf_{n\to\infty} \mathcal{G}_n^{(p)}(u_n) \geqslant \mathcal{G}_\infty^{(p)}(u).$$

Indeed, taking the liminf on both sides of (4.3) and using Theorem 3.2 we have

$$\liminf_{n\to\infty} \mathcal{G}_n^{(p)}(u_n) \geqslant \liminf_{n\to\infty} \frac{1}{\alpha_n} \mathcal{F}_{\varepsilon_n'}^{(p)}(v_n) \geqslant \mathcal{G}_\infty^{(p)}(u),$$

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

$$\limsup_{n\to\infty} \mathcal{G}_n^{(p)}(u_n) \leqslant \mathcal{G}_\infty^{(p)}(u).$$

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

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 $\mathcal{G}_\infty^{(p)}(u) < +\infty$ via a diagonalisation argument.

Case 1. Assume that u is a polyhedral function (see Definition 2.16). Let $w_n \in L^1(X, \mu)$ be a recovery sequence for the Γ-convergence of $\mathcal{F}_{\varepsilon_n'}^{(p)}$ to $\mathcal{G}_\infty^{(p)}$ where $\varepsilon_n' = \frac{\varepsilon_n}{\hat{\alpha}_n}$ and $\hat{\alpha}_n$ is a sequence we fix shortly (see step 2), i.e. we assume $w_n \to u$ in $L^1(X, \mu)$ and $\lim_{n\to\infty} \mathcal{F}_{\varepsilon_n'}^{(p)}(w_n) = \mathcal{G}_\infty^{(p)}(u)$. Let $u_n(x_i) = n \int_{T_n^{-1}(x_i)} w_n(x)\, \mathrm{d}x$ where T_n is any sequence of transport maps satisfying the conclusions of Theorem 2.8.

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 $\zeta_n \gg \sqrt{\frac{\|T_n - \mathrm{Id}\|_{L^\infty}}{\varepsilon_n}}$. By Theorem 3.2 we can assume that w_n are Lipschitz continuous with $\mathrm{Lip}(w_n) \leqslant \frac{1}{\zeta_n \varepsilon_n}$ and $w_n(x) = u(x)$ for all x satisfying $\mathrm{dist}(x, \partial^*\{u = 1\}) > \frac{\varepsilon_n}{\zeta_n}$. Hence, $v_n(x) = w_n(x) = u(x)$, for n sufficiently large, for all x such that $\mathrm{dist}(x, \partial^*\{u = 1\}) > \frac{3\varepsilon_n}{\zeta_n}$.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

Therefore,

$$\begin{aligned}\|v_n - w_n\|_{L^1(X)} &= \sum_{i=1}^n \int_{T_n^{-1}(x_i)} |v_n(x) - w_n(x)|\,\mathrm{d}x \\ &\leqslant n \sum_{i=1}^n \int_{T_n^{-1}(x_i)} \int_{T_n^{-1}(x_i)} |w_n(y) - w_n(x)|\,\mathrm{d}y\,\mathrm{d}x \\ &\leqslant \frac{2\|T_n - \mathrm{Id}\|_{L^\infty}}{\zeta_n \varepsilon_n n} \#\left\{ i \,:\, \mathrm{dist}(x_i, \partial^*\{u=1\}) \leqslant \frac{3\varepsilon_n}{\zeta_n} \right\}.\end{aligned}$$

Now,

$$\begin{aligned}\#\left\{ i \,:\, \mathrm{dist}(x_i, \partial^*\{u=1\}) \leqslant \frac{3\varepsilon_n}{\zeta_n} \right\} &= n\mu\left(\left\{ x \,:\, \mathrm{dist}(T_n(x), \partial^*\{u=1\}) \leqslant \frac{3\varepsilon_n}{\zeta_n} \right\} \right) \\ &\leqslant n c_2 \mathrm{Vol}\left(\left\{ x \,:\, \mathrm{dist}(x, \partial^*\{u=1\}) \leqslant \frac{4\varepsilon_n}{\zeta_n} \right\} \right) \\ &= O\left(\frac{n\varepsilon_n}{\zeta_n} \right).\end{aligned}$$

Hence,

$$\|v_n - w_n\|_{L^1(X)} = o(\varepsilon_n). \tag{4.4}$$

Moreover,

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

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

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

$$\eta\left(\frac{T_n(x) - T_n(z)}{\varepsilon_n} \right) \leqslant \hat{c}_n \eta\left(\frac{\hat{\alpha}_n(x-z)}{\varepsilon_n} \right). \tag{4.5}$$

To show the validity of (4.5) we use the following subclaim: for all $\hat{\delta} > 0$ sufficiently small there exists $\hat{\alpha}_{\hat{\delta}}, \hat{c}_{\hat{\delta}} > 0$ such that $\hat{\alpha}_{\hat{\delta}} \to 1, \hat{c}_{\hat{\delta}} \to 1$, as $\hat{\delta} \to 0$, and, for any $\hat{a}, \hat{b} \in \mathbb{R}^d$, it holds

$$|\hat{a} - \hat{b}| < \hat{\delta} \quad \Rightarrow \quad \hat{c}_{\hat{\delta}} \eta(\hat{\alpha}_{\hat{\delta}} \hat{b}) \geqslant \eta(\hat{a}). \tag{4.6}$$

Then (4.5) can be obtained as follows: for any $n \in \mathbb{N}$ take

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

and let $\hat{\alpha}_n, \hat{c}_n$ be the numbers given by the subclaim for which (4.6) 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 $|\hat{a} - \hat{b}| \leqslant \delta_n$, we infer (4.5).

To prove the subclaim, we let α_δ, c_δ be as in Assumption (C4) and let $\hat{\delta} > 0$. Without loss of generality we assume that $\inf_{\gamma \in (0,1]} \alpha_\gamma \in (0, \infty)$. We choose $\delta := \min\left\{ 1, \frac{\hat{\delta}}{\inf_{s \in (0,1]} \alpha_s} \right\}$, trivially

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

$\delta \to 0$ as $\hat{\delta} \to 0$. We assume that $\frac{\hat{\delta}}{\inf_{\gamma\in(0,1]} \alpha_\gamma} \leqslant 1$. 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, $|a - b| \leqslant \frac{\hat{\delta}}{\alpha_\delta} \leqslant \frac{\hat{\delta}}{\inf_{\gamma\in(0,1]} \alpha_\gamma} = \delta$,

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

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

Step 3. Using Lemma 1 and (4.5) we get

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

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

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

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

$$|a|^p \leqslant (1 + \delta)|b|^p + C_\delta |a - b|^p. \tag{4.7}$$

Inequality (4.7) follows by noticing that, for every given $\delta > 0$, the function $x \mapsto \frac{|x|^p - 1 - \delta}{|x-1|^p}$ is bounded in $[0, \infty)$, and then setting $x = \frac{|a|}{|b|}$. Moreover, for all $p \geqslant 1$ and all $a, b \in \mathbb{R}$, it holds

$$|a + b|^p \leqslant 2^{p-1} \left(|a|^p + |b|^p\right). \tag{4.8}$$

Fix $\delta > 0$. Using (4.7) and (4.8) we infer

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

where

$$d_n := \frac{2^{p-1}\hat{c}_n C_\delta}{\hat{\alpha}_n^{d+1}\varepsilon_n'} \int_X \int_X \eta_{\varepsilon_n'}(x - z) \left(|v_n(x) - w_n(x)|^p + |v_n(z) - w_n(z)|^p\right) \rho(x)\rho(z)\, \mathrm{d}x\, \mathrm{d}z.$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

We show $d_n \to 0$. We have that

$$\begin{aligned} d_n &\leqslant \frac{2^p \hat{c}_n C_\delta c_2^2}{\hat{\alpha}_n^{d+1} \varepsilon_n'} \int_X |v_n(x) - w_n(x)|^p \, \mathrm{d}x \int_{\mathbb{R}^d} \eta(x) \, \mathrm{d}x \\ &\leqslant \frac{2^{2p-1} \hat{c}_n C_\delta c_2^2}{\hat{\alpha}_n^{d+1} \varepsilon_n'} \|v_n - w_n\|_{L^1(X)} \int_{\mathbb{R}^d} \eta(x) \, \mathrm{d}x \\ &= O\left(\frac{C_\delta \|v_n - w_n\|_{L^1(X)}}{\varepsilon_n} \right), \end{aligned}$$

where we used the fact that $\hat{c}_n \to 1$, $\hat{\alpha}_n \to 1$ as $n \to \infty$. From Equation (4.4) we get that

$$d_n \to 0, \tag{4.9}$$

as $n \to \infty$.

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

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

then $b_n \to 0$ by (4.4).

Combining the above estimates we have,

$$\begin{aligned} \mathcal{G}_n^{(p)}(u_n) \leqslant \frac{\hat{c}_n(1+\delta)}{\hat{\alpha}_n^{d+1} \varepsilon_n'} \int_X \int_X \eta_{\varepsilon_n'}(x-z) |w_n(x) - w_n(z)|^p \rho(x) \rho(z) \, \mathrm{d}x \, \mathrm{d}z \\ + \frac{1}{\varepsilon_n} \int_X V(w_n(x)) \rho(x) \, \mathrm{d}x + b_n + d_n. \end{aligned} \tag{4.10}$$

Step 4. We now conclude as follows. Using (4.10) we get

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

for $\mathcal{F}_{\varepsilon_n'}^{(p)}$ defined as in (3.1) with $s_n := \frac{\hat{c}_n}{\hat{\alpha}_n^d}$. Taking the limsup on both sides and using (4.9), together with Theorem 3.2 we have

$$\limsup_{n \to \infty} \mathcal{G}_n^{(p)}(u_n) \leqslant \limsup_{n \to \infty} (1+\delta) \mathcal{F}_{\varepsilon_n'}(w_n) \leqslant (1+\delta) \mathcal{G}_\infty^{(p)}(u).$$

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 2.17 together with Lemma 3 and Theorem 2.18 we can find a sequence of polyhedral functions $\{u^{(m)}\}_{m=1}^\infty$ such that

$$\|u - u^{(m)}\|_{L^1} \leqslant \frac{1}{m}, \qquad \mathcal{G}_\infty^{(p)}(u^{(m)}) \leqslant \mathcal{G}_\infty^{(p)}(u) + \frac{1}{m}.$$

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

$$\limsup_{n \to \infty} \mathcal{G}_n^{(p)}(u_n^{(m)}) \leqslant \mathcal{G}_\infty^{(p)}(u^{(m)}),$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

where $u_n^{(m)}$ is the recovery sequence for $u^{(m)}$ in the Γ-convergence $\mathcal{G}_\infty^{(p)} = \Gamma\text{-}\lim_{n\to\infty} \mathcal{F}_{\varepsilon_n'}^{(p)}$. For each $m \in \mathbb{N}$ let $n_m \in \mathbb{N}$ be such that

$$\mathcal{G}_n^{(p)}(u_n^{(m)}) \leqslant \mathcal{G}_\infty^{(p)}(u^{(m)}) + \frac{1}{m} \quad \text{and} \quad \|u_n^{(m)} \circ T_n - u^{(m)}\|_{L^1} \leqslant \frac{1}{m}$$

for all $n \geqslant n_m$. 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,

$$\limsup_{n\to\infty} \mathcal{G}_n^{(p)}(u_n) = \limsup_{m\to\infty} \sup_{n\in[n_m,n_{m+1})} \mathcal{G}_n^{(p)}(u_n^{(m)}) \leqslant \limsup_{m\to\infty} \left(\mathcal{G}_\infty^{(p)}(u^{(m)}) + \frac{1}{m}\right) \leqslant \mathcal{G}_\infty^{(p)}(u).$$

Similarly,

$$\lim_{n\to\infty} \|u_n \circ T_n - u\|_{L^1} \leqslant \limsup_{m\to\infty} \sup_{n\in[n_m,n_{m+1})} \left(\|u_n^{(m)} \circ T_n - u^{(m)}\|_{L^1} + \frac{1}{m}\right) \leqslant \lim_{m\to\infty} \frac{2}{m} = 0,$$

therefore u_n converges to u in TL^1. Hence u_n is a recovery sequence for u. This completes the proof. □

5 Convergence of minimisers with data fidelity

In this section we prove Corollary 1.

Proof of Corollary 1. In view of Theorem 2.21 and Proposition 3 it is enough to prove that, for any $u \in L^1(X, \mu)$ and any $\{u_n\}_{n=1}^\infty$ with $u_n \in L^1(X_n)$ such that $u_n \to u$ in $TL^1(X)$ it holds that

$$\lim_{n\to\infty} \mathcal{K}_n(u_n) = \mathcal{K}_\infty(u).$$

We can restrict ourselves to sequences $u_n \to u$ satisfying

$$\sup_{n\in\mathbb{N}} \mathcal{G}_n^{(p)}(u_n) < +\infty. \tag{5.1}$$

Let

$$v_n(x) := \begin{cases} u_n(x) & \text{if } x \in X_n(u_n), \\ 1 & \text{otherwise,} \end{cases}$$

where

$$X_n(u_n) := \{x \in X_n \,:\, |u_n(x)|^q \leqslant R_V\},$$

and R_V is as in Assumption (B3).

Step 1. We claim that

$$\lim_{n\to\infty} |\mathcal{K}_n(u_n) - \mathcal{K}_n(v_n)| \to 0, \tag{5.2}$$

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

as $n \to \infty$. Indeed we have that

$$\lim_{n\to\infty} |\mathcal{K}_n(u_n) - \mathcal{K}_n(v_n)| = \lim_{n\to\infty} \left| \frac{1}{n} \sum_{x_i \notin X_n(u_n)} (k_n(x_i, u_n(x_i)) - k_n(x_i, 1)) \right|$$
$$= \beta \lim_{n\to\infty} \frac{1}{n} \sum_{x_i \notin X_n(u_n)} (3 + |u_n(x_i)|^q),$$

where in the last equality we used Assumption (D2). Now,

$$\frac{1}{n} \sum_{x_i \notin X_n(u_n)} 1 \leqslant \frac{1}{nR_V} \sum_{x_i \notin X_n(u_n)} |u_n(x_i)|^q \leqslant \frac{1}{nR_V\tau} \sum_{x_i \notin X_n(u_n)} V(u_n(x_i)) \leqslant \frac{\varepsilon_n \mathcal{G}_n^{(p)}(u_n)}{\tau R_V}.$$

Using (5.1) we conclude (5.2).

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

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

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

Step 3. We show that $\mathcal{K}_n(v_n) \to \mathcal{K}_\infty(u)$. 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 relabelled) such that $v_n(T_n(x)) \to u(x)$ for almost every $x \in X$. Using Assumption (D3) we get

$$\lim_{n\to\infty} k_n\,(T_n(x), v_n(T_n(x))) = k_\infty(x, u(x)). \tag{5.3}$$

Moreover it holds

$$k_n(T_n(x), v_n(T_n(x))) \leqslant \beta\,(1 + |v_n(T_n(x))|^q) \leqslant \beta\,(1 + R_V). \tag{5.4}$$

Using (5.3) and (5.4) and applying the Lebesgue's dominated convergence theorem we get

$$\lim_{n\to\infty} \mathcal{K}_n(v_n) = \lim_{n\to\infty} \int_X k_n(T_n(x), v_n(T_n(x)))\,\rho(x)\,\mathrm{d}x$$
$$= \int_X k_\infty(x, u(x))\rho(x)\,\mathrm{d}x$$
$$= \mathcal{K}_\infty(u). \tag{5.5}$$

Since any subsequence of v_n has a further subsequence such that $\lim_{n\to\infty} \mathcal{K}_n(v_n) = \mathcal{K}_\infty(u)$, we can conclude the convergence is over the full sequence.

Step 4. Using (5.2) and (5.5) we conclude that

$$\lim_{n\to\infty} \mathcal{K}_n(u_n) \to \mathcal{K}_\infty(u)$$

as required. □

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

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. Both authors would like to thank the referees for valuable feedback that improved the paper.

Conflicts of interest

None.

References

[1] ALBERTI, G. & BELLETTINI, G. (1998) A nonlocal anisotropic model for phase transitions: Asymptotic behaviour of rescaled energies. *Eur. J. Appl. Math.* **9**, 261–284.

[2] ALBERTI, G. & BELLETTINI, G. (1998) A nonlocal anisotropic model for phase transitions I. The optimal profile problem. *Math. Ann.* **310**, 527–560.

[3] AMBROSIO, L., FUSCO, N. & PALLARA, D. (2000) *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York.

[4] ARORA, S., RAO, S. & VAZIRANI, U. (2009) Expander flows, geometric embeddings and graph partitioning. *J. ACM* **56**, Art. 5, 37.

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

[6] BARROSO, A. C. & FONSECA, I. (1994) Anisotropic singular perturbations the vectorial case. *Proc. R. Soc. Edinburgh Sect. A: Math.* **124**, 527–571.

[7] BERTOZZI, A. L. & FLENNER, A. (2012) Diffuse interface models on graphs for classification of high dimensional data. *Multiscale Model. Simul.* **10**, 1090–1118.

[8] BOUCHITTE, G. (1990) Singular perturbations of variational problems arising from a two-phase transition model. *Appl. Math. Optim.* **21**, 289–314.

[9] BOURGAIN, J., BREZIS, H. & MIRONESCU, P. (2001) Another look at Sobolev spaces. In: J. L. Menaldi, E. Rofman and A. Sulem (editors), *Optimal control and partial differential equations. In honour of Professor Alain Bensoussan's 60th birthday. Proceedings of the conference, Paris, France, December 4, 2000*, IOS Press, Amsterdam; Ohmsha, Tokyo, pp. 439–455.

[10] BOYKOV, Y., VEKSLER, O. & ZABIH, R. (2001) Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 1222–1239.

[11] BRAIDES, A. (2002) Γ-*Convergence for Beginners*, Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford.

[12] BRAIDES, A. & YIP, N. K. (2012) A quantitative description of mesh dependence for the discretization of singularly perturbed nonconvex problems. *SIAM Journal on Numerical Analysis* **50**, 1883–1898.

[13] BRESSON, X. & LAURENT, T. (2012) Asymmetric Cheeger cut and application to multi-class unsupervised clustering. Technical report UCLA CAM Report 12–27, Los Angeles.

[14] BRESSON, X., LAURENT, T., UMINSKY, D. & VON BRECHT, J. H. (2012) Convergence and energy landscape for Cheeger cut clustering. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS)*, pp. 1385–1393.

[15] BRESSON, X., LAURENT, T., UMINSKY, D. & VON BRECHT, J. H. (2013) Multiclass total variation clustering. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS)*, pp. 1421–1429.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

[16] Bresson, X., Tai, X.-C., Chan, T. F. & Szlam, A. (2014) Multi-class transductive learning based on l^1 relaxations of Cheeger cut and Mumford-Shah-Potts model. *J. Math. Imaging Vis.* **49**, 191–201.
[17] Calatroni, L., van Gennip, Y., Schönlieb, Rowland, H. M. & Flenner, A. (2017) Graph clustering, variational image segmentation methods and Hough transform scale detection for object measurement in images. *J. Math. Imaging Vis.* **57**, 269–291.
[18] Chambolle, A., Giacomini, A. & Lussardi, L. (2010) Continuous limits of discrete perimeters. *M2AN Math. Modell. Numer. Anal.* **44**, 207–230.
[19] Dal Maso, G. (1993) *An Introduction to* Γ*-Convergence*, Birkhäuser, Basel.
[20] Davis, E. & Sethuraman, S. (2018) Consistency of modularity clustering on random geometric graphs. *Ann. Appl. Probab.* **28**, 2003–2062.
[21] De Giorgi, E. & Franzoni, T. (1975) Su un tipo di convergenza variazionale. *Atti della Accademia Nazionale dei Lincei. Rendiconti. Classe di Scienze Fisiche, Matematiche e Naturali* **58**, 842–850.
[22] Dudley, R. M. (2010) *Real Analysis and Probability*, Cambridge University Press, Cambridge.
[23] Dunlop, M. M., Slepčev, D., Stuart, A. M. & Thorpe, M. (2018) *Large data and zero noise limits of graph-based semi-supervised learning algorithms*, https://arxiv.org/abs/1805.09450.
[24] Egger, A. & Huotari, R. (1990) Rate of convergence of the discrete pólya algorithm. *J. Approx. Theory* **60**, 24–30.
[25] Esedoğlu, S. & Otto, F. (2015) Threshold dynamics for networks with arbitrary surface tensions. *Commun. Pure Appl. Math.* **68**, 808–864.
[26] Fonseca, I. & Müller, S. (1993) Relaxation of quasiconvex functionals in $\mathrm{BV}(\Omega, \mathbf{R}^p)$ for integrands $f(x, u, \nabla u)$. *Arch. Ration. Mech. Anal.* **123**, 1–49.
[27] Fonseca, I. & Popovici, C. (2005) Coupled singular perturbations for phase transitions. *Asymptotic Anal.* **44**, 299–325.
[28] Fonseca, I. & Tartar, L. (1989) The gradient theory of phase transitions for systems with two potential wells. *Proc. R. Soc. Edinburgh Sect. A: Math.* **111**, 89–102.
[29] Gangbo, W. & McCann, R. J. (1996) The geometry of optimal transportation. *Acta Math.* **177**, 113–161.
[30] Garcia-Cardona, C., Merkurjev, E., Bertozzi, A. L., Flenner, A. & Percus, A. G. (2014) Multiclass data segmentation using diffuse interface methods on graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1600–1613.
[31] García-Trillos, N. (2016) Variational limits of k-NN graph based functionals on data clouds, https://arxiv.org/abs/1607.00696.
[32] García Trillos, N. & Murray, R. (2017) A new analytical approach to consistency and overfitting in regularized empirical risk minimization. *Eur. J. Appl. Math.* **28**, 886–921.
[33] García Trillos, N. & Slepčev, D. (2015) On the rate of convergence of empirical measures in ∞-transportation distance. *Can. J. Math.* **67**, 1358–1383.
[34] García Trillos, N. & Slepčev, D. (2016) Continuum limit of total variation on point clouds. *Arch. Ration. Mech. Anal.* **220**, 193–241.
[35] García Trillos, N. & Slepčev, D. (2018) A variational approach to the consistency of spectral clustering. *Appl. Comput. Harmon. Anal.* **45**, 239–281.
[36] García Trillos, N., Slepčev, D., von Brecht, J., Laurent, T. & Bresson, X. (2016) Consistency of Cheeger and ratio graph cuts. *J. Mach. Learn. Res.* **17**, 6268–6313.
[37] Kohn, R. & Sternberg, P. (1989) Local minimizers and singular perturbations. *Proc. R. Soc. Edinburgh Sect. A: Math.* **111**, 69–84.
[38] Kyng, R., Rao, A., Sachdeva, S. & Spielman, D. A. (2015) Algorithms for Lipschitz learning on graphs. In: *Proceedings of the 28th Conference on Learning Theory*, JMLR: Workshop and Conference Proceedings, Vol. 40:1–34, pp. 1190–1223.
[39] Merkurjev, E., Kostić, T. & Bertozzi, A. L. (2013) An MBO scheme on graphs for classification and image processing. *SIAM J. Imaging Sci.* **6**, 1903–1930.
[40] Modica, L. (1987) The gradient theory of phase transitions and the minimal interface criterion. *Arch. Ration. Mech. Anal.* **98**, 123–142.
[41] Modica, L. & Mortola, S. (1977) Un esempio di Γ-convergenza. *Bollettino dell'Unione Matematica Italiana B* **14**, 285–299.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press

[42] NG, A. Y., JORDAN, M. I. & WEISS, Y. (2001) On spectral clustering: Analysis and an algorithm. In: *Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS)*, pp. 849–856.
[43] OWEN, N. C. & STERNBERG, P. (1991) Nonconvex variational problems with anisotropic perturbations. *Nonlinear Anal.* **16**, 705–719.
[44] PENROSE, M. (2003) *Random Geometric Graphs*. Oxford Studies in Probability, Vol. 5, Oxford University Press, Oxford.
[45] PONCE, A. C. (2004) A new approach to Sobolev spaces and connections to Γ-convergence. *Calculus Var. Partial Differ. Equ.* **19**, 229–255.
[46] SAVIN, O. & VALDINOCI, E. (2012) Γ-convergence for nonlocal phase transitions. *Annales de l'Institut Henri Poincaré. Analyse Non Linéaire* **29**, 479–500.
[47] SHI, J. & MALIK, J. (2000) Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905.
[48] SLEPČEV, D. & THORPE, M. (2017) Analysis of *p*-Laplacian regularization in semi-supervised learning, https://arxiv.org/abs/1707.06213.
[49] STERNBERG, P. (1988) The effect of a singular perturbation on nonconvex variational problems. *Arch. Ration. Mech. Anal.* **101**, 209–260.
[50] SZLAM, A. & BRESSON, X. (2009) A total variation-based graph clustering algorithm for cheeger ratio cuts. Technical report UCLA CAM Report 9–68, Los Angeles, pp. 1–12.
[51] SZLAM, A. & BRESSON, X. (2010) Total variation and Cheeger cuts. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML)*, Omnipress, USA, pp. 1039–1046.
[52] THORPE, M., PARK, S., KOLOURI, S., ROHDE, G. K. & SLEPČEV, D. (2017) A transportation L^p distance for signal analysis. *J. Math. Imaging Vis.* **59**, 187–210.
[53] THORPE, M. & SLEPČEV, D. (2017, in preparation) Transportation L^p distances: Properties and extensions.
[54] THORPE, M. & THEIL, F. (2017, to appear) Asymptotic analysis of the Ginzburg-Landau functional on point clouds. *Proc. R. Soc. Edinburgh Sect. A: Math.* https://arxiv.org/abs/1604.04930.
[55] VAN GENNIP, Y. & BERTOZZI, A. L. (2012) Γ-convergence of graph Ginzburg-Landau functionals. *Adv. Differ. Equ.* **17**, 1115–1180.
[56] VAN GENNIP, Y. & SCHÖNLIEB, C.-B. (2017) Introduction: Big data and partial differential equations. *Eur. J. Appl. Math.* **28**, 877–885.
[57] VILLANI, C. (2003) *Topics in Optimal Transportation*. Graduate Studies in Mathematics, Vol. 58, American Mathematical Society, Providence, RI.

https://doi.org/10.1017/S0956792518000645 Published online by Cambridge University Press