IOP Publishing | London Mathematical Society Nonlinearity

Nonlinearity **33** (2020) 3846–3888 https://doi.org/10.1088/1361-6544/ab81ee

Mumford–Shah functionals on graphs and their asymptotics

Marco Caroccia[1], Antonin Chambolle[2] and Dejan Slepčev[3,4]

[1] Dipartimento di Matematica, Università di Roma 'Tor Vergata', 00133 Rome, Italy
[2] CMAP, École Polytechnique 91128 Palaiseau Cedex, France
[3] Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, United States of America

E-mail: caroccia.marco@gmail.com, antonin.chambolle@polytechnique.fr and slepcev@math.cmu.edu

Received 17 June 2019, revised 3 March 2020
Accepted for publication 20 March 2020
Published 5 June 2020

Abstract

We consider adaptations of the Mumford–Shah functional to graphs. These are based on discretizations of nonlocal approximations to the Mumford–Shah functional. Motivated by applications in machine learning we study the random geometric graphs associated to random samples of a measure. We establish the conditions on the graph constructions under which the minimizers of graph Mumford–Shah functionals converge to a minimizer of a continuum Mumford–Shah functional. Furthermore we explicitly identify the limiting functional. Moreover we describe an efficient algorithm for computing the approximate minimizers of the graph Mumford–Shah functional.

Keywords: nonlocal variational problems, variational problems with randomness, discrete to continuum limit, asymptotic consistency, Gamma convergence, regression

Mathematics Subject Classification numbers: 49J55, 62G20, 65N12.

(Some figures may appear in colour only in the online journal)

1. Introduction

Our investigation of graph based Mumford–Shah functionals is motivated by problems arising in machine learning. Given a point cloud in a Euclidean space with (noisy) real-valued labels, or an undirected graph with labeled vertices, we investigate a model to denoise the labels while allowing for jumps (discontinuities) in label values. As does the classical Mumford–Shah for images, the functional we consider allows one to identify the locations of sharp transitions of

[4] Author to whom any correspondence should be addressed.

label values. Our primary focus is on graphs arising as neighborhood graphs of point clouds in Euclidean spaces of dimension two or higher, where we can carry out rigorous analysis. However some of the functionals we study can be formulated purely in the setting of weighted graphs and may be useful in applications.

The model we study is based on ideas from image processing and go back to the celebrated Mumford and Shah [MS89] variational model for image segmentation. To adapt the Mumford–Shah functional to point clouds and graphs we rely on the work of Gobbino [Gob98] and Gobbino and Mora [GM01] who introduced a family of nonlocal models which approximate the Mumford–Shah functional. A discrete version of Gobbino's work, on a uniform square/cubic lattice, was then studied in [Cha99]. Ruf [Ruf17] has recently adapted these nonlocal models to random discrete setting and studied them in the setting of stochastic lattices. Here we study such functionals in the setting of random geometric, and related, graphs relevant to machine learning.

General graph setting. Consider an undirected weighted graph with vertices $V = \{1, \dots, n\}$ and edge weights matrix $W = [w_{ij}]_{i,j=1,\dots,n}$. The edge weights are considered to be nonnegative and symmetric. Let $f : V \to \mathbb{R}$ be the observed noisy labels. Let $\zeta : [0, \infty) \to [0, \infty)$ be concave and such that $\zeta(0) = 0$, $0 < \zeta'(0) < \infty$. We define the *graph Mumford–Shah* functional acting on $u : V \to \mathbb{R}$ as

$$\mathcal{GMS}_f(u) := \frac{\lambda}{n}\sum_{i=1}^{n} |u_i - f_i|^2 + \frac{1}{\varepsilon n^2}\sum_{i,j=1}^{n} \zeta\left(\frac{1}{\varepsilon}|u_i - u_j|^2\right) w_{ij}. \tag{1.1}$$

We note that when the differences $u_i - u_j$ are relatively small the functional is similar to the graph dirichlet energy, while for large values of $u_i - u_j$ the functional saturates and in some ways considers u to be discontinuous over the edge. It then just penalizes the size of the set of discontinuities. Minimizing the functional allows one to find the sharp transitions in the data by detecting edges where $u_i - u_j$ is large compared to ε. That is the parameter $\varepsilon > 0$ sets the scale for what differences of the values are considered 'large'. We note that the functional is nonconvex.

Geometric graph setting. We now consider the setting of point clouds and the random geometric graphs generated by them. The ability to measure the distance between vertices allow us to create a larger family of graph Mumford–Shah functionals. Let $V_n = \{x_1, \dots, x_n\}$ be a set of points in $\mathbb{R}^d$. The points x_i are typically random samples of a measure describing the data distribution, but this interpretation is not essential in defining the functional. Given these points we define a graph by setting the edge weights to be $w_{i,i} = 0$ and for $i \neq j$

$$w_{ij} = \varepsilon^{-d}\eta\left(\frac{|x_i - x_j|}{\varepsilon}\right) =: \eta_\varepsilon(|x_i - x_j|) \tag{1.2}$$

where η is a nonnegative, nonincreasing function which decays to 0 faster than a specified algebraic rate. Let $f : V_n \to \mathbb{R}$ be the observed noisy labels and let ζ be as in the graph setting above. For $p \in [1, d)$ and $q \in [0, p-1]$ we define the *graph Mumford–Shah* functional acting on $u : V_n \to \mathbb{R}$ as

$$\mathcal{GMS}_{f,\varepsilon,n}(u) := \frac{\lambda}{n}\sum_{i=1}^{n} |u(x_i) - f_i|^2 + \frac{1}{\varepsilon}\frac{1}{n^2}\sum_{i,j=1}^{n} \zeta\left(\varepsilon^{1-p+q}\frac{|u(x_i) - u(x_j)|^p}{|x_i - x_j|^q}\right)\eta_\varepsilon(|x_i - x_j|). \tag{1.3}$$

We note that taking $q = 0$ reduces this functional to the one considered in the pure graph setting.

We rigorously study of the asymptotics of $\mathcal{GMS}_{\varepsilon,n}(u)$ as $n \to \infty$ and $\varepsilon \to 0$ and establish in theorem 2.1 that its minimizers converge to minimizers of a Mumford–Shah functional posed in a continuum Euclidean domain. We note that related results for a stochastic lattice model have been obtained recently by Ruf [Ruf17], see remark 2.5. The conditions of the theorem 2.1 are optimal in terms of scaling of ε_n on n for which the convergence holds for all dimensions $d \geqslant 2$. To show the result we follow the general strategy of [GTS16] and use a number of results of calculus of variations, in particular the works of Gobbino and Mora [GM01, Gob98]. There are two notable advances:

(a) We introduce a strategy to overcome the issues that arise from the lack of control of the denominator in (1.3). Namely the discrepancy in the quotients inside of ζ can be large if the standard tools to compare the discrete and continuum functionals using a transport map are used directly. In remark 4.6 we outline the steps we subsequently take to overcome this difficulty.

(b) Unlike in [GTS16], our results have optimal scaling in 2D. Using lemmas 3.1 and 3.2 we develop an approach to Γ-convergence that uses a more relaxed way to compare the discrete and continuum measures. In particular the approach outlined at the beginning of section 4.2 would allow one to obtain optimal estimates for total variation, Laplacian, and p-Laplacian functionals considered in [GTS16, GTS18, ST19] respectively. We note that for the graph total variation optimal estimates in 2D were recently obtained by Müller and Penrose [MP18]. The approach here is simpler, but does use the insight of Müller and Penrose that binning at an intermediate scale can be advantageous.

Organization. In section 1.1 we review the works on related problems, primarily on the mathematical aspects of related data science questions. In section 2 we introduce the graph based and the continuum functionals and state the main results. In section 3 we recall the mathematical notion of Γ-convergence and its main properties and we recall the TL^p space and its main properties. We introduce the relaxed way to compare measures with the empirical measures of their samples. In section 4 we prove the main results on Γ-convergence, while in section 5 we prove the accompanying compactness result. In section 7 we describe an algorithm for computing the approximate (local) minimizers of the graph Mumford–Shah functional and perform numerical experiments on synthetic data to showcase its properties and on real-estate sales data to highlight its applicability in prediction problems. In appendices A and B we prove two technical results needed in section 4.

1.1. Related works

Here we review the related models in data analysis. The basic background about the Mumford–Shah functional has been listed in the introduction. Further mathematical works which serve as the basis for our proofs are recalled as we present our approach in sections 3 and 4.

Regularizing and denoising functions given on graphs has been studied in variety of contexts in machine learning. Here we focus on regularizations which still allow for the jumps in the regularized function. There are two lines of research which have led to such functionals. One, as is the case with our approach, is inspired by models in image processing where variational approaches have been widely used for image denoising and segmentation. Particularly relevant in the context of imaging are the works of Chan and Vese [CV01, VC02], who proposed a piecewise constant simplification of the Mumford–Shah functional and have shown its effectiveness in image segmentation, and Rudin *et al* [ROF92] who proposed a TV

(total variation) based regularization for the image denoising. In analogy with Chan and Vese, [HLPB13] Hu *et al* formulated the piecewise-constant Mumford–Shah functional on graphs. They also developed an efficient numerical approach to compute the minimizers and used it to study a (multi-class) classification problem. An ROF functional on graphs, with L^1 fidelity term, was studied by Trillos and Murray [GTM17].

TV based regularizations have also been developed in the statistics community. Mammen and van de Geer [MvdG97] have considered it in the setting of nonparametric regression and have shown that the TV regularization provides an estimator that achieves the optimal min–max recovery rate in one dimension over noisy samples of functions in the unit ball with respect to the BV norm. TV based regularizations in higher dimensions have been considered by Tibshirani *et al* [TSR+05] who call the functional fused LASSO. Hütter and Rigollet [HR16], show that, up to logarithms, in dimension $d \geqslant 2$, the TV regularization on grids achieves the optimal min–max rate over the unit ball with respect to the BV norm. Recently, Padilla *et al* [PSCW18] show for random geometric graphs and for KNN graphs that up to logarithms, in dimension $d \geqslant 2$, TV regularization again achieves the optimal min–max rate.

The work of Hallac *et al* [HLB15] extends the fused LASSO to the graph setting and considers some further functionals which are closely related to the graph Mumford–Shah functional we consider here. In particular the initial models of the paper deal with convex functionals which include graph total-variation based terms, and are thus called 'network LASSO'. The second part of the paper modifies the total-variation term, which leads to nonconvex functionals. Here we interpret some of these nonconvex functionals, in particular model (7) of [HLB15]), as the graph-based Mumford–Shah functional, which, together with our asymptotic results, explains the behavior of these models. Wang *et al* [WSST16] consider higher order total variation regularizers on graphs. We also note that the use of total variation penalization for signal denoising and filtering has also been considered in the signal processing community, see for example the work of Chen *et al* [CSMK15].

2. Setting and main results

2.1. Continuum Mumford–Shah functional and its nonlocal approximation

In their celebrated paper [MS89], Mumford and Shah proposed a variational approach for image segmentation. Given a domain $\Omega \subset \mathbb{R}^d$ and a potentially noisy image with intensity f they sought to approximate it by a piecewise smooth function u, whose discontinuities delineate the segments of the image.

We recall their functional using the formulation in the space of special functions of bounded variation. For background on spaces of (special) functions of bounded variation we refer the reader to the book [AFP00]. For $u \in \mathrm{SBV}(\Omega)$

$$MS_f(u) := \lambda \int_\Omega |u - f|^2 \, \mathrm{d}x + \int_\Omega |\nabla u|^2 \mathrm{d}x + \mathcal{H}^{d-1}(S_u) \tag{2.1}$$

where $f \in L^\infty(\Omega)$ is the noisy image, ∇u is the absolutely continuous (in the measure theoretic sense, and with respect to the Lebesgue measure) part of the gradient Du (which is a measure) of the function u, S_u is the jump set of u, and $\mathcal{H}^{d-1}$ is the $(d-1)-$dimensional Hausdorff measure. The first term of the functional ensures the closeness of the approximation u to the original image f, while the next two terms reward the regularity of u. The idea is that natural images are piecewise smooth, but often do have jumps in intensity between different

regions. Thus the functional rewards the regularity of u, while still allowing for jumps in the intensity.

Thanks to the work of Ambrosio in [Amb89] and to the lower-semicontinuity of MS_f with respect to the topology of the space $\mathrm{SBV}(\Omega)$, the direct method of calculus of variation ensures us that a minimum $u_0 \in \mathrm{SBV}(\Omega)$ for the functional (2.2) is always attained.

For the considerations we have in mind the fidelity term $\lambda \int_\Omega |u-f|^2 \mathrm{d}x$ is quite straightforward to treat. Hence, for readability, we introduce the functional without it and focus mainly on this functional:

$$MS(u) := \int_\Omega |\nabla u|^2 \mathrm{d}x + \mathcal{H}^{d-1}(S_u). \tag{2.2}$$

As shown in [BDM97] any functional of the form of (2.2) cannot be approximated in the sense of Γ-convergence by *local* integral functionals of the type

$$\int_\Omega h_\varepsilon(\nabla u(x))\mathrm{d}x$$

where $u \in W^{1,2}(\Omega)$. De Giorgi conjectured that the Mumford–Shah functional can be approximated by nonlocal functionals. The conjecture was proved by Gobbino in [Gob98], who showed that (2.2) can be approximated by the functionals

$$G_\varepsilon(u) := \frac{1}{\varepsilon^{d+1}} \int_{\mathbb{R}^d \times \mathbb{R}^d} \arctan\left(\frac{|u(y)-u(x)|^2}{|y-x|}\right) \mathrm{e}^{-\frac{|y-x|^2}{\varepsilon^2}} \mathrm{d}x\mathrm{d}y \tag{2.3}$$

defined for $u \in L^1_{\mathrm{loc}}(\Omega)$. He shows that for appropriate dimensional constants θ, σ

$$\Gamma\text{-}\lim_{\varepsilon\to 0} G_\varepsilon = \theta \int_\Omega |\nabla u|^2 \mathrm{d}x + \sigma \mathcal{H}^{d-1}(S_u)$$

where the Γ-limit is considered with respect to L^1 topology. The work in [Gob98] has been then generalized in [GM01] to functionals defined on $\mathrm{SBV}(\Omega)$ of the form

$$F(u) := \int_\Omega \varphi\left(|\nabla u(x)|\right) \mathrm{d}x + \int_{S_u} \psi(|u^+(x) - u^-(x)|) \mathrm{d}\mathcal{H}^{n-1}(x) \tag{2.4}$$

where $u^+(x)$ and $u^-(x)$ denote the so-called approximate lim inf and lim sup of u at the point x:

$$u^+(x) = \sup\left\{ t \in \mathbb{R} : \lim_{r\to 0+} \frac{1}{r^n} |\{y \in B(x,r) : u(y) > t\}| > 0 \right\}. \tag{2.5}$$

They show that for suitable φ, ψ the functional can be approximated in the Γ-convergence sense with the family of non-local functionals of the form

$$F_\varepsilon(u) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi_{|x-y|}\left(\frac{|u(x)-u(y)|}{|x-y|}\right) \eta_\varepsilon(x-y) \mathrm{d}x\mathrm{d}y \tag{2.6}$$

where $\{\varphi_\varepsilon\}_\varepsilon$ is a family of functions related to φ, ψ and $\{\eta_\varepsilon\}_{\varepsilon>0} \subset L^1(\Omega)$ is a kernel.

2.2. Point cloud Mumford–Shah functional

The above nonlocal approximation to the Mumford–Shah functional can be adapted to the graph setting. We consider the setting of random geometric graphs formulated on random samples of a measure μ with density ρ, which describes the underlying data distribution. Consider an open, bounded set with Lipschitz boundary Ω. The density ρ is assumed to satisfy: $\rho \in C^1(\Omega) \cap C^0(\overline{\Omega})$ and

$$0 < c \leqslant \min_{x\in\overline{\Omega}} \rho(x) \leqslant \max_{x\in\overline{\Omega}} \rho(x) \leqslant C < \infty. \tag{2.7}$$

We consider $\zeta : [0,\infty) \to [0,\infty)$ such that

(A1) ζ is concave and differentiable at 0;
(A2) ζ is non decreasing;
(A3) $\zeta'(0) < \infty$ and

$$\Theta := \lim_{x\to\infty} \zeta(x). \tag{2.8}$$

We fix $p \geqslant 1$, $q \in [0,p)$ and we assume that the kernel $\eta : [0,\infty) \to [0,\infty)$ satisfies

(B1) η is a nonincreasing L^1 function, non identically 0;
(B2) $0 < \int_0^\infty (t^d + t^{p-q+d-1})\eta(t)\mathrm{d}t < \infty$.

In the sequel, we always assume the functions η, ζ and ρ to satisfy the above assumptions.

Let $x_1, \ldots, x_n \in \Omega$ a set of n i.i.d. random points on Ω chosen according to the probability measure $\mu = \rho \mathrm{d}x$. The empirical measure of the sample is defined by

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

Given a Borel measure σ on Ω, the space $L^p(\Omega,\sigma)$ is the space of equivalence classes of measurable functions $u : \Omega \to \mathbb{R}$ with $\int_\Omega |u|^p \mathrm{d}\sigma$ finite. Notice that, under this assumption on ρ, we have that $L^1(\Omega;\rho) = L^1(\Omega)$. For that reason we often write $u \in L^1(\Omega)$ in place of $u \in L^1(\Omega;\rho)$.

The graph Mumford–Shah functional we devote the most attention to is the functional (1.3) without the fidelity term. Namely for a function $u \in L^1(\Omega;\mu_n)$ let

$$\mathcal{GMS}_{\varepsilon,n}(u) := \frac{1}{\varepsilon}\frac{1}{n^2}\sum_{i,j=1}^n \zeta\left(\varepsilon^{1-p+q}\frac{|u(x_i)-u(x_j)|^p}{|x_i-x_j|^q}\right)\eta_\varepsilon(|x_i-x_j|). \tag{2.9}$$

Here $\eta_\varepsilon(s) := \varepsilon^{-d}\eta(s/\varepsilon)$.

2.3. Main results

In theorem 2.1 we prove the Γ-convergence of the graph Mumford–Shah functional (2.9) as $n \to \infty$ and $\varepsilon_n \to 0$ (at allowable rate) towards the following continuum Mumford–Shah functional:

$$MS_{\eta,\zeta}(u;\rho) := \vartheta_\eta(p,q)\zeta'(0)\int_\Omega |\nabla u(x)|^p \rho(x)^2 \mathrm{d}x + \sigma_\eta \Theta \int_{S_u} \rho(y)^2 \mathrm{d}\mathcal{H}^{d-1}(y) \tag{2.10}$$

defined for all $u \in \mathrm{SBV}^p(\Omega)$ and where Θ is defined by (2.8) and

$$\begin{cases} \vartheta_\eta(p,q) & := 2\omega_{d-1}\dfrac{\Gamma(p/2+1/2)\Gamma(d/2+1/2)}{\Gamma(p/2+d/2)}\displaystyle\int_0^\infty t^{p-q+d-1}\eta(t)\mathrm{d}t \\ \sigma_\eta & := 2\omega_{d-1}\displaystyle\int_0^\infty t^d\eta(t)\mathrm{d}t. \end{cases} \tag{2.11}$$

The Γ convergence is the TL^1 sense, recalled in section 3.1. We point out that assumption (B2) on η is the one that guarantees the finiteness of $\sigma_\eta, \vartheta_\eta(p,q)$.

Theorem 2.1 (Γ-*convergence*). *Let Ω be an open, bounded set with Lipschitz boundary and ρ be a probability density satisfying (2.7). Consider ζ, η satisfying the assumptions (A1)–(A3) and (B1)–(B2).*

Let $\{x_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random points chosen accordingly to the density ρ and $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 such that

$$\lim_{n\to\infty}\frac{(\log(n))^{1/d}}{\varepsilon_n n^{1/d}} = 0 \quad \textit{for } d \geqslant 2. \tag{2.12}$$

Then $\mathcal{GMS}_{\varepsilon_n,n}$, defined in (2.9), $\Gamma-$converges to $MS_{\eta,\zeta}(\cdot;\rho)$, defined in (2.2), in the TL^1 sense.

We refer to [GTS16] for detailed introduction of the TL^p topology. For the reader's convenience we retrieve the main concepts in section 3.1 below.

Remark 2.2. The condition (2.12) of theorem 2.1 comes from the following fact. Given random samples $\{x_1,\dots,x_n\}$ as above, we show in lemma 3.1 that there exists a sequence of probability measures $\tilde{\mu}_n$, absolutely continuous with respect to Lebesgue measure such that $\frac{\mathrm{d}\tilde{\mu}_n}{\mathrm{d}\mu} \rightrightarrows 1$ and whose ∞-Wasserstein distance from the empirical measure of the sample μ_n is decaying faster than ε_n. More precisely, there exist $T_n : \Omega \to \{x_1,\dots,x_n\}$ transport maps between $\tilde{\mu}_n = \rho_n\mathcal{L}^d$ and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, such that

$$\lim_{n\to\infty}\frac{\|T_n - Id\|_\infty}{\varepsilon_n} = 0. \tag{2.13}$$

In section 5 we discuss compactness of the functionals. In particular we establish the following:

Theorem 2.3. *Let Ω, ρ, ζ, η, and $x_i, i = 1,\dots,n$ satisfy the assumptions of theorem 2.1. Consider a sequence of $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfying (2.12). If $u_n \in L^\infty(\Omega;\mu_n)$ satisfy*

$$\sup_{n\in\mathbb{N}}\{\|u_n\|_\infty + \mathcal{GMS}_{\varepsilon_n,n}(u_n)\} < \infty,$$

then the sequence $\{(\mu_n,u_n)\}_{n\in N}$ is TL^1-relatively compact.

Remark 2.4. Note that we ask for an L^∞ bound on the sequence, instead of a weaker L^1 bound (as was done in [GTS16, theorem 1.2]). Here L^1 bound would not be sufficient as we show in section 5 and remark 5.1. On the other hand since the signal f in (2.1) is bounded in applications, the minimizers are also bounded by the same bound, and hence the L^∞ boundedness of the sequence of interest is assured.

Remark 2.5. Recently Ruf [Ruf17] has studied the convergence of graph Mumford–Shah functionals on random lattices to the continuum Mumford–Shah functional. These interesting results are closely related, but also substantially different both in terms of their nature

and the techniques used. One difference is the nature of randomness of the structure considered. Here we consider random samples, or in fact any discrete sets of points whose empirical measures weakly approximate the continuum measure. Ruf considers random lattices in 2D, which have precise requirements of the distribution of points. The disordered structure of the points we allow forces us to require that the typical degree of a vertex converges to infinity faster than log$n$, while Ruf is able to work with graphs of bounded degree. On the flip side we identify the Γ-limit explicitly, while, due to the graph construction, Ruf only identifies the functional up to a constant. In a sense he is able to work under homogenization type graph-behavior where compactness arguments show that a Γ-limit exists without fully identifying it.

2.3.1. Convergence of functionals with the fidelity term.. The theorems 2.1 and 2.3 enable us to show the convergence of the Mumford–Shah functional with the fidelity term as well. We establish two results. The first one is in the setting without noise. In order to be able to evaluate the signal at sample points we assume that f is a bounded piecewise continuous function, that is that the set of discontinuities J_f is of finite $d-1$ dimensional Hausdorff measure, $\mathcal{H}^{d-1}(J_f) < \infty$.

Corollary 2.6. *Let Ω, ρ, ζ, η, and $x_i, i = 1, \dots, n$ satisfy the assumptions of theorem 2.1 and assume that $p \geqslant q+1$. Assume $f : \Omega \to \mathbb{R}$ is a bounded, piecewise continuous function. Consider a sequence of $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfying (2.12). Then the functional $\mathcal{GMS}_{f,\varepsilon,n}$ defined in (1.3), considered with $f_i = f(x_i)$ for $i = 1, \dots, n$, Γ-converges in TL^2 topology to $MS_{\eta,\zeta}(u;\rho) + \lambda \int |u-f|^2 \rho(x)\mathrm{d}x$, where $MS_{\eta,\zeta}(u;\rho)$ is defined in (2.10). Furthermore any sequence of minimizers u_n of $\mathcal{GMS}_{f,\varepsilon,n}$, converge along a subsequence to a minimizer of $MS_{\eta,\zeta}(u;\rho) + \lambda \int |u-f|^2 \rho(x)\mathrm{d}x$.*

We note that due to the fidelity term the topology of Γ convergence in the corollary is TL^2 instead of TL^1. We remark that the change of the topology when considering the fidelity term was not needed in [Gob98] since Gobbino could rely on the Fubini's theorem. However due the fact that we also deal with the discrete-to-continuum passage, the stronger topology is needed.

More importantly and more interestingly we are able to establish the convergence of minimizers of the graph Mumford–Shah functional when the labels are noisy. We note that the limit is a minimizer of a deterministic variational problem, even though the amount of noise does not vanish as $n \to \infty$.

Corollary 2.7. *Let Ω, ρ, ζ, η satisfy the assumptions of theorem 2.1 and assume that $p \geqslant q+1$. Assume $f : \Omega \to \mathbb{R}$ is a bounded, piecewise continuous function. Let β be a measure on $\mathbb{R}$ modeling the noise. We assume β has compact support and mean zero. Let (x_i, y_i) for $i = 1, \dots, n$ be i.i.d. samples of the measure $\mu \times \beta$. Consider a sequence of $\{\varepsilon_n\}_{n\in\mathbb{N}}$ satisfying (2.12).*

Then the functional $\mathcal{GMS}_{f,\varepsilon,n}$ defined in (1.3), considered with

$$f_i = f(x_i) + y_i \quad \text{for } i = 1, \dots, n,$$

Γ-converges in TL^2 topology to $MS_{\eta,\zeta}(u;\rho) + \lambda \int |u-f|^2 \rho(x)\mathrm{d}x + \mathrm{Var}(\beta)$, where $MS_{\eta,\zeta}(u;\rho)$ is defined in (2.10). Furthermore any sequence of minimizers u_n of $\mathcal{GMS}_{f,\varepsilon,n}$, converge along a subsequence to a minimizer of $MS_{\eta,\zeta}(u;\rho) + \lambda \int |u-f|^2 \rho(x)\mathrm{d}x$.

We make several observations. Note that while the minimizers of the random functional converge to the minimizers of a deterministic functional, and that the limit of the minimizers does not depend on the amount of noise. In a sense noise does not create a bias. The randomness

affects the limit in that the actual values of the random discrete functionals are higher when there is more noise which leads to the presence of Var(β) in the deterministic limit. We note that while we do not allow for Gaussian noise this is purely for technical reasons, to make the proof of compactness easier. On the other hand we do not require the noise to have continuous density with respect to Lebesgue measure.

Remark 2.8. Let us contrast the result of corollary 2.7 to results on min–max recovery rates in nonparametric regression that we mentioned in the introduction (see [HR16, MvdG97, PSCW18, WSST16]). In the setting of regression one is concerned with recovering a function $f^\dagger$ in some functional class (e.g., BV unit ball) whose noisy samples are available. Thus the fidelity term is made stronger as $n \to \infty$. Namely λ in (1.3) is taken to infinity at appropriate rate as n increases. The works obtain rates at which minimizers of functionals like (1.3), with TV regularization instead of the Mumford–Shah term, converge to $f^\dagger$. We conjecture that for the functions $f^\dagger$ in BV ball the functional (1.3) also achieves the optimal min–max rate. One difference between the Mumford–Shah and the TV regularization is that the Mumford–Shah one does not decrease the contrast over sharp edges as TV regularization does. See example 7.1. In the context of regression our contribution is that by taking the limit $n \to \infty$ as λ is fixed we shed the light on what is the precise amount of regularization introduced by the Mumford–Shah term at finite λ.

2.3.2. Extension to data on a manifold. In machine learning it is often relevant to consider data that lie in a potentially high dimensional space, but have an intrinsic low dimensional structure. Here we remark that it is straightforward to extend our results to the setting where data are sampled from a measure μ whose support is a d-dimensional C^2 manifold, without boundary, $\mathcal{M}$, embedded in $\mathbb{R}^D$ for some $D > d$. We require that the measure μ has a continuous density ρ with respect to the volume form of the manifold $\mathcal{M}$: $\mathrm{d}\mu = \rho \mathrm{dVol}_{\mathcal{M}}$. The form of the graph Mumford–Shah functional remains the same, while the only change in the limiting functional (2.10) is that ∇u is replaced by the manifold gradient $\mathrm{grad}_{\mathcal{M}} u$. We note that the scaling of η_ε and the definition of σ_η depend on the intrinsic dimension d, but not the ambient dimension D. Full details of how related statements are extended to the manifold setting can be found in [GTGHS18].

2.3.3. Extension to vector valued functions. We note that in machine learning it is also natural to consider functions on graphs which are vector valued. The graph Mumford–Shah functional (1.3) can be considered for functions u with values in $\mathbb{R}^m$. In fact such functionals have been used in the work of Hallac *et al* [HLB15]. We do not rigorously treat the limits of vector-valued functionals in this paper. Nevertheless we remark that we expect that the Γ-limit of the vector valued graph Mumford–Shah functional for $p = 2$ is the following Mumford–Shah type functional: for $u \in \mathrm{SBV}(\Omega)^m$

$$MS^m_{\eta,\zeta}(u;\rho) := \vartheta_\eta(2,q)\zeta'(0)\int_\Omega |\nabla u(x)|^2 \rho(x)^2 \mathrm{d}x + \sigma_\eta \Theta \int_{S_u} \rho(y)^2 \mathrm{d}\mathcal{H}^{d-1}(y)$$

where S_u is the union of jump sets for each of coordinate functions, $u = (u_1, \ldots, u_m)$ and $S_u = \bigcup_{j=1}^m S_{u_i}$, and where Θ is defined by (2.8) and ϑ_η and σ_η by (2.11). We furthermore expect that one can prove this using similar techniques that we use. We note that this would require a careful verification of the slicing argument in multiple dimensions.

3. Γ-convergence and topology in the space of configuration

Given a sequence of functionals $F_n : X \to \mathbb{R}$ and a metrizable (and separable) topology $\mathcal{T}$ on X we say that $F_n\Gamma$-converges at $F : X \to \mathbb{R}$ with respect to the topology $\mathcal{T}$ if the following two conditions are satisfied:

(a) For every sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ such that $x_n \xrightarrow{\mathcal{T}} x$ it holds that

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

(b) For all $x \in X$ there exists a sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ such that $x_n \xrightarrow{\mathcal{T}} x$ and for which

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

In this case we write

$$\Gamma\text{-}\lim_{n\to 0} F_n = F.$$

Notice that, if $\Gamma\text{-}\lim_{n\to 0} F_n = F$ then the following assertions also hold:

(a) F is lower semi-continuous and

$$\begin{aligned} F(x) &= \inf\left\{ \liminf_{n\to\infty} F_n(x_n) \,\middle|\, \{x_n\}_{n\in\mathbb{N}} \subset X, \quad x_n \xrightarrow{\mathcal{T}} x \right\} \\ &= \inf\left\{ \limsup_{n\to\infty} F_n(x_n) \,\middle|\, \{x_n\}_{n\in\mathbb{N}} \subset X, \quad x_n \xrightarrow{\mathcal{T}} x \right\}; \end{aligned}$$

(b) If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of minima of F_n on X, namely

$$F_n(x_n) = \min_{y\in X} \{F_n(y)\},$$

converging to some $x \in X$ in the topology $\mathcal{T}$ then x is a minimum of F on X:

$$F(x) = \min_{y\in X} \{F(y)\}.$$

3.1. The TL^p topology: brief overview

The TL^p space has been introduced in [GTS16]. Given a bounded open set Ω let $\mathcal{P}(\Omega)$ be the set of Borel probability measures on $\overline{\Omega}$. The space $TL^p(\Omega)$ is defined by

$$TL^p(\Omega) := \{(\mu, f) : \mu \in \mathcal{P}(\Omega)\, f \in L^p(\Omega;\mu)\}. \tag{3.1}$$

Given $(\mu,f),(\nu,g) \in TL^p(\Omega)$ their TL^p distance is defined as follows

$$d_{TL^p}((\mu,f),(\nu,g)) := \begin{cases} \inf\limits_{\pi\in\Gamma(\mu,\nu)} \left(\displaystyle\int_{\Omega\times\Omega} |x-y|^p + |f(x)-g(y)|^p \mathrm{d}\pi(x,y) \right)^{\frac{1}{p}} & \text{if } p \in [1,\infty), \\ \inf\limits_{\pi\in\Gamma(\mu,\nu)} \operatorname{ess\,sup}_{(x,y)\in\operatorname{supp}\pi} |x-y| + |f(x)-g(y)| & \text{if } p = \infty \end{cases}$$

where the infimum is taken among all transport plans π (i.e. couplings) between μ and ν:

$$\Gamma(\mu,\nu) := \{\gamma \in \mathcal{P}(\Omega\times\Omega) : (\forall A \text{ Borel})\, \gamma(A\times\Omega) = \mu(A),\ \gamma(\Omega\times A) = \nu(A)\}.$$

Given $\mu \in \mathcal{P}(\Omega)$ and a measurable mapping $T : \Omega \to \Omega$, we recall that $\nu = T_{\#}\mu$ is the *push-forward* of μ by T, namely the measure on Ω such that for any A Borel $\nu(A) = \mu(T^{-1}(A))$. A consequence of the definition is the following change of variables identity

$$\int_\Omega f(T(x))\mathrm{d}\mu(x) = \int_\Omega f(y)\mathrm{d}\nu(y). \tag{3.2}$$

Well-known results of the theory of optimal transportation, [Bre87] for $p = 2$, [GM96] for $p < \infty$ and [CDPJ08], for $p = \infty$, provide that when μ is absolutely continuous with respect to $\mathcal{L}^d$ then there exists an optimal transport map between μ and ν, namely $T : \Omega \to \Omega$ such that $T_{\#}\mu = \nu$ and

$$\begin{aligned} d_p^p(\mu,\nu) &:= \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\Omega\times\Omega} |x-y|^p \mathrm{d}\gamma(x,y) = \int_\Omega |T(x)-x|^p \mathrm{d}\mu(x) \quad \text{when } p<\infty, \\ d_\infty(\mu,\nu) &:= \inf_{\gamma\in\Gamma(\mu,\nu)} \operatorname{ess\,sup}_{(x,y)\in\mathrm{supp}(\gamma)} |x-y| = \operatorname{ess\,sup}_{x\in\mathrm{supp}(\mu)} |T(x)-x| \quad \text{when } p=\infty. \end{aligned} \tag{3.3}$$

In particular the transport plan induced by T, namely $\pi := (Id, T)_{\#}\mu$, is optimal. The distance $d_p(\mu,\nu)$ we define in the expressions above is called the p-transportation metric (also referred to as the p-Wasserstein distance).

When considering convergence of a sequence (μ_n, f_n) toward (μ, f), the following sufficient criterion will be useful. We say that a sequence of transportation maps is *stagnating* if

$$T_{n\#}\mu = \mu_n \quad \text{and} \quad \|I - T_n\|^p_{L^p(\mu)} = \int_\Omega |x - T_n(x)|^p \,\mathrm{d}\mu(x) \to 0 \tag{3.4}$$

as $n \to \infty$. To show TL^p convergence it thus suffices to find a stagnating sequence of transportation maps such that $\int |f(x) - f_n(T_n(x))|^p \mathrm{d}\mu(x)$ converges to zero as $n \to \infty$.

We now introduce the new results that allow us to obtain the optimal scaling of ε_n for Γ-convergence in 2D. Namely while $d_\infty(\mu, \mu_n) \sim \frac{(\ln n)^{3/4}}{\sqrt{n}}$ when $d = 2$ we introduce an auxiliary measure $\tilde{\mu}_n$ which is absolutely continuous with respect to Lebesgue measure and satisfies both that $d_\infty(\tilde{\mu}_n, \mu_n) \sim \sqrt{\frac{(\ln n)}{n}}$ when $d = 2$ and that its density with respect to measure μ uniformly converges to 1. These two facts are enough to pass to the limit in the functionals we consider, and many others (like total variation or dirichlet energy). Lemma 3.2 is a technical result needed to transfer the TL^p convergence to the desired measures.

Lemma 3.1. *Let μ be a probability measure with continuous density ρ, supported on $\overline{\Omega}$, where Ω is a bounded open set with Lipschitz boundary in $\mathbb{R}^d$, $d \geqslant 2$ and which satisfies the assumption (2.7). Let ε_n be a sequence of positive numbers converging to zero and satisfying (2.12). Let $\{x_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random points chosen according to the density ρ, and let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Then there exists a sequence of probability measures $\tilde{\mu}_n$ which are absolutely continuous with respect to the measure μ and satisfy*

(*a*) *Almost surely $\ell_n := d_\infty(\mu_n, \tilde{\mu}_n) \ll \varepsilon_n$*
(*b*) *As $n \to \infty$, $\frac{\mathrm{d}\tilde{\mu}_n}{\mathrm{d}\mu}$ almost surely converges to 1 uniformly on $\overline{\Omega}$.*

Proof. Let us first consider the case that $\Omega = [0, 1]^d$. Given a sequence ε_n satisfying the assumptions of the lemma let $\{b_n\}_{n\in\mathbb{N}}$ and $\{c_n\}_{n\in\mathbb{N}}$ be increasing sequences of positive numbers such that $\left(b_n \frac{\ln n}{n}\right)^{-1/d}$ is integer and

$$b_n \to \infty, \quad \varepsilon_n \gg \left(b_n \frac{\ln n}{n}\right)^{1/d} \quad \text{and} \quad c_n^2 \gg b_n \gg c_n \quad \text{as } n \to \infty. \tag{3.5}$$

Let, for $n \geqslant 2$,

$$\delta^d = b_n \frac{\ln n}{n} \quad \text{and} \quad t = c_n \frac{\ln n}{n}.$$

Note that they both t and δ depend on n but we decided to omit the index n to lighten the notation. We divide $[0,1]^d$ into $m = \left(b_n \frac{\ln n}{n}\right)^{-1}$ disjoint cubes $K_j, j = 1, \ldots, m$ with side length δ. Note that the probability p_j that a point x_i is in the box K_j is equal to $\mu(K_j)$ and that

$$c\delta^d \leqslant p_j \leqslant C\delta^d.$$

Bernstein's inequality [Ber24] gives

$$\begin{aligned} P\left(|\mu_n(K_j) - p_j| \geqslant t\right) &< 2\,\exp\left(-\frac{\frac{1}{2}nt^2}{p_j(1-p_j) + \frac{1}{3}t}\right) < 2\,\exp\left(-\frac{nt^2}{3p_j}\right) \\ &\leqslant 2\,\exp\left(-\frac{c_n^2 \ln n}{3Cb_n}\right) = 2n^{-\frac{c_n^2}{3Cb_n}}. \end{aligned} \tag{3.6}$$

It follows, by union bound, that the probability that in all boxes $|\mu_n(K_j) - p_j| < t$ satisfies

$$P(\{(\forall j = 1, \ldots, m)\, |\mu_n(K_j) - p_j| < t) \geqslant 1 - m2n^{-\frac{c_n^2}{3Cb_n}} \geqslant 1 - n^{-2}, \tag{3.7}$$

for all n large enough. By Borel–Cantelli lemma we conclude that almost surely for n sufficiently large for all boxes it holds that $|\mu_n(K_j) - p_j| < t$.

Define the measure $\tilde{\mu}_n$ as follows:

$$d\tilde{\mu}_n = \sum_{j=1}^{m} \mathbb{1}_{K_j} \frac{\mu_n(K_j)}{\delta^d}\, \mathrm{d}x.$$

Since $\tilde{\mu}_n(K_j) = \mu_n(K_j)$, the distance $d_\infty(\tilde{\mu}_n, \mu_n)$ is at most the diameter of the boxes, namely

$$d_\infty(\tilde{\mu}_n, \mu_n) \leqslant \sqrt{d}\,\delta.$$

For large n and arbitrary $x \in \Omega$ let K_j be such that $x \in K_j$. Using (3.6) we obtain

$$\begin{aligned} \left|\frac{d\tilde{\mu}_n}{d\mu}(x) - 1\right| &\leqslant \left|\frac{\frac{\mu_n(K_j)}{\delta^d} - \rho(x)}{\rho(x)}\right| \\ &\leqslant \frac{|\mu_n(K_j) - \mu(K_j)|}{\delta^d\,\rho(x)} + \frac{\int_{K_j} |\rho(z) - \rho(x)|\,\mathrm{d}z}{\delta^d\,\rho(x)} \\ &\leqslant \frac{t}{c\delta^d} + \frac{1}{c}\omega\left(\sqrt{d}\,\delta\right) \leqslant \frac{c_n}{cb_n} + \frac{1}{c}\omega\left(\sqrt{d}\,\delta\right), \end{aligned}$$

where ω is the modulus of continuity of ρ. The uniform convergence follows since the terms on the right-hand side converge to zero.

Extending the argument to general Ω with smooth boundary is straightforward using the partition procedure detailed in section 3 of [GTS15]. To general case of Ω with Lipschitz boundary

can be reduced to domains with smooth boundary using the result of Ball and Zarnescu [BZ17], as was done is the step 2 of the proof of theorem 1.2 in [GTS15]. □

Lemma 3.2. *Let μ be a probability measure with density ρ supported on a compact set in $\mathbb{R}^d$. Let $\tilde{\mu}_n$ be a sequence of probability measures which are absolutely continuous with respect to μ such that*

$$\frac{d\tilde{\mu}_n}{d\mu} \to 1 \quad \textit{uniformly on the support of } \mu.$$

Assume $f_n \to f$ in $L^p(\mu)$ as $n \to \infty$. Then

$$(\tilde{\mu}_n, f_n) \to (\mu, f) \quad \textit{in } TL^p \textit{ as } n \to \infty. \tag{3.8}$$

Proof. From the assumption that $\frac{d\tilde{\mu}_n}{d\mu}$ uniformly converges to 1 follows that the Lévy–Prokhorov metric between $\tilde{\mu}_n$ and μ converges to zero. Since the Lévy–Prokhorov and the p-transportation metric, d_p, defined by (3.3), both metrize the weak convergence of measures on compact sets we conclude that $d_p(\mu, \tilde{\mu}_n) \to 0$ as $n \to \infty$. Thus there exists a sequence of transportation plans $\pi_n \in \Pi(\mu, \tilde{\mu}_n)$ such that

$$\int |x - y|^p \mathrm{d}\pi_n(x, y) \to 0 \quad \text{as } n \to \infty.$$

Since Lipschitz continuous functions are dense in $L^p(\mu)$ there exists a sequence of Lipschitz continuous functions g_m which converges to f in $L^p(\mu)$. Let ρ_n be the Lebesgue density of $\tilde{\mu}_n$. Since $\frac{\rho_n}{\rho}$ uniformly converges to 1, there exists n_1 such that for all $n \geqslant n_1$, $\frac{1}{2} \leqslant \frac{\rho_n}{\rho} \leqslant 2$. For $n \geqslant n_1$

$$\int |f_n(y) - f(x)|^p \mathrm{d}\pi_n(x, y) \leqslant 2^p \left(\int |f_n(y) - f(y)|^p \mathrm{d}\tilde{\mu}_n(y) + \int |f(y) - f(x)|^p \mathrm{d}\pi_n(x, y) \right).$$

We estimate the terms separately:

$$\int |f_n(y) - f(y)|^p \mathrm{d}\tilde{\mu}_n(y) \leqslant 2 \int |f_n(y) - f(y)|^p \mathrm{d}\mu(y),$$

which converges to zero as $n \to \infty$.

$$\begin{aligned} \int |f(y) - f(x)|^p \mathrm{d}\pi_n(x, y) &\lesssim \int |f(y) - g_m(y)|^p + |g_m(y) - g_m(x)|^p + |g_m(x) - f(x)|^p \mathrm{d}\pi_n(x, y) \\ &\lesssim \|f - g_m\|^p_{L^p(\mu)} + \int \mathrm{Lip}(g_m)|x - y|^p \pi_n(x, y). \end{aligned}$$

We observe that the right hand can be made arbitrarily small by taking m large enough and then taking n sufficiently large. Consequently $\int |f_n(y) - f(x)|^p \mathrm{d}\pi_n(x, y)$ converges to zero as $n \to \infty$, which implies that (3.8) holds. □

4. Proof of the Γ-convergence (theorem 2.1)

We prove theorem 2.1 by separately proving the Γ-liminf bound and by building a recovery sequence. The proof of Γ-liminf bound relies on slicing (i.e. one dimensional decomposition) as used by Gobbino [Gob98] and the techniques of [GTS16] to deal with randomness of the sample. However since the spatial coordinates appear also in a denominator within the

functional (2.9), the way [GTS16] deals with space is not precise enough and new ideas are needed to overcome this challenge. These are discussed in section 4.3. In the subsection below we introduce notation and present the one-dimensional slicing of the continuum Mumford–Shah functional.

4.1. One dimensional slicing

We repeatedly use the following computation, which we sketch for a generic function f.

Let $f : \Omega \times \Omega \to \mathbb{R}$ be an integrable function. Then, for $\varepsilon > 0$ it holds that

$$\begin{aligned}\int_{\Omega\times\Omega} f(x,y)\mathrm{d}x\mathrm{d}y &= \varepsilon^d \int_\Omega \mathrm{d}x \int_{\frac{\Omega-x}{\varepsilon}} f(x, x+\varepsilon\xi)\mathrm{d}\xi \\ &= \varepsilon^d \int_{\mathbb{R}^d} \mathrm{d}\xi \int_{\Omega\cap(\Omega-\varepsilon\xi)} f(x, x+\varepsilon\xi)\mathrm{d}x \end{aligned} \tag{4.1}$$

where we exploited the identity

$$\mathbb{1}_\Omega(x)\mathbb{1}_{(\Omega-x)/\varepsilon}(\xi) = \mathbb{1}_{\mathbb{R}^d}(\xi)\mathbb{1}_{\Omega\cap(\Omega-\varepsilon\xi)}(x)$$

and then Fubini's theorem (here $\mathbb{1}_E(x)$ stands for the characteristic function of the set E and takes value 1 for $x \in E$ and 0 otherwise). Given $A \subseteq \mathbb{R}^d$ and $\xi \in \mathbb{R}^d$ we define for $z \in \xi^\perp$ the one dimensional slice

$$[A]_z := \{t \in \mathbb{R} : z + t\xi/|\xi| \in A\}. \tag{4.2}$$

Above we are omitting the dependence on ξ, since it will be clear from the context. We proceed to consider one dimensional slices as follows

$$\begin{aligned}\int_{\Omega\times\Omega} f(x,y)\mathrm{d}x\mathrm{d}y &= \varepsilon^d \int_{\mathbb{R}^d} \mathrm{d}\xi \int_{\Omega\cap(\Omega-\varepsilon\xi)} f(x, x+\varepsilon\xi)\mathrm{d}x \\ &= \varepsilon^d \int_{\mathbb{R}^d} \mathrm{d}\xi \int_{\xi^\perp} \mathrm{d}z \int_{[\Omega\cap(\Omega-\varepsilon\xi)]_z} f\left(z + t\frac{\xi}{|\xi|}, z + (t+\varepsilon|\xi|)\frac{\xi}{|\xi|}\right)\mathrm{d}t. \end{aligned} \tag{4.3}$$

When using one dimensional decompositions we will make several uses of the following notation. Given $g : \mathbb{R}^d \to \mathbb{R}$, $\xi \in \mathbb{R}^d$ and $z \in \xi^\perp$ we define

$$g_\xi(t;z) := g\left(z + t\frac{\xi}{|\xi|}\right). \tag{4.4}$$

We now state and prove two technical lemmas that we use in the sequel.

Lemma 4.1. *Let $u \in \mathrm{SBV}(\Omega)^p$. Then*

$$\begin{aligned} MS_{\eta,\zeta}(u;\rho) = \zeta'(0)&\int_{\mathbb{R}^d} |\xi|^{p-q}\eta(|\xi|)\mathrm{d}\xi \int_{\xi^\perp} \mathrm{d}z \int_{[\Omega]_z} \left|u'_\xi(t;z)\right|^p \rho_\xi(t;z)^2\mathrm{d}t \\ &+ \Theta\int_{\mathbb{R}^d} |\xi|\eta(|\xi|)\mathrm{d}\xi \int_{\xi^\perp} \mathrm{d}z \int_{S_{u_\xi(\cdot;z)}} \rho_\xi(t;z)^2\mathrm{d}\mathcal{H}^0(t) \end{aligned} \tag{4.5}$$

$$
\begin{aligned}
=&\zeta'(0)\int_{\mathbb{R}^d}|\xi|^{p-q}\eta(|\xi|)\mathrm{d}\xi\int_\Omega\left|\nabla u(x)\cdot\frac{\xi}{|\xi|}\right|^p\rho(x)^2\mathrm{d}x\\
&+\Theta\int_{\mathbb{R}^d}|\xi|\eta(|\xi|)\mathrm{d}\xi\int_{S_u}\left|N_u(y)\cdot\frac{\xi}{|\xi|}\right|\rho(y)^2\mathrm{d}\mathcal{H}^{d-1}(y).
\end{aligned}
\tag{4.6}
$$

where $N_u(y)$ is any vector field normal to S_u.

Proof. We can rewrite the right-hand side of (4.5) as follows

$$
\begin{aligned}
&\zeta'(0)\int_{\mathbb{R}^d}|\xi|^{p-q}\eta(|\xi|)\mathrm{d}\xi\int_{\xi^\perp}\mathrm{d}z\int_{[\Omega]_z}\left|u'_\xi(t;z)\right|^p\rho_\xi(t;z)^2\mathrm{d}t\\
&\quad+\Theta\int_{\mathbb{R}^d}|\xi|\eta(|\xi|)\mathrm{d}\xi\int_{\xi^\perp}\mathrm{d}z\int_{S_{u_\xi(\cdot;z)}}\rho_\xi(t;z)^2\mathrm{d}\mathcal{H}^0(t)\\
&=\zeta'(0)\int_{\mathbb{R}^d}|\xi|^{p-q}\eta(|\xi|)\mathrm{d}\xi\int_{\xi^\perp}\mathrm{d}z\int_{[\Omega]_z}\left|\nabla u\left(z+t\frac{\xi}{|\xi|}\right)\cdot\frac{\xi}{|\xi|}\right|^p\rho\left(z+t\frac{\xi}{|\xi|}\right)^2\mathrm{d}t\\
&\quad+\Theta\int_{\mathbb{R}^d}|\xi|\eta(|\xi|)\mathrm{d}\xi\int_{\xi^\perp}\mathrm{d}z\int_{S_{u_\xi(\cdot;z)}}\rho\left(z+t\frac{\xi}{|\xi|}\right)^2\mathrm{d}\mathcal{H}^0(t)\\
&=\zeta'(0)\int_{\mathbb{R}^d}|\xi|^{p-q}\eta(|\xi|)\mathrm{d}\xi\int_\Omega\left|\nabla u(x)\cdot\frac{\xi}{|\xi|}\right|^p\rho(x)^2\mathrm{d}x\\
&\quad+\Theta\int_{\mathbb{R}^d}|\xi|\eta(|\xi|)\mathrm{d}\xi\int_{S_u}\left|N_u(y)\cdot\frac{\xi}{|\xi|}\right|\rho(y)^2\mathrm{d}\mathcal{H}^{d-1}(y).
\end{aligned}
$$

The last equality above follows from the Coarea formula, using a well known relation between the one-dimensional slices of an SBV function and the total length of its jump set (see for instance [Amb89]). Notice that

$$
\begin{aligned}
\int_{\mathbb{R}^d}|\xi|\eta(|\xi|)\left|N_u(y)\cdot\frac{\xi}{|\xi|}\right|\mathrm{d}\xi&=\int_0^\infty t\mathrm{d}t\int_{\partial B_t}\eta(t)\left|N_u(y)\cdot\frac{\xi}{t}\right|\mathrm{d}\mathcal{H}^{d-1}(\xi)\\
&=\int_0^\infty t^d\eta(t)\mathrm{d}t\int_{\partial B_1}|N_u(y)\cdot v|\,\mathrm{d}\mathcal{H}^{d-1}(v)\\
&=2\omega_{d-1}\int_0^\infty t^d\eta(t)\mathrm{d}t=\sigma_\eta
\end{aligned}
$$

where we have exploited the relation

$$
\int_{\partial B_1}|N_u(y)\cdot v|\,\mathrm{d}\mathcal{H}^{d-1}(v)=2\omega_{d-1}.
$$

Moreover

$$
\begin{aligned}
&\int_{\mathbb{R}^d}|\xi|^{p-q}\eta(|\xi|)\left|\nabla u(x)\cdot\frac{\xi}{|\xi|}\right|^p\mathrm{d}\xi\\
&\quad=\int_0^\infty t^{p-q+d-1}\eta(t)\mathrm{d}t\int_{\partial B_1}|\nabla u(x)\cdot v|^p\mathrm{d}\mathcal{H}^{d-1}(v)
\end{aligned}
$$

$$= 2|\nabla u(x)|^p \omega_{d-1} \frac{\Gamma(p/2+1/2)\Gamma(d/2+1/2)}{\Gamma(p/2+d/2)} \int_0^\infty t^{p-q+d-1}\eta(t)\mathrm{d}t$$
$$= |\nabla u(x)|^p \vartheta_\eta(p,q)$$

where we made use of

$$\int_{\partial B_1} |w\cdot v|^p \mathrm{d}\mathcal{H}^{d-1}(v) = 2|w|^p \omega_{d-1} \frac{\Gamma(p/2+1/2)\Gamma(d/2+1/2)}{\Gamma(p/2+d/2)}.$$

In particular

$$\begin{aligned}
&\zeta'(0)\int_{\mathbb{R}^d} |\xi|^{p-q}\eta(|\xi|)\mathrm{d}\xi \int_{\xi^\perp} \mathrm{d}z \int_{[\Omega]_z} \left|u'_\xi(t;z)\right|^p \rho_\xi(t;z)^2 \mathrm{d}x \\
&\quad + \Theta \int_{\mathbb{R}^d} |\xi|\eta(|\xi|)\mathrm{d}\xi \int_{\xi^\perp} \mathrm{d}z \int_{S_{u_\xi(\cdot;z)}} \rho_\xi(t;z)^2 \mathrm{d}\mathcal{H}^0(t) \\
&= \zeta'(0)\int_{\mathbb{R}^d} |\xi|^{p-q}\eta(|\xi|)\mathrm{d}\xi \int_\Omega \left|\nabla u(x)\cdot \frac{\xi}{|\xi|}\right|^p \rho(x)^2 \mathrm{d}x \\
&\quad + \Theta \int_{\mathbb{R}^d} |\xi|\eta(|\xi|)\mathrm{d}\xi \int_{S_u} \left|N_u(y)\cdot \frac{\xi}{|\xi|}\right| \rho(y)^2 \mathrm{d}\mathcal{H}^{d-1}(y) \\
&= \vartheta_\eta(p,q)\zeta'(0)\int_\Omega |\nabla u(x)|^p \rho(x)^2 \mathrm{d}x \\
&\quad + \sigma_\eta \Theta \int_{S_u} \rho(y)^2 \mathrm{d}\mathcal{H}^{d-1}(y) = MS_{\eta,\zeta}(u;\rho).
\end{aligned}$$

□

4.2. Auxiliary functionals

We introduce two auxiliary functionals: $\overline{\mathcal{GMS}}$ and $\mathcal{GAMS}$. The first one is motivated by the calculation below and allows us to switch the auxiliary measure $\tilde{\mu}_n$, constructed in lemma 3.1 by the measure μ. The functionals $\mathcal{GAMS}$ moves a step further towards to local, limiting, functional by replacing the integral over the product measure by one with the weight $\rho(x)^2$. We will first establish the lim inf and lim sup bounds on an auxiliary energy $\mathcal{GAMS}$, and then, by exploiting lemma 4.2 apply these bounds to $\overline{\mathcal{GMS}}$. Thanks to (4.7) we can then prove the statement of theorem 2.1 for $\mathcal{GMS}$.

Consider the setting of theorem 2.1 and let $\tilde{\mu}_n$ be the measures constructed in lemma 3.1. Let $T_n : \Omega \to \Omega$ be the d_∞ optimal transport from $\tilde{\mu}_n$ to μ_n. Let $\ell_n = d_\infty(\tilde{\mu}_n, \mu_n) = \|T - Id\|_{L^\infty}$. By exploiting the change of variable (3.2), we can rewrite the Mumford–Shah functional on the point clouds in the following integral form:

$$\begin{aligned}
\mathcal{GMS}_{\varepsilon,n}(u) &:= \frac{1}{\varepsilon}\frac{1}{n^2}\sum_{i,j=1}^n \zeta\left(\varepsilon^{1-p+q}\frac{|u(x_i)-u(x_j)|^p}{|x_i-x_j|^q}\right)\eta_\varepsilon(|x_i-x_j|) \\
&= \frac{1}{\varepsilon}\int_{\Omega\times\Omega} \zeta\left(\varepsilon^{1-p+q}\frac{|u(x)-u(y)|^p}{|x-y|^q}\right)\eta_\varepsilon(|x-y|)\mathrm{d}\mu_n(x)\mathrm{d}\mu_n(y)
\end{aligned}$$

$$= \frac{1}{\varepsilon}\int_{\Omega\times\Omega}\zeta\left(\varepsilon^{1-p+q}\frac{|u(x)-u(y)|^p}{|x-y|^q}\right)\eta_\varepsilon(|x-y|)\mathrm{d}(T_{n\#}\tilde{\mu}_n)(x)\mathrm{d}(T_{n\#}\tilde{\mu}_n)(y)$$

$$= \frac{1}{\varepsilon}\int_\Omega\int_\Omega\zeta\left(\varepsilon^{1-p+q}\frac{|u(T_n(x))-u(T_n(y))|^p}{|T_n(x)-T_n(y)|^q}\right)\eta_\varepsilon(|T_n(x)-T_n(y)|)\mathrm{d}\tilde{\mu}_n(x)\mathrm{d}\tilde{\mu}_n(y)$$

$$\leqslant \sup_{x\in\Omega}\left|\frac{\rho_n}{\rho}\right|^2\frac{1}{\varepsilon}\int_\Omega\int_\Omega\zeta\left(\varepsilon^{1-p+q}\frac{|u(T_n(x))-u(T_n(y))|^p}{|T_n(x)-T_n(y)|^q}\right)\eta_\varepsilon(|T_n(x)-T_n(y)|)\mathrm{d}\rho(x)\mathrm{d}\rho(y).$$

Analogously

$$\mathcal{GMS}_{\varepsilon,n}(u)\geqslant\inf_{x\in\Omega}\left|\frac{\rho_n}{\rho}\right|^2\frac{1}{\varepsilon}\int_\Omega\int_\Omega\zeta\left(\varepsilon^{1-p+q}\frac{|u(T_n(x))-u(T_n(y))|^p}{|T_n(x)-T_n(y)|^q}\right)\eta_\varepsilon(|T_n(x)-T_n(y)|)\mathrm{d}\rho(x)\mathrm{d}\rho(y).$$

In the light of the above computation, we consider three different functionals, which, as we show, share the same Γ-limit. We consider

$$\mathcal{GMS}_{\varepsilon,n}(u) := \frac{1}{\varepsilon}\frac{1}{n^2}\sum_{i,j=1}^n\zeta\left(\varepsilon^{1-p+q}\frac{|u(x_i)-u(x_j)|^p}{|x_i-x_j|^q}\right)\eta_\varepsilon(|x_i-x_j|)$$

$$\overline{\mathcal{GMS}}_{\varepsilon,n}(u) := \frac{1}{\varepsilon}\int_\Omega\int_\Omega\zeta\left(\frac{|u(T_n(x))-u(T_n(y))|^p}{\varepsilon^{p-q-1}|T_n(x)-T_n(y)|^q}\right)\eta_\varepsilon(|T_n(x)-T_n(y)|)\mathrm{d}\rho(x)\mathrm{d}\rho(y)$$

$$\mathcal{GAMS}_{\varepsilon,n}(u) := \frac{1}{\varepsilon}\int_\Omega\int_\Omega\zeta\left(\frac{|u(T_n(x))-u(T_n(y))|^p}{\varepsilon^{p-q-1}|T_n(x)-T_n(y)|^q}\right)\eta_\varepsilon(|T_n(x)-T_n(y)|)\rho(x)^2\mathrm{d}x\mathrm{d}y.$$

Notice that since $\frac{\rho_n}{\rho}$ uniformly converges to 1 (by lemma 3.1)

$$\Gamma\text{-}\lim_{n\to+\infty}\mathcal{GMS}_{\varepsilon_n,n} = \Gamma\text{-}\lim_{n\to+\infty}\overline{\mathcal{GMS}}_{\varepsilon_n,n}. \tag{4.7}$$

The next lemma shows that, in case of a compactly supported kernel η, the auxiliary energy $\mathcal{GAMS}$ is asymptotically equivalent to $\overline{\mathcal{GMS}}$.

Lemma 4.2. *Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be any sequence satisfying (2.12). Le $\{u_n\}_{n\in\mathbb{N}}$ be a sequence in $L^1(\Omega;\mu_n)$ and assume that the kernel η is compactly supported. Then*

$$\liminf_{n\to\infty}\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = \liminf_{n\to\infty}\mathcal{GAMS}_{\varepsilon_n,n}(u_n)$$

$$\limsup_{n\to\infty}\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = \limsup_{n\to\infty}\mathcal{GAMS}_{\varepsilon_n,n}(u_n)$$

Proof. Let T_n be the transport maps as above and let

$$\mathcal{R}_n(u_n) := \frac{1}{\varepsilon_n}\int_{\Omega\times\Omega}\zeta\left(\frac{|u_n(T_n(x))-u_n(T_n(y))|^p}{\varepsilon_n^{p-q-1}|T_n(x)-T_n(y)|^q}\right)\eta_{\varepsilon_n}(|T_n(x)-T_n(y)|)[\rho(y)-\rho(x)]\rho(x)\mathrm{d}x\mathrm{d}y.$$

Notice that

$$\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = \mathcal{GAMS}_{\varepsilon_n,n}(u_n) + \mathcal{R}_n(u_n).$$

Moreover, due to the properties of ρ, we can find a constant $L>1$ such that

$$\frac{1}{L}\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u)\leqslant\mathcal{GAMS}_{\varepsilon_n,n}(u)\leqslant L\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u)$$

for all $u \in L^1(\Omega;\mu_n)$ and for all $n \in \mathbb{N}$. In particular

$$\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = 0 \;\Leftrightarrow\; \mathcal{GAMS}_{\varepsilon_n,n}(u_n) = 0$$
$$\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = +\infty \;\Leftrightarrow\; \mathcal{GAMS}_{\varepsilon_n,n}(u_n) = +\infty.$$

For the purpose of our proof we can hence restrict to

$$0 < \mathcal{GAMS}_{\varepsilon_n,n}(u_n) < +\infty \quad \text{for all}\, n \in \mathbb{N}.$$

since otherwise the seek equality trivially hold. With this assumption in mind we consider the finite ratio $\mathcal{D}_n := \mathcal{R}_n(u_n)/\mathcal{GAMS}_{\varepsilon_n,n}(u_n)$ so that

$$\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = \mathcal{GAMS}_{\varepsilon_n,n}(u_n)(1+\mathcal{D}_n). \tag{4.8}$$

Since η is compactly supported, and since

$$|T_n(x) - T_n(y)| \geqslant |x-y| - 2\ell_n$$

we have that $\eta_{\varepsilon_n}(|T_n(x) - T_n(x)|) = 0$ for all x, y such that $|x-y| \geqslant M\varepsilon_n$ for some $M \in \mathbb{R}$. Moreover if $|x-y| < M\varepsilon_n$ we have $|\rho(x) - \rho(y)| \leqslant \mathrm{Lip}(\rho)M\varepsilon_n$. Thus, since ρ is bounded from below, we get

$$\mathcal{R}_n(u_n) \leqslant \frac{\mathrm{Lip}(\rho)M\varepsilon_n}{c}\mathcal{GAMS}_{\varepsilon_n,n}(u_n).$$

Hence

$$\lim_{n\to\infty} \mathcal{D}_n = 0$$

that, combined with (4.8) completes the proof. □

4.3. The liminf inequality

This subsection is devoted to the proof of the liminf inequality claim of the Γ-convergence of theorem 2.1:

Proposition 4.3. *Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be any sequence satisfying (2.12). Let $u_n \in L^1(\Omega;\mu_n)$, $u \in L^1(\Omega)$ such that $(\mu_n, u_n) \to (\mu, u)$ in TL^1. Then*

$$\liminf_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u_n) \geqslant MS_{\eta,\zeta}(u;\rho).$$

The proof relies on the following result, which can be obtained following ideas of [GM01, corollary 3.3]. Its proof is provided in the appendix A.

Lemma 4.4. *Let $A \subset \mathbb{R}$ be a finite union of intervals and $f \in C^1(A) \cap C^0(\overline{A})$ be such that*

$$0 < c \leqslant \min_{x\in\overline{A}}\{f(x)\} \leqslant \max_{x\in\overline{A}}\{f(x)\} \leqslant C < \infty.$$

Define for every $\xi \in \mathbb{R}^d$ the one-dimensional functional

$$E_\delta(u;\xi,A) := \frac{1}{\delta}\int_A \zeta\left(\frac{|u(x+\delta|\xi|) - u(x)|^p}{\delta^{p-1}|\xi|^q}\right) f(x)\mathrm{d}x \tag{4.9}$$

and

$$E(u;\xi,A) := \zeta'(0)|\xi|^{p-q}\int_A |u'(x)|^p f\,\mathrm{d}x + \Theta|\xi|\int_{S_u\cap A} f(y)\mathrm{d}\mathcal{H}^0(y) \tag{4.10}$$

for $u \in \mathrm{SBV}^p(\mathbb{R})$. *Then*

$$\liminf_{\delta\to 0} E_\delta(u_\delta;\xi,A) \geqslant E(u;\xi,A) \tag{4.11}$$

for all $\{u_\delta\}_{\delta>0} \subset L^1(A)$, $u \in \mathrm{SBV}^p(A)$ *such that* $u_\delta \to u$ *in* L^1.

We also need the following lemma.

Lemma 4.5. *Let* $u_n : \{x_1,\dots,x_n\} \to \mathbb{R}$ *be a sequence of function* $L^p(\mu_n)$ *such that* $(\mu_n,u_n)\to(\mu,u)$ *in* TL^1. *Let* $\tilde{\mu}_n$ *be the sequence of measures provided by lemma* 3.1 *and consider* $T_n : \mathbb{R}^d \to \mathbb{R}^d$ *to be the associated transport maps between* $\tilde{\mu}_n$ *and* μ_n. *Fix* $\xi \in \mathbb{R}^d$, $z \in \xi^\perp$ *and define the sectional functions*

$$t \mapsto \tilde{u}_{n,\xi}(t;z) := u_n \circ T_n\left(z + t\frac{\xi}{|\xi|}\right).$$

Then up to extracting a subsequence (*not relabeled*), *for* $\mathcal{H}^{d-1}$*-a.e.* $z \in \xi^\perp$ *it holds*

$$\tilde{u}_{\xi,n}(\cdot;z) \to u_\xi(\cdot;z) \quad \text{in } L^1(\mathbb{R};\rho_\xi(\cdot;z)\mathcal{L}^1)$$

Proof. Thanks to the TL^1 convergence of the sequence (μ_n,u_n) and to the properties of our measures $\tilde{\mu}_n$ given by lemma 3.1 we can infer $(\tilde{\mu}_n, u_n\circ T_n)\to(\mu,u)$ in TL^1. In particular thanks also to [GTS16, assertion 5, proposition 3.12]) we know that

$$\lim_{n\to\infty}\int_{\mathbb{R}^d}|u_n\circ T_n(x)-u(x)|\rho(x)\mathrm{d}x = 0. \tag{4.12}$$

We can rearrange and estimate the above integral as

$$\begin{aligned}\int_{\mathbb{R}^d}|u_n\circ T_n(x)-u(x)|\rho(x)\mathrm{d}x &= \int_{-\infty}^{\infty}\mathrm{d}t\int_{\left\{\frac{x\cdot\xi}{|\xi|}=t\right\}}|u_n\circ T_n(x)-u(x)|\rho(x)\mathrm{d}\mathcal{H}^{n-1}(x)\\ &= \int_{-\infty}^{\infty}\mathrm{d}t\int_{\xi^\perp}|\tilde{u}_{\xi,n}(t;z)-u_\xi(t;z)|\rho_\xi(t;z)\mathrm{d}\mathcal{H}^{n-1}(z)\\ &= \int_{\xi^\perp}\mathrm{d}\mathcal{H}^{n-1}(z)\int_{-\infty}^{\infty}|\tilde{u}_{\xi,n}(t;z)-u_\xi(t;z)|\rho_\xi(t;z)\mathrm{d}t.\end{aligned}$$

Using (4.12), this goes to zero as $n\to\infty$, so that up to a subsequence the term under the integral goes to zero for a.e. $z\perp\xi$. □

Remark 4.6. The strategy adopted to prove the lim inf inequality is based on the integral form of the functional $\mathcal{GMS}$ (more precisely on its asymptotic counterpart $\mathcal{GAMS}$). Substantially we exploit such an integral form and the monotonicity of η,ζ to compare our energy with the energy

$$\int_{\mathbb{R}^d}\eta(|\xi|)\mathrm{d}\xi\int_{\xi^\perp}E_{\varepsilon_n}(\tilde{u}_{\xi,n}(\cdot;z);\xi;\Omega)\mathrm{d}\mathcal{H}^{d-1}(z)$$

in order to apply lemmas 4.5 and 4.4. To achieve this goal we need to compare $|T_n(x) - T_n(y)|$ and $|x - y|$, as well as $\frac{1}{|T_n(x)-T_n(y)|}$ and $\frac{1}{|x-y|}$. Notice that $|T_n(x) - T_n(y)| \approx |x - y| \pm 2\ell_n$ and in particular due to the fact that $\ell_n \ll \varepsilon_n$ it is not difficult to compare $|T_n(x) - T_n(y)|$ to $|x - y|$ for couples such that $|x - y| \approx \varepsilon_n$. The problems arise for all those points x, y that lie very close to one another. The strategy we adopt to overcome this difficulty is to create a hole of fixed size r around the origin in the kernel η. That is to replace η by

$$\eta^r(t) \coloneqq \eta(t)(1 - \mathbb{1}_{[0,r)}(t))$$

and thus to neglect all the, small, contributions we cannot compare and then progressively recover the full energy by shrinking the hole in a limit process at the end of our proof. In the sequel we will often write $\mathcal{GMS}_{\varepsilon_n,n}(u_n; r), \mathcal{GAMS}_{\varepsilon_n,n}(u_n; r)$ to denote the corresponding energies having η^r in place of η.

Proof of proposition 4.3. We can assume, without loss of generality, that

$$\sup_{n\in\mathbb{N}} \{\mathcal{GMS}_{\varepsilon_n,n}(u_n)\} < \infty. \tag{4.13}$$

Notice that it is enough to assume that η is compactly supported. Indeed we can always replace η with $\eta_R \coloneqq \eta(t)\mathbb{1}_{[0,R)}(t)$ and notice that, by meaning of lemma 4.2 and (4.7) we have

$$\liminf_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u_n) \geqslant \liminf_{n\to\infty} \overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) \geqslant \liminf_{n\to\infty} \mathcal{GAMS}_{\varepsilon_n,n}(u_n; \eta_R)$$

where $\mathcal{GAMS}_{\varepsilon_n,n}(u_n; \eta_R)$ denotes the usual energy $\mathcal{GAMS}_{\varepsilon_n,n}$ with η_R in place of η. In particular if we can prove that, for compactly supported kernel, it holds

$$\liminf_{n\to\infty} \mathcal{GAMS}_{\varepsilon_n,n}(u_n; \eta_R) \geqslant MS_{\eta_R,\zeta}(u)$$

then the continuity of the constants in MS allows us to send R to infinity and recover

$$\liminf_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u_n) \geqslant MS_{\eta,\zeta}(u).$$

We thus focus on proving the theorem for a compactly supported kernel η.

With this assumption in mind we invoke again lemma 4.2 and (4.7) to infer

$$\liminf_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u_n) \geqslant \liminf_{n\to\infty} \mathcal{GAMS}_{\varepsilon_n,n}(u_n; r) \quad \text{for all } r > 0. \tag{4.14}$$

For the reader's convenience in what follows we write ε and $\ell \coloneqq \|T_n - Id\|_\infty$ in place of ε_n and ℓ_n by omitting the dependence on n. Thanks to (2.13) we have

$$\frac{|x-y|}{\varepsilon} - \frac{2\ell}{\varepsilon} \leqslant \frac{|T_n(x) - T_n(y)|}{\varepsilon} \leqslant \frac{|x-y|}{\varepsilon} + \frac{2\ell}{\varepsilon}. \tag{4.15}$$

On the set $\{|T_n(x) - T_n(y)| \geqslant r\varepsilon\}$ we thus have

$$\frac{|x-y|}{1 + \frac{2\ell}{\varepsilon r}} \leqslant |T_n(x) - T_n(y)| \leqslant \frac{|x-y|}{1 - \frac{2\ell}{\varepsilon r}}.$$

In particular, since the function $t \to \zeta(a/t^q)\eta^r_\varepsilon(t)$ is non-increasing for all $\varepsilon, r, a \in \mathbb{R}$, we conclude that

$$\mathcal{GAMS}_{\varepsilon,n}(u_n;r) \geqslant \frac{1}{\varepsilon}\int_{\Omega\times\Omega}\zeta\left(\frac{|u_n(T_n(x))-u_n(T_n(y))|^p}{\frac{\varepsilon^{p-q-1}}{\left(1-\frac{2\ell}{\varepsilon r}\right)^q}|x-y|^q}\right)\eta_\varepsilon^r\left(\frac{|x-y|}{1-\frac{2\ell}{\varepsilon r}}\right)\rho(x)^2\mathrm{d}x\mathrm{d}y$$

$$= \frac{1}{\varepsilon}\int_{\Omega\times\Omega}\zeta\left(\frac{|u_n(T_n(x))-u_n(T_n(y))|^p}{\frac{\varepsilon^{p-q-1}}{\left(1-\frac{2\ell}{\varepsilon r}\right)^{q-p+1}}\left(1-\frac{2\ell}{\varepsilon r}\right)^{1-p}|x-y|^q}\right)\eta^r\left(\frac{|x-y|}{\varepsilon\left(1-\frac{2\ell}{\varepsilon r}\right)}\right)\varepsilon^{-d}\rho(x)^2\mathrm{d}x\mathrm{d}y$$

$$\geqslant \frac{1}{\varepsilon}\int_{\Omega\times\Omega}\zeta\left(\frac{|u_n(T_n(x))-u_n(T_n(y))|^p}{\left[\varepsilon\left(1-\frac{2\ell}{\varepsilon r}\right)\right]^{p-q-1}|x-y|^q}\right)\eta^r\left(\frac{|x-y|}{\varepsilon\left(1-\frac{2\ell}{\varepsilon r}\right)}\right)\varepsilon^{-d}\rho(x)^2\mathrm{d}x\mathrm{d}y.$$

By setting

$$\delta = \delta_\varepsilon := \varepsilon\left(1-\frac{2\ell}{\varepsilon r}\right)$$

and

$$\mathcal{G}_{\delta,n}(u_n,r) := \frac{1}{\delta}\int_{\Omega\times\Omega}\zeta\left(\frac{|u_n(T_n(x))-u_n(T_n(y))|^p}{\delta^{p-q-1}|x-y|^q}\right)\eta_\delta^r\left(|x-y|\right)\rho(x)^2\mathrm{d}x\mathrm{d}y$$

$$= \frac{1}{\delta}\int_{\mathbb{R}^d}\eta^r\left(|\xi|\right)\mathrm{d}\xi\int_{\Omega\cap(\Omega-\delta\xi)}\zeta\left(\frac{|u_n(T_n(x+\delta\xi))-u_n(T_n(x))|^p}{\delta^{p-1}|\xi|^q}\right)\rho(x)^2\mathrm{d}x$$

(where we applied the change of variable (4.1)) we obtain

$$\left(1-\frac{2\ell}{\varepsilon r}\right)^{-d-1}\mathcal{GAMS}_{\varepsilon,n}(u;r) \geqslant \mathcal{G}_{\delta,n}(u_n,r). \tag{4.16}$$

Notice that, since $\ell \ll \varepsilon$, we have $\delta = \delta_\varepsilon \to 0$ and $\left(1-\frac{2\ell}{\varepsilon r}\right)^{-d-1} \to 1$. We now focus our attention on the energy $\mathcal{G}_{\delta,n}$ and more precisely on

$$\mathcal{E}_{\delta,n}(u_n;\xi) := \frac{1}{\delta}\int_{\Omega\cap(\Omega-\delta\xi)}\zeta\left(\frac{|u_n(T_n(x+\delta\xi))-u_n(T_n(x))|^p}{\delta^{p-1}|\xi|^q}\right)\rho(x)^2\mathrm{d}x.$$

Clearly

$$\mathcal{G}_{\delta,n}(u_n,r) = \int_{\mathbb{R}^d}\eta^r(|\xi|)\mathcal{E}_{\delta,n}(u_n;\xi)\mathrm{d}\xi. \tag{4.17}$$

Fix ξ and consider an open bounded set with finite perimeter and smooth boundary $A \subset\subset \Omega$ and notice that, for δ small enough, we have

$$A \subset \Omega \cap (\Omega - \delta\xi).$$

In particular

$$\mathcal{E}_{\delta,n}(u_n;\xi) \geqslant \frac{1}{\delta}\int_A\zeta\left(\frac{|u_n(T_n(x+\delta\xi))-u_n(T_n(x))|^p}{\delta^{p-1}|\xi|^q}\right)\rho(x)^2\mathrm{d}x$$

$$= \frac{1}{\delta}\int_{\xi^\perp}\mathrm{d}\mathcal{H}^{n-1}(z)\int_{[A]_z}\zeta\left(\frac{|\tilde{u}_{\xi,n}(t+\delta|\xi|;z)-\tilde{u}_{\xi,n}(t;z)|^p}{\delta^{p-1}|\xi|^q}\right)\rho_\xi(t;z)^2\mathrm{d}t$$

$$= \int_{\xi^\perp}E_\delta(\tilde{u}_{\xi,n}(\cdot;z);\xi,[A]_z)\mathrm{d}\mathcal{H}^{d-1}(z).$$

Thanks to the co-area formula we have that, for $\mathcal{H}^{d-1}$ almost every $z \in \xi^{\perp}$, the set $[A]_z$ must be a finite union of open intervals. Let now n_k be the subsequence achieving

$$\liminf_{n\to+\infty} \mathcal{E}_{\delta,n}(u_n;\xi) = \lim_{k\to+\infty} \mathcal{E}_{\delta,n_k}(u_{n_k};\xi)$$

By applying lemma 4.5 and the boundedness of ρ we conclude that, up to extracting a further subsequence (not relabeled), for $\mathcal{H}^{d-1}$ almost every $z \in \xi^{\perp}$ we have

$$\lim_{k\to+\infty} \int_{[A]_z} |\tilde{u}_{\xi,n_k}(t;z) - u_\xi(t;z)|\mathrm{d}t \to 0.$$

Henceforth, by applying lemma 4.4, for $\mathcal{H}^{d1}$-a.e. $z \in \xi^{\perp}$ we can infer that

$$\liminf_{k\to+\infty} E_\delta(\tilde{u}_{\xi,n_k}(\cdot;z);\xi,[A]_z) \geqslant E(u_\xi(\cdot;z);\xi,[A]_z) \tag{4.18}$$

being E the one dimensional functional defined in (4.10). An application of Fatou's lemma yields then

$$\liminf_{n\to+\infty} \mathcal{E}_{\delta,n}(u_n;\xi) = \lim_{k\to+\infty} \mathcal{E}_{\delta,n_k}(u_{n_k};\xi) \geqslant \int_{\xi^{\perp}} \liminf_{k\to+\infty} E_\delta(\tilde{u}_{\xi,n_k}(\cdot;z);\xi,[A]_z)\mathrm{d}\mathcal{H}^{d-1}(z)$$

which, combined with (4.18), leads to

$$\begin{aligned}\liminf_{n\to\infty} \mathcal{E}_{\delta,n}(u_n;\xi) \geqslant\ & |\xi|^{p-q}\zeta'(0)\int_{\xi^{\perp}}\int_{[A]_z} |u'_\xi(t;z)|^p \rho_\xi(t;z)^2 \mathrm{d}t\mathrm{d}\mathcal{H}^{d-1}(z) \\ & + \Theta|\xi| \int_{\xi^{\perp}}\int_{S_{u_\xi(\cdot;z)}\cap[A]_z} \rho_\xi(t,z)^2 \mathrm{d}\mathcal{H}^0(t)\mathrm{d}\mathcal{H}^{d-1}(z).\end{aligned} \tag{4.19}$$

In particular, from (4.6), (4.19), (4.17) and Fatou's lemma we get

$$\begin{aligned}\liminf_{n\to\infty} \mathcal{G}_{\delta,n}(u_n;r) \geqslant\ & \zeta'(0)\int_{\mathbb{R}^d} |\xi|^{p-q}\eta^r(|\xi|)\mathrm{d}\xi \int_A \left|\nabla u(x)\cdot\frac{\xi}{|\xi|}\right|^p \rho(x)^2\mathrm{d}x \\ & + \Theta\int_{\mathbb{R}^d} |\xi|\eta^r(|\xi|)\mathrm{d}\xi \int_{S_u\cap A} \left|N_u(y)\cdot\frac{\xi}{|\xi|}\right| \rho(y)^2\mathrm{d}\mathcal{H}^{d-1}(y)\end{aligned}$$

which, being holds for all $A \subset\subset \Omega$. Consequently

$$\begin{aligned}\liminf_{n\to\infty} \mathcal{G}_{\delta,n}(u_n;r) \geqslant\ & \zeta'(0)\int_{\mathbb{R}^d} |\xi|^{p-q}\eta^r(|\xi|)\mathrm{d}\xi \int_\Omega \left|\nabla u(x)\cdot\frac{\xi}{|\xi|}\right|^p \rho(x)^2\mathrm{d}x \\ & + \Theta\int_{\mathbb{R}^d} |\xi|\eta^r(|\xi|)\mathrm{d}\xi \int_{S_u} \left|N_u(y)\cdot\frac{\xi}{|\xi|}\right| \rho(y)^2\mathrm{d}\mathcal{H}^{d-1}(y).\end{aligned}$$

Thanks to (4.6), we have that

$$\liminf_{n\to\infty} \mathcal{GAMS}_{\varepsilon_n,n}(u_n;r) \geqslant \liminf_{n\to\infty} \mathcal{G}_{\delta_n,n}(u_n;r) \geqslant MS_{\eta^r,\zeta}(u;\rho).$$

By exploiting once more the continuity in r of the quantities $\vartheta_{\eta^r}(p,q)$, σ_{η^r} we can take the limit as $r \to 0$. This, considering also (4.14), completes the proof. □

4.4. The limsup inequality

We now prove the limsup inequality part of theorem 2.1:

Proposition 4.7. *Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be any sequence satisfying (2.12). Let $u \in \mathrm{SBV}(\Omega)$. Then there exists $u_n \in L^1(\Omega;\mu_n)$ such that $(\mu_n, u_n) \to (\mu, u)$ in TL^1 and*

$$\limsup_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u_n) \leqslant MS_{\eta,\zeta}(u;\rho).$$

We prove the proposition by providing a recovery sequence for regular functions and argue by approximation. We start by showing how to recover the energy of a function $u \in \mathrm{SBV}(\mathbb{R}^d)$ having the following properties:

(H1) $\overline{S_u}$ is the union of a finite number of $(d-1)$-dimensional simplexes, $\mathcal{H}^{d-1}(\overline{S_u}\backslash S_u) = 0$;
(H2) $u \in C^\infty(\Omega\backslash\overline{S_u}) \cap W^{1,\infty}(\Omega\backslash\overline{S_u})$;
(H3) $MS_{\eta,\zeta}(u) < \infty$.

We then use of the following density theorem which is a consequence of a well known result of Cortesani and Toader [CT99].

Theorem 4.8. *Let Ω be an open bounded set with Lipschitz boundary and $u \in \mathrm{SBV}(\Omega)^p$. Then there exists a sequence of function $u_j \in \mathrm{SBV}(\Omega)^p$ satisfying (H1), (H2) and (H3) such that:*

(a) $\limsup_{j\to\infty} \int_{S_{u_j}} \rho(y)^2 \mathrm{d}\mathcal{H}^{d-1}(y) \leqslant \int_{S_u} \rho(y)^2 \mathrm{d}\mathcal{H}^{d-1}(y)$;
(b) $\nabla u_j L^p \nabla u$ *and* $u_j \xrightarrow{L^1} u$, *where ∇u is, as before, the absolutely continuous part of the gradient Du.*

The following lemma is used to compare the energy of $u \circ T_n$ with the energy of u.

Lemma 4.9. *Let $\{x_i\}_{i=1}^n$ be a sequence of i.i.d. points chosen according to the density ρ. Let $\tilde{\mu}_n$ be the measures provided by lemma 3.1 and $T_n : \Omega \to \{x_1,\dots,x_n\}$ be the transport maps between $\tilde{\mu}_n$ and μ_n. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence satisfying (2.12). For any $u \in \mathrm{SBV}(\Omega)$ satisfying (H1)–(H2), $\xi \in \mathbb{R}^d$, $\ell > 0$ and $\varepsilon > 0$ define*

$$(S_u)_\ell := \{x \in \Omega : d(x, S_u) \leqslant \ell\},$$
$$(S_u - \varepsilon\xi)_\ell = \{x \in \Omega : d(x, S_u - \varepsilon\xi) \leqslant \ell\},$$
$$D(\varepsilon,\ell) := (S_u)_\ell \cup (S_u - \varepsilon\xi)_\ell.$$

Let $\ell_n = \|T_n - Id\|_\infty$. Then for any $x \in \Omega\backslash D(\varepsilon_n, \ell_n)$ it holds

$$|u(T_n(x+\varepsilon_n\xi)) - u(T_n(x))| \leqslant |u(x+\varepsilon_n\xi) - u(x)| + 2\ell_n\|\nabla u\|_\infty. \tag{4.20}$$

Moreover

$$\lim_{n\to\infty} \frac{|D(\varepsilon_n,\ell_n)|}{\varepsilon_n} = 0. \tag{4.21}$$

The proof is presented in appendix B.

Proposition 4.10. *Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be any sequence satisfying (2.12). Let $u \in \mathrm{SBV}(\Omega)$ satisfying (H1)–(H3). Then there exists a sequence of function $\{u_n\}_{n\in\mathbb{N}} \subset L^1(\Omega;\mu_n)$ such that $(\mu_n, u_n) \to (\mu, u)$ in TL^1 and*

$$\limsup_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u_n) \leqslant MS_{\eta,\zeta}(u;\rho). \tag{4.22}$$

Proof. We start by noticing that, since Ω has Lipschitz boundary, for any polyhedral set $A \supset\supset \Omega$, there exists an extension $\tilde{u} \in \mathrm{SBV}(\mathbb{R}^d)$, still satisfying hypothesis (H1)–(H3), such that $|D\tilde{u}|(\partial\Omega) = 0$ and $\tilde{u} = 0$ outside A (see for example [AFP00, proposition 3.21]). We thus fix $A \supset \Omega$ and consider such an extension (still denoted, with a slight abuse of notation, by u). We also extend $\rho(x) = 0$ on $\mathbb{R}^d \backslash \overline{\Omega}$. For every n, consider $\tilde{\mu}_n$ of lemma 3.1. Let T_n, as before be the d_∞ optimal transport map between $\tilde{\mu}_n$ and μ_n and $\ell_n = \|T - I_d\|_{L^\infty(\Omega)}$. Notice that, since $MS_{\eta,\zeta}(u;\rho) < \infty$ we can infer that, for $\mathcal{L}^d$-a.e. $\xi \in \mathbb{R}^d$ and $\mathcal{H}^{d-1}$-a.e. $z \in \xi^\perp$ it holds

$$\int_{\mathbb{R}} |u'_\xi(t;z)|\rho_\xi(t;z)^2 \mathrm{d}t + \int_{S_{u_\xi(\cdot;z)}} \rho_\xi(t;z)^2 \mathrm{d}\mathcal{H}^0(t) < \infty. \tag{4.23}$$

We define $u_n : \{x_1, \dots, x_n\} \to \mathbb{R}$ as

$$u_n(x_i) = \begin{cases} u(x_i) & \text{if } x_i \in \Omega \backslash S_u; \\ u^+(x_i) & \text{if } x_i \in S_u \end{cases} \tag{4.24}$$

where u^+ is defined in (2.5). We now divide the proof in two steps.

Step one: lim sup *bound on* $\mathcal{GAMS}$. We first prove that the lim sup bound holds for the auxiliary energy

$$\mathcal{GAMS}_{\varepsilon_n,n}(u_n) = \frac{1}{\varepsilon_n} \int_{\Omega\times\Omega} \zeta\left(\frac{|u_n(T_n(x)) - u_n(T_n(y))|^p}{\varepsilon_n^{p-q-1}|T(x) - T_n(y)|^q} \right) \eta_{\varepsilon_n}(|T_n(x) - T_n(y)|)\rho(x)^2 \mathrm{d}x\mathrm{d}y$$

From now on, we will omit, as in other proofs, the dependence on n of ε_n and ℓ_n. Define, for $t \in \mathbb{R}$, the kernel $\overline{\eta}(t) := \eta(\max\{t - 2\ell/\varepsilon, 0\})$ (where we are omitting to explicitly denote the dependence on ε). Since

$$|x - y| - 2\ell \leqslant |T_n(x) - T_n(y)| \leqslant |x - y| + 2\ell$$

and since η is non-increasing, we deduce

$$\eta\left(\frac{|T_n(x) - T_n(y)|}{\varepsilon}\right) \leqslant \eta\left(\max\left\{\frac{|x-y|}{\varepsilon} - \frac{2\ell}{\varepsilon}, 0\right\}\right) \leqslant \overline{\eta}\left(\frac{|x-y|}{\varepsilon}\right).$$

Since ζ is non-decreasing we have

$$\begin{aligned} \mathcal{GAMS}_{\varepsilon,n}(u_n) &\leqslant \frac{1}{\varepsilon} \int_{\Omega\times\Omega} \zeta\left(\varepsilon^{1-p+q} \frac{|u_n(T_n(x)) - u_n(T_n(y))|^p}{(|x-y| - 2\ell)_+^q} \right) \overline{\eta}_\varepsilon(|x-y|)\rho(x)^2 \mathrm{d}x\mathrm{d}y \\ &\leqslant \frac{1}{\varepsilon} \int_{\mathbb{R}^d} \mathrm{d}\xi \int_{\Omega\cap(\Omega-\varepsilon\xi)} \zeta\left(\frac{|u_n(T_n(x+\varepsilon\xi)) - u_n(T_n(x))|^p}{\varepsilon^{p-1}(|\xi| - 2\ell/\varepsilon)_+^q} \right) \overline{\eta}(|\xi|)\rho(x)^2 \mathrm{d}x. \end{aligned}$$

For any fixed $\xi \in \mathbb{R}^d$ set $\Omega'_\varepsilon := \Omega \cap (\Omega - \varepsilon\xi)$ (where we omit the dependence of ω'_ε on ξ) and consider

$$\mathcal{F}_{\varepsilon,n}(u_n;\xi) = \frac{1}{\varepsilon} \int_{\Omega'_\varepsilon} \zeta\left(\frac{|u_n(T_n(x+\varepsilon\xi)) - u_n(T_n(x))|^p}{\varepsilon^{p-1}(|\xi| - 2\ell/\varepsilon)^q} \right) \rho(x)^2 \mathrm{d}x$$

so that

$$\mathcal{GAMS}_{\varepsilon,n}(u_n) \leqslant \int_{\mathbb{R}^d} \overline{\eta}(|\xi|)\mathcal{F}_\varepsilon(u_n;\xi)\mathrm{d}\xi. \tag{4.25}$$

Let $D(\varepsilon,\ell)$ be the set defined in lemma 4.9. Notice that

$$\mathcal{F}_{\varepsilon,n}(u_n;\xi)=\frac{1}{\varepsilon}\int_{\Omega'_\varepsilon\setminus D(\varepsilon,\ell)}\zeta\left(\frac{|u_n(T_n(x+\varepsilon\xi))-u_n(T_n(x))|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho(x)^2\mathrm{d}x \tag{4.26}$$

$$+\frac{1}{\varepsilon}\int_{D(\varepsilon,\ell)}\zeta\left(\frac{|u_n(T_n(x+\varepsilon\xi))-u_n(T_n(x))|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho(x)^2\mathrm{d}x. \tag{4.27}$$

The second integral (4.27) can be easily estimated as

$$\frac{1}{\varepsilon}\int_{D(\varepsilon,\ell)}\zeta\left(\frac{|u_n(T_n(x+\varepsilon\xi))-u_n(T_n(x))|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho(x)^2\mathrm{d}x\leqslant C^2\Theta\frac{|D(\varepsilon,\ell)|}{\varepsilon} \tag{4.28}$$

which, thanks to (4.21), is decaying to 0 as $n\to\infty$ (recall that $\varepsilon=\varepsilon_n\to 0$).

Let us now treat the first term (4.26) in the light of lemma 4.9. Since ζ is non-decreasing

$$\frac{1}{\varepsilon}\int_{\Omega'_\varepsilon\setminus D(\varepsilon,\ell)}\zeta\left(\frac{|u_n(T_n(x+\varepsilon\xi))-u_n(T_n(x))|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho(x)^2\mathrm{d}x$$

$$\leqslant\frac{1}{\varepsilon}\int_{\Omega'_\varepsilon}\zeta\left(\frac{|\,|u(x+\varepsilon\xi)-u(x))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho(x)^2\mathrm{d}x. \tag{4.29}$$

From now on, we use the same arguments of of the proof of Gobbino [Gob98, theorem 3.4, proposition 3.5, theorem 3.6], suitably adapted to our situation (see also [GM01]). By slicing along $\xi^\perp$ we get

$$\frac{1}{\varepsilon}\int_{\Omega'_\varepsilon}\zeta\left(\frac{|\,|u(x+\varepsilon\xi)-u(x))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho(x)^2\mathrm{d}x$$

$$=\frac{1}{\varepsilon}\int_{\xi^\perp}\mathrm{d}\mathcal{H}^{d-1}(z)\int_{[\Omega'_\varepsilon]_z}\zeta\left(\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho_\xi(t;z)^2\mathrm{d}t.$$

and, for the sake of clarity we introduce the notation

$$F_\varepsilon(u_\xi(\cdot;z);\xi,[\Omega'_\varepsilon]_z):=\frac{1}{\varepsilon}\int_{[\Omega'_\varepsilon]_z}\zeta\left(\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho_\xi(t;z)^2\mathrm{d}t.$$

We define

$$[S]_z:=\{t\in[\Omega'_\varepsilon]_z:[t,t+\varepsilon|\xi|)\cap S_{u_\xi(\cdot;z)}\neq\emptyset\} \tag{4.30}$$

and, for a fixed $\xi\in\mathbb{R}^d\setminus\{0\}$, $z\in\xi^\perp$, we split once again $F_\varepsilon(u_\xi(\cdot;z);\xi,[\Omega]_z)$ as (notice that, for any fixed ξ we can find $\varepsilon>0$ small enough such that $|\xi|-2\ell/\varepsilon>0$)

$$\frac{1}{\varepsilon}\int_{[\Omega'_\varepsilon]_z\setminus[S]_z}\zeta\left(\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho_\xi(t;z)^2\mathrm{d}t$$

$$+\frac{1}{\varepsilon}\int_{[S]_z}\zeta\left(\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^{p-1}(|\xi|-2\ell/\varepsilon)_+^q}\right)\rho_\xi(t;z)^2\mathrm{d}t$$

$$\leqslant\frac{\zeta'(0)}{(|\xi|-2\ell/\varepsilon)^q}\int_{[\Omega'_\varepsilon]_z\setminus[S]_z}\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^p}\rho_\xi(t;z)^2\mathrm{d}t \tag{4.31}$$

$$+\Theta\frac{1}{\varepsilon}\int_{[S]_z}\rho_\xi(t;z)^2\mathrm{d}t \tag{4.32}$$

Notice that, if $t\in[\Omega'_\varepsilon]_z\backslash[S]_z$, we have

$$\begin{aligned}\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^p}&\leqslant\varepsilon^{-p}\left[\int_0^{\varepsilon|\xi|}\left(|u'_\xi(t+\tau;z)|+\frac{\ell}{\varepsilon|\xi|}\|\nabla u\|_\infty\right)\mathrm{d}\tau\right]^p\\&\leqslant\varepsilon^{-1}|\xi|^{p-1}\int_0^{\varepsilon|\xi|}\left[|u'_\xi(t+\tau;z)|+\frac{\ell}{\varepsilon|\xi|}\|\nabla u\|_\infty\right]^p\mathrm{d}\tau\\&=\varepsilon^{-1}|\xi|^{p-1}\int_t^{t+\varepsilon|\xi|}\left[|u'_\xi(s;z)|+\frac{\ell}{\varepsilon|\xi|}\|\nabla u\|_\infty\right]^p\mathrm{d}s\end{aligned}$$

hence

$$\begin{aligned}&\frac{\zeta'(0)}{(|\xi|-2\ell/\varepsilon)^q}\int_{[\Omega'_\varepsilon]_z\backslash[S]_z}\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^p}\rho_\xi(t;z)^2\mathrm{d}t\\&\quad\leqslant\frac{\zeta'(0)|\xi|^{p-1}\varepsilon^{-1}}{(|\xi|-2\ell/\varepsilon)^q}\int_{\mathbb{R}}\int_t^{t+\varepsilon|\xi|}\left[|u'_\xi(s;z)|+\frac{\ell}{\varepsilon|\xi|}\|\nabla u\|_\infty\right]^p\mathrm{d}s\rho_\xi(t;z)^2\mathrm{d}t\\&\quad=\frac{\zeta'(0)|\xi|^{p-1}\varepsilon^{-1}}{(|\xi|-2\ell/\varepsilon)^q}\int_{\mathbb{R}}\left[|u'_\xi(s;z)|+\frac{\ell}{\varepsilon|\xi|}\|\nabla u\|_\infty\right]^p\int_{s-\varepsilon|\xi|}^s\rho_\xi(t;z)^2\mathrm{d}t\mathrm{d}s\end{aligned}\tag{4.33}$$

where we used the identity

$$\mathbb{1}_{\mathbb{R}}(t)\mathbb{1}_{[t,t+\varepsilon|\xi|]}(s)=\mathbb{1}_{\mathbb{R}}(s)\mathbb{1}_{[s-\varepsilon|\xi|,s]}(t).$$

Notice now that, for $\mathcal{L}^1$-a.e. $s\in\mathbb{R}$, we have

$$\lim_{\varepsilon\to0}(\varepsilon|\xi|)^{-1}\left[|u'_\xi(s;z)|+\frac{\ell}{\varepsilon|\xi|}\|\nabla u\|_\infty\right]^p\int_{s-\varepsilon|\xi|}^s\rho_\xi(t;z)^2\mathrm{d}t=|u'_\xi(s;z)|^p\rho_\xi(s;z)^2.$$

In particular by exploiting the dominated convergence theorem (the sequence is dominated by twice its limit for example, which, for $\mathcal{L}^d$-a.e. $\xi\mathbb{R}^d$ is summable for $\mathcal{H}^{d-1}$-a.e. $z\in\xi^\perp$ due to (4.23)) we obtain for $\mathcal{L}^d$-a.e. fixed ξ and $\mathcal{H}^{d-1}$-a.e. fixed $z\in\xi^\perp$

$$\begin{aligned}&\lim_{\varepsilon\to0^+}\frac{\zeta'(0)}{(|\xi|-2\ell/\varepsilon)^q}\int_{[\Omega'_\varepsilon]_z\backslash[S]_z}\frac{|\,|u_\xi(t+\varepsilon|\xi|;z)-u_\xi(t;z))|+2\ell\|\nabla u\|_\infty|^p}{\varepsilon^p}\rho_\xi(t;z)^2\mathrm{d}t\\&\quad\leqslant\zeta'(0)|\xi|^{p-q}\int_{\mathbb{R}}|u'_\xi(s;z)|^p\rho_\xi(s;z)^2\mathrm{d}s\\&\quad=\zeta'(0)|\xi|^{p-q}\int_{[\Omega]_z}|u'_\xi(s;z)|^p\rho_\xi(s;z)^2\mathrm{d}s\end{aligned}\tag{4.34}$$

since ρ is defined to be zero outside Ω. Let ω be the modulus of continuity of ρ^2 on $\overline{\Omega}$. That is, for $r>0$

$$\omega(r)=\sup\{|\rho^2(x)-\rho^2(y)|:x,y\in\overline{\Omega},\quad|x-y|<r\}.$$

From the definition of $[S]_z$ in (4.30) follows that

$$\begin{aligned}\frac{\Theta}{\varepsilon}\int_{[S]_z}\rho_\xi(t;z)^2\mathrm{d}t &\leqslant \frac{\Theta}{\varepsilon}\int_{S_{u_\xi(\cdot;z)}}\int_{(y-\varepsilon|\xi|,y]\cap[\Omega'_\varepsilon]_z}\rho_\xi(t;z)^2\,\mathrm{d}t\,\mathrm{d}\mathcal{H}^0(y)\\ &\leqslant \Theta|\xi|\int_{S_{u_\xi(\cdot;z)}\cap[\Omega'_\varepsilon]_z}\rho_\xi(y;z)^2+\omega(\varepsilon|\xi|)\,\mathrm{d}\mathcal{H}^0(y)\\ &\leqslant \Theta|\xi|\int_{S_{u_\xi(\cdot;z)}\cap[\Omega]_z}\rho_\xi(y;z)^2+\omega(\varepsilon|\xi|)\,\mathrm{d}\mathcal{H}^0(y) \qquad (4.35)\end{aligned}$$

where the last inequality follows from the fact that

$$[\Omega_\varepsilon]_z=[\Omega\cap(\Omega-\varepsilon\xi)]_z\subset[\Omega]_z.$$

By collecting together (4.34) and (4.35) we get, for $\mathcal{L}^d$-a.e. $\xi\in\mathbb{R}^d$ and for $\mathcal{H}^{d-1}$-a.e $z\in\xi^\perp$, that

$$\begin{aligned}\limsup_{\varepsilon\to0}F_\varepsilon(u_\xi(\cdot;z);\xi,[\Omega'_\varepsilon]_z)\leqslant\zeta'(0)|\xi|^{p-q}&\int_{[\Omega]_z}|u'_\xi(s;z)|^p\rho_\xi(s;z)^2\mathrm{d}s\\ &+\Theta|\xi|\int_{S_{u_\xi(\cdot;z)}\cap[\Omega]_z}\rho_\xi(t;z)^2\mathrm{d}\mathcal{H}^0(t).\end{aligned}$$

In particular we can apply the reverse Fatou's lemma (again for every ξ, F_ε is dominated by twice its limit which is summable in z due to $MS_{\eta,\zeta}(u)<\infty$) and conclude that

$$\begin{aligned}\limsup_{\varepsilon\to0}\int_{\xi^\perp}F_\varepsilon(u_\xi(\cdot;z);\xi,[\Omega'_\varepsilon]_z)\mathrm{d}\mathcal{H}^{d-1}(z)\leqslant\zeta'(0)|\xi|^{p-q}&\int_{\xi^\perp}\mathrm{d}\mathcal{H}^{d-1}(z)\int_{[\Omega]_z}|u'_\xi(s;z)|^p\rho_\xi(s;z)^2\mathrm{d}s\\ &+\Theta|\xi|\int_{\xi^\perp}\mathrm{d}\mathcal{H}^{d-1}(z)\int_{S_{u_\xi(\cdot;z)}\cap[\Omega]_z}\rho_\xi(t;z)^2\mathrm{d}\mathcal{H}^0(t).\end{aligned}$$

By collecting (4.26)–(4.29) the definition of F_ε and lemma 4.9 we get, for $\mathcal{L}^d$-a.e. $\xi\in\mathbb{R}^d$, that

$$\begin{aligned}\limsup_{n\to\infty}\mathcal{F}_{\varepsilon,n}(u_n;\xi)\leqslant\zeta'(0)|\xi|^{p-q}&\int_{\xi^\perp}\mathrm{d}\mathcal{H}^{d-1}(z)\int_{[\Omega]_z}|u'_\xi(s;z)|^p\rho_\xi(s;z)^2\mathrm{d}s\\ &+\Theta|\xi|\int_{\xi^\perp}\mathrm{d}\mathcal{H}^{d-1}(z)\int_{S_{u_\xi(\cdot;z)}\cap[\Omega]_z}\rho_\xi(t;z)^2\mathrm{d}\mathcal{H}^0(t).\end{aligned}$$

A further application of the reverse Fatou's lemma on (4.25), combined with the fact that $\overline{\eta}(t)=\eta(\max\{t-2\ell/\varepsilon,0\})\to\eta(t)$ in L^1 as $\varepsilon\to0$, leads to

$$\begin{aligned}\limsup_{n\to\infty}\mathcal{GAMS}_{\varepsilon_n,n}(u_n)\leqslant\zeta'(0)&\int_{\mathbb{R}^d}\eta(|\xi|)|\xi|^{p-q}\mathrm{d}\xi\int_{\xi^\perp}\mathrm{d}z\int_{[\Omega]_z}|u'_\xi(s;z)|^p\rho_\xi(s;z)^2\mathrm{d}s\\ &+\Theta\int_{\mathbb{R}^d}\eta(|\xi|)|\xi|\mathrm{d}\xi\int_{\xi^\perp}\mathrm{d}z\int_{S_{u_\xi(\cdot;z)}\cap[\Omega]_z}\rho_\xi(t;z)^2\mathrm{d}\mathcal{H}^0(t)\end{aligned}$$

which, thanks to (4.5) achieves the proof of step one.

Step two: lim sup *bound on* $\mathcal{GMS}$. Consider $\eta_M:=\mathbb{1}_{[0,M)}(t)\eta(t)$ and notice that, by exploiting the notation of the proof of lemma 4.2, we have

$$\overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = \mathcal{GAMS}_{\varepsilon_n,n}(u_n) + \mathcal{R}_n(u_n)$$
$$= \mathcal{GAMS}_{\varepsilon_n,n}(u_n) + \mathcal{R}_n(u_n;\eta_M) + \mathcal{R}_n(u_n;\eta-\eta_M)$$

where, with $\mathcal{R}_n(u_n;\eta_M)$, $\mathcal{R}_n(u_n;\eta-\eta_M)$ we mean the energy $\mathcal{R}_n(u_n)$ with η^M, $\eta-\eta^M$ in place of η. Since $\mathcal{R}_n(u_n;\eta_M) = \overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n;\eta_M) - \mathcal{GAMS}_{\varepsilon_n,n}(u_n;\eta_M)$, by virtue of lemma 4.2 we have

$$\lim_{n\to\infty} \mathcal{R}_n(u_n;\eta_M) = 0.$$

From the other side, since ρ is bounded from above and below we have that

$$|\mathcal{R}_n(u_n;\eta-\eta_M)| \leqslant C\mathcal{GAMS}_{\varepsilon_n,n}(u_n;\eta-\eta_M)$$

for a universal constant C. Thanks to the step one and to proposition 4.3 we thus have

$$\lim_{n\to\infty} |\mathcal{R}_n(u_n;\eta-\eta_M)| \leqslant C\left[\lim_{n\to\infty}\mathcal{GAMS}_{\varepsilon_n,n}(u_n) - \lim_{n\to\infty}\mathcal{GAMS}_{\varepsilon_n,n}(u_n;\eta_M)\right]$$
$$= C(MS_{\eta,\zeta}(u) - MS_{\eta_M,\zeta}(u)).$$

Since $MS_{\eta,\zeta}(u) < +\infty$ by taking the limit as $M\to\infty$ and by exploiting the continuity of the constants in MS we get

$$\lim_{M\to\infty}\lim_{n\to\infty} |\mathcal{R}_n(u_n;\eta-\eta_M)| = 0,$$

yielding

$$\lim_{n\to+\infty} \mathcal{R}_n(u_n) = 0.$$

In particular, by invoking (4.7), we reach

$$\limsup_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u_n) \leqslant \limsup_{n\to\infty} \overline{\mathcal{GMS}}_{\varepsilon_n,n}(u_n) = \limsup_{n\to\infty} \mathcal{GAMS}_{\varepsilon_n,n}(u_n) \leqslant MS_{\eta,\zeta}(u).$$

□

Proof of proposition 4.7. Assume now that $u\in \mathrm{SBV}(\Omega)$. Let u_j be the sequence given by theorem 4.8. Then $u_j\to u$ in L^1 which means that $d_{TL^1}((\mu,u_j),(\mu,u))\to 0$. Set (up to a subsequence)

$$L := \limsup_{j\to\infty} MS_{\eta,\zeta}(u_j).$$

Notice that thanks to assertions (a) and (b) of theorem 4.8 we have that $L\leqslant MS_{\eta,\zeta}(u)$. For all $k\in\mathbb{N}$, consider j_k such that

$$d_{TL^1}((\mu,u_{j_k}),(\mu,u)) \leqslant 1/(2k)$$
$$MS_{\eta,\zeta}(u_{j_k}) \leqslant L + 1/(2k).$$

In particular it also holds that

$$MS_{\eta,\zeta}(u_{j_k}) \leqslant L + 1/(2k) \leqslant MS_{\eta,\zeta}(u) + 1/(2k).$$

For every j_k chosen as above let $\{u^n_{j_k}\}_{n\in\mathbb{N}}$ be the sequence given by proposition 4.10 relative to u_{j_k}. By exploiting proposition 4.3 we can infer

$$MS_{\eta,\zeta}(u_{j_k}) = \lim_{n\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(u^n_{j_k}).$$

Choose now n_k such that

$$d_{TL^1}((\mu_n, u^n_{j_k}), (u_{j_k}, \mu)) < 1/(2k)$$
$$\mathcal{GMS}_{\varepsilon_n,n}(u^n_{j_k}) < MS_{\eta,\zeta}(u_{j_k}) + 1/(2k) \quad \text{for all } n \geqslant n_k$$
$$\left(< MS_{\eta,\zeta}(u) + \frac{1}{k} \quad \text{for all } n \geqslant n_k \right).$$

Define now the following recovery sequence

$$w_n := u^n_{j_k}, \quad \text{if } n \in [n_k, n_{k+1}),\ k \in \mathbb{N} \tag{4.36}$$

This means that, for any $n \in [n_k, n_{k+1})$ we have

$$d_{TL^1}((\mu_n, u^n_{j_k}), (\mu, u)) < 1/k,$$
$$\mathcal{GMS}_{\varepsilon_n,n}(u^n_{j_k}) \leqslant MS_{\eta,\zeta}(u) + \frac{1}{k}.$$

Implying

$$d_{TL^1}((\mu_n, w_n), (\mu, u)) < 1/k, \quad \text{for all } n \in [n_k, n_{k+1})$$
$$\mathcal{GMS}_{\varepsilon_n,n}(w_n) \leqslant MS_{\eta,\zeta}(u) + \frac{1}{k} \quad \text{for all } n \in [n_k, n_{k+1}).$$

In particular, $(\mu_n, w_n) \to (\mu, u)$ in TL^1 and

$$\limsup_{k\to\infty} \mathcal{GMS}_{\varepsilon_n,n}(w_n) \leqslant MS_{\eta,\zeta}(u).$$

□

5. Proof of the compactness result (theorem 2.3)

This section is devoted to the proof of theorem 2.3 that establishes a compactness result for sequences of functions with uniformly bounded $\mathcal{GMS}_{\varepsilon_n,n}$ where ε_n is any sequence satisfying (2.12).

Remark 5.1. Let us point out that, in contrast to [GTS16] where an L^1 bound is assumed, our compactness theorem 2.3 requires an L^∞ bound on the sequence u_n. Namely due to the fact that in $\mathcal{GMS}_{\varepsilon_n,n}(u_n)$ the differences in u_n are inside a bounded concave function ζ a uniform bound on $\mathcal{GMS}_{\varepsilon_n,n}(u_n)$ is, in general, not translatable into a uniform bound on $GTV_{\varepsilon,n}$. This is not just a technical issue and in fact an L^1-type bound is not sufficient for compactness. Here we provide a counterexample to compactness if one only assumes an L^1 bound on u_n. Choose $\rho = 1$ and $\Omega = Q$ the unit cube centered at 0. Let $\{Q^k_i\}_{i=1}^{2^{kd}}$ be a di-adic division of Q in cubes of edges-size 2^{-k} and let $\{x^k_i\}_{i=1}^{2^{kd}}$ be the uniform grid given by the baricenter of each cube Q^k_i. Consider the sequence of functions $u_k : \{x^k_i\}_{i=1}^{2^{kd}} \to \mathbb{R}$ defined as

$$u_k(x^k_j) := \frac{\mathbb{1}_{B_{r_k}(0)}(x^k_j)}{\omega_d r_k^d}$$

with $r_k := 2^{-k/2}$. Notice that $x^k_i \in B_{r_k}(0)$ implies $Q^k_i \subset B_{2r_k}(0)$ and thus

$$\#(\{i : x^k_i \in B_{r_k}(0)\}) \leqslant \#(\{i : Q^k_i \subset B_{2r_k}(0)\}) = 2^d \omega_d 2^{dk/2}.$$

On the other hand

$$\#(\{i : Q_i^k \subset B_{r_k}(0)\}) = 2^{kd}\left|\bigcup_{Q_i^k \subset B_{r_k}(0)} Q_j^k\right| \geqslant 2^{kd}|B_{r_k/2}(0)| = 2^{-d}\omega_d 2^{kd/2}.$$

Since $Q_i^k \subset B_{r_k}(0)$ implies $x_i^k \in B_{r_k}(0)$, we conclude

$$2^{-d}\omega_d 2^{dk/2} \leqslant \#(\{i : x_i^k \subset B_{r_k}(0)\}) \leqslant 2^d \omega_d 2^{dk/2}.$$

In particular, setting $\nu_k := 2^{-kd}\sum_{i=1}^{2^{kd}} \delta_{x_i^k}$, we have

$$\int_Q u_k \mathrm{d}\nu_k = 2^{-kd}\sum_{i=1}^{2^{kd}} u_k(x_i^k) = 2^{-kd}\frac{\#(\{i : x_i^k \in B_{r_k}(0)\})}{\omega_d r_k^d} = \frac{\#(\{i : x_i^k \in B_{r_k}(0)\})}{\omega_d 2^{kd/2}}$$

and so

$$2^{-d} \leqslant \int_Q u_k \mathrm{d}\nu_k \leqslant 2^d \quad \text{for all } k \in \mathbb{N}. \tag{5.1}$$

This means that $u_k \in L^1(Q;\nu_k)$ and that

$$\sup_{k\in\mathbb{N}}\{\|u_k\|_{L^1}\} \leqslant 2^d \tag{5.2}$$

Consider now $\varepsilon_k := 2^{-k\alpha}$ for some $1/2 < \alpha < 1$ and notice that it satisfies (2.12), since

$$\lim_{k\to\infty} \frac{\log(n_k)^{1/d}}{\varepsilon_k n_k^{1/d}} = 0$$

(here $n_k = 2^{kd}$ and we are also considering $d > 2$). Now we choose ζ as

$$\zeta(x) = \begin{cases} x & \text{for } x \leqslant 1, \\ 1 & \text{for } x \geqslant 1. \end{cases} \tag{5.3}$$

With all these choices in mind, for any kernel η satisfying the assumptions (B1) and (B2), we can conclude

$$\begin{aligned}\mathcal{GMS}_{\varepsilon_k,n_k}(u_k) &= \frac{1}{\varepsilon_k n_k^2}\sum_{i,j=1}^{n_k} \zeta\left(\varepsilon_k^{1-p+q}\frac{|u_k(x_i^k) - u_k(x_j^k)|^p}{|x_i^k - x_j^k|^q}\right)\eta_{\varepsilon_k}(|x_i^k - x_j^k|) \\ &\leqslant \frac{2}{\varepsilon_k n_k^2}\sum_{x_i^k \in B_{r_k}(0)^c}\ \sum_{x_j^k \in B_{r_k}(0)} \eta_{\varepsilon_k}(|x_i^k - x_j^k|) = 2GTV_{\varepsilon_k,n_k}(\mathbb{1}_{B_{r_k}}(0)).\end{aligned}$$

Notice that $\sup_k\{GTV_{\varepsilon_k,n_k}(\mathbb{1}_{B_{r_k}}(0))\} < \infty$ and so

$$\sup_{k\in\mathbb{N}}\{\mathcal{GMS}_{\varepsilon_k,n_k}(u_k)\} < \infty. \tag{5.4}$$

By collecting (5.2) and (5.4) we are finally lead to

$$\sup_{k\in\mathbb{N}}\{\|u_k\|_{L^1} + \mathcal{GMS}_{\varepsilon_k,n_k}(u_k)\} < \infty.$$

Nonetheless we cannot expect any sort of TL^1 compactness for the sequence (ν_k, u_k). Indeed, the only possible pointwise limit for $u_k \circ T_{n_k} : Q \to \{0, r_k^{-d}/\omega_d\}$ can be $u = 0$ but

$$\int_Q u_k(T_{n_k}(x))\mathrm{d}x = \int_Q u_k \mathrm{d}\nu_k \geqslant 2^{-d} > 0$$

because of (5.1).

Our proof is based on the approach to compactness for general non-local functionals established in [GM01, theorem 5.1].

Proposition 5.2. *Let $p \geqslant 1$ and consider an open set $A \subset \mathbb{R}^d$ with Lipscthiz boundary. Let $u_\varepsilon \in L^1(\mathbb{R}^d)$ be a sequence of function such that*

$$\inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|u_\varepsilon(x) - u_\varepsilon(y)|}{|x-y|} J_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y + \|u_\varepsilon\|_\infty\right\} < +\infty$$

where J is any kernel such that $\{\xi : J(\xi) > c\}$ has non-empty interior for some $c > 0$. Then the sequence u_ε is compact in $L^1(\mathbb{R}^d)$.

Before proceeding to the proof of theorem 2.3 and in order to apply proposition 5.2 (which holds for functions defined on the whole $\mathbb{R}^d$) we need the following extension lemma in the same spirit of [GTS16, lemma 4.4].

Lemma 5.3. *Suppose that Ω is abounded open set with C^2 boundary. Let η be a compactly supported, non-increasing kernel which is not identically equal to zero. Let $\{u_\varepsilon\}_{\varepsilon>0} \subset L^1(\Omega)$ be a sequence. Then there exists a sequence of function $\{v_\varepsilon\}_{\varepsilon>0} \subset L^1(\mathbb{R}^d)$ such that*

(*a*) *$v_\varepsilon = u_\varepsilon$ $\mathcal{L}^d$-a.e. on Ω;*

(*b*) *There exists a kernel J^η such that $\{\xi : J^\eta(|\xi|) > c\}$ has not empty interior for some $c > 0$ and such that if*

$$\sup_{\varepsilon>0}\left\{\int_{\Omega\times\Omega} \frac{|u_\varepsilon(x) - u_\varepsilon(y)|}{|x-y|}\eta_\varepsilon(|x-y|)\left(\frac{|x-y|}{\varepsilon}\right)^p \mathrm{d}x\mathrm{d}y + \|u_\varepsilon\|_\infty\right\} < \infty,$$

then

$$\sup_{\varepsilon>0}\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|v_\varepsilon(x) - v_\varepsilon(y)|}{|x-y|} J^\eta_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y + \|v_\varepsilon\|_\infty\right\} < \infty.$$

Proof. Since Ω has C^2 boundary we can find $\delta > 0$ for which the projection operator $x \to Px \in \Omega$ is well defined on $U := \{x \in \mathbb{R}^d | d(x, \Omega) \leqslant \delta\}$ and satisfies

$$|x - Px| = d(x, \Omega).$$

We moreover consider a smooth cut off function $\tau(s) \leqslant 1$, such that $\tau(s) = 1$ for $s \leqslant \delta/8$ and $\tau(s) = 0$ for $s \geqslant \delta/4$ and we consider the reflection $Rx := 2Px - x$. Set also

$$W := \{z \in \mathbb{R}^d\backslash\Omega | d(x, \Omega) < \delta/4\}$$
$$V := \{z \in \mathbb{R}^d\backslash\Omega | d(x, \Omega) < \delta/8\}.$$

It has been shown in the proof of [GTS16, lemma 4.4] that

$$\frac{1}{4}|x-y| \leqslant |Rx - Ry| \leqslant 4|x-y| \quad \text{for all } x, y \in W; \tag{5.5}$$

$$|Rx - y| \leqslant 2|x - y| \quad \text{for all } x \in W,\ y \in \Omega; \tag{5.6}$$

Since η can be extended continuously at 0 with $\eta(0) > 0$, up to decreasing the value of δ, we can also guarantee that

$$J^{\eta}(t) := t^{p}\eta\,(4t)\left(1 - \mathbb{1}_{[0,\delta]}(t)\right)$$

is such that $\{t|J^{\eta}(t) > c\}$ has not empty interior for some $c > 0$. Notice also that

$$|Rx - y| \geqslant \frac{3}{4}|x - y| \quad \text{for } x, y \in W,\ |x - y| \geqslant \delta. \tag{5.7}$$

In the light of this fact, we introduce the functions $\tilde{v}_{\varepsilon} := u_{\varepsilon}(Rx)$ and $v_{\varepsilon}(x) := \tau(|Px - x|)\tilde{v}_{\varepsilon}(x)$. Clearly (a) is satisfied. Notice that for $y \in \Omega$

$$\frac{|x - y|}{\varepsilon} \geqslant \frac{d(x, \Omega)}{\varepsilon}.$$

Thus, if $d(x, \Omega) \geqslant \delta/8$, for ε small enough and thanks to the fact that η is compactly supported, we conclude that

$$\eta_{\varepsilon}(4|x - y|) = 0.$$

Consequently

$$\begin{aligned}
&\int_{(\mathbb{R}^d\setminus\Omega)\times\Omega} \frac{|v_{\varepsilon}(x) - v_{\varepsilon}(y)|}{|x - y|} J^{\eta}_{\varepsilon}(|x - y|)\mathrm{d}x\mathrm{d}y \\
&\quad = \int_{(\mathbb{R}^d\setminus\Omega)\times\Omega} \frac{|v_{\varepsilon}(x) - v_{\varepsilon}(y)|}{|x - y|} \eta_{\varepsilon}(4|x - y|)\left(\frac{|x - y|}{\varepsilon}\right)^{p}\left(1 - \mathbb{1}_{[0,\delta)}\left(\frac{|x - y|}{\varepsilon}\right)\right)\mathrm{d}x\mathrm{d}y \\
&\quad = \int_{\substack{V\times\Omega \\ \{|x-y|\geqslant\delta\varepsilon\}}} \frac{|\tilde{v}_{\varepsilon}(x) - \tilde{v}_{\varepsilon}(y)|}{|x - y|} \eta_{\varepsilon}(4|x - y|)\left(\frac{|x - y|}{\varepsilon}\right)^{p}\mathrm{d}x\mathrm{d}y \\
&\quad \leqslant C\int_{\substack{V\times\Omega \\ \{|x-y|\geqslant\delta\varepsilon\}}} \frac{|u_{\varepsilon}(Rx) - u_{\varepsilon}(y)|}{|Rx - y|} \eta_{\varepsilon}(4|x - y|)\left(\frac{|Rx - y|}{\varepsilon}\right)^{p}\mathrm{d}x\mathrm{d}y \\
&\quad \leqslant C\int_{V\times\Omega} \frac{|u_{\varepsilon}(Rx) - u_{\varepsilon}(y)|}{|Rx - y|} \eta_{\varepsilon}(2|Rx - y|)\left(\frac{|Rx - y|}{\varepsilon}\right)^{p}\mathrm{d}x\mathrm{d}y \\
&\quad \leqslant C\int_{\Omega\times\Omega} \frac{|u_{\varepsilon}(z) - u_{\varepsilon}(y)|}{|z - y|} \eta_{\varepsilon}(2|z - y|)\left(\frac{|z - y|}{\varepsilon}\right)^{p}\mathrm{d}z\mathrm{d}y \\
&\quad \leqslant C\int_{\Omega\times\Omega} \frac{|u_{\varepsilon}(z) - u_{\varepsilon}(y)|}{|z - y|} \eta_{\varepsilon}(|z - y|)\left(\frac{|z - y|}{\varepsilon}\right)^{p}\mathrm{d}z\mathrm{d}y.
\end{aligned} \tag{5.8}$$

Where we have exploited (5.5)–(5.7) the change of variable $Rx = z$ and the fact that R is bi-Lipscthiz on W. From the other side, for $(x, y) \in W \times W$ we have

$$\frac{|v_\varepsilon(x)-v_\varepsilon(y)|}{|x-y|} \leqslant \frac{\tau(|Px-x|)(|\tilde{v}_\varepsilon(x)-\tilde{v}_\varepsilon(y)|)+\tau(|Px-x|)-\tau(|Py-y|)|\tilde{v}_\varepsilon(y)|}{|x-y|}$$
$$\leqslant \frac{|u_\varepsilon(Rx)-u_\varepsilon(Ry)|}{|x-y|}+\|v_\varepsilon\|_\infty \mathrm{Lip}(\tau) \leqslant \frac{1}{4}\frac{|u_\varepsilon(Rx)-u_\varepsilon(Ry)|}{|Rx-Ry|}+\|v_\varepsilon\|_\infty \mathrm{Lip}(\tau).$$

Moreover

$$\eta_\varepsilon(4|x-y|)\left(\frac{|x-y|}{\varepsilon}\right)^p\left(1-\mathbb{1}_{[0,\delta)}\left(\frac{|x-y|}{\varepsilon}\right)\right) \leqslant \eta_\varepsilon(4|x-y|)\left(\frac{|x-y|}{\varepsilon}\right)^p$$
$$\leqslant C\eta_\varepsilon(|Rx-Ry|)\left(\frac{|Rx-Ry|}{\varepsilon}\right)^p.$$

In particular with the same change of variable as above we achieve

$$\begin{aligned}&\int_{(\mathbb{R}^d\setminus\Omega)\times(\mathbb{R}^d\setminus\Omega)}\frac{|v_\varepsilon(x)-v_\varepsilon(y)|}{|x-y|}J^\eta_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y\\ &\leqslant C\left(1+\int_{\Omega\times\Omega}\frac{|u_\varepsilon(x)-u_\varepsilon(y)|}{|x-y|}\eta_\varepsilon(|x-y|)\left(\frac{|x-y|}{\varepsilon}\right)^p\mathrm{d}x\mathrm{d}y\right).\end{aligned} \tag{5.9}$$

By collecting (5.8), (5.9) and the definition of v_ε we prove (b). □

Proof of theorem 2.3. Since ρ is always bounded from above and below, without loss of generality we can assume $\rho = 1$. Moreover we can always assume that η is compactly supported since, by replacing η with $\eta\mathbb{1}_{[0,M]}$ for suitable M, we are decreasing the energy. Without loss of generality we may also assume that η is supported on $[0,1]$. Moreover, as usual, we will omit the dependence on n of the sequences ε_n and $\ell_n = \|T_n - Id\|_\infty$.

Due to the properties of ζ we can always find real constants ζ_2 and $c_2 > 0$ such that

$$\begin{cases}\dfrac{\zeta'(0)}{2} \leqslant \dfrac{\zeta(t)}{t} & \text{for } t \leqslant \zeta_2\\ c_2 \leqslant \zeta(t) & \text{for } t \geqslant \zeta_2.\end{cases} \tag{5.10}$$

Set $\tilde{u}_n(x) := u_n \circ T_n$, where $T_n : \Omega \to \{x_1,\dots,x_n\}$ is the map that transports $\tilde{\mu}_n$ to μ_n; the measures given by lemma 3.1. We define

$$A_\varepsilon := \left\{(x,y)\in\Omega\times\Omega \;\middle|\; \varepsilon^{1-p+q}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|^p}{|T_n(x)-T_n(y)|^q} \geqslant \zeta_2\right\}. \tag{5.11}$$

We immediately see that

$$\mathcal{GMS}_{\varepsilon,n}(u_n) \geqslant \frac{1}{\varepsilon}\int_{A_\varepsilon}\zeta\left(\varepsilon^{1+q-p}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|^p}{|T_n(x)-T_n(y)|^q}\right)\eta_\varepsilon(|T_n(x)-T_n(y)|)\mathrm{d}x\mathrm{d}y$$
$$\geqslant \frac{c_2}{\varepsilon}\int_{A_\varepsilon}\eta_\varepsilon(|T_n(x)-T_n(y)|)\mathrm{d}x\mathrm{d}y.$$

Moreover since η is non increasing and non identically 0 we can find a positive $r > 0$ such that $\eta(t+r)\mathbb{1}_{[0,1)}(t)$ is not identically 0. Set $\bar{\eta}(t) := \eta(t+r)\mathbb{1}_{[0,1]}(t)$ and notice that $\bar{\eta}$ is still a non increasing kernel, supported on $[0, 1-r]$. Since for ε small enough we can always guarantee that

$$\frac{|T_n(x)-T_n(y)|}{\varepsilon} \leqslant \frac{|x-y|}{\varepsilon}+r \Rightarrow \eta\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon}\right) \geqslant \bar{\eta}\left(\frac{|x-y|}{\varepsilon}\right).$$

We can also infer, for $|x-y|/\varepsilon \leqslant (1-r)$

$$\left(\frac{|x-y|}{\varepsilon}\right)^p \bar{\eta}_\varepsilon(|x-y|) \leqslant (1-r)^{p-1}\frac{|x-y|}{\varepsilon}\eta_\varepsilon(|T_n(x)-T_n(y)|).$$

Thus we can conclude

$$\left(\int_{A_\varepsilon}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|}{|x-y|}\left(\frac{|x-y|}{\varepsilon}\right)^p\bar{\eta}_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y\right)^p \leqslant C\|u_n\|_\infty^p \mathcal{GMS}_{\varepsilon,n}(u_n)^p \tag{5.12}$$

for a universal constant $C>0$. For the remaining part we notice the following thing. On $(\Omega\times\Omega)\backslash A_\varepsilon$ it holds

$$\zeta\left(\varepsilon^{1-p+q}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|^p}{|T_n(x)-T_n(y)|^q}\right) \geqslant \varepsilon\frac{\zeta'(0)}{2}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|^p}{|x-y|^p}\frac{\varepsilon^q}{|T_n(x)-T_n(y)|^q}\left(\frac{|x-y|}{\varepsilon}\right)^p$$

and for ε small enough we have

$$\frac{\varepsilon^q}{|T_n(x)-T_n(y)|^q} \geqslant \frac{1}{\left(\frac{|x-y|}{\varepsilon}+\frac{2\ell}{\varepsilon}\right)^q} \geqslant \frac{1}{\left(\frac{|x-y|}{\varepsilon}+1\right)^q}.$$

This yields, by recalling that $\bar{\eta}_\varepsilon(|x-y|)=0$ for $|x-y|\geqslant(1-r)\varepsilon$, that

$$\begin{aligned}\mathcal{GMS}_{\varepsilon,n}(u_n) &\geqslant \frac{\zeta'(0)}{2}\int_{(\Omega\times\Omega)\backslash A_\varepsilon}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|^p}{|x-y|^p}\frac{\left(\frac{|x-y|}{\varepsilon}\right)^p}{\left(\frac{|x-y|}{\varepsilon}+1\right)^q}\bar{\eta}_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y\\ &\geqslant \Lambda^{1-p}C\left(\int_{(\Omega\times\Omega)\backslash A_\varepsilon}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|}{|x-y|}\left(\frac{|x-y|}{\varepsilon}\right)^p\bar{\eta}_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y\right)^p\end{aligned} \tag{5.13}$$

where

$$\begin{aligned}\Lambda &:= \int_{(\Omega\times\Omega)\backslash A_\varepsilon}\left(\frac{|x-y|}{\varepsilon}\right)^p\bar{\eta}_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y \leqslant \int_{\Omega\times\Omega}\left(\frac{|x-y|}{\varepsilon}\right)^p\bar{\eta}_\varepsilon(|x-y|)\mathrm{d}x\mathrm{d}y\\ &\leqslant 2|\Omega|\int_{\mathbb{R}^d}|\xi|^p\bar{\eta}(|\xi|)\mathrm{d}\xi \leqslant 2|\Omega|\int_{B_1(0)}|\xi|^p\bar{\eta}(|\xi|)\mathrm{d}\xi < +\infty.\end{aligned}$$

By collecting (5.12) and (5.13) we conclude that

$$\sup_{n\in\mathbb{N}}\left\{\int_{\Omega\times\Omega}\frac{|\tilde{u}_n(x)-\tilde{u}_n(y)|}{|x-y|}\left(\frac{|x-y|}{\varepsilon_n}\right)^p\bar{\eta}_{\varepsilon_n}(|x-y|)\mathrm{d}x\mathrm{d}y+\|\tilde{u}_n\|_\infty\right\} < +\infty. \tag{5.14}$$

We now divide the proof in three steps.

Step one: Ω *has* C^2 *boundary*. In this case, by applying lemma 5.3 we can find a sequence of $\{v_n\}_{n\in\mathbb{N}} \subset L^1(\mathbb{R}^d)$ such that $v_n = \tilde{u}_n$ $\mathcal{L}^d$-a.e. on Ω. Moreover, due to (5.14) there exists a kernel $J^{\bar{\eta}}$ such that $\{\xi : J^{\bar{\eta}}(|\xi|) \geqslant c\}$ has not empty interior for some $c > 0$ and for which

$$\sup_{n\in\mathbb{N}} \left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|v_n(x) - v_n(y)|}{|x-y|} J^{\bar{\eta}}_{\varepsilon_n}(|x-y|)\mathrm{d}x\mathrm{d}y + \|v_n\|_\infty \right\} < +\infty.$$

Then, by applying proposition 5.2 we deduce that $\{v_n\}_{n\in\mathbb{N}}$ is compact in $L^1(\mathbb{R}^d)$ and thus $\{\tilde{u}_n\}_{n\in\mathbb{N}}$ is compact in $L^1(\Omega)$.

Step two: Ω *has Lipschitz boundary*. Thanks to [BZ17, remark 5.3] there exists a bi-Lipscthiz map $\Psi : \hat{\Omega} \to \Omega$ where $\hat{\Omega}$ is a domain with smooth boundary. Consider $\hat{u}_n \coloneqq \tilde{u}_n \circ \Psi : \hat{\Omega} \to \mathbb{R}$. Clearly $\|\hat{u}_n\|_\infty \leqslant \|\tilde{u}_n\|_\infty$. Moreover

$$\begin{aligned}
&\int_{\hat{\Omega}\times\hat{\Omega}} \frac{|\hat{u}_n(x) - \hat{u}_n(y)|}{|x-y|} \left(\frac{|x-y|}{\varepsilon_n}\right)^p \bar{\eta}_\varepsilon(\mathrm{Lip}(\Psi)^{-1}|x-y|)\mathrm{d}x\mathrm{d}y \\
&\quad\leqslant C\int_{\Omega\times\Omega} \frac{|\tilde{u}_n(x) - \tilde{u}_n(y)|}{|\Psi(x) - \Psi(y)|} \left(\frac{|\Psi(x) - \Psi(y)|}{\varepsilon_n}\right)^p \bar{\eta}_{\varepsilon_n}(\mathrm{Lip}(\Psi)^{-1}|\Psi(x) - \Psi(y)|)\mathrm{d}x\mathrm{d}y \\
&\quad\leqslant C\int_{\Omega\times\Omega} \frac{|\tilde{u}_n(x) - \tilde{u}_n(y)|}{|x-y|} \left(\frac{|x-y|}{\varepsilon_n}\right)^p \bar{\eta}_{\varepsilon_n}(|x-y|)\mathrm{d}x\mathrm{d}y.
\end{aligned}$$

By exploiting (5.14), lemma 5.3 and by arguing as in Step one we conclude that $\{\hat{u}_n\}_{n\in\mathbb{N}}$ is compact in $L^1(\hat{\Omega})$. Since Ψ is bi-Lipschitz, a simple change of variable shows that $\{\tilde{u}_n\}_{n\in\mathbb{N}}$ is compact in $L^1(\Omega)$.

Step three: *compactness of* (μ_n, u_n) *in* TL^1. Thanks to steps one and two we obtained that $\{\tilde{u}_n \coloneqq u_n \circ T_n\}_{n\in\mathbb{N}}$ is compact in $L^1(\Omega)$ and converges to some u up to a subsequence. Thanks to lemma 3.2 we deduce that $(\tilde{\mu}_n, u_n \circ T_n) \to (\mu, u)$ in TL^1 as well. In particular, since $d_\infty(\mu_n, \tilde{\mu}_n) \to 0$, we have

$$\begin{aligned}
d_{TL^1}((\mu_n, u_n), (\mu, u)) &\leqslant d_{TL^1}((\mu_n, u_n), (\tilde{\mu}_n, u_n \circ T_n) + d_{TL^1}((\tilde{\mu}_n, u_n \circ T_n), (\mu, u)) \\
&= d_\infty(\mu_n, \tilde{\mu}_n) + d_{TL^1}((\tilde{\mu}_n, u_n \circ T_n), (\mu, u)) \to 0.
\end{aligned}$$

□

6. Proofs of corollaries

We now prove the corollary 2.6.

Proof. Let $u_n : V_n \to \mathbb{R}$ be a sequence of functions such that (μ_n, u_n) converges in TL^2 towards (μ, u). Let $\overline{T}_n$ be the ∞-optimal transport map between μ and μ_n. To show the lim-inf inequality needed for Γ-convergence, it suffices to establish the convergence of the fidelity term, as the other terms are same as in theorem 2.1. Note that

$$\frac{1}{n}\sum_{i=1}^n |u_n(x_i) - f(x_i)|^2 = \int_\Omega |u_n \circ \overline{T}_n - f \circ \overline{T}_n|^2 \rho(x)\,\mathrm{d}x.$$

We claim that since f is piecewise continuous $f \circ \overline{T}_n$ converges to f in $L^2(\mu)$. Namely let J_f be the set of discontinuities of f and let $\tilde{J}_f = J_f \cup \partial\Omega$. Let $\tilde{J}_{f,\delta} = \{x \in \Omega : d(x, \tilde{J}_f) < \delta\}$. Since $\mathcal{H}^{d-1}(\tilde{J}_f \cup \partial\Omega) < \infty$, $\mu(\tilde{J}_{f,\delta}) \to 0$ as $\delta \to 0$. To establish the convergence let $\varepsilon > 0$. Let δ be such that $4\|f\|_{L^\infty}\mu(\tilde{J}_{f,\delta}) < \varepsilon$. Let n be so large that $\|\overline{T}_n - I\|_{L^\infty} < \frac{1}{2}\delta$. Since f is uniformly continuous on $\Omega\backslash\tilde{J}_{f,\delta/2}$ and $\|\overline{T}_n - I\|_{L^\infty} \to 0$ as $n \to \infty$, $f \circ \overline{T}_n$ converges uniformly to f on $\Omega\backslash\tilde{J}_{f,\delta}$. Therefore for all n large enough

$$\int_\Omega |f - f\circ \overline{T}_n|^2\rho(x)\,\mathrm{d}x \leqslant 2\mu(\tilde{J}_{f,\delta}|)\|f\|_{L^\infty} + \int_{\Omega\setminus\tilde{J}_{f,\delta}} |f - f\circ\overline{T}_n|^2\rho(x)\,\mathrm{d}x < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.$$

Since $(\mu_n, u_n) \to (\mu, u)$ in TL^2, and $\overline{T}_n$ is a stagnating sequence of transport maps we conclude that we conclude that $u_n \circ \overline{T}_n \to u$ in $L^2(\mu)$. Combining with the convergence for $f\circ\overline{T}_n$ obtained above we conclude that $\int_\Omega |u_n\circ\overline{T}_n - f\circ\overline{T}_n|^2\rho(x)\,\mathrm{d}x \to \int_\Omega |u-f|^2\rho(x)\,\mathrm{d}x$ as $n\to\infty$.

Establishing the limsup inequality is straightforward by using the same approximation argument and recovery sequence as in the proof of theorem 2.1.

To establish the compactness of the sequence of minimizers, let u_n be a minimizer of $\mathcal{GMS}_{f,\varepsilon,n}$. By truncation it is immediate that $\|u_n\|_{L^\infty(\mu_n)} \leqslant \|f\|_{L^\infty(\mu)}$. Therefore the compactness claim of the theorem 2.3 implies that (μ_n, u_n) converges along a subsequence in TL^1 to (μ, u) for some $u \in L^\infty(\mu)$. The boundedness in L^∞ implies, via interpolation, that the convergence is in TL^2. The fact that u is a minimizer follows from Γ-convergence. □

We now prove the corollary 2.7.

Proof. Let us first establish the liminf inequality. Let $u_n : V_n \to \mathbb{R}$ be a sequence of functions such that (μ_n, u_n) converges in TL^2 towards (μ, u). Given the results of theorem 2.1 and assumptions on β, it suffices to show that the fidelity term converges, that is that

$$\frac{1}{n}\sum_{i=1}^n |u_n(x_i) - f(x_i) - y_i|^2 \to \iint_{\Omega\times\mathbb{R}} |u(x) - f(x) - y|^2\rho(x)\,\mathrm{d}x\mathrm{d}\beta(y) \tag{6.1}$$

as $n\to\infty$. Note that $\frac{1}{n}\sum_{i=1}^n |u_n(x_i) - f(x_i)|^2 \to \int_\Omega |u(x)-f(x)|^2\rho(x)\,\mathrm{d}x$ follows from the proof of corollary 2.6. The fact that $\frac{1}{n}\sum_{i=1}^n \delta_{y_i}$ weakly converges to β follows from Glivenko–Cantelli lemma. Due to boundedness of moments we conclude that $\frac{1}{n}\sum_{i=1}^n \delta_{y_i}$ converges to β is $q-$Wasserstein distance for all $q \geqslant 1$. Using the boundedness of β we conclude that $\frac{1}{n}\sum_{i=1}^n |y_i|^2$ converges to $\int_{\mathbb{R}} y^2 d\beta(y) = \iint_{\Omega\times\mathbb{R}} y^2\rho(x)\,\mathrm{d}x\mathrm{d}\beta(y)$ since $\int_\Omega \rho(x)\mathrm{d}x = 1$. Note that $\gamma_n := \frac{1}{n}\sum_{i=1}^n \delta_{(x_i,y_i)}$ converges to $\gamma := \mu\times\beta$ in Wasserstein distance, again due to Glivenko–Cantelli lemma. Let κ_n be the optimal transport plan for the Wasserstain distance between γ and γ_n. Let $\pi_n := \Pi_{1,3\sharp}\kappa_n$ where $\Pi_{1,3}$ is the projection to the first and the third variable. By definition π_n is a stagnating sequence of transport plans. Therefore, by proposition 3.12 of [GTS16], since $(\mu_n, u_n - f)TL^2(\mu, u-f)$ by the proof of corollary 2.6, $\iiiint |x-\tilde{x}|^2 + |u_n(\tilde{x}) - f(\tilde{x}) - u(x) + f(x)|^2\mathrm{d}\kappa_n(x,y,\tilde{x},\tilde{y}) \to 0$ as $n\to\infty$. Thus $(\gamma_n, u_n - f)TL^2(\gamma, u-f)$ as $n\to\infty$. Similarly $(\gamma_n, y) \xrightarrow{TL^2} (\gamma, y)$ as $n\to\infty$. Consequently $(\gamma_n, (u_n - f)y) \xrightarrow{TL^1} (\gamma, (u-f)y)$ as $n\to\infty$. Thus

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n (u_n(x_i) - f(x_i))y_i - \iint_{\Omega\times\mathbb{R}} (u(x) - f(x))y\,\rho(x)\,\mathrm{d}x\mathrm{d}\beta(y)\\ = \iiiint (u_n(\tilde{x}) - f(\tilde{x}))\tilde{y} - (u(x) - f(x))y\,\mathrm{d}\kappa_n(x,y,\tilde{x},\tilde{y}) \to 0 \quad \text{as } n\to\infty.\end{aligned}$$

Combining with the limits above establishes (6.1).

The proofs of limsup inequality, compactness and the converge of minimizers are as before. □

7. Numerical algorithm and experiments

Here we desribe an efficient numerical algorithm for computing the, approximate, minimizers of the graph Mumford–Shah functional and illustrate its behavior on a real world set of data. We note that similar functionals were minimized using the ADMM algorithm by Hallac *et al* by [HLB15]. Here we minimize (1.1), where ζ is non convex. We use a standard ;iterated reweighted least square' approach which in this context dates back at least to [GR92] (*cf* also the implementation in [Cha99]). In our case, the idea is to perform several iteration, linearizing each time the problem with respect to ζ around the previous value.

This can be presented as follows: we assume that ζ is concave, with $\zeta'(0) = 1$ and $\zeta(+\infty) = 1$, for instance $\zeta(t) = t/(1+t)$ for $t > 0$. Then, using the Legendre transform, one can write for $t > 0 \zeta(t) = \min_{z\in[0,1]} zt + \Psi(z)$ for some convex function Ψ. (One has $\zeta(t) = -\Psi^*(-t)$ and $\Psi(z) = \max_t \zeta(t) - tz$, where Ψ^* denotes the classical convex conjugate of Ψ.) The minimum (if unconstrained) is reached at z which solves $t + \Psi'(z) = 0$, hence $z = (\Psi^*)'(-t) = \zeta'(t)$.

We consider the edge weights given by kernel $\eta(s) = \mathrm{e}^{-\frac{s^2}{2\sigma^2}}$ where σ is a parameter that can be tuned. Minimizing (1.1) is equivalent to solving:

$$\min_{u,z} \sum_{i=1}^{n} |u(x_i) - f_i|^2 + \frac{1}{\lambda\varepsilon n} \sum_{i,j=1}^{n} \left(z_{i,j} \frac{1}{\varepsilon} |u(x_i) - u(x_j)|^2 + \Psi(z_{i,j}) \right) \frac{1}{\varepsilon^d} \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2\varepsilon^2} \right) \tag{7.1}$$

where the new variable $(z_{i,j})$ is defined on the active edges. This is computed by alternatively minimizing the problem with respect to u and z: in u, the problem is quadratic and can be minimized efficiently, depending on the graph, by inverting the graph Laplacian (plus identity) or a conjugate gradient method. In z, the solution is explicitly given by

$$z_{i,j} = \zeta'\left(\frac{|u(x_i) - u(x_j)|^2}{\varepsilon} \right).$$

In practice, we have implemented the following cases:

$$\zeta(t) = \frac{2}{\pi} \arctan\left(\frac{\pi t}{2}\right), \qquad \zeta'(t) = \frac{1}{1 + \frac{\pi^2 t^2}{4}} \tag{7.2}$$

$$\zeta(t) = \sqrt{\delta^2 + t}, \qquad \zeta'(t) = \frac{1}{2\sqrt{\delta^2 + t}}, \tag{7.3}$$

$$\zeta(t) = t, \qquad \zeta'(t) = 1, \tag{7.4}$$

The choice (7.3) leads as $\delta \to 0$ to a consistent approximation of the graph total variation which was first proposed in [VO96]. The choice $\zeta(t) = t$ corresponds to regularization by Dirichlet energy, which corresponds to (unnormalized) graph Laplacian. Our implementation is available on bitbucket: https://bitbucket.org/AntoninCham/ms_on_graphs/

7.1. A synthetic example

We consider denoising and detecting edges in the signal given by piecewise linear function u on domain $[0,1]^2$, shown on figure 1(a). The signal is sampled at 10 000 points, X_{10000}, and corrupted by Gaussian noise with variance 0.2. We build the graph using η as in (7.1) with $\sigma = 5$ and $\varepsilon = 0.0225$ and with the maximum number of neighbors $k = 8$. We considered three

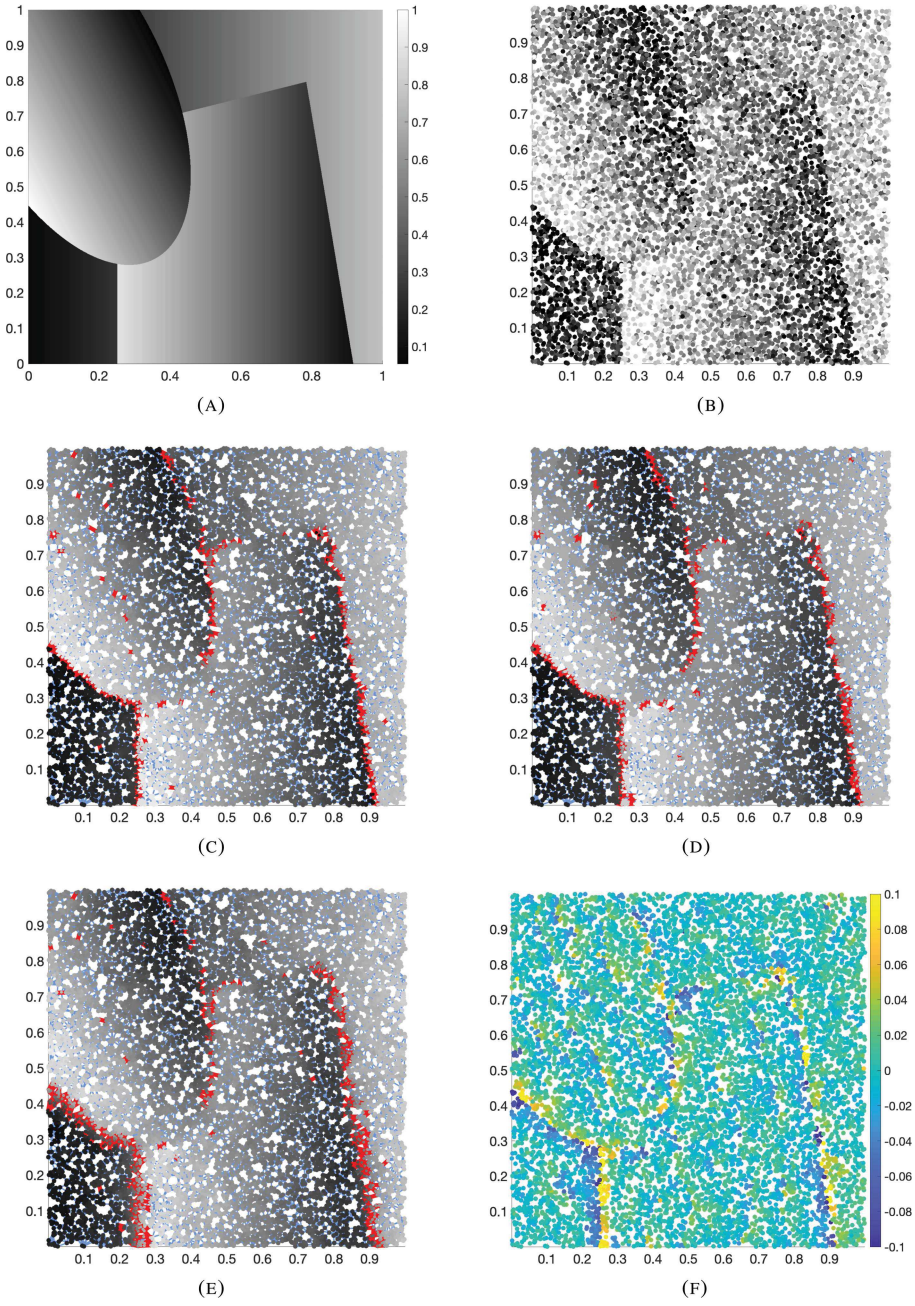

Figure 1. Denoising (regression) and edge detection. (A) Noiseless function. (B) Function sampled at 10 000 random points and corrupted by Gaussian noise with $\sigma = 0.2$. (C) Minimizer u_{MS} of (1.1) for $\lambda = 162$. Edges with jump over 0.075 are red. (D) Minimizer u_{TV} of the graph TV functional for $\lambda = 438$. Edges with jump over 0.14 are red. (E) Minimizer u_L for ζ given by (7.4) with $\lambda = 248$. Edges with jump over 0.09 are red. (F) The difference $u_{\mathrm{MS}} - u_{\mathrm{TV}}$ tends to be positive on the upper side of jumps and negative on the lower side.

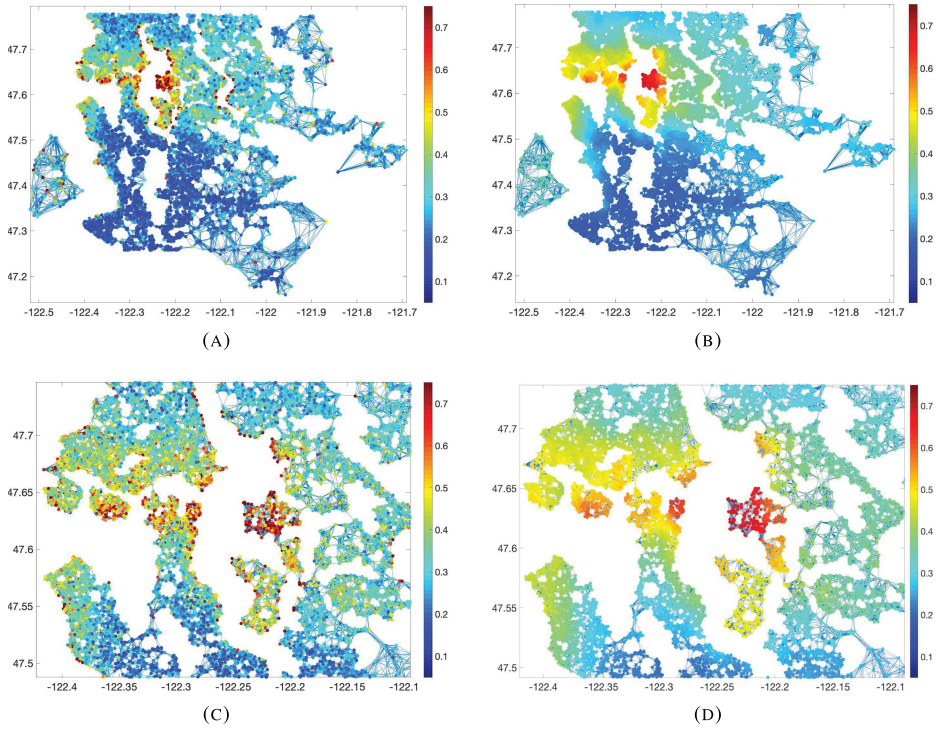

Figure 2. Denoising of housing prices. The maximum price per square foot. (A) The recorded price per square foot. (B) The minimizer of the graph Mumford–Shah functional computed. (C) Detail of the image above. (D) Detail of the image above.

models for denoising and edge detection given by ζ in (7.2)–(7.4). Namely on figure 1(c) we display the computed minimizer of the graph Mumford–Shah for ζ given by (7.2) and $\lambda = 162$. On figure 1(d) we display the minimizer of the approximation of the graph TV functional for ζ given by (7.3) with $\delta = 0.001$ and $\lambda = 438$. On figure 1(e) we show the minimizer of the functional with dirichlet regularization, corresponding to ζ given by (7.4) and $\lambda = 248$. For comparison, for each of the models we display the result for parameter λ which minimizes the $L^1(\mu_{10000})$ error between the minimizer and the clean signal u restricted to X_{10000}. The errors observed for optimal lambdas were $\|u_{MS} - u\|_{L^1} = 0.0258$, $\|u_{MS} - u\|_{L^1} = 0.0297$, and $\|u_L - u\|_{L^1} = 0.0392$. We note that the recovery by Mumford–Shah is somewhat better than for graph TV. We think that the main reason is that the graph TV tends to decrease contrast (as is well known in image processing, see page 30 of [CCC+10]) while the Mumford–Shah does not have this bias.

We also observe to what extent the minimizers recover the edges of the domains by labeling the graph edges that have a relatively large difference between values at the nodes. These are shown in red on the plots. The critical jump size was set manually for visually the best results for each model. We note that Mumford–Shah and TV give similar results, while the Laplacian smoothing blurs the edges as expected. Taking the difference between the minimizers $u_{\mathrm{MS}} - u_{\mathrm{TV}}$ shows, on figure 1(f) that indeed jumps across the edges are typically larger for the Mumford–Shah minimizers that for total-variation regularization.

7.2. Denoising housing prices

Here we present an example of minimization of Mumford–Shah functionals on graphs arising from real-world data samples. This example is given as an illustration of the nature of minimizers.

We consider denoising the real estate prices in King County, WA. The housing prices in the period May 2014 to May 2015 are obtained from the Kaggle website: https://kaggle.com/harlfoxem/housesalesprediction.

We removed the geographical outliers (east of longitude $-121.68°$) and data rows missing square footage. The recorded price per square foot is shown on the left. This left 21 594 usable records. The maximum price per square foot was \$810.14. We normalized the input prices per square foot by dividing by the maximal price. On figure 2 we present the computed minimizers of the graph Mumford–Shah functional with $\varepsilon = 0.04$, $\lambda = 14$, and $\sigma = 1$. We also allow one to limit the maximal degree of a vertex considered, which we set to $k = 15$. On a 2018 Macbook Pro, the computation takes 31 s, including the construction of the graph. The denoised data allow one to visualize by how much the typical price per square foot depends on the location.

Acknowledgments

The collaboration was supported by European Commission H2020 grant number 777826 (NoMADS). The work of Caroccia has been partially supported by Fundação para a Ciência e Tecnologia through the Carnegie Mellon–Portugal Program under Grant 18316.1.5004440. Part of the work has been completed under the grant "Calcolo delle variazioni, Equazione alle derivate parziali, Teoria geometrica della misura, Trasporto ottimo" co-founded by Scuola Normale Superiore and the University of Florence. Chambolle is grateful to CNRS for support. Slepcev is thankful to the NSF for its support (grants DMS-1516677 and DMS-1814991). He is also grateful to the Center for Nonlinear Analysis of CMU.

Appendix A. Proof of lemma 4.4

The proof of lemma 4.4 is obtained as a slight modification of the proof contained in [GM01] for $f = 1$. In particular we apply [GM01, theorems 3.1 and 3.2] on the family of functions

$$\varphi_\varepsilon(r) := \frac{|\xi|}{\varepsilon}\zeta\left(\frac{\varepsilon r^p}{|\xi|^{q-p+1}}\right).$$

Indeed we note the following facts

(a) $\varphi_{\delta|\xi|}\left(\frac{|u(x+\delta|\xi|)-u(x)|}{\delta|\xi|}\right) = \frac{1}{\delta}\zeta\left(\frac{|u(x+\delta|\xi|)-u(x)|^p}{\delta^{p-1}|\xi|^q}\right)$;
(b) $\lim_{\varepsilon\to 0+}\varphi_\varepsilon(r) = \zeta'(0)|\xi|^{p-q}r^p$;
(c) $\lim_{\varepsilon\to 0+}\varepsilon\varphi_\varepsilon(r/\varepsilon) = \Theta|\xi|$.

In particular, by combining (a)–(c) and with a slight variation of the proof of [GM01, theorem 3.1] (as in [Cha99, section 3.2]) we conclude that

$$\liminf_{\delta\to 0+}\frac{1}{\delta}\int_A \zeta\left(\frac{|u(x+\delta|\xi|)-u(x)|^p}{\delta^{p-1}|\xi|^q}\right) \geqslant \zeta'(0)|\xi|^{p-q}\int_A |u'(x)|^p \mathrm{d}x + \Theta|\xi|\mathcal{H}^0(S_u\cap A). \tag{A.1}$$

We now proceed to the proof of lemma 4.4.

Proof of lemma 4.4. For every $k \in \mathbb{N}$ consider a partition of A in small intervals $\{I_j^k\}_{j=1}^{|A|/k}$. Then

$$\frac{1}{\delta}\int_A \zeta\left(\frac{|u(x+\delta|\xi|)-u(x)|^p}{\delta^{p-1}|\xi|^q}\right) f(x)\mathrm{d}x \geqslant \sum_{j=1}^{|A|/k} \min_{I_j^k}\{f\}\frac{1}{\delta}\int_A \zeta\left(\frac{|u(x+\delta|\xi|)-u(x)|^p}{\delta^{p-1}|\xi|^q}\right)\mathrm{d}x.$$

In particular, by applying (A.1) on each intervals I_j^k we reach

$$\begin{aligned}\liminf_{\delta\to 0}\frac{1}{\delta}\int_A \zeta\left(\frac{|u(x+\delta|\xi|)-u(x)|^p}{\delta^{p-1}|\xi|^q}\right) f(x)\mathrm{d}x \geqslant \zeta'(0)|\xi|^{p-q}\sum_{j=1}^{|A|/k}\int_{I_j^k}|u'(x)|^p \min_{I_j^k}\{f\}\mathrm{d}x \\ + \sum_{j=1}^{|A|/k}\Theta|\xi|\int_{S_u\cap I_j^k}\min_{I_j^k}\{f\}\mathrm{d}\mathcal{H}^0(y).\end{aligned}$$

Since f is a Lipschitz function we now notice that, given $\varepsilon > 0$, we can find ρ such that

$$|x-y| < \rho \Rightarrow |f(x)-f(y)| < \varepsilon.$$

In particular, for any fixed $\varepsilon > 0$, we can find a $k \in \mathbb{N}$ big enough such that

$$\min_{I_j^k}\{f\} \geqslant \max_{I_j^k}\{f\} - \varepsilon \geqslant f(x) - \varepsilon \quad \text{for all } x \in I_j^k$$

Thus,

$$\begin{aligned}\liminf_{\delta\to 0}\frac{1}{\delta}\int_A \zeta\left(\frac{|u(x+\delta|\xi|)-u(x)|^p}{\delta^{p-1}|\xi|^q}\right) f(x)\mathrm{d}x &\geqslant \zeta'(0)|\xi|^{p-q}\sum_{j=1}^{|A|/k}\int_{I_j^k}|u'(x)|^p \min_{I_j^k} f(x)\mathrm{d}x \\ &\quad + \sum_{j=1}^{|A|/k}\Theta|\xi|\int_{S_u\cap I_j^k}\min_{I_j^k} f(x)\mathrm{d}\mathcal{H}^0(y) \\ &\geqslant \zeta'(0)|\xi|^{p-q}\int_A |u'(x)|^p(f(x)-\varepsilon)\mathrm{d}x \\ &\quad + \Theta|\xi|\int_{S_u\cap A}(f(y)-\varepsilon)\mathrm{d}\mathcal{H}^0(y).\end{aligned}$$

Since the above holds for arbitrarily small positive ε, we conclude that (4.11) holds. □

Appendix B. Proof of lemma 4.9

Proof. Note that for all $x \in \Omega\backslash D(\varepsilon_n, \ell_n)$ it holds that

(a) $\{x + s(T_n(x) - x) : s \in [0,1]\} \cap S_u = \emptyset$.
(b) $\{x + \varepsilon_n\xi + s(T_n(x+\varepsilon_n\xi) - x + \varepsilon\xi) : s \in [0,1]\} \cap S_u = \emptyset$.

Indeed, assume by contradiction that $x + t_0(T_n(x) - x) \in S_u$. Then $d(x, S_u) \leqslant t_0\|T_n(x) - x\| \leqslant t_0\ell_n$ which would imply $x \in (S_u)_{\ell_n}$. Thus (a) holds. Analogously assume that for some $t_0 \in [0,1]$ we have $x + \varepsilon_n\xi + t_0(T_n(x+\varepsilon_n\xi) - x + \varepsilon_n\xi) \in S_u$. Then

$$d(x, S_u - \varepsilon_n\xi) \leqslant \|x - (x + \varepsilon_n\xi + t_0(T_n(x + \varepsilon_n\xi) - x + \varepsilon_n\xi) - \varepsilon_n\xi)\|$$
$$= t_0\|(T_n(x + \varepsilon_n\xi) - x + \varepsilon_n\xi)\| \leqslant \ell_n$$

again contradicting $x \in \Omega\backslash D(\varepsilon_n, \ell_n)$. In particular

$$|u(T_n(x + \varepsilon\xi)) - u(T_n(x))| \leqslant |u(x + \varepsilon\xi) - u(x)| + |u(T_n(x + \varepsilon\xi)) - u(x + \varepsilon\xi)| + |u(T_n(x)) - u(x)|$$

and, since u is regular outside S_u,

$$|u(T_n(x + \varepsilon\xi)) - u(x + \varepsilon\xi)| \leqslant \ell_n \int_0^1 |\nabla u((x + \varepsilon\xi)s + (1 - s)T_n(x + \varepsilon\xi))|\mathrm{d}s$$
$$|u(T_n(x)) - u(x)| \leqslant \ell_n \int_0^1 |\nabla u(xs + (1 - s)T_n(x)))|\mathrm{d}s,$$

which is proving (4.20). In order to prove (4.21) we just notice that

$$|(Su - \varepsilon_n\xi)_{\ell_n}| = |(Su)_{\ell_n}|$$

and that, since S_u is a ployhedral set, for big n

$$|(Su)_{\ell_n}| = 2\ell_n\mathcal{H}^{d-1}(S_u) + o(\ell_n).$$

This, combined with (2.13), implies (4.21). □

ORCID iDs

Marco Caroccia https://orcid.org/0000-0002-9200-8856
Antonin Chambolle https://orcid.org/0000-0002-9465-4659
Dejan Slepčev https://orcid.org/0000-0002-7600-1144

References

[AFP00] Ambrosio L, Fusco N and Pallara D 2000 *Functions of Bounded Variation and Free Discontinuity Problems* vol 254 (Oxford: Clarendon)
[Amb89] Ambrosio L 1989 A compactness theorem for a new class of functions of bounded variation *Boll. della Unione Mat. Ital.* **3** 857–81
[BDM97] Braides A and Dal Maso G 1997 Non-local approximation of the Mumford-Shah functional *Calc. Var. Partial Differ. Equ.* **5** 293–322
[Ber24] Bernstein S 1924 On a modification of Chebyshev's inequality and of the error formula of Laplace *Ann. Sci. Inst. Sav. Ukraine, Sect. Math* **1** 38–49
[Bre87] Brenier Y 1987 Décomposition polaire et réarrangement monotone des champs de vecteurs *C. R. Acad. Sci., Paris* I **305** 805–8
[BZ17] Ball J M and Zarnescu A 2017 Partial regularity and smooth topology-preserving approximations of rough domains *Calc. Var. Partial Differ. Equ.* **56** 13
[CCC+10] Chambolle A, Caselles V, Cremers D, Novaga M and Pock T 2010 An introduction to total variation for image analysis *Theoretical Foundations and Numerical Methods for Sparse Recovery* (Radon Ser. Comput. Appl. Math. vol 9) (Berlin: Walter de Gruyte) pp 263–340
[CDPJ08] Champion T, De Pascale L and Juutinen P 2008 The ∞-Wasserstein distance: local solutions and existence of optimal transport maps *SIAM J. Math. Anal.* **40** 1–20
[Cha99] Chambolle A 1999 Finite-differences discretizations of the Mumford-Shah functional *ESAIM: Math. Modelling Numer. Anal.* **33** 261–88

[CSMK15] Chen S, Sandryhaila A, Moura J M F and Kovačević J 2015 Signal recovery on graphs: variation minimization *IEEE Trans. Signal Process.* **63** 4609–24
[CT99] Guido C and Toader R 1999 A density result in sbv with respect to non-isotropic energies *Nonlinear Anal. Theory Methods Appl.* **38** 585–604
[CV01] Chan T F and Vese L A 2001 Active contours without edges *IEEE Trans. Image Process.* **10** 266–77
[GM96] Gangbo W and McCann R J 1996 The geometry of optimal transportation *Acta Math.* **177** 113–61
[GM01] Gobbino M and Mora M G 2001 Finite-difference approximation of free-discontinuity problems *Proc. R. Soc. Edinburgh* A **131** 567–95
[Gob98] Gobbino M 1998 Finite difference approximation of the Mumford-Shah functional *Commun. Pure Appl. Math.* **51** 197–228
[GR92] Geman D and Reynolds G 1992 Constrained restoration and the recovery of discontinuities *IEEE Trans. Pattern Anal. Mach. Intell.* **14** 367–83
[GTGHS19] García Trillos N, Gerlach M, Hein M and Slepčev D 2019 Error estimates for spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace-Beltrami operator *Found. Comput. Math.* (https://doi.org/10.1007/s10208-019-09436-w)
[GTM17] Garcia Trillos N and Murray R 2017 A new analytical approach to consistency and overfitting in regularized empirical risk minimization *Eur. J. Appl. Math.* **28** 886–921
[GTS15] Garcia Trillos N and Slepčev D 2015 On the rate of convergence of empirical measures in ∞-transportation distance *Can. J. Math.* **67** 1358–83
[GTS16] Garcia Trillos N and Slepčev D 2016 Continuum limit of total variation on point clouds *Arch. Ration. Mech. Anal.* **220** 193–241
[GTS18] Garcia Trillos N and Slepčev D 2018 A variational approach to the consistency of spectral clustering *Appl. Comput. Harmon. Anal.* **45** 239–81
[HLB15] Hallac D, Leskovec J and Boyd S 2015 Network lasso: clustering and optimization in large graphs Proc. of the 21th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD '15 (New York: ACM) pp 387–96
[HLPB13] Hu H, Laurent T, Porter M A and Bertozzi A L 2013 A method based on total variation for network modularity optimization using the MBO scheme *SIAM J. Appl. Math.* **73** 2224–46
[HR16] Hütter J-C and Rigollet P 2016 Optimal rates for total variation denoising Conf. on Learning Theory pp 1115–46
[MP18] Müller T and Penrose M D 2018 Optimal cheeger cuts and bisections of random geometric graphs (arXiv:1805.08669)
[MS89] Mumford D and Shah J 1989 Optimal approximations by piecewise smooth functions and associated variational problems *Commun. Pure Appl. Math.* **42** 577–685
[MvdG97] Mammen E and van de Geer S 1997 Locally adaptive regression splines *Ann. Stat.* **25** 387–413
[PSCW18] Padilla O H M, James S, Chen Y and Daniela M W 2018 Adaptive non-parametric regression with the k-nn fused lasso (arXiv:1807.11641)
[ROF92] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
[Ruf17] Ruf M 2017 Discrete stochastic approximations of the Mumford-Shah functional (arXiv:1710.05571)
[ST19] Slepčev D and Thorpe M 2019 Analysis of p-Laplacian regularization in semi-supervised learning *SIAM J. Math. Anal.* **51** 2085–2120
[TSR+05] Tibshirani R, Saunders M, Rosset S, Zhu J and Knight K 2005 Sparsity and smoothness via the fused lasso *J. Roy. Stat. Soc. B Stat. Methodol.* **67** 91–108
[VC02] Vese L A and Chan T F 2002 A multiphase level set framework for image segmentation using the Mumford and Shah model *Int. J. Comput. Vis.* **50** 271–93
[VO96] Vogel C R and Oman M E 1996 Iterative methods for total variation denoising *SIAM J. Sci. Comput.* **17** 227–38
[WSST16] Wang Y-X, James S, Smola A J and Tibshirani R J 2016 Trend filtering on graphs *J. Mach. Learn. Res.* **17** 1–41